University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Michał D. Lemańczyk

Recurrence of stochastic processes in some concentration of measure and entropy problems

PhD dissertation

Supervisor
dr hab. Joanna Kułaga-Przymus, prof. UMK
Faculty of Mathematics and Computer Science
Nicolaus Copernicus University in Toruń

December 2020

Author's declaration:
I hereby declare that this dissertation is my own work.

December 2020

Michał Lemańczyk

Michał D. Lemańczyk

Supervisor's declaration:
This dissertation is ready to be reviewed.

December 2020

Joanna Kułaga-Przymus

dr hab. Joanna Kułaga-Przymus, prof. UMK

Contents

Part I

Introduction

This thesis is about the role played by return time processes in probability theory and in dynamical systems. We show how they allow us to obtain Bernstein-type inequalities for additive functionals of general Markov chains. We demonstrate how they provide a criterion for the inability of the retrieval of a lost signal. Moreover, we explain how they can be used to solve the task of finding the entropy of multiplicative convolution of measures which leads to an explicit formula for the topological pressure of $\mathscr{B}$-free systems. Apart from that we address closely related problems such as the absence of the Gibbs property by the measure of maximal entropy.

The purpose of this part is twofold. Firstly (in Chapter 1), we would like to introduce the reader to the notion of a stochastic process seen from both points of view, the probabilistic and ergodic one. We explain how these perspectives can be combined and used to better grasp the behaviour of random processes. We explain some differences between these approaches as well. Let us add that apart from this basic information given in this part, for the convenience of the reader, in Appendix A, we formulate (and prove) in a probabilistic manner selected standard ergodic theorems concerning stochastic processes. Secondly (in Chapter 2), we present the essence of our results. Beware that in that part, for brevity's sake, some of our theorems are skipped or presented in simplified versions. Moreover, some basic notions may not be explained. Thus, Chapter 2 should be treated as a foretaste of full demonstration made in Part II, where all our results are grouped thematically (and presented in separate chapters) and whole necessary theory is developed.

Let us add that apart from Appendix A mentioned above, at the end of this thesis, we provide some additional supplementary chapters. In Appendix B we consider the theory of tail σ-algebras of processes (moreover, the Pinsker σ-algebra is discussed). In Appendix C we present basic facts concerning Besicovitch and Prokhorov metrics (results from this part are used in the proof of the formula for the topological pressure of $\mathscr{B}$-free systems, see Section 4.2.3). In Appendix D we recall standard facts concerning Orlicz norms (they appear naturally in concentration inequalities, see Chapters 5 and 6). In Appendix E we show how Markov-like properties can be established for the split chain. Although, many results from Appendix E are well-known to specialists, they are hard to find in the literature (especially in the provided form) and we believe that they deserve to be presented with their full proofs (the Markov property of a random block process constitutes one of the main ingredients to the proof of our result concerning Bernstein inequality for Markov chains, see Section 6.2).

This thesis is based on one (submitted) preprint: [A1] (for the summary of our results thence see Section 2.2) and two published articles: [A2, A3] (see Sections 2.3 and 2.1, respectively). Apart from these papers, we generalized the formula for the entropy rate of multiplicative convolution from [A3] to the case in which processes are stationary (see Theorem 3.2.1). Moreover, we established new results concerning a topological pressure of $\mathscr{B}$-free systems (see Section 2.1).

[A1] J. Kułaga-Przymus and M.D. Lemańczyk. Entropy rate of product of independent processes. *Preprint: arXiv:2004.07648*, 2020.

[A2] J. Kułaga-Przymus and M.D. Lemańczyk. Hereditary subshifts whose measure of maximal entropy has no Gibbs property. *To appear in Colloquium Mathematicum, arXiv:2004.07643*, 2020.

[A3] M.D. Lemańczyk. General Bernstein-like inequality for additive functionals of Markov chains. *Journal of Theoretical Probability*, 2020.

Chapter 1

Preliminaries

The aim of this chapter is to explain the concept of a stochastic process, introduce our general mixed probabilistic-ergodic setup and present our notation and conventions.

The organization is as follows. Firstly, we recall some basic facts and definitions from the probability field (Section 1.1). Secondly, we show how to perceive a stochastic process in the light of dynamical systems (Section 1.2). Next, we give a general setup in which we usually work (Section 1.3). At the end we provide some auxiliary notation (Section 1.4).

1.1 Stochastic processes: probabilistic approach

For simplicity's sake, throughout this thesis, we assume that all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where by a ***random variable*** X we mean any measurable function taking values in some measurable **state space** $\mathcal{X}$ (for short, we write $X \in \mathcal{X}$). Moreover, a ***discrete stochastic process*** (for short, a ***process***) $\mathbf{X} = (X_i)_{i\in T}$, where we use only time set $T = \mathbb{N} = \{0, 1, \dots\}$ or $T = \mathbb{Z}$, is just a family of random variables X_i taking values in a **common** state space $\mathcal{X}$. Note that these definitions ensure that every process is a random variable (we always consider the product measurable structure on $\mathcal{X}^T$).

Sometimes, we replace the underlying probability measure $\mathbb{P}$ by its conditioned version, $\mathbb{P}_A(\cdot) = \mathbb{P}(\cdot \cap A)/\mathbb{P}(A)$, where $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, or more generally, by some other probability measure $\mathbb{Q}$. In particular, $\mathbb{E}_A$ and $\mathbb{E}_\mathbb{Q}$ stand for the expectation taken with respect to $\mathbb{P}_A$ or $\mathbb{Q}$ respectively. For convenience's sake, sometimes we write A, B instead of $A \cap B$ for any $A, B \in \mathcal{F}$: for example, $\mathbb{E}_{A,B}$ stands for $\mathbb{E}_{A\cap B}$.

For any random variable $X \in \mathcal{X}$ and underlying measure $\mathbb{Q}$ on Ω we write $\mathcal{L}_\mathbb{Q}(X) = \mu$ for the ***distribution of*** X with respect to $\mathbb{Q}$, that is $\mu(A) = \mathbb{Q}(X \in A)$ holds for all measurable $A \subset \mathcal{X}$. For brevity's sake, $\mathcal{L}(X) = \mathcal{L}_\mathbb{P}(X)$. The expression $X \overset{\mathbb{Q}}{\sim} \mu$ is an equivalent of $\mathcal{L}_\mathbb{Q}(X) = \mu$, whereas for any random variable Y, $X \sim Y$ should be translated as $\mathcal{L}(X) = \mathcal{L}(Y)$. Sometimes we use the expression "under measure $\mathbb{Q}$" to indicate that we consider random variables on the modified space $(\Omega, \mathcal{F}, \mathbb{Q})$ instead of usual $(\Omega, \mathcal{F}, \mathbb{P})$.

Although, many problems in the probability field are stated in terms of a specific process $\mathbf{X}$, in fact, they depend only on the distribution of $\mathbf{X}$. Take for example the well-known task of establishing concentration inequalities for the tails of sums of centered random variables,

$$\mathbb{P}(X_0 + \dots + X_n \geq t),$$

where $t \in \mathbb{R}$ and $X_i \in \mathbb{R}$. Clearly, this problem can be reformulated just in terms of marginals $(X_0, \dots, X_n) \sim \mu_n$ of $\mu \sim \mathbf{X} = (X_i)_{i\in\mathbb{N}}$. Thus, it is justified and natural to say that $\mathbf{X}$ and $\mathbf{Y}$ are "equivalent" if $\mathbf{X} \sim \mathbf{Y}$. A natural question arises: why one should ever bother to use the language of random variables instead that of measures? It turns out that in many situations notions such as independence or coupling of random variables give a powerful insight and intuition which helps to better conceive objects such as Shannon's entropy or Pinsker's σ-algebra.

At the end let us introduce the notion of a **canonical process**. The underlying idea is very simple and is also broadly used in ergodic theory, cf. (1.2.3) below. Firstly, we consider a special case of underlying probability space, namely, $\Omega = \mathcal{X}^{\mathbb{Z}}$ and some probability measure μ on Ω. Secondly, for any $i \in \mathbb{Z}$, we define a random variable X_i to be the projection on the i'th coordinate, that is $X_i(\mathbf{x}) = x_i$, where $\mathbf{x} = (x_i)_{i\in\mathbb{Z}}$. Now, the ***canonical process*** is given by

$$\mathbf{X} = (X_i)_{i\in\mathbb{Z}}.$$

Note that $\mathbf{X} \sim \mu$ under $\mathbb{P} = \mu$ and thus every random process can be realized (in the sense of distribution) as the canonical one.

1.2 Stochastic processes: dynamical approach

In this section we show how stochastic processes are related to the field of dynamical systems. Let us introduce first the basic notions from dynamics. Let $(X, \mathcal{B}, \mu)$ be a standard probability Borel space, $f\colon X \to X$ be an invertible, bi-measurable, μ-preserving map, i.e. $\mu(f^{-1}A) = \mu(A)$ for any $A \in \mathcal{B}$. Then the quadruple $(X, \mathcal{B}, \mu, f)$ is called a ***measure-theoretic dynamical system***. Such systems often arise from ***topological dynamical systems***, i.e. by taking a compact metric space X, with a homeomorphism f. Space X is then equipped with the σ-algebra of Borel subsets $\mathcal{B} = \mathcal{B}(X)$ and the existence of invariant measures on $(X, \mathcal{B})$ follows from the Krylov-Bogolyubov theorem (one can also prove it using the compactness of the space of probability measures on $(X, \mathcal{B})$ considered with the weak*-topology). Sometimes one needs to study non-invertible systems and only assumes that either f is measurable (in the measure-theoretic case) or that f is continuous (in the topological case), skipping the assumption of the invertibility of f.

We are particularly interested in the class of systems which are known as **subshifts**. Let $\mathcal{X}$ be a **measurable** space and let T stand for either $\mathbb{Z}$ or $\mathbb{N}$. The map $S\colon \mathcal{X}^T \to \mathcal{X}^T$ given by

$$S\,(x_i)_{i\in T} = (x_{i+1})_{i\in T} \tag{1.2.1}$$

is called the ***left shift*** and the pair $(\mathcal{X}^T, S)$ is called a ***full shift***. Assume now that $\mathcal{X}$ is a **topological** space and consider the product topology on $\mathcal{X}^T$. This immediately makes S (and its inverse for $T = \mathbb{Z}$) continuous. Most frequently, $\mathcal{X}$ is compact or even finite (sometimes we need to go beyond this setting and consider $\mathcal{X}$ countable). As soon as $\mathcal{X}$ is (at most) countable, we always equip it with the discrete topology. We say that $(\mathscr{X}, S)$ is a ***subshift***, whenever $\mathscr{X} \subset \mathcal{X}^T$ is closed and satisfies $S\mathscr{X} \subset \mathscr{X}$ (if it is clear from the context that we deal with a subshift then sometimes we just write $\mathscr{X}$ for short instead of $(\mathscr{X}, S)$). Notice that as soon as $\mathcal{X}$ is a compact metric space, this results in a topological dynamical system as defined above. Moreover, if μ is an S-invariant measure on $(\mathscr{X}, \mathcal{B})$, where $\mathcal{B}$ is the σ-algebra of Borel subsets, then its topological support $\operatorname{supp}\mu$ is closed and S-invariant and $\mathscr{X} = \operatorname{supp}\mu$ is a subshift. Recall also that S can be interpreted as an operator on the space of measurable functions (as the composition map $Sf = f \circ S$) or on the space of probability measures (as the push-forward map $S\mu(\mathbf{A}) = \mu(S^{-1}\mathbf{A})$). Note that for simplicity's sake, we use the same letter S in all these interpretations instead, e.g. S_* for the push-forward.

Let $\mathbf{X} = (X_i)_{i\in T}$, where $X_i \in \mathcal{X}$, be a stochastic process. Recall that the left shift S acts naturally on processes via

$$S\mathbf{X} = (X_{i+1})_{i\in T},$$

the process $S\mathbf{X}$ is called ***shifted*** and $\mathbf{X}$ is ***stationary*** if $\mathbf{X} \sim S\mathbf{X}$ (most processes which we consider, especially ones arising from a dynamical context – see below – will be stationary and we will say it explicitely if we deal with a non-stationary one). Now, notice that if $\mathbf{X}$ is stationary then $S\mu = \mu$, and process $\mathbf{X}$ induces a measure-theoretic dynamical system

$$\mathcal{Q} = (\mathcal{X}^T, \mathcal{B}, \mu, S). \tag{1.2.2}$$

Can we reverse the above procedure? In fact, every $\mathcal{Q}$ as in (1.2.2) yields a whole **bunch of stationary processes** $\mathbf{Y} = (Y_i)_{i\in T}$, where $Y_n \in \mathcal{Y}$ and $\mathcal{Y}$ is a measurable space. One of them (the most basic) is given by the canonical one. More generally, let $g\colon \mathcal{X}^T \to \mathcal{Y}$ be measurable and let $\mathbf{Y} = (Y_i)_{i\in T}$ be given by

$$Y_n = S^n g(\mathbf{X}),\ n \in T. \tag{1.2.3}$$

Every such process $\mathbf{Y}$ is called a ***factor*** of $\mathbf{X}$. Let us stress that this definition of a factor is consistent with the one from dynamical systems. Recall that for two dynamical systems $\mathcal{Q}_i = (X_i, \mathcal{B}_i, \mu_i, f_i)$, where $i \in \{1,2\}$, $\mathcal{Q}_2$ is said to be a ***factor of*** $\mathcal{Q}_1$ if there exists a measurable map $\pi\colon X_1 \to X_2$ satisfying $\pi \circ T_1 = T_2 \circ \pi$ and $\pi\mu_1 = \mu_2$. If additionally π is invertible and bi-measurable then we say that $\mathcal{Q}_1$ is ***isomorphic to*** $\mathcal{Q}_1$. The consistency of the definition of the factor of a process follows now from the fact that if $\pi\colon \mathcal{X}^T \to \mathcal{Y}^T$ satisfies $\pi \circ S = S \circ \pi$ then π must be of the form $(S^n g)_{n\in T}$, where $g\colon \mathcal{X}^T \to \mathcal{Y}$ is equal to the zero coordinate of π.

Recall that in the probabilistic setup we said that two processes are "equivalent" if they have the same distribution. How one should interpret the isomorphism of random processes in the dynamical context? It is the most natural to say that two stationary stochastic process are equivalent if the corresponding quadruples $\mathcal{Q}$ (as in (1.2.2)) are isomorphic in the sense of dynamical systems. Thus, $\mathbf{X}$ ***is isomorphic to*** $\mathbf{Y}$ given by (1.2.3) iff $\pi = (S^n g)_{n\in T} : \mathcal{X}^T \to \mathcal{Y}^T$ is invertible and bi-measurable. In a slightly informal way, this means that every realization $\mathbf{x}$ of $\mathbf{X}$ enables us to reconstruct the realization $\mathbf{y}$ of $\mathbf{Y}$ and vice versa.

At last, let us give some examples highlighting the differences between the dynamical and probabilistic definitions of equivalence of processes. The famous theorem due to Ornstein and Friedman [40] says that if two weak Bernoulli systems have the same entropy then they are isomorphic. In particular, every mixing Markov chain $\mathbf{M}$ is isomorphic to some i.i.d. process $\mathbf{X}$. On the other hand, clearly, if $\mathbf{M}$ is not degenerated then $\mathbf{M}$ and $\mathbf{X}$ cannot have the same distribution. Reversly, if we take $\Omega = [0,1]$, $\mathcal{F}$ is the Borel σ-algebra and $\mathbb{P} = \lambda$ is the Lebesgue measure and consider $X(t) = 2t \mod 1$ and $Y(t) = t$ and almost surely constant processes $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ and $\mathbf{Y} = (Y_i)_{i\in\mathbb{Z}}$, where $X_i = X$ and $Y_i = Y$ for all $i \in \mathbb{Z}$ then $\mathbf{X} \sim \mathbf{Y}$ (because $X \sim Y \sim \lambda$). On the other hand, there is no bi-measurable bijection π such that $\pi\mathbf{X} = \mathbf{Y}$ because $\mathbf{Y}$ is 1-1 whereas $\mathbf{X}$ is not. In short, each of these notions of equivalence is quite different. However, as we will see in this thesis, both these approaches contribute a great deal of knowledge concerning stochastic processes and when combined together, constitute a powerful tool which leads to many non-trivial results.

1.3 Mixed ergodic-probabilistic setup

In this section we introduce basic notation that is used throughout this thesis.

1.3.1 Static notions

In this part we introduce basic notions and notation related to sequences. They arise naturally both when one deals with stochastic processes and subshifts. As no dynamics is needed to define these notions, we put them in the most abstract context that is of our interest.

Let $\mathscr{X} \subset \mathcal{X}^T$. The state space $\mathcal{X}$ is sometimes called the ***alphabet*** of $\mathscr{X}$ (especially if $|\mathcal{X}| < \infty$). The elements of $\mathcal{X}$ are denoted by small letters, e.g. $x \in \mathcal{X}$, whereas **sequences** are thickened, e.g. $\mathbf{x} = (x_i)_{i\in T} \in \mathscr{X}$. An analogous convention is used for subsets: $A \subset \mathcal{X}$ and $\mathbf{A} \subset \mathscr{X}$. For convenience's sake we "upgrade" numbers 0 and 1 to ***constant sequences*** $\mathbf{0}$ and $\mathbf{1}$ respectively. Usually, we assume tacitly that all considered subsets and functions are measurable.

If $\mathcal{X} \subset \mathbb{R}$ then

$$\operatorname{supp} \mathbf{x} = \{i \in T \mid x_i \neq 0\}$$

stands for the ***support of*** $\mathbf{x}$.

Given a sequence $\mathbf{x} \in \mathscr{X}$ or a process $\mathbf{X} = (X_i)_{i\in T}$ and a finite set of indices $N \subset T$, where $N = \{i_1, \ldots, i_k\}$, $i_1 < i_2 < \cdots < i_k$, we define

$$x_N = (x_{i_1}, \ldots, x_{i_n}), \qquad X_N = (X_{i_1}, \ldots, X_{i_n})$$

(with obvious modifications when N is infinite). Let us add that if $N = \varnothing$ then it is convenient to think about $X_\varnothing$ as about a constant random variable.

For any k, $l \in \mathbb{Z}$ we introduce ***integer intervals*** (as soon as there is no confusion with the usual definition of an interval)

$$[k,l] = \{k, k+1, \ldots, l\}, \qquad (-\infty, k] = \{\ldots, k-2, k-1, k\}, \qquad [l, \infty) = \{l, l+1, l+2, \ldots\},$$

where we interpret $[k,l] = \varnothing$ as soon as $k > l$.

All standard operations valid on the real numbers are vectorized in a natural way (that is coordinatewise), resulting in their counterparts for sequences and random processes. For example, $\mathbf{x}+\mathbf{y} = \mathbf{w}$ where $w_i = x_i + y_i$ or $\mathbf{X} \leq \mathbf{Y}$ if $X_i \leq Y_i$ almost surely for all $i \in T$.

We denote by $\mathcal{L}_{\mathscr{X}}$ the ***language of*** $\mathscr{X}$, that is the family of all words appearing in $\mathscr{X}$. Recall that w is called ***a word*** (***over the alphabet*** $\mathcal{X}$) if $w = (w_0, w_1, \ldots, w_{n-1}) \in \mathcal{X}^n$ for some n and that w appears in $\mathscr{X}$ if there exists $\mathbf{x} \in \mathscr{X}$ and $i \in T$ such that $w = x_{[i,i+n-1]}$ (in other words, w is a substring of some $\mathbf{x} \in \mathscr{X}$). In that case $|w| = n$ is called the ***length of*** w. Furthermore, $\mathcal{L}_{\mathscr{X}}^{(n)} \subset \mathcal{L}_{\mathscr{X}}$ stands for the subset of the language of $\mathscr{X}$ consisting of words of length n. Sometimes we speak of ***blocks*** instead of words.

If $k \in T$ and w is a word over $\mathcal{X}$ then the corresponding ***cylinder*** set is given by

$$[w]_k = \left\{\mathbf{x} \in \mathscr{X} \mid x_{[k,k+|w|-1]} = w\right\}. \tag{1.3.1}$$

For brevity's sake we put $[w] = [w]_0$. Sometimes we identify words w with the corresponding cylinders $[w]$. In most cases it leads to no confusion and significantly clarifies writing. Thus, for example, for measure μ on $\mathscr{X}$, we can write $\mu(1)$ instead of $\mu([1])$.

For every $x \in \mathcal{X}$ we denote by $\#_x(w)$ ***the number of*** x ***which appear in*** w, that is

$$\#_x(w) = |\{i \in [0, n-1] \mid w_i = x\}|,$$

where $n = |w|$.

1.3.2 Dynamical notions

Measures When one speaks of stochastic processes or dynamical systems, measures come naturally into play. In both, ergodic and probabilistic approach, one usually identifies objects that are the same up to measure zero and often we do so tacitly. Frequently (where it leads to no confusion), we omit measurability details, e.g. we write that a measure is defined on some set assuming that this set is equipped with a measurable σ-field (usually the Borel one, if the set is a topological space). Sometimes, however, we need to be more precise and then some extra notation is provided.

Fix a measure ν on a measurable space $(\mathscr{X}, \mathcal{B})$. For any $A, B \in \mathcal{B}$ we say that B ***contains*** A ***up to*** ν ***(and write*** $A \overset{\nu}{\subset} B$***)*** if there is some Z such that $A \backslash Z \subset B$ and $\nu(Z) = 0$. Furthermore, $A \overset{\nu}{=} B$ if $A \overset{\nu}{\subset} B \overset{\nu}{\subset} A$. A similar convention is used for families of sets $\mathcal{A}$, $\mathcal{B}$ (including σ-algebras). More precisely we write $\mathcal{A} \overset{\nu}{\subset} \mathcal{B}$ if for every $A \in \mathcal{A}$ there is $B \in \mathcal{B}$ such that $A \overset{\nu}{\subset} B$. In case of σ-algebras, sometimes we say that $\mathcal{G} \subset \mathcal{H}$ in ***the sense of measure algebras*** if $\mathcal{G} \overset{\nu}{\subset} \mathcal{H}$ and the underlying measure ν is clear from the context. The ***set of all probability measures*** on $\mathscr{X}$ is denoted by $\mathcal{P} = \mathcal{P}_{\mathscr{X}}$.

Let now $\mathscr{X}$ be a subshift. As we have seen, any choice of $\mu \in \mathcal{P} = \mathcal{P}_{\mathscr{X}}$ results in a stochastic process $\mathbf{X}$ (with distribution μ). The ***subset of*** $\mathcal{P}$ ***of shift-invariant measures*** is denoted by

$$\mathcal{M} = \mathcal{M}_{\mathscr{X}} = \{\mu \in \mathcal{P}_{\mathscr{X}} \mid S\mu = \mu\}.$$

We omit index $\mathscr{X}$ as soon as the underlying subshift is clear from the context. Sometimes, associating a process with its distribution, we write $\mathbf{X} \in \mathcal{M}$.

Note that if $\mathcal{X}$ is **compact** then there is a natural construction of shift invariant measures. Indeed, recall that given a sequence $\mathbf{x} \in \mathscr{X}$ and $n \in \mathbb{N}$,

$$\delta_{\mathbf{x},n} = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{S^i\mathbf{x}} \tag{1.3.2}$$

is known as an ***empirical measure***. Since $\mathcal{M}_{\mathscr{X}}$ is compact, probability measures $\delta_{\mathbf{x},n}$ converge weakly along some subsequence to a probability measure $\nu \in \mathcal{P}$. Clearly, by the very definition, $\nu \in \mathcal{M}$. In such case we say that $\mathbf{x}$ is ***quasi-generic*** for ν. If the convergence holds along the whole sequence of natural numbers then $\mathbf{x}$ is ***generic*** for ν.

It is well-known that $\mathcal{M}$ is a convex set, the extreme points of which are ergodic measures. Recall that $\mu \in \mathcal{M}$ is ***ergodic*** (or S-***ergodic***) if the invariant σ-field,

$$\mathcal{I} = \left\{\mathbf{A} \subset \mathscr{X} \mid S^{-1}\mathbf{A} = \mathbf{A}\right\}$$

is μ-***trivial***, that is, $\mathcal{I} \stackrel{\mu}{=} \{\varnothing, \mathscr{X}\}$.

Induced shift Recall that if $\mathbf{X} \sim \mu \in \mathcal{M}$ then the quadruple $(\mathscr{X}, \mathcal{B}, \mu, S)$ (sometimes abbrieviated to $(\mathscr{X}, \mu, S)$ or even to $(\mathscr{X}, S)$ or $\mathscr{X}$) constitutes a measure-theoretic dynamical system (cf. (1.2.2)). Given $\mathbf{A} \subset \mathscr{X}$ with $\mu(\mathbf{A}) > 0$, we introduce now the corresponding ***induced subshift***

$$\mathcal{Q}_{\mathbf{A}} = (\mathscr{X}, S_{\mathbf{A}}, \mu_{\mathbf{A}}), \tag{1.3.3}$$

where the ***induced shift*** $S_{\mathbf{A}}$ and ***first return time*** $n_{\mathbf{A}}$ are given by

$$S_{\mathbf{A}}\mathbf{x} = S^{n_{\mathbf{A}}(\mathbf{x})}\mathbf{x}, \qquad n_{\mathbf{A}}(\mathbf{x}) = \inf\{k \geq 1 \mid S^k\mathbf{x} \in \mathbf{A}\}.$$

Note that $S_{\mathbf{A}}$ naturally acts on processes via

$$S_{\mathbf{A}}\mathbf{X} = \left(X_{i+\tau_{\mathbf{A}}}\right), \qquad \tau_{\mathbf{A}} = \tau_{\mathbf{A}}(\mathbf{X}) = \inf\left\{n \geq 1 \mid S^n\mathbf{X} \in \mathbf{A}\right\}.$$

In the case of random processes we call $S_{\mathbf{A}}$ the ***random shift***. Furthermore, we say that $S_{\mathbf{A}}\mathbf{X}$ is a ***randomly shifted process***. A process (or measure) $\mathbf{X} \sim \mu$ is $S_{\mathbf{A}}$-invariant if $S_{\mathbf{A}}\mathbf{X} \stackrel{\mathbb{P}_{\mathbf{X}\in\mathbf{A}}}{\sim} \mathbf{X}$ (equivalently, $S_{\mathbf{A}}\mu_{\mathbf{A}} = \mu_{\mathbf{A}}$). Similarly, $\mathbf{X}$ is $S_{\mathbf{A}}$-ergodic if $\{\mathbf{B} \mid S_{\mathbf{A}}^{-1}\mathbf{B} = \mathbf{B}\} \stackrel{\mu_{\mathbf{A}}}{=} \{\varnothing, \mathscr{X}\}$, that is the $S_{\mathbf{A}}$-invariant σ-field is $\mu_{\mathbf{A}}$-trivial. It is well-known that if $S\mu = \mu$ then $S_{\mathbf{A}}\mu_{\mathbf{A}} = \mu_{\mathbf{A}}$ and that $\mu_{\mathbf{A}}$ is $S_{\mathbf{A}}$-ergodic as soon as μ is S-ergodic. For more information and proofs of these facts we refer to Appendix A.

Thus, we can summarize our usual general mixed setup as

$$\mathbf{X} = (X_i)_{i\in T} \in \mathscr{X} \subset \mathcal{X}^T, \qquad X_i \in \mathcal{X}, \qquad S\mathbf{X} \sim \mathbf{X}, \qquad S_{\mathbf{A}}\mathbf{X} \stackrel{\mathbb{P}_{\mathbf{X}\in\mathbf{A}}}{\sim} \mathbf{X},$$

where $\mathscr{X} \subset \mathcal{X}^T$ is a subshift.

Couplings and joinings Fix two measures μ and ν on some sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. We define a ***coupling*** of μ and ν as a measure π on $\mathcal{X} \times \mathcal{Y}$ with ***marginals*** μ and ν, that is $\pi(\mathcal{X} \times \cdot) = \nu$ and $\pi(\cdot \times \mathcal{Y}) = \mu$. We denote the set of all such couplings π by $\mathcal{C}(\mu, \nu)$. Sometimes, for brevity's sake, we slightly abuse notation by writing $(X, Y) \in \mathcal{C}(\mu, \nu)$ instead of $\mathcal{L}((X, Y)) \in \mathcal{C}(\mu, \nu)$. Now let $\mathbf{X} \sim \mu$ and $\mathbf{Y} \sim \nu$ be stationary processes on $\mathcal{X}^T$ and $\mathcal{Y}^T$ respectively. We say that $(\mathbf{X}, \mathbf{Y}) = ((X_i, Y_i))_{i\in T} \sim \pi$ is a ***joining of*** μ ***and*** ν ***(***$\mathbf{X}$ ***and*** $\mathbf{Y}$***)*** if $\pi \sim (\mathbf{X}, \mathbf{Y}) \in \mathcal{C}(\mu, \nu)$ and $(\mathbf{X}, \mathbf{Y})$ is stationary. The set of all such joinings $\pi \sim (\mathbf{X}, \mathbf{Y})$ is denoted by $\in \mathcal{J}(\mu, \nu)$ (or $\mathcal{J}(\mathbf{X}, \mathbf{Y})$).

Generating partitions At the end let us say something about **generating partitions** which link ergodic objects with their probabilistic counterparts. Fix some measurable (at most countable) partitions $\mathscr{A}$ and $\mathscr{B}$ of a subshift $\mathscr{X}$. Recall that for any $i, j \in T$, where $i < j$,

$$\mathscr{A}_i^j = \bigvee_{k=i}^{j} S^{-k}\mathscr{A}, \tag{1.3.4}$$

where for any two partitions $\mathscr{A}$ and $\mathscr{B}$, $\mathscr{A} \vee \mathscr{B} = \{\mathbf{A} \cap \mathbf{B} \mid \mathbf{A} \in \mathscr{A}, \mathbf{B} \in \mathscr{B}\}$. We say that $\mathscr{A}$ is ***generating*** if the Borel σ-field $\mathcal{B}$ is generated by

$$\begin{cases} \displaystyle\bigvee_{k=0}^{\infty} S^{-k}\mathscr{A} = \sigma\left(\bigcup_{k=0}^{\infty}\bigvee_{j=0}^{k} S^{-k}\mathscr{A}\right), & \text{if } T = \mathbb{N}, \\ \displaystyle\bigvee_{k=-\infty}^{\infty} S^{-k}\mathscr{A} = \sigma\left(\bigcup_{k=0}^{\infty}\bigvee_{j=-k}^{k} S^{-k}\mathscr{A}\right), & \text{if } T = \mathbb{Z}. \end{cases}$$

If $\mathscr{X} \subset \mathcal{X}^T$ and $|\mathcal{X}|$ is at most countable then the partition

$$\{[x] \mid x \in \mathcal{X}\} \tag{1.3.5}$$

is a generating one. This observation immediately gives the following conclusions. The entropy rate (for the definition see Section 3.1.2) is a special case of Kolmogorov-Sinai entropy. This follows from the well-known fact that the supremum in the definition of Kolmogorov-Sinai entropy is attained on any generating partition (see for example [32], Section 4.2). Moreover, one-sided tail σ-algebras of processes coincide with the Pinsker algebra (see Chapter B.1). At last but not least, two dynamical systems are isomorphic if the corresponding codings for generating partitions are isomorphic.

1.4 Some additional notation

As usual, for convenience sake, we abbreviate $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. Usually, the log symbol stands either for $\log_e$ or $\log_2$ (in every section or statement we make it clear which one we use). Sometimes we go further and to avoid some annoying technicalities, we put for example $\log x$ to be equal $\log_e(x \vee e)$.

For any sequence of numbers $(a_n)_{n\in\mathbb{N}}$, we write $a_n \xrightarrow{n\to\infty} a$ if $\lim\limits_{n\to\infty} a_n = a$. When the time is clear from the context we just write $a_n \to a$. Furthermore, $a_n \sim b_n$ if $\frac{a_n}{b_n} \to 1$ when $n \to \infty$.

Given random variables X and Y we write $X \amalg Y$ if X ***is independent of*** Y. For a sequence of random variables $(X_i)_{i\in\mathbb{N}}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$, we use $X_n \to X$ if X_n **converges $\mathbb{P}$-almost surely** (abbreviated $\mathbb{P}$ a.s. or even a.s. if $\mathbb{P}$ is clear from the context) to X. Sometimes, in order to stress that a.s. convergence is meant with respect to some other probability measure $\mathbb{Q}$ on Ω, we use $X_n \xrightarrow{\mathbb{Q}\ a.s.} X$. Furthermore, $X_n \Rightarrow X$ stands for the **convergence in distribution**. Some other types of convergence are announced by explicit writings, for example $f_n \xrightarrow{L_p(\mathbb{P})} f$ means that f_n convergence in $L_p(\mathbb{P})$ norm to f.

We use the following norms: $\|\cdot\|_{TV}$ **total variation norm on the space of signed finite measures**, $\|\cdot\|_\infty$, depending on the context, either the **supremum norm** or the L_∞ **norm**.

Unconventionally, given a subshift $\mathscr{X}$, $\mathbf{H} = \mathbf{H}_{\mathscr{X}}$ always stands for the **topological entropy of $\mathscr{X}$**. Moreover, for any $\mu \in \mathcal{M}_{\mathscr{X}}$, $\mathbf{H}(\mu)$ denotes the **Sinai-Kolmogorov entropy of μ**. This notation is motivated by the compatibility with the symbols we use for entropy in probabilistic setting. For now, let us only say that if $\mathcal{X}$ is countable then $\mathbf{H}(\mu) = \mathbf{H}(\mathbf{X})$ for any $\mathbf{X}$ such that $\mathbf{X} \sim \mu$, where by $\mathbf{H}(\mathbf{X})$ we mean the **entropy rate** of $\mathbf{X}$.

At the end, let us stress the difference between $X \in \mathcal{X}$ and $X \in \mathcal{G}$ where $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-field and $\mathcal{X}$ is the state space of X. In the latter we mean that X is measurable with respect to $\mathcal{G}$.

Chapter 2

Summary of our results

As it has already been mentioned, in this section we give just a taste of our main results. In particular, in order to avoid introducing some additional theory or objects, some of the results are given in a slightly weakened form. Furthermore, we take for granted that the reader is familiar with the basic notions from both, ergodic and probability theory. In particular, we skip some standard definitions and facts. For the detailed introduction and full exposition of particular results we refer to part two of this thesis, where each subject is treated thoroughly.

Let us add that Sections 2.1 and 2.2 are highly correlated with each other. On the one hand, the results from Section 2.1 concerning entropy and topological pressure are more general than their counterparts from Section 2.1. On the other hand, $\mathscr{B}$-free systems from Section 2.1 were our main motivation behind obtaining theorems from Section 2.2. In particular, many comments from Section 2.1 refer to Section 2.2 and vice versa.

2.1 Gibbs measures, topological pressure and $\mathscr{B}$-free systems

2.1.1 Introduction

The study of $\mathscr{B}$-free systems partly arises from the interest in the properties of the Möbius function $\boldsymbol{\mu}\colon \mathbb{N} \to \{-1,0,1\}$, whose square $\boldsymbol{\mu}^2$ (extended to $\mathbb{Z}$ symmetrically) is the characteristic function of square-free integers, i.e. numbers not divisible by the square of any prime. Given $\mathscr{B} \subset \mathbb{N}\setminus\{1\}$, let $\eta = \mathbb{1}_{\mathcal{F}_{\mathscr{B}}}$, where $\mathcal{F}_{\mathscr{B}} = \mathbb{Z}\setminus\bigcup_{b\in\mathscr{B}} b\mathbb{Z}$. The corresponding dynamical system $\mathscr{X}_\eta$ (called a ***$\mathscr{B}$-free system***) is defined as the orbit closure of $\eta \in \{0,1\}^{\mathbb{Z}}$ under the left shift S (i.e. we deal with a subshift). The square-free system $(X_{\boldsymbol{\mu}^2}, S)$ given by $\boldsymbol{\mu}^2$ is a topological factor of the subshift given by $\boldsymbol{\mu}$ itself. Sarnak in his seminal lectures on the randomness of the Möbius function [92] formulated certain statements about the square-free system which extend in a natural way to general $\mathscr{B}$-free systems. One of the open problems stated back then was the intrinsic ergodicity of the square-free system, i.e. the problem of whether $(X_{\boldsymbol{\mu}^2}, S)$ has exactly one measure realizing the topological entropy. It was resolved by Peckner in [85] and later extended to the general case in [61, 34]. A natural question arose, whether the measure of maximal entropy has the Gibbs property (as it is often the case in many natural situations, including sofic systems [101], i.e. factors of subshifts of finite type). Peckner in [85] provided the negative answer in the square-free case. However, his proof relied on non-trivial number-theoretic facts on the primes (and an explicit formula for the Mirsky measure of a block) and thus he asked if his result extends to general $\mathscr{B}$-free systems. Our main result gives the positive answer to Peckner's question. In fact, we are able to give a more general criterion (applicable beyond the $\mathscr{B}$-free systems) based on the notions of topological entropy and (topological) density of ones which ensures the absence of Gibbs property (see Theorem 2.1.2).

Furthermore, we study the closely related problem of finding the topological pressure for $\mathscr{B}$-free systems. Computing the topological pressure of general dynamical systems is a non-trivial task – explicit formulas are available only in some cases for special potentials. A classical example is a Walter's result for subshifts of finite type [99], where the topological pressure is roughly given by the greatest eigenvalue of an appropriate matrix or its generalization, the Ruelle-Perron-Frobenius operator. Even though $\mathscr{B}$-free systems can be approximated by sofic systems, it does not seem to us that this can be used to solve the problem under consideration. One reason is that the size of matrices coming into play grows very rapidly. This makes it difficult to find algebraic relations useful in the problem of computing the topological pressure. The probabilistic approach completely avoids these obstacles and results in an explicit (and relatively easy to compute!) formula for the desired quantity for continuous potentials. Moreover, it allows us to establish the uniqueness of the equilibrium measure for potentials depending only on one coordinate, which extends the fact of the intrinsic ergodicity of $\mathscr{B}$-free systems.

2.1.2 Basic notions and notation

Recall that a subshift $(\mathscr{X}, S)$, where $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$, is ***hereditary*** if for every $W \in \mathcal{L}$ and $W' \leq W$ we have $W' \in \mathcal{L}$, where $\mathcal{L} = \mathcal{L}_{\mathscr{X}}$ stands for the language of $\mathscr{X}$. Moreover, given a subshift $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$, the ***hereditary closure*** of $\mathscr{X}$ is defined via

$$\widetilde{\mathscr{X}} = \{\mathbf{z} \in \{0,1\}^{\mathbb{Z}} : \mathbf{z} \leq \mathbf{x} \text{ for some } \mathbf{x} \in \mathscr{X}\}.$$

The ***topological entropy*** $\mathbf{H}$ of $\mathscr{X}$ is given by

$$\mathbf{H} = \mathbf{H}_{\mathscr{X}} = \lim_{n\to\infty} \frac{1}{n} \log\left(\left|\mathcal{L}^{(n)}\right|\right).$$

(here and later in this chapter, log stands for $\log_2$). Similarly, for any $\nu \in \mathcal{M}$ the ***Kolmogorov-Sinai entropy*** $\mathbf{H}(\nu)$ can be computed as

$$\mathbf{H}(\nu) = \lim_{n\to\infty} \frac{1}{n} \mathbf{H}_\nu\left(\mathcal{L}^{(n)}\right), \qquad \mathbf{H}_\nu\left(\mathcal{L}^{(n)}\right) = -\sum_{w\in\mathcal{L}^{(n)}} \nu(w)\log\nu(w).$$

An ergodic measure $\kappa \in \mathcal{M}^e_{\mathscr{X}}$ is said to have the ***Gibbs property*** if there exists a constant $a > 0$ such that

$$\kappa(C) \geq a \cdot 2^{-|C|\mathbf{H}_{\mathscr{X}}}, \qquad \forall_{C\in\mathcal{L},\, \kappa(C)>0}. \tag{2.1.1}$$

Remark 2.1.1. Note that if $\mathbf{H}_{\mathscr{X}} = 0$ and κ has the Gibbs property then it must be purely atomic. In that sense, when one considers the Gibbs property, the most interesting cases arise when the underlying subshift has positive topological entropy.

For any $\mu, \nu \in \mathcal{M}$, we say that $\kappa = \mu \overset{\text{ind.}}{*} \nu$ is ***the independent multiplicative convolution of ν and μ*** if $\kappa \sim \mathbf{X}\cdot\mathbf{Y}$ where $\mathbf{X} \sim \mu$, $\mathbf{Y} \sim \nu$ and $\mathbf{X} \amalg \mathbf{Y}$.

Now, we introduce some notions concerning densities of ones. We define the ***density of ones*** for $\mathscr{X}$ and its measure equivalent for $\nu \in \mathcal{M}$ by

$$\mathbf{D} = \mathbf{D}_{\mathscr{X}} = \lim_{n\to\infty} \frac{1}{n} \max_{W\in\mathcal{L}^{(n)}} \#_1 W, \qquad \mathbf{D}_\nu = \lim_{n\to\infty} \frac{1}{n} \max_{W\in\mathcal{L}^{(n)},\, \nu(W)>0} \#_1 W,$$

respectively. Note that the limits exist due to the subadditivity of appropriate sequences. Furthermore, one can show that $\sup_{\nu\in\mathcal{M}_{\mathscr{X}}} \mathbf{D}_\nu = \mathbf{D}_{\mathscr{X}}$. Any measure ν which realizes the supremum is called ***ones-saturated***. Clearly, if ν is of full topological support then it is ones-saturated.

Finally, for any $0 \leq p \leq 1$, let us introduce the ***family of Bernoulli measures*** $B_p \sim \mathbf{B}^{(p)}$, where $\mathbf{B}^{(p)} = \left(B_i^{(p)}\right)_{i\in\mathbb{Z}}$ is an i.i.d. Bernoulli process with parameter p, that is, $\mathbb{P}\left(B_i^{(p)} = 1\right) = p = 1 - \mathbb{P}\left(B_i^{(p)} = 0\right)$.

2.1.3 Results

Our main result is the following one.

Theorem 2.1.2. *Fix a subshift $(\mathscr{X}, S)$, where $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$, and suppose that $\nu \in \mathcal{M}^e_{\mathscr{X}}$ is ones-saturated and non-atomic. If $\mathbf{D}_{\widetilde{\mathscr{X}}} = \mathbf{H}_{\widetilde{\mathscr{X}}}$ then $\kappa = B_{\frac{1}{2}} \overset{\text{ind.}}{*} \nu$ **does not** have the Gibbs property (2.1.1).*

Remark 2.1.3. One can easily show that $\mathbf{D}_{\widetilde{\mathscr{X}}} \leq \mathbf{H}_{\widetilde{\mathscr{X}}}$ always holds (cf. Proposition 4.1.14).

Recall that given a $\mathscr{B}$-free system $(\mathscr{X}_{\boldsymbol{\eta}}, S)$, the associated Mirsky measure $\nu_{\boldsymbol{\eta}}$ is a natural invariant measure which is (quasi-)generated by the characteristic function of the $\mathscr{B}$-free integers (if $\mathscr{B}$ is pairwise coprime with summable series of reciprocals then η is generic for ν_η; this clearly includes the square-free case): for the details on the Mirsky measure we refer to Section 4.1.7. Moreover, the measure of maximal entropy for $(\widetilde{\mathscr{X}}_\eta, S)$ is unique and given by $B_{\frac{1}{2}} \overset{\text{ind.}}{*} \nu_\eta$. It is a classical fact in the theory of cut-and-project sets that for any $\mathscr{B}$, the Mirsky measure $\nu_{\boldsymbol{\eta}}$ is ones-saturated for $(\mathscr{X}_{\boldsymbol{\eta}}, S)$ (see, e.g., Theorem 4 and Corollary 4 in [57], cf. Chapter 7 in [7] as well; combine these facts with Theorem 4.1.9). Alternatively, one can show that every shift-invariant measure on $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$ can be expressed as a multiplicative convolution of a certain measure with the Mirsky measure (Theorem 4.1.23) and thus, by Theorem 4.1.9, the Mirsky measure must be of maximal density of ones (and thus, ones-saturated). Furthermore, we always have $\mathbf{D}_{\mathscr{X}_\eta} = \mathbf{H}_{\widetilde{\mathscr{X}}_\eta}$ (see Proposition K in [34] or combine Theorems 2.1.16 and 4.1.9 below with the fact that the Mirsky measure is of maximal density of ones). Thus, Theorem 2.1.2 immediately results in the positive answer to the question asked by Peckner in [85].

Corollary 2.1.4. *Let $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$. Suppose that the Mirsky measure ν_η is not periodic. Then the measure of maximal entropy of $(\widetilde{\mathscr{X}}_\eta, S)$ does not have the Gibbs property* (2.1.1).

Remark 2.1.5. We will say a few words about the (non-)periodicity of $\nu_{\boldsymbol{\eta}}$ later in this section. For the precise description of sets $\mathscr{B}$ for which $\nu_{\boldsymbol{\eta}}$ is periodic, see Corollary 4.1.36.

Apart from $\mathscr{B}$-free systems, Theorem 2.1.2 allows us to obtain results for some other intrinsically ergodic subshifts. Recall that a subshift $\mathscr{X}$ is ***uniquely ergodic*** if there is exactly one shift-invariant (and thus ergodic) measure on $\mathscr{X}$.

Corollary 2.1.6. *If $(\mathscr{X}, S)$, where $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$, is uniquely ergodic and $\mathbf{H}_{\mathscr{X}} = 0$ then $B_{\frac{1}{2}} \overset{ind.}{*} \nu$ has no Gibbs property whenever the unique invariant measure ν is non-atomic.*

Let us recall the definition of a Sturmian dynamical system. Consider a real number $\alpha \in (0,1)$ and a sequence $\mathbf{c}_\alpha = (c_\alpha(n))_{n\in\mathbb{Z}}$, where

$$c_\alpha(n) = \mathbb{1}_{[0,1-\alpha)}(n\alpha \bmod 1). \tag{2.1.2}$$

The corresponding ***Sturmian system*** is given by $\mathscr{X}_\alpha := \overline{\{S^i \mathbf{c}_\alpha \mid i \in \mathbb{Z}\}} \subset \{0,1\}^{\mathbb{Z}}$ (considered with the left shift S). Since irrational rotation dynamical systems are uniquely ergodic and of zero topological entropy, Sturmian systems as codings (with controlled discontinuities) of rotation dynamical systems, must inherit these properties. We will deal with its hereditary closure, called ***hereditary Sturmian system***, i.e. $\widetilde{\mathscr{X}}_\alpha$. As shown in [61], such subshifts are intrinsically ergodic. Their measure of maximal entropy is of the form $B_{\frac{1}{2}} * \nu$, where ν is the unique invariant measure for the underlying Sturmian system. Therefore Corollary 2.1.6 yields the following result.

Corollary 2.1.7. *If $(\widetilde{\mathscr{X}}, S)$ is a hereditary Sturmian system then its measure of maximal entropy has no Gibbs property.*

As a byproduct, we also prove several results on $\mathscr{B}$-free systems that are of independent interest. In particular, we prove the converse to a recent result by Keller [56], thus obtaining a dynamical characterization of an important arithmetical notion of tautness. Before we state it, recall that a set $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$ is ***taut*** (see [49]) if for every $b \in \mathscr{B}$,

$$\boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}}) > \boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}\setminus\{b\}}),$$

where for any set $N \subset \mathbb{Z}$, $\boldsymbol{\delta}(N) = \lim_{n\to\infty} \frac{1}{\log n} \sum_{i=1}^{n} \frac{1}{i} \mathbb{1}_{i\in N}$ stands for its logarithmic density (or rather the logarithmic density of $N \cap \mathbb{N}$), as soon as it exists (which is the case for the sets of multiples, as proved by Davenport and Erdös in [29]). Recall also that whenever the natural density of a set $N \subset \mathbb{Z}$ exists then so does the logarithmic density and these two quantities are equal (for an easy argument see Remark 4.1.4).

Theorem 2.1.8. *Let $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$. If the corresponding Mirsky measure ν_η is of full support $\mathscr{X}_\eta$ then $\mathscr{B}$ is taut.*

Furthermore, we obtain an explicit formula for the topological pressure of $\mathscr{B}$-free systems. Recall that, given a potential φ and a finite alphabet $\mathcal{X} = \{x_1, \dots, x_k\}$, the ***topological pressure of a subshift*** $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$ is given by

$$\mathscr{P}_{\mathscr{X},\varphi} = \lim_{n\to\infty} \frac{1}{n} \log_2 \sum_{\mathbf{A}_i \in \mathscr{P}_n} 2^{\sup_{\mathbf{A}_i} S_n\varphi(\mathbf{x})},$$

where $\mathscr{P}_n = \{[x_1], \dots, [x_k]\}_0^{n-1}$, $S_n = \sum_{i=0}^{n-1} S^i$ (for any partition $\mathscr{C}$ of $\mathscr{X}$, $\mathscr{C}_i^j = \bigvee_{k=i}^j S^{-k}\mathscr{C}$). In particular, if $\varphi = 0$ is the zero potential then we recover the definition of the topological entropy. Furthermore, the following ***variational principle*** holds,

$$\mathscr{P}_{\mathscr{X},\varphi} = \sup_{\mathbf{X} \in \mathcal{M}_{\mathscr{X}}} \left[\mathbf{H}\left(\mathbf{X}\right) + \mathbb{E}\varphi(\mathbf{X}) \right]. \tag{2.1.3}$$

Every measure (process) $\mathbf{X} \in \mathcal{M}_{\mathscr{X}}$ which realizes the above supremum is called an ***equilibrium measure*** (if $\varphi = 0$ then such measure $\mathbf{X}$ is known as a ***measure of maximal entropy*** and a system which admits a unique measure of maximal entropy is called ***intrinsically ergodic***).

Let us note that for any $\mathscr{B}$-free system $\mathscr{X}_{\boldsymbol{\eta}}$, every measure $\mathbf{X} \in \mathcal{M}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}}$ has a representation of the form $\mathbf{Z} \cdot \mathbf{Y}$ for some stationary process $(\mathbf{Z}, \mathbf{Y}) \in \mathcal{M}_{\{0,1\}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}}}$, where $\mathbf{Y}$ is distributed according to the Mirsky measure $\nu_{\boldsymbol{\eta}}$, (see Theorem 4.1.23). Therefore the variational principle (2.1.3) can be rewritten as

$$\mathscr{P}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}},\varphi} = \sup_{(\mathbf{Z},\mathbf{Y}) \in \mathcal{M}_{\{0,1\}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}}}, \mathbf{Y} \sim \nu_{\boldsymbol{\eta}}} \left[\mathbf{H}\left(\mathbf{Z} \cdot \mathbf{Y}\right) + \mathbb{E}\varphi(\mathbf{Z} \cdot \mathbf{Y}) \right]. \tag{2.1.4}$$

The Mirsky measure $\nu_{\boldsymbol{\eta}}$ corresponding to a $\mathscr{B}$-free system $\mathscr{X}_{\boldsymbol{\eta}}$ can be either periodic or not (depending on the structure of $\mathscr{B}$). Clearly, if $\mathscr{B}$ is finite then the corresponding $\mathscr{X}_{\boldsymbol{\eta}}$ is periodic and so is its unique invariant measure, that is, the Mirsky measure. On the other hand if $|\mathscr{B}| = \infty$ then it can still happen that the corresponding Mirsky measure $\nu_{\boldsymbol{\eta}}$ is periodic. Indeed, for example, consider two sets: $\mathscr{B} = \{2\}$ and $\mathscr{B} = 2\mathscr{P}$, where $\mathscr{P}$ stands for the set of primes. Clearly, in the first case, the set of $\mathscr{B}$-free integers is equal to odd numbers and in the second one, to the union of odd numbers and set $\{-2, 2\}$. However, in both these cases the Mirsky measure is the same and periodic (recall that the Mirsky measure is generated by the indicator of $\mathscr{B}$-free numbers whenever this sequence is generic). Furthermore, the problems we study (and the way we approach them) make us look at $\mathscr{X}_{\boldsymbol{\eta}}$ from the point of view of $\nu_{\boldsymbol{\eta}}$. This is why the (non-)periodicity of $\nu_{\boldsymbol{\eta}}$ is of our interest, rather than the (in)finiteness of $\mathscr{B}$.

We discuss the periodic case first, as it is much easier than the non-periodic one. Let $\mathbf{Y} \sim \nu_{\boldsymbol{\eta}}$ stand for the Mirsky measure associated with $\mathscr{B}$. Let $p \in \mathbb{N}$ be the ***period*** of $\mathbf{Y}$, that is the smallest natural number such that $S^p\mathbf{Y} = \mathbf{Y}$ and $m = \sum_{i=1}^p Y_i$ be the ***number of ones contained in that period***. A combination of Theorem 2.2.18 with the variational principle (2.1.4) immediately gives the following formula for the pressure of $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$.

Theorem 2.1.9. *Consider a $\mathscr{B}$-free system $\mathscr{X}_{\boldsymbol{\eta}}$ with the corresponding Mirsky measure $\mathbf{Y}$, where $|\mathscr{B}| < \infty$. Let p be the period of $\mathbf{Y}$. Then for any potential φ depending on at most p consecutive coordinates,*

$$\mathscr{P}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}},\varphi} = \frac{1}{p} \log_2 \left[\sum_{z_{[0,m-1]} \in \mathcal{X}^m} 2^{p\Phi(z_{[0,m-1]})} \right], \tag{2.1.5}$$

where $m = \sum_{i=1}^p Y_i$ and $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$, that is, Φ is the upgrade of φ given by (3.2.24).

Remark 2.1.10. In this section we intentionally omit the precise definition of the upgrade of potential φ (via process $\mathbf{Y}$) – despite being easy in concept it is burdensome to present in a short way (for a quick introduction to this object we refer to Definition 2.2.16; a thorough analysis is done in (3.2.24)). For now, let us only mention that the function Φ is completely determined by the returns of process $\mathbf{Y}$ to state 1.

Now, we turn to the case of non-periodic Mirsky measure.

Theorem 2.1.11. *For any $\mathscr{B}$-free system $\mathscr{X}_{\boldsymbol{\eta}}$ such that the corresponding Mirsky measure $\mathbf{Y} \sim \nu_{\boldsymbol{\eta}}$ is **not** periodic and a continuous potential $\varphi\colon \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$,*

$$\mathscr{P}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}},\varphi} = \mathbb{P}\left(Y_0 = 1\right) + \sup_{\mathbf{z}\in\mathcal{X}^{\mathbb{Z}}} \Phi(\mathbf{z}), \tag{2.1.6}$$

where $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$ is the upgrade of φ given by (3.2.24).

Remark 2.1.12. Note that the assumptions of Theorem 2.1.11 exclude $\nu_{\boldsymbol{\eta}} = \delta_{\mathbf{1}}$ (in which the corresponding $\mathscr{B}$-free shift becomes the full shift) and thus our result does not say anything about the topological pressure of the full shift $\{0,1\}^{\mathbb{Z}}$.

Remark 2.1.13. Let us also add that the proof of the above theorem relies on both, the explicit formula for the entropy of multiplicative convolution of measures (which we provide in Theorem 2.2.5 below) and the appropriate choice of **periodic approximation** of $\mathscr{X}_{\boldsymbol{\eta}}$ (which, for every $K \in \mathbb{N}$, is given by the $\mathscr{B}$-free subshift associated with $\mathcal{M}_{\{b\in\mathscr{B}:b\leq K\}}$; in particular, we can apply Theorem 2.1.9 to approximate the pressure of non-periodic case). This approach is highly suggested by the result of Davenport-Erdös (see [29]):

$$\boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}}) = \underline{\mathbf{d}}(\mathcal{M}_{\mathscr{B}}) = \lim_{K\to\infty} \mathbf{d}(\mathcal{M}_{\{b\in\mathscr{B}:b\leq K\}}),$$

where for any set $N \subset \mathbb{Z}$, $\mathbf{d}(N), \underline{\mathbf{d}}(N)$ stands for the density (lower density resp.) of N (or rather $N \cap \mathbb{N}$), which implies some convergence results (see for example Corollary C.0.5).

Remark 2.1.14. Let us explain (informally) why knowing Theorem 2.1.9 one may expect a result like Theorem 2.1.11. The intuition can be summarized in two observations. Firstly, by the previous remark, a general non-periodic case can be approximated by periodic ones for which Theorem 2.1.9 can be applied. Secondly, the expression in (2.1.5) can be rewritten as (a logarithm of) the l_p norm of an appropriate vector. It is well-known that as $p \to \infty$ such norms converge to the l_∞ norm (this is why we obtain the supremum in the formula from (2.1.6)). Of course this reasoning does not explain the appearance of $\mathbb{P}\left(Y_0 = 1\right)$ in (2.1.6) and some detailed convergence analysis of (2.1.5) as $p \to \infty$ must be done.

Remark 2.1.15. In view of Theorem 2.1.11 which provides an explicit formula for the topological pressure of a $\mathscr{B}$-free system, it would be interesting to describe any of equilibrium measures for $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$. So far we know only that such an example can be obtained as a weak limit of certain multiplicative convolutions of Gibbs-like i.i.d. processes with (periodic) approximations of the Mirsky measure (see (3.2.33) in Theorem 3.2.32). However, we know nothing about properties of this limiting process. In particular, it remains open if (or more precisely, under which conditions) a system like in Theorem 2.1.11 admits only one equilibrium measure (we know only that this happens if the underlying potential depends on one coordinate, see Theorem 2.1.16 below).

At the end we present the result which extends the fact of the intrinsic ergodicity of the hereditary closure of a $\mathscr{B}$-free system.

Theorem 2.1.16. *Suppose that a continuous potential $\varphi\colon \{0,1\}^{\mathbb{Z}} \to \mathbb{R}$ depends only on one coordinate. Then the topological pressure of the hereditary closure of $\mathscr{X}_{\boldsymbol{\eta}}$ is given by*

$$\mathscr{P}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}},\varphi} = (1-d)\varphi(0) + d\log_2\left(2^{\varphi(0)} + 2^{\varphi(1)}\right), \tag{2.1.7}$$

where $d = \nu_{\boldsymbol{\eta}}(1)$. Furthermore, there is a unique equilibrium measure for φ, which is given by

$$\mathbf{G}\cdot\mathbf{Y}, \tag{2.1.8}$$

where $\mathbf{Y} \sim \nu_{\boldsymbol{\eta}}$, $\mathbf{Y} \amalg \mathbf{G}$ and G is an i.i.d. binary process such that $\mathbb{P}\left(G_i = j\right) = 2^{\varphi(j)} / \left[2^{\varphi(0)} + 2^{\varphi(1)}\right]$ for $j \in \{0,1\}$ (so G_i is the Gibbs measure associated with φ).

Remark 2.1.17. Compare the formula from (2.1.7) to the well-known one for the pressure of the full shift $\mathscr{X} = \{0,1\}^{\mathbb{Z}}$ for the potentials depending on one coordinate (say on the zero coordinate), namely,

$$\mathscr{P}_{\mathscr{X},\varphi} = \log_2 \left(2^{\varphi(0)} + 2^{\varphi(1)}\right),$$

which easily follows from

$$\mathcal{P}_{\mathscr{X},\varphi} = \sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}} [\mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(X_0)] \leq \sup_{X_0\in\{0,1\}} [\mathbf{H}(X_0) + \mathbb{E}\varphi(X_0)] = \log_2 \left(2^{\varphi(0)} + 2^{\varphi(1)}\right)$$

(the last step follows from a standard calculation made for example below equation (3) in [20]). In particular, we easily see that the supremum in the variational formula for $\mathscr{X} = \mathscr{P}_{\{0,1\}^{\mathbb{Z}},\varphi}$ is attained at an i.i.d. process $\mathbf{G}$ from Theorem 2.1.16.

Thus, the case of the full shift can be treated as a special case of Theorem 2.1.16 for $d = 1$ (that is, the case in which $\nu_{\boldsymbol{\eta}} = \delta_{\mathbf{1}}$).

2.2 Entropy of multiplicative convolution, topological pressure and retrieving lost signal

2.2.1 Introduction

Let $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ and $\mathbf{Y} = (Y_i)_{i\in\mathbb{Z}}$ be finitely-valued real processes such that $(\mathbf{X},\mathbf{Y})$ is stationary. Assume additionally that $\mathbf{Y}$ is ergodic, $Y_i \in \{0,1\}$ for $i \in \mathbb{Z}$ and $\mathbb{P}(Y_0 = 1) > 0$ (the last assumption, $\mathbf{Y} \neq \mathbf{0}$, is made to avoid some degenerate cases when one considers $\mathbf{X}\cdot\mathbf{Y}$). Recall that for any finitely-valued stationary process $\mathbf{Z}$,

$$\mathbf{H}(\mathbf{Z}) = \lim_{n\to\infty} \frac{1}{n}\mathbf{H}\left(Z_{[1,n]}\right) = \mathbf{H}\left(Z_0 \mid Z_{(-\infty,-1]}\right)$$

stands for the ***entropy rate***.

In this part we investigate the entropy rate of multiplicative convolution process $\mathbf{X}\cdot\mathbf{Y}$, that is $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$. Apart from obtaining an explicit formula for $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$, we study the following questions posed in a slightly weaker form in [61] (Question 1 therein):

1. Is there a general formula for the entropy rate $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$?
2. Do we always have $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) > 0$ whenever $\mathbf{H}(\mathbf{X}) > 0$?
3. Can we have $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) = \mathbf{H}(\mathbf{X}) > 0$?

There is, though, another problem related to the process $\mathbf{X}\cdot\mathbf{Y}$. Before we explain it, recall the famous problem of filtering a noisy signal by Furstenberg (from 1967). The fundamental question asked in [41] was when one can retrieve a signal $\mathbf{Z}$ from the perturbed one $\mathbf{Z}+\mathbf{W}$, where $\mathbf{Z}$ and $\mathbf{W}$ are real-valued stationary processes. To solve this problem Furstenberg introduced the notion of **absolute disjointness of processes** (which is a much stronger property than the independence of processes) and showed that this property is sufficient for extracting $\mathbf{Z}$ from $\mathbf{Z}+\mathbf{W}$ (in fact Furstenberg needed some additional condition of integrability but it was shown much later by Garbit [44] that it is redundant). In the same spirit one may interpret $\mathbf{X}\cdot\mathbf{Y}$ as a lost signal (recall that $Y_i \in \{0,1\}$) and ask when it is possible/impossible to retrieve $\mathbf{X}$ from $\mathbf{X}\cdot\mathbf{Y}$. Clearly, if $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) < \mathbf{H}(\mathbf{X})$ then one cannot hope to get $\mathbf{X}$, thus, it is natural to reformulate Question 3 in the following way:

3'. Is there a natural criterion for $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) < \mathbf{H}(\mathbf{X})$ subject to $\mathbf{H}(\mathbf{X}) > 0$?

Remark 2.2.1. Note that if the state spaces of processes $\mathbf{X}$ and $\mathbf{Y}$ were contained in the set of positive real numbers then an application of logarithm to $\mathbf{X}\cdot\mathbf{Y}$ would transform the problem of retrieving of a lost signal into Furstenberg's filtering problem $\mathbf{Z}+\mathbf{W}$, where $\mathbf{Z} = \log\mathbf{X}$ and $\mathbf{W} = \log\mathbf{Y}$. In that sense it is important that we allow Y_i's to be zero.

Remark 2.2.2. Notice that in the classical Furstenberg problem the signal $\mathbf{X}$ is represented by a zero entropy (i.e. deterministic) process. Contrary to this, in our setting $\mathbf{X}$ is non-deterministic and it is perturbed by a deterministic process $\mathbf{Y}$, so the interpretation from the classical situation does not fully apply. Nevertheless, there is a clear analogy of these two settings and the problem of entropy loss seems to be of independent interest (see also the next remark).

Remark 2.2.3. Let us add that a similar (in fact, a much more general) problem of retrieving signal was studied by Furtenberg, Peres and Weiss in [42]. More precisely, they stated the following question. Let $\mathbf{X}^{(i)} = \left(X_j^{(i)}\right)_{j\in\mathbb{Z}}$, where $i \in \mathbb{N}$, be a family of processes and $\mathbf{U}$ be $\mathbb{N}$-valued process. Suppose that all these processes are jointly stationary. Define

$$\mathbf{X}^{(\mathbf{U})} = \left(X_i^{(U_i)}\right)_{i\in\mathbb{Z}} \tag{2.2.1}$$

(thus, informally, $\mathbf{U}$ chooses among the family of processes). When one can retrieve $\mathbf{U}$ from $\mathbf{X}^{(\mathbf{U})}$? In order to answer this question the authors of [42] introduced the notion of **double disjointness** of processes. We say that process $\mathbf{A}$ is ***double disjoint (DD) from*** $\mathbf{B}$ if every self-joining of $\mathbf{A}$ is absolutely disjoint from $\mathbf{B}$. In other words, if $(\mathbf{A}', \mathbf{A}'', \mathbf{B}')$ is a stationary process such that $\mathbf{A}', \mathbf{A}'' \sim \mathbf{A}$ and $\mathbf{B}' \sim \mathbf{B}$ then $(\mathbf{A}', \mathbf{A}) \amalg \mathbf{B}'$. The most basic example of DD processes arises when we take $\mathbf{A}$ of zero entropy rate (then clearly, every self-coupling of $\mathbf{A}$ has zero entropy) and $\mathbf{B}$ which has the trivial tail-σ-field (let us add that in fact if $\mathbf{A}$ is DD from $\mathbf{B}$ then **necessarily** $\mathbf{H}(\mathbf{A}) = 0$ and $\mathbf{B}$ is ergodic). Now, the main result of [42] can be summarized (roughly) as follows. Suppose that $\mathbf{X}^{(i)}$ for $i \in \mathbb{N}$ and $\mathbf{U}$ are jointly stationary. If $\mathbf{U}$ is DD from each $\mathbf{X}^{(i)}$ for $i \in \mathbb{N}$ then one can retrieve $\mathbf{U}$ from $\mathbf{X}^{(\mathbf{U})}$.

Let us explain how to fit this theorem to our setting from Question 3'. Consider two processes $\mathbf{X}^{(i)}$, for $i \in \{0, 1\}$, where

$$X_j^{(i)} = iX_j \tag{2.2.2}$$

and take $\mathbf{U} = \mathbf{Y}$. Then $\mathbf{X}^{(\mathbf{U})} = \mathbf{X} \cdot \mathbf{Y}$ and the theorem states that we can retrieve $\mathbf{Y}$ from $\mathbf{X} \cdot \mathbf{Y}$ as soon as $\mathbf{Y}$ is DD from $\mathbf{X}$. Note that, since we assume that $\mathbf{H}(\mathbf{X}) > 0$, we cannot exchange the role of $\mathbf{X}$ and $\mathbf{Y}$ in above reasoning. In this sense, the problem we address is complementary to the one studied in [42].

Remark 2.2.4. We construct just one example in which $\mathbf{X}$ can be retrieved from $\mathbf{X} \cdot \mathbf{Y}$ (see Example 2.2.11). It might be interesting to provide some description of cases (which intuitively are fairly rare) in which it can be done, however, we do not study this problem in this thesis.

A natural generalization of the notion of entropy is that of the topological pressure. Suppose that a subshift $(\mathscr{X}, S)$ has the following "multiplicative convolution" property. There is a measure $\mathbf{Y} \in \mathcal{M}_{\mathscr{X}}$, satisfying $\mathbf{H}(\mathbf{Y}) = 0$, such that

$$\mathbf{Z} \in \mathcal{M}_{\mathscr{X}} \quad \Leftrightarrow \quad \mathbf{Z} = \mathbf{X} \cdot \mathbf{Y} \quad \text{for some stationary process } (\mathbf{X}, \mathbf{Y}). \tag{2.2.3}$$

For example if we deal with the full shift $\mathscr{X} = \{0,1\}^{\mathbb{Z}}$ then clearly we can take $\mathbf{Y} \sim \delta_{\mathbf{1}}$ to obtain all elements of $\mathcal{M}_{\mathscr{X}}$ as described above in (2.2.3). A more elaborate example (fundamental for us!) comes from the theory of $\mathscr{B}$-free systems. It has been proved in [61, 34] (see Theorem 4.1.23 and also our simple proof of this result on page 66) that in this case all members of $\mathcal{M}_{\widetilde{\mathscr{X}_\eta}}$ are of the above form for $\mathbf{Y}$ distributed according to the Mirsky measure ν_η. This triggers the following question:

4. What is the topological pressure of systems satisfying (2.2.3)? Are the corresponding equilibrium measures unique?

In fact, our main motivation behind all these questions comes from the theory of $\mathscr{B}$-free systems. This was also the setting from [61] alluded to above and Questions 1-3 were formulated in this very context. Moreover, Question 4 is just as natural for this class. We have already seen some results corresponding to these problems in the preceeding section. Now, we turn to the abstract setting.

2.2.2 Results: entropy of convolution

Recall that we assume that $(\mathbf{X},\mathbf{Y})$ is a stationary finitely-valued process, $X_i \in \mathbb{R}$, $\mathbf{Y}$ is ergodic, $Y_i \in \{0,1\}$ and $\mathbb{P}(Y_0=1)>0$. For any two finitely-valued stationary processes $\mathbf{Z}$ and $\mathbf{W}$,

$$\mathbf{H}(\mathbf{Z}|\mathbf{W}) = \lim_{n\to\infty}\frac{1}{n}\mathbf{H}\left(Z_{[1,n]} \mid W_{[1,n]}\right) = \mathbf{H}\left(Z_0 \mid Z_{(-\infty,-1]},\mathbf{W}\right)$$

stands for the ***relative entropy rate***. Note that the limit exists due to the subadditivity of $n \mapsto \mathbf{H}\left(Z_{[1,n]} \mid W_{[1,n]}\right)$. The second equality is an easy consequence of the chain rule for Shannon's entropy. The Shannon's entropy chain rule can be used as well to obtain ***the chain rule for the entropy rate***, namely,

$$\mathbf{H}((\mathbf{Z},\mathbf{W})) = \mathbf{H}(\mathbf{Z}|\mathbf{W}) + \mathbf{H}(\mathbf{W}).$$

Since by the very definition $\mathbf{H}(\mathbf{Z}|\mathbf{W}) \le \mathbf{H}(\mathbf{Z})$, if $\mathbf{U}$ is a finitely-valued stationary process such that $\mathbf{H}(\mathbf{U})=0$, then

$$\mathbf{H}((\mathbf{Z},\mathbf{W},\mathbf{U})) = \mathbf{H}((\mathbf{Z},\mathbf{W})), \qquad \mathbf{H}((\mathbf{Z},\mathbf{U})|\mathbf{W}) = \mathbf{H}(\mathbf{Z}\mid\mathbf{W}) = \mathbf{H}(\mathbf{Z}\mid(\mathbf{W},\mathbf{U})).$$

Therefore, if $\mathbf{H}(\mathbf{Y})=0$ (which is the case of our main interest) then

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) = \mathbf{H}(\mathbf{X}\cdot\mathbf{Y}\mid\mathbf{Y}).$$

Furthermore, it seems that in general $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}\mid\mathbf{Y})$ is much easier to handle than $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$ and hence, unlike in Questions 1, 2, 3, 3'., all our main theorems will be expressed in terms of relative entropy rate (with respect to $\mathbf{Y}$).

Let $\mathbf{R}=\mathbf{R}(\mathbf{Y})=(R_i)_{i\in\mathbb{Z}}$ be the ***return process***, i.e. the process of consecutive ***arrival times*** of $\mathbf{Y}$ to the state 1:

$$R_i = \begin{cases} \inf\{j\ge 0 : Y_j=1\}, & i=0,\\ \inf\{j> R_{i-1} : Y_j=1\}, & i\ge 1,\\ \sup\{j< R_{i+1} : Y_j=1\}, & i\le -1. \end{cases} \tag{2.2.4}$$

Our main result provides an explicit formula for the entropy rate of multiplicative convolution.

Theorem 2.2.5 (Answer to Question 1)**.** *Under our standing assumptions, if* $\mathbf{H}(\mathbf{X})>0$ *then*

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}\mid\mathbf{Y}) = \mathbb{P}(Y_0=1)\,\mathbf{H}_{Y_0=1}\left(X_0 \mid X_{\{R_{-1},R_{-2},\cdots\}},\mathbf{Y}\right). \tag{2.2.5}$$

If additionally $\mathbf{X}\amalg\mathbf{Y}$ *then*

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}\mid\mathbf{Y}) = \mathbb{P}(Y_0=1)\,\mathbb{E}_{Y_0=1}\mathbf{H}\left(X_0 \mid X_{\{r_{-1},r_{-2},\cdots\}}\right)|_{r_{-i}=R_{-i}}. \tag{2.2.6}$$

Remark 2.2.6. In order to calculate the integral from the right hand side of (2.2.6) one must take the following steps. Firstly, for almost $\mathbb{P}_{Y_0=1}$ every realization of our return process $\mathbf{R}$ we calculate $\mathbb{E}_{Y_0=1}\mathbf{H}\left(X_0 \mid X_{\{r_{-1},r_{-2},\cdots\}}\right)$, thus obtaining some function $f(r_{(-\infty,-1]})$. Secondly, we find $\mathbb{E}_{Y_0=1}f(R_{(-\infty,-1]})$.

Recall that if $\mathbf{H}(\mathbf{Y})=0$ then $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}\mid\mathbf{Y})=\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$. Therefore, the above theorem gives a formula for $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$, as soon as $\mathbf{H}(\mathbf{Y})=0$. As a consequence, we immediately get the following result.

Corollary 2.2.7 (Answer to Question 2)**.** *Under our standing assumptions, if we assume additionally that* $\mathbf{H}(\mathbf{Y})=0<\mathbf{H}(\mathbf{X})$ *and* $\mathbf{X}\amalg\mathbf{Y}$ *then*

$$\mathbb{P}(Y_0=1)\,\mathbf{H}(\mathbf{X}) \le \mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) \le \mathbb{P}(Y_0=1)\,\mathbf{H}(X_0).$$

If one drops the independence assumption $\mathbf{X}\amalg\mathbf{Y}$ then it might happen that $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})=0$ which complements the answer to Question 2. (Take for example processes $\mathbf{X}=\mathbf{Z}\cdot(\mathbf{1}-\mathbf{W})$ and $\mathbf{Y}=\mathbf{W}$, where $\mathbf{Z}\amalg\mathbf{W}$, $W_i\in\{0,1\}$ and $\mathbf{H}(\mathbf{Z})>0=\mathbf{H}(\mathbf{W})$.)

Beside the explicit formula from Theorem 2.2.5 we obtain the following "drop bound" on $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$ when $\mathbf{X}$ is independent of $\mathbf{Y}$.

Theorem 2.2.8. *Under our standing assumptions, if we assume additionally that* $\mathbf{X} \amalg \mathbf{Y}$ *and* $\mathbf{H}(\mathbf{X}) > 0$*, then*

$$\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) \leq \mathbf{H}(\mathbf{X}) - \mathbb{P}(Y_0 = 1)^2 \, \mathbb{E}_{Y_0=1} \mathbf{H}\left(X_{[1,r_1)} \mid X_{(-\infty,0]}, X_{\{r_1,r_2,\ldots\}}\right)|_{r_i=R_i}. \tag{2.2.7}$$

As a direct consequence of this theorem we obtain the following criterion for the drop of entropy.

Corollary 2.2.9 (Answer to Question 3'.)**.** *Under our standing assumptions, if we assume additionally that* $\mathbf{X} \amalg \mathbf{Y}$ *and* $\mathbf{H}(\mathbf{X}) > 0$ *then*

$$\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) \leq \mathbf{H}(\mathbf{X}) - \mathbb{P}(Y_0 = 1)^2 \sum_{k=1}^{\infty} \mathbb{P}_{Y_0=1}(R_1 = k) \, \mathbf{H}\left(X_{[1,k)} \mid X_{(-\infty,0] \cup [k,\infty)}\right). \tag{2.2.8}$$

Recall now that two discrete random variables X and Y with joint distribution $p_{X,Y}(x,y)$ and margins $p_X(x)$ and $p_Y(y)$ respectively, are ***ε-independent*** if

$$\sum_{x,y} |p_{X,Y}(x,y) - p_X(x) p_Y(y)| < \varepsilon.$$

Furthermore, a stationary finitely valued process $\mathbf{X}$ is ***weak Bernoulli*** (or equivalently ***absolutely regular*** cf. Section B.3.4) if the past and future become ε-independent if separated by a gap g, that is, given $\varepsilon > 0$ there is a gap $g \in \mathbb{N}$ such that for any $k \geq 0$ and $m > 0$, the random vectors $X_{[g,g+m]}$ and $X_{[-k,0]}$ are ε-independent (see [93] page 233). Let us mention that the weak Bernoulli property is stronger than the very weak Bernoulli property (for the definition see [93], page 232). Moreover, a process $\mathbf{X}$ is very weak Bernoulli iff it is isomorphic to some i.i.d. process.

It is well-known that if a process is absolutely regular then the ***double tail σ-field***

$$\mathcal{T}_{double} = \bigcap_{i \geq 0} \sigma\left(X_{(-\infty,-i]}, X_{[i,\infty)}\right)$$

must be trivial. In order to see it, recall that every absolutely regular process must be mixing – even very weak Bernoulli processes are mixing, see Theorem IV.2.1 in [93], page 230. In particular, it is ergodic, i.e. its σ-field of invariant sets is trivial. Furthermore, the property of being weak Bernoulli can be defined in terms of β-mixing coefficients and it corresponds to the convergence $\beta_n \to 0$ (see Section B.3.4). It remains to apply theorem by Berbee, see Theorem B.3.2.

Thus, if $\mathbf{X}$ is stationary and absolutely regular then, due to the continuity of conditional Shannon's entropy (with respect to conditioning), it is clear that for all sufficiently big $k \in \mathbb{N}$,

$$\begin{aligned} \mathbf{H}\left(X_{[1,k)} \mid X_{(-\infty,0]\cup[k,\infty)}\right) &\geq \mathbf{H}\left(X_{k/2} \mid X_{(-\infty,0]\cup[k,\infty)}\right) = \mathbf{H}\left(X_0 \mid X_{(-\infty,-k/2]\cup[k/2,\infty)}\right) \\ &\approx \mathbf{H}(X_0 \mid \mathcal{T}_{double}) = \mathbf{H}(X_0) > 0 \end{aligned}$$

(for the sake of simplicity we assumed that k is even) and thus Corollary 2.2.9 immediately yields the following result.

Corollary 2.2.10. *Apart from our standing assumptions, assume additionally that* $\mathbf{X} \amalg \mathbf{Y}$, $\mathbf{H}(\mathbf{X}) > 0$, $\mathbf{X}$ *is weak Bernoulli and* $\mathbb{P}_{Y_0=1}(R_1 = k) > 0$ *for infinitely many* $k \in \mathbb{N}$*. Then we observe the phenomenon of the drop of entropy,* $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) < \mathbf{H}(\mathbf{X})$*.*

Finally, it is not so hard to come up with the following example.

Example 2.2.11 (Answer to Question 3)**.** Let $(\xi_i)_{i \in \mathbb{Z}}$ be a sequence of i.i.d. random variables such that $\mathbb{P}(\xi_0 = 0) = \mathbb{P}(\xi_0 = 1) = \frac{1}{2}$, an arbitrary (relabeling) 1-1 function $F\colon \{0,1\}^2 \to \{1,2,3,4\}$ and put $X_i = F(\xi_i, \xi_{i+1})$. Furthermore, let $\mathbf{Y}$ be independent of $\mathbf{X}$ and $\mathbf{Y} \sim \frac{1}{2}(\delta_{\mathbf{x}} + \delta_{S\mathbf{x}})$, where $x_{2i} = 0 = 1 - x_{2i+1}$ for $i \in \mathbb{Z}$. Since $\mathbf{X}$ is a Markov chain and F is 1-1, we have $\mathbf{H}(\mathbf{X}) = \mathbf{H}(X_1 \mid X_0) = \mathbf{H}(\xi_1, \xi_2 \mid \xi_0, \xi_1) = \mathbf{H}(\xi_2 \mid \xi_0, \xi_1) = \mathbf{H}(\xi_2) = \log 2$. Moreover, $\mathbb{P}_{Y_0=1}(R_{-1} = 2) = 1$ and therefore by (2.2.6), $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) = \frac{1}{2}\mathbf{H}(X_0 \mid X_{-2}) = \frac{1}{2}\mathbf{H}(X_0) = \log 2$, where we used the fact that X_0 is independent of X_{-2}. Summing it up,

$$\mathbf{H}(\mathbf{X}) = \mathbf{H}(\mathbf{X} \cdot \mathbf{Y}). \tag{2.2.9}$$

We can even strengthen (2.2.9). Note that since F does not attain value 0, we can retrieve **both**, $\mathbf{X}$ and $\mathbf{Y}$ from $\mathbf{X} \cdot \mathbf{Y}$. Indeed, any zero coordinate in $\mathbf{X} \cdot \mathbf{Y}$ immediately determines $\mathbf{Y}$. Furthermore, by the very definition of $\boldsymbol{\xi}$, $\boldsymbol{\xi}$ (and thus $\mathbf{X}$) can be reconstructed as soon as we know odd or even coordinates in $\mathbf{X}$. Hence, as soon as we get $\mathbf{Y}$ from the $\mathbf{X} \cdot \mathbf{Y}$, the process $\mathbf{X}$ is easily found.

Remark 2.2.12. Note that $\mathbf{X}$ from the previous example is very weak Bernoulli (it is easily checked from the very definition that $\mathbf{X}$ is absolutely regular (see Section B.3.4) which implies that $\mathbf{X}$ is very weak Bernoulli), that is, it is isomorphic to some i.i.d. process $\mathbf{X}'$. However, for such process $\mathbf{X}'$, by Corollary 3.2.3, $\mathbf{H}(\mathbf{X}' \cdot \mathbf{Y}) < \mathbf{H}(\mathbf{X}')$ (as soon as $\mathbf{Y} \neq \mathbf{1}$) and thus, in particular, the signal cannot be retrieved. In that sense, the problem of retrieving signal is probabilistic (we care for the distribution of $\mathbf{X}$) and not ergodic (we cannot allow one to take isomorphism).

Remark 2.2.13. In the previous example $\mathbf{Y}$ was periodic, in particular, $\mathbb{P}_{Y_0=1}(R_1 \leq K) = 1$ for some $K \in \mathbb{N}$. It would be interesting to know whether a similar phenomenon is possible with $\mathbb{P}_{Y_0=1}(R_1 > K) > 0$ for every $K \in \mathbb{N}$.

2.2.3 Results: topological pressure

Let $(\mathscr{X}, S)$ be a subshift and $\varphi \colon \mathscr{X} \to \mathbb{R}$ be a **continuous** function which we call a ***potential***. A potential is called ***local*** if it depends only on finitely many coordinates. In our setting one can show that the ***topological pressure***, $\mathscr{P}_{\mathscr{X},\varphi}$ equals (note that we use base 2 in all logarithms and exponentials)

$$\mathscr{P}_{\mathscr{X},\varphi} = \lim_{n\to\infty} \frac{1}{n} \log_2 \sum_{\mathbf{A}\in\mathcal{L}^{(n)}} 2^{\sup_{\mathbf{A}} S_n\varphi(\mathbf{x})},$$

where $S_n = \sum_{i=1}^{n} S^i$. It is well-known that the following ***variational principle (VP)*** holds (see [99], Theorem 4.1),

$$\mathscr{P}_{\mathscr{X},\varphi} = \sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}} \left[\mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})\right]. \tag{2.2.10}$$

Note that the map $\mathbf{X} \to \mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})$ is upper semi-continuous (in the weak topology) and thus there is always some optimal $\mathbf{X}$ attaining the supremum in (2.2.10), called an ***equilibrium measure***. Motivated by the VP, for any subset of invariant measures $\mathcal{N} \subset \mathcal{M}(\mathcal{X}^{\mathbb{Z}})$, we define

$$V_{\mathcal{N},\varphi} = \sup_{\mathbf{X}\in\mathcal{N}} \left[\mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})\right]. \tag{2.2.11}$$

Fix a random stationary process $\nu \sim \mathbf{Y} = (Y_i)_{i\in\mathbb{Z}} \in \{0,1\}^{\mathbb{Z}}$ satisfying $\mathbf{H}(\mathbf{Y}) = 0$ and assume that $\mathbf{X}$ is a **real process**, that is $X_i \in \mathbb{R}$. Inspired by the $\mathscr{B}$-free systems (the reader can think about $\nu \sim \mathbf{Y}$ as about a Mirsky measure), let us consider the family

$$\mathcal{N}_{\mathbf{Y}} = \left\{\mathbf{X} \cdot \mathbf{Y} \mid (\mathbf{X}, \mathbf{Y}) \in \mathcal{M}_{\mathcal{X}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}}}\right\}. \tag{2.2.12}$$

Our aim is to find the solution to the following variational problem:

$$V_{\mathcal{N}_{\mathbf{Y}},\varphi} = \sup_{\mathbf{X}\cdot\mathbf{Y}\in\mathcal{N}_{\mathbf{Y}}} \left[\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) + \mathbb{E}\varphi(\mathbf{X} \cdot \mathbf{Y})\right], \tag{2.2.13}$$

which in case of "multiplicative convolution" spaces (cf. the discussion above Question 4.) coincides with the topological pressure of the system (recall that the hereditary closure of a $\mathscr{B}$-free system is an example of such space, see (2.1.4)).

Let us now present our results. We have three types of theorems. The first one (Theorem 2.2.14) concerns the case in which φ depends only on one coordinate. This is clearly the simplest possible extension in comparison to studying topological entropy. The second result (Theorem 2.2.18) holds for (sufficiently) local potentials and periodic processes. In Theorem 2.2.19 we deal with arbitrary continuous potentials and processes which can be approximated in a certain way by periodic ones. The case when the limit process is itself periodic is here excluded – this is essential for our methods to work (notice that m and ℓ in Theorem 2.2.18 are not completely arbirary).

Theorem 2.2.14 (Potential depending only on one coordinate)**.** *Fix some stationary process* $\mathbf{Y}$*. Suppose that the potential* φ *is local and depends only on the first coordinate, that is* $\varphi(\mathbf{x}) = \varphi(x_0)$*. Then*

$$V_{\mathcal{N}_{\mathbf{Y}},\varphi} = \sup_{\mathbf{X}\cdot\mathbf{Y}\in\mathcal{N}_{\mathbf{Y}}} \left[\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y}\right) + \mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y})\right] = (1-d)\varphi(0) + d\log_2\left(\sum_{x\in\mathcal{X}} 2^{\varphi(x)}\right),$$

where $d = \mathbb{P}\left(Y_0 = 1\right)$ *stands for the density of ones. Furthermore, if* $\mathbf{X}$ *attains the above supremum then* $\mathbf{X}\cdot\mathbf{Y} \sim \mathbf{G}\cdot\mathbf{Y}$*, where* G *is an i.i.d. process such that* $\mathbb{P}\left(G_i = x\right)$ *is proportional to* $2^{\varphi(x)}$ *(so* G_i *is the Gibbs measure associated with* φ*).*

Remark 2.2.15. Note that an application of the above theorem with $\mathbf{Y} = \mathbf{1}$ (that is $d = 1$) yields the corresponding result for the full shift (cf. Remark 2.1.17).

Given a finite subset of real numbers $\mathcal{X}$ (containing zero), a potential $\varphi : \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ and a stationary binary process $\mathbf{Y}$, there is a natural operation

$$\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$$

(where Φ is a certain measurable function defined still on $\mathcal{X}^{\mathbb{Z}}$) which we call a $\mathbf{Y}$***-upgrade of*** $\boldsymbol{\varphi}$ and is used in theorems below (cf. Remark 2.1.10 and recall Theorems 2.1.9 and 2.1.11, where we intentionally omitted the precise definition). We will show now how to construct Φ, still omitting all technicalities, but giving more flavour of what is happening here (for the details and more information we refer the reader to (3.2.24)).

Definition 2.2.16 ($\mathbf{Y}$***-upgrade of a potential*** $\boldsymbol{\varphi}$)**.** Let $\varphi\colon \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ be a continuous potential, where $0 \in \mathcal{X} \subset \mathbb{R}$. In order to give the reader some intuition behind $\mathbf{Y}$-upgrade of φ, we start with a toy example of $\mathbf{Y}$. Afterwards, we generalize it to the periodic case of $\mathbf{Y}$ and at the end we explain briefly how the general case of $\mathbf{Y}$ is treated.

Let $\mathbf{Y}$ be distributed according to $\frac{1}{2}\left(\delta_{(01)^\infty} + \delta_{(10)^\infty}\right)$. Note that then with equal probabilities $1/2$, the associated return (to the state 1) process $\mathbf{R}$ (recall (2.2.4)) is equal to either odd or even integers. In such case we define the $\mathbf{Y}$-upgrade of φ via

$$\Phi(\mathbf{z}) = \frac{1}{2}\left[\varphi(\ldots, z_{-1}, 0, \underbrace{z_0}_{0-coor.}, 0, \ldots) + \varphi(\ldots, 0, z_{-1}, \underbrace{0}_{0-coor.}, z_0, \ldots)\right]. \tag{2.2.14}$$

More generally, take some 0-1 word $w = (w_0, \ldots, w_{\ell-1}) \in \{0,1\}^\ell$ of length $\ell \in \mathbb{N}$ such that $w_0 = 1$. Let $\mathbf{Y}$ be the corresponding w-periodic stationary process (in other words $\mathbf{Y} \sim \frac{1}{\ell}\sum_{i=1}^{\ell} S^i\delta_{w^\infty}$, where $w^\infty \in \{0,1\}^{\mathbb{Z}}$ and $w_i^\infty = w_{i \bmod \ell}$). In that case, the $\mathbf{Y}$-upgrade is given by

$$\Phi = \frac{1}{\ell}\sum_{i=1}^{\ell} S^i \varphi_{\mathbf{r}^{(w)}}, \tag{2.2.15}$$

where the sequence of integers $\mathbf{r}^{(w)}$ is equal to the consecutive positions of ones in w^∞ (in particular, $r_0^{(w)} = 0$ and $r_i^{(w)} = k$ iff $w_k^\infty = 1$) and for any strictly increasing sequence of integers $\mathbf{r}$, $\varphi_{\mathbf{r}}$ is given by

$$\varphi_{\mathbf{r}}(\mathbf{z}) = \varphi\left(\ldots, 0^{r_{-1}-r_{-2}-1}, \underbrace{z_{-1}}_{r_{-1}-coor.}, 0^{r_0-r_{-1}-1}, \underbrace{z_0}_{r_0-coor.}, 0^{r_1-r_0-1}, \underbrace{z_1}_{r_1-coor.}, 0^{r_2-r_1-1}, \ldots\right).$$

In order to get a better grasp on the definition of $\varphi_{\mathbf{r}}$ note that on the right hand side of (2.2.14) the first summand equals to $\varphi_{2\mathbb{Z}}(\mathbf{z})$ and the second one to $\varphi_{2\mathbb{Z}+1}(\mathbf{z})$ and thus (2.2.15) extends the definition given in (2.2.14).

Notice now that (2.2.15) can be rewritten as

$$\Phi = \mathbb{E}\varphi_{\mathbf{R}}, \tag{2.2.16}$$

where $\mathbb{E}$ denotes the Bochner integral and the return process $\mathbf{R}$ is given by (2.2.4). We use formula (2.2.16) to extend the definition of Φ to general processes.

Remark 2.2.17. In the theorems below, given a state space $\mathcal{X} \subset \mathbb{R}$ and a binary process $\mathbf{Y}$, $\varphi(\mathbf{X} \cdot \mathbf{Y})$ must be well-defined for any process $\mathbf{X} \in \mathcal{X}^{\mathbb{Z}}$. In particular, if $\mathcal{X}$ does not contain 0 then we must (somehow artificially) define φ on $\mathcal{X} \cup \{0\}$ (and not just on $\mathcal{X}$).

Recall that a word w is called ***primitive*** if there is no word u and natural number $n \geq 2$ such that $w = u^n$.

Theorem 2.2.18 (Periodic case). *Fix some 0-1 primitive word w of length ℓ containing $m = \#_1(w)$ of ones. Let $\mathbf{Y}$ be the corresponding w-periodic stationary process. If $\varphi : (\mathcal{X} \cup \{0\})^{\mathbb{Z}} \to \mathbb{R}$ is local and depends only on $[0, \ell - 1]$ coordinates then*

$$V_{\mathcal{N}_{\mathbf{Y}},\varphi} = \frac{1}{\ell} \log_2 \left[\sum_{z_{[0,m-1]} \in \mathcal{X}^m} 2^{\ell \Phi(z_{[0,m-1]})} \right],$$

where $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$.

Theorem 2.2.18 leads to the following result (cf. Remark 2.1.14) which constitutes the crucial part of the proof of Theorem 2.1.11 (the formula for the topological pressure for $\mathscr{B}$-free systems).

Theorem 2.2.19 (Weak limits of periodic). *Assume that the sequence of w_n-periodic processes $\mathbf{Y}^{(n)}$, where w_n are finite primitive 0-1 words satisfying $\#_1(w_n) \to \infty$, converges weakly to $\mathbf{Y}$. Then for any continuous potential φ: $(\mathcal{X} \cup \{0\})^{\mathbb{Z}} \to \mathbb{R}$,*

$$V_{\mathcal{N}_{\mathbf{Y}^{(n)}},\varphi} \to \mathbb{P}(Y_0 = 1) \log |\mathcal{X}| + \sup_{\mathbf{z} \in \mathcal{X}^{\mathbb{Z}}} \Phi(\mathbf{z}),$$

where $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$.

2.3 Concentration for m-dependent random variables and Markov chains

2.3.1 Introduction

In this section we establish Bernstein type concentration inequalities for Markov chains and m-dependent sequences. Let us start with recalling the structure of such inequality in the simplest i.i.d. case.

Theorem 2.3.1 (Classical Bernstein inequality). *If $(\xi_i)_i$ is a sequence of i.i.d. real centered random variables such that $\|\xi_i\|_\infty \leq M$ then for $\sigma^2 = \mathbb{E}\xi_i^2$ and any $t > 0$,*

$$\mathbb{P}\left(\left| \sum_{i=1}^{n} \xi_i \right| \geq t \right) \leq 2 \exp\left(-\frac{t^2}{2n\sigma^2 + \frac{2}{3}Mt} \right). \tag{2.3.1}$$

Let us now analyze (slightly informally) the right hand side of (2.3.1). Note that

$$2 \exp\left(-\frac{t^2}{2n\sigma^2 + \frac{2}{3}Mt} \right)$$

as a function of t exhibits two types of behavior: **for "small" t, *the Gaussian one*** (of order $\exp(-ct^2)$ for some $c > 0$), namely,

$$2 \exp\left(-\frac{t^2}{2n\sigma^2} \right) \tag{2.3.2}$$

and **for "large" t**, the ***exponential one*** (of order $\exp(-ct)$ for some $c > 0$), namely

$$2 \exp\left(-\frac{t}{\frac{2}{3}M} \right). \tag{2.3.3}$$

By "large" and "small" t we mean the ranges of t for which one of the terms, $2n\sigma^2$ and $\frac{2}{3}Mt$, "strongly" dominates the other. (From now on, we write (informally) $a \gg b$ to say that a "is much greater" than b.) Clearly, $2n\sigma^2 \gg \frac{2}{3}Mt$ if t is sufficiently small and $2n\sigma^2 \ll \frac{2}{3}Mt$ if t is large enough. The "(2.3.2) part" of Bernstein's inequality is usually called ***the Gaussian part of the Bernstein inequality***. Let us now explain the name.

Assume that η_i are i.i.d. Gaussian random variables with zero mean and variance σ^2, that is, $\eta_i \sim \mathcal{N}(0, \sigma^2)$. It is classical that in such case

$$\sum_{i=1}^n \eta_i \sim \mathcal{N}(0, n\sigma^2)$$

and, moreover,

$$\mathbb{P}\left(\left|\sum_{i=1}^n \eta_i\right| \geq t\right) \approx 2\exp\left(-\frac{t^2}{2n\sigma^2}\right).$$

Thus, if $2n\sigma^2 \gg \frac{2}{3}Mt$ then Theorem 2.3.1 just says that the worst case arises when $\xi_i = \eta_i$, in other words,

$$\mathbb{P}\left(\left|\sum_{i=1}^n \xi_i\right| \geq t\right) \lesssim \mathbb{P}\left(\left|\sum_{i=1}^n \eta_i\right| \geq t\right). \tag{2.3.4}$$

Let us address now the problem of optimality of the Bernstein inequality (2.3.1). By the central limit theorem (CLT),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \quad \Rightarrow \quad \mathcal{N}(0, \sigma^2)$$

which roughly can be rewritten as

$$\sum_{i=1}^n \xi_i \sim \mathcal{N}(0, n\sigma^2).$$

This observation combined with (2.3.4) shows that the Bernstein inequality (2.3.1) is in fact an (asymptotically) optimal concentration inequality (at least when it comes to the Gaussian part). Later on, we will (slightly imprecisely) refer to this fact by saying that the ***Bernstein inequality is optimal***.

Now, let us abandon the i.i.d. setup. Suppose that $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ is an arbitrary real-valued process such that $\sup_i \|X_i\|_\infty < \infty$, $\mathbb{E}X_i = 0$, for which we want to establish a Bernstein-like inequality. Assume additionally that the CLT holds for $\mathbf{X}$, that is,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \quad \Rightarrow \quad \mathcal{N}(0, \sigma_\infty^2) \tag{2.3.5}$$

for some non-negative number $\sigma_\infty^2 \geq 0$ which we call ***the asymptotic variance***. Now, if $\mathbf{X}$ is sufficiently strongly mixing then one should be able to express σ_∞^2 as

$$\sigma_\infty^2 = \lim_{n\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right) = \lim_{n\to\infty} \frac{1}{n}\operatorname{Var}\left(\sum_{i=1}^n X_i\right). \tag{2.3.6}$$

Furthermore, in the vein of classical Bernstein inequality, it is natural to expect that for such processes the following analog of (2.3.1) should hold with an appropriate choice of a constant $C \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2n\sigma_\infty^2 + CMt}\right), \tag{2.3.7}$$

where $M = \sup_i \|X_i\|_\infty$. Note that in order to ensure that (2.3.7) "is able" to reflect the CLT behaviour (2.3.5) (in other words, that $2n\sigma_\infty^2 \gg CMt$ for sufficiently large n), one must insist on C being $o(n)$ as $n \to \infty$. On the other hand, generally, one can allow C to depend on some properties of $\mathbf{X}$ (as soon as $C = o(n)$). This is the case for example for Markov chains, where C depends on the starting point, the transition probability and is of order $\log n$ (for the details see Theorem 2.3.5 below). Similarly to the i.i.d. case, we say (slightly imprecisely) that the ***Bernstein inequality is optimal***, if the inequality from (2.3.7) reflects the CLT behaviour (2.3.5).

In this part we consider the problem of obtaining versions of Bernstein inequality (2.3.7) for both, (stationary) m-dependent sequences and general Markov chains. Recall that a process $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ is m-dependent for $m \in \mathbb{N}$ if

$$X_{(-\infty,k]} \amalg X_{[k+m+1,\infty)}, \qquad \forall_{k\in\mathbb{Z}}. \tag{2.3.8}$$

The class of m-dependent random variables was studied in many papers including [1, 73, 19, 51, 59, 97] but it seems that the problem of optimal Bernstein inequality has not been addressed so far as opposed to the case of the class of Markov chains, where many types of concentration inequalities have been established, see [4, 5, 11, 12, 24, 30, 31, 43, 54, 68, 67, 75, 74, 83, 91, 102]. Let us mention that 1-dependent sequences are strongly related to Markov chains, due to the splitting method (see Section 6.1.11), which splits a Markov chain into 1-dependent blocks. In particular, a version of Bernstein's inequality for 1-dependent processes yields (almost immediately) some for Markov chains, but not vice versa. On the other hand, there is a conjecture (still open, for the details see Section 5.4) which says that every 1-dependent stationary process is in fact a 1-factor of a 1-dependent Markov chain which (if true) would set a nice correspondence between these two classes of processes.

2.3.2 Results: m-dependent processes

Let $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ be a stationary m-dependent sequence (recall (2.3.8)) of bounded and centered random variables. It is easy to check that in this case the ***asymptotic variance*** (cf. (2.3.6)) is given by

$$\sigma_\infty^2 = \lim_{n\to\infty} \frac{1}{n} \mathrm{Var}\,(X_1 + X_2 + \cdots + X_n) = \mathbb{E}X_0^2 + 2\sum_{i=1}^{m} \mathbb{E}X_0X_i. \tag{2.3.9}$$

Since our main result is quite technical, let us postpone its precise formulation to part two of this thesis and now present only its consequence formulated for 1-factors of m-dependent l-Markov chains which are of special interest in the theory of m-dependent sequences (and of Markov chains as well). Recall that a process $\mathbf{X} \in \mathcal{X}^{\mathbb{Z}}$ is a k***-factor*** of $\mathbf{Y} \in \mathcal{Y}^{\mathbb{Z}}$ (cf. (1.2.3)) if there is a function $f\colon \mathcal{X}^k \to \mathcal{Y}$ such that

$$Y_i = f(X_i, X_{i+1}, \dots, X_{i+k-1}).$$

A process $\mathbf{Y}$ is called l***-Markov chain*** if, for any $k \in \mathbb{Z}$, given the present $X_{[k,k+l-1]}$, the future, $X_{(k+l,\infty]}$ is independent of the past $X_{(-\infty,k-1]}$.

Theorem 2.3.2. *Let $X_i = f(Y_i)$, where f is a bounded measurable function and $\mathbf{Y} = (Y_i)$ is a stationary m-dependent l-Markov chain. Then*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) \leq 2(m+1)\exp\left(-\frac{t^2}{c_{m,l}(n+m+l)\sigma_\infty^2 + d_m tM}\right), \tag{2.3.10}$$

where $c_{m,l} = 2(1+\frac{3}{2\log(2m+2l)})^2(m+l)$, $d_{m,l} = \frac{4}{3}(1+\frac{3}{2\log(2m+2l)})(m+l)(m+1)$, $M = \|X_i\|_\infty$ and σ_∞^2 is as in (2.3.9).

Remark 2.3.3. The novelty of Theorem 2.3.2 arises from the use of the asymptotic variance σ_∞^2 in the Gaussian part of the Bernstein inequality instead of $\sigma^2 = \mathbb{E}X_0^2$, that is the variance of a single random variable. In fact, it is an easy task to obtain the version of (2.3.10) with σ_∞^2 replaced by σ^2. What are the relations between σ_∞^2 and σ^2? On the one hand, due to the Schwarz inequality, we always have $\sigma_\infty^2 \leq (m+1)\sigma^2$. On the other hand, it may happen that $\sigma_\infty^2 \ll \sigma^2$, that is, σ_∞^2 can be arbitrarily small compared to σ^2 (in fact, the extreme case $\sigma_\infty^2 = 0 < \sigma^2$ can occur). In the latter case our theorem provides a much sharper inequality than one with σ_∞^2 replaced by σ^2. Moreover, our inequality (2.3.10) is an optimal one (up to constants depending on l and m).

Remark 2.3.4. Note that if we could prove that every m-dependent stationary process is in fact a k-factor (for some $k \in \mathbb{N}$ depending on m) factor of a 1-dependent Markov chain then the above theorem would imply that the optimal (up to constants depending on m) Bernstein inequality holds for **all** m-dependent stationary processes.

2.3.3 Results: Markov chains

In this part we assume that $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ is a Markov chain defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in a measurable space $(\mathcal{X}, \mathcal{B})$ (we require $\mathcal{B}$ to be countably generated), with a transition function $P\colon \mathcal{X} \times \mathcal{B} \to [0, 1]$. Moreover, we assume that $\mathbf{X}$ is **ψ-irreducible**, **aperiodic** and **admits a unique invariant probability measure** π (for an introduction to this notions we refer the reader to Section 6.1). As usual, for any initial distribution μ on $\mathcal{X}$ we write $\mathbb{P}_\mu(\mathbf{X} \in \cdot)$ for the distribution of the chain with X_0 distributed according to the measure μ. In order to shorten the notation we use $\mathbb{P}_x$ instead of $\mathbb{P}_{\delta_x}$, where δ_x denotes the Dirac mass at x.

We say that $\mathbf{X}$ is ***geometrically ergodic*** if there exists a positive number $\rho < 1$ and a real function $G\colon \mathcal{X} \to \mathbb{R}$ belonging to $L_1(\pi)$ such that for every starting point $x \in \mathcal{X}$ and $n \in \mathbb{N}$,

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \le G(x)\rho^n, \tag{2.3.11}$$

where $\|\cdot\|_{TV}$ stands for the total variation norm of a measure and $P^n(\cdot, \cdot)$ is the n-step transition function of the chain.

Our main result is the following. (Below, for convenience sake, we set $\log(\cdot) = \ln(\cdot \vee e)$, where $\ln(\cdot)$ is the natural logarithm.)

Theorem 2.3.5 (Bernstein-like inequality for Markov chains)**.** *Let $\mathbf{X}$ be a geometrically ergodic Markov chain with state space $\mathcal{X}$ and let π be its unique stationary probability measure. Moreover, let $f\colon \mathcal{X} \to \mathbb{R}$ be a bounded measurable function such that $\mathbb{E}_\pi f = 0$. Furthermore, let $x \in \mathcal{X}$. Then we can find constants $K, \tau > 0$ depending only on x and the transition probability $P(\cdot, \cdot)$ such that for all $t > 0$,*

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \le K \exp\left(-\frac{t^2}{32n\sigma^2_{Mrv} + \tau t\|f\|_\infty \log n}\right),$$

where

$$\sigma^2_{Mrv} = \mathrm{Var}_\pi(f(X_0)) + 2\sum_{i=1}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_i)) \tag{2.3.12}$$

denotes the asymptotic variance of the process $(f(X_i))_i$.

Remark 2.3.6. The constants K and τ are explicit and can be found in Theorem 6.2.4. More general versions of Theorem 2.3.5 are available in Theorems 6.2.1 and 6.2.3, where the assumption of boundedness of X_i's is replaced by the integrability in the Orlicz norm.

Now, we make general comments on Theorem 2.3.5, to see how our result fits in a broader picture. Recall the classical Bernstein inequality in the bounded case from Theorem 2.3.1. The CLT for Markov chains (see [21, 80, 77] or Section 6.1.9) guarantees that under the assumptions and notation of Theorem 2.3.5 the sums $\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f(X_i)$ converge in distribution to the normal distribution $\mathcal{N}(0, \sigma^2_{Mrv})$. Thus, the inequality obtained in Theorem 2.3.5 reflects (up to constants) the asymptotic normal behavior of the sums $\frac{1}{\sqrt{n}}\sum f(X_i)$, similarly as the classical Bernstein inequality in the i.i.d. context does. Furthermore, the term $\log n$ which appears in our inequality is necessary. More precisely, one can show that if the following inequality holds for all $t > 0$:

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \le const \cdot \exp\left(-\frac{t^2}{const \cdot n\sigma^2 + const(x) \cdot a_n t\|f\|_\infty}\right) \tag{2.3.13}$$

for some $a_n = o(n)$ and $\sigma \in \mathbb{R}$ ($const$'s stand for some absolute constants whereas $const(x)$ depends only on x and the Markov chain) then one must have $\sigma^2 \ge const \cdot \sigma^2_{Mrv}$. Moreover, it is known that for some geometrically ergodic Markov chains, a_n must grow at least logarithmically with n (see [4], Section 3.3).

Concentration inequalities for Markov chains and processes have been thoroughly studied in the literature, the (non-comprehensive) list of works concerning this topic includes [4, 5, 11, 12, 24, 30, 31, 43, 54, 68, 67, 75, 74, 83, 91, 102]. Some results are devoted to concentration for general functions of the chain (they are usually obtained under various Lipschitz or bounded difference type conditions), others specialize to additive functionals, which are the object of study in our case. Tail inequalities for additive functionals are usually counterparts of Hoeffding or Bernstein inequalities. The former ones do not take into account the variance of the additive functional and are expressed in terms of $\|f\|_\infty$ only. They can be often obtained as special cases of concentration inequalities for general function (see, e.g., [30, 83, 91]). Bernstein type estimates of the form (2.3.13) are considered, e.g., in [4, 5, 11, 12, 24, 31, 43, 68, 67, 75, 74, 83, 102] and use various variance proxies σ^2, which do not necessarily coincide with the limiting variance σ^2_{Mrv}. In the continuous time case, inequalities of Bernstein type for the natural counterpart of the additive functional, involving asymptotic variance have been obtained under certain spectral gap or Lyapunov type conditions in [43, 68]. For discrete time Markov chains, inequalities obtained in [4, 5, 12, 24, 31] by the regeneration method give (2.3.13) (under various types of ergodicity assumptions and with various parameters a_n) with σ^2, which coincides with σ^2_{Mrv} only under the additional assumption of strong aperiodicity of the chain. On the other hand, the articles [75, 74, 91, 102] provide more general results, available for non-necessarily Markovian sequences of random variables, satisfying various types of mixing conditions. The variance proxies σ^2 that are used in these references are close to the asymptotic variance, however in general do not coincide with it. For instance, the inequality obtained in [75], which is valid in particular for geometrically ergodic chains, uses (in our notation) $\sigma^2 = \mathrm{Var}_\pi(f(X_0)) + 2\sum_{i=1}^\infty |\mathrm{Cov}_\pi(f(X_0), f(X_i))|$. Comparing this with (2.3.12), one can see that $\sigma^2_{Mrv} \le \sigma^2$. In fact, one can construct examples when the ratio betweeen the two quantities is arbitrarily large or even $\sigma^2_{Mrv} = 0$ and $\sigma^2 > 0$. The reference [102] provides an inequality for uniformly geometrically ergodic processes, involving a certain implicitly defined variance proxy σ^2_n, which may be bounded from above by σ^2 from [75] or by $\mathrm{Var}_\pi(f(X_0)) + C\|f\|_\infty \mathbb{E}_\pi|f(X_0)|$, where C is a constant depending on the mixing properties of the process. For a fixed process, in the non-degenerate situation, when the asymptotic variance is non-zero, it can be substituted for σ^2_n at the cost of introducing additional multiplicative constants, depending on the chain and the function f.

To the best of our knowledge, Theorem 2.3.5 is therefore the first tail inequality available for general geometrically ergodic Markov chains (not necessarily strongly aperiodic), which (up to universal constants) reflects the correct limiting Gaussian behavior of additive functionals. The problem of obtaining an inequality of this type was posed in [5]. Let us remark that the quantitative investigation of problems related to the Central Limit Theorems for general aperiodic Markov chains seems to be substantially more difficult than for chains which are strongly aperiodic. For instance optimal strong approximation results are still known only in the latter case [76].

Part II

Results and proofs

The aim of this part is to present all our results which are gathered together, sorted thematically and presented in separated, specialized chapters. Each chapter contains (at the very beginning in sections called "Background") all necessary notions, definitions and facts required for full understanding of our results. Then in the next chapters called "The results" we present our main theorems. There is one exception to this rule. Namely, in Chapter 4, Sections 4.1.5 and 4.1.7 which treat about $\mathscr{B}$-free systems, we develop some new theory concerning notions of "density of ones" and Mirsky measure. We justify this irregularity by the fact that these results are either kind of additional or serve just as a tool used for proofs of our main theorems.

Chapter 3

Entropy and topological pressure

The notions of entropy and topological pressure are classical and can be found in almost every information theory book. We recommend [47] for entropy in a more probabilistic approach, [45] Chapter 14 for the ergodic point of view on entropy and [100] Chapter 7 for the basic properties of topological pressure.

3.1 Background

3.1.1 Shannon Entropy

In this part we give a brief summary of definitions and basic facts concerning different kinds of entropy. Unless stated otherwise, all random variables are ***discrete*** (countably valued). If $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ and a set A is such that $\mathbb{P}(A) > 0$ then we define the ***Shannon entropy*** and the ***Shannon conditional entropy*** via

$$\mathbf{H}_A(X) = -\sum_{x \in \mathcal{X}} \mathbb{P}_A(X = x) \log \mathbb{P}_A(X = x), \qquad \mathbf{H}_A(X \mid Y) = \sum_{y \in \mathcal{Y}} \mathbb{P}_A(Y = y)\, \mathbf{H}_{Y=y,A}(X), \quad (3.1.1)$$

respectively. If $\mathbb{P}(A) = 1$ then we write $\mathbf{H}(X)$ and $\mathbf{H}_Y(X)$ instead of $\mathbf{H}_A(X)$ and $\mathbf{H}_A(X \mid Y)$. Sometimes it is convenient to extend the definition of $\mathbf{H}(X \mid Y)$ to $\mathbf{H}(X \mid \mathcal{G})$ where $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$. In order to do so we need to recall the notion of regular conditional distribution. Given random variables $Z \in \mathcal{Z}$, $W \in \mathcal{W}$ and a sub-σ-field $\mathcal{G} \subset \mathcal{F}$ we say that $p_{Z|W}(\cdot,\cdot)$ or $p_{Z|\mathcal{G}}(\cdot,\cdot)$ is a ***regular conditional distribution*** of Z given W or of Z given $\mathcal{G}$, respectively, if the following holds:

- For every $\omega \in \Omega$ ($w \in \mathcal{W}$ resp.), function $p_{Z|\mathcal{G}}(\cdot, \omega)$ ($p_{Z|W}(\cdot, w)$ resp.) is a probability measure on Ω.
- For every measurable $A \subset \Omega$, function $p_{Z|\mathcal{G}}(A, \cdot)$ ($p_{Z|W}(A, \cdot)$ resp.) is measurable. Furthermore, $p_{Z|\mathcal{G}}(A, \cdot) = \mathbb{P}(Z \in A \mid \mathcal{G})$ ($p_{Z|W}(A, W) = \mathbb{P}(Z \in A \mid W)$ resp.).

Now, if X is discrete and there exists a regular conditional distribution $p_{X|\mathcal{G}}$ of X given $\mathcal{G}$, then we put

$$\mathbf{H}(X \mid \mathcal{G}) = \mathbb{E}\mathbf{H}\left(p_{X \mid \mathcal{G}}(\cdot, \omega)\right) = \int \mathbf{H}\left(p_{X \mid \mathcal{G}}(\cdot, \omega)\right) d\mathbb{P}(\omega).$$

One easily checks that if $\mathcal{G} = \sigma(Y)$ (here Y need not be discrete) then $\mathbf{H}(X \mid \mathcal{G}) = \mathbf{H}(X \mid Y) = \mathbb{E}\mathbf{H}\left(p_{X \mid Y}(\cdot, y)\right)_{|y=Y}$, where $p_{X \mid Y}$ is a regular conditional distribution of X given Y. Sometimes slightly informally we write $\mathbf{H}_{Y=y}(X)$ for $\mathbf{H}\left(p_{X \mid Y}(\cdot, y)\right)_{|y=Y}$ and thus $\mathbf{H}(X \mid Y) = \mathbb{E}\left[\mathbf{H}_{Y=y}(X)\right]_{|y=Y}$. One can check that the conditional counterpart of this formula holds, namely,

$$\mathbf{H}(X \mid Y, Z) = \mathbb{E}[\mathbf{H}_{Y=y}(X \mid Z)]_{|y=Y} \quad (3.1.2)$$

for arbitary Y, Z such that the regular conditional distribution $p_{X,Z|Y}$ exists. Note that for any y, $\mathbf{H}_{Y=y}(X \mid Z) = \mathbf{H}\left(X^{(y)} \mid Z^{(y)}\right)$ where $(X^{(y)}, Z^{(y)}) \sim p_{X,Z|Y}(\cdot, y)$. Moreover, we have the following properties:

- **Invariance under relabeling.** If $f\colon \mathcal{X} \to \mathcal{Y}$ is a bijection then $\mathbf{H}(X) = \mathbf{H}(f(X))$ (a similar fact holds in conditional setting).
- **Non-negativity.** We have $\mathbf{H}(X \mid \mathcal{G}) \geq 0$, with equality iff $X \in \mathcal{G}$ (see Proposition 14.18 in [45]).
- **Upper bound.** If $X \in \mathcal{X}$ and $|\mathcal{X}| < \infty$ then $\mathbf{H}(X) \leq \log |\mathcal{X}|$. This follows from the well-known fact that the Shannon entropy (treated as a function on the probability simplex in $\mathbb{R}_+^{|\mathcal{X}|}$) is strictly concave and invariant with respect to permutations of its arguments.
- **Monotonicity in conditioning.** If $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ are sub-σ-algebras then we have $\mathbf{H}(X \mid \mathcal{G}) \leq \mathbf{H}(X \mid \mathcal{H})$. Furthermore, $\mathbf{H}(X \mid \mathcal{G}) = \mathbf{H}(X)$ iff X is independent of $\mathcal{G}$ (see Proposition 14.18 in [45]).
- **Continuity.** If $\mathcal{G}_n \searrow \mathcal{G}$ or $\mathcal{G}_n \nearrow \mathcal{G}$ then $\mathbf{H}(X \mid \mathcal{G}_n) \nearrow \mathbf{H}(X \mid \mathcal{G})$ or $\mathbf{H}(X \mid \mathcal{G}_n) \searrow \mathbf{H}(X \mid \mathcal{G})$, respectively (see Theorem 14.28 in [45]).
- **Chain rule.** We have $\mathbf{H}(X, Y \mid \mathcal{G}) = \mathbf{H}(X \mid \mathcal{G}) + \mathbf{H}(Y \mid \mathcal{G}, X)$ (see Proposition 14.16 in [45]). In particular, for any function f, $\mathbf{H}(X, f(X) \mid \mathcal{G}) = \mathbf{H}(X \mid \mathcal{G})$.
- **Decrease of entropy under quantization of argument.** For any function f, $\mathbf{H}(f(X) \mid \mathcal{G}) \leq \mathbf{H}(X \mid \mathcal{G})$ (see Proposition 14.18 in [45]).

3.1.2 Entropy rate

Fix stationary processes $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ and $\mathbf{Y} = (Y_i)_{i\in\mathbb{Z}}$ with at most countable alphabets such that $\mathbf{H}(X_0), \mathbf{H}(Y_0) < \infty$ and recall that in such a case ***the entropy rate*** and ***the relative entropy rate*** are given by

$$\mathbf{H}(\mathbf{X}) = \lim_{n\to\infty} \frac{1}{n}\mathbf{H}\left(X_{[0,n-1]}\right) = \inf_{n\in\mathbb{N}} \frac{1}{n}\mathbf{H}\left(X_{[0,n-1]}\right) = \lim_{n\to\infty} \mathbf{H}\left(X_n \mid X_{[0,n-1]}\right) = \mathbf{H}\left(X_0 \mid X_{(-\infty,-1]}\right)$$

and

$$\mathbf{H}(\mathbf{X} \mid \mathbf{Y}) = \lim_{n\to\infty} \frac{1}{n}\mathbf{H}\left(X_{[0,n-1]} \mid Y_{[0,n-1]}\right) = \inf_{n\in\mathbb{N}} \frac{1}{n}\mathbf{H}\left(X_{[0,n-1]} \mid Y_{[0,n-1]}\right),$$

respectively. Moreover, the following well-known facts hold:

- **Affinity.** Let $\mathbf{X}^{(i)}$, $i \in \mathbb{N}$, be a family of stationary processes and a random variable θ be independent of $\mathbf{X}^{(i)}$ for any i, where $\mathbb{P}(\theta = i) = p_i$, $\sum_i p_i = 1$. If $\mathbf{H}(\theta) < \infty$ then $\mathbf{H}\left(\mathbf{X}^{(\theta)}\right) = \sum_i p_i \mathbf{H}\left(\mathbf{X}^{(i)}\right)$. Indeed, it is enough to notice that $\mathbf{H}\left(X^{(\theta)}_{[0,n]} \mid \theta\right) \leq \mathbf{H}\left(X^{(\theta)}_{[0,n]}\right) \leq \mathbf{H}\left(X^{(\theta)}_{[0,n]} \mid \theta\right) + \mathbf{H}(\theta)$ and recollect the definition of the Shannon entropy. Alternatively, see Theorem 14.25 in [45].
- **Upper semi-continuity.** Suppose that $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}^{(\infty)}$ and, for any $n \in \mathbb{N} \cup \{\infty\}$ and $i \in \mathbb{N}$, $X_i^{(n)} \in \mathcal{X}$ with $|\mathcal{X}| < \infty$. Then $\limsup_{n\to\infty} \mathbf{H}\left(\mathbf{X}^{(n)}\right) \leq \mathbf{H}(\mathbf{X})$. This immediately follows from a combination of the following facts.
  - We have $\mathbf{H}(\mathbf{X}) = \inf_n \mathbf{H}\left(X_{[1,n]}\right)/n$.
  - The function $\mathbf{X} \to \mathbf{H}\left(X_{[1,n]}\right)/n$ is continuous (in the weak topology).
  - The infimum of continuous functions is upper semi-continuous.
- **Chain rule.** We have $\mathbf{H}((\mathbf{X}, \mathbf{Y})) = \mathbf{H}(\mathbf{Y}) + \mathbf{H}\left(X_0 \mid X_{(-\infty,-1]},\ \mathbf{Y}\right)$. In particular,

  $$\mathbf{H}(\mathbf{X}) \vee \mathbf{H}(\mathbf{Y}) \leq \mathbf{H}((\mathbf{X}, \mathbf{Y})) \leq \mathbf{H}(\mathbf{X}) + \mathbf{H}(\mathbf{Y}).$$

  This is an easy consequence of the following decomposition:

  $$\mathbf{H}\left(X_{[1,n]}, Y_{[1,n]}\right) = \mathbf{H}\left(Y_{[1,n]}\right) + \mathbf{H}\left(X_{[1,n]} \mid Y_{[1,n]}\right) = \mathbf{H}\left(Y_{[1,n]}\right) + \sum_{i=1}^{n} \mathbf{H}\left(X_i \mid Y_{[1,n]}, X_{[1,i)}\right)$$

  and the fact that $\mathbf{H}\left(X_i \mid Y_{[1,n]}, X_{[1,i)}\right) = \mathbf{H}\left(X_0 \mid Y_{[1-i,n-i]}, X_{[1-i,0)}\right) \xrightarrow{n\to\infty} \mathbf{H}\left(X_0 \mid X_{[-\infty,0)},\ \mathbf{Y}\right)$ uniformly in i satisfying $\log n \leq i \leq n - \log n$.

Let us say a few words about zero entropy processes. Firstly, by the chain rule, if $\mathbf{H}(\mathbf{Y}) = 0$ then $\mathbf{H}(\mathbf{X} \mid \mathbf{Y}) = \mathbf{H}(\mathbf{X})$. Secondly, if $\mathcal{T}_{past} = \bigcap_{i\geq 0} \sigma\left(Y_{(-\infty,-i]}\right)$ stands for the past tail σ-field of $\mathbf{Y}$ then the following conditions are equivalent:

- $\mathbf{H}(\mathbf{Y}) = 0$.
- For every $i \in \mathbb{N}$, $\mathbf{H}\left(Y_{[-i,i]} \mid \mathcal{T}_{past}\right) = 0$.
- For every $i \in \mathbb{N}$, $Y_{[-i,i]} \in \mathcal{T}_{past}$.

Intuitively, $\mathbf{Y}$ is a zero entropy process iff the knowledge of $\mathcal{T}_{past}$ determines the whole process $\mathbf{Y}$. Let us explain the only non-trivial implication in the above list of conditions, namely, the fact that $\mathbf{H}(\mathbf{Y}) = 0$ implies $\mathbf{H}\left(Y_{[-i,i]} \mid \mathcal{T}_{past}\right) = 0$. Note that by the subadditivity of Shannon's entropy and the stationarity of $\mathbf{Y}$, it is enough to prove that $\mathbf{H}(Y_0 \mid \mathcal{T}_{past}) = 0$. Using the continuity property (in conditioning) of the conditional Shannon entropy, we easily reduce this problem to demonstration of the following fact: $\mathbf{H}\left(Y_0 \mid Y_{(-\infty,-k]}\right) = 0$ for any $k \in \mathbb{N}$. But this is clear because due to the monotonicity property, the chain rule and the stationarity of $\mathbf{Y}$ we have $\mathbf{H}\left(Y_0 \mid Y_{(-\infty,-k]}\right) \leq \mathbf{H}\left(Y_{[-k+1,0]} \mid Y_{(-\infty,-k]}\right) = k\mathbf{H}(\mathbf{Y}) = 0$. More information on tail σ-algebras is included in Appendix B.

3.1.3 Topological entropy and measures of maximal entropy

Let $(\mathscr{X}, S)$, where $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$, be a subshift over a finite alphabet $\mathcal{X}$. Recall that the ***topological entropy of*** $(\mathscr{X}, S)$ is given by

$$\mathbf{H}_{\mathscr{X}} = \mathbf{H} = \lim_{n\to\infty} \frac{1}{n} \left|\mathcal{L}^{(n)}\right| = \inf_{n\in\mathbb{N}} \frac{1}{n} \left|\mathcal{L}^{(n)}\right|, \tag{3.1.3}$$

where $\mathcal{L}^{(n)}$ consists of words of length n which appear in $\mathscr{X}$. Moreover, by the variational principle,

$$\mathbf{H} = \sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}} \mathbf{H}(\mathbf{X}). \tag{3.1.4}$$

Any measure which attains the supremum in (3.1.4) is called a ***measure of maximal entropy***. Furthermore, we say that dynamical system $(\mathscr{X}, S)$ is ***intrinsically ergodic*** if there is exactly one measure of maximal entropy. Note that the measure of maximal entropy always exists due to the upper semi-continuity of the entropy rate.

3.1.4 Topological pressure and equilibrium measures

The notion of the topological pressure is a natural extension of that of topological entropy.

Let $(\mathscr{X}, S)$, where $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$, be a subshift over a finite alphabet $\mathcal{X} = \{x_1, \ldots, x_k\}$ for some $k \in \mathbb{N}$. Furthermore, let $\varphi\colon \mathscr{X} \to \mathbb{R}$ be a continuous function which we call a ***potential***. A potential is called ***local*** if it depends only on finitely many coordinates. Recall that the ***topological pressure of*** $(\mathscr{X}, S)$ is given by

$$\mathscr{P}_{\mathscr{X},\varphi} = \lim_{n\to\infty} \frac{1}{n} \log_2 \sum_{\mathbf{A}_i\in\mathscr{P}_n} 2^{\sup_{\mathbf{A}_i} S_n\varphi(\mathbf{x})},$$

where $\mathscr{P}_n = \{[x_1], \ldots, [x_k]\}_0^{n-1}$ and, for any partition $\mathcal{C}$ of $\mathscr{X}$, $\mathcal{C}_i^j = \bigvee_{k=i}^j S^{-k}\mathcal{C}$ and $S_n = \sum_{i=0}^{n-1} S^i$. In particular, if $\varphi = 0$ then we recover the definition of the topological entropy. Moreover, the following ***variational principle*** holds:

$$\mathcal{P}_{\mathscr{X},\varphi} = \sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}} \left[\mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})\right]. \tag{3.1.5}$$

Note that $\mathbf{X} \to \mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})$ is upper semi-continuous (recall that we assume that $\mathcal{X}$ is finite and that then $\mathbf{X} \to \mathbf{H}(\mathbf{X})$ itself is upper semi-continuous), thus, there is always an optimal process $\mathbf{X}$ attaining the supremum in (3.1.5). Such $\mathbf{X}$ (in fact, its distribution) is called an ***equilibrium measure***.

3.1.5 Periodic processes

Let u, v, w be finite 0-1 words. Then $u \cdot v$ for the ***concatenation*** of u and v whereas $|w|$ stands for the ***length of*** w. Let $|w| = n$ and denote by w^∞ the sequence given by $w_i^\infty = w_{i \bmod n}$ for $i \in \mathbb{Z}$. We say that $\mathbf{Y}$ is w***-periodic*** if

$$\mathbb{P}\left(Y_{[0,n-1]} = w^{(i)}\right) = \frac{1}{n}, \qquad \mathbb{P}\left(Y_{[nd,n(d+1))]} = Y_{[0,n-1]}\right) = 1$$

for all $d \in \mathbb{Z}$, where $w^{(i)} = w[i, n-1] \cdot w[0, i-1]$ stands for the i***'th cyclic shift of*** w. In other words, $\mathbf{Y}$ is a start randomization of deterministic process w^∞.

If $\mathbf{Y}$ is w-periodic then by the very definition of the entropy rate, $\mathbf{H}(\mathbf{Y}) = 0$. Moreover, clearly, $\mathbf{Y}$ is a $|w|$-Markov chain.

3.1.6 Disjointness of processes by Furstenberg

Recall that stationary finitely-valued processes $\mathbf{X}$ and $\mathbf{Y}$ are ***disjoint in the sense of Furstenberg*** (or ***absolutely disjoint***) if there is a unique stationary coupling of $\mathbf{X}$ and $\mathbf{Y}$, namely, the independent one (see the celebrated paper [41]). This notion still plays one of the main roles in ergodic theory and is one of the most important concepts used in the field of dynamical systems.

It is well-known that if $\mathbf{H}(\mathbf{X}), \mathbf{H}(\mathbf{Y}) > 0$ then $\mathbf{X}$ and $\mathbf{Y}$ are **not** absolutely disjoint (this result goes back to Furstenberg [41], Theorem I.1; it is also a consequence of Sinai's and Orstein's theorems). On the other hand, if $\mathbf{H}(\mathbf{Y}) = 0$ and $\mathbf{X}$ has trivial past tail $\mathcal{T}_{past} = \bigcap_{n \geq 0} \sigma\left(X_{(\infty,-n]}\right)$ then the only possible joining is the independent one (Furstenberg [41] notices that this is an interpretation of a result attributed to Pinsker by Rokhlin [89], see the discussion in [41] following Theorem I.2). We recall that more information on tail σ-algebras is included in Appendix B.

3.2 Results

3.2.1 Notation and basic assumptions

In this part we assume the following:

$$\begin{aligned} &1. \text{ Stochastic processes } \mathbf{X} = (X_i)_{i\in\mathbb{Z}} \text{ and } \mathbf{Y} = (Y_i)_{i\in\mathbb{Z}} \text{ have } \textbf{finite real} \text{ alphabets.} \\ &2. \text{ The process } (\mathbf{X},\mathbf{Y}) = ((X_i,Y_i))_{i\in\mathbb{Z}} \text{ is stationary.} \\ &3. \ \mathbf{Y} \text{ is ergodic and, for every } i \in \mathbb{Z}, \text{ we have } Y_i \in \{0,1\}. \\ &4. \text{ In order to avoid some degenerate cases, } 0 < \mathbb{P}(Y_0 = 1) < 1. \end{aligned} \tag{3.2.1}$$

It is essential for us to introduce the ***return process*** $\mathbf{R} = (R_i)_{i\in\mathbb{Z}}$ given by the return times of $\mathbf{Y}$ to the state 1,

$$R_i = \begin{cases} \inf\{j \geq 0 : Y_j = 1\}, & i = 0, \\ \inf\{j > R_{i-1} : Y_j = 1\}, & i \geq 1, \\ \sup\{j < R_{i+1} : Y_j = 1\}, & i \leq -1. \end{cases}$$

Note that all R_i's are well defined due to the ergodicity of $\mathbf{Y}$. For basic properties of $\mathbf{R}$ we refer to Appendix A. The aim of this section is to explore the quantity $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$. Let us recall that if one additionally assumes that $\mathbf{H}(\mathbf{Y}) = 0$ (this is the case in our motivation example, where $\mathbf{Y}$ corresponds to the Mirsky measure ν_η, see Section 4.1.7) then $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) = \mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y})$. It turns out that in general $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y})$ is slightly easier to handle and thus our main results are stated in terms of $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y})$ instead of $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$.

3.2.2 Entropy of multiplicative convolution

Our main result gives an explicit formula for $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y})$.

Theorem 3.2.1. *Under our standing conditions* (3.2.1), *we have*

$$\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y}\mid\mathbf{Y}\right)=\mathbb{P}\left(Y_0=1\right)\mathbf{H}_{Y_0=1}\left(X_0\mid X_{\{R_{-1},R_{-2},\cdots\}},\mathbf{Y}\right).$$

If additionally $\mathbf{X}\amalg\mathbf{Y}$ *then*

$$\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y}\mid\mathbf{Y}\right)=\mathbb{P}\left(Y_0=1\right)\mathbb{E}_{Y_0=1}\mathbf{H}\left(X_0\mid X_{\{r_{-1},r_{-2},\cdots\}}\right)|_{r_{-i}=R_{-i}}.$$

Remark 3.2.2. The integral $\mathbb{E}_{Y_0=1}\mathbf{H}\left(X_0\mid X_{\{r_{-1},r_{-2},\cdots\}}\right)|_{r_{-i}=R_{-i}}$ should be understood as follows. Firstly, for any sequence $\mathbf{r}$ (in fact, for any realization of $\mathbf{R}$) we calculate $f(\mathbf{r}):=\mathbf{H}\left(X_0\mid X_{\{r_{-1},r_{-2},\cdots\}}\right)$. Secondly, we find $\mathbb{E}f(\mathbf{R})$.

Proof. Using the chain rule and the stationarity of $\mathbf{Y}$ we get

$$\mathbf{H}\left(M_{[0,n]}\mid Y_{[0,n]}\right)=\sum_{i=0}^{n}\mathbf{H}\left(M_i\mid Y_{[0,n]},M_{[0,i-1]}\right)=\sum_{i=0}^{n}\sum_{j=0}^{1}\mathbb{P}\left(Y_i=j\right)\mathbf{H}_{Y_i=j}\left(M_i\mid Y_{[0,n]},M_{[0,i-1]}\right)$$

$$=\mathbb{P}\left(Y_0=0\right)\sum_{i=0}^{n}\mathbf{H}_{Y_i=0}\left(M_i\mid Y_{[0,n]},M_{[0,i-1]}\right)+\mathbb{P}\left(Y_0=1\right)\sum_{i=0}^{n}\mathbf{H}_{Y_i=1}\left(M_i\mid Y_{[0,n]},M_{[0,i-1]}\right).$$

Clearly, if $Y_i=0$ then $M_i=0$ and $\mathbf{H}_{Y_i=0}\left(M_i\mid Y_{[0,n]},M_{[0,i-1]}\right)=0$ whereas if $Y_i=1$ then $M_i=X_i$. Therefore,

$$\mathbf{H}\left(M_{[0,n]}\mid Y_{[0,n]}\right)=\mathbb{P}\left(Y_0=1\right)\sum_{i=0}^{n}\mathbf{H}_{Y_i=1}\left(X_i\mid Y_{[0,n]},M_{[0,i-1]}\right).$$

Now, the stationarity of $(\mathbf{X},\mathbf{Y})$ implies that

$$\mathbf{H}\left(M_{[0,n]}\mid Y_{[0,n]}\right)=\mathbb{P}\left(Y_0=1\right)\sum_{i=0}^{n}\mathbf{H}_{Y_0=1}\left(X_0\mid Y_{[-i,n-i]},M_{[-i,-1]}\right).$$

Since

$$(Y_{[-i,n-i]},M_{[-i,-1]})=(Y_{[-i,n-i]},0^{R_{-S_{-i}}-1},X_{R_{-S_{-i}}},0^{R_{-S_{-i}+1}-R_{-S_{-i}}-1},X_{R_{-S_{-i}+1}},\dots,X_{R_{-1}},0^{R_{-1}-1}),$$

where $S_{-i}=\sum_{k=-i}^{-1}Y_k$, provides the same amount of information as $(Y_{[-i,n-i]},X_{R_{[-S_{-i},-1]}})$, that is there exists a bijection between these vectors, using the invariance under relabelling, we arrive at

$$\mathbf{H}\left(\mathbf{M}\mid\mathbf{Y}\right)=\lim_{n\to\infty}\frac{1}{n}\mathbf{H}\left(M_{[0,n]}\mid Y_{[0,n]}\right)=\mathbb{P}\left(Y_0=1\right)\lim_{n\to\infty}\frac{1}{n}\sum_{i=\log n}^{n-\log n}\mathbf{H}_{Y_0=1}\left(X_0\mid Y_{[-i,n-i]},X_{R_{[-S_{-i},-1]}}\right).$$

(Note that in the above sum we restricted our attention to i satisfying $\log n\le i\le n-\log n$. We can do this because all summands are bounded by $\mathbf{H}\left(X_0\right)$ and we normalize the sum by $1/n$.) It is remains to notice that, by the ergodicity of $\mathbf{Y}$, $S_{-i}\to\infty$ a.s. as $i\to\infty$ and

$$\mathbf{H}_{Y_0=1}\left(X_0\mid Y_{[-i,n-i]},X_{\{R_{-1},R_{-2},\dots,R_{-S_i}\}}\right)\xrightarrow{n\to\infty}\mathbf{H}_{Y_0=1}\left(X_0\mid\mathbf{Y},X_{\{R_{-1},R_{-2},\cdots\}}\right),$$

uniformly in i satisfying $\log n\le i\le n-\log n$.

The independent case easily follows from the dependent one. Namely, using the definition of conditional entropy (recall (3.1.2)),

$$\mathbf{H}_{Y_0=1}\left(X_0\mid\mathbf{Y},X_{\{R_{-1},R_{-2},\cdots\}}\right)=\int\mathbf{H}_{\mathbf{Y}=\mathbf{y}}\left(X_0\mid X_{\{R_{-1},R_{-2},\cdots\}}\right)d\mu_{\mathbf{Y}}(\mathbf{y})$$

where $\mu_{\mathbf{Y}}=\mathcal{L}_{\mathbb{P}_{Y_0=1}}(\mathbf{Y})$. Clearly,

$$\mathbf{H}_{\mathbf{Y}=\mathbf{y}}\left(X_0\mid X_{\{R_{-1},R_{-2},\cdots\}}\right)=\mathbf{H}_{\mathbf{Y}=\mathbf{y}}\left(X_0\mid X_{\{r_{-1},r_{-2},\cdots\}}\right)=\mathbf{H}\left(X_0\mid X_{\{r_{-1},r_{-2},\cdots\}}\right),$$

where in the last inequality we have used $\mathbf{X}\amalg\mathbf{Y}$ (the sequence (r_i) as usually stands for the support of realization $\mathbf{y}$ of $\mathbf{Y}$). ■

Using standard properties of entropy, Theorem 3.2.1 immediately implies the following bounds on $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y})$.

Corollary 3.2.3 (Bounds). *Under our standing assumptions* (3.2.1),

$$\mathbb{P}(Y_0 = 1)\, \mathbf{H}_{Y_0=1}\left(X_0 \mid X_{(-\infty,-1]}, \mathbf{Y}\right) \leq \mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) \leq \mathbb{P}(Y_0 = 1)\, \mathbf{H}_{Y_0=1}(X_0), \tag{3.2.2}$$

which, under additional condition $\mathbf{X} \amalg \mathbf{Y}$, *simplifies to*

$$\mathbb{P}(Y_0 = 1)\, \mathbf{H}(\mathbf{X}) \leq \mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) \leq \mathbb{P}(Y_0 = 1)\, \mathbf{H}(X_0). \tag{3.2.3}$$

Remark 3.2.4. Note that the right inequality in (3.2.3) goes to zero as soon as $\mathbb{P}(Y_0 = 1) \to 0$. In particular, if we keep $\mathbf{X}$ fixed, $\mathbf{H}(\mathbf{Y}) = 0$ and $\mathbf{Y}$ has sufficiently small density of ones $\mathbb{P}(Y_0 = 1)$ then $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) < \mathbf{H}(\mathbf{X})$ and we cannot retrieve $\mathbf{X}$ from $\mathbf{X} \cdot \mathbf{Y}$.

One may wonder when the bounds given in (3.2.2) are attained and when they are strict. In order to check this, for simplicity's sake, we assume that $\mathbf{X} \amalg \mathbf{Y}$ and thus consider (3.2.3). In that case it turns out that the lower bound is attained on the class of exchangeable processes $\mathbf{X}$, the upper bound on i.i.d. processes and for non-trivial Markov chains we have both inequalities strict.

Corollary 3.2.5 (Lower bound attained). *Suppose additionally to* (3.2.1) *that* $\mathbf{X}$ *is exchangeable and* $\mathbf{X} \amalg \mathbf{Y}$. *Then* $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) = \mathbb{P}(Y_0 = 1)\, \mathbf{H}(\mathbf{X})$.

Proof. It follows from the exchangeability of $\mathbf{X}$ that for any negative distinct times r_{-i}, $i \in \mathbb{N}$,

$$\mathbf{H}\left(X_0 \mid X_{\{r_{-1}, r_{-2}, \dots\}}\right) = \mathbf{H}\left(X_0 \mid X_{\{-1,-2,\dots\}}\right) = \mathbf{H}(\mathbf{X}).$$

It remains to use Theorem 3.2.1. ■

Corollary 3.2.6 (Upper bound attained). *Suppose additionally to* (3.2.1) *that* $\mathbf{X}$ *is i.i.d. and* $\mathbf{X} \amalg \mathbf{Y}$. *Then* $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) = \mathbb{P}(Y_0 = 1)\, \mathbf{H}(X_0)$.

Proof. It is enough to recall that every i.i.d. process $\mathbf{X}$ is exchangeable and use the previous corollary. ■

Corollary 3.2.7 (Strict bounds). *Suppose additionally to* (3.2.1) *that* $\mathbf{X}$ *is a Markov chain and* $\mathbf{X} \amalg \mathbf{Y}$. *Then*

$$\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) = \mathbb{P}(Y_0 = 1) \sum_{k=1}^{\infty} \mathbb{P}_{Y_0=1}(R_1 = k)\, \mathbf{H}(X_k \mid X_0).$$

Proof. It is enough to use Theorem 3.2.1 and recall that a process $\mathbf{X}$ is a Markov chain if, for every time $i \in \mathbb{Z}$, conditionally on X_i, $X_{(-\infty,i-1]}$ is independent of $X_{[i+1,\infty)}$. ■

Given a stationary process $\mathbf{Y}$, we define the corresponding set of convolution measures:

$$\mathcal{N}_{\mathbf{Y}} = \{\mathbf{X} \cdot \mathbf{Y}' \mid (\mathbf{X}, \mathbf{Y}') \in \mathcal{M}_{\mathcal{X}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}}},\ \mathbf{Y}' \sim \mathbf{Y}\}. \tag{3.2.4}$$

Remark 3.2.8. Beware that for simplicity's sake, further on, (sligtly imprecisely) we denote the members of $\mathcal{N}_{\mathbf{Y}}$ by $\mathbf{X} \cdot \mathbf{Y}$ (we just suppress the prim in $\mathbf{Y}'$). It does not affect the correctness of our proofs and results because, in fact, we do care about the distributions and not their particular realizations.

Now, we will show that there is exactly one member of $\mathcal{N}_{\mathbf{Y}}$ which realizes the maximal entropy

$$\sup_{\mathbf{Z} \in \mathcal{N}_{\mathbf{Y}}} \mathbf{H}(\mathbf{Z}).$$

Thus, if a subshift $(\mathscr{X}, S)$ has the property that for some process $\mathbf{Y}$, $\mathcal{N}_{\mathbf{Y}} = \mathcal{M}_{\mathscr{X}}$ then $(\mathscr{X}, S)$ must be intrinsically ergodic. For example, this is the case for $\mathscr{B}$-free shifts (see Theorem 4.1.23 below). In particular, Corollary 3.2.9 gives a new proof of intrinsic ergodicity of $\mathscr{B}$-free systems (proved before in various settings in [85, 61, 34]).

Corollary 3.2.9 (Convolution intrinsic ergodicity)**.** *Let $\mathcal{N}_{\mathbf{Y}}$ be as in (3.2.4). Then there is exactly one convolution measure $\mathbf{X} \cdot \mathbf{Y} \in \mathcal{N}_{\mathbf{Y}}$ which realizes*

$$\sup_{\mathbf{X}\cdot\mathbf{Y}\in\mathcal{N}_{\mathbf{Y}}} \mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}\right).$$

Moreover, it is given by $\mathbf{B}\cdot\mathbf{Y}$, where $\mathbf{B}$ is an i.i.d. process independent of $\mathbf{Y}$ such that $\mathbb{P}\left(B_i = x\right) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$.

Remark 3.2.10. A similar result (with almost the same proof) holds for the topological pressure for potentials depending only on one coordinate, see Theorem 3.2.21.

Remark 3.2.11. For the proof of Corollary 3.2.9, we will need a standard argument concerning conditional mean values. Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. If $Z \amalg \mathcal{G}$ and $W \in \mathcal{G}$ then

$$\mathbb{E}\left(g(Z,W)|\mathcal{G}\right) = G(W), \qquad G(w) = \mathbb{E}g(Z,w).$$

Proof of Corollary 3.2.9. By Theorem 3.2.1, we have

$$\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}\right) = \mathbb{P}\left(Y_0=1\right)\mathbf{H}_{Y_0=1}\left(X_0 \mid X_{R_{(-\infty,-1]}}, \mathbf{Y}\right) \le \mathbb{P}\left(Y_0=1\right)\mathbf{H}_{Y_0=1}\left(X_0\right) \le \mathbb{P}\left(Y_0=1\right)\log|\mathcal{X}|.$$

Note that these inequalities become equalities iff (conditionally on $Y_0 = 1$)

$$X_0 \amalg \left(X_{R_{(-\infty,-1]}}, \mathbf{Y}\right) \tag{3.2.5}$$

and

$$\mathbb{P}_{Y_0=1}\left(X_i = x\right) = 1/|\mathcal{X}| \qquad \forall_{x\in\mathcal{X}}. \tag{3.2.6}$$

Clearly, $\mathbf{B}$ satisfies these conditions.

Now, we show that if a process $\mathbf{X}$ has properties (3.2.5) and (3.2.6) then $\mathbf{X}\cdot\mathbf{Y} \sim \mathbf{B}\cdot\mathbf{Y}$ under $\mathbb{P}_{Y_0=1}$. Clearly, it is enough to show that for any local (i.e. depending on finitely many coordinates) potential f, we have

$$\mathbb{E}f(\mathbf{X}_-\cdot\mathbf{Y}_-) = \mathbb{E}f(\mathbf{B}_-\cdot\mathbf{Y}_-), \tag{3.2.7}$$

where for any sequence $\mathbf{x}$, $\mathbf{x}_-$ stands for $x_{(-\infty,0]}$. Before we show (3.2.7), we need some auxiliary concepts. For every continuous function f depending on coordinates $(-\infty,0]$ and an increasing sequence of non-positive integers $\mathbf{r}_- = (\ldots, r_{-1}, r_0)$, we define

$$f_{\mathbf{r}_-}(z_{(-\infty,0]}) = f\left(\ldots, \underbrace{z_{-1}}_{r_{-1}-coord.}, 0^{r_0-r_{-1}-1}, \underbrace{z_0}_{r_0-coord.}, 0^{-r_0}\right), \tag{3.2.8}$$

where we interpret 0^k for $k \le 0$ as the empty word. It is easily checked that $(\mathbf{r}_-, \mathbf{z}_-) \to f_{\mathbf{r}_-}(\mathbf{z}_-)$ is a measurable map ($\mathbf{z}_- \to f_{\mathbf{r}_-}(\mathbf{z}_-)$ is continuous and $\mathbf{r}_- \to f_{\mathbf{r}_-}(\mathbf{z}_-)$ is measurable, cf. upcoming Remark 3.2.23). Moreover, given a word $x \in \mathcal{X}^k$, let

$$f^{(x)}(\mathbf{z}_-) = f(\mathbf{z}_-, x). \tag{3.2.9}$$

As usual, $\mathbf{M} = \mathbf{X}\cdot\mathbf{Y}$.

We prove (3.2.7) in two steps. Firstly, we show it holds on $\{Y_0 = 1\}$ and then on $\{Y_0 = 0\}$. In fact, on $\{Y_0 = 1\}$ we show a little stronger formula than (3.2.7), namely

$$\mathbb{E}_{Y_0=1}f(\mathbf{X}_-\cdot\mathbf{Y}_-)g(\mathbf{Y}) = \mathbb{E}_{Y_0=1}f(\mathbf{B}_-\cdot\mathbf{Y}_-)g(\mathbf{Y}) \tag{3.2.10}$$

for any bounded measurable $g\colon \{0,1\}^{\mathbb{Z}} \to \mathbb{R}$, This enhanced version enables us to use an inductive argument (on the number of coordinates f depends on) and is used in the proof of (3.2.7) on $\{Y_0 = 0\}$

as well. To see (3.2.10), notice that by using first (3.2.8) and then the tower property of conditional mean value, we obtain

$$
\begin{aligned}
E &:= \mathbb{E}_{Y_0=1} f(\mathbf{M}_-) g(\mathbf{Y}) = \mathbb{E}_{Y_0=1} g(\mathbf{Y}) f_{(R_{(-\infty,-1]},0)} \left(X_{R_{(-\infty,-1]}}, X_0 \right) \\
&= \mathbb{E}_{Y_0=1} g(\mathbf{Y}) \mathbb{E}_{Y_0=1} \left(f_{(R_{(-\infty,-1]},0)} \left(X_{R_{(-\infty,-1]}}, X_0 \right) \mid \mathbf{Y}, X_{R_{(-\infty,-1]}} \right).
\end{aligned}
$$

Now, due to Remark 3.2.11 (with $Z := X_0$, $W := (R_{(-\infty,-1]}, X_{R_{(-\infty,-1]}})$ and $\mathcal{G} := (\mathbf{Y}, X_{R_{(-\infty,-1]}}))$ and conditions (3.2.5) and (3.2.6), we arrive at Now, due to Remark 3.2.11 (with $Z := X_0$, $W := (R_{(-\infty,-1]}, X_{R_{(-\infty,-1]}})$ and $\mathcal{G} := \sigma(\mathbf{Y}, X_{R_{(-\infty,-1]}})$), we have

$$
\begin{aligned}
&\mathbb{E}_{Y_0=1}(f_{(R_{(-\infty,-1]},0)}(X_{R_{(-\infty,-1]}}, X_0) | \mathbf{Y}, X_{R_{(-\infty,-1]}}) \\
&\qquad = \sum_{x \in \mathcal{X}} \mathbb{P}_{Y_0=1}(X_0 = x) f_{(R_{(-\infty,-1]},0)}(X_{R_{(-\infty,-1]}}, x) = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} f_{(R_{(-\infty,-1]},0)}(X_{R_{(-\infty,-1]}}, x),
\end{aligned}
$$

where the last step follows from (3.2.5) and (3.2.6). Therefore, using our introduced notation (3.2.9), we conclude that

$$
|\mathcal{X}| E = \sum_{x \in \mathcal{X}} \mathbb{E}_{Y_0=1} g(\mathbf{Y}) f^{(x)}_{R_{(-\infty,-1]}} \left(X_{R_{(-\infty,-1]}} \right).
$$

Furthermore,

$$
|\mathcal{X}| E = \sum_{x \in \mathcal{X}} \mathbb{E}_{Y_0=1} g(\mathbf{Y}) f^{(0^{-R_{-1}-1}x)} \left(M_{(-\infty,R_{-1}]} \right) = \sum_{x \in \mathcal{X}} \sum_{k=1}^{\infty} \mathbb{E}_{Y_0=1} g(\mathbf{Y}) f^{(0^{k-1}x)} \left(M_{(-\infty,-k]} \right) \mathbb{1}_{R_{-1}=-k}.
$$

Since $(\mathbf{X}, \mathbf{Y})$ is stationary,

$$
\begin{aligned}
|\mathcal{X}| E &= \sum_{x \in \mathcal{X}} \sum_{k=1}^{\infty} \mathbb{E}_{Y_{-k}=1} g(\mathbf{Y}) f^{(0^{k-1}x)} \left(M_{(-\infty,-k]} \right) \mathbb{1}_{R_{-1}=-k, Y_0=1} \\
&= \sum_{x \in \mathcal{X}} \sum_{k=1}^{\infty} \mathbb{E}_{Y_0=1} g(S^k \mathbf{Y}) f^{(0^{k-1}x)} \left(\mathbf{M}_- \right) \mathbb{1}_{R_1=k}.
\end{aligned}
$$

(Let us mention that in order to see that the last equality holds notice that $S^k\{R_{-1} = -k, Y_0 = 1\} = \{Y_0 = 1, R_1 = k\}$). Summing it up, we have shown that for any f and g,

$$
\begin{aligned}
\mathbb{E}_{Y_0=1} f(\mathbf{M}_-) g(\mathbf{Y}) &= \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \sum_{k=1}^{\infty} \mathbb{E}_{Y_0=1} f^{(0^{k-1}x)} \left(\mathbf{M}_- \right) \left[g(S^k \mathbf{Y}) \mathbb{1}_{R_1=k} \right] \\
&= \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \sum_{k=1}^{\infty} \mathbb{E}_{Y_0=1} f_{k,x} \left(\mathbf{M}_- \right) g_k(\mathbf{Y})
\end{aligned}
$$

for some function $f_{k,x}$, g_k. Note that for any arbitrary k and x, $f_{k,x}$ depends on a strictly smaller number of coordinates than f does. Thus, an easy inductive argument gives (3.2.10) whenever f is local. Now standard approximation arguments imply that (3.2.10) holds for all bounded measurable functions f depending on $(-\infty, 0]$ coordinates.

It remains to check what happens on $\{Y_0 = 0\}$, namely, whether

$$
\mathbb{E}_{Y_0=0} f(\mathbf{M}_-) = \mathbb{E}_{Y_0=0} f(\mathbf{B}_- \cdot \mathbf{Y}_-),
$$

where once more $\mathbf{M} = \mathbf{X} \cdot \mathbf{Y}$. Using (3.2.10) and the stationarity of $(\mathbf{X}, \mathbf{Y})$, we obtain

$$
\begin{aligned}
\mathbb{E}_{Y_0=0} f(\mathbf{M}_-) &= \sum_{k=1}^{\infty} \mathbb{E}_{Y_0=0} f^{(0^k)} \left(M_{(-\infty,-k]} \right) \mathbb{1}_{R_{-1}=-k} = \sum_{k=1}^{\infty} \mathbb{E}_{Y_{-k}=1} f^{(0^k)} \left(M_{(-\infty,-k]} \right) \mathbb{1}_{R_{-1}=-k, Y_0=1} \\
&= \sum_{k=1}^{\infty} \mathbb{E}_{Y_0=1} f^{(0^k)} \left(\mathbf{M}_- \right) \mathbb{1}_{R_1=k} \overset{(3.2.10)}{=} \sum_{k=1}^{\infty} \mathbb{E}_{Y_0=1} f^{(0^k)} \left(\mathbf{B}_- \cdot \mathbf{Y}_- \right) \mathbb{1}_{R_1=k} = \mathbb{E}_{Y_0=0} f(\mathbf{B}_- \cdot \mathbf{Y}_-).
\end{aligned}
$$

■

3.2.3 When a lost signal cannot be retrieved?

Now, we turn to the phenomenon of the entropy drop. Our main result is as follows.

Theorem 3.2.12. *Assume additionally to* (3.2.1) *that* $\mathbf{X} \amalg \mathbf{Y}$. *Then*

$$\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}\right) \leq \mathbf{H}\left(\mathbf{X}\right) - \mathbb{P}\left(Y_0 = 1\right)^2 \mathbb{E}_{Y_0=1}\mathbf{H}\left(X_{[1,r_1)} \mid X_{(-\infty,0]}, X_{\{r_1,r_2,\ldots\}}\right)|_{r_i=R_i}.$$

We divide the proof of the above theorem into three steps. We start with a technical Lemma 3.2.13, which sets up the ground for further analysis. The proof of Theorem 3.2.12 is then concluded in Lemma 3.2.14.

Proof of Theorem 3.2.12

Suppose that (3.2.1) holds and we have $\mathbf{X} \amalg \mathbf{Y}$.

Lemma 3.2.13. *We have*

$$\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}\right) = \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\mathbb{1}_{S_n>0}\mathbf{H}\left(X_{r_0}, X_{r_1}, \ldots, X_{r_{s_n-1}}\right)_{|r_i=R_i, s_n=S_n},$$

where $S_n = \sum_{i=0}^n Y_i$.

Proof. Since for any $k \in \mathbb{Z}$, on the event $Y_k = 0$, we have $M_k \equiv 0$, it follows that

$$\mathbf{H}\left(M_{[0,n]} \mid Y_{[0,n]}\right) = \mathbb{P}\left(S_n > 0\right)\sum_{y_{[0,n]}}\mathbb{P}_{S_n>0}\left(Y_{[0,n]} = y_{[0,n]}\right)\mathbf{H}_{Y_{[0,n]}=y_{[0,n]}}\left(M_{[0,n]}\right).$$

Moreover, if $s_n = \sum_{i=0}^n y_i > 0$ then (by the indepedence of $\mathbf{X}$ and $\mathbf{Y}$)

$$\mathbb{P}_{Y_{[0,n]}=y_{[0,n]}}\left(M_{[0,n]} = m_{[0,n]}\right) = \mathbb{P}\left(X_{r_0} = m_{r_0}, \ldots, X_{r_{s_n-1}} = m_{r_{s_n-1}}\right),$$

whenever $m_{[0,n]}$ and $y_{[0,n]}$ are such that $y_i = 0$ implies $m_i = 0$. Hence,

$$\mathbf{H}_{Y_{[0,n]}=y_{[0,n]}}\left(M_{[0,n]}\right) = \mathbf{H}\left(X_{r_0}, \ldots, X_{r_{s_n-1}}\right),$$

which results in

$$\begin{aligned}\mathbf{H}\left(M_{[0,n]} \mid Y_{[0,n]}\right) &= \mathbb{P}\left(S_n > 0\right)\mathbb{E}_{S_n>0}\mathbf{H}\left(X_{r_0}, \ldots, X_{r_{s_n-1}}\right)_{|r_i=R_i, s_n=S_n}\\ &= \mathbb{E}\mathbb{1}_{S_n>0}\mathbf{H}\left(X_{r_0}, \ldots, X_{r_{s_n-1}}\right)_{|r_i=R_i, s_n=S_n}.\end{aligned}$$

This completes the proof. ∎

Notice now that

$$\frac{1}{n}\mathbf{H}\left(X_{r_0}, \ldots, X_{r_{s_n-1}}\right) = \frac{1}{n}\mathbf{H}\left(X_{[0,n]}\right) - \frac{1}{n}\mathbf{H}\left(X_{[0,n]\setminus\{r_0,\ldots,r_{s_n-1}\}} \mid X_{r_0}, \ldots, X_{r_{s_n-1}}\right),$$

$\lim_{n\to\infty}\frac{1}{n}\mathbf{H}\left(X_{[0,n]}\right) = \mathbf{H}\left(\mathbf{X}\right)$ and that (by the ergodicity of $\mathbf{Y}$) we have $\mathbb{1}_{S_n>0} \to 1$ a.s. Thus, in order to conclude the proof of Theorem 3.2.12, it remains to find $\lim\limits_{n\to\infty}\mathbb{E}\mathbb{1}_{S_n>0}H(n,\mathbf{R})$ where

$$H(n,\mathbf{r}) = \frac{1}{n}\mathbf{H}\left(X_{[0,n]\setminus\{r_0,\ldots,r_{s_n-1}\}} \mid X_{r_0}, \ldots, X_{r_{s_n-1}}\right), \quad \mathbf{r} = (r_i)_{i\in\mathbb{Z}}.$$

Clearly,

$$\mathbb{E}\mathbb{1}_{S_n>0}H(n,\mathbf{R}) \geq \mathbb{P}\left(Y_0 = 1\right)\mathbb{E}_{Y_0=1}H(n,\mathbf{R})$$

and $H(n,\mathbf{R})$ is bounded, so if we show that under $\mathbb{P}_{Y_0=1}$, we have

$$\lim_{n\to\infty}H(n,\mathbf{R}) = \mathbb{P}(Y_0 = 1)\mathbb{E}_{Y_0=1}\mathbf{H}\left(X_{[1,r_1-1]} \mid X_{(-\infty,0]}, X_{\{r_1,r_2,\ldots\}}\right)|_{r_i=R_i} \tag{3.2.11}$$

then the proof is concluded. This will be done in the following lemma using the chain rule and Maker's ergodic theorem.

Lemma 3.2.14. *Equality* (3.2.11) *holds under* $\mathbb{P}_{Y_0=1}$.

Proof. Fix $\mathbf{y}$ and $n \in \mathbb{N}$. By the chain rule, we get

$$nH(n,\mathbf{r}) = \underbrace{\mathbf{H}\left(X_{[0,r_0-1]} \mid X_{\{r_0,\dots,r_{s_n-1}\}}\right)}_{\Sigma_1(n)} + \underbrace{\mathbf{H}\left((X_{[r_{s_n-1}+1,n]} \mid X_{r_{s_n-1}}\right)}_{\Sigma_3(n)}$$
$$+ \underbrace{\sum_{i=0}^{s_n-2} \mathbf{H}\left(X_{[r_i+1,r_{i+1}-1]} \mid X_{[0,r_i]}, X_{\{r_{i+1},\dots,r_{s_n-1}\}}\right).}_{\Sigma_2(s_n-1)}$$

We will deal first with the summands $\Sigma_1(n)$ and $\Sigma_3(n)$. Clearly,

$$\frac{1}{n}\Sigma_1(n) \le \frac{1}{n}\mathbf{H}\left(X_{[0,r_0-1]}\right) \le \frac{r_0}{n} H(X_0) \to 0$$

when $n \to \infty$. Since $s_n = s_{r_{s_n-1}}$, $\frac{s_t}{t} \to \mathbb{P}(Y_0 = 1) > 0$ as $\mathbb{N} \ni t \to \infty$ (by the ergodicity of $\mathbf{Y}$) and $r_{s_n-1} \to \infty$, it follows that

$$\frac{\Sigma_3(n)}{n} \le \frac{n - r_{s_n-1}}{n} H(X_0) = \left(1 - \frac{r_{s_n-1}}{s_{r_{s_n-1}}} \cdot \frac{s_n}{n}\right) H(X_0) \to 0.$$

In order to deal with $\Sigma_2(s_n - 1)$, notice that

$$\frac{1}{n}\Sigma_2(s_n-1) = \frac{s_n}{n}\frac{1}{s_n}\Sigma_2(s_n-1).$$

Because of $\frac{s_n}{n} \to \mathbb{P}(Y_0 = 1)$, to conclude the proof, it suffices to show that

$$\lim_{n\to\infty} \frac{1}{n}\Sigma_2(n) = \mathbf{H}\left(X_{[1,r_1-1]} \mid X_{(-\infty,0]}, X_{\{r_1,r_2,\dots\}}\right)|_{r_i=R_i}.$$

Using the stationarity of $\mathbf{X}$, for $t_i = r_i - r_{i-1}$, we obtain

$$\begin{aligned}\Sigma_2(n) &= \sum_{i=0}^{n-1} \mathbf{H}\left(X_{[r_i+1,r_{i+1}-1]} \mid X_{[0,r_i]}, X_{\{r_{i+1},\dots,r_n\}}\right)\\ &= \sum_{i=0}^{n-1} \mathbf{H}\left(X_{[1,t_{i+1}-1]} \mid X_{[-r_i,0]}, X_{\{t_{i+1},\dots,t_{i+1}+\dots+t_n\}}\right).\end{aligned}$$

We would like to apply Maker's ergodic theorem (see (A.2.5)) to study the above sum. However, we cannot do it directly due to the term $X_{[-r_i,0]}$ appearing in the conditional entropies. This obstacle will be overcome by estimating each summand from below and above.

Fix $k \in \mathbb{N}$. Then for every i such that $r_i \ge k$ and for every $j \in \mathbb{N}$, we have

$$H_{\infty,j}\left(t_{[i+1,\infty)}\right) \quad \le \quad \mathbf{H}\left(X_{[1,t_{i+1}-1]} \mid X_{[-r_i,0]}, X_{\{t_{i+1},\dots,t_{i+1}+\dots t_{i+j}\}}\right) \quad \le \quad H_{k,j}\left(t_{[i+1,\infty)}\right), \tag{3.2.12}$$

where $H_{k,j}(t_{i+1}, t_{i+2}, \dots) = \mathbf{H}\left(X_{[1,t_{i+1}-1]} \mid X_{(-k,0]}, X_{\{t_{i+1},\dots,t_{i+1}+\dots t_{i+j}\}}\right)$ for $k \in \mathbb{Z} \cup \{\infty\}$. Clearly,

$$H_{k,j}(t_1, t_2, \dots) \xrightarrow{j\to\infty} H_k(t_1, t_2, \dots) := \mathbf{H}\left(X_{[1,t_1-1]} \mid X_{(-k,0]}, X_{\{t_1,t_1+t_2,\dots\}}\right).$$

Recall that $\mathbf{T} = (T_i)_{i\in\mathbb{Z}}$, where $T_i = R_i - R_{i-1}$, is the inter-arrival process. Furthermore, $\mathbf{T}$ is stationary and ergodic under $\mathbb{P}_{Y_0=1}$ (for more information on this process see Appendix A). By the entropy chain rule and Kac's lemma,

$$\sup_{k,j\in\mathbb{N}} H_{k,j}(T_{[1,\infty)}) \le \mathbf{H}(X_0)\, T_1 \in L_1(\mathbb{P}_{Y_0=1}). \tag{3.2.13}$$

Therefore, Maker's ergodic theorem implies that, for every $k \in \mathbb{N} \cup \{\infty\}$, $\mathbb{P}_{Y_0=1}$ a.s., we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} H_{k,n-i}\left(t_{i+1}, t_{i+2}, \dots\right) \to \mathbb{E}_{Y_0=1} H_k\left(T_1, T_2, \dots\right). \tag{3.2.14}$$

Using (3.2.12), it follows from the definition of Σ_2 (and the chain rule) that

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1} H_{\infty,n-i}(t_{i+1}, t_{i+2}, \dots) \leq \frac{1}{n}\Sigma_2(n) &\leq \frac{t_1+\dots+t_k}{n} H(X_0) + \frac{1}{n}\sum_{i=k}^{n-1} H_{k,n-i}(t_{i+1}, t_{i+2}, \dots) \\ &\leq \frac{t_1+\dots+t_k}{n} H(X_0) + \frac{1}{n}\sum_{i=0}^{n-1} H_{k,n-i}(t_{i+1}, t_{i+2}, \dots),\end{aligned}$$

with $\frac{t_1+\dots+t_k}{n} H(X_0) \xrightarrow{n\to\infty} 0$. Thus, due to (3.2.14),

$$\mathbb{E}_{Y_0=1} H_\infty\left(T_1, T_2, \dots\right) \leq \lim_{n\to\infty} \frac{1}{n}\Sigma_2(n) \leq \mathbb{E}_{Y_0=1} H_k\left(T_1, T_2, \dots\right).$$

Notice that $H_k \to H_\infty$ as $k \to \infty$. Hence, combining (3.2.13) and the bounded convergence theorem, we obtain

$$\lim_{n\to\infty} \frac{1}{n}\Sigma_2(n) = \mathbb{E}_{Y_0=1} H_\infty\left(T_1, T_2, \dots\right)$$

$\mathbb{P}_{Y_0=1}$ a.s. which is exactly (3.2.3). ■

A consequence of Theorem 3.2.12. When $\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y}\right) < \mathbf{H}\left(\mathbf{X}\right)$?

Theorem 3.2.12 immediately yields the following corollary.

Corollary 3.2.15 (Drop of entropy). *Assume additionally to* (3.2.1) *that* $\mathbf{X} \amalg \mathbf{Y}$. *Then*

$$\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}\right) \leq \mathbf{H}\left(\mathbf{X}\right) - \mathbb{P}\left(Y_0 = 1\right)^2 \sum_{k=1}^{\infty} \mathbb{P}_{Y_0=1}\left(R_1 = k\right) \mathbf{H}\left(X_{[1,k)} \mid X_{(-\infty,0]\cup[k,\infty)}\right).$$

Note that this corollary immediately allows us to create a criterion for $\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y}\right) < \mathbf{H}\left(\mathbf{X}\right)$ (assuming that $\mathbf{H}\left(\mathbf{X}\right) > 0$). For example, it is enough to ensure that the following conditions hold:

- For infinitely many $k \in \mathbb{N}$, $\mathbb{P}_{Y_0=1}\left(R_1 = k\right) > 0$.
- For some $k_0 \in \mathbb{N}$, $\mathbf{H}\left(X_{[-k_0,k_0]} \mid \mathcal{T}_{double}\right) > 0$, where $\mathcal{T}_{double} := \bigcap_{i\geq 0} \sigma\left(X_{(-\infty,-i]\cup[i,\infty)}\right)$.

In order to see this it is enough to notice that if $\mathbf{H}\left(X_{[-k_0,k_0]} \mid \mathcal{T}_{double}\right) > 0$ then (by the continuity in conditioning), for some sufficiently big $K \in \mathbb{N}$, $\mathbf{H}\left(X_{[-k_0,k_0]} \mid X_{(-\infty,-K]\cup(K,\infty]}\right) > 0$. Thus, by the stationarity of $\mathbf{X}$, for $k \geq 10\max(K, k_0)$, $\mathbf{H}\left(X_{[1,k)} \mid X_{(-\infty,0]\cup[k,\infty)}\right) \geq \mathbf{H}\left(X_{[-k_0,k_0)} \mid X_{(-\infty,-K]\cup[K,\infty)}\right) > 0$.

Note that the second requirement is met if $\mathcal{T}_{double} = \mathcal{T}_{past} = \bigcap_{i\geq 0} \sigma\left(X_{(-\infty,-0]}\right)$ (in the sense of measure algebras). Indeed, in such a case, we can take $k_0 = 0$ and then $\mathbf{H}\left(X_0 \mid \mathcal{T}_{double}\right) = \mathbf{H}\left(X_0 \mid \mathcal{T}_{past}\right) > 0$.

In particular, we can observe the phenomenon of the drop of entropy $\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y}\right) < \mathbf{H}\left(\mathbf{X}\right)$ if for infinitely many $k \in \mathbb{N}$, $\mathbb{P}_{Y_0=1}\left(R_1 = k\right) > 0$ and any of the following conditions holds.

- $\mathbf{X}$ is weak Bernoulli, that is $\beta_n(\mathbf{X}) \to 0$ (for the definition and properties of β-coefficients, see Section B.3.4). In this case $\mathcal{T}_{double}$ is trivial (cf. Theorem B.3.2).
- $\mathbf{X}$ is a Markov chain such that $\mathbf{H}\left(\mathbf{X}\right) > 0$. Here $\mathcal{T}_{double} = \mathcal{T}_{tail}$ (see (B.3.3)).
- $\mathbf{X}$ is a non-trivial exchangeable process. Here $\mathcal{T}_{double} = \mathcal{T}_{tail}$ (see (B.3.1)).
- $\mathbf{X}$ is a non-trivial ergodic process and $\beta_n(\mathbf{X}) < 1$ for some $n \in \mathbb{N}$. Here $\mathcal{T}_{double} = \mathcal{T}_{tail}$ (see Theorem B.3.2).

Note that the last condition contains the first one which we included in our list because of the outstanding role played by the weak Bernoulli processes in ergodic and probability theories.

Thus, in some cases of interest we observe the phenomenon of the drop of entropy under multiplicative convolution. However, we already know by Corollary 3.2.3 that if $\mathbf{X} \amalg \mathbf{Y}$ then $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) \geq \mathbb{P}(Y_0 = 1)\, \mathbf{H}(\mathbf{X})$. In other words the drop is controlled by the quantity $\mathbb{P}(Y_0 = 1)$. The situation changes completely if we allow some dependence between $\mathbf{X}$ and $\mathbf{Y}$: it can happen that

$$\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) = 0 = \mathbf{H}(\mathbf{Y}) < \mathbf{H}(\mathbf{X}). \tag{3.2.15}$$

Example 3.2.16. Take a stationary process $(\mathbf{X}, \mathbf{Y})$ and notice that $\mathbf{X}$ can be viewed in two separate parts, namely, on support of $\mathbf{Y}$ and on support of $\mathbf{1} - \mathbf{Y}$. Heuristically, when $\mathbf{X}$ is multiplied by $\mathbf{Y}$, the part of $\mathbf{X}$ on the support of $\mathbf{Y}$ remains unchaged, whereas the part of $\mathbf{X}$ on the support of $\mathbf{1} - \mathbf{Y}$ dies. Therefore, if $\mathbf{X}$ is such that its entropy is zero on the support of $\mathbf{Y}$ and positive on the support of $\mathbf{1} - \mathbf{Y}$ then we have (3.2.15).

To make this argument precise, for every joining $(\mathbf{Z}, \mathbf{W}, \mathbf{U})$, where $\mathbf{U}$ is binary, consider

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{Z} + (\mathbf{1} - \mathbf{U})\mathbf{W}. \tag{3.2.16}$$

Notice that $(\mathbf{A}, \mathbf{U})$ is stationary and every stationary process $(\mathbf{X}, \mathbf{Y})$ can be realized as $(\mathbf{A}, \mathbf{U})$, just by taking $\mathbf{W} = \mathbf{Z} = \mathbf{X}$, $\mathbf{U} = \mathbf{Y}$. Now, if we assume that $\mathbf{W} \amalg \mathbf{U} \amalg \mathbf{Z}$ and $\mathbf{H}(\mathbf{Z}) = \mathbf{H}(\mathbf{U}) = 0 < \mathbf{H}(\mathbf{W})$ and $\mathbf{U} \neq \mathbf{1}$ then by the sub-additivity of entropy rate

$$\mathbf{H}(\mathbf{A} \cdot \mathbf{U}) = \mathbf{H}(\mathbf{U} \cdot \mathbf{Z}) \leq \mathbf{H}(\mathbf{U}) + \mathbf{H}(\mathbf{Z}) = 0 \tag{3.2.17}$$

and by Corollary 3.2.3,

$$\mathbf{H}(\mathbf{A}) = \mathbf{H}(\mathbf{A} \mid \mathbf{Z}, \mathbf{U}) = \mathbf{H}((\mathbf{1} - \mathbf{U})\mathbf{W} \mid \mathbf{Z}, \mathbf{U}) = \mathbf{H}((\mathbf{1} - \mathbf{U})\mathbf{W} \mid \mathbf{U}) \overset{\text{Coro.3.2.3}}{\geq} \mathbb{P}(U_0 = 0)\, \mathbf{H}(\mathbf{W}) > 0$$

(to see the second equality use the definition of the conditional entropy rate and then apply relabelling invariance).

At the end let us look at the quantity $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y})$ in the light of ergodic theory. Up to the end of this section we will denote by $\mu \overset{ind.}{*} \nu$ the distribution of $\mathbf{X} \cdot \mathbf{Y}$ where $\mathbf{X} \sim \mu$ and $\mathbf{Y} \sim \nu$ are independent. Thus, $\mu \overset{ind.}{*} \nu = M(\mu \otimes \nu)$ where $M(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$. Recall that $\mathbf{H}(\mu)$ stands for Sinai-Kolmogorov entropy of μ and we have $\mathbf{H}(\mu) = \mathbf{H}(\mathbf{Z})$ for any $\mathbf{Z} \sim \mu$. Before we start, let us recall a well-known technical fact.

Remark 3.2.17. Let $(H_1, \langle \cdot, \cdot \rangle_1), (H_2, , \langle \cdot, \cdot \rangle_2)$ be Hilbert spaces. Recall that the tensor product of H_1 and H_2 is the pair (H, φ), where H is a Hilbert space equipped with scalar product $\langle \cdot, \cdot \rangle$ and $\varphi \colon H_1 \times H_2 \to H$ is a bilinear mapping satisfying two conditions:

1. the closed linear span of vectors of the form $\varphi(x, y)$ is equal to H;
2. $\langle \varphi(x_1, y_1), \varphi(x_2, y_2) \rangle = \langle x_1, x_2 \rangle_1 \langle y_1, y_2 \rangle_2$ for any $x_1, x_2 \in H_1$ and y_1, y_2 in H_2 (in particular, $\|\phi(x, y)\| = \|x\|_{H_1} \|y\|_{H_2}$).

Usually, $\varphi(x, y)$ is denoted by $x \otimes y$, whereas H by $H_1 \otimes H_2$.

Suppose now that $G_2 \subset H_2$ is a closed subspace of H_2 and for some $x \in H_1$ and $y \in H_2$, $x \otimes y \in H_1 \otimes G_2$, where $x \neq 0$. Then necessarily

$$y \in G_2. \tag{3.2.18}$$

Indeed, let $y = y_0 + y_0'$, with $y_0 \in G_2$ and $y_0' \in G_2^{\perp}$. By our assumption this implies that $H_1 \otimes G_2' \ni x \otimes y_0' = x \otimes (y - y_0) \in H_1 \otimes G_2$. But $H_1 \otimes G_2$ and $H_1 \otimes G_2'$ are orthogonal (firstly, approximate arbitrary elements of these spaces using property 1 and then combine property 2 with the bilinearity of tensor product). Hence $x \otimes y_0' = 0$, $\|x \otimes y_0'\| = \|x\|_{H_1} \|y_0'\|_{H_2} = 0$ and since $\|x\|_{H_1} \neq 0$ we must have $y_0' = 0$.

Let us apply this result in the special case where for $i \in \{1,2\}$, $H_i = L_2(X_i, \mathcal{B}_i, \mu_i)$ and $(X_i, \mathcal{B}_i, \mu_i)$ are standard Borel probability spaces. Define $\varphi(f_1, f_2) = f_1 \cdot f_2$ and note that then due to Fubini's theorem we can assume that $H = H_1 \otimes H_2 = L_2(X_1 \times X_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_1 \otimes \mu_2)$. If for $i \in \{1,2\}$, $\mathcal{C}_i$'s are sub-σ-algebras of $\mathcal{B}_i$'s then $G_i = L_2(X_i, \mathcal{C}_i, \mu_i)$'s are closed subspaces of H_i's. Therefore, by (3.2.18), if $B_i \in \mathcal{B}_i$ then

$$B_1 \times B_2 \overset{\mu_1 \otimes \mu_2}{\in} \mathcal{C}_1 \otimes \mathcal{C}_2 \quad \Leftrightarrow \quad \mathbb{1}_{B_1 \times B_2} = \mathbb{1}_{B_1}\mathbb{1}_{B_2} \in G_1 \otimes G_2 \quad \Leftrightarrow \quad \forall_i\ \mathbb{1}_{B_i} \in G_i \quad \Leftrightarrow \quad \forall_i\ B_i \overset{\mu_i}{\in} \mathcal{C}_i$$

and thus

$$B_1 \times B_2 \overset{\mu_1 \otimes \mu_2}{\in} \mathcal{C}_1 \otimes \mathcal{C}_2 \qquad \Leftrightarrow \qquad B_1 \overset{\mu_1}{\in} \mathcal{C}_1,\ B_1 \overset{\mu_2}{\in} \mathcal{C}_2. \tag{3.2.19}$$

By Corollary 3.2.3, if $\mathbf{X} \amalg \mathbf{Y}$ then $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) \geq \mathbb{P}(Y_0 = 1)\, \mathbf{H}(\mathbf{X})$. Let us show how ergodic theory enables us to prove a slightly weaker condition $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) > 0$.

Proposition 3.2.18. *Assume that $\mu, \nu \in \mathcal{M}^e(\{0,1\}^{\mathbb{Z}}, S)$ satisfy $\mathbf{H}(\nu) = 0$ with $\nu \neq \delta_{(\ldots 0,0,0\ldots)}$ and $\mathbf{H}(\mu) > 0$. Then $\mathbf{H}\left(\mu \overset{ind.}{*} \nu\right) > 0$.*

Proof. Consider $(\{0,1\}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}}, \mu \otimes \nu, S \times S)$ and denote by $\Pi(\mu) \subset \mathcal{B}$ the Pinsker σ-algebra of μ. Recall that for (X_i, μ_i, T_i), $i = 1,2$, we have the corresponding relation between the Pinsker σ-algebras: $\Pi(\mu_1 \otimes \mu_2, T_1 \times T_2) = \Pi(\mu_1, T_1) \otimes \Pi(\mu_2, T_2)$, see, e.g. [46]. It follows that

$$\Pi(\mu \otimes \nu) = \Pi(\mu) \otimes \mathcal{B}. \tag{3.2.20}$$

Let $C = \{x \in \{0,1\}^{\mathbb{Z}} \mid x_0 = 1\}$ and suppose that $\mathbf{H}\left(\mu \overset{ind.}{*} \nu\right) = 0$, i.e. $\Pi(\mu \overset{ind.}{*} \nu) = \mathcal{B}$. Therefore, additionally using (3.2.20), we obtain

$$M^{-1}(\mathcal{B}) = M^{-1}(\Pi(\mu \overset{ind.}{*} \nu)) \subset \Pi(\mu \otimes \nu) = \Pi(\mu) \otimes \mathcal{B}$$

and it follows that

$$C \times C = M^{-1}C \in \Pi(\mu) \otimes \mathcal{B}$$

(even though $C \times C = M^{-1}C$ is an equality between sets, we think of it up to sets of measure zero). Hence, for C on the first coordinate in $C \times C$, we have $C \in \Pi(\mu)$ (see (3.2.19)). Since $\{C, C^c\}$ is a generating partition, $\Pi(\mu) = \mathcal{B} \pmod{\mu}$ and it follows immediately that $\mathbf{H}(\mu) = 0$. ■

3.2.4 Topological pressure for "convolution systems"

Let $(\mathscr{X}, S)$, where $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$, be a subshift over a finite alphabet $\mathcal{X} = \{x_1, \ldots, x_k\}$ for some $k \in \mathbb{N}$ and $\varphi\colon \mathscr{X} \to \mathbb{R}$ be a continuous potential.

Remark 3.2.19. Without loss of generality we assume that every potential φ is defined **on the whole space** $\mathcal{X}^{\mathbb{Z}}$. Indeed, since every subshift $\mathscr{X}$ is (by definition) closed, the Tietze expansion theorem ensures that every **continuous** $\varphi\colon \mathscr{X} \to \mathbb{R}$ can be extended to the full shift (with supremum norm preserved).

Remark 3.2.20. Suppose that φ is local and depends on $[k,l]$ coordinates, where $k, l \in \mathbb{Z}$ and $k \leq l$. Then, slightly abusing our notation, we frequently write $\varphi(x_{[k,l]})$ for $\varphi(\mathbf{y})$, where $\mathbf{y}$ is such that $y_{[k,l]} = x_{[k,l]}$.

Recall (see Section 3.1.4) that the following **variational principle** holds:

$$\mathscr{P}_{\mathscr{X},\varphi} = \sup_{\mathbf{X} \in \mathcal{M}_{\mathscr{X}}} \big[\mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})\big]. \tag{3.2.21}$$

Motivated by (3.2.21), for any subset of invariant measures $\mathcal{N} \subset \mathcal{M}_{\mathscr{X}}$, we define

$$V_{\mathcal{N},\varphi} = \sup_{\mathbf{X} \in \mathcal{N}} \big[\mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})\big].$$

There is a special class of $\mathcal{N}$ we are interested in, namely, given stationary 0-1 process $\mathbf{Y}$ and finite **real** alphabet $\mathcal{X} \subset \mathbb{R}$, we define

$$\mathcal{N}_{\mathbf{Y}} = \left\{ \mathbf{X} \cdot \mathbf{Y} \mid (\mathbf{X}, \mathbf{Y}) \in \mathcal{M}_{\mathcal{X}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}}} \right\}.$$

Recall that such class of measures has been already considered in (3.2.4). Let us also add that the inspiration to study such families comes from $\mathscr{B}$-free systems (for the details see Section 4.2.3).

In order to warm up we consider firstly potentials depending only on one coordinate. In that case we have the following generalization of Corollary 3.2.9.

Theorem 3.2.21. *Fix a stationary process $\mathbf{Y}$ satisfying $\mathbf{H}(\mathbf{Y}) = 0$. Suppose that potential φ depends only on the first coordinate, that is $\varphi(\mathbf{x}) = \varphi(x_0)$. Then*

$$V_{\mathcal{N}_{\mathbf{Y}},\varphi} = \sup_{\mathbf{X}\cdot\mathbf{Y} \in \mathcal{N}_{\mathbf{Y}}} \left[\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) + \mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y}) \right] = (1-d)\varphi(0) + d \log_2 \left(\sum_{x\in\mathcal{X}} 2^{\varphi(x)} \right), \tag{3.2.22}$$

where $d = \mathbb{P}(Y_0 = 1)$. Furthermore, if $\mathbf{X}$ attains the above supremum then $\mathbf{X}\cdot\mathbf{Y} \sim \mathbf{G}\cdot\mathbf{Y}$ where G is an i.i.d. process (thus, $\mathbf{G} \amalg \mathbf{Y}$) such that $\mathbb{P}(G_i = x)$ is proportional to $2^{\varphi(x)}$.

Remark 3.2.22. The common distribution of G_i's is called a ***Gibbs measure*** (associated with φ). For the purpose of proof below recall that if we fix a finite alphabet $\mathcal{X}$ then the Gibbs measure $\mathbf{G}$ realizes the supremum

$$\sup_{X\in\mathcal{X}} \left[\mathbf{H}(X) + \mathbb{E}\varphi(X) \right]$$

and this supremum equals to $\log_2 \sum_{x\in\mathcal{X}} 2^{\varphi(x)}$ (see for example a calculation below equation (3) in [20]).

Proof. Firstly consider the case of $\mathbf{H}(\mathbf{X}) > 0$. Using Theorem 3.2.1, properties of entropy (conditioning decreases entropy) and Remark 3.2.22 we get

$$\begin{aligned} &\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) + \mathbb{E}\varphi(X_0Y_0) \overset{\text{Thm 3.2.1}}{=} d\mathbf{H}_{Y_0=1}\left(X_0 \mid X_{\{R_{(-\infty,-1]}}, \mathbf{Y} \right) + \mathbb{E}\varphi(X_0Y_0) \\ &\le (1-d)\varphi(0) + d\left[\mathbf{H}_{Y_0=1}(X_0) + \mathbb{E}_{Y_0=1}\varphi(X_0) \right] \overset{\text{Rema.3.2.22}}{\le} (1-d)\varphi(0) + d\log_2 \left(\sum_{x\in\mathcal{X}} 2^{\varphi(x)} \right). \end{aligned}$$

Furthermore, the first (second, resp.) inequality above becomes an equality iff conditionally on $Y_0 = 1$, X_0 is independent of $\mathbf{Y}$ and $X_{R_{(-\infty,-1]}}$ (we have $\mathbb{P}_{Y_0=1}(X_0 = x) = \frac{2^{\varphi(x)}}{\sum_{y\in\mathcal{X}} 2^{\varphi(y)}}$ for all $x \in \mathcal{X}$, resp.). Clearly, $\mathbf{G}$ satisfies both these conditions which immediately yields (3.2.22).

Now if $\mathbf{H}(\mathbf{X}) = 0$ then $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) = 0$ and

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) + \mathbb{E}\varphi(X_0Y_0) = \mathbb{E}\varphi(X_0Y_0) = (1-d)\varphi(0) + d\mathbb{E}_{Y_0=1}\varphi(X_0) \le (1-d)\varphi(0) + d\max_{\mathcal{X}} \varphi.$$

Clearly, $\max_{\mathcal{X}} \varphi \le \log_2 \left(\sum_{x\in\mathcal{X}} 2^{\varphi(x)} \right)$.

The proof of part of the uniqueness of the distribution of $\mathbf{X}\cdot\mathbf{Y}$ goes along the same lines as in Corollary 3.2.9 (just change letter $\mathbf{B}$ to $\mathbf{G}$ and use analogous arguments as in Corollary 3.2.9). ■

What happens if our potential φ depends on more than one coordinate? Can we get some concise formula like in (3.2.22)? It seems that there is no explicit expression for general local potentials. Take for example the case when $\mathbf{Y} = \mathbf{1}$ is the constant process. Then $\mathcal{N}_{\mathbf{Y}} = \mathcal{M}_{\mathcal{X}^{\mathbb{Z}}}$ and there are no evident closed-form expressions for the topological pressure. One can use the result by Walters which expresses the topological pressure in terms of greatest eigenvalue of some matrix (see Lemma 4.7 in [99]). Notice that finding explicit formulas from Walter's Lemma 4.7 is a tedious task which, for big n, if φ depends on n coordinates, is impossible to perform either by humans or computers. However, it turns out that there are some examples of $\mathbf{Y}$ and φ for which one can answer positively our questions. Roughly, one can give explicit formula for $V_{\mathcal{N}_{\mathbf{Y}},\varphi}$ if either $\mathbf{Y}$ is periodic and φ does not depend on "too many" coordinates (more precisely, if p is the period of $\mathbf{Y}$ then φ can depend on at most p coordinates; for the details see Theorem 3.2.32) or $\mathbf{Y}$ can be approximated in the weak topology by periodic processes (here φ can be arbitrary; see Theorem 3.2.34). In fact, both these cases arose during our studies of $\mathscr{B}$-free systems.

In order to deal with potentials depending on more than one coordinate, it will be convenient to introduce the following notion of **Y-upgrade of potential** φ. Recall that with every 0-1 process $\mathbf{Y}$ we can associate the return process $\mathbf{R}$ as in (2.2.4). Note that either of $\mathbf{R}$ and $\mathbf{Y}$ determines the other. Consider some $\mathbf{r} = (r_i)_{i\in\mathbb{Z}}$ (think about $\mathbf{r}$ as about a realization of $\mathbf{R}$) and put

$$\varphi_{\mathbf{r}}(\mathbf{z}) = \varphi\left(\dots, 0^{r_{-1}-r_{-2}-1}, \underbrace{z_{-1}}_{r_{-1}-coor.}, 0^{r_0-r_{-1}-1}, \underbrace{z_0}_{r_0-coor.}, 0^{r_1-r_0-1}, \underbrace{z_1}_{r_1-coor.}, 0^{r_2-r_1-1}, \dots\right).$$

Thus, $\mathbf{r}$ determines the slots in which sequence $\mathbf{z}$ is put. For example $\varphi_{\mathbf{1}} = \varphi$ and $\varphi_{\mathbf{0}} = \varphi(\mathbf{0})$. If φ is local and depends only on coordinates $[-m, m]$ for some $m \in \mathbb{N}$ then we will sometimes use an alternative definition corresponding to finite sequence $\mathbf{r}$. Namely, given a set $A = \{i_1, i_2, \dots, i_l\} \subset [-m, m]$, where $i_1 < \cdots i_{j-1} < 0 \le i_j < \cdots < i_k$, we put

$$\varphi_A\left(z_{[-m,m]}\right) = \varphi(\dots, \underbrace{z_{-1}}_{i_{j-1}-coor.}, 0^{i_j-i_{j-1}}, \underbrace{z_0}_{i_j-coor.}, 0^{i_{j+1}-i_j}, \underbrace{z_1}_{i_{j+1}-coor.}, \dots). \tag{3.2.23}$$

Thus, for example $\varphi_{\varnothing} = \varphi(0^{2m+1})$, $\varphi_{[-m,m]} = \varphi$, $\varphi_{\{0\}}(z_{[-m,m]}) = \varphi(0^m, z_0, 0^m)$. Note that we slightly abused notation by identifying $\varphi(\mathbf{x})$ and $\varphi(x_{[-m,m]})$ via $\varphi(x_{[-m,m]}) = \varphi(\mathbf{y})$ for any $\mathbf{y}$ such that $y_{[-m,m]} = x_{[-m,m]}$.

Now, we define ***Y-upgrade of potential*** φ, $\Phi = \Phi_{\mathbf{Y}}$, as

$$\Phi = \mathbb{E}\varphi_{\mathbf{R}}, \tag{3.2.24}$$

where $\mathbb{E}$ stands here for the Bochner's integral ($\mathbf{r} \to \varphi_{\mathbf{r}}$ acts on increasing bilateral integer-valued sequences and has image in the space of real continuous functions on $\mathcal{X}^{\mathbb{Z}}$, equipped with the supremum norm). For brevity's sake, we denote this procedure by

$$\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi. \tag{3.2.25}$$

Remark 3.2.23. Note that $\mathbf{r} \to \varphi_{\mathbf{r}}$ is measurable. Indeed, if φ is local, then it is easy to see that $\mathbf{r} \to \varphi_{\mathbf{r}}$ is a continuous function. In the general case, $\mathbf{r} \to \varphi_{\mathbf{r}}$ is measurable. Indeed, since the space of local functions is dense in the supremum norm, we can find a sequence of local functions $\varphi^{(n)}$ such that $\left\|\varphi - \varphi^{(n)}\right\|_\infty \le 1/n$ for $n \in \mathbb{N}$. By the definition of $\varphi_{\mathbf{r}}$, this implies $\sup_{\mathbf{r}} \left\|\varphi_{\mathbf{r}} - \varphi_{\mathbf{r}}^{(n)}\right\|_\infty \le 1/n$. Hence the mapping $\mathbf{r} \to \varphi_{\mathbf{r}}$ is a pointwise limit of continuous function, thus, it is measurable.

Remark 3.2.24. Clearly, $\|\varphi_{\mathbf{r}}\|_\infty \le \|\varphi\|_\infty$. Therefore, by the previous Remark 3.2.23, $\varphi_{\mathbf{R}}$ is Bochner-integrable, $\mathbb{E}\varphi_{\mathbf{R}}$ is well-defined and $\|\mathbb{E}\varphi_{\mathbf{R}}\|_\infty \le \|\varphi\|_\infty$.

Remark 3.2.25. Suppose that, $\mathbf{Y}^{(n)} \Rightarrow \mathbf{Y}$ and let $\mathbf{R}^{(n)}$ be are defined as in (2.2.4) with $\mathbf{Y}$ replaced by $\mathbf{Y}^{(n)}$. If $\varphi \overset{\mathbf{Y}^{(n)}}{\rightsquigarrow} \Phi^{(n)}$ and $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$ then

$$\Phi^{(n)} = \mathbb{E}\varphi_{\mathbf{R}^{(n)}} \xrightarrow{\|\cdot\|_\infty} \mathbb{E}\varphi_{\mathbf{R}} = \Phi.$$

Indeed, if φ is local then this is trivial. Otherwise, for each $\varepsilon > 0$, there is a local potential $\varphi^{(\varepsilon)}$ which satisfies $\left\|\varphi - \varphi^{(\varepsilon)}\right\|_\infty \le \varepsilon$ and then

$$\|\mathbb{E}\varphi_{\mathbf{R}^{(n)}} - \mathbb{E}\varphi_{\mathbf{R}}\|_\infty \le \left\|\mathbb{E}\varphi_{\mathbf{R}^{(n)}} - \mathbb{E}\varphi^{(\varepsilon)}_{\mathbf{R}^{(n)}}\right\|_\infty + \left\|\mathbb{E}\varphi_{\mathbf{R}} - \mathbb{E}\varphi^{(\varepsilon)}_{\mathbf{R}}\right\|_\infty + \left\|\mathbb{E}\varphi^{(\varepsilon)}_{\mathbf{R}^{(n)}} - \mathbb{E}\varphi^{(\varepsilon)}_{\mathbf{R}^{(n)}}\right\|_\infty. \tag{3.2.26}$$

It remains to notice that the first two terms are bounded by ε, whereas the third one goes to 0 when $n \to \infty$.

Example 3.2.26 (Local case)**.** Consider a **local** potential φ depending only on coordinates $[-m,m]$ for some $m \in \mathbb{N}$. How does Φ defined by (3.2.24) look like? It is not hard to see that

$$\Phi(\mathbf{z}) = \sum_{A \subset [-m,m]} \mathbb{P}\left(Y_A = 1, Y_{A^c} = 0\right) \varphi_A\left(z_{[-m,m]}\right), \tag{3.2.27}$$

where $A^c = [-m,m] \backslash A$. Since each φ_A is continuous, it follows from (3.2.27) that so must be Φ. Moreover, Φ is **local** and depends (at most) on coordinates $[-m,m]$.

Remark 3.2.27. Let us give a short glossary related to **binary words**. Let

$$w = (w_0, \ldots, w_{\ell-1}) \in \{0,1\}^\ell.$$

- Recall that $|w| = \ell$ stands for the ***length of*** w. Moreover, $\#_1(w)$ denotes the ***number of ones appearing in*** w, that is $\#_1(w) = \sum_{i=0}^{\ell-1} w_i$.
- We **upgrade** w **to sequence** $w^\infty \in \{0,1\}^{\mathbb{Z}}$ via $w_i^\infty = w_{i \bmod \ell}$.
- If w is non-zero then $\mathbf{r}^{(w)} = \left(r_i^{(w)}\right)_{i \in \mathbb{Z}} \in \mathbb{Z}^{\mathbb{Z}}$ denotes the ***sequence of positions of ones in*** w^∞. (If w is clear from the context we will omit the upper index and write $\mathbf{r}$ and r_i instead of $\mathbf{r}^{(w)}$ and $r_i^{(w)}$ respectively.) More precisely, we set $r_0 = \inf\{k \geq 0 \mid w_k = 1\}$, $r_i < r_{i+1}$ for $i \in \mathbb{Z}$ and for any $k \in \mathbb{Z}$, $w_k = 1$ iff $k = r_i$ for some $i \in \mathbb{Z}$. Note that this definition is consistent with the definition of the return time process (2.2.4) where $\mathbf{Y} = w^\infty$. Clearly, $\mathbf{r}$ is periodic, with period equal to $\#_1(w)$.
- We call w ***primitive*** if there exists no word u and $k \geq 2$ such that $w = u^k$.
- At last, for any $0 \leq i \leq \ell - 1$ we define the ***cyclic shift of word*** w by $w^{(i)} = w_{[i,\ell-1]} \cdot w_{[0,i-1]}$.

Remark 3.2.28. One of our main motivations behind considering such upgrade of potentials lies in the following easy observation

$$\mathbb{E}\varphi(\mathbf{X} \cdot \mathbf{Y}) = \mathbb{E}\mathbb{E}\left(\varphi(\mathbf{X} \cdot \mathbf{Y}) | \mathbf{Y}\right) = \mathbb{E}\mathbb{E}\left(\varphi_{\mathbf{R}}\left[(X_{R_i})_{i \in \mathbb{Z}}\right] | \mathbf{Y}\right),$$

which follows from the tower property of conditional expectation and the definition of $\varphi_{\mathbf{R}}$.

Example 3.2.29. Let $\mathbf{Y}$ be periodic with $\mathbf{Y} \sim \frac{1}{2}\left(\delta_{(01)^\infty} + \delta_{(10)^\infty}\right)$. Then with equal probabilities 1/2, $\mathbf{R}$ is either a sequence of odd or even integers. Thus, for any potential φ, if $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$, then

$$\Phi(\mathbf{z}) = \frac{1}{2}\left[\varphi(\ldots, z_{-1}, 0, \underbrace{z_0}_{0-coor.}, 0, \ldots) + \varphi(\ldots, 0, z_{-1}, \underbrace{0}_{0-coor.}, z_0, \ldots)\right]. \tag{3.2.28}$$

More generally, take $w = (w_0, \ldots, w_{\ell-1}) \in \{0,1\}^\ell$ with $w_0 = 1$. Let $\mathbf{Y}$ be the corresponding w-periodic stationary process (so $\mathbf{Y}$ arises as the start randomization of deterministic process w^∞). In that case, if $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$, then the expansion of the integral in (3.2.24) gives

$$\Phi = \frac{1}{\ell} \sum_{i=1}^{\ell} S^i \varphi_{\mathbf{r}^{(w)}}, \tag{3.2.29}$$

where the $\mathbf{r}^{(w)}$ are the positions of ones in w^∞.

There is another reason why we care for upgrades of potentials and it is contained in the following lemma.

Lemma 3.2.30. *Let w be a binary word of length $|w| = \ell$. Suppose that $(\mathbf{X}, \mathbf{Y}) \in (\mathbb{R} \times \mathbb{R})^{\mathbb{Z}}$ is a stationary finitely-valued process such that $\mathbf{Y}$ is w-periodic. Then*

$$\mathbb{E}\varphi(\mathbf{X} \cdot \mathbf{Y}) = \mathbb{E}_{Y_{[0,\ell-1]}=w} \Phi\left((X_{r_i})_{i \in \mathbb{Z}}\right), \tag{3.2.30}$$

where $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$ and $\mathbf{r} = (r_i)_{i \in \mathbb{Z}}$ are indices of ones in w^∞.

Proof. Let $w = (w_0, \dots, w_{\ell-1}) \in \mathcal{X}^\ell$ and recall that $w^{(i)} = w_{[i,\ell-1-i]} \cdot w_{[0,i-1]}$ stands for the i'th cyclic shift of w. Using the fact that $(\mathbf{X}, \mathbf{Y})$ is stationary we get

$$\mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y}) = \frac{1}{\ell}\sum_{i=1}^{\ell} \mathbb{E}_{Y_{[-i,-i+\ell-1]}=w}\varphi\left(\mathbf{X}\cdot\left(w^{(i)}\right)^{\infty}\right) = \frac{1}{\ell}\sum_{i=1}^{\ell} \mathbb{E}_{Y_{[0,\ell-1]}=w}\varphi\left(S^i\mathbf{X}\cdot\left(w^{(i)}\right)^{\infty}\right)$$

$$= \frac{1}{\ell}\sum_{i=1}^{\ell} \mathbb{E}_{Y_{[0,\ell-1]}=w}\varphi\left(S^i\left(\mathbf{X}\cdot w^{\infty}\right)\right) \overset{(3.2.29)}{=} \mathbb{E}_{Y_{[0,\ell-1]}=w}\Phi\left((X_{r_i})_{i\in\mathbb{Z}}\right).$$

■

In order to present our main result concerning the periodic case, we need one more easy corollary of Theorem 3.2.1.

Corollary 3.2.31. *Let w be a binary word of length $|w| = \ell$, with $\#_1(w) = m$. Suppose that $(\mathbf{X}, \mathbf{Y})$ is a stationary finitely-valued process, such that $\mathbf{H}(\mathbf{X}) > 0$ and $\mathbf{Y}$ is w-periodic. Then*

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) = \frac{1}{\ell}\mathbf{H}_{Y_{[0,\ell-1]}=w}\left(X_{\{r_{m-1},r_{m-2},\dots,r_0\}} \mid X_{\{r_{-1},r_{-2},\dots\}}\right), \tag{3.2.31}$$

where $\mathbf{r} = (r_i)_{i\in\mathbb{Z}}$ are indices of ones in w^∞.

Proof. Without loss of generality we can assume that $w_0 = 1$. For brevity's sake for any set $Z \subset \mathbb{Z}$ and $k \in \mathbb{Z}$ denote $Z \pm k = \{z \pm k \mid z \in Z\}$. Using the properties of conditional entropy, the fact that $\mathbf{Y}$ is periodic and then the stationarity of $(\mathbf{X}, \mathbf{Y})$ we get

$$\mathbf{H}_{Y_0=1}\left(X_0 \mid X_{R_{(-\infty,-1]}}, \mathbf{Y}\right) = \sum_{i=0}^{m-1} \mathbb{P}_{Y_0=1}\left(Y_{[0,\ell-1]} = w^{(r_i)}\right)\mathbf{H}_{Y_{[0,\ell-1]}=w^{(r_i)}}\left(X_0 \mid X_{r_{(-\infty,i-1]}-r_i}\right)$$

$$= \frac{1}{m}\sum_{i=0}^{m-1} \mathbf{H}_{Y_{[-r_i,-r_i+\ell-1]}=w}\left(X_0 \mid X_{r_{(-\infty,i-1]}-r_i}\right) = \frac{1}{m}\sum_{i=0}^{m-1} \mathbf{H}_{Y_{[0,\ell-1]}=w}\left(X_{r_i} \mid X_{r_{(-\infty,i-1]}}\right),$$

which after the application of the chain rule gives the desired result. ■

Lemma 3.2.30 and Corollary 3.2.31 are the main ingredients of the proof of the following theorem.

Theorem 3.2.32. *Let w be a primitive binary word of length $|w| = \ell$, with $\#_1(w) = m$. Suppose that $(\mathbf{X}, \mathbf{Y}) \in (\mathbb{R}\times\mathbb{R})^{\mathbb{Z}}$ is a stationary finitely-valued process, such that $\mathbf{H}(\mathbf{X}) > 0$ and $\mathbf{Y}$ is w-periodic. If $\varphi\colon (\mathcal{X}\cup\{0\})^{\mathbb{Z}} \to \mathbb{R}$ is local and depends only on coordinates $[0, \ell-1]$ then*

$$V_{\mathcal{N}_{\mathbf{Y}},\varphi} = \frac{1}{\ell}\log_2\left[\sum_{z_{[0,m-1]}\in\mathcal{X}^m} 2^{\ell\Phi(z_{[0,m-1]})}\right]. \tag{3.2.32}$$

Moreover, the supremum defining $V_{\mathcal{N}_{\mathbf{Y}},\varphi}$ is attained by any pair $(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X}$ conditionally on $Y_{[0,\ell-1]} = w$ is such that $\left(X_{[i\ell,(i+1)\ell-1]}\right)$ is an i.i.d. process satisfying

$$\mathbb{P}\left(X_{r_{[0,m-1]}} = w_{[0,m-1]}\right) = \frac{2^{\ell\Phi(w_{[0,m-1]})}}{\sum_{z_{[0,m-1]}} 2^{\ell\Phi(z_{[0,m-1]})}}. \tag{3.2.33}$$

Proof. Let $\mathbf{r}$ be the sequence of positions of ones in w^∞. Firstly we consider the case of $\mathbf{H}(\mathbf{X}) > 0$. Then, by Lemma 3.2.30 and Corollary 3.2.31,

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) = \frac{1}{\ell}\mathbf{H}_{Y_{[0,\ell-1]}=w}\left(X_{r_{[0,m-1]}} \mid X_{r_{(-\infty,-1]}}\right), \quad \mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y}) = \mathbb{E}_{Y_{[0,\ell-1]}=w}\Phi\left(X_{r_{[0,m-1]}}\right).$$

Thus, we must now deal with the following problem

$$\sup_{\mathbf{X}\cdot\mathbf{Y}\in\mathcal{N}_{\mathbf{Y}}} \frac{1}{\ell}\mathbf{H}_{Y_{[0,\ell-1]}=w}\left(X_{r_{[0,m-1]}} \mid X_{r_{(-\infty,-1]}}\right) + \mathbb{E}_{Y_{[0,\ell-1]}=w}\Phi\left(X_{r_{[0,m-1]}}\right).$$

Note that we can absorb the constant $1/\ell$ by considering $\frac{1}{\ell}\varphi$ instead of φ. Hence, substituting $Z_i = X_{r_i}$ reduces our problem to showing that

$$\sup_{\mathbf{Z}} \mathbf{H}\left(Z_{[0,m-1]} \mid Z_{(-\infty,-1]}\right) + \mathbb{E}\Phi\left(Z_{[0,m-1]}\right) = \log_2\left[\sum_{z_{[0,m-1]}} 2^{\Phi(z_{[0,m-1]})}\right],$$

where the supremum is taken over processes $\mathbf{Z}$ that are stationary under S^m (here we use the assumption of primitivity of w!). Thus, equivalently, after another substitution, namely $U_i = Z_{[im,(i+1)m-1]}$, we must deal with

$$\sup_{\mathbf{U}-\text{stationary}} \mathbf{H}\left(U_0 \mid U_{(-\infty,-1]}\right) + \mathbb{E}\Phi\left(U_0\right) = \log_2\left[\sum_u 2^{\Phi(u)}\right]. \tag{3.2.34}$$

But this is a standard problem:

$$\mathbf{H}\left(U_0 \mid U_{(-\infty,-1]}\right) + \mathbb{E}\Phi\left(U_0\right) \leq \mathbf{H}\left(U_0\right) + \mathbb{E}\Phi\left(U_0\right) \leq \log_2\left[\sum_u 2^{\Phi(u)}\right], \tag{3.2.35}$$

where the last inequality becomes equality iff (recall Remark 3.2.22)

$$\mathbb{P}\left(U_0 = u_0\right) = \frac{2^{\Phi(u_0)}}{\sum_{\mathbf{u}} 2^{\Phi(u)}}. \tag{3.2.36}$$

At last, one can obtain equalities in (3.2.35) iff $\mathbf{U}$ is an i.i.d. process satisfying (3.2.36).

It is remains to notice that if $\mathbf{H}(\mathbf{X}) = 0$ then $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) = 0$ and by Lemma 3.2.30 $\mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y}) \leq \max_{x\in\mathcal{X}^m}\Phi(x)$ which is smaller than the right-hand side expression of (3.2.31). ∎

Remark 3.2.33 (Continuity (in potential) of variational problem in supremum norm)**.** Recall that

$$\mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y}) = \mathbb{E}\mathbb{E}\left(\varphi_{\mathbf{R}}\left[(X_{R_i})_{i\in\mathbb{Z}}\right]|\mathbf{Y}\right).$$

Thus,

$$\left|V_{\mathcal{N}_{\mathbf{Y}},\varphi} - V_{\mathcal{N}_{\mathbf{Y}},\varphi}\right| \leq \left|\mathbb{E}\mathbb{E}\left(\left(\varphi_{\mathbf{R}} - \psi_{\mathbf{R}}\right)\left[(X_{R_i})_{i\in\mathbb{Z}}\right]|\mathbf{Y}\right)\right| \leq \sup_{\mathbf{r}} \left\|\varphi_{\mathbf{r}} - \psi_{\mathbf{r}}\right\|_\infty \leq \left\|\varphi - \psi\right\|_\infty.$$

Now, we will show what happens if $\mathbf{Y}$ can be approximated by w_n-periodic processes $\mathbf{Y}^{(n)}$ with w_n primitive and such that $|w_n| \to \infty$.

Theorem 3.2.34. *Assume that the sequence of w_n-periodic processes $\mathbf{Y}^{(n)}$, where w_n are finite primitive 0-1 words satisfying $|\#_1(w_n)| \to \infty$, converges weakly to $\mathbf{Y}$. Then for any continuous potential $\varphi\colon (\mathcal{X}\cup\{0\})^{\mathbb{Z}} \to \mathbb{R}$,*

$$V_{\mathcal{N}_{\mathbf{Y}^{(n)}},\varphi} \to \mathbb{P}\left(Y_0 = 1\right)\log|\mathcal{X}| + \sup_{\mathbf{z}\in\mathcal{X}^{\mathbb{Z}}} \Phi(\mathbf{z}), \tag{3.2.37}$$

where $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$.

Proof. Since for any $\mathbf{Z}$, $\varphi \to \Phi$ and $\varphi \to V_{\mathcal{N}_{\mathbf{Z}},\varphi}$ are continuous in supremum norm (recall Remark 3.2.24 and Remark 3.2.33 resp.), without loss of generality, we can assume that φ is local and depends on $[0, N]$ coordinates. Let $\varphi \overset{\mathbf{Y}^{(n)}}{\rightsquigarrow} \Phi^{(n)}$, $\ell_n = |w_n|$ and $m_n = \#_1(w_n)$. Then by Theorem 3.2.32 we get that

$$V_{\mathcal{N}_{\mathbf{Y}^{(n)}},\varphi} = \frac{1}{\ell_n}\log_2\left[\sum_{z_{[0,m_n-1]}} 2^{\ell_n\Phi^{(n)}(z_{[0,m_n-1]})}\right]$$

Since Φ_n depends only on $[0, N-1]$ coordinates and $m_n \to \infty$, for sufficiently big n, we have

$$V_{\mathcal{N}_{\mathbf{Y}^{(n)}},\varphi} = \frac{1}{\ell_n}\log_2\left[\sum_{z_{[0,m_n-1]}} 2^{\ell_n\Phi^{(n)}(z_{[0,N-1]})}\right] = \frac{m_n - N}{\ell_n}\log_2|\mathcal{X}| + \frac{1}{\ell_n}\log_2\left[\sum_{z_{[0,N-1]}} 2^{\ell_n\Phi^{(n)}(z_{[0,N-1]})}\right]$$

Therefore, using $\frac{m_n}{\ell_n} = \mathbb{P}\left(Y_0^{(n)} = 1\right) \to \mathbb{P}\left(Y_0 = 1\right)$, Remark 3.2.25 and the standard fact that ℓ_p norms converge to the ℓ_∞ norm (as $p\to\infty$), we get our result. ∎

Chapter 4

$\mathscr{B}$-free systems

Sets of multiples of a given set $\mathscr{B} \subset \mathbb{N}$ and their complements ($\mathscr{B}$-free sets) were studied already in the 30's by numerous mathematicians from the number-theoretic viewpoint (see, e.g. [28, 22, 14, 29, 29, 35]). The most prominent example here is the set of ***square-free integers***, i.e. the set of integers not divisible by the square of any prime. The dynamical approach was initiated by Sarnak in his seminal lectures [92]. He proposed to study the dynamical system given by the orbit closures of the Möbius function $\boldsymbol{\mu}$ and its square $\boldsymbol{\mu}^2$ under the left shift S in $\{-1,0,1\}^{\mathbb{Z}}$ (note that $\boldsymbol{\mu}^2$ is nothing but the characteristic function of the set of non-negative square-free integers). These ideas were later extended to general sets of $\mathscr{B}$-free numbers, resulting in a class of dynamical systems called $\mathscr{B}$-free systems. See, e.g. [3], where the basic dynamical tools were developed or [34]. Last, but not least, let us mention that Sarnak's dynamical approach was motivated by the random-like behaviour of the Möbius function. He formulated a conjecture on the orthogonality of $\boldsymbol{\mu}$ to all deterministic sequences (arising from topological dynamical systems of zero topological entropy) [92]. This conjecture is weaker than the celebrated conjecture of Chowla on the absence of autocorrelations of $\boldsymbol{\mu}$ (for an ergodic-theoretic proof suggested already in [92], see [2], cf. also [86, 96]). Thus, we deal with a very active area of study, lying at the verge of ergodic theory and number theory (one of the break-throughs made on the number-theoretic side was made by Matomäki, Radziwiłł and Tao [72]). For examples and more background, we refer the reader, e.g., to [38] and [60].

Since the topological entropy of the square-free system is positive, a natural question arose whether this system is intrinsically ergodic. It was answered by Peckner in [85]. Later, this result was extended to general $\mathscr{B}$-free systems [61, 34]. Peckner also showed that the measure of maximal entropy **fails** to have Gibbs property. However, his proof relies on non-trivial number-theoretic facts (more precisely, on the explicit formula for the Mirsky measure of blocks and some classical estimates concerning the squares of primes). Thus, he asked if his result extends to the general case of $\mathscr{B}$ such that elements of $\mathscr{B}$ are coprime and $\sum_{b\in\mathscr{B}} 1/b < \infty$. We recall that in such case the corresponding $\mathscr{B}$-free subshift is hereditary. Our main result gives the positive answer to this problem. Furthermore, it is proven using different kinds of arguments than those from [85]. In fact, we formulate a more general criterion based on notions of topological entropy and (topological) density of ones which ensures the absence of Gibbs property (see Theorem 4.2.3).

4.1 Background

4.1.1 $\mathscr{B}$-free subshift

For a subset of positive integers $\mathscr{B} \subset \mathbb{N}\setminus\{1\}$, consider respectively, the ***set of multiples*** and the ***set of $\mathscr{B}$-free numbers***

$$\mathcal{M}_{\mathscr{B}} = \bigcup_{b\in\mathscr{B}} b\mathbb{Z}, \qquad \mathcal{F}_{\mathscr{B}} = \mathbb{Z}\setminus\mathcal{M}_{\mathscr{B}}. \tag{4.1.1}$$

Let $\boldsymbol{\eta} = \boldsymbol{\eta}(\mathscr{B}) = \mathbb{1}_{\mathcal{F}_{\mathscr{B}}}$ stand for the ***characteristic function of*** $\mathcal{F}_{\mathscr{B}}$ (thus, $\boldsymbol{\eta}$ is just a binary bilateral sequence) and define the ***$\mathscr{B}$-free subshift*** by setting

$$\mathscr{X}_{\boldsymbol{\eta}} = \overline{\{S^k\boldsymbol{\eta} \mid k\in\mathbb{Z}\}} \subset \{0,1\}^{\mathbb{Z}}. \tag{4.1.2}$$

Remark 4.1.1. We tacitly assume that $\mathscr{B}$ is ***primitive*** in the sense that if k and l are distinct members of $\mathscr{B}$ then $k \not| l$ and $l \not| k$. Note that if $\mathscr{B}$ is not primitive then we can throw away some elements of $\mathscr{B}$ (namely, those which are multiples of the others) obtaining a primitive set $\mathscr{C}$ in such a way that the set of multiples does not change, that is $\mathcal{M}_{\mathscr{B}} = \mathcal{M}_{\mathscr{C}}$.

4.1.2 Hereditary subshifts

A subshift $(\mathscr{X}, S)$ with language $\mathcal{L}$, where $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$, is called ***hereditary*** if

$$w \in \mathcal{L},\ w' \leq w \quad \Rightarrow \quad w' \in \mathcal{L}.$$

Moreover, given a subshift $\mathscr{X}$, the ***hereditary closure*** of $\mathscr{X}$ is defined by

$$\widetilde{\mathscr{X}} = \{\mathbf{z} \in \{0,1\}^{\mathbb{Z}} \mid \mathbf{z} \leq \mathbf{x} \text{ for some } \mathbf{x} \in \mathscr{X}\}.$$

It follows immediately that $\mathscr{X}$ is hereditary iff $\widetilde{\mathscr{X}} = \mathscr{X}$. Examples of hereditary systems include many $\mathscr{B}$-free systems, spacing shifts [65], beta shifts ([87], for the proof of heredity, see [62]), bounded density shifts [95] or some shifts of finite type. Most of them are intrinsically ergodic (i.e. they have a unique measure of maximal entropy), see [25] for beta shifts and [84] for a subclass of bounded density shifts (for other listed examples, to our best knowledge, intrinsic ergodicity remains open). See also [58, 62].

There is another subshift of $\{0,1\}^{\mathbb{Z}}$ closely related to $\mathscr{X}_{\boldsymbol{\eta}}$, namely $\mathscr{X}_{\mathscr{B}}$, known as ***$\mathscr{B}$-admissible subshift*** and defined by

$$\mathbf{x} \in \mathscr{X}_{\mathscr{B}} \quad \Leftrightarrow \quad |\text{supp } \mathbf{x} \bmod b| < b \quad \forall_{b \in \mathscr{B}}. \tag{4.1.3}$$

By the very definition, $\mathscr{X}_{\mathscr{B}}$ is hereditary. Since for an arbitrary $\mathscr{B}$, $0 \notin \text{supp } \boldsymbol{\eta}$, we immediately get $\boldsymbol{\eta} \in \mathscr{X}_{\mathscr{B}}$. Thus, $\mathscr{X}_{\boldsymbol{\eta}} \subset \mathscr{X}_{\mathscr{B}}$ and it follows that $\mathscr{X}_{\boldsymbol{\eta}} \subset \widetilde{\mathscr{X}_{\boldsymbol{\eta}}} \subset \mathscr{X}_{\mathscr{B}}$. It is not always true that $\mathscr{X}_{\boldsymbol{\eta}}$ is hereditary, but if $\mathscr{B}$ is a co-prime set such that $\sum_{b \in \mathscr{B}} 1/b < \infty$ (if both these conditions are satisfied, we say that $\mathscr{B}$ is ***Erdös***) then this is the case. In fact, we then even have $\mathscr{X}_{\boldsymbol{\eta}} = \widetilde{\mathscr{X}_{\boldsymbol{\eta}}} = \mathscr{X}_{\mathscr{B}}$ (see Remark 3.11 in [34]).

Example 4.1.2 (Square-free system)**.** If we put

$$\mathscr{B} = \mathscr{P}^2 = \{p^2 : p \in \mathscr{P}\}, \qquad \mathscr{P} = \{\text{prime numbers}\} \tag{4.1.4}$$

then $\mathcal{F}_{\mathscr{B}}$ is the ***set of square-free integers***. The characteristic function of $\mathcal{F}_{\mathscr{B}}$ is the square $\boldsymbol{\mu}^2$ of the Möbius function $\boldsymbol{\mu}$ extended to $\mathbb{Z}$ in the natural symmetric way, $\boldsymbol{\mu}(-n) = \boldsymbol{\mu}(n)$. Recall that $\boldsymbol{\mu}(n) = (-1)^k$ if n is a product of $k \geq 1$ distinct primes, $\boldsymbol{\mu}(1) = 1$ and $\boldsymbol{\mu}(n) = 0$ if $n \in \mathbb{N}$ is not square-free. Since $\mathscr{P}^2$ is Erdös, $\mathscr{X}_{\boldsymbol{\eta}} = \widetilde{\mathscr{X}_{\boldsymbol{\eta}}} = \mathscr{X}_{\mathscr{P}^2}$.

4.1.3 Taut and Behrend sets

Recall that given a subset of integers $N \subset \mathbb{Z}$, the ***upper and lower logarithmic density of*** N are defined as

$$\overline{\boldsymbol{\delta}}(N) = \limsup_{n \to \infty} \frac{1}{\log n} \sum_{i=1}^{n} \frac{1}{i} \mathbb{1}_{i \in N}, \qquad \underline{\boldsymbol{\delta}}(N) = \liminf_{n \to \infty} \frac{1}{\log n} \sum_{i=1}^{n} \frac{1}{i} \mathbb{1}_{i \in N},$$

respectively. If $\overline{\boldsymbol{\delta}}(N) = \underline{\boldsymbol{\delta}}(N)$ then we say that N ***has logarithmic density*** and denote this quantity by $\boldsymbol{\delta}(N)$. Similarly, the classical (***upper/lower***) ***density of*** N are given by

$$\overline{\mathbf{d}}(N) = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{i \in N}, \qquad \underline{\mathbf{d}}(N) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{i \in N}, \qquad \mathbf{d}(N) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{i \in N},$$

respectively.

Remark 4.1.3. It turns out that the classical density of $\mathcal{M}_{\mathscr{B}}$ does not always exist (see [13]) unlike its logarithmic counterpart (see [29]). More precisely, it was shown in [29] that for any $\mathscr{B}$,

$$\boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}}) = \underline{\mathbf{d}}(\mathcal{M}_{\mathscr{B}}) = \lim_{K \to \infty} \mathbf{d}(\mathcal{M}_{\{b \in \mathscr{B} : b \leq K\}}). \tag{4.1.5}$$

Remark 4.1.4 (Relations between $\boldsymbol{\delta}$ and $\mathbf{d}$). Let us fix some $N \subset \mathbb{N}$ and let $s_k = \sum_{i=1}^{k} \mathbb{1}_{i \in N}$. Then, using the summation by parts we get

$$\overline{\boldsymbol{\delta}}_n := \frac{1}{\log n} \sum_{i=1}^{n} \frac{1}{i} \mathbb{1}_{i \in \mathbb{N}} = \frac{1}{\log n} \sum_{i=1}^{n} \frac{1}{i}(s_{i+1} - s_i) = \frac{1}{\log n} \sum_{i=1}^{n-1} \frac{1}{i} \frac{s_{i+1}}{i+1} + \frac{s_{n+1}}{(n+1)\log n} - \frac{s_1}{\log n}.$$

Thus, roughly, the partial sum of logarithmic type $\overline{\boldsymbol{\delta}}_n$ can be treated as convex combinations of $s_1/1, s_2/2, \ldots, s_n/n$ .This immediately leads to the following conclusion:

$$\underline{\mathbf{d}}(N) \leq \underline{\boldsymbol{\delta}}(N) \leq \overline{\boldsymbol{\delta}}(N) \leq \overline{\mathbf{d}}(N).$$

A set $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$ is ***taut*** (see [49]) if for every $b \in \mathscr{B}$,

$$\boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}}) > \boldsymbol{\delta}(\mathcal{M}_{\mathscr{B} \setminus \{b\}}). \tag{4.1.6}$$

It was shown in [34] (see Corollary 2.31 therein) that the tautness of $\mathscr{B}$ implies the following property

$$\boldsymbol{\delta}(\mathcal{M}_{\mathscr{B} \cup \{a\}}) = \boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}}) \qquad \Rightarrow \qquad a \in \mathcal{M}_{\mathscr{B}}. \tag{4.1.7}$$

A set $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$ is said to be ***Behrend*** (see [49]) if $\boldsymbol{\delta}(\mathcal{F}_{\mathscr{B}}) = 0$. Each infinite subset of primes whose sums of reciprocals is infinite is Behrend (see (0.69) in [49]). Take $a, r \in \mathbb{N}$ with $\gcd(a, r) = 1$. Dirichlet proved that $a\mathbb{Z} + r$ contains infinitely many primes and $\sum_{p \in (a\mathbb{Z}+r) \cap \mathscr{P}} 1/p = \infty$. Thus,

$$\gcd(a, r) = 1 \quad \Rightarrow \quad a\mathbb{Z} + r \text{ is Behrend.} \tag{4.1.8}$$

4.1.4 Entropy

Firstly, for convenience of the reader (we want this chapter to be self-contained), we recall (and rewrite in ergodic manner) some basic facts and definitions concerning entropy. Given a subshift $(\mathscr{X}, S)$, its ***topological entropy*** $\mathbf{H} = \mathbf{H}_{\mathscr{X}}$ (see Section 3.1.3) can be computed as

$$\mathbf{H} = \lim_{n \to \infty} \frac{1}{n} \log \left| \mathcal{L}^{(n)} \right| = \inf_{n \in \mathbb{N}} \frac{1}{n} \log \left| \mathcal{L}^{(n)} \right|. \tag{4.1.9}$$

Furthermore, for any $\nu \in \mathcal{M}_{\mathscr{X}}$ the ***measure entropy*** (cf. Section 3.1.2) $\mathbf{H}(\nu)$ is given by

$$\mathbf{H}(\nu) = \lim_{n \to \infty} \frac{1}{n} \mathbf{H}_{\nu}\left(\mathcal{L}^{(n)}\right) = \inf_{n \in \mathbb{N}} \frac{1}{n} \mathbf{H}_{\nu}\left(\mathcal{L}^{(n)}\right), \tag{4.1.10}$$

where $\mathbf{H}_{\nu}\left(\mathcal{L}^{(n)}\right) = -\sum_{w \in \mathcal{L}^{(n)}} \nu(w) \log(\nu(w))$ denotes the Shannon entropy with respect to the partition of $\mathscr{X}$ given by $\mathcal{L}^{(n)}$ (cf. Section 3.1.1). It is well-known that $\mathbf{H}(\nu)$ and $\mathbf{H}$ are related via ***variational principle***

$$\mathbf{H} = \sup_{\nu \in \mathcal{M}} \mathbf{H}(\nu). \tag{4.1.11}$$

Every measure ν such that $\mathbf{H}(\nu) = \mathbf{H}$ is called ***measure of maximal entropy***. If there is only one measure of maximal entropy then we say that $\mathscr{X}$ is ***intrinsically ergodic***.

Remark 4.1.5. Since every probabilistic measure on a finite set can be interpreted as a probabilistic vector $\boldsymbol{p} = (p_1, \ldots, p_n)$ it is natural to extend the definition of Shannon's entropy to such vectors via

$$\mathbf{H}(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log_2 p_i. \tag{4.1.12}$$

In particular, if $p \in [0, 1]$ and $\boldsymbol{p} = (p, 1-p)$ then we write

$$\mathbf{H}(p) = \mathbf{H}(\boldsymbol{p}) = -p \log(p) - (1-p)\log(1-p) \tag{4.1.13}$$

for ***binary entropy function***. Notice that $\mathbf{H}(p)$ is symmetric with respect to $1/2$, strictly increasing (decreasing resp.) on $[0, 1/2]$ ($[1/2, 1]$ resp.).

4.1.5 Densities of ones

In this part, given a subshift $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$, we discuss how to measure the density of occurrences of ones in $\mathscr{X}$. On the one hand we do it in light of the topology of $\mathscr{X}$ ($\mathbf{d}_{\mathscr{X}}$ and $\mathbf{D}_{\mathscr{X}}$), on the other hand, from the point of view of a fixed measure $\nu \in \mathcal{M}_{\mathscr{X}}$ ($\mathbf{d}_\nu$ and $\mathbf{D}_\nu$). More precisely, for $\nu \in \mathcal{M} = \mathcal{M}_{\mathscr{X}}$, we define

$$\mathbf{d}_{\mathscr{X}} = \sup_{\nu\in\mathcal{M}} \mathbf{d}_\nu, \quad \mathbf{d}_\nu = \nu([1]), \quad \mathbf{D}_{\mathscr{X}} = \lim_{n\to\infty} \frac{\max_{W\in\mathcal{L}_n} \#_1 W}{n}, \quad \mathbf{D}_\nu = \lim_{n\to\infty} \frac{\max_{W\in\mathcal{L}_n,\, \nu(W)>0} \#_1 W}{n}.$$

When the underlying space $\mathscr{X}$ is clear from the context we skip the index $\mathscr{X}$ in $\mathbf{D}_{\mathscr{X}}$ and $\mathbf{d}_{\mathscr{X}}$. Notice that both $\mathbf{D}$ and $\mathbf{D}_\nu$ are well defined since the sequences $\left(\max_{w\in\mathcal{L}_n,\, \nu(W)>0} \#_1(w)\right)_n$ and $(\max_{w\in\mathcal{L}_n} \#_1 w)_n$ are sub-additive. In particular, we can replace lim's by inf's.

We call a measure $\nu \in \mathcal{M}_{\mathscr{X}}$ a ***maximal density measure*** if $\mathbf{d}_\nu = \mathbf{d}$ and ***ones-saturated*** if $\mathbf{D}_\nu = \mathbf{D}$.

Remark 4.1.6. Notice that a measure of maximal density always exists. Indeed, $f = \mathbb{1}_1$ is continuous and thus so is $\nu \mapsto \nu(1) = \nu(f)$.

At last recall that for any $N \subset \mathbb{N}$ its ***upper Banach density*** is given by

$$\overline{\mathbf{d}}_B(N) = \limsup_{n-m\to\infty} \frac{1}{n-m+1} \sum_{i=m}^{n} \mathbb{1}_{i\in N}.$$

In order to present our main theorem concerning relations between different notions of density of ones, we need to recall a folklore result. For reader's convenience we provide its proof.

Proposition 4.1.7. *Let $(\mathscr{X}, S)$ be a subshift. Let $\mathbf{x} \in \mathscr{X}$ and let $\mathbf{A} \subset \mathscr{X}$ be a clopen set. Then*

$$\overline{\mathbf{d}}_B(\{n \in \mathbb{N} \mid S^n\mathbf{x} \in \mathbf{A}\}) = \sup\{\nu(\mathbf{A}) \mid \nu \in \mathcal{M} \text{ such that } \nu(\overline{\{S^n\mathbf{x} \mid n \in \mathbb{Z}\}}) = 1\}.$$

Proof. Firstly, we show that there exists ν such that

$$\nu(\overline{\{S^n\mathbf{x} \mid n\in\mathbb{Z}\}}) = 1, \qquad \overline{\mathbf{d}}_B(\{n \in \mathbb{N} \mid S^n\mathbf{x} \in \mathbf{A}\}) = \nu(\mathbf{A}).$$

Fix $\mathbf{x} \in \mathscr{X}$ and let $(m_k) \subset \mathbb{N}$ be a sequence such that

$$\overline{\mathbf{d}}_B(\{n \in \mathbb{N} \mid S^n\mathbf{x} \in \mathbf{A}\}) = \lim_{k\to\infty} \frac{1}{k} \left|\{m_k \le n \le m_k + k - 1 \mid S^n\mathbf{x} \in \mathbf{A}\}\right|. \tag{4.1.14}$$

Let $\mathbf{x}^{(k)} = S^{m_k}\mathbf{x}$ and

$$\nu_k = \frac{1}{k} \sum_{i=0}^{k-1} \delta_{S^i\mathbf{x}^{(k)}}.$$

Without loss of generality we may assume that $\nu_k \Rightarrow \nu$. Notice that ν is S-invariant and concentrated on the orbit closure of $\mathbf{x}$ under S. Rewriting (4.1.14) and using the fact that $\mathbb{1}_{\mathbf{A}}$ is continuous, we obtain

$$\overline{\mathbf{d}}_B(\{n \in \mathbb{N} \mid S^n\mathbf{x} \in \mathbf{A}\}) = \lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k} \mathbb{1}_{\mathbf{A}}(S^i\mathbf{x}^{(k)}) = \lim_{k\to\infty} \int \mathbb{1}_{\mathbf{A}}\, d\nu_k = \nu(\mathbf{A}).$$

Now we show that if $\nu \in \mathcal{M}$ is such that $\nu(\overline{\{S^n\mathbf{x} \mid n \in \mathbb{Z}\}}) = 1\}$ then $\nu(\mathbf{A})$ cannot exceed $\overline{\mathbf{d}}_B(\{n \in \mathbb{N} \mid S^n\mathbf{x} \in \mathbf{A}\})$. Using the ergodic decomposition, it is clear that it suffices to prove it for ν ergodic. For any ergodic ν that is concentrated on the orbit closure of $\mathbf{x}$ under S, we can find a generic point $\mathbf{y}$ in the orbit closure of $\mathbf{x}$. In particular, one can find m_k such that

$$\{0 \le i \le k-1 \mid S^{m_k+i}\mathbf{x} \in \mathbf{A}\} = \{0 \le i \le k-1 \mid S^i\mathbf{y} \in \mathbf{A}\}$$

(recall that $\mathbf{A}$ is clopen). It follows that

$$\begin{aligned}\overline{\mathbf{d}}_B(\{n \in \mathbb{N} \mid S^n\mathbf{x} \in \mathbf{A}\}) &\ge \lim_{k\to\infty} \frac{1}{k} \left|\{m_k \le n \le m_k + k - 1 \mid S^n\mathbf{x} \in \mathbf{A}\}\right| \\ &= \lim_{k\to\infty} \frac{1}{k} \left|\{0 \le n \le k-1 \mid S^n\mathbf{y} \in \mathbf{A}\}\right| = \nu(\mathbf{A})\end{aligned}$$

which completes the proof. ∎

Let us make a (side) remark on Proposition 4.1.7 and recall a result of a very similar flavour, the proof of which, goes along the same lines as that of Proposition 4.1.7. Surprisingly, it seems that none of the results implies the other.

Theorem 4.1.8 (Theorem 2.6 in [10]). *Let (X,T) be a topological dynamical system and let $x \in X$. The following conditions are equivalent:*

- *the point $x \in X$ is essentially recurrent, i.e. for any neighbohood U_x of x the set of visits $\{n \in \mathbb{Z} : T^n x \in U_x\}$ has positive upper Banach density;*
- *the orbit closure of x under T is measure saturated, i.e. for every nonempty open subset U of the orbit closure of x, there exists an invariant measure μ with $\mu(U) > 0$.*

Now we are ready to present our main result of this section.

Theorem 4.1.9. *For any $\nu \in \mathcal{M}^e_{\mathscr{X}}$,*

$$\boldsymbol{d}_\nu \leq \boldsymbol{D}_\nu \leq \boldsymbol{D} = \boldsymbol{d}. \tag{4.1.15}$$

Proof. Taking into account the additional supremum over all $\mathbf{x} \in \mathscr{X}$ and applying Proposition 4.1.7 to $\mathbf{A} = [1]$, we obtain that

$$\mathbf{D} = \sup_{\nu \in \mathcal{M}} \nu(1) = \sup_{\nu \in \mathcal{M}_{\mathscr{X}}} \mathbf{d}_\nu = \mathbf{d}. \tag{4.1.16}$$

Moreover, denoting $\mathscr{Y}_\nu = \mathrm{supp}(\nu) \subset \mathscr{X}$, Proposition 4.1.7 yields

$$\mathbf{D}_\nu = \lim_{n\to\infty} \frac{1}{n} \max_{w \in \mathcal{L}_n(\mathscr{Y}_\nu)} \#_1(w) = \sup_{\mathbf{x} \in \mathscr{Y}_\nu} \overline{\mathbf{d}}_B(\{n \in \mathbb{N} \mid x_n = 1\}) = \sup\{\mu(1) \mid \mu(\mathscr{Y}_\nu) = 1\} = \sup_{\mu \in \mathcal{M}(\mathscr{Y}_\nu)} \mathbf{d}_\mu.$$

It follows immediately that $\mathbf{d}_\nu \leq \mathbf{D}_\nu \leq \mathbf{D} = \mathbf{d}$. ∎

Remark 4.1.10. Clearly, each measure of full support is ones-saturated. In other words, if $\mathbf{D}_\nu < \mathbf{D}$ then ν cannot be of full support. Moreover, it follows from Theorem 4.1.9 that also each measure of maximal density is ones-saturated, whence, by Remark 4.1.6, a ones-saturated measure always exists.

Example 4.1.11. Let us present some examples which show that the inequalities in Theorem 4.1.9 can be sharp.

- $(\mathbf{d}_\nu < \mathbf{D}_\nu)$ Consider the full shift $\mathscr{X} = \{0,1\}^{\mathbb{Z}}$ and the Bernoulli measure $\nu = B_p$ where $p = \nu([1]) \in (0,1)$. Then $\mathbf{D}_\nu = 1 > \mathbf{d}_\nu = p$.
- $(\mathbf{D}_\nu < \mathbf{D})$ Consider the full shift $\mathscr{X} = \{0,1\}^{\mathbb{Z}}$ and $\nu = \frac{1}{2}(\delta_{(01)^\infty} + S\delta_{(01)^\infty})$. In this case $\mathbf{D}_\nu = \frac{1}{2} < \mathbf{D} = 1$.
- $(\mathbf{d}_\nu < \mathbf{D}_\nu < \mathbf{D})$ Consider the full shift $\mathscr{X} = \{0,1\}^{\mathbb{Z}}$, a measure $\kappa \sim \mathbf{B}^{(p)} \cdot \mathbf{Y}$, where $\mathbf{Y} \sim \nu \sim \frac{1}{2}(\delta_{(01)^\infty} + S\delta_{(01)^\infty})$, $\mathbf{Y} \amalg \mathbf{B}^{(p)}$ and $\mathbf{B}^{(p)}$ is a Bernoulli process with parameter $p \in [0,1]$. Then $\mathbf{d}_\kappa = \frac{p}{2} < \mathbf{D}_\kappa = \frac{1}{2} < \mathbf{D} = 1$ as soon as $p < 1$.

Density vs entropy

What are the relations between topological entropy $\mathbf{H} = \mathbf{H}_{\mathscr{X}}$ and densities? Let us start with the following lemma.

Lemma 4.1.12. *Let $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$ be a subshift. Suppose that $\boldsymbol{d} \leq \frac{1}{2}$. Then*

$$\mathbf{H} \leq \mathbf{H}(\boldsymbol{d}). \tag{4.1.17}$$

In particular, if $\boldsymbol{d} \to 0$ then $\mathbf{H} \to 0$.

Proof. For $\mathbf{d}=\frac{1}{2}$ the inequality $\mathbf{H}\leq\mathbf{H}(\mathbf{d})$ is obvious. Fix $\mathbf{d}<\frac{1}{2}$. It follows from the right hand side of (4.1.15) that for every $\varepsilon>0$ and sufficiently large $n\geq 1$, every allowed block of length n has at most $(\mathbf{d}+\varepsilon)n$ ones. Thus, for such n's

$$|\mathcal{L}_n|\leq\sum_{i=0}^{(\mathbf{d}+\varepsilon)n}\binom{n}{i}\leq 2^{n\mathbf{H}(\mathbf{d}+\varepsilon)}.$$

As a consequence, $\mathbf{H}\leq\mathbf{H}(\mathbf{d}+\varepsilon)$ for any $\varepsilon>0$. Taking $\varepsilon\to 0$, gives (4.1.17). ∎

Example 4.1.13. It turns out that the relations between $\mathbf{H}$ and $\mathbf{d}$ can be arbitrary.

- ($\mathbf{d}=\mathbf{H}$) Just take the full shift $\mathscr{X}=\{0,1\}^{\mathbb{Z}}$. Then $\mathbf{d}=\mathbf{D}=\mathbf{H}=1$.
- ($\mathbf{d}>\mathbf{H}$) For each zero entropy subshift $\mathscr{X}$ admitting an invariant measure different from $\delta_{\mathbf{0}}$, we have $0=\mathbf{H}<\mathbf{d}$.
- ($\mathbf{d}<\mathbf{H}$) We will construct a whole family of examples indexed by $0<p<1/2$. (Think about p to be very close to $1/2$.) Using the Jewett-Krieger theorem we may find a uniquely ergodic subshift $(\mathscr{X},\nu,S)$ measure-theoretically isomorphic to the Bernoulli shift $(\{0,1\}^{\mathbb{Z}},B_p,S)$. By the variational principle (4.1.11),

$$\mathbf{H}_{\mathscr{X}}\overset{(4.1.11)}{=}\mathbf{H}\left(\nu\right)=\mathbf{H}\left(B_p\right)=\mathbf{H}\left(p\right). \tag{4.1.18}$$

  On the other hand

$$p\leq\mathbf{d}_{\mathscr{X}}\leq 1-p. \tag{4.1.19}$$

  Indeed, by (4.1.10),

$$\mathbf{H}\left(\nu(1)\right)\overset{(4.1.10)}{\geq}\mathbf{H}\left(\nu\right)=\mathbf{H}\left(p\right).$$

  Due to the shape of binary entropy function $\mathbf{H}$, it follows that

$$p\leq\nu\left(1\right)=\mathbf{d}_{\nu}=\mathbf{d}_{\mathscr{X}}\leq 1-p.$$

  If we take $0<p<1/2$ such that $1-p<\mathbf{H}(p)$, (4.1.18) with (4.1.19) imply $\mathbf{d}<\mathbf{H}$.

Density, entropy and hereditary closures

Let us now see the relations between entropies and densities of ones if the hereditary closure of $\mathscr{X}$ is taken into account. For brevity's sake denote

$$\mathbf{d}=\mathbf{d}_{\mathscr{X}},\quad \mathbf{D}=\mathbf{D}_{\mathscr{X}},\quad \mathbf{H}=\mathbf{H}_{\mathscr{X}},\quad \widetilde{\mathbf{d}}=\mathbf{d}_{\widetilde{\mathscr{X}}},\quad \widetilde{\mathbf{D}}=\mathbf{D}_{\widetilde{\mathscr{X}}},\quad \widetilde{\mathbf{H}}=\mathbf{H}_{\widetilde{\mathscr{X}}}. \tag{4.1.20}$$

We have the following easy observations:

$$\mathbf{d}=\widetilde{\mathbf{d}}=\mathbf{D}=\widetilde{\mathbf{D}}\leq\widetilde{\mathbf{H}}\leq\mathbf{H}+\mathbf{d}\text{ and }\mathbf{H}\leq\widetilde{\mathbf{H}}. \tag{4.1.21}$$

To see this, notice that using the identity $\mathbf{D}=\mathbf{d}$ from (4.1.15), one obtains immediately that $\mathbf{d}=\widetilde{\mathbf{d}}=\mathbf{D}=\widetilde{\mathbf{D}}$. The inequality $\widetilde{\mathbf{H}}\geq\mathbf{H}$ follows from the very definition of topological entropy. Furthermore, we have the following observation.

Proposition 4.1.14. *We have*

$$\boldsymbol{d}_{\mathscr{X}}\leq\mathbf{H}_{\widetilde{\mathscr{X}}}\leq\mathbf{H}_{\mathscr{X}}+\boldsymbol{d}_{\mathscr{X}}. \tag{4.1.22}$$

Proof. Recall that $\mathbf{d}_{\mathscr{X}}=\mathbf{D}_{\mathscr{X}}=\mathbf{D}_{\widetilde{\mathscr{X}}}$. If C_n is a ones-maximal block of length n then the "downgrading argument" (here we refer to the fact that, given $\mathbf{x}\in\widetilde{\mathscr{X}}$, due to the heredity, we can replace 1's by 0's in $\mathbf{x}$ still remaining in $\widetilde{\mathscr{X}}$) yields

$$2^{\#_1(C_n)}\leq\left|\mathcal{L}^{(n)}_{\widetilde{\mathscr{X}}}\right|\leq\left|\mathcal{L}^{(n)}_{\mathscr{X}}\right|2^{\#_1(C_n)}.$$

Taking logarithms, dividing by n and passing to limits $n\to\infty$ results in

$$\mathbf{H}_{\mathscr{X}}\leq\mathbf{H}_{\widetilde{\mathscr{X}}}=\lim_{n\to\infty}\frac{1}{n}\log_2\left|\mathcal{L}^{(n)}_{\widetilde{\mathscr{X}}}\right|\leq\lim_{n\to\infty}\frac{1}{n}\log_2\left|\mathcal{L}^{(n)}_{\mathscr{X}}\right|+\lim_{n\to\infty}\frac{1}{n}\#_1\left(C_n\right)=\mathbf{H}_{\mathscr{X}}+\mathbf{D}_{\mathscr{X}}=\mathbf{H}_{\mathscr{X}}+\mathbf{d}_{\mathscr{X}},$$

which concludes the proof. ∎

Now we give some examples in the class of **topological Markov chains** which show that one cannot hope (in general) for other relations save these in (4.1.21).

Given a family $\mathcal{F} \subset \bigcup_{i=1}^{\infty}\{0,1\}^i$ of blocks, by $\mathscr{X}_{\mathcal{F}}$ we denote the set of all $\mathbf{x} \in \{0,1\}^{\mathbb{Z}}$ such that no block from $\mathcal{F}$ appears in $\mathbf{x}$ (hence, $\mathcal{F} \cap \mathcal{L}(\mathscr{X}_{\mathcal{F}}) = \varnothing$). A subshift $(\mathscr{X}, S)$ is said to be of ***finite type*** (SFT) (or a ***topological Markov chain***) if $\mathscr{X} = \mathscr{X}_{\mathcal{F}}$ for a certain finite family of blocks.

Remark 4.1.15. Note that if $\mathcal{F}$ satisfies: $C \in \mathcal{F}, C' \geq C \Rightarrow C' \in \mathcal{F}$, then $(\mathscr{X}_{\mathcal{F}}, S)$ is hereditary.

We make use of some facts from the theory of SFTs given in [70].

Example 4.1.16. Consider the ***golden mean subshift*** $\mathscr{X} = \mathscr{X}_{\{11\}}$. By Remark 4.1.15, $\mathscr{X}$ is hereditary. Moreover, by Example 4.1.4 in [70] and (4.1.15),

$$\mathbf{H} = \log_2 \frac{1+\sqrt{5}}{2} \approx 0.69 \qquad > \qquad \mathbf{d} = \frac{1}{2}. \tag{4.1.23}$$

Now, we present a SFT that is not hereditary and satisfies

$$\mathbf{H} < \mathbf{d} \text{ and } \mathbf{d} = \widetilde{\mathbf{d}} < \widetilde{\mathbf{H}}. \tag{4.1.24}$$

Example 4.1.17. Consider $\mathcal{F} = \{00, 111\}$ and $\mathscr{X} = \mathscr{X}_{\mathcal{F}}$. We claim that (4.1.24) is valid. Firstly, we show that $\mathbf{H} < \mathbf{d}$. Note that $\mathcal{F}' = \{000, 001, 100, 111\}$ is the full list of forbidden blocks of length 3 and $\mathscr{X}_{\mathcal{F}} = \mathscr{X}_{\mathcal{F}'}$. Now, the admissible blocks in $\mathscr{X}_{\mathcal{F}}$ of length 2 are 11, 10 and 01. Hence, the adjacent matrix A for this subshift is given by

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

and since A^4 has all entries positive, A is aperiodic, that is, $\mathscr{X}_{\mathcal{F}'}$ is irreducible. It follows that $\mathbf{H} = \log \lambda$, where λ is the Perron-Frobenius eigenvalue of A. Since the characteristic polynomial equals $t^3 - t - 1$, we get $\lambda \approx 1.32$ and

$$\mathbf{H} \approx \log(1.32) \approx 0.4.$$

Moreover, $\mathbf{d} = 2/3$ (consider $\mathbf{x} = \ldots 011.011011\ldots \in \mathscr{X}_{\mathcal{F}}$), which results in $\mathbf{H} < \mathbf{d}$.

Now, we turn to the proof of $\widetilde{\mathbf{d}} < \widetilde{\mathbf{H}}$. The crucial observation is that

$$\mathscr{Y} = \mathscr{X}_{\{111,1001\}} \subset \widetilde{\mathscr{X}}_{\mathcal{F}}. \tag{4.1.25}$$

Assume for a moment that (4.1.25) is true. Then, we have $\widetilde{\mathbf{H}} \geq \mathbf{H}_{\mathscr{Y}}$, so in order to show $\widetilde{\mathbf{H}} > \widetilde{\mathbf{d}}$, it is enough to bound $\mathbf{H}_{\mathscr{Y}}$ from below. We claim that

$$\mathbf{H}_{\mathscr{Y}} \approx 0.76. \tag{4.1.26}$$

In order to see (4.1.26), notice that 3-admissible blocks in $\mathscr{X}_{\{111,1001\}}$ are

$$000, 100, 010, 001, 110, 101, 011.$$

Hence, the adjacent matrix equals

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Now $A^7 > 0$, so A is aperiodic. It remains to calculate $\log \lambda$, where λ is the Perron-Frobenius eigenvalue of A, which is approximately 0.76.

We turn to the proof of (4.1.25). For $\mathbf{y} \in \mathscr{Y}$, we need to find $\mathbf{x} \in \mathscr{X}$ with $\mathbf{y} \leq \mathbf{x}$. We begin by setting $\mathbf{x} = \mathbf{y}$. Now, suppose that somewhere on $\mathbf{x}$ we see a block of the form

$$B = 1\underbrace{00\dots0}_{\ell}1. \tag{4.1.27}$$

By the definition of $\mathscr{Y}$, either $\ell = 1$ or $\ell \geq 3$. If $\ell = 1$, we do nothing. If $\ell \geq 3$ and is even, we replace B by $A = 1\underbrace{01\dots10}_{\ell}1$. If $\ell \geq 3$ and is odd, we replace B by $A = 1\underbrace{0110101010\dots1010}_{\ell}1$. We apply this procedure to all occurences of blocks of the form (4.1.27). It is easy to see that, as a result, we obtain a point $\mathbf{x}$ with the desired properties.

4.1.6 Sofic systems

Fix a finite alphabet $\mathcal{X}$ and let (G, L) be a labeled graph, i.e. G is a graph with edge set E and the labeling $\mathsf{L}\colon E \to \mathcal{X}$. Then if $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$ arises by reading the labels along the paths on G is called ***sofic*** (this term was coined by Weiss [101] and there are several equivalent ways to define sofic subshifts, see also [70]). We consider only the case of $\mathcal{X} = \{0, 1\}$.

Remark 4.1.18. Let us mention that the class of sofic shifts is precisely the class of factors of subshifts of finite type (also called topological Markov chains) given by continuous local (i.e. depending on finitely many coordinates) maps. For more information on this subject we refer to [70] (Chapters 2 and 3).

Notice that for a sofic subshift $\mathscr{X} \subset \{0, 1\}^{\mathbb{Z}}$, the subshift $\widetilde{\mathscr{X}}$ is also sofic. Indeed, take a corresponding labeled graph (G, L) for $\mathscr{X}$ and define $(\widetilde{G}, \widetilde{\mathsf{L}})$ as follows: for each edge in G labeled with 1 add an extra edge between the same vertices and label it with 0. Clearly, the subshift resulting by reading the labels along the paths in the new graph is nothing but $\widetilde{\mathscr{X}}$. Recall also that a finite union of sofic shifts remains sofic (to see this, it suffices to consider the corresponding graphs and take their disjoint union).

Remark 4.1.19. Let us now consider $\mathscr{B}$-free subshifts. It was shown in [34] that that for each finite $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$, both $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$ and $\mathscr{X}_{\mathscr{B}}$ are sofic. A simpler way to prove this is to notice that if $\mathscr{B}$ is finite then $\boldsymbol{\eta}$ is periodic. This immediately gives that $\mathscr{X}_{\boldsymbol{\eta}}$ is sofic and by the above discussion so is $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$. Moreover, $\mathscr{X}_{\mathscr{B}}$ is a finite union of the following form:

$$\mathscr{X}_{\mathscr{B}} = \bigcup_{b \in \mathscr{B}} \bigcup_{0 \leq r_b \leq b-1} \mathscr{X}_{(r_b : b \in \mathscr{B})},$$

where $x \in \mathscr{X}_{(r_b:b\in\mathscr{B})}$ iff $(\operatorname{supp} x \bmod b) \cap (b\mathbb{Z} + r_b) = \emptyset$ for each $b \in \mathscr{B}$. Notice also that $\mathscr{X}_{(r_b:b\in\mathscr{B})}$ is the hereditary closure of the subshift generated by the periodic point $x_{(r_b:b\in\mathscr{B})}$ whose support equals $\mathbb{Z} \setminus (\bigcup_{b\in\mathscr{B}}(b\mathbb{Z} + r_b))$. Thus, we can apply here the same argument as for $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$.

4.1.7 Mirsky measure

Instead of definition A central role in the theory $\mathscr{B}$-free systems is played by the **Mirsky measure** $\nu_{\boldsymbol{\eta}}$. Instead of giving the definition of $\nu_{\boldsymbol{\eta}}$, let us recall here some of its properties. In the Erdös case, $\boldsymbol{\eta}$ is a generic point for $\nu_{\boldsymbol{\eta}}$ (see [3]), i.e. we are interested in the frequency of blocks appearing in $\boldsymbol{\eta}$ (for $\boldsymbol{\eta} = \boldsymbol{\mu}^2$ they were first studied by Mirsky in [78, 79]). In general, $\boldsymbol{\eta}$ may fail to be a generic point, cf. (4.1.3). However, if (N_k) is an increasing sequence of integers realizing the lower density of $\mathcal{M}_{\mathscr{B}}$, i.e.

$$\boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}}) = \underline{d}(\mathcal{M}_{\mathscr{B}}) = \lim_{k\to\infty} \frac{1}{N_k} |[1, N_k] \cap \mathcal{M}_{\mathscr{B}}|,$$

then $\boldsymbol{\eta}$ is quasi-generic along N_k for the Mirsky measure [34] (Theorem 4.1 therein).

Remark 4.1.20. If we deal with a finite set $\mathscr{B} = \{b_1, \dots, b_n\}$ then $\boldsymbol{\eta}$ is a **periodic** point and its period is a divisor of the least common multiple of b_i's. It follows immediately that $X_{\boldsymbol{\eta}}$ is also finite and the unique shift-invariant probability measure on $X_{\boldsymbol{\eta}}$ is given by $\nu_{\boldsymbol{\eta}} = \frac{1}{N} \sum_{i=0}^{N-1} S^i \delta_{\boldsymbol{\eta}}$.

Remark 4.1.21. In the usual approach, the Mirsky measure is defined as the image of the Haar measure via a certain coding from an odometer corresponding to $\mathscr{B}$ to $\{0,1\}^{\mathbb{Z}}$. This coding first appeared in [3] in the Erdös case. For the details of the general case, we refer the reader to [34]. In particular, it follows immediately that the Mirsky measure is of zero entropy as we deal with a factor of a (uniquely ergodic) zero entropy system (in fact, the factoring map establishes an isomorphim, see Theorem F in [34]).

Entropy and intrinsic ergodicity It was shown in [34] (Proposition K) that the topological entropy of $(S, X_{\boldsymbol{\eta}})$ equals $\boldsymbol{\delta}(\mathcal{F}_{\mathscr{B}})$ (this extends the earlier results from [85, 3]). Moreover $(S, \widetilde{\mathscr{X}}_{\boldsymbol{\eta}})$ is ***intrinsically ergodic*** (i.e. has only one measure of maximal entropy) and its unique measure of maximal entropy is given by

$$\nu_{\boldsymbol{\eta}} \overset{\text{ind.}}{*} B_{\frac{1}{2}} = M(\nu_{\boldsymbol{\eta}} \otimes B_{\frac{1}{2}}),$$

where $B_{\frac{1}{2}}$ stands for the Bernoulli measure on $\{0,1\}^{\mathbb{Z}}$ with parameter $1/2$ and $M(\mathbf{x},\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ (see [34] and the earlier papers [85, 61] or combine Theorem 4.1.23 recalled below with Corollary 3.2.9).

Remark 4.1.22. In the notation of stochastic processes the unique measure of maximal entropy equals by $\mathbf{B}^{(1/2)} \cdot \mathbf{Y}^{(\boldsymbol{\eta})}$, where $\mathbf{Y}^{(\boldsymbol{\eta})} \sim \nu_{\boldsymbol{\eta}}$ and $\mathbf{B}^{(1/2)}$ stands for the i.i.d. Bernoulli process with parameter $1/2$ (recall Corollary 3.2.9 and combine it with upcoming Theorem 4.1.23 below). Note that, due to the fact that $\mathbf{B}$ has the trivial tail σ-algebra and $\mathbf{H}\left(\mathbf{Y}^{(\boldsymbol{\eta})}\right) = 0$ there is only one stationary coupling of these processes, namely the independent one (see Theorem I.2 in [41]).

It was shown in [61] (in the Erdös case and later, in [34], in the general case) that all invariant measures for $\widetilde{X}_{\boldsymbol{\eta}}$ are of the following special form.

Theorem 4.1.23. *For any $\mu \in \mathcal{M}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}}$, there exists $\rho \in \mathcal{M}_{\{0,1\}^{\mathbb{Z}} \times \mathscr{X}_{\boldsymbol{\eta}}}$ such that $\rho|_{\mathscr{X}_{\boldsymbol{\eta}}} = \nu_{\boldsymbol{\eta}}$ and $M\rho = \mu$, where $M(\mathbf{x},\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$.*

Remark 4.1.24. In terms of stochastic processes the above theorem means that if $\mathbf{Z} \in \mathcal{M}^{e}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}}$ then we can find a stationary and ergodic process $(\mathbf{X}, \mathbf{Y}^{(\boldsymbol{\eta})}) = \left(\left(X_i, Y_i^{(\boldsymbol{\eta})}\right)\right)_{i\in\mathbb{Z}}$ such that $\mathbf{Y}^{(\boldsymbol{\eta})} \sim \nu_{\boldsymbol{\eta}}$ and $\mathbf{X}\cdot\mathbf{Y}^{(\boldsymbol{\eta})} \sim \mathbf{Z}$.

Taut case Taut $\mathscr{B}$-free sets are of big importance in the theory of $\mathscr{B}$-free systems. It turns out that they carry the information about all invariant measures for **all** $\mathscr{B}$-free systems (cf. (4.1.29)). More precisely, we have the following.

Theorem 4.1.25 (Theorem C and Theorem 4.5 in [34])**.** *For each $\mathscr{B} \subset \mathbb{N}\setminus\{1\}$ there exists a unique taut $\mathscr{B}'$ such that*

$$\mathcal{F}_{\mathscr{B}'} \subset \mathcal{F}_{\mathscr{B}} \text{ and } \nu_{\boldsymbol{\eta}} = \nu_{\boldsymbol{\eta}'}, \tag{4.1.28}$$

where $\nu_{\boldsymbol{\eta}}$ and $\nu_{\boldsymbol{\eta}'}$ stand for the Mirsky measures for $\mathscr{X}_{\boldsymbol{\eta}}$ and $\mathscr{X}_{\boldsymbol{\eta}'}$ respectively.

In particular, combining this result with Theorem 4.1.23, one sees immediately that

$$\mathcal{M}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}} = \mathcal{M}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}'}}. \tag{4.1.29}$$

Moreover, recently, Keller proved the following.

Theorem 4.1.26 ([56])**.** *If $\mathscr{B}$ is taut then $\nu_{\boldsymbol{\eta}}$ has full support in $\mathscr{X}_{\boldsymbol{\eta}}$.*

Our results on the Mirsky measure

This section consists of three parts. First we give a short proof of Theorem 4.1.23 using the notion of generic points. Then, we prove the converse of Theorem 4.1.26. Finally, we describe all sets $\mathscr{B}$ for which the Mirsky measure is atomic.

Proof of Theorem 4.1.23. **Step 1. $|\mathscr{B}| < \infty$.**
Note that in this case $\mathscr{X}_{\boldsymbol{\eta}}$ is periodic (and thus uniquely ergodic) and the unique measure of maximal entropy $\nu_{\boldsymbol{\eta}}$ is periodic. In other words, in order to finish this step, it is enough to prove the following.

Proposition 4.1.27. *Suppose that $(\mathscr{Y}, S)$ is a uniquely ergodic subshift of $\{0,1\}^{\mathbb{Z}}$. Let us denote by $\nu \in \mathcal{M}_{\mathscr{Y}}$ the unique S-invariant measure. Then, for any $\mu \in \mathcal{M}^e_{\widetilde{\mathscr{Y}}}$ there exists $\rho \in \mathcal{M}^e_{(\{0,1\}^{\mathbb{Z}}\times\mathscr{Y}, S\times S)}$ such that $M\rho = \mu$ and $\rho_{|\mathscr{Y}} = \nu$.*

Remark 4.1.28. Note that $\rho_{|\mathscr{Y}} = \nu$ is automatic because by the very definition $\rho_{|\mathscr{Y}} \in \mathcal{M}_{\mathscr{Y}} = \{\nu\}$.

Proof of Proposition 4.1.27. Let $\mathbf{z} \in \widetilde{\mathscr{Y}}$ be a generic point for μ. Then there exists $\mathbf{y} \in \mathscr{Y}$ such that $\mathbf{z} \leq \mathbf{y}$. Moreover, $\mathbf{y}$ must be generic for the unique S-invariant ν on $\mathscr{Y}$. Let $\mathbf{x} \in \{0,1\}^{\mathbb{Z}}$ be such that $\mathbf{x} \cdot \mathbf{y} = \mathbf{z}$. Notice that $(\mathbf{x}, \mathbf{y})$ is quasi-generic for some measure $\rho \in \mathcal{M}_{(\{0,1\}^{\mathbb{Z}}\times\mathscr{Y}, S\times S)}$ satisfying $\rho_{|\mathscr{Y}} = \nu$. Moreover, since $\mathbf{x} \cdot \mathbf{y} = \mathbf{z}$, $M\rho = \mu$. In order to complete the proof, it suffices to use the ergodic decomposition of ρ (the image of a convex combination of measures is a convex combination of their images, with the same coefficients). ■

Step 2. $|\mathscr{B}| = \infty$.
Let $\nu_{\boldsymbol{\eta}}$ be the Mirsky measure associated with the $\mathscr{B}$-free system $\mathscr{X} := \mathscr{X}_{\boldsymbol{\eta}}$ and $\mathscr{B} = \{b_1, b_2, \dots\}$, where $b_1 < b_2 < \cdots$. For any $k \in \mathbb{N}$ define $\mathscr{B}_k = \{b_1, \dots, b_k\}$ and consider $\mathscr{X}_k := \mathscr{X}_{\boldsymbol{\eta}_k}$ with the corresponding Mirsky measure $\nu_k := \nu_{\boldsymbol{\eta}_k}$, where $\boldsymbol{\eta}_k = \mathbb{1}_{\mathcal{F}_{\mathscr{B}_k}}$. Clearly, $\boldsymbol{\eta} \leq \boldsymbol{\eta_k}$ and thus $\widetilde{\mathscr{X}} \subset \widetilde{\mathscr{X}}_k$. Take $\mu \in \mathcal{M}^e_{\widetilde{\mathscr{X}}} \subset \mathcal{M}_{\widetilde{\mathscr{X}}_k}$ and let $\mathbf{x} \in \widetilde{\mathscr{X}}$ be a generic point for μ. Since $\mathbf{x} \in \widetilde{\mathscr{X}}_k$, we can find $i \in \mathbb{N}$ such that $\mathbf{x} \leq S^i \boldsymbol{\eta}_k$. As $\mathbf{x}$ is generic (for μ) iff $S^{-i}\mathbf{x}$ is (for μ), in what follows, we assume without loss of generality, that $i = 0$. Thus, $\mathbf{x} = \mathbf{y}_k \cdot \boldsymbol{\eta}_k$ for some $\mathbf{y}_k \in \{0,1\}^{\mathbb{Z}}$. Now the $(\mathbf{y}_k, \boldsymbol{\eta}_k)$ is quasi-generic for some $\rho_k \in \mathcal{M}_{\{0,1\}^{\mathbb{Z}}\times\mathscr{X}_k}$ satisfying $\rho_{k|\mathscr{X}_k} = \nu_k$ and $M\rho_k = \mu$ (the latter property of ρ_k follows from the fact that $\mathbf{x} = \mathbf{y}_k \cdot \boldsymbol{\eta}_k$ is quasi-generic for $M\rho_k$ and generic for μ). Passing to a subsequence we can assume that $\rho_k \Rightarrow \rho$ for some $\rho \in \mathcal{M}_{\{0,1\}^{\mathbb{Z}}\times\widetilde{\mathscr{X}}}$ (note that $\bigcap_k \widetilde{\mathscr{X}}_k = \widetilde{\mathscr{X}}$). Therefore,

$$\mu = M(\rho_k) \Rightarrow M(\rho), \qquad \nu_k = \rho_{k|\widetilde{\mathscr{X}}_k} \Rightarrow \rho_{|\widetilde{\mathscr{X}}}. \tag{4.1.30}$$

Thus, our next step in the proof of Theorem 4.1.23 is the following lemma.

Lemma 4.1.29. *We have*

$$\nu_k \Rightarrow \nu_{\boldsymbol{\eta}}.$$

Proof. For simplicity's sake let $\boldsymbol{\eta}_\infty := \boldsymbol{\eta}$. Recall that $\boldsymbol{\eta}_k$ is generic for $\mathbf{Y}_{\boldsymbol{\eta}_k}$ for $k < \infty$ and quasi-generic for $k = \infty$. Thus, we can choose a common subsequence (n_i) such that for every $k \in \mathbb{N} \cup \{\infty\}$,

$$\frac{1}{n_i} \sum_{j=0}^{n_i} \delta_{S^j \boldsymbol{\eta}_k} \Rightarrow \nu_k \tag{4.1.31}$$

when $i \to \infty$. It remains to use Corollary C.0.5 along with the fact (4.1.5). ■

At the end we need to pass from ergodic measures $\mu \in \mathcal{M}^e_{\widetilde{\mathscr{X}}}$ to the non-ergodic ones. Let

$$A := \{\rho \in \mathcal{M}(X_\eta \times \{0,1\}^{\mathbb{Z}}) : \rho|_{X_\eta} = \nu_\eta\}.$$

Notice that A is a closed subset of $\mathcal{M}(\{0,1\}^{\mathbb{Z}} \times \{0,1\}^{\mathbb{Z}})$ and thus, it consitutes a compact metric space. What we have proved so far can be written down in this notation as

$$\mathcal{M}^e(\widetilde{X}_\eta) \subset M(A) \subset \mathcal{M}(\widetilde{X}_\eta).$$

Since M (as the push-forward on measures) is continuous, it follows that the image of A via M is measurable (as it is compact). Moreover, for any measure $\kappa \in M(A)$ its inverse image $M^{-1}\kappa$ is closed, whence compact. Thus, we can apply the Arsenin-Kunugui theorem on measurable selection (see e.g.

Theorem 18.18 in [55]) and it follows that there exists a measurable map $\iota\colon M(A) \to A$ such that $M \circ \iota$ is the identity map on $M(A)$. Now, fix $\kappa \in \mathcal{M}(\widetilde{X}_\eta)$ and consider its ergodic decomposition

$$\kappa = \int_{\mathcal{M}(\widetilde{X})} \kappa_\gamma dP(\gamma)$$

(where measure P is concentrated on the set of ergodic measures, but it is defined globally on the set of all invariant measures). By the measurability of ι, we can define

$$\rho := \int_{\mathcal{M}(\widetilde{X})} \iota(\kappa_\gamma) dP(\gamma)$$

(since $\gamma \mapsto \kappa_\gamma$ is measurable, so is $\gamma \mapsto \iota \circ \kappa_\gamma$). Since $\rho \in A$, it follows that $\rho|_{X_\eta} = \nu_\eta$. Moreover, we have $M_*(\int \rho_\gamma \, dP(\gamma)) = \int M_*(\rho_\gamma) \, dP(\gamma)$. This finishes the proof of Theorem 4.1.23.

■

Now, we turn to the converse of Theorem 4.1.26.

Theorem 4.1.30. *Let $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$. If the Mirsky measure $\nu_{\boldsymbol{\eta}}$ is of full support $\mathscr{X}_{\boldsymbol{\eta}}$ then $\mathscr{B}$ is taut.*

Proof. Let $\mathscr{B}$ be non-taut and let $\mathscr{B}'$ be the corresponding taut set, as in (4.1.28). Then

$$\mathscr{X}_{\boldsymbol{\eta}'} \subsetneq \mathscr{X}_{\boldsymbol{\eta}}. \tag{4.1.32}$$

Suppose for a moment that we have already proved (4.1.32). We know that $\nu_{\boldsymbol{\eta}} = \nu_{\boldsymbol{\eta}'}$ (cf. (4.1.28)). Moreover, by Theorem 4.1.26, $\mathscr{X}_{\boldsymbol{\eta}'}$ is the support of $\nu_{\boldsymbol{\eta}'}$. It follows immediately from (4.1.32) that the support of $\nu_{\boldsymbol{\eta}}$ (equal to $\mathscr{X}_{\boldsymbol{\eta}'}$) is not full.

Now we turn to the proof of (4.1.32). We will prove first that $\mathscr{X}_{\boldsymbol{\eta}'} \subset \mathscr{X}_{\boldsymbol{\eta}}$. By Theorem 4.1.26, $\nu_{\boldsymbol{\eta}'}$ is of full support $\mathscr{X}_{\boldsymbol{\eta}'}$, i.e. each block appearing in $\boldsymbol{\eta}'$ is of positive $\nu_{\boldsymbol{\eta}'}$-measure. By (4.1.28), we have $\nu_{\boldsymbol{\eta}} = \nu_{\boldsymbol{\eta}'}$, i.e. each block appearing in $\boldsymbol{\eta}'$ is of positive $\nu_{\boldsymbol{\eta}}$-measure. Since $\boldsymbol{\eta}$ is a quasi-generic point for $\nu_{\boldsymbol{\eta}}$, each block of positive $\nu_{\boldsymbol{\eta}}$-measure appears in $\boldsymbol{\eta}$. Therefore, each block appearing in $\boldsymbol{\eta}'$ appears also on $\boldsymbol{\eta}$, which gives $\mathscr{X}_{\boldsymbol{\eta}'} \subset \mathscr{X}_{\boldsymbol{\eta}}$.

Suppose now that $\mathscr{X}_{\boldsymbol{\eta}'} = \mathscr{X}_{\boldsymbol{\eta}}$. In particular, we have $\boldsymbol{\eta} \in \mathscr{X}_{\boldsymbol{\eta}'} \subset \mathscr{X}_{\mathscr{B}'}$. Therefore, for each $b' \in \mathscr{B}'$, there exists $1 \leq r' \leq b'$ such that $\mathcal{F}_{\mathscr{B}} \cap (b'\mathbb{Z} + r') = \emptyset$, i.e. $b'\mathbb{Z} + r' \subset \mathcal{M}_{\mathscr{B}}$. Let $d = \gcd(b', r')$. For $b'' = b'/d$, $r'' = r'/d$, we have

$$d(b''\mathbb{Z} + r'') \subset \mathcal{M}_{\mathscr{B}} \subset \mathcal{M}_{\mathscr{B}'}.$$

It follows by this and by (4.1.8) that $\boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}'}) = \boldsymbol{\delta}(\mathcal{M}_{\mathscr{B}' \cup \{d\}})$. By (4.1.7), we obtain $d \in \mathcal{M}_{\mathscr{B}'}$. Hence, there exists $b''' \in \mathscr{B}'$ such that $b'''|d$, i.e. we have $b''' \mid d \mid b'$. Thus, by the primitivity of $\mathscr{B}'$, we obtain $b''' = d = b'$. Therefore, $r' = b'$ and we conclude that $b'\mathbb{Z} \subset \mathcal{M}_{\mathscr{B}}$. Since $b' \in \mathscr{B}'$ was arbitrary, it follows that $\mathcal{M}_{\mathscr{B}} = \mathcal{M}_{\mathscr{B}'}$. Now, it remains to use the primitivity of $\mathscr{B}$ and $\mathscr{B}'$ to conclude that $\mathscr{B} = \mathscr{B}'$. This yields a contradiction and completes the proof.

■

Remark 4.1.31. It is a classical fact in the theory of cut-and-project sets that for any $\mathscr{B}$, the Mirsky measure $\nu_{\boldsymbol{\eta}}$ is a measure of maximal density for $(\mathscr{X}_{\boldsymbol{\eta}}, S)$ (see e.g. Theorem 4 and Corollary 4 in [57], cf. Chapter 7 in [7] as well; alternatively see Corollary 4.1.32 below). Therefore, in order to obtain a maximal density measure without full support, it is enough to consider a $\mathscr{B}$-free system that is not taut and take its Mirsky measure. Furthermore, we always have $\mathbf{d}_{\mathscr{X}_{\boldsymbol{\eta}}} = \mathbf{H}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}}$ (see Proposition K in [34]).

Corollary 4.1.32. *The Mirsky measure $\nu = \nu_{\boldsymbol{\eta}}$ has maximal density in $\mathscr{X}_{\boldsymbol{\eta}}$. If $\mathscr{B}$ is not taut then ν is not of full support.*

Proof. Firstly, we show that the Mirsky measure is of maximal density of ones. By Theorem 4.1.25, there exists a unique taut $\mathscr{B}'$ such that

$$\mathcal{F}_{\mathscr{B}'} \subset \mathcal{F}_{\mathscr{B}} \qquad \text{and} \qquad \nu = \nu',$$

where $\nu' = \nu_{\eta'}$ stands for the Mirsky measures for $\mathscr{X}_{\eta'}$. By Keller's theorem (see Theorem 4.1.26), ν' must be of full support. In particular, ν' is of maximal density (recall Theorem 4.1.9). Using, (4.1.29) we get

$$d_\nu = d_{\nu'} = \sup_{\mu \in \mathcal{M}_{\widetilde{\mathscr{X}}_{\eta'}}} d_\mu = \sup_{\mu \in \mathcal{M}_{\widetilde{\mathscr{X}}_{\eta}}} d_\mu.$$

If $\mathscr{B}$ is not taut then by Theorem 4.1.30 ν cannot have full support. ■

As an immediate consequence (recall Theorem 4.1.9) we obtain the following fact.

Corollary 4.1.33. *The Mirsky measure is ones-saturated.*

Remark 4.1.34. In [33], it was shown that $\mathscr{B}$-free systems that are minimal, are necessarily taut. Notice that this also follows immediately from Corollary 4.1.30, as in minimal systems all invariant measures have full support.

Now, we describe all sets $\mathscr{B}$ for which the Mirsky measure is atomic.

Proposition 4.1.35. *The Mirsky measure ν_η is atomic if and only if the taut set $\mathscr{B}'$ given by (4.1.28) is finite.*

Proof. Clearly, if $\mathscr{B}'$ is finite then the corresponding Mirsky measure is atomic. We will prove now the other implication. In view of (4.1.28), we can assume that $\mathscr{B}$ itself is taut, and we need to prove that in this case $\mathscr{B}$ is finite. But if $\mathscr{B}$ is taut then by Theorem F in [34] the measure-theoretic dynamical system $(\mathscr{X}_\eta, \nu_\eta, S)$ is isomorphic to a rotation on a certain compact Abelian group considered with Haar measure. However, Haar measure has an atom if and only if the group is finite. Since the group is given by the inverse limit of cyclic groups $\mathbb{Z}/\mathrm{lcm}(\{b \in \mathscr{B} : b \leq K\})$, $K \geq 1$, $\mathscr{B}$ itself is finite. ■

Corollary 4.1.36. *The Mirsky measure ν_η is atomic if and only if for some $k, \ell \geq 1$,*

$$\mathscr{B} = c_1\mathscr{B}_1 \cup \cdots \cup c_k\mathscr{B}_k \cup \{c_1', \ldots, c_\ell'\}, \tag{4.1.33}$$

with $\mathscr{B}_1, \ldots, \mathscr{B}_k$ being Behrend.

Proof. Let $\mathscr{B}'$ be as in (4.1.28). It follows by the construction of the taut set $\mathscr{B}'$ in Section 4.2 in [34] that either

$$\mathscr{B}' = (\mathscr{B} \setminus (c_1\mathbb{Z} \cup \ldots \cup c_n\mathbb{Z})) \cup \{c_1, \ldots, c_n\} \tag{4.1.34}$$

and

$$\mathscr{B} = (\mathscr{B} \setminus (c_1\mathbb{Z} \cup \ldots \cup c_n\mathbb{Z})) \cup (c_1\mathscr{B}_1 \cup \cdots \cup c_n\mathscr{B}_n) \tag{4.1.35}$$

for some $n \geq 1$ and some Behrend sets $\mathscr{B}_1, \ldots, \mathscr{B}_n$ or

$$\mathscr{B}' = (\mathscr{B} \setminus \bigcup_{n \geq 1} c_n\mathbb{Z}) \cup \{c_n : n \geq 1\}$$

and

$$\mathscr{B} = (\mathscr{B} \setminus \bigcup_{n \geq 1} c_n\mathbb{Z}) \cup \bigcup_{n \geq 1} c_n\mathscr{B}_n$$

for some Behrend sets $\mathscr{B}_n, n \geq 1$.

The finiteness of $\mathscr{B}'$ means that (4.1.34) and (4.1.35) hold for some $n \geq 1$. In particular, the set $\mathscr{B} \setminus (c_1\mathbb{Z} \cup \cdots \cup c_n\mathbb{Z})$ is finite, i.e. (4.1.33) holds. ■

4.1.8 Multiplicative convolution of measures

Recall that $M(\mathbf{x},\mathbf{y})=\mathbf{x}\cdot\mathbf{y}$ and given subshifts $\mathscr{X},\mathscr{Y}\subset\mathbb{R}^{\mathbb{Z}}$ and measures $\nu\in\mathcal{M}_{\mathscr{X}}$ and $\mu\in\mathcal{M}_{\mathscr{Y}}$, the ***independent multiplicative convolution of*** ν and μ is defined by $\nu\overset{\text{ind.}}{*}\mu=M(\nu\otimes\mu)$. We already know that $(\widetilde{\mathscr{X}}_{\boldsymbol{\eta}},S)$ is intrinsically ergodic and the measure of maximal entropy for $(\widetilde{\mathscr{X}}_{\boldsymbol{\eta}},S)$ is equal to $\nu_{\boldsymbol{\eta}}\overset{\text{ind.}}{*}B_{\frac{1}{2}}$ where $B_{\frac{1}{2}}$ stands for the Bernoulli measure on $\{0,1\}^{\mathbb{Z}}$ with parameter $1/2$.

In this part we will study measure of the form

$$\kappa=\nu\overset{\text{ind.}}{*}B_{\frac{1}{2}},\tag{4.1.36}$$

where $\nu\in\mathcal{M}^e_{\mathscr{X}}$. Later, we will apply these general facts to $\nu_{\boldsymbol{\eta}}\overset{\text{ind.}}{*}B_{\frac{1}{2}}$.

It is not hard to see that $\kappa=\nu\overset{\text{ind.}}{*}B_{\frac{1}{2}}$ is of full support as soon as ν is. Moreover, $\kappa=\nu\overset{\text{ind.}}{*}B_{\frac{1}{2}}$ is ergodic whenever ν is. Indeed, κ is a factor of the product of a mixing and an ergodic system.

Lemma 4.1.37. *Let $\nu\in\mathcal{M}_{\mathscr{X}}$. Then for $\kappa=\nu\overset{\text{ind.}}{*}B_{\frac{1}{2}}$ and each $C\in\mathcal{L}_{\widetilde{\mathscr{X}}}$ we have*

$$\kappa(C)=\sum_{\mathcal{L}_{\mathscr{X}}\ni C'\geq C}\nu(C')\cdot 2^{-\#_1(C')}.\tag{4.1.37}$$

Proof. Let $\kappa=\nu\overset{\text{ind.}}{*}B_{\frac{1}{2}}\sim\mathbf{M}=\mathbf{B}\cdot\mathbf{Y}$ where process $\mathbf{B}$ has the symmetric Bernoulli distribution and is independent of $\mathbf{Y}$. Then, due to $\mathbf{B}\amalg\mathbf{Y}$, the conditioning on $Y_{[0,n]}$ gives

$$\mathbb{P}\left(M_{[0,n]}=m_{[0,n]}\right)=\mathbb{E}\mathbb{E}\left(\mathbb{1}_{M_{[0,n]}=m_{[0,n]}}\mathbb{1}_{Y_{[0,n]}\geq m_{[0,n]}}|Y_{[0,n]}\right)=\mathbb{E}\mathbb{1}_{Y_{[0,n]}\geq m_{[0,n]}}2^{\sum_{i=0}^n Y_i}$$

which is equivalent to the desired formula. ∎

Remark 4.1.38. Notice that $\kappa=\nu\overset{\text{ind.}}{*}B_{\frac{1}{2}}$ is descreasing in the sense that for any two words of length n, w_1, w_2 such that $w_1\leq w_2$, we have

$$\kappa\left(w_1\right)\geq\kappa\left(w_2\right).\tag{4.1.38}$$

4.1.9 Ones-maximal blocks

In our proof of absence of Gibbs property, the main role is played by the family of ones-maximal blocks. We say that a block $C\in\mathcal{L}_{\mathscr{X}}$ is ***ones-maximal*** if

$$\#_1\left(C\right)=\max_{W\in\mathcal{L}_{|C|}(\mathscr{X})}\#_1\left(W\right).\tag{4.1.39}$$

Analogously, for any measure $\nu\in\mathcal{M}_{\mathscr{X}}$, we call block $C\in\mathcal{L}_{\mathscr{X}}$ ν-***ones-maximal*** if

$$\#_1\left(C\right)=\max_{W\in\mathcal{L}_{|C|}(\mathscr{X}),\ \nu(W)>0}\#_1\left(W\right).\tag{4.1.40}$$

Remark 4.1.39. Notice that if C is ν-ones-maximal (or ones-maximal) then (4.1.37) simplifies to

$$\nu\overset{\text{ind.}}{*}B_{\frac{1}{2}}(C)=\nu(C)\cdot 2^{-\#_1(C)}.\tag{4.1.41}$$

4.1.10 Gibbs property

Measure $\kappa\in\mathcal{M}^e_{\mathscr{X}}$ is said to have the ***Gibbs property*** if there exists a constant $a>0$ such that

$$\kappa(C)\geq a\cdot 2^{-|C|\mathbf{H}_{\mathscr{X}}}\tag{4.1.42}$$

for all blocks $C\in\mathcal{L}_{\mathscr{X}}$ having **positive** κ-measure.

Remark 4.1.40. Let us recall here that the notion of Gibbs measures comes from statistical physics [90, 64] and it corresponds to the idea of equilibrium states of complicated physical systems. They turned out to be an interesting object also from the point of view of dynamics and have played an important role in ergodic theory (see, e.g., [17, 94]). Given a finite alphabet $\mathcal{X}$ and a (Hölder) continuous potential $\varphi\colon \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ and a subshift $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$, a measure $\mu_\varphi \in \mathcal{M}_{\mathscr{X}}$ is called a ***Gibbs measure*** for φ, whenever there exist constants $\mathscr{P} = \mathscr{P}_{\varphi,\mathscr{X}} \geq 0$ and $b = b(\varphi, \mathscr{X}) > 0$ such that for every $\mathbf{x} \in \mathscr{X}$

$$b^{-1} \leq \frac{\mu_\varphi(x_{[0,n-1]})}{2^{\sum_{k=0}^{n-1}\varphi(S^k\mathbf{x})-n\mathscr{P}}} \leq b \text{ for any } n \geq 1. \tag{4.1.43}$$

One can show that the above constant $\mathscr{P}$ is equal to the topological pressure of $\mathscr{X}$ (with underlying potential φ). In particular, if $\varphi = 0$ then $\mathscr{P} = \mathbf{H}_{\mathscr{X}}$. Moreover, if we consider just the lowerbound in (4.1.43) with $\varphi = 0$ then we arrive at the definition of the Gibbs property (recall (4.1.42)).

Let us now explain our motivation to study the Gibbs property defined above. In many natural situations, like sofic systems [101] or systems enjoying particular specification properties and beyond (see [25, 26] and the references therein), there is a unique measure of maximal entropy and it enjoys the Gibbs property or a weakening of it. More than that, by a result of B. Weiss [101], if κ satisfies the Gibbs property and is a measure of maximal entropy, then $(\mathscr{X}, S)$ is intrinsically ergodic. We are interested in examples, where (4.1.42) fails, but the system under consideration remains intrinsically ergodic. This yields natural classes of positive entropy intrinsically ergodic systems different from many known so far.

By the variational principle for entropy (4.1.11), if $\mathbf{H} = 0$ then $\mathbf{H}(\kappa) = 0$ for any $\kappa \in \mathcal{M}$. In general, if $\mathbf{H} > 0$, it is hard to say for which κ we have $\mathbf{H}(\kappa) > 0$. However, we have the following simple observation.

Proposition 4.1.41. *Suppose that $\kappa \in \mathcal{M}$ has full support and satisfies Gibbs property* (4.1.42). *Then*

$$\mathbf{H}(\kappa) \geq a\mathbf{H}.$$

Proof. Without a loss of generality we can assume that $\mathbf{H} > 0$. Let $\ell_n = |\mathcal{L}_n|$. Notice that (4.1.9) implies that $\log \ell_n \geq n\mathbf{H}$ for any $n \in \mathbb{N}$, i.e. we have

$$\ell_n \geq 2^{n\mathbf{H}}. \tag{4.1.44}$$

Moreover, the function $x \mapsto -x\log x$ is increasing for $x \leq 1/2$. Due to the full support of κ and the Gibbs property (4.1.42), we obtain

$$-\sum_{W\in\mathcal{L}_n} \kappa(W)\log\kappa(W) \geq \sum_{W\in\mathcal{L}_n,\ \kappa(W)\leq 1/2} a2^{-n\mathbf{H}}\left[n\mathbf{H} - \log(a)\right]. \tag{4.1.45}$$

Since only one atom of the partition given by $\mathcal{L}_n$ can have the measure larger than $\frac{1}{2}$, it follows that

$$\sum_{W\in\mathcal{L}_n,\ \kappa(W)\leq 1/2} a2^{-n\mathbf{H}}\left[n\mathbf{H} - \log(a)\right] \geq (\ell_n - 1)a2^{-n\mathbf{H}}\left[n\mathbf{H} - \log(a)\right]. \tag{4.1.46}$$

Now, we apply (4.1.44) to get

$$(\ell_n - 1)a2^{-n\mathbf{H}}\left[n\mathbf{H} - \log a\right] \geq a\left(2^{n\mathbf{H}} - 1\right)2^{-n\mathbf{H}}\left[n\mathbf{H} - \log a\right] = a\left(1 - 2^{-n\mathbf{H}}\right)\left(n\mathbf{H} - \log a\right). \tag{4.1.47}$$

Combining (4.1.45), (4.1.46) and (4.1.47), we obtain

$$\mathbf{H}(\kappa) \xleftarrow[n\to\infty]{} \frac{-\sum_{W\in\mathcal{L}_n}\kappa(W)\log(\kappa(W))}{n} \geq a\left(1 - 2^{-n\mathbf{H}}\right)\left(\mathbf{H} - \frac{\log(a)}{n}\right) \xrightarrow[n\to\infty]{} a\mathbf{H}$$

and the result follows. ■

Remark 4.1.42. If $\mathscr{X} = \{0,1\}^{\mathbb{Z}}$ and κ is an ergodic measure of full support with the Gibbs property then $\kappa = B_{\frac{1}{2}}$. Indeed, the inequality in (4.1.42) can be rewritten as $\kappa(C) \geq a \cdot B_{\frac{1}{2}}(C)$ for each block C. In particular, $B_{\frac{1}{2}} \ll \kappa$ and the claim follows from the ergodicity of κ and $B_{\frac{1}{2}}$.

Remark 4.1.43. Note also that if $\mathbf{H} = 0$ then κ cannot have the Gibbs property unless it is purely atomic.

Since $\kappa = \nu \overset{\text{ind.}}{*} B_{\frac{1}{2}}$ is of our special interest, at the end, let us give the following fact concerning the rate of convergence in the formula for topological entropy.

Proposition 4.1.44. *Suppose we can find $\nu \in \mathcal{M}_{\mathscr{X}}$ such that $\kappa = \nu \overset{\text{ind.}}{*} B_{\frac{1}{2}}$ satisfies Gibbs property (4.1.42) and has full support. Let us denote $|\mathcal{L}_{\mathscr{X}}^{(n)}|$ by ℓ_n. Then for every $n \in \mathbb{N}$,*

$$0 \leq \frac{\log(\ell_n)}{n} - \mathbf{H} \leq \frac{1}{n}\log\left(\frac{1}{a}\right).$$

Proof. It follows from the "decreasing property" (4.1.38) that for any $n \in \mathbb{N}$ there exists a maximal word (in the sense of the coordinatewise order) $W_n^{min} \in \mathcal{L}_{\mathscr{X}}^{(n)}$ such that for every $W_n \in \mathcal{L}_{\mathscr{X}}^{(n)}$, we have $\kappa(W_n) \geq \kappa\left(W_n^{min}\right)$. Then

$$\kappa\left(W_n^{min}\right) \leq \frac{1}{\ell_n}.$$

Now, taking advantage of the Gibbs property, we get

$$a2^{-\mathbf{H}n} \leq \kappa(W_n^{min}) \overset{\text{Rema. 4.1.39}}{=} \nu\left(W_n^{min}\right)2^{-|W_n^{min}|} \leq 2^{-\log(\ell_n)}.$$

Thus, $a \leq 2^{-n\left[\frac{\log(\ell_n)}{n} - \mathbf{H}\right]}$ and finally $n\left[\frac{\log(\ell_n)}{n} - \mathbf{H}\right] \leq \log\left(\frac{1}{a}\right)$, which gives the desired rate of convergence. ■

4.2 Results

4.2.1 Gibbs property in periodic case

Theorem 4.2.1. *Suppose that $\nu \in \mathcal{M}_{\mathscr{X}}^{e}$ is purely atomic. Then $\kappa = \nu \overset{\text{ind.}}{*} B_{\frac{1}{2}}$ has the Gibbs property.*

Proof. Since ν is atomic, it follows immediately that ν is concentrated on a finite orbit, i.e. there exists $\mathbf{x} \in \mathscr{X}$ and $k \geq 1$ with $S^k\mathbf{x} = \mathbf{x}$ and we have

$$\nu = \frac{1}{k}(\delta_{\mathbf{x}} + \delta_{S\mathbf{x}} + \cdots + \delta_{S^{k-1}\mathbf{x}}).$$

Thus, since in the definition of Gibbs property we must check only what happens on the support of ν, we may assume that $\mathscr{X} = \{\mathbf{x}, S\mathbf{x}, \ldots, S^{k-1}\mathbf{x}\}$. It follows from Section 3.2.1 in [61] (or from our Corollary 3.2.9) that the (unique) measure of maximal entropy for $(\widetilde{\mathscr{X}}, S)$ is of the form

$$\kappa = \nu \overset{\text{ind.}}{*} B_{\frac{1}{2}}.$$

Now, $(\widetilde{\mathscr{X}}, S)$ as the hereditary closure of finite subshift is sofic. Therefore, its measure of maximal entropy has the Gibbs property. ■

Remark 4.2.2. As a matter of fact, if $\mathbf{x} \in \{0,1\}^{\mathbb{Z}}$ is periodic of period $k \geq 1$ and $\mathscr{X} = \{S^j\mathbf{x} \colon j = 0, \ldots, k-1\}$ then $\mathbf{H}_{\widetilde{\mathscr{X}}} = \mathbf{d}_{\mathscr{X}} =: \mathbf{d}$ and $(\widetilde{\mathscr{X}}, S)$ is intrinsically ergodic with $\kappa = \nu \overset{\text{ind.}}{*} B_{\frac{1}{2}}$ being the measure of maximal entropy (combine Corollary 3.2.9 with Theorem 3.2.1). By the monotonicity property (4.1.38), we need to check (4.1.42) only for the maximal blocks and for such, by (4.1.41), we obtain

$$\kappa(B) = \nu(B)2^{-\#_1(B)} \geq \frac{1}{k}2^{-n\mathbf{d}} = \frac{1}{k}2^{-n\mathbf{H}_{\widetilde{\mathscr{X}}}},$$

so κ has the Gibbs property and one can take $a = \frac{1}{k}$ in (4.1.42).

4.2.2 Absence of Gibbs property

Let us present our main result.

Theorem 4.2.3. *Fix $(\mathscr{X}, S)$ and suppose that $\nu \in \mathcal{M}^e_{\mathscr{X}}$ is ones-saturated and non-atomic. If $\mathbf{D} = \mathbf{H}_{\widetilde{\mathscr{X}}}$ then $\kappa = B_{\frac{1}{2}} \overset{\text{ind.}}{*} \nu$* ***does not*** *have Gibbs property.*

Proof. Let us start with an easy observation. Let $\nu \in \mathcal{M}_{\mathscr{X}}$ and $a > 0$. Suppose that a sequence of blocks C_n satisfies $|C_n| \nearrow \infty$ and $\nu(C_n) \geq a$. Then there exists a subsequence (n_k) such that $\bigcap_{k\geq 1} C_{n_k} \neq \emptyset$. Moreover, we have $\nu(\{x\}) \geq a$ for $\{x\} = \bigcap_{k\geq 1} C_{n_k}$. Indeed, for any $k \geq 1$, there exists $B \in \mathcal{L}_k(\mathscr{X})$ such that for infinitely many $n \in \mathbb{N}$, we have $C_n[0, k-1] = B$. Now, it is enough to apply a diagonal procedure to find the required (n_k) for which $\nu(\bigcap_{k\geq 1} C_{n_k}) = \nu(\{x\}) \geq a$.

Now, for $n \in \mathbb{N}$, let $C_n \in \mathcal{L}_n$ be ν-ones-maximal. Define $o_n = \#_1(C_n)$ and $\widetilde{\mathbf{H}} = \mathbf{H}_{\widetilde{\mathscr{X}}}$. Due to Lemma 4.1.37, for $\kappa = \nu \overset{\text{ind.}}{*} B_{\frac{1}{2}}$, we have

$$\kappa(C_n) \cdot 2^{n\widetilde{\mathbf{H}}} = \nu(C_n) \cdot 2^{n\widetilde{\mathbf{H}} - o_n}.$$

Using (4.1.9), we get $\mathbf{d} = \mathbf{D} = \mathbf{D}_\nu \leq o_n/n$. Therefore, using the assumption that $\widetilde{\mathbf{H}} = \mathbf{d}$, we obtain $n\widetilde{\mathbf{H}} - o_n \leq 0$. If we could find some $a > 0$ for which $\nu(C_n) \geq a$ holds for infinitely many $n \in \mathbb{N}$ then the observation made on the very beginning would imply that ν is **not** atomless. This is not possible because we assumed otherwise. The proof is concluded. ■

Recall that for the $\mathscr{B}$-free systems we have $d_{\mathscr{X}_\eta} = \mathbf{H}_{\widetilde{\mathscr{X}_\eta}}$. Moreover, the Mirsky measure ν_η is of maximal density. Furthermore, $\nu_\eta \overset{\text{ind.}}{*} B_{\frac{1}{2}}$ is the unique measure of maximal entropy. Therefore, the above theorem immediately answers the question asked by Peckner concerning $\mathscr{B}$-free subshifts. More precisely, we have the following.

Corollary 4.2.4. *Let $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$. Suppose that the Mirsky measure ν_η is not atomic. Then the (unique) measure of maximal entropy of $(\widetilde{\mathscr{X}}_\eta, S)$* ***does not have*** *the Gibbs property.*

Theorem 4.2.3 goes beyond the $\mathscr{B}$-free context. For example, if $(\mathscr{X}, S)$ is of zero topological entropy, it follows from Lemma 2.2.16 in [61] (or our Proposition 4.1.14) that $\mathbf{d} = \mathbf{H}_{\widetilde{\mathscr{X}}}$. Thus, as a consequence of Theorem 4.2.3, we obtain the following result.

Corollary 4.2.5. *If $(\mathscr{X}, S)$ is uniquely ergodic and $\mathbf{H}_{\mathscr{X}} = 0$, then $B_{\frac{1}{2}} \overset{\text{ind.}}{*} \nu$ has no Gibbs property whenever the unique invariant measure ν is non-atomic.*

At last but not least, in [61], Sturmian sequences are discussed (we refer the reader to [61]). It is proved that the hereditary closure of the system given by any Sturmian sequence yields an intrinsically ergodic system whose measure of maximal entropy is of the form $\nu \overset{\text{ind.}}{*} B_{\frac{1}{2}}$. Moreover, in this case we also have $\mathbf{d} = \widetilde{\mathbf{H}}$. Using again Theorem 4.2.3 we obtain the following corollary.

Corollary 4.2.6. *If $(\widetilde{\mathscr{X}}, S)$ is a Sturmian hereditary system then its measure of maximal entropy has no Gibbs property.*

4.2.3 Topological pressure

Remark 4.2.7. At the beginning let us note that the following results are very recent and are not a part of any preprint. Moreover, we are still working on many aspects of this subject. Here, we would like to explain the motivation behind Theorem 3.2.34.

Let ν_η be the Mirsky measure associated with a $\mathscr{B}$-free system $(\mathscr{X}_\eta, S)$. Moreover, if $\mathscr{B} = \{b_1, b_2, \ldots\}$ is infinite with $b_1 < b_2 < \cdots$, we define its natural approximations $\mathscr{B}_k = \{b_1, \ldots, b_k\}$ where $k \in \mathbb{N}$. Thus, for every $k \in \mathbb{N}$ we can consider $\mathscr{X}_{\eta_k}$ and the associated Mirsky measure ν_{η_k}, where $\eta_k = \mathbb{1}_{\mathcal{F}_{\mathscr{B}_k}}$. Recall that each measure ν_{η_k} is periodic. It is intuitively clear that $\widetilde{\mathscr{X}}_{\eta_k}$ can be treated as a approximation for $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$, thus, in particular, it should be true that the topological pressure of $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}_k}$ converges to that of $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$. Furthermore, we already know that $\nu_{\boldsymbol{\eta}_k} \Rightarrow \nu$ (recall Lemma 4.1.29).

Now we make the above argument strict. For brevity's sake let $\boldsymbol{\eta}_\infty = \boldsymbol{\eta}$. Firstly, we take care of the convergence of the topological pressure of approximations $\mathscr{X}_{\boldsymbol{\eta}_k}$. The crucial observation is contained in the fact that $\mathscr{X}_{\boldsymbol{\eta}_k}$ is a **descending** sequence of sets.

Lemma 4.2.8. *Let $\mathscr{X}_k$ be a decreasing sequence of subshifts, that is $\mathscr{X}_k \supset \mathscr{X}_{k+1}$. Then for any upper semi-continuous potential φ,*

$$\mathscr{P}_{\mathscr{X}_k,\varphi} \to \mathscr{P}_{\mathscr{X},\varphi}, \tag{4.2.1}$$

where $\mathscr{X} = \bigcap_{k\geq 1} \mathscr{X}_k$.

Proof. By a basic monotonicity property (with respect to the underlying space) of topological pressure, $\mathscr{P}_{\mathscr{X}_k,\varphi} \geq \mathscr{P}_{\mathscr{X}_{k+1},\varphi}$ and thus

$$\liminf_{k\to\infty} \mathscr{P}_{\mathscr{X}_k,\varphi} \geq \mathscr{P}_{\mathscr{X},\varphi} \tag{4.2.2}$$

On the other hand, if μ_k are the equilibrium states for φ (on $\mathscr{X}_k$) then, due to the upper semi-continuity of entropy rate and variational principle,

$$\limsup_{k\to\infty} \mathscr{P}_{\mathscr{X}_k,\varphi} = \limsup_{k\to\infty} \left[\mathbf{H}(\mu_k) + \int \varphi d\mu_k\right] \leq \mathbf{H}(\mu) + \int \varphi d\mu \leq \mathscr{P}_{\mathscr{X},\varphi}, \tag{4.2.3}$$

where, without loss of generality, we assumed that $\mu_k \Rightarrow \mu$ for some μ. ■

Since $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}_k} \supset \widehat{\mathscr{X}}_{\boldsymbol{\eta}_{k+1}}$, for any $k \in \mathbb{N}$ and $\bigcap_{k\geq 1} \widetilde{\mathscr{X}}_{\boldsymbol{\eta}_k} = \widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$, we immediately get the following result.

Lemma 4.2.9. *For every upper semi-continuous potential φ,*

$$\mathscr{P}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}_k},\varphi} \to \mathscr{P}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}_\infty},\varphi}. \tag{4.2.4}$$

The second observation concerns the weak convergence of $\mathbf{Y}_{\boldsymbol{\eta}_k}$. (The lemma below has been already proven in the ergodic setting in Lemma 4.1.29. However, for convenience of the reader we recall here its short proof.)

Lemma 4.2.10. *We have*

$$\mathbf{Y}_{\boldsymbol{\eta}_k} \Rightarrow \mathbf{Y}_{\boldsymbol{\eta}_\infty}. \tag{4.2.5}$$

Proof. Recall that $\boldsymbol{\eta}_k$ is generic for $\mathbf{Y}_{\boldsymbol{\eta}_k}$ for $k < \infty$ and quasi-generic for $k = \infty$. Thus, we can choose a common subsequence (n_i) such that for every $k \in \mathbb{N} \cup \{\infty\}$,

$$\frac{1}{n_i} \sum_{j=0}^{n_i} \delta_{S^j \boldsymbol{\eta}_k} \Rightarrow \mathbf{Y}_{\boldsymbol{\eta}_k} \tag{4.2.6}$$

when $i \to \infty$. It remains to use Corollary C.0.5 along with (4.1.5). ■

Let us now present our main theorem of this section.

Theorem 4.2.11. *For any $\mathscr{B}$-free system $(\mathscr{X}_{\boldsymbol{\eta}}, S)$ and a continuous potential $\varphi\colon \{0,1\}^{\mathbb{Z}} \to \mathbb{R}$ we have*

$$\mathscr{P}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}},\varphi} = \mathbb{P}(Y_0 = 1) + \sup_{\mathbf{z}\in\{0,1\}^{\mathbb{Z}}} \Phi(\mathbf{z}),$$

where $\nu_{\boldsymbol{\eta}} \sim \mathbf{Y}$ is the Mirsky measure and $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$ is the upgrade of φ given by (3.2.24).

Proof. By Theorem 4.1.23, for any $k \in \mathbb{N} \cup \{\infty\}$, all measures $\mu \in \mathcal{M}_{\widetilde{\mathscr{X}}_{\eta_k}}$ are of the form

$$\mu \sim \mathbf{X} \cdot \mathbf{Y}_k,$$

where $\mathbf{Y}_k \sim \nu_{\boldsymbol{\eta}_k}$, $\mathbf{X} \in \mathcal{M}_{\{0,1\}^{\mathbb{Z}}}$ and $(\mathbf{X}, \mathbf{Y}_k)$ is stationary. In other words,

$$\mathcal{N}_{\mathbf{Y}_k} = \mathcal{M}_{\widetilde{\mathscr{X}}_{\eta_k}}. \tag{4.2.7}$$

Recall also that by Lemma 4.2.10, we have $\mathbf{Y}_k \Rightarrow \mathbf{Y}$. Therefore, using the variational principle, (4.2.7) and Theorem 3.2.34, we obtain

$$\mathscr{P}_{\widetilde{\mathscr{X}}_{\eta_k},\varphi} = \sup_{\mu \in \mathcal{M}_{\widetilde{\mathscr{X}}_{\eta_k}}} \mathbf{H}(\mu) + \int \varphi d\mu = \sup_{\mu \in \mathcal{N}_{\mathbf{Y}_k}} \mathbf{H}(\mu) + \int \varphi d\mu \to \mathbb{P}(Y_0 = 1) + \sup_{\mathbf{z}} \Phi(\mathbf{z}).$$

On the other hand, by Lemma 4.2.9, for every continuous potential φ, $\mathscr{P}_{\widetilde{\mathscr{X}}_{\eta_k},\varphi} \to \mathscr{P}_{\widetilde{\mathscr{X}}_{\boldsymbol{\eta}},\varphi}$, which concludes the proof. ∎

4.3 Open questions

In view of Theorem 4.2.11 which provides an explicit formula for the topological pressure of a $\mathscr{B}$-free system, it would be interesting to describe any of the corresponding equilibrium measures for $\widetilde{\mathscr{X}}_{\boldsymbol{\eta}}$. So far we known only, that such an example can be obtained as a weak limit of certain multiplicative convolutions of Gibbs-like i.i.d. processes with (periodic) approximations of the Mirsky measure (recall (3.2.33) in Theorem 3.2.32). However, we know nothing about properties of this limiting process. In particular, it remains open if (or more precisely, under which conditions) a system like in Theorem 4.2.11 admits only one equilibrium measure (we know only that this happens if the underlying potential depends on one coordinate, recall Theorem 3.2.21).

Chapter 5

m-dependent random variables

5.1 Introduction

The class of m-dependent random variables was studied in many papers including [1, 73, 19, 51, 59, 97]. Such variables can be treated as a middle ground between the classical case of independent random variables and strongly mixing processes (like α-mixing or β-mixing sequences, see Sections B.3.3 and B.3.4). They are highly correlated with Markov chains, due to the splitting method (see Section 6.1.11), which allows one to split a Markov chain into one-dependent blocks (in fact, this is why we started to study this class of random variables). Many facts and theorems for independent random variables are still valid in the m-dependent case, take for example the central limit theorem or the strong law of large numbers. However, there are many questions which remain unanswered (see Section 5.4). Among them there is one concerning finding an "optimal" Bernstein inequality, which we explore in this section.

5.2 Background

5.2.1 Definitions and examples

Recall that a process $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ is ***m-dependent*** if for any $k \in \mathbb{Z}$, $(X_i)_{i\leq k}$ is independent of $(X_i)_{i\geq k+m+1}$. Thus, for example, if $m = 0$ then $\mathbf{X}$ is an ***independent process***. Let us give now some examples.

Example 5.2.1 (**Block factors of stochastic processes**)**.** Consider an independent process $\boldsymbol{\xi} = (\xi_i)_{i\in\mathbb{Z}}$ where $\xi_i \in \mathcal{X}$ and a measurable function $f\colon \mathcal{X}^{m+1} \to \mathcal{Y}$. Put $X_i = f(\xi_i, \xi_{i+1}, \ldots, \xi_{i+m-1})$. Any such processes $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ is called an ***m-block factor of an i.i.d. process***. Clearly, by the very definitions, $\mathbf{X}$ is an $m-1$-dependent process. Moreover, if $\boldsymbol{\xi}$ is stationary then so is $\mathbf{X}$. More generally, if $X_i = f(Y_{[i,i+m-1]})$ for some process $\mathbf{Y}$ then we say that $\mathbf{X}$ is an ***m-block factor of*** $\mathbf{Y}$.

Example 5.2.2 (**m-dependent Markov chains**)**.** Let $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ be a Markov chain on a finite state space $\mathcal{X}$ and let $P = [p(x,y)]_{x,y\in\mathcal{X}}$ be its transition matrix. Then $\mathbf{X}$ is m-dependent iff P^{m+1} has identical rows. It can be shown that every stationary one-dependent Markov chain with $|\mathcal{X}| \leq 4$ is in fact a two-block factor (see [1] Corollary below Theorem 3). Moreover, Matúš in [73] gives an explicit example of a 5-state stationary Markov chain which is one-dependent but cannot be expressed as a 2-block factor of an i.i.d. process (see Consequence in [73]).

Remark 5.2.3. For a long time there was a conjecture that every 1-dependent process is in fact a two-block factor of some i.i.d. process. As mentioned above, this is not true. In fact, Matúš showed that it is not true even for Markov chains. Moreover, the authors in [19] gave an explicit construction of a stationary one-dependent process (in fact, a whole family of such processes) that are not m-factors (of an i.i.d. process) for any $m \in \mathbb{N}$. Recently, in [51] the authors presented a natural class of coloring processes which are m-dependent but are not k-factors (of an i.i.d. process) for any k. For example they showed that there exists 1-dependent 4-coloring of $\mathbb{Z}$ (part of Theorem 1 in [51]) and no r-block factor (of an i.i.d. process) q-coloring exists for any r and q (Proposition 2 in [51]).

Example 5.2.4 (Renewal processes)**.** Aaronson in [1] showed that every stationary, one-dependent renewal process is in fact a 2-block factor.

Example 5.2.5 (See [59])**.** For an explicit easy yet non-trivial example of one-dependent process consider mutually independent $B_i \sim Bern(p)$, $\xi_i \sim Rade(1/2)$, that is $\mathbb{P}(B_i = 1) = 1 - \mathbb{P}(B_i = 0) = p$ and $\mathbb{P}(\xi_i = 1) = 1 - \mathbb{P}(\xi_i = -1) = 1/2$ and put

$$X_i = B_i\xi_i + (1 - B_i)\xi_{i-1}\xi_{i-2}. \tag{5.2.1}$$

Clearly $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ is stationary and $X_i \sim B_i\xi_i + (1-B_i)\xi_{i-1}\xi_{i-2} \sim B_i\xi_i + (1-B_i)\xi_{i-1} \sim Rade(1/2)$. One can show that X_i is pairwise independent. More surprisingly, this 3-block factor is also a one-dependent process (see Corollary 2 in [59]). Best to our knowledge it is not known if it can be expressed as a 2-block factor (of an i.i.d. process). Furthermore, $\mathbf{X}$ is not a Markov chain of any order.

Example 5.2.6 (Longest alternating sequence)**.** The authors in [52] showed (among many other results) that the length of the longest alternating sequence of a random uniform permutation can be expressed as a 3-block factor of an i.i.d. process (see Proposition 2.2 or equation (4) therein).

5.2.2 Bernstein inequality. Introduction

Let us recall the classical Bernstein inequality for bounded functions.

Theorem 5.2.7 (Classical Bernstein inequality)**.** *If $(\xi_i)_i$ is a sequence of i.i.d. centered random variables such that $\|\xi_i\|_\infty \le M$ then for any $t > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n \xi_i\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{2n\sigma^2 + \frac{2}{3}tM}\right), \tag{5.2.2}$$

where $\sigma^2 = \mathbb{E}\xi_i^2$.

Notice that if $tM \ll n\sigma^2$ then (5.2.2) reflects the CLT behaviour of the partial sums $\sum_{i=1}^n \xi_i$ which should be of order $2\exp\left(-\frac{t^2}{2n\sigma^2}\right)$. For this reason, we speak of the ***Gaussian part*** of Bernstein's inequality to refer to this part of the right-hand side of (5.2.2). We also say that (5.2.2) is optimal, meaning that its Gaussian part is optimal. (We refer (slightly imprecisely) to the remaining part of (5.2.2) as to the ***Poisson part*** even though it is not of "Poisson order" $t\log t$ for large t.)

Consider **a stationary m-dependent, bilateral process $\mathbf{X}$** such that $\mathbb{E}X_i = 0$ and $\|X_i\|_\infty \le M < \infty$ for each $i \in \mathbb{Z}$. The ***asymptotic variance*** σ_∞^2 is given by

$$\sigma_\infty^2 = \lim_{n\to\infty} \frac{1}{n} Var(X_1 + \cdots + X_n) = \mathbb{E}X_1^2 + 2\sum_{i=1}^m \mathbb{E}X_1X_i. \tag{5.2.3}$$

Thus, $(1/\sqrt{n})(X_1 + \cdots + X_n) \Rightarrow \mathcal{N}(0, \sigma_\infty^2)$ and the "ideal Bernstein inequality" for the partial sums should be of the form

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{2n\sigma_\infty^2 + cMt}\right), \tag{5.2.4}$$

where σ_∞^2 is as in (5.2.3) and c is some numerical constant. We do not have any counterexample to (5.2.4) but, at the same time, we do not have a slightest idea how to show (5.2.4).

Note that in the Gaussian part of (5.2.4) we have used the asymptotic variance σ_∞^2 (5.2.3) and not the variance of a single random variable X_i, $\sigma^2 = \mathbb{E}X_i^2$. In general, these two variances can be quite different. Firstly, by Hölder's inequality, we always have $\sigma_\infty^2 \le (m+1)\sigma^2$. If $\mathbb{E}X_1X_i \ge 0$ for all i, then clearly, $\sigma_\infty^2 \ge \sigma^2$. However, if $\mathbb{E}X_1X_i \le 0$ for all i then σ_∞^2 can be arbitrarily small when compared to σ^2. For the extreme example of this phenomenon consider an i.i.d. process $\boldsymbol{\xi} = (\xi_i)_{i\in\mathbb{Z}}$ and $X_i = \xi_i - \xi_{i-1}$. In this case $\sigma_\infty^2 = 0$, whereas $\sigma^2 = 2\mathbb{E}\xi_i^2$. In fact, it turns out (cf. [53]) that the reverse is true, that is if for a 1-dependent, bounded stationary process $(X_i)_{i\in\mathbb{N}}$ we have $\sigma_\infty^2 = 0$ then there exists an i.i.d. process $(\xi_i)_{i\in\mathbb{N}}$ such that $X_i = \xi_{i+1} - \xi_i$.

Now we show how one can immediately get a version of (5.2.4) with σ_∞^2 replaced by $\sigma^2 = \mathbb{E}X_i^2$. Clearly, since $\mathbf{X}$ is m-dependent and stationary, processes $\left(X_{i(m+1)+k}\right)_{i\in\mathbb{Z}}$, where $0 \le k \le m$, are i.i.d. Thus, splitting the sum $\sum_{i=1}^n X_i$ into $(m+1)$ sums of the form $\sum_{i=1}^{\lfloor n/(m+1)-k\rfloor} X_{i(m+1)+k}$, using the union bound and applying the classical Bernstein's inequality (5.2.2), we obtain

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge t\right) &\le 2(m+1)\exp\left(-\frac{(t^2/(m+1))^2}{2\lceil n/(m+1)\rceil\sigma^2 + \frac{2}{3}Mt/(m+1)}\right)\\ &\le 2(m+1)\exp\left(-\frac{t^2}{2(m+1)(n+m+1)\sigma^2 + \frac{2}{3}(m+1)Mt}\right).\end{aligned} \tag{5.2.5}$$

Note that unlike in (5.2.4), the Poisson and Gaussian parts of right hand-side of (5.2.5), depend on m.

5.3 Results

This section is divided into two parts. In the first one (Section 5.3.1) we consider arbitrary $m \ge 1$ but we restrict our attention to **bounded random variables**. This additional assumption allows us to present the key ideas in a simplified form. However, sometimes (e.g. in Chapter 6 where we study the phenomenon of concentration of measure for additive functionals of Markov chains) **more general integrability conditions** are necessary. Here, the case $m = 1$ is of particular interest (cf. Section 6.1.11, where we describe the splitting method for Markov chains). We provide the corresponding versions of the Bernstein inequality under this extra assumption (see Section 5.3.2).

5.3.1 Bernstein inequality for bounded bounded random variables, $m \ge 1$

Best to our knowledge, it remains open if the optimal Bernstein inequality (5.2.4) holds for arbitrary m-dependent stationary, bounded random variables X_i, $i \in \mathbb{N}$. Nevertheless, the argument used in our paper [66] results in a nearly optimal (up to constants depending on m) version of (5.2.4) in some cases of interest, including functions of m-dependent Markov chains and k-block factors of i.i.d. processes. The whole idea is based on the observation that we can replace X_i's by Z_i's in such a way that $\sum_{i=1}^n X_i \approx \sum_{i=1}^n Z_i$, Z_i are k-dependent (relations between k and m may depend on the very case) and, most importantly, $\mathrm{Var}\,(Z_i) = \sigma_\infty^2$ (thus the variance of a single random variable Z_i is equal to the asymptotic variance of process $\mathbf{X}$). To construct Z_i, we use a filtration satisfying certain technical properties.

Lemma 5.3.1 (Variance of modification of X_i's)**.** *Let $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ be a stationary centered square-integrable process and $m \in \mathbb{N}$. Suppose that we can find a filtration $\mathcal{G} = (\mathcal{G}_i)_{i\in\mathbb{Z}}$ such that X_i is $\mathcal{G}_i$-measurable for all $i \in \mathbb{N}$ and for*

$$Z_i = \sum_{j=i}^{i+m}\left[\mathbb{E}\left(X_j|\mathcal{G}_i\right) - \mathbb{E}\left(X_j|\mathcal{G}_{i-1}\right)\right] \tag{5.3.1}$$

we have the following:

1. *$(Z_i)_{i\ge1}$ is stationary.*
2. *For any $1 \le p \le m$ process $\left(\mathbb{E}\left(X_{i+p}|\mathcal{G}_i\right)\right)_{i\ge1}$ is stationary.*
3. *For any $i \ge 1$, $\mathcal{G}_{i-1}$ is independent of X_{i+m}.*
4. *In case of $m > 1$ let for any $0 \le p \le q \le m+1$ process $\left(X_{i+q}\mathbb{E}\left(X_{i+p}|\mathcal{G}_i\right)\right)_{i\in\mathbb{Z}}$ be stationary.*

Then

$$\mathbb{E}Z_i^2 = \sigma_\infty^2 = \mathbb{E}X_1^2 + 2\sum_{j=2}^{m+1}\mathbb{E}X_1X_j. \tag{5.3.2}$$

Remark 5.3.2. For brevity's sake, every filtration $\mathcal{G}$ which satisfies conditions 1-4 from Lemma 5.3.1 will be called ***nice*** (***for*** $\mathbf{X}$).

Before we proceed with the proof of this lemma let us give examples for which a nice filtration can be easily found.

Example 5.3.3 (m**-block factors**)**.** Let $\mathbf{X}$ be an $(m+1)$-block factor of an i.i.d. process, i.e. $X_i = f(\xi_i, \xi_{i+1}, \ldots, \xi_{i+m})$ for $i \in \mathbb{N}$ and a measurable and bounded function f. Clearly, $(X_i)_{i\in\mathbb{Z}}$ is m-dependent and stationary. Define a filtration

$$\mathcal{F}_i = \sigma\left(\xi_{(-\infty,i+m])}\right). \tag{5.3.3}$$

We show now that $\mathbf{Z}$ given by (5.3.1) is an $(m+1)$-block factor of $\boldsymbol{\xi}$. More precisely,

$$Z_i = F\left(\xi_{[i,i+m]}\right)$$

for some measurable function F which does **not** depend on i.

By the stationarity of $\boldsymbol{\xi}$, there exist functions F_k for $0 \le k \le m$ such that

$$\mathbb{E}(X_j|\mathcal{F}_i) = \mathbb{E}\left(f(\xi_{[j,j+m]}|\xi_{(-\infty,i+m]}\right) = \mathbb{E}\left(f(\xi_{[j,j+m]})|\xi_{[j,i+m]}\right) = F_{j-i}(\xi_{[j,i+m]}).$$

for $i \le j \le i+m$. We also have

$$Z_i = \sum_{j=i}^{i+m} \mathbb{E}\left(X_j|\mathcal{F}_i\right) - \mathbb{E}\left(X_j|\mathcal{F}_{i-1}\right) = \sum_{j=0}^{m} \mathbb{E}\left(X_{j+i}|\mathcal{F}_i\right) - \mathbb{E}\left(X_{j+i}|\mathcal{F}_{i-1}\right) = F(\xi_i, \xi_{i+1}, \ldots, \xi_{i+m}).$$

It is now a pure routine to check that $(\mathcal{F}_i)$ is nice.

Example 5.3.4 (l**-Markov chains**)**.** Let $\mathbf{X}$ be a 1-block factor of a stationary m-dependent l-Markov chain $\mathbf{Y}$, with $l \ge m$, i.e. $X_i = f(Y_i)$, $i \in \mathbb{N}$, for a measurable and bounded function f. Consider the natural filtration associated with $\mathbf{Y}$,

$$\mathcal{F}_i = \sigma\left(Y_{(-\infty,i]}\right).$$

We show now that $\mathbf{Z}$ given by (5.3.1) is an l-block factor of $\mathbf{Y}$. More precisely,

$$Z_i = F(Y_{[i-l-1,i]})$$

for some measurable function F which does **not** depend on i. In particular, $\mathbf{Z}$ is an $(m+l-1)$-dependent stationary process.

By the l-Markov property and the stationarity of $\mathbf{Y}$, for $0 \le k \le m$, there exist functions F_k such that

$$\mathbb{E}(X_j|\mathcal{F}_i) = \mathbb{E}\left(f(Y_j)|Y_{[i-l,i]}\right) = F_{j-i}(Y_{[i-l,i]}) \tag{5.3.4}$$

for $i \le j \le i+m$. Thus,

$$Z_i = \sum_{j=i}^{i+m} \mathbb{E}\left(X_j|\mathcal{F}_i\right) - \mathbb{E}\left(X_j|\mathcal{F}_{i-1}\right) = F(Y_{[i-l-1,i]}).$$

Once again, it is easy to check that $(\mathcal{F}_i)$ is nice.

Proof of Lemma 5.3.1. For brevity's sake, let $\mathbb{E}_i^p(X) = [\mathbb{E}\left(X|\mathcal{G}_i\right)]^p$ for any random variable X, $p = 1, 2$ and $i \in \mathbb{Z}$. Notice that by assumption 1, Z_i share the same distribution and thus it is enough to prove (5.3.2) for $i = 1$. We have

$$\begin{aligned}\mathbb{E}Z_1^2 = \mathbb{E}\left(\sum_{j=1}^{m+1} \mathbb{E}_1(X_j) - \mathbb{E}_0(X_j)\right)^2 &= \sum_{j=1}^{m+1} \mathbb{E}\left(\mathbb{E}_1(X_j) - \mathbb{E}_0(X_j)\right)^2 \\ + 2 \sum_{1\le j<j'\le m+1} \mathbb{E}\left(\mathbb{E}_1(X_j) - \mathbb{E}_0(X_j)\right)&\left(\mathbb{E}_1(X_{j'}) - \mathbb{E}_0(X_{j'})\right) = I + 2II.\end{aligned}$$

Clearly, if we show that $I = \mathbb{E}X_1^2$ and $II = \sum_{2\le j\le m+1} \mathbb{E}X_1X_j$ then the proof will be concluded.

Firstly, we consider I. For any $i \in \mathbb{Z}$ and $j \in \mathbb{N}$ put $v_j = \mathbb{E}\mathbb{E}_i^2(X_{i+j})$ and notice that v_j is well-defined (independent of i) due to assumption 2. Moreover, since $X_1 \in \mathcal{G}_1$ and we have assumption 3, we get that $v_0 = \mathbb{E}X_1^2$ and $v_{m+1} = 0$. Hence, using properties of conditional expectation, for any $j \ge 0$, we obtain

$$\mathbb{E}\left(\mathbb{E}_1(X_j) - \mathbb{E}_0(X_j)\right)^2 = \mathbb{E}\mathbb{E}_1^2(X_j) + \mathbb{E}\mathbb{E}_0^2(X_j) - 2\mathbb{E}\mathbb{E}_1(X_j)\mathbb{E}_0(X_j) = \mathbb{E}\mathbb{E}_1^2(X_j) - \mathbb{E}\mathbb{E}_0^2(X_j) = v_{j-1} - v_j.$$

and thus $I = \sum_{j=1}^{m+1} v_{j-1} - v_j = v_0 - v_{m+1} = \mathbb{E}X_1^2$.

To see the formula for II, note that for any $1 \le j < j' \le m+1$ we have

$$\begin{aligned}&\mathbb{E}\left(\mathbb{E}_1(X_j) - \mathbb{E}_0(X_j)\right)\left(\mathbb{E}_1(X_{j'}) - \mathbb{E}_0(X_{j'})\right) = \mathbb{E}X_{j'}\mathbb{E}_1(X_j) - \mathbb{E}X_j\mathbb{E}_0(X_{j'}) - \mathbb{E}X_{j'}\mathbb{E}_0(X_j) + \mathbb{E}X_j\mathbb{E}_0(X_{j'})\\ &= \mathbb{E}X_{j'}\mathbb{E}_1(X_j) - \mathbb{E}X_{j'}\mathbb{E}_0(X_j) = \mathbb{E}X_{j'}\mathbb{E}_1(X_j) - \mathbb{E}X_{j'+1}\mathbb{E}_1(X_{j+1}),\end{aligned}$$

where in the last equality we have used assumption 4 (note that if $m = 1$ then $j = 1$, $j' = 2$ and instead of assumption 4 one can use property 3; this is why assumption 4 is redundant in this case). Therefore

$$\begin{aligned} II &= \sum_{1\le j<j'\le m+1} \left[\mathbb{E}X_{j'}\mathbb{E}_1(X_j) - \mathbb{E}X_{j'+1}\mathbb{E}_1(X_{j+1})\right]\\ &= \sum_{1\le j<j'\le m+1} \mathbb{E}X_{j'}\mathbb{E}_1(X_j) - \sum_{2\le j<j'\le m+2} \mathbb{E}X_{j'}\mathbb{E}_1(X_j)\\ &= \sum_{2\le j'\le m+1} \mathbb{E}X_{j'}\mathbb{E}_1(X_1) - \sum_{2\le j\le m+1} \mathbb{E}X_{m+2}\mathbb{E}_1(X_j) = \sum_{2\le j'\le m+1} \mathbb{E}X_1X_{j'} + 0,\end{aligned}$$

where in the last line we have used assumption 3 and the fact that X_i is $\mathcal{G}_i$-measurable. ■

Now, we show how to obtain a version of the Bernstein inequality (5.2.4), using process $\mathbf{Z}$ defined in (5.3.1).

Theorem 5.3.5. *For $i \in \mathbb{Z}$ let X_i be bounded centered random variables satisfying $\|X_i\|_\infty \le M < \infty$. Suppose that a filtration $\mathcal{G} = (\mathcal{G}_i)_{i\in\mathbb{Z}}$ is such that $X_i \in \mathcal{G}_i$ and $\mathbf{Z} = (Z_i)_{i\in\mathbb{Z}}$ given by*

$$Z_i = \sum_{j=i}^{i+m} \left[\mathbb{E}\left(X_j|\mathcal{G}_i\right) - \mathbb{E}\left(X_j|\mathcal{G}_{i-1}\right)\right]$$

satisfies the assumptions 1-4 from Lemma 5.3.1. If $\mathbf{Z}$ is k-dependent for some $k \in \mathbb{N}$ then

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge t\right) \le 2(k+1)\exp\left(-\frac{t^2}{c_k(n+1+k)\sigma_\infty^2 + d_{k,m}tM}\right), \tag{5.3.5}$$

where $c_k = 2(1 + \frac{3}{2\log(2k+2)})^2(k+1)$, $d_{k,m} = \frac{4}{3}(1 + \frac{3}{2\log(2k+2)})(k+1)(m+1)$.

Remark 5.3.6. Note that the right hand-side of (5.3.5) is monotonic with respect to σ_∞^2. Thus (for $k \sim m$), in terms of the Gaussian parts, (5.3.5) is stronger than (5.2.5) only when $\sigma_\infty^2 \le \sigma^2$.

Proof. Once more, for brevity's sake, denote $\mathbb{E}_i(X) = \mathbb{E}\left(X|\mathcal{G}_i\right)$ for any random variable X and $i \in \mathbb{Z}$. Firstly, notice that

$$\sum_{i=1}^n X_i = \sum_{i=1}^n Z_i - \sum_{j=1}^m \left[\mathbb{E}_n\left(X_{n+j}\right) - \mathbb{E}_0\left(X_j\right)\right]. \tag{5.3.6}$$

Indeed,

$$\begin{aligned}&\sum_{i=1}^n Z_i = \sum_{i=1}^n\sum_{j=i}^{i+m}\left[\mathbb{E}_i\left(X_j\right) - \mathbb{E}_{i-1}\left(X_j\right)\right] = \sum_{i=1}^n\sum_{j=i}^{i+m}\mathbb{E}_i\left(X_j\right) - \sum_{i=0}^{n-1}\sum_{j=i+1}^{i+m+1}\mathbb{E}_i\left(X_j\right)\\ &= \sum_{i=1}^{n-1} X_i - \sum_{i=1}^{n-1}\mathbb{E}_i\left(X_{i+m+1}\right) + \sum_{j=n}^{n+m}\mathbb{E}_n\left(X_j\right) - \sum_{j=1}^{m+1}\mathbb{E}_0\left(X_j\right) = \sum_{i=1}^n X_i + \sum_{j=1}^m\left[\mathbb{E}_n\left(X_{n+j}\right) - \mathbb{E}_0\left(X_j\right)\right].\end{aligned}$$

Notice that without loss of generality, $t > d_{k,m}M$ (otherwise the right hand-side of (5.3.5) exceeds one). Define $\epsilon > 0$ by $(1-\epsilon)^{-1} = 1 + \frac{3}{2\log(2k+2)}$ and notice that $t\epsilon > 2Mm \geq \| \sum_{j=1}^{m} \mathbb{E}\left(X_{n+j}|\mathcal{F}_n\right) - \mathbb{E}\left(X_j|\mathcal{F}_0\right) \|_\infty$. Now, using (5.3.6) we get

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) \leq \mathbb{P}\left(\left|\sum_{i=1}^{n} Z_i\right| \geq (1-\epsilon)t\right) + \mathbb{P}\left(\left|\sum_{j=1}^{m} \mathbb{E}\left(X_{n+j}|\mathcal{F}_n\right) - \mathbb{E}\left(X_j|\mathcal{F}_0\right)\right| \geq \epsilon t\right) \\
&= \mathbb{P}\left(\left|\sum_{i=1}^{n} Z_i\right| \geq (1-\epsilon)t\right) \leq \sum_{j=0}^{k} \mathbb{P}\left(\left|\sum_{i=1,\ldots,n,\ (k+1)|(i-j)} Z_i\right| \geq t(1-\epsilon)/(k+1)\right).
\end{aligned} \tag{5.3.7}$$

Applying the classical Bernstein inequality (5.2.2) for i.i.d. sequences and the k-dependence of (Z_i) we obtain

$$\begin{aligned}
\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) &\leq 2(k+1)\exp\left(-\frac{\left(\frac{t(1-\epsilon)}{k+1}\right)^2}{2\lceil n/(k+1)\rceil\sigma_\infty^2 + \frac{2}{3}2(m+1)M\frac{t(1-\epsilon)}{k+1}}\right) \\
&\leq 2(k+1)\exp\left(-\frac{t^2}{2(1-\epsilon)^{-2}(k+1)(n+k+1)\sigma_\infty^2 + \frac{4}{3}(1-\epsilon)^{-1}(k+1)(m+1)M}\right).
\end{aligned}$$

■

Combining Theorem 5.3.5 with the observations made in Examples 5.3.3 and 5.3.4, we immediately get the following two corollaries.

Corollary 5.3.7. *Let $\mathbf{X} = (f(Y_i))_{i\in\mathbb{Z}}$ be a factor of a stationary m-dependent l-Markov chain $\mathbf{Y}$, with $l \geq m$. Then*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) \leq 2(m+l+1)\exp\left(-\frac{t^2}{c_{m,l}(n+m+l)\sigma_\infty^2 + d_m tM}\right), \tag{5.3.8}$$

where $c_{m,l} = 2(1+\frac{3}{2\log(2m+2l)})^2(m+l)$, $d_{m,l} = \frac{4}{3}(1+\frac{3}{2\log(2m+2l)})(m+l)(m+1)$.

Corollary 5.3.8. *Let $\mathbf{X}$ be an m-block factor of an i.i.d. process. Then (5.3.8) holds with constants $c_{m,l}$ and $d_{m,l}$ replaced by $c_{m,1}$ and $d_{m,1}$, respectively.*

Remark 5.3.9. In both above corollaries the constants c and d depend (at least) on m, whereas in the "ideal Bernstein inequality" (5.2.4) there is no such dependence. The following natural question arises: are there any natural examples indicating that the dependence on m is essential? Notice that due to the union bound used in the proof, our method is not sufficient to get rid of this dependence.

5.3.2 Bernstein inequality for unbounded random variables, $m = 1$

In this section, we present two versions of Bernstein inequality for one-dependent random variables: for suprema of sums and randomly stopped sums. They are later used in the proofs of our main theorems concerning additive functional of Markov chains (see, for example, Theorem 6.2.1). As usual,

$$\sigma_\infty^2 = \mathbb{E}X_1^2 + 2\mathbb{E}X_1X_2$$

stands for the asymptotic variance of one-dependent stationary process $\mathbf{X}$. Recall that

$$\|X\|_{\psi_\alpha} = \inf\left\{c > 0 \mid \mathbb{E}\exp\left(\frac{|X|^\alpha}{c^\alpha}\right) \leq 2\right\}.$$

stands for the exponential Orlicz norm (for more information, see Chapter D).

For convenience of the reader, let us start with recalling two versions of the Bernstein inequality for suprema of independent random variables: the bounded case and the case of Orlicz integrable random variables.

Theorem 5.3.10 (Classical Bernstein inequality; supremum case). *If $(\xi_i)_i$ is a sequence of i.i.d. centered random variables such that $\|\xi_i\|_\infty \le M$ then for any $t > 0$,*

$$\mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k}\xi_i\right| \ge t\right) \le 2\exp\left(-\frac{t^2}{2n\sigma^2 + \frac{2}{3}t\|\xi_i\|_\infty}\right),$$

where $\sigma^2 = \mathbb{E}\xi_i^2$.

Proof. It is just a special case of the Azuma-Bernstein inequality. See e.g. Theorem A in [36]. Alternatively, one can combine a Chernoff-like type of argument along with the Doob maximal inequality for martingales. ∎

Now, we show how Bernstein's inequality changes if we replace the boundedness condition $\|X_i\|_\infty < \infty$ by the integrability with respect to the Orlicz norm.

Lemma 5.3.11. *Let $(\xi_i)_{i\ge 0}$ be i.i.d. sequence of random variables such that $\|\xi_i\|_{\psi_\alpha} \le c$ for some $c > 0$ and $0 < \alpha \le 1$. If $U_i = \xi_i \mathbb{1}_{|\xi_i|>M}$ then for $\lambda = (2^{1/\alpha}c)^{-1}$,*

$$\mathbb{E}\exp\left(\lambda^\alpha \sum_{i=0}^{n-1}\left(|U_i|^\alpha + (\mathbb{E}|U_i|)^\alpha\right)\right) \le \exp(8). \tag{5.3.9}$$

Furthermore, if $\mathbb{E}\xi_i = 0$ then for any $t > 0$ and $n \in \mathbb{N}$

$$\mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k}\xi_i\right| > t\right) \le \exp(8)\exp\left(-\frac{t^\alpha}{2(6c)^\alpha}\right) + 2\exp\left(-\frac{t^2}{\frac{72}{25}n\sigma^2 + \frac{8}{5}tM}\right),$$

where $M = c(3\alpha^{-2}\log n)^{\frac{1}{\alpha}}$ and $\sigma^2 = \mathbb{E}\xi_i^2$.

Proof. The first part of the lemma (5.3.9) is just the content of Lemma 4.1 in [5].

Now, we prove the tail inequality for $\sum_{i=1}^{n}\xi_i$. Fix $p = 1/6$, define $B_i = \xi_i \mathbb{1}_{|\xi_i|\le M}$, $\overline{B_i} = B_i - \mathbb{E}B_i$, $\overline{U_i} = U_i - \mathbb{E}U_i$ and notice that $\xi_i = \overline{B_i} + \overline{U_i}$. Therefore, the union bound implies that

$$\mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k}\xi_i\right| > t\right) \le \mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k}\overline{U_i}\right| > tp\right) + \mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k}\overline{B_i}\right| > t(1-p)\right).$$

Firstly, we take care of the unbounded part. Using the Markov inequality, $\alpha \le 1$ and (5.3.9),

$$\mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k}\overline{U_i}\right| > tp\right) \le \exp\left(-\frac{t^\alpha p^\alpha}{2c^\alpha} + 8\right) = \exp\left(-\frac{t^\alpha}{2(6c)^\alpha} + 8\right). \tag{5.3.10}$$

Now, we turn to the bounded part. Notice that $\mathbb{E}\overline{B_i}^2 \le \mathbb{E}B_i^2 \le \sigma^2$. Therefore, Theorem 5.3.10 yields (recall that $p = 1/6$)

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\overline{B_i}\right| > t(1-p)\right) \le 2\exp\left(-\frac{t^2(1-p)^2}{2n\sigma^2 + \frac{4}{3}t(1-p)M}\right) = 2\exp\left(-\frac{t^2}{\frac{72}{25}n\sigma^2 + \frac{8}{5}tM}\right).$$

The proof is concluded. ∎

Now we turn to the one-dependent case.

Lemma 5.3.12 (Bernstein inequality for suprema of partial sums). *Let $(X_i)_{i\ge 0}$ be a 1-dependent sequence of centered random variables such that $\|X_i\|_{\psi_\alpha} \le c$ for some $\alpha \in (0,1]$ and $c > 0$. Assume that there exists a filtration $(\mathcal{F}_i)_{i\ge 0}$ such that for*

$$Z_i = X_i + \mathbb{E}\left(X_{i+1}|\mathcal{F}_i\right) - \mathbb{E}\left(X_i|\mathcal{F}_{i-1}\right) \tag{5.3.11}$$

we have the following:

0. X_i is $\mathcal{F}_i$-measurable,

1. $(Z_i)_{i\geq 1}$ is stationary,

2. $(Z_i)_{i\geq 1}$ is m-dependent with $m=1$ or $m=2$,

3. $(\mathbb{E}(X_i|\mathcal{F}_{i-1}))_{i\geq 1}$ is stationary,

4. For any $i\geq 1$, $\mathcal{F}_{i-1}$ is independent of X_{i+1}.

Then $\mathbb{E}Z_i^2=\sigma_\infty^2$ and

$$\|Z_i\|_{\psi_\alpha}\leq c(8/\alpha)^{\frac{1}{\alpha}}, \tag{5.3.12}$$

where $\|Z_i\|_{\psi_\alpha}$ stands for the exponential Orlicz norm of Z_i (see Appendix D). Moreover, for any $t>0$ and $n\in\mathbb{N}$,

$$\mathbb{P}\left(\sup_{1\leq k\leq n}\left|\sum_{i=1}^k X_i\right|>t\right)\leq K_m\exp\left(-\frac{t^\alpha}{u_m c^\alpha}\right)\ +L_m\exp\left(-\frac{t^2}{v_{n,m}\sigma_\infty^2+w_{n,m}t}\right) \tag{5.3.13}$$

where $u_m=\frac{16\cdot 8^\alpha(m+1)^\alpha}{\alpha}$, $v_{n,m}=5(m+1)(n+m+1)$, $w_{n,m}=2(m+1)(24\alpha^{-3}\log n)^{\frac{1}{\alpha}}c$, $K_m=2(m+1)\exp(8)$ and $L_m=2(m+1)$.

Remark 5.3.13. Note that Z_i in (5.3.11) are defined in exactly the same manner as Z_i's from Lemma 5.3.5 for $m=1$.

Proof. Assume for a moment that (5.3.12) holds. We show now how to combine Lemma 5.3.11 and (5.3.12) to obtain (5.3.13). Firstly, notice that the assumptions of Lemma 5.3.1 are satisfied and thus $\mathbb{E}Z_i^2=\sigma_\infty^2$. Moreover, (5.3.13) is trivial unless $t\geq w_{n,m}\log(2(m+1))$ (as the right-hand side exceeds 1). Therefore from now on we consider only t satisfying this lower bound. In particular, setting $p=1/5$, we have $t\geq\frac{2}{p}(2/\alpha)^{\frac{1}{\alpha}}c$ and $t\geq 4^{\frac{1}{\alpha}}\frac{2c}{p}(\log n)^{\frac{1}{\alpha}}$. Using the union bound and assumption 3, we get (denoting for brevity $\mathbb{E}_i(\cdot)=\mathbb{E}(\cdot|\mathcal{F}_i)$)

$$\begin{aligned}\mathbb{P}\left(\sup_{1\leq k\leq n}\left|\sum_{i=1}^k X_i\right|>t\right)&\leq\mathbb{P}\left(\sup_{1\leq k\leq n}\left|\sum_{i=1}^k Z_i\right|>t(1-p)\right)+\mathbb{P}\left(\sup_{1\leq i\leq n}|\mathbb{E}_iX_{i+1}-\mathbb{E}_0X_1|>tp\right)\\&\leq\mathbb{P}\left(\sup_{1\leq k\leq n}\left|\sum_{i=1}^k Z_i\right|>t(1-p)\right)+2\mathbb{P}\left(\sup_{1\leq i\leq n}|\mathbb{E}_{i-1}X_i|>\frac{tp}{2}\right).\end{aligned} \tag{5.3.14}$$

By another application of the union bound together with the stationarity of $(\mathbb{E}_{i-1}X_i)_i$ (cf. assumption 3) and Lemma D.0.5, we obtain

$$2\mathbb{P}\left(\sup_{1\leq i\leq n}|\mathbb{E}_{i-1}X_i|>\frac{tp}{2}\right)\leq 2n\mathbb{P}\left(|\mathbb{E}_0X_1|>\frac{tp}{2}\right)\leq 12n\exp\left(-\frac{p^\alpha t^\alpha}{2(2c)^\alpha}\right).$$

Notice that

$$12n\exp\left(-\frac{p^\alpha t^\alpha}{2(2c)^\alpha}\right)=12\left[n\exp\left(-\frac{p^\alpha t^\alpha}{4(2c)^\alpha}\right)\right]\exp\left(-\frac{p^\alpha t^\alpha}{4(2c)^\alpha}\right)\leq 12\exp\left(-\frac{p^\alpha t^\alpha}{4(2c)^\alpha}\right),$$

where the inequality is a consequence of the estimate $t\geq 4^{\frac{1}{\alpha}}\frac{2c}{p}(\log n)^{\frac{1}{\alpha}}$. It follows that

$$2\mathbb{P}\left(\sup_{1\leq i\leq n}|\mathbb{E}_{i-1}X_i|>\frac{pt}{2}\right)\leq 12\exp\left(-\frac{p^\alpha t^\alpha}{4(2c)^\alpha}\right)=12\exp\left(-\frac{t^\alpha}{4(10c)^\alpha}\right). \tag{5.3.15}$$

In order to deal with $\mathbb{P}\left(|\sum_{i=1}^n Z_i|>t(1-p)\right)$, we split this sum into $m+1$ parts and use the union bound:

$$\mathbb{P}\left(\sup_{1\leq k\leq n}\left|\sum_{i=1}^k Z_i\right|>t(1-p)\right)\leq\sum_{j=0}^m\mathbb{P}\left(\sup_{1\leq k\leq n}\left|\sum_{1\leq i\leq k,m+1|i-j} Z_i\right|>\frac{t(1-p)}{m+1}\right).$$

Now, to each summand on the right-hand side of the above inequality we will apply the estimate for the independent case obtained at the beginning of this proof. Setting $M = (24\alpha^{-3}\log n)^{\frac{1}{\alpha}} c$ and taking into account (5.3.2) from Lemma 5.3.1, we obtain

$$\begin{aligned}
\frac{1}{m+1}\mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k} Z_i\right| > t(1-p)\right) &\le \frac{1}{m+1}\sum_{j=0}^{m}\mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{1\le i\le k, m+1|i-j} Z_i\right| > \frac{t(1-p)}{m+1}\right)\\
&\le \exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(8(m+1)c)^\alpha}\right) + 2\exp\left(-\frac{(1-p)^2t^2}{\frac{72}{25}(m+1)\left[(n+m+1)\,\sigma_\infty^2 + \frac{8}{5}(1-p)tM\right]}\right)\\
&\le \exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(8(m+1)c)^\alpha}\right) + 2\exp\left(-\frac{t^2}{(m+1)\left[5\,(n+m+1)\,\sigma_\infty^2 + 2tM\right]}\right). \qquad (5.3.16)
\end{aligned}$$

Finally using (5.3.14), (5.3.15) and (5.3.16) we get

$$\begin{aligned}
\mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k} X_i\right| > t\right) \le& 12\exp\left(-\frac{t^\alpha}{4(10c)^\alpha}\right) + (m+1)\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(8(m+1)c)^\alpha}\right)\\
&+ 2(m+1)\exp\left(-\frac{t^2}{5(m+1)\,(n+m+1)\,\sigma_\infty^2 + 2(m+1)tM}\right).
\end{aligned}$$

To conclude (5.3.13) it is now enough to note that the second summand on the right-hand side above dominates the first one.

To finish the proof of the lemma it remains to show the upperbound on Orlicz norm of Z_i, i.e. (5.3.12). Using the triangle inequality (cf. Lemma D.0.1) twice and then Lemma D.0.3, we obtain

$$\begin{aligned}
\|Z_i\|_{\psi_\alpha} &\le 2^{\frac{1}{\alpha}-1}\|X_i\|_{\psi_\alpha} + 2^{\frac{1}{\alpha}-1}\|\mathbb{E}_i X_{i+1} - \mathbb{E}_0 X_1\|_{\psi_\alpha} \le 2^{\frac{1}{\alpha}}\|X_i\|_{\psi_\alpha} + 2^{\frac{2}{\alpha}-1}\|\mathbb{E}_0 X_1\|_{\psi_\alpha}\\
&\le 2^{\frac{1}{\alpha}}\|X_i\|_{\psi_\alpha} + 2^{\frac{2}{\alpha}-1}(2/\alpha)^{\frac{1}{\alpha}}\|X_1\|_{\psi_\alpha} \le \|X_1\|_{\psi_\alpha}\left(2^{\frac{1}{\alpha}} + \frac{1}{2}(8/\alpha)^{\frac{1}{\alpha}}\right) \le c(8/\alpha)^{\frac{1}{\alpha}}. \qquad (5.3.17)
\end{aligned}$$

This concludes the proof of the lemma. ■

Now, a similar argument to that given in the bounded case (recall Example 5.3.4) combined with Lemma 5.3.12 (applied with $m = 2$) immediately yields the following corollary (we omit numerical calculations).

Corollary 5.3.14. *Let $(X_i)_{i\ge 0}$ be a 1-dependent stationary Markov chain and f be such that $\mathbb{E}f(X_i) = 0$ and $\|f(X_i)\|_{\psi_\alpha} \le c$ for some $\alpha \in (0,1]$ with $c > 0$. Then for any $t > 0$ and $n \in \mathbb{N}$,*

$$\mathbb{P}\left(\sup_{1\le k\le n}\left|\sum_{i=1}^{k} f(X_i)\right| > t\right) \le 6\exp\left(\frac{-\alpha t^\alpha}{16\cdot(24c)^\alpha} + 8\right) + 6\exp\left(\frac{-t^2}{15(n+3)\sigma_\infty^2 + 6(24\alpha^{-3}\log n)^{\frac{1}{\alpha}} ct}\right).$$

Now we turn to the case of random length sums. In the proof of Lemma 5.3.16 below we will need the following fact.

Lemma 5.3.15. *Fix independent random variables $(\gamma_i)_{0\le i\le l-1}$ such that $\mathbb{E}\gamma_i = 0$, $\sigma^2 = \mathbb{E}\gamma_i^2$ and $\|\gamma_i\|_{\psi_\alpha} \le v$ for some $v > 0$. Let $B := v(3\alpha^{-2}\log(l))^{1/\alpha}$. Moreover, assume that T is a bounded stopping time (with respect to some filtration $\mathcal{G}_i \supset \sigma(\gamma_1, \gamma_1, \ldots, \gamma_{i-1})$ such that γ_i is independent of $\mathcal{G}_i$). Then for any $a > 0$ and $t \ge 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^{T}\gamma_{i-1}\right| > t\right) \le e^8\exp\left(-\frac{t^\alpha}{2\left(2+\sqrt{2}\right)^\alpha c^\alpha}\right) + 2^{\frac{3}{2}}\exp\left(-\frac{t^2}{8a\sigma^2 + 2\sqrt{2}\mu t}\right),$$

where

$$\mu = \max\left(\frac{8B}{3}, 2\sigma\sqrt{\|(T-a)_+\|_{\psi_1}}\right).$$

Proof. It is just a reformulation of Proposition 4.4. ii) from [5] with $\epsilon := 1$, $p := \frac{\sqrt{2}}{\sqrt{2}-1}$ and $q := \sqrt{2}$. ■

Lemma 5.3.16 (Bernstein inequality for random sums). *Let $(X_i)_{i\geq 0}$ be a 1-dependent sequence of centered random variables such that $\|X_i\|_{\psi_\alpha} \leq c$ for some $\alpha \in (0,1]$ and $c \geq 1$. Moreover, let $N \leq n \in \mathbb{N}$ be an $\mathbb{N}$-valued bounded random variable. Assume that we can find a filtration $\mathcal{F} = (\mathcal{F}_i)_{i\geq 0}$ such that for $Z_i = X_i + \mathbb{E}(X_{i+1}|\mathcal{F}_i) - \mathbb{E}(X_i|\mathcal{F}_{i-1})$ we have the following:*

0. *X_i is $\mathcal{F}_i$ measurable,*
1. *N is a stopping time with respect to $\mathcal{F}$,*
2. *$(Z_i)_{i\geq 1}$ is stationary,*
3. *For each $j \in \mathbb{N}$ process $(Z_i)_{i\geq j+3}$ is independent of $\mathcal{F}_j$,*
4. *$(\mathbb{E}(X_i|\mathcal{F}_{i-1}))_{i\geq 1}$ is stationary,*
5. *$\mathcal{F}_{i-1}$ is independent of X_{i+1} for all $i \geq 1$.*

Then for any $t > 0$ and $a > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} X_i\right| > t\right) \leq 4\exp(8)\exp\left(-\frac{t^\alpha}{uc^\alpha}\right) + 9\exp\left(-\frac{t^2}{v\sigma_\infty^2 + wt}\right), \tag{5.3.18}$$

where $u = \frac{16\cdot 26^\alpha}{\alpha}$, $v = 102a$, $w = 14M\max\left(2, \sqrt{\|\left(\lceil N/3\rceil - a + 1\right)_+\|_{\psi_1}}\right)$ and $M = c(24\alpha^{-3}\log n)^{\frac{1}{\alpha}}$.

Proof. Observe that 0. and 3. imply the 2-dependence of process $(Z_i)_{i\geq 1}$. Therefore, filtration $\mathcal{F}$ satisfies all the assumptions of Lemma 5.3.12 and thus (5.3.2) holds. Note also that without loss of generality we may assume that $t \geq w\log 9$ (otherwise the right-hand side of (5.3.18) is at least one). Fix $s = (8\sqrt{2}\log 9)^{-1}$. Using the union bound and setting $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot|\mathcal{F}_i)$, we get

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} X_i\right| > t\right) \leq \mathbb{P}\left(\left|\sum_{i=1}^{N} Z_i\right| > t(1-s)\right) + 2\mathbb{P}\left(\sup_{1\leq i\leq n}|\mathbb{E}_{i-1}X_i| > \frac{ts}{2}\right). \tag{5.3.19}$$

Now, using Lemma D.0.5 and the inequalities $ts/2 \geq c\left(\frac{2}{\alpha}\right)^{\frac{1}{\alpha}}$, $t \geq w\log 9$ and $n\exp\left(-\frac{(st)^\alpha}{4(2c)^\alpha}\right) \leq 1$, we obtain

$$2\mathbb{P}\left(\sup_{1\leq i\leq n}|\mathbb{E}_{i-1}X_i| > \frac{st}{2}\right) \leq 2n\mathbb{P}\left(|\mathbb{E}_0 X_1| > \frac{st}{2}\right) \leq 12\exp\left(-\frac{(st)^\alpha}{4(2c)^\alpha}\right). \tag{5.3.20}$$

Next, we take care of the other term on the right-hand side of (5.3.19). Firstly we split the sum:

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} Z_i\right| > t(1-s)\right) \leq \sum_{j=0}^{2}\mathbb{P}\left(\left|\sum_{1\leq i\leq N,\ 3|(i+j)} Z_i\right| > \frac{t(1-s)}{3}\right). \tag{5.3.21}$$

Now, we will consider the jth summand of the above sum. Let us take $r = \frac{3}{8\sqrt{2}\log(9)}$ and notice that there exists function $f_j : \mathbb{N} \to \mathbb{N}$ such that for any $n \in \mathbb{N}$, $\left\lfloor \frac{n}{3}\right\rfloor \leq f_j(n) \leq \left\lceil \frac{n}{3}\right\rceil$ and

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{1\leq i\leq N,\ 3|i+j} Z_i\right| > t(1-s)/3\right) &= \mathbb{P}\left(\left|\sum_{1\leq i\leq f_j(N)} Z_{3i-j}\right| > t(1-s)/3\right)\\ &\leq \mathbb{P}\left(\left|\sum_{1\leq i\leq\lceil N/3\rceil+1} Z_{3i-j}\right| > t(1-r)(1-s)/3\right) + \mathbb{P}\left(2\sup_{k\leq n+6}|Z_k| > tr(1-s)/3\right).\end{aligned} \tag{5.3.22}$$

Due to $\|Z_i\|_{\psi_\alpha} \le c(8/\alpha)^{\frac{1}{\alpha}}$ (cf. (5.3.12)) and Lemma D.0.4 along with $t \ge w\log(9)$, $n \ge 2$ (for $n = 1$ there is nothing to prove), we get

$$\begin{aligned}\mathbb{P}\left(2\sup_{k\le n+6}|Z_k| > \frac{tr(1-s)}{3}\right) &\le (n+6)\mathbb{P}\left(|Z_k| > \frac{tr(1-s)}{3}\right)\\ &\le 2(n+6)\exp\left(-\frac{\alpha(tr(1-s))^\alpha}{8(3c)^\alpha}\right) \le 2\exp\left(-\frac{\alpha(tr(1-s))^\alpha}{16(3c)^\alpha}\right).\end{aligned} \tag{5.3.23}$$

To handle the first summand on the right-hand side of (5.3.22), let us fix j and put $\gamma_i = Z_{3i+3-j}$, $\mathcal{G}_i = \mathcal{F}_{3i-j}$, $T = \lceil N/3+1\rceil \le \lceil n/3\rceil + 1$. Using the assumptions on the filtration $\mathcal{F}$ and (5.3.2) it is straightforward to check that the following properties hold:

1. γ_i are independent,
2. $\mathbb{E}\gamma_i = 0$, $\mathbb{E}\gamma_i^2 = \sigma_\infty^2$, $\|\gamma_i\|_{\psi_\alpha} \le c(8/\alpha)^{\frac{1}{\alpha}}$,
3. γ_{i-1} is $\mathcal{G}_i$ measurable,
4. γ_i is independent of $\mathcal{G}_i$,
5. T is a stopping time with respect to the filtration $\mathcal{G}_i$.

This is precisely the setting of Lemma 5.3.15 which implies that for any $a > 0$,

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{1\le i\le\lceil N/3\rceil+1} Z_{3i-j}\right| > t(1-r)(1-s)/3\right)\\ &\le \exp(8)\exp\left(-\frac{(t(1-r)(1-s))^\alpha}{2(3(2+\sqrt{2})\hat{c})^\alpha}\right) + 3\exp\left(-\frac{(t(1-r)(1-s))^2}{72a\sigma_\infty^2 + 6\sqrt{2}\mu(1-r)(1-s)t}\right),\end{aligned} \tag{5.3.24}$$

where

$$\mu = \max\left(\frac{8M}{3}, 2\sigma_\infty\sqrt{\|\left(\lceil N/3\rceil - a + 1\right)_+\|_{\psi_1}}\right), \qquad \hat{c} = c\left(\frac{8}{\alpha}\right)^{\frac{1}{\alpha}}.$$

Using (5.3.2), Lemma D.0.2 with $Y = \frac{\alpha Z^\alpha}{8c^\alpha}$ and $\beta = \frac{2}{\alpha}$, together with the gamma function estimate $\Gamma(x) \le \left(\frac{x}{2}\right)^{x-1}$ for $x \ge 2$ (see Theorem 1 in [69]), we get

$$\sigma_\infty^2 = \mathbb{E}Z_1^2 \le 2c^2\left(\frac{8}{\alpha}\right)^{\frac{2}{\alpha}}\Gamma\left(\frac{2}{\alpha}+1\right) \le \frac{4}{\alpha}c^2\left(\frac{8}{\alpha}\right)^{\frac{2}{\alpha}}\Gamma\left(\frac{2}{\alpha}\right) \le 4c^2\left(\frac{8}{\alpha^2}\right)^{\frac{2}{\alpha}},$$

which implies that $\sigma_\infty \le \frac{2}{3}M$ and, as a consequence,

$$\mu \le \frac{4}{3}Mb, \text{ where } b = \max\left(2, \sqrt{\|\left(\lceil N/3\rceil - a + 1\right)_+\|_{\psi_1}}\right).$$

Therefore, (5.3.24) reduces to

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{1\le i\le\lceil N/3\rceil+1} Z_{3i-j}\right| > t(1-r)(1-s)/3\right)\\ &\le \exp(8)\exp\left(-\frac{(t(1-r)(1-s))^\alpha}{2(3(2+\sqrt{2})\hat{c})^\alpha}\right) + 3\exp\left(-\frac{(t(1-r)(1-s))^2}{72a\sigma_\infty^2 + 8\sqrt{2}Mb(1-r)(1-s)t}\right).\end{aligned}$$

Combining the above inequality with (5.3.19)–(5.3.23), we obtain

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^N X_i\right| > t\right) &\le 12\exp\left(-\frac{(st)^\alpha}{4(2c)^\alpha}\right) + 6\exp\left(-\frac{\alpha(tr(1-s))^\alpha}{16(3c)^\alpha}\right)\\ &+ 3\exp(8)\exp\left(-\frac{(t(1-r)(1-s))^\alpha}{2(3(2+\sqrt{2})\hat{c})^\alpha}\right) + 9\exp\left(-\frac{(t(1-r)(1-s))^2}{72a\sigma_\infty^2 + 8\sqrt{2}Mb(1-r)(1-s)t}\right).\end{aligned}$$

To conclude, it is now enough to recall that $r = 3(8\sqrt{2}\log(9))^{-1}$, $s = (8\sqrt{2}\log 9)^{-1}$ and do some elementary calculations. ■

Once more we easily get the following corollary for one-factors of one-dependent Markov chains.

Corollary 5.3.17. *Let $(X_i)_{i\geq 0}$ be a 1-dependent stationary Markov chain and f be such that $\mathbb{E}f(X_i) = 0$ and $\|f(X_i)\|_{\psi_\alpha} \leq c$ for some $\alpha \in (0,1]$ and $c \geq 1$. Moreover, let $N \leq n \in \mathbb{N}$ be a bounded stopping time with respect to the natural filtration $\mathcal{F}_i = \sigma(X_{[0,i]})$. Then for any $t > 0$ and $a > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} X_i\right| > t\right) \leq 4\exp(8)\exp\left(-\frac{t^\alpha}{uc^\alpha}\right) + 9\exp\left(-\frac{t^2}{v\sigma_\infty^2 + wt}\right), \tag{5.3.25}$$

where $u = \frac{16\cdot 26^\alpha}{\alpha}$, $v = 102a$, $w = 14M\max\left(2, \sqrt{\|\left(\lceil N/3\rceil - a + 1\right)_+\|_{\psi_1}}\right)$ and $M = c(24\alpha^{-3}\log n)^{\frac{1}{\alpha}}$.

5.4 Open questions

Assume that $\mathbf{X}$ is a stationary, 1-dependent sequence of random variables $\mathcal{X}_i \in \mathcal{X}$.

Is it true that every such $\mathbf{X}$ is, in fact, a k-block factor of a stationary Markov chain $\mathbf{M}$, where M_i's belong to some **countable state space** $\mathcal{M}$? Clearly, in this problem we must assume that $\mathcal{X}$ is at most countable. Furthermore, if we drop the assumption about the countability of $\mathcal{M}$ then the answer to this question is trivial since we can always take $M_i = X_{(-\infty,i]}$.

Can $\mathbf{X}$ always be expressed as a k-factor of a **one-dependent** stationary Markov chain $\mathbf{M}$ (here, unlike in the previous question, we allow an arbitrary state space $\mathcal{M}$)? Note that in view of Corollary 5.3.7, the positive answer to this question would immediately imply general, optimal up to numerical constants, version of the Bernstein inequality for **all** one-dependent stationary processes $\mathbf{X}$.

Chapter 6

Markov Chains on general spaces

In this section we show how the combination of the classical splitting method for Markov chains and our versions of Bernstein inequalities for one-dependent sequences allows us to obtain an "optimal" Bernstein-type inequalities for general Markov chains.

6.1 Background

6.1.1 Definitions

In this part $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ always stands for a time homogeneous Markov chain (not necessarily stationary) defined on $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in a countably generated measurable space $(\mathcal{X}, \mathcal{B})$ and equipped with a transition probability $P(x, A) : \mathcal{X} \times \mathcal{B} \to [0, 1]$. For any initial distribution μ on $\mathcal{X}$, we write $\mathbb{P}_\mu (\mathbf{X} \in \cdot)$ for the distribution of the chain with X_0 distributed according to the measure μ. More precisely, for all $n \in \mathbb{N}$ and arbitrary measurable sets A_i,

$$\mathbb{P}_\mu \left(X_{[0,n]} \in \bigtimes_{i=0}^{n} A_i \right) = \int_{\mathbf{X}^{n+1}} \mathbb{1}_{x_{[0,n-1]} \in \times_{i=0}^{n-1} A_i} P(x_{n-1}, A_n) \ldots P(x_1, dx_2) P(x_0, dx_1) \mu(dx_0). \tag{6.1.1}$$

For simplicity's sake we use $\mathbb{P}_x = \mathbb{P}_{\delta_x}$ where δ_x is the Dirac measure at x. In particular, under $\mathbb{P}_x$, $X_0 = x$ almost surely. Denoting by $\mathbb{E}_\mu$ the expectation taken with respect to $\mathbb{P}_\mu$ we easily extend (6.1.1) to

$$\mathbb{E}_\mu \left(f(X_{[n,\infty)}) | \mathcal{F}^n \right) = \mathbb{E}_{X_n} f(X_{[n,\infty)}) = \left[\mathbb{E}_{x_n} f(x_n, X_{[n+1,\infty)}) \right]_{x_n = X_n},$$

where f is any integrable (product measurable) function and for any $i < j$, $\mathcal{F}_i^j = \sigma(X_i, \ldots, X_j)$ (we abbreviate $\mathcal{F}^n = \mathcal{F}_0^n$). Sometimes, when the distribution of the integrand does not depend on the choice of starting distribution μ, we express this fact by omitting the index, that is writing $\mathbb{P}$ instead of $\mathbb{P}_\mu$ (the same convention is used for $\mathbb{E}$). Since we consider a discrete time Markov processes it is well-known (Proposition 3.4.6 in [77]) that the ***strong Markov property holds***, that is for all initial distributions μ, all real integrable function f and all stopping times τ (with respect to the filtration $\mathcal{F}^n$),

$$\mathbb{E}_\mu \left(f(X_{[\tau,\infty)}) | \mathcal{F}_0^\tau \right) = \mathbb{E}_{X_\tau} f(X_{[\tau,\infty)}) \tag{6.1.2}$$

$\mathbb{P}_\mu$ almost surely on the set $\{\tau < \infty\}$ (by definition, τ is a stopping time with respect to the filtration $(\mathcal{F}^k)_k$ if for all $k \in \mathbb{N}$, $\mathbb{P}(\tau \le k) \in \mathcal{F}^k$).

Recall that the n-step transition kernel is defined recursively by

$$P^n(x, A) = \begin{cases} \delta_x(A), & \text{if } n = 0, \\ \displaystyle\int P^{n-1}(y, A) P(x, dy), & \text{if } n \ge 1. \end{cases}$$

In other words, $P^n(x, A) = \mathbb{P}_x (X_n \in A)$, that is, if $\mathbf{X}$ starts from x then the probability that $X_n \in A$ is equal to $P^n(x, A)$. Due to the Markov property, it is intuitive (see Theorem 3.4.2 in [77]) that for any $0 \le m \le n$ we have ***Chapman–Kolmogorov formula***

$$P^n(x, A) = \int P^{n-m}(y, A) P^m(x, dy). \tag{6.1.3}$$

Heuristically, in order to transport $\mathbf{X}$ from x to A using n steps, alternatively, we can start from x, at time m visit some $y \in \mathcal{X}$ (at this point $\mathbf{X}$ forgets about the past due to the Markov property) and move in the $n-m$ remaining steps to the set A.

6.1.2 Irreducibility

For any set $A \subset \mathcal{X}$ let

$$\tau_A = \inf\{n \geq 1 \mid X_n \in A\}$$

be the ***first return time to the set*** A. Given a (not necessarily probability) measure φ on $\mathcal{X}$, we say that $\mathbf{X}$ is φ-irreducible if for every x

$$\varphi(A) > 0 \quad \Rightarrow \quad \mathbb{P}_x(\tau_A < \infty) > 0.$$

If such a φ exists then one can show (Theorem 4.0.1 in [77]) that there is essentially one ***"maximal" irreducibility measure*** ψ in the following sense.

- For every $x \in \mathcal{X}$, $\psi(A) > 0$ iff $\mathbb{P}_x(\tau_A < \infty) > 0$.
- If $\psi(A) = 0$ then $\psi\big(\{y \mid \mathbb{P}_y(\tau_A < \infty) > 0\big) = 0$.
- If A is such that $\psi(\mathcal{X} \backslash A) = 0$ then $A = A_\psi \cup Z$ where $\psi(Z) = 0$ and for all $x \in A_\psi$, $P(x, A_\psi) = 1$.
- Moreover, if ψ is a maximal irreducibility measure then all irreducibility measures are absolutely continuous with respect to ψ (see Proposition 4.2.2 in [77]). In particular, any two maximal irreducibility measures are equivalent (they have the same measure zero sets). Thus, if such ψ exists, it makes sense to define
$$\mathcal{B}_+ = \{A \mid \psi(A) > 0\} \tag{6.1.4}$$
the ***family of sets of positive*** ψ***-measure***.

In the statements like "$\mathbf{X}$ is ψ-irreducible" we always tacitly assume that ψ is a maximal irreducibility measure.

6.1.3 Recurrence, transience and Harris recurrence

Define

$$\eta_A = \sum_{i=1}^{\infty} \mathbb{1}_{X_n \in A}. \tag{6.1.5}$$

We say that A is ***uniformly transient*** if there exists $M < \infty$ such that $\sup_{x \in A} \mathbb{E}_x \eta_A \leq M$. A set A is called ***recurrent*** if $\mathbb{E}_x \eta_A = \infty$ for all $x \in A$. If $\mathbf{X}$ is ψ-irreducible then one can show (Theorem 8.0.1 in [77]) that either every set $A \in \mathcal{B}_+$ (recall (6.1.4)) is recurrent (in this case we call $\mathbf{X}$ ***recurrent***) or there is a countable cover of $\mathcal{X}$ with uniformly transient sets (in this case we say that $\mathbf{X}$ is ***transient***). A set A is called ***Harris recurrent*** if

$$\mathbb{P}_x(\eta_A = \infty) = 1, \qquad \forall_{x \in A}. \tag{6.1.6}$$

A ψ-irreducible chain $\mathbf{X}$ is ***Harris recurrent*** if every $A \in \mathcal{B}_+$ is Harris recurrent. One can verify that A is Harris recurrent just by checking if $\mathbb{P}_x(\tau_A < \infty) = 1$ for every $x \in A$ (this is a content of Theorem 9.1.1 in [77]). In some sense this is clear due to the strong Markov property of $\mathbf{X}$ (cf. (6.1.2)). Unlike in the case of a countable $\mathcal{X}$ it is not true in general that every recurrent chain is Harris recurent (see the example from the the second paragraph of Section 9.1.2 in [77]) Nonetheless, one can show (Theorem 9.1.5 in [77]) that if $\mathbf{X}$ is recurrent then $\mathcal{X} = H \sqcup T$ where H is a ***maximal Harris set*** (the precise definition of such a set is given above Theorem 9.1.5 in [77]) and T is transient.

6.1.4 Minorization condition

We say that a Markov chain $\mathbf{X}$ satisfies ***a minorization condition*** if there exists a set $C \in \mathcal{B}$ (called a ***small set***), a probability measure ν on $\mathcal{X}$ (a ***small measure***), a constant $\delta > 0$ and a positive integer $m \in \mathbb{N}$ such that

$$P^m(x, B) \geq \delta\nu(B) \qquad \forall_{x\in C}\,\forall_{B\in\mathcal{B}}. \tag{6.1.7}$$

Although at first glance the minorization condition (6.1.7) might look a little technical it plays a central role in the analysis of general Markov chains. For example, it allows one to formulate the notion of period of the chain $\mathbf{X}$. Moreover, it serves as a basic tool when one introduces the split chain of $\mathbf{X}$, which will be essential for us to obtain concentration inequalities for additive functionals of $\mathbf{X}$.

Furthermore, one can show that for ψ-irreducible Markov chains a minorization condition is always satisfied. More precisely, due to Theorem 5.2.2 in [77], if $\mathbf{X}$ is ψ-irreducible and $A \in \mathcal{B}_+$ then there exists $m \geq 1$, small measure ν and a set $C \subset A$ such that $C \in \mathcal{B}_+$, $\nu(C) > 0$ and (6.1.7) holds.

At the end let us note that (6.1.7) is a generalization of the notion of **atom**. We say that A is an ***atom*** if there is a probability measure ν such that

$$P(x, \cdot) = \nu, \qquad \forall_{x\in A}. \tag{6.1.8}$$

If additionally $A \in \mathcal{B}_+$ (recall (6.1.4)) then A is is called an ***accessible atom***. Unlike in the case of minorization condition (6.1.7), it turns out that it is not true that accessible atoms always exist (even for ψ-irreducible Markov chains). However, (6.1.7) can be used to construct a ***pseudo-atom*** (see upcoming Section 6.1.11). The existence of an atom simplifies many proofs and ideas. For example, a consideration of consecutive return times of $\mathbf{X}$ to A leads to **the regeneration technique** (in order to get some intuition on this subject, see upcoming Section 6.1.10). Roughly, these return times split $\mathbf{X}$ into independent (random length) blocks for which many well-known techniques from the theory of independent random variables can be applied.

6.1.5 Periodicity

We follow Section 5.4.3 from [77]. Suppose that $\mathbf{X}$ is ψ-irreducible and satisfies the minorization condition (6.1.7) where $\psi(C) > 0$. Consider set

$$\{n \in \mathbb{N},\ n \geq 1 \mid \exists_{\delta_n > 0} \quad (6.1.7) \text{ is satisfied with } m = n,\ \nu = \delta_n\nu,\ C = C\}. \tag{6.1.9}$$

By (6.1.3) this set is closed under addition and thus it contains natural "period" given by the GCD p. Moreover, for a sufficiently large $k \in \mathbb{N}$ all pk belong to this set. Furthermore, it turns out that the definition of p does not depend on the choice of a small set C. By Theorem 5.4.4. in [77], one can find sets $D_0, \ldots, D_{p-1}$ such that for every $0 \leq i \leq p-1$ and $x \in D_i$

$$P(x, D_{i+1 \mod p}) = 1, \qquad \psi\left(\mathcal{X}\setminus \bigcup_{0\leq i\leq p-1} D_i\right) = 0. \tag{6.1.10}$$

The cycle D_i is "maximal" in the sense that if $E_1, \ldots, E_q$ satisfy the analog of (6.1.10) then necessarily $q|p$. If $q = p$ then after relabeling one can assume that $D_i \overset{\psi}{=} E_i$.

We call p the ***period of*** $\mathbf{X}$. If $p = 1$ then we say that $\mathbf{X}$ is ***aperiodic***. Note that if the minorization condition (6.1.7) is satisfied with $m = 1$ then by the very definition 1 belongs to the set (6.1.9) and the chain is aperiodic. In such a case we call $\mathbf{X}$ ***strongly aperiodic***.

6.1.6 Invariant measures

Recall that π (not necessarily finite) is invariant for $\mathbf{X}$ if $\pi P = \pi$, that is

$$\pi(A) = \int P(y, A)d\pi(y), \qquad \forall_{A\subset\mathcal{X}}.$$

If $\mathbf{X}$ is recurrent then one can show (Theorem 10.0.1 in [77]) that there is a unique invariant measure π for $\mathbf{X}$. Moreover, π is a maximal irreducibility measure and for any A such that $\pi(A) > 0$,

$$\pi(B) = \int_A \mathbb{E}_x \sum_{i=1}^{\tau_A} \mathbb{1}_{X_n \in B} d\pi(x).$$

Furthermore, if we can find a petit set C (for the definition of a petit set see the beginning of Section 5.5.2 in [77], let us only note that every small set is petit) such that $\sup_{x\in C} \mathbb{E}_x \tau_C < \infty$ then π is finite (in general π is only σ-finite).

A recurrent Markov chain $\mathbf{X}$ is called ***positive***, if $\pi(\mathcal{X}) < \infty$ and ***null*** otherwise. If $\mathbf{X}$ is Harris recurrent and positive then we say that $\mathbf{X}$ is ***positive Harris***.

If $\mathbf{X}$ is positive then it must be recurrent (Theorem 10.1.1 in [77]). For chains admitting atoms we have the following positivity criterion (Theorem 10.2.2 in [77]). If $\mathbf{X}$ is ψ-irreducible and admits an accessible atom A then $\mathbf{X}$ is positive iff $\mathbb{E}_A \tau_A < \infty$. Here, by $\mathbb{E}_A$, we mean $\mathbb{E}_x$ for arbitrary $x \in A$ (recall (6.1.8)). In that case (cf. the Kac's formula (A.3.2))

$$\pi(A) = \frac{1}{\mathbb{E}_A \tau_A}.$$

As an immediate corollary we get (Proposition 10.2.3 in [77]) that if $\mathcal{X}$ is countable and $\mathbf{X}$ is positive recurrent irreducible Markov chain on $\mathcal{X}$ then the unique invariant probability measure π is given by $\pi_x = 1/\mathbb{E}_x \tau_x$ for any $x \in \mathcal{X}$.

Furthermore, due to Theorem 10.4.9 in [77], if $\mathbf{X}$ is recurrent and π is its unique stationary measure then π equivalent to a maximal irreducibility measure ψ. In other words, if $\mathbf{X}$ is recurrent then π can be taken as a maximal irreducibility measure. In particular, if $\mathbf{X}$ is recurrent then an application of Theorem 5.2.1 from [77] in case $\varphi = \pi$ gives the existence of $C \subset \mathcal{X}$ such that $\pi(C) > 0$ and the conditional version π_C (π conditioned on C) satisfies a minorization condition (6.1.7) for some $m \in \mathbb{N}$ and $\delta > 0$.

6.1.7 Different kinds of "ergodicities"

This time we follow [21] (surroundings of equation (1.6)). In this part we assume tacitly that $\mathbf{X}$ is ψ-irreducible.

We say that $\mathbf{X}$ is:

- **ergodic** if it is positive Harris recurrent and aperiodic or equivalently, if $\mathbf{X}$ is positive and for all starting points $x \in \mathcal{X}$, $\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \to 0$.
- **ergodic of order** 2 if for any $A \in \mathcal{B}_+$ (recall (6.1.4)) $\mathbb{E}_\pi \tau_A < \infty$ or equivalently if $\mathbf{X}$ is ergodic and $\sum_{i=1}^n \int \left\|P^i(x,\cdot) - \pi(\cdot)\right\|_{TV} d\pi(x) < \infty$ (for this equivalence, see Theorem 4.1 in [21]).
- **geometrically ergodic** if there are $r > 1$ such that $\sum_{i=1}^n r^i \int \left\|P^i(x,\cdot) - \pi(\cdot)\right\|_{TV} d\pi(x) < \infty$ or equivalently that there is $r < 1$ and a function $G : \mathcal{X} \to \mathbb{R}$ such that $G \in L_1(\pi)$ such that $\left\|P^i(x,\cdot) - \pi(\cdot)\right\|_{TV} \leq G(x) r^n$ for all $x \in \mathcal{X}$.
- **uniformly (geometrically) ergodic** if $\sup_{x\in\mathcal{X}} \|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \to 0$. (In fact, since the sequence $a_n = \|P^n(x,\cdot) - \pi(\cdot)\|_{TV}$ is sub-multiplicative and $a_n \to 0$, a_n goes to zero geometrically fast, that is, there are $0 \leq c < 1$ and $K \in \mathbb{R}_+$ such that $a_n \leq K c^n$ for all $n \in \mathbb{N}$.)

Thus,

$$\text{unif. erg.} \quad \Rightarrow \quad \text{geom. erg.} \quad \Rightarrow \quad \text{erg. of order 2} \quad \Rightarrow \quad \text{erg.}$$

Remark 6.1.1 (Periodic case)**.** Suppose that $\mathbf{X}$ is positive Harris recurrent with period p. Then (see Theorem 13.3.4 in [77]) for **any** initial distribution λ we have

$$\lim_{n\to\infty} \left\| \frac{1}{p} \sum_{i=0}^{p-1} P^{np+i}(x,\cdot) d\lambda(x) - \pi(\cdot) \right\|_{TV} = 0. \tag{6.1.11}$$

If $\mathbf{X}$ is just positive recurrent then there is a set Z such that $\pi(Z) = 0$ and (6.1.11) holds for any initial distribution λ which satisfies $\lambda(Z) = 0$.

6.1.8 Strong law of large numbers (SLLN)

This part is based on Chapter 17 from [77]. Firstly, recall that by Birkhoff's ergodic theorem if π is a stationary distribution for a Markov chain $\mathbf{X}$ then for any function $f\colon \mathcal{X} \to \mathcal{Y}$ such that $f(\mathbf{X}) \in L_1(\mathbb{P}_\pi)$,

$$\frac{1}{n}\sum_{i=1}^{n} f(S^i\mathbf{X}) \xrightarrow[L_1(\mathbb{P}_\pi)]{\mathbb{P}_\pi\ a.s.} f_\infty(\mathbf{X}) = \mathbb{E}_{\mathbb{P}_\pi}\left(f|\mathcal{I}\right)(\mathbf{X})$$

where $\mathcal{I}$ is the invariant σ-field of $\mathcal{X}$. It turns out $f_\infty(\mathbf{X})$ can be expressed as a function of merely X_0 (under $\mathbb{P}_\pi$). In other words,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f(S^i\mathbf{X}) \xrightarrow[L_1(\mathbb{P}_\pi)]{\mathbb{P}_\pi\ a.s.} g_\infty(X_0) \tag{6.1.12}$$

for some function $g_\infty\colon \mathcal{X} \to \mathbb{R}$. This follows from the fact that f_∞ is an S-invariant function, that is $S^k f_\infty(\mathbf{X}) \overset{\mathbb{P}_\pi}{=} f_\infty(\mathbf{X})$ for all $k \in \mathbb{N}$, and the following lemma.

Lemma 6.1.2. *Suppose that $f_\infty\colon \mathcal{X}^{\mathbb{N}} \to \mathbb{R}$ is an S-invariant function in $L_1(\mathbb{P}_\pi)$. Then there exists $g_\infty\colon \mathcal{X} \to \mathbb{R}$ such that*

$$f_\infty(\mathbf{X}) \overset{\mathbb{P}_\pi}{=} g_\infty(X_k) \qquad \forall_{k\in\mathbb{N}}. \tag{6.1.13}$$

Proof. Consider a function

$$h_{f_\infty}(x) = \mathbb{E}_x f_\infty(\mathbf{X}).$$

We prove now that one may take $g_\infty = h_{f_\infty}$. To this end denote $\mathcal{F}^k = \sigma\left(X_{[0,k]}\right)$. Using the fact that $S^k f(\mathbf{X}) \overset{\mathbb{P}_\pi}{=} f(\mathbf{X})$ (note that this implies that for π-almost every $x \in \mathcal{X}$, $S^k f(\mathbf{X}) \overset{\mathbb{P}_x}{=} f(\mathbf{X})$) along with the Markov property, we get

$$h_{f_\infty}(X_k) = \mathbb{E}_{X_k} f_\infty(\mathbf{X}) \overset{\mathbb{P}_\pi}{=} \mathbb{E}_{X_k} f_\infty(S^k\mathbf{X}) \overset{\mathbb{P}_\pi}{=} \mathbb{E}\left(f_\infty(S^k\mathbf{X})|X_{[0,k]}\right) = \mathbb{E}\left(f_\infty(\mathbf{X})|\mathcal{F}^k\right).$$

Therefore, under $\mathbb{P}_\pi$, $(h_{f_\infty}(X_k), \mathcal{F}^k)$ is an integrable martingale such that

$$h_{f_\infty}(X_k) \xrightarrow{\mathbb{P}_\pi\ a.s.} f_\infty(\mathbf{X}).$$

Since, additionally, $(h_{f_\infty}(X_k))_{k\in\mathbb{N}}$ is stationary (under $\mathbb{P}_\pi$), for all k, $h_{f_\infty}(X_0) \overset{\mathbb{P}_\pi}{=} h_{f_\infty}(X_k) \overset{\mathbb{P}_\pi}{=} f_\infty(\mathbf{X})$. Indeed,

$$\begin{aligned}
&\mathbb{P}\left(|h_{f_\infty}(X_0) - h_{f_\infty}(X_n)| \geq \varepsilon\right) = \lim_{i\to\infty}\mathbb{P}\left(|h_{f_\infty}(X_i) - h_{f_\infty}(X_{i+n})| \geq \varepsilon\right)\\
&\leq \lim_{i\to\infty}\mathbb{P}\left(|h_{f_\infty}(X_i) - f(\mathbf{X})| \geq \frac{\varepsilon}{2}\right) + \lim_{i\to\infty}\mathbb{P}\left(|h_{f_\infty}(X_{i+n}) - f(\mathbf{X})| \geq \frac{\varepsilon}{2}\right) = 0.
\end{aligned}$$

■

Remark 6.1.3. Note that for π almost every x,

$$\frac{1}{n}\sum_{i=1}^{n} f(S^i\mathbf{X}) \xrightarrow{\mathbb{P}_x\ a.s.} g_\infty(x).$$

Indeed, it is enough to recall that $\mathbb{P}_\pi = \int \mathbb{P}_x d\pi(x)$ and by (6.1.12),

$$1 = \mathbb{P}_\pi\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f(S^i\mathbf{X}) = g_\infty(X_0)\right) = \int \mathbb{P}_x\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f(S^i\mathbf{X}) = g_\infty(x)\right) d\pi(x).$$

Remark 6.1.4. At the end let us give a following fact. Suppose that a stationary distribution π exists. Then one can show (Theorem 17.1.7 in [77]) that $\mathbf{X}$ is positive Harris iff for any $f \in L_1(\pi, \mathcal{X})$ and initial distribution λ,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow{\mathbb{P}_\lambda\ a.s.} \int f d\pi.$$

6.1.9 Central limit theorem (CLT)

Assume that $\mathbf{X}$ is ergodic and $f\colon \mathcal{X} \to \mathbb{R}$ is such that $\mathbb{E}_\pi f = 0$, $\mathbb{E}_\pi f^2 < \infty$ and the sum of covariances $\sum_{i=1}^{\infty} \mathrm{Cov}_\pi\left(f(X_0), f(X_i)\right)$ converges. Then by Theorem 3.1 in [21], for any initial distribution λ,

$$\frac{\sum_{i=1}^{n} f(X_i)}{\sqrt{n}} \Rightarrow N(0, \sigma^2_{Mrv}) \tag{6.1.14}$$

for some $\sigma^2_{Mrv} \geq 0$. If we strengthen our assumptions and assume additionally that $\sum_{i=1}^{\infty} f(\cdot)P^i f(\cdot)$ converges in $L_1(\pi)$ then we can identify σ^2_{Mrv} as

$$\sigma^2_{Mrv} = \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}_\pi\left(f(X_1) + \cdots + f(X_n)\right) = \mathrm{Var}_\pi\left(f(X_0)\right) + \sum_{k=1}^{\infty} \mathrm{Cov}_\pi\left(f(X_0), f(X_k)\right). \tag{6.1.15}$$

It turns out (Theorem 4.1 in [21]) that if $\mathbf{X}$ is ergodic then $\mathbf{X}$ is ergodic of order 2 iff (6.1.14) holds for every bounded and π-centered function f, where σ^2_{Mrv} is given by (6.1.15). Furthermore, if $f\colon \mathcal{X} \to \mathbb{R}$ is such that $\mathbb{E}_\pi f = 0$, $\mathbb{E}_\pi f^2 < \infty$ and $\mathbf{X}$ is uniformly ergodic then CLT (6.1.14) holds with (6.1.15) (see Theorem 4.3 in [21]).

6.1.10 Split chain by Athreya-Ney: intuition

The aim of this part is to give an intuitive description of the regeneration technique via regeneration times by Athreya and Ney (see [6]) which was invented independently by Nummelin (see [80]). To do so in a user-friendly way, we only sketch an idea omitting many details which we provide in the next section. To this end suppose that $\mathbf{X}$ is strongly aperiodic, that is (6.1.7) holds with $m = 1$. Moreover, let for all $x \in \mathcal{X}$,

$$\mathbb{P}_x\left(\tau_C < \infty\right) = 1 \tag{6.1.16}$$

(as we mentioned before, (6.1.16) implies that $\mathbf{X}$ visits C infinitely often $\mathbb{P}_x$ almost surely).

In order to split $\mathbf{X}$ into independent parts, firstly, one proves that there is a random time $\tau \geq 1$ (a **regeneration time**) such that

$$\mathbb{P}_x\left(X_{n+1} \in A, \tau = n\right) = \nu(A \cap C)\mathbb{P}_x\left(\tau = n\right) \tag{6.1.17}$$

for all n and A. Now, (6.1.17) implies that at time τ the Markov chain regenerates, that is starts anew accordingly to small measure ν forgetting about what happened in past. In particular, $X_{[0,\tau]}$ is independent of $X_{[\tau+1,\infty)}$. Now, one can repeat this procedure just by finding an analog of τ for the process $X_{[\tau+1,\infty)}$ and so on and so forth. In short, this technique allows to split the chain into random length blocks Ξ_i (for $i \geq 0$) such that $(\Xi_i)_{i\geq 1}$ is stationary and independent.

Assume that a minorization condition (6.1.7) holds. In order to construct τ as in (6.1.17) we run $\mathbf{X}$ until it hits the small set C (cf. (6.1.16)). Let's say, it happens at time k and a place x. Then with probability δ we distribute X_{k+1} according to ν and with probability $1-\delta$ to

$$Q(x,\cdot) = \frac{1}{1-\delta}\left[P(x,\cdot) - \delta\nu(\cdot)\right].$$

We repeat this procedure every time $\mathbf{X}$ enters C. Since each time we do so there is an independent positive probability δ of choosing $X_{k+1} \sim \nu$, it is should be intuitively clear that eventually at some step we will distribute X_{n+1} according to ν. The first time it happens serves as the definition of τ.

6.1.11 General splitting of the chain

In this part we will introduce the splitting method in its full strength and in far more detailed way. This part is based on [77], Section 17.3.1. However, there are some slight differences between our exposition and the one in [77]. Furthermore, we postpone some proofs concerning Markov-like properties of processes connected with the split chain to Appendix E.

Fix C, m, ν and $\delta > 0$ as in (6.1.7). The minorization condition allows us to redefine the chain $\mathbf{X}$ together with an auxiliary regeneration structure. More precisely we start with a **splitting of the space** $\mathcal{X}$ into two identical copies on level 0 and 1 namely we consider

$$\overline{\mathcal{X}} = \mathcal{X} \times \{0, 1\}.$$

Now we split $\mathbf{X}$ in the following way. We consider a process $\overline{\mathbf{X}}$ defined on $\overline{\mathcal{X}}^{\mathbb{N}}$ (usually called the ***split chain***) given by

$$\overline{\mathbf{X}} = \left(\overline{X}_i\right)_{i \geq 0} = (\mathbf{X}, \mathbf{Y}) = (X_i, Y_i)_{i \geq 0}.$$

Remark 6.1.5. For simplicity's sake, we slightly abuse our notation by denoting the first coordinate of the split chain with the same letter as we used for the initial Markov chain. However, in the end, it turns out that the first coordinate of the split chain $\overline{\mathbf{X}}$ has the same distribution as the starting Markov chain $\mathbf{X}$ which justifies this convention.

The random variables Y_i take values in $\{0, 1\}$ and should be interpreted as **indicators of levels on which $\overline{\mathbf{X}}$ is at the very moment**. Now, for any $A_1, \ldots, A_m \in \mathcal{B}$, $k \in \mathbb{N}$ and $i \in \{0, 1\}$, set

$$\begin{aligned}
&\mathbb{P}\left(Y_{km} = i, X_{[km+1,(k+1)m]} \in \bigtimes_{i=1}^{m} A_i \mid \mathcal{F}_{km}^{\mathbf{X}}, \mathcal{F}_{km-m}^{\mathbf{Y}}, X_{km} = x\right) \\
&\qquad = \mathbb{P}\left(Y_0 = i, X_{[1,m]} \in \bigtimes_{i=1}^{m} A_i \mid X_0 = x\right) \\
&\qquad = \int_{A_1} \cdots \int_{A_m} r(x, x_m, i) P(x_{m-1}, dx_m) P(x_{m-2}, dx_{m-1}) \ldots P(x, dx_1),
\end{aligned} \tag{6.1.18}$$

where

$$r(x, y, i) = \begin{cases} \mathbb{1}_{x \in C}\, r(x, x_m), & \text{if } i = 1, \\ 1 - \mathbb{1}_{x \in C}\, r(x, x_m), & \text{if } i = 0, \end{cases}, \qquad r(x, y) = \frac{\delta \nu(dy)}{P^m(x, dy)}. \tag{6.1.19}$$

Moreover, for any $k, i \in \mathbb{N}$ such that $km < i < (k+1)m$ we put

$$Y_i = Y_{km}. \tag{6.1.20}$$

Since the above definition are far from being pleasant, let us give some words of explanation. For the clarity of this presentation, here and later on, we omit the measurability details.

Firstly, the Radon derivative $r(x, y)$ in (6.1.19) is well defined due to (6.1.7). Moreover, (6.1.7) implies that $r(x, y) \leq 1$.

When it comes to the **level process Y**, which is the second coordinate of $\overline{\mathbf{X}}$, (6.1.18) defines only Y_i when i is a multiple of m. Thus, one needs to provide a definition of the remaining Y_i's and this is done in (6.1.20) just by saying that if $\overline{\mathbf{X}}$ at time km was on level Y_{km} then it remains on this level up to time $(k+1)m$.

Notice that, by (6.1.18), $Y_{km} = 1$ enforces X_{km} to fall into the small set C. Moreover, (6.1.18) implies that in that case

$$\mathcal{L}\left(X_{mk+m} \mid Y_{mk} = 1\right) = \nu.$$

Summing it up, if the split chain $\overline{\mathbf{X}}$, at time km, is on level $Y_{km} = 1$ (thus $X_{km} \in C$) then, at time $km + m$, $\overline{\mathbf{X}}$ **regenerates** and starts anew from ν. Thus, if for convenience sake we put $\tau_{-1} = -m$ and then for $i \geq 0$ we define τ_i to be the i'th time when the second coordinate (level coordinate) hits 1, namely

$$\tau_i = \min\{k > \tau_{i-1} \mid Y_k = 1,\ m|k\}, \tag{6.1.21}$$

we obtain a desired **regeneration structure** for $\mathbf{X}$ (cf. (6.1.17)). In particular, we can split $\overline{\mathbf{X}}$ into random length blocks which are "nearly independent", that is, one-dependent. More precisely, we introduce ***the random block process***

$$\boldsymbol{\Xi} = \left(\Xi_i\right)_{i \geq 0}, \quad \Xi_i = X_{[\tau_{i-1}+m, \tau_i+m-1]}, \tag{6.1.22}$$

where we consider each Ξ_i as a random variable with values in the disjoint union $\bigsqcup_{j\geq 0} \mathcal{X}^j$. It turns out that the **random block process $\mathbf{\Xi}$ is a one dependent Markov chain such that** $(\Xi_i)_{i\geq 1}$ is **stationary** (see [21], Corollary 2.4). These properties of $\mathbf{\Xi}$ will be of crucial importance when we consider concentration inequalities for additive functionals of Markov chains.

Remark 6.1.6. If $m = 1$ then one can show that the "Athrey-Ney"'s random times introduced in (6.1.17) are in fact return times of $\overline{\mathbf{X}}$ to the **atom** $C \times \{1\}$ (that is $\tau = \tau_{C\times\{1\}}$ in (6.1.17)).

Now, we present some ideas which clarify why $\mathbf{\Xi}$ is a Markov chain (see (6.1.24) below). Although, in general **the split chain $\overline{\mathbf{X}}$ is not a Markov chain** one should think about it as about a Markov-like process possessing the Markov-like property (6.1.18), which easily generalizes to

$$\mathbb{E}\left(F\left(X_{[km+1,\infty)}, Y_{[km,\infty)}\right) \mid \mathcal{F}^{\mathbf{X}}_{km}, \mathcal{F}^{\mathbf{Y}}_{km-m}\right) = \mathbb{E}\left(F\left(X_{[km+1,\infty)}, Y_{[km,\infty)}\right) \mid X_{km}\right), \tag{6.1.23}$$

where F is a product measurable bounded function. This, in turn, immediately leads to the fact that the m**-vectorized split** chain $\overline{\mathbf{X}}^{(m)} = \left(\overline{X}_{[im,im+m-1]}\right)_{i\in\mathbb{N}}$ **is a Markov chain**. Even more,

$$\mathbb{E}\left(F\left(\overline{X}^{(m)}_{[k,\infty)}\right) \mid \overline{X}^{(m)}_{[0,k-1]}\right) = \mathbb{E}\left(F\left(\overline{X}^{(m)}_{[k,\infty)}\right) \mid \overline{X}^{(m)}_{k-1}\right) = \mathbb{E}\left(F\left(\overline{X}^{(m)}_{[k,\infty)}\right) \mid X_{mk-m}, X_{mk-1}, Y_{mk-m}\right).$$

By the strong Markov property of m-vectorized split chain $\overline{\mathbf{X}}^{(m)}$ it follows that **$\mathbf{\Xi}$ is a Markov chain**. In fact, we can get even more. Namely, for any product measurable function F

$$\mathbb{E}\left(F\left(\Xi_{[i,\infty)}\right) | \Xi_{[0,i-1]}\right) = \mathbb{E}\left(F\left(\Xi_{[i,\infty)}\right) | \Xi_{i-1}\right) = \mathbb{E}\left(F\left(\Xi_{[i,\infty)}\right) | \mathrm{pr}_m\left(\Xi_{i-1}\right)\right), \tag{6.1.24}$$

where $\mathrm{pr}_m : \bigsqcup_{j\geq m} \mathcal{X}^j \to \mathcal{X}^m$ is the projection on m-last coordinates,

$$\mathrm{pr}_m\left(x_0, \ldots, x_j\right) = \left(x_{j-m+1}, \ldots, x_j\right). \tag{6.1.25}$$

Although in general (save the case $m = 1$) Ξ_i's **are not independent** one can show that **the lengths of** Ξ_i given by $|\Xi_i| = \tau_i - \tau_{i-1}$ **are independent** (for $i \geq 0$). Moreover, $(\tau_i - \tau_{i-1})_{i\geq 1}$ is stationary.

At the end of this section let us give a remark concerning initial distributions for the split chain. In order to be able to set the initial distribution for the split chain $\overline{\mathbf{X}}$ for arbitrary probability measure μ on $\mathcal{X}$, we define ***the split measure*** μ^* on $\overline{\mathcal{X}}$ via

$$\mu^*(A \times \{i\}) = \begin{cases} (1-\delta)\mu(C \cap A) + \mu(A \cap C^c), & \text{if } i = 0, \\ \delta\mu(C \cap A), & \text{if } i = 1. \end{cases} \tag{6.1.26}$$

Such definition ensures that $(X_0, Y_0) \sim \mu^*$ as soon as $X_0 \sim \mu$. For convenience sake, for any $x \in \mathcal{X}$, we write $\mathbb{P}_{x^*}$ instead of $\mathbb{P}_{\delta^*_x}$.

6.1.12 Asymptotic variances

During the upcoming proofs we will meet two types of asymptotic variances: σ^2_{Mrv} associated with the process $f(\mathbf{X})$ and σ^2_∞ associated with process $f(\mathbf{\Xi})$. The first one, defined as

$$\sigma^2_{Mrv} = \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}\left(f(X_{[0,n-1]})\right) = \mathrm{Var}_\pi(f(X_0)) + 2\sum_{i\geq 1}\mathrm{Cov}_\pi(f(X_0), f(X_i)), \tag{6.1.27}$$

is exactly the variance of the limiting normal distribution of the sequence $\frac{1}{\sqrt{n}}\sum_{i=1}^n f(X_i)$. The second one,

$$\sigma^2_\infty = \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}\left(f(\Xi_1) + \cdots + f(\Xi_n)\right) = \mathbb{E}f(\Xi_1)^2 + 2\mathbb{E}f(\Xi_1)f(\Xi_2),$$

is the variance of the limiting normal distribution of the sequence $\frac{1}{\sqrt{n}}\sum_{i=1}^n f(\Xi_i)$. Both asymptotic variances are very closely linked via the formula

$$\sigma^2_\infty = \sigma^2_{Mrv}\mathbb{E}(\tau_1 - \tau_0) = \sigma^2_{Mrv} m\delta^{-1}\pi(C)^{-1}. \tag{6.1.28}$$

For the proof of this formula we refer to [77] (see (17.32), page 434).

6.1.13 Additive functionals

Recall that the our aim is to bound the tail probabilities of **additive functionals of X**

$$f\left(X_{[0,n-1]}\right) = f(X_0) + \cdots + f(X_{n-1}). \tag{6.1.29}$$

It is convenient to extend every real function $f\colon \mathcal{X} \to \mathbb{R}$ to $f\colon \bigsqcup_{i\geq 0} \mathcal{X}^i \to \mathbb{R}$ via

$$f(x_{[0,n]}) = \sum_{i=0}^{n} f(x_i), \qquad f(\varnothing) = 0. \tag{6.1.30}$$

Suppose that the majorization condition (6.1.7) hold. The splitting technique allows us to split the sum from (6.1.29) into a random number of random length blocks. More precisely, (recall the regeneration times τ_i from (6.1.21)) let

$$N = N_{n-1} = \inf\{i \geq 0 \mid \tau_i + m - 1 \geq n - 1\} \tag{6.1.31}$$

stand for the ***number of regenerations up to time*** $n-1$. Note that if $N \geq 1$ then split (6.1.29) into the three parts

$$f\left(X_{[0,n-1]}\right) = f(\Xi_0) + \left[\sum_{i=1}^{N} f(\Xi_i)\right] - f(X_{[n,\tau_N+m-1]}). \tag{6.1.32}$$

By properties of the random blocks Ξ_i, one immediately concludes that process $(f(\Xi_i))_{i\geq 1}$ is **a one-block factor of a stationary, one-dependent Markov chain**. Moreover, for any $i \geq 1$ and starting measure μ (see Theorem 17.3.1 in [77], page 435)

$$\mathbb{E}_\mu f(\Xi_i) = \mathbb{E}_\nu f(\Xi_0) = \delta^{-1}\pi(C)^{-1} m \int f d\pi. \tag{6.1.33}$$

As a direct consequence, for any $i \geq 1$, we have $\mathbb{E}_\mu|\Xi_i| = \mathbb{E}_\mu(\tau_i - \tau_{i-1}) = \delta^{-1}\pi(C)^{-1}m$ and if $\mathbb{E}_\pi f(X_0) = 0$ then $\mathbb{E}_\mu f(\Xi_i) = 0$.

Clearly, by (6.1.32), the main difficulty in obtaining some tail inequality for (6.1.29) resides in getting such for the middle term $\sum_{i=1}^{N} f(\Xi_i)$. There are two natural ways to do it. The first one relies on the combination of the following observation

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} f(\Xi_i)\right| \geq t\right) \leq \mathbb{P}\left(\sup_{1\leq k\leq K}\left|\sum_{i=1}^{k} f(\Xi_i)\right| \geq t\right) + \mathbb{P}\left(N > K\right), \qquad \forall_{K\in\mathbb{N}} \tag{6.1.34}$$

with Lemma 5.3.12. The second one is just an application of Lemma 5.3.16. In both cases one needs to provide some exponential bounds on tails of N. This is done in the upcoming Section 6.1.14.

6.1.14 Bounds on the number of regenerations

In this part we provide bounds on the tail of number of regenerations N (recall (6.1.31)). To do so, we need a notion of the exponential Orlicz norm. Recall that for any random variable X and $\alpha > 0$ the ***exponential Orlicz's (quasi-) norm*** is defined as

$$\|X\|_{\psi_\alpha} = \inf\left\{c > 0 \mid \mathbb{E}\exp\left(\frac{|X|^\alpha}{c^\alpha}\right) \leq 2\right\}. \tag{6.1.35}$$

Let us stress that, unlike in the case of $\alpha \geq 1$, if $\alpha < 1$ then $\|\cdot\|_{\psi_\alpha}$ is just a quasi-norm. For basic properties of these quasi-norms we refer to Appendix D.

In what follows we deal with various underlying measures on the state space $\overline{\mathcal{X}}$. In order to stress the dependence of the Orlicz norm on the initial distribution μ of the chain $\mathbf{X}$ we sometimes write $\|\cdot\|_{\psi_\alpha,\mu}$ instead of $\|\cdot\|_{\psi_\alpha}$.

Firstly, we need the ψ_1 version of the Bernstein inequality, which follows easily from the classical moment version of this inequality (see, e.g., Lemma 2.2.11 in [98]), by observing that for $k \geq 2$, $\mathbb{E}|\xi|^k \leq k!\|\xi\|_{\psi_1}^k = k!M^{k-2}v/2$, where $M = \|\xi\|_{\psi_1}$, $v = 2\|\xi\|_{\psi_1}^2$.

Lemma 6.1.7 (ψ_1 Bernstein's inequality)**.** *If $(\xi_i)_i$ is a sequence of independent centered random variables such that $\sup_i \|\xi_i\|_{\psi_1} \le \tau$, then*

$$\mathbb{P}\left(\sum_{i=1}^{n} \xi_i \ge t\right) \le \exp\left(-\frac{t^2}{4n\tau^2 + 2\tau t}\right).$$

Now, we turn to the bounds on the number of regenerations. Recall that if the distribution of a random variable does not depend on the starting distribution μ then instead of $\mathbb{P}_\mu$ we write $\mathbb{P}$ omitting the subscript.

Lemma 6.1.8. *If $\|\tau_1 - \tau_0\|_{\psi_1} \le d$ then for any $p > 0$ and starting distribution μ,*

$$\mathbb{P}_\mu\left(N > \left\lceil (1+p)n\left[\mathbb{E}(\tau_1 - \tau_0)\right]^{-1}\right\rceil\right) \le \exp\left(-\frac{pn\mathbb{E}(\tau_1-\tau_0)}{K_p d^2} + 1\right), \tag{6.1.36}$$

where $K_p = L_p + 16/L_p$ and $L_p = \frac{16}{p} + 20$. Moreover, the function $p \to K_p$ is decreasing on $\mathbb{R}_+$ (in particular $K_p \ge K_\infty = \frac{104}{5}$) and if $p = 2/3$ then $\frac{1}{p}K_p \le 67$.

Proof. For convenience sake, let $T_i = \tau_i - \tau_{i-1}$ for $i \ge 1$. Firstly, notice that without loss of generality we may assume that $np \ge L_p\mathbb{E}T_1$. Indeed, otherwise, using $\mathbb{E}T_1 \le d$ we obtain

$$\exp\left(-\frac{pn\mathbb{E}T_1}{K_p d^2} + 1\right) \ge \exp\left(-\frac{L_p\mathbb{E}^2T_1}{K_p d^2} + 1\right) \ge \exp\left(1 - \frac{L_p}{K_p}\right) \ge 1.$$

Thus, from now on we consider n such that $np \ge L_p\mathbb{E}T_1$. For $A = (1+p)n\left[\mathbb{E}T_1\right]^{-1} \ge 1$ we get

$$\begin{aligned} &\mathbb{P}_\mu(N > \lceil A\rceil) \le \mathbb{P}(\tau_{\lceil A\rceil} - \tau_0 \le n) \le \mathbb{P}\left(\sum_{i=0}^{\lceil A\rceil - 1} T_{i+1} - \mathbb{E}T_{i+1} \le n - A\mathbb{E}T_1\right) \\ &= \mathbb{P}\left(\sum_{i=0}^{\lceil A\rceil - 1} T_{i+1} - \mathbb{E}T_{i+1} \le n - (1+p)n\right) = \mathbb{P}\left(\sum_{i=0}^{\lceil A\rceil - 1} T_{i+1} - \mathbb{E}T_{i+1} \le -np\right). \end{aligned} \tag{6.1.37}$$

Clearly, $\|T_{i+1} - \mathbb{E}T_{i+1}\|_{\psi_1} \le 2d$. Using Lemma 6.1.7, $\mathbb{E}T_1 \le d$ and $np \ge L_p\mathbb{E}T_1$ we get

$$\begin{aligned} &\mathbb{P}_\mu\left(N > \left\lceil (1+p)n\left[\mathbb{E}T_1\right]^{-1}\right\rceil\right) \le \exp\left(-\frac{p^2n^2}{4(A+1)4d^2 + 4dnp}\right) \\ &= \exp\left(-\frac{pn\mathbb{E}T_1}{16d^2\left(\frac{1+p}{p} + \frac{\mathbb{E}T_1}{pn}\right) + 4d\mathbb{E}T_1}\right) \\ &\le \exp\left(-\frac{pn\mathbb{E}T_1}{16d^2\left(\frac{1+p}{p} + \frac{1}{L_p}\right) + 4d^2}\right) = \exp\left(-\frac{pn\mathbb{E}T_1}{K_p d^2}\right) \le \exp\left(1 - \frac{pn\mathbb{E}T_1}{K_p d^2}\right), \end{aligned}$$

which finishes the proof of (6.1.36). The required properties of K_p follow from easy computations. ∎

The following lemma is a standard consequence of the tail estimates given in Lemma 6.1.8. Its proof, based on integration by parts, is analogous to that of Lemma 5.4 in [5].

Lemma 6.1.9. *Suppose that $\|\tau_1 - \tau_0\|_{\psi_1} \le d$ for some $d > 0$. Then for any $p > 0$,*

$$\left\|\left(N - (1+p)n\left[\mathbb{E}(\tau_1 - \tau_0)\right]^{-1}\right)_+\right\|_{\psi_1} \le \frac{4K_p d^2}{\left[\mathbb{E}(\tau_1 - \tau_0)\right]^2} \le \frac{4K_p d^2}{m^2},$$

where $K_p = L_p + \frac{16}{L_p}$ and $L_p = \frac{16}{p} + 20$. Moreover, $\frac{d^2K_p}{\left[\mathbb{E}(\tau_1-\tau_0)\right]^2} \ge K_p \ge K_\infty$.

Proof. Put $a = (1+p)n\left[\mathbb{E}(\tau_1-\tau_0)\right]^{-1}$, $c = \mathbb{E}(\tau_1-\tau_0)$ and $b = 2\frac{d^2K_p}{c^2} \geq 2K_p \geq 2K_\infty$. Then

$$\begin{aligned}\mathbb{E}\exp\left(\frac{(N-a)_+}{b}\right) &= 1+\int_0^\infty e^t\mathbb{P}\left(\frac{(N-a)_+}{b} > t\right)dt = 1+\int_0^\infty e^t\mathbb{P}\left(N > tb+a\right)dt \\ &= 1+\int_0^\infty e^t\mathbb{P}\left(N > \lceil tb+a-1\rceil\right)dt \leq e^{\frac{1}{b}} + \int_{1/b}^\infty e^t\mathbb{P}\left(N > \lceil tb+a-1\rceil\right)dt.\end{aligned}$$

Note that $b \geq 2K_p \geq 2K_\infty$ and thus $\exp(1/b) \leq \exp(\frac{1}{2K_\infty})$. To bound the above integral denote $\hat{p} = p + \frac{tb-1}{n}c$ and note that for $t \geq 1/b$ we have $\hat{p} \geq p > 0$ and $K_{\hat{p}} \leq K_p$. Therefore applying Lemma 6.1.8 with $p = \hat{p}$, we get

$$\int_{1/b}^\infty e^t\mathbb{P}\left(N > \lceil tb+a-1\rceil\right)dt \leq \int_{1/b}^\infty \exp\left(-\frac{\hat{p}nc}{K_{\hat{p}}d^2}+t+1\right)dt \leq \int_{1/b}^\infty \exp\left(-\frac{\hat{p}nc}{K_pd^2}+t+1\right)dt.$$

Now, due to the definition of $\hat{p}$,

$$-\frac{\hat{p}nc}{K_pd^2} = -\frac{2c^2t\frac{d^2K_p}{c^2}+pnc-c^2}{K_pd^2} = -2t+\frac{c^2-pnc}{K_pd^2} \leq -2t+\frac{1}{K_p} \leq -2t+\frac{1}{K_\infty}$$

which gives

$$\int_{1/b}^\infty e^t\mathbb{P}\left(N > \lceil tb+a-1\rceil\right)dt \leq \int_0^\infty \exp\left(-2t+\frac{1}{K_\infty}+t+1\right) = \exp\left(1+\frac{1}{K_\infty}\right).$$

Thus, using $K_\infty = \frac{104}{5}$ we conclude that

$$\mathbb{E}\exp\left(\frac{(N-a)_+}{b}\right) \leq \exp\left(\frac{1}{2K_\infty}\right) + \exp\left(1+\frac{1}{K_\infty}\right) \leq 4.$$

In order to finish the proof it is enough to apply the Jensen inequality. ■

6.2 Results

Before we formulate our main result let us introduce and explain the role of the following parameters:

$$\mathbf{a} = \left\|\sum_{k=0}^{\tau_0/m}|\Theta_k|\right\|_{\psi_\alpha,\mathbb{P}_{x^*}}, \quad \mathbf{b} = \left\|\sum_{k=0}^{\tau_0/m}|\Theta_k|\right\|_{\psi_\alpha,\mathbb{P}_{\pi^*}}, \quad \mathbf{c} = \|f(\Xi_1)\|_{\psi_\alpha}, \quad \mathbf{d} = \|\tau_1-\tau_0\|_{\psi_1}, \tag{6.2.1}$$

where $\Theta_k = \sum_{i=0}^{m-1} f(X_{km+i})$. Recall our "random block" decomposition

$$f\left(X_{[0,n-1]}\right) = \underbrace{f(\Xi_0)}_{I} + \underbrace{\left[\sum_{i=1}^N f(\Xi_i)\right]}_{II} - \underbrace{f(X_{[n,\tau_N+m-1]})}_{III}. \tag{6.2.2}$$

The parameter $\mathbf{a}$ (resp. $\mathbf{b}$) allows us to estimate the I (III) term on the right-hand side of (6.2.2), whereas the parameters $\mathbf{c}$ and $\mathbf{d}$ are used to control the middle term II. We note that $\mathbf{d}$ quantifies the geometric ergodicity of $\mathbf{X}$ and is finite as soon as $\mathbf{X}$ is geometrically ergodic. Moreover, all these parameters can be bounded for example by means of drift conditions widely used in the theory of Markov chains (see Remark 6.2.2). Finally, let us remind that $\sigma^2_{Mrv} = \mathrm{Var}_\pi(f(X_0)) + 2\sum_{i=1}^\infty \mathrm{Cov}_\pi(f(X_0), f(X_i))$ denotes the asymptotic variance of normalized partial sums of the process $(f(X_i))_i$.

We are now ready to formulate the first of our main results (recall the definitions of the small set C, m and δ from the minorization condition (6.1.7)).

Theorem 6.2.1. *Let $\mathbf{X}$ be a geometrically ergodic Markov chain and π be its unique stationary probability measure. Let $f\colon \mathcal{X} \to \mathbb{R}$ be a measurable function such that $\mathbb{E}_\pi f = 0$ and let $\alpha \in (0,1]$. Moreover, assume for simplicity that $m|n$. Then for all $x \in \mathcal{X}$ and $t > 0$,*

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \le 2\exp\left(-\frac{t^\alpha}{(23\mathbf{a})^\alpha}\right) + 2\left[\delta\pi(C)\right]^{-1}\exp\left(-\frac{t^\alpha}{(23\boldsymbol{b})^\alpha}\right) \\ + 6\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(27\boldsymbol{c})^\alpha}\right) + 6\exp\left(-\frac{t^2}{30n\sigma^2_{Mrv} + 8tM}\right) + \exp(1)\exp\left(-\frac{nm}{67\delta\pi(C)\boldsymbol{d}^2}\right), \tag{6.2.3}$$

where σ^2_{Mrv} denotes the asymptotic variance for the process $(f(X_i))_i$ given by (6.1.27), the parameters $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$ are defined by (6.2.1) and $M = \boldsymbol{c}(24\alpha^{-3}\log n)^{\frac{1}{\alpha}}$.

Remark 6.2.2. For the conditions under which $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are finite we refer to [5], where the authors give bounds on $\mathbf{a}, \mathbf{b}, \mathbf{c}$ under the classical drift conditions. If f is bounded then one easily shows that

$$\max(\mathbf{a}, \mathbf{b}) \le 2D\|f\|_\infty, \qquad \mathbf{c} \le D\|f\|_\infty, \tag{6.2.4}$$

where $D = \max\left(\mathbf{d}, \|\tau_0\|_{\psi_1,\,\mathbb{P}_{x^*}}, \|\tau_0\|_{\psi_1,\,\mathbb{P}_{\pi^*}}\right)$. For computable bounds on D we refer to [8].

Let us note that in Theorem 6.2.1 the right-hand side of the inequality does not converge to 0 when t tends to infinity (one of the terms depends on n but not on t). Usually in applications t is of order at most n and the other terms dominate on the right-hand side of the inequality, so this does not pose a problem. Nevertheless one can obtain another version of Theorem 6.2.1, namely

Theorem 6.2.3. *Under the assumptions and notation of Theorem 6.2.1 we have*

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \le 2\exp\left(-\frac{t^\alpha}{(54\mathbf{a})^\alpha}\right) + 2\left[\delta\pi(C)\right]^{-1}\exp\left(-\frac{t^\alpha}{(54\boldsymbol{b})^\alpha}\right) \\ + 4\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(27\boldsymbol{c})^\alpha}\right) + 6\exp\left(-\frac{t^2}{37(1+p)n\sigma^2_{Mrv} + 18M\boldsymbol{d}\sqrt{K_p}t}\right), \tag{6.2.5}$$

where $K_p = L_p + 16/L_p$ and $L_p = \frac{16}{p} + 20$.

It is well-known that for geometrically ergodic chains $\|\tau_0\|_{\psi_1,\,\mathbb{P}_{x^*}}$, $\|\tau_0\|_{\psi_1,\,\mathbb{P}_{\pi^*}}$, $\|\tau_1 - \tau_0\|_{\psi_1} < \infty$ (see [8] for constructive estimates). Therefore (6.2.4) and Theorem 6.2.1 lead to

Theorem 6.2.4. *Let $\mathbf{X}$ be a geometrically ergodic Markov chain and π be its unique stationary, probability measure. Let $f\colon \mathcal{X} \to \mathbb{R}$ be a bounded, measurable function such that $\mathbb{E}_\pi f = 0$. Fix $x \in \mathcal{X}$. Moreover assume that $\|\tau_0\|_{\psi_1,\delta_x^*}$, $\|\tau_0\|_{\psi_1,\pi^*}$, $\|\tau_1 - \tau_0\|_{\psi_1} \le D$. Then for all $t > 0$,*

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \le K\exp\left(-\frac{t^2}{32n\sigma^2_{Mrv} + 433t\delta\pi(C)\|f\|_\infty D^2\log n}\right), \tag{6.2.6}$$

where σ^2_{Mrv} is the asymptotic variance of $(f(X_i))_i$ and $K = \exp(10) + 2\delta^{-1}\pi(C)^{-1}$.

Remark 6.2.5. *Theorem 6.2.4 implies our main Theorem 2.3.5 from Part I with constants $K = \left(\exp(10) + 2\delta^{-1}\pi(C)^{-1}\right)$ and $\tau = 433\delta\pi(C)D^2$.*

6.2.1 Proofs of the main results

In this section we prove our main results. The structure of proofs of Theorems 6.2.1 and 6.2.3 is similar, and they contain a common part, which we present in Sections 6.2.2 and 6.2.3. The proof of Theorem 6.2.1 is concluded in Section 6.2.4 whereas that of Theorem 6.2.3 in Section 6.2.5. Theorem 6.2.4 is obtained as a corollary to Theorem 6.2.1 in Section 6.2.6.

Let us thus pass to the proofs of Theorems 6.2.1 and 6.2.3. Recall that $m|n$. The argument is based on the approach of [4] and [5] (see also [24] and [31]) and relies on the decomposition

$$\left|\sum_{i=0}^{n-1} f(X_i)\right| \le H_n + M_n + T_n, \tag{6.2.7}$$

where

$$H_n = \left|\sum_{i=0}^{\tau_0/m} \Theta_i \mathbb{1}_{N>0} + \mathbb{1}_{N=0} \sum_{i=0}^{n/m-1} \Theta_i\right|, \quad M_n = \left|\sum_{i=1}^{N} f(\Xi_i)\right|, \quad T_n = \left|\mathbb{1}_{N>0} \sum_{k=n}^{\tau_N+m-1} f(X_k)\right|,$$
$$N = \inf\{i \ge 0 \mid \tau_i + m - 1 \ge n - 1\}.$$

The proof is divided into three main steps. In the first two (common for both theorems) we get easy bounds on tails of H_n and T_n. The main, third step is devoted to obtaining two different estimates on the tail of M_n. To this end we use Lemmas 5.3.12, 6.1.8 (for the proof of Theorem 6.2.1) and Lemmas 5.3.16, 6.1.9 (for Theorem 6.2.3).

6.2.2 Estimate on H_n

Using $\{N = 0\} \subset \{\tau_0 \ge n - m\}$, the definition of $\mathbf{a}$ (see (6.2.1)) and Lemma D.0.4 we get

$$\begin{aligned}\mathbb{P}_{x^*}(H_n > t) &\le \mathbb{P}_{x^*}\left(\mathbb{1}_{N>0} \sum_{i=0}^{\tau_0/m} |\Theta_i| + \mathbb{1}_{N=0} \sum_{i=0}^{n/m-1} |\Theta_i| > t\right) \le \mathbb{P}_{x^*}\left(\sum_{i=0}^{\tau_0/m} |\Theta_i| > t\right) \\ &\le 2\exp\left(-\frac{t^\alpha}{\mathbf{a}^\alpha}\right).\end{aligned} \tag{6.2.8}$$

6.2.3 Estimate on T_n

By repeating verbatim the easy argument presented in the proof of Theorem 5.1 in [5], we obtain

$$\mathbb{P}\left(|T_n| > t\right) \le 2\left[\delta\pi(C)\right]^{-1} \exp\left(-\frac{t^\alpha}{\mathbf{b}^\alpha}\right). \tag{6.2.9}$$

We skip the details.

6.2.4 Proof of Theorem 6.2.1

Recall that $M = \mathbf{c}(24\alpha^{-3}\log n)^{\frac{1}{\alpha}}$ and note that without loss of generality we can assume that $t \ge 8M\log 6$. Otherwise (6.2.3) is trivial as the right hand side is greater than or equal to 1. Fix $p = 2/3$. We have $(A := \lceil (p+1)n(\mathbb{E}(\tau_1 - \tau_0))^{-1}\rceil)$

$$\begin{aligned}\mathbb{P}\left(M_n \ge t\right) &= \mathbb{P}\left(M_n \ge t,\ N \le A\right) + \mathbb{P}\left(M_n \ge t, N > A\right) \\ &\le \mathbb{P}\left(\sup_{1\le k\le A}\left|\sum_{i=1}^{k} f(\Xi_i)\right| \ge t\right) + \mathbb{P}\left(N > A\right).\end{aligned} \tag{6.2.10}$$

To control the first summand on the right-hand side of the above inequality we apply Corollary 5.3.14, $X_i := \Xi_i$, $c := \mathbf{c}$ and $n := A$ obtaining

$$\begin{aligned}P &:= \mathbb{P}\left(\sup_{1\le k\le A}\left|\sum_{i=1}^{k} f(\Xi_i)\right| \ge t\right) = \mathbb{P}\left(\sup_{1\le k\le A}\left|\sum_{i=1}^{k} F(\Xi_{i+1})\right| \ge t\right) \\ &\le 6\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24\mathbf{c})^\alpha}\right) + 6\exp\left(-\frac{t^2}{15\left(\lceil (p+1)n(\mathbb{E}(\tau_1-\tau_0))^{-1}\rceil + 3\right)\sigma_\infty^2 + 6tM}\right)\end{aligned}$$

$$\leq 6\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24\mathbf{c})^\alpha}\right)+6\exp\left(-\frac{t^2}{15\left((p+1)n(\mathbb{E}(\tau_1-\tau_0))^{-1}+4\right)\sigma_\infty^2+6tM}\right). \quad (6.2.11)$$

Recall that by (6.1.28), $\sigma_\infty^2=\sigma_{Mrv}^2\mathbb{E}(\tau_1-\tau_0)$. We will now obtain a comparison between σ_∞^2 and tM, which will allow us to reduce the above estimate to one in which the subgaussian coefficient is expressed only in terms of σ_{Mrv}^2. Thanks to Lemma D.0.2 applied to $Y:=(f(\Xi_1)/\mathbf{c})^\alpha$ and $\beta:=2/\alpha$, we have

$$\sigma_\infty^2\leq 3\mathbb{E}f(\Xi_1)_1^2\leq 3\mathbf{c}^2\Gamma(2/\alpha+1)\leq 3\mathbf{c}^2(2/\alpha)^{\frac{2}{\alpha}+1},$$

where the last inequality is a consequence of equation 4 in [69]. Moreover, recalling the definition of M and using the assumption $t\geq 8\log(6)M$, we obtain

$$tM\geq 8\log(6)M^2=8\log(6)\mathbf{c}^2(24\alpha^{-3}\log(n))^{\frac{2}{\alpha}}\geq 16\cdot 8\log(6)\mathbf{c}^2(2/\alpha)^{\frac{2}{\alpha}+1}\geq 76\sigma_\infty^2.$$

The last inequality in combination with (6.2.11) yields

$$\begin{aligned}P&\leq 6\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24\mathbf{c})^\alpha}\right)+6\exp\left(-\frac{t^2}{15(p+1)n\sigma_{Mrv}^2+7tM}\right)\\&\leq 6\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24\mathbf{c})^\alpha}\right)+6\exp\left(-\frac{t^2}{25n\sigma_{Mrv}^2+7tM}\right).\end{aligned} \quad (6.2.12)$$

Thus, in order to get a bound on $\mathbb{P}(M_n>t)$ it suffices to estimate the second term on the right-hand side of (6.2.10). To this end we use Lemma 6.1.8 with $p=2/3$ and $d=\mathbf{d}$ obtaining

$$\mathbb{P}\left(N>\lceil(1+p)n(\mathbb{E}(\tau_1-\tau_0))^{-1}\rceil\right)\leq\exp(1)\exp\left(-\frac{n\mathbb{E}(\tau_1-\tau_0)}{67\mathbf{d}^2}\right).$$

In combination with (6.2.10) and (6.2.12) this gives

$$\mathbb{P}(M_n\geq t)\leq 6\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24\mathbf{c})^\alpha}\right)+6\exp\left(-\frac{t^2}{25n\sigma_{Mrv}^2+7tM}\right)+\exp\left(-\frac{n\mathbb{E}(\tau_1-\tau_0)}{67\mathbf{d}^2}+1\right).$$

Combining the above inequality with (6.2.8) and (6.2.9), we get

$$\begin{aligned}&\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1}f(X_i)\right|>t\right)\leq\mathbb{P}\left(H_n\geq\frac{1-\sqrt{5/6}}{2}t\right)+\mathbb{P}\left(M_n\geq\sqrt{5/6}t\right)+\mathbb{P}\left(T_n\geq\frac{1-\sqrt{5/6}}{2}t\right)\\&\leq 2\exp\left(-\frac{t^\alpha}{(23\mathbf{a})^\alpha}\right)+2\left[\delta\pi(C)\right]^{-1}\exp\left(-\frac{t^\alpha}{(23\mathbf{b})^\alpha}\right)+\exp(1)\exp\left(-\frac{n\mathbb{E}(\tau_1-\tau)}{67\mathbf{d}^2}\right)\\&\quad+6\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(27\mathbf{c})^\alpha}\right)+6\exp\left(-\frac{t^2}{30n\sigma_{Mrv}^2+8tM}\right).\end{aligned}$$

In order to finish the proof of Theorem 6.2.1 it is enough to recall that $\mathbb{E}(\tau_1-\tau_0)=\delta^{-1}\pi(C)^{-1}m$.

6.2.5 Proof of Theorem 6.2.3

Recall that $M=\mathbf{c}(24\alpha^{-3}\log n)^{\frac{1}{\alpha}}$ and let $p>0$ be a parameter which will be fixed later on. We are going to apply Corollary 5.3.17, $X_i:=\Xi_{i+1}$, $c:=\mathbf{c}$, $\mathcal{F}_i:=\sigma\{\Xi_j\mid 0\leq j\leq i+1\}$. Clearly, N is a stopping time with respect to $\mathcal{F}$. Let $a=(1+p)\frac{n}{3}\left[\mathbb{E}(\tau_1-\tau_0)\right]^{-1}$. By Lemma 6.1.9 we get

$$\begin{aligned}\left\|(\lceil N/3\rceil-a+1)_+\right\|_{\psi_1}&\leq\frac{1}{3}\left\|\left(N-(1+p)n(\mathbb{E}(\tau_1-\tau_0))^{-1}\right)_+\right\|_{\psi_1}+\frac{2}{\log 2}\\&\leq\frac{4}{3}\mathbf{d}^2K_p+\frac{2}{\log 2}\leq\left(\frac{4}{3}+\frac{7}{50}\right)\mathbf{d}^2K_p,\end{aligned}$$

where the last inequality follows from (recall the definition of K_∞ from Lemma 6.1.8)

$$\frac{7}{50}K_p \geq \frac{7}{50}K_\infty = \frac{7}{50}\cdot\frac{104}{5} \geq \frac{2}{\log 2}.$$

Therefore $\max\left(2, \sqrt{\|\left(\lceil N/3\rceil - a + 1\right)_+\|_{\psi_1}}\right) \leq \sqrt{4/3 + 7/50}\sqrt{K_p}\cdot\mathbf{d}$ and we get that for arbitrary $p > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} f(\Xi_i)\right| > t\right) \leq 4\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(26\mathbf{c})^\alpha}\right) + 9\exp\left(-\frac{t^2}{34(1+p)\sigma^2_{Mrv} + 17M\mathbf{d}t\sqrt{K_p}}\right).$$

Using the above inequality together with (6.2.8), (6.2.9) we obtain

$$\begin{aligned}\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) &\leq \mathbb{P}\left(H_n \geq \frac{t}{54}\right) + \mathbb{P}\left(M_n \geq \frac{26t}{27}\right) + \mathbb{P}\left(T_n \geq \frac{t}{54}\right)\\ &\leq 2\exp\left(-\frac{t^\alpha}{(54\mathbf{a})^\alpha}\right) + 2\left[\delta\pi(C)\right]^{-1}\exp\left(-\frac{t^\alpha}{(54\mathbf{b})^\alpha}\right) + 4\exp(8)\exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(27\mathbf{c})^\alpha}\right)\\ &\quad + 9\exp\left(-\frac{t^2}{37(1+p)\sigma^2_{Mrv} + 18M\mathbf{d}t\sqrt{K_p}}\right)\end{aligned}$$

which concludes the proof of Theorem 6.2.3.

6.2.6 Proof of Theorem 6.2.4.

Let us denote $\|f\|_\infty$ by M and notice that for $t > nM$ the left-hand side of (6.2.6) vanishes, so we may assume that $t \leq nM$. Using (6.2.4), one can easily see that if $m|n$ then Theorem 6.2.1 applied with $\alpha = 1$ implies that

$$\begin{aligned}\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) &\leq \left(2 + 2\left[\delta\pi(C)\right]^{-1}\right)\exp\left(-\frac{t}{46DM}\right) + 6\exp(8)\exp\left(-\frac{t}{432DM}\right)\\ &\quad + 6\exp\left(-\frac{t^2}{30n\sigma^2_{Mrv} + 192tDM}\right) + \exp(1)\exp\left(-\frac{nm}{67\delta\pi(C)D^2}\right).\end{aligned} \tag{6.2.13}$$

The assumption $t \leq nM$ yields

$$\exp\left(-\frac{nm}{67\delta\pi(C)D^2}\right) \leq \exp\left(-\frac{tm}{67\delta\pi(C)MD^2}\right),$$

which plugged into (6.2.13) gives, after some elementary calculations, that (recall $K = \exp(10) + 2\left[\delta\pi(C)\right]^{-1}$)

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \leq K\exp\left(-\frac{t^2}{30n\sigma^2_{Mrv} + 432tD^2M\delta\pi(C)\log n}\right), \tag{6.2.14}$$

proving the theorem in the special case $m|n$.

Now we consider the case $m\nmid n$. Define $\lceil n\rceil_m$ to be the smallest integer greater or equal to n, which is divisible by m. Notice that without loss of generality we can assume that $t > 4330D^2M\delta\pi(C)$ (otherwise the assertion of the theorem is trivial as the right-hand side of (6.2.6) exceeds one). Since $D^2\delta\pi(C) > m$ (recall $\mathbb{E}(\tau_1 - \tau_0) = \delta^{-1}\pi(C)^{-1}m$), this implies that $t \geq 4330Mm$. Moreover, as $t \leq nM$, we also obtain that $n \geq 4330m$.

Thus, for $p = 1/4330$ we have $\left|\sum_{i=n}^{\lceil n\rceil_m} f(X_i)\right| \leq Mm \leq pt$, and as a consequence

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \leq \mathbb{P}_x\left(\left|\sum_{i=0}^{\lceil n\rceil_m - 1} f(X_i)\right| > (1-p)t\right). \tag{6.2.15}$$

Now using (6.2.14) and the inequality $n > 4330m$, we get

$$\begin{aligned}
\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) &\leq K \exp\left(-\frac{t^2}{31\lceil n\rceil_m \sigma_{Mrv}^2 + 433tD^2 M\delta\pi(C)\log n}\right) \\
&\leq K \exp\left(-\frac{t^2}{31(n+m)\sigma_{Mrv}^2 + 433tD^2 M\delta\pi(C)\log n}\right) \\
&\leq K \exp\left(-\frac{t^2}{32n\sigma_{Mrv}^2 + 433tD^2 M\delta\pi(C)\log n}\right).
\end{aligned}$$

This concludes the proof of Theorem 6.2.4.

Part III

Appendixes

Appendix A

Probability view on ergodic theorems

The aim of this chapter is to extend the mixed ergodic-probabilistic setting introduced in Section 1.3 and to present all basic ergodic theorems expressed in the language of stochastic processes. In Section A.2, we include a concise dictionary of such facts. Even though these results are nowadays a folklore knowledge (especially for researchers involved in ergodic theory), we discuss them in detail and include proofs of most of them in Section A.3. We believe that our unusual and definitely less frequent way of presentation of ergodic facts via stochastic processes nicely corresponds to the content of this thesis and deserves to be included.

In this part, for simplicity's sake, we assume that all processes are bilateral, that is $T = \mathbb{Z}$. Recall that in this case the left shift S is invertible.

A.1 Notation and definitions

Let us fix a stochastic process $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ and some measurable $\mathbf{A} \subset \mathcal{X}^{\mathbb{Z}}$ with $\mathbb{P}(\mathbf{X} \in \mathbf{A}) > 0$. Firstly, we slightly extend the definition of the ***return process*** given in (2.2.4) to

$$\mathbf{R} = \mathbf{R}^{(\mathbf{A})} = \left(R_i^{(\mathbf{A})}\right)_{i\in\mathbb{Z}}, \qquad R_i^{(\mathbf{A})} = \begin{cases} \inf\left\{k \geq 0 \mid S^k\mathbf{X} \in \mathbf{A}\right\}, & i = 0, \\ \inf\left\{k \geq R_{i-1}^{(\mathbf{A})} \mid S^k\mathbf{X} \in \mathbf{A}\right\}, & i \geq 1, \\ \sup\left\{k < R_{i+1}^{(\mathbf{A})} \mid S^k\mathbf{X} \in \mathbf{A}\right\}, & i \leq -1, \end{cases} \tag{A.1.1}$$

where we use the following conventions: $\inf \varnothing = \infty$ and $\sup \varnothing = -\infty$. Thus, for example, if $R_k = \infty$ for some $k \in \mathbb{Z}$ then $R_l = \infty$ for all $l \geq k$.

Note that for any $k \in \mathbb{Z}$, the random time R_k determines ***the randomly shifted processes*** (defined on $\{-\infty < R_k < \infty\}$),

$$\mathbf{X}^{(\mathbf{A},k)} = \left(X_i^{(\mathbf{A},k)}\right)_{i\in\mathbb{Z}}, \qquad X_i^{(\mathbf{A},k)} = X_{i+R_k^{(\mathbf{A})}}. \tag{A.1.2}$$

Furthermore, with every return process $\mathbf{R}^{(\mathbf{A})}$ we can associate the corresponding ***inter-arrival process***,

$$\mathbf{T} = \mathbf{T}^{(\mathbf{A})} = \left(T_k^{(\mathbf{A})}\right)_{k\in\mathbb{Z}}, \qquad T_k^{(\mathbf{A})} = R_k^{(\mathbf{A})} - R_{k-1}^{(\mathbf{A})}. \tag{A.1.3}$$

Here we use the following convention. As soon as $R_k^{(\mathbf{A})} < \infty$ or $R_{k-1}^{(\mathbf{A})} > -\infty$ then $T_k^{(\mathbf{A})}$ is well-defined. Otherwise, we put $T_k^{(\mathbf{A})} = \infty$.

When the set $\mathbf{A}$ (and the reference process $\mathbf{X}$) are clear from the context, we abbreviate, $\mathbf{R} = \mathbf{R}^{(\mathbf{A})}$, $\mathbf{X}^{(k)} = \mathbf{X}^{(\mathbf{A},k)}$, $\mathbf{T} = \mathbf{T}^{(\mathbf{A})}$.

Let us now pass to the ergodic setting and explain how the counterpart of (A.1.1) in dynamical systems looks like. To this end, given $\mathbf{X}$ and $\mathbf{A}$, we construct the corresponding ***induced dynamical system***

$$\mathcal{Q}_{\mathbf{A}} = (\mathcal{X}^T, S_{\mathbf{A}}, \mathcal{B}, \mu_{\mathbf{A}}), \tag{A.1.4}$$

where $\mu_{\mathbf{A}}$ stand for the conditioned version of μ, $S_{\mathbf{A}}$ for the ***induced shift***

$$S_{\mathbf{A}}\colon \{n_{\mathbf{A}} < \infty\} \to \mathcal{X}^{\mathbb{Z}}, \qquad S_{\mathbf{A}}\mathbf{x} = S^{n_{\mathbf{A}}(\mathbf{x})}(\mathbf{x}),$$

and the ***first return time function*** $n_{\mathbf{A}}\colon \mathbf{A} \to \mathbb{N} \cup \{\infty\}$ is given by

$$n_{\mathbf{A}}(\mathbf{x}) = \inf\{n \geq 1 \mid S^n\mathbf{x} \in \mathbf{A}\}. \tag{A.1.5}$$

One can show that $n_{\mathbf{A}} \overset{\mu_{\mathbf{A}}}{<} \infty$ (this is the content of the Poincaré recurrence theorem, see Theorem 2.11 in [27]) and $S\mu = \mu$ implies $S_{\mathbf{A}}\mu_{\mathbf{A}} = \mu_{\mathbf{A}}$ (cf. Lemma 2.43 in [27]). Therefore, indeed, (A.1.4) constitutes a dynamical system.

How exactly this ergodic setting (A.1.4) is connected with (A.1.2)? Firstly, note that $S_{\mathbf{A}}$ acts naturally on processes via

$$S_{\mathbf{A}}\mathbf{X} = (X_{i+\tau_{\mathbf{A}}}), \qquad \tau_{\mathbf{A}} = \inf\{n \geq 1 \mid S^n\mathbf{X} \in \mathbf{A}\}. \tag{A.1.6}$$

Clearly, if $\mathbf{X} \sim \mu$ then $\mu_{\mathbf{A}}$ corresponds to the distribution of $\mathbf{X}$ under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$. Moreover, by the Poincaré recurrence theorem, $\tau_{\mathbf{A}} \overset{\mu_{\mathbf{A}}}{<} \infty$, and thus, under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$, $\tau_{\mathbf{A}} = R_1^{(\mathbf{A})} < \infty$. In particular, $S_{\mathbf{A}}\mathbf{X} = \mathbf{X}^{(\mathbf{A},1)}$ is well defined on $\{\mathbf{X} \in \mathbf{A}\}$. Furthermore, $S_{\mathbf{A}}\mu_{\mathbf{A}} = \mu_{\mathbf{A}}$ is equivalent to $S_{\mathbf{A}}\mathbf{X} \sim \mathbf{X}$ under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$. Summing it up, **under** $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$, for any $k \in \mathbb{Z}$, process $\mathbf{X}^{(\mathbf{A},\mathbf{k})}$ is well defined and

$$\mathbf{X} \sim \mu_{\mathbf{A}}, \qquad \tau_{\mathbf{A}} = R_1^{(\mathbf{A})}, \qquad S_{\mathbf{A}}\mathbf{X} = \mathbf{X}^{(\mathbf{A},\mathbf{1})}, \qquad R_k^{(\mathbf{A})} < \infty, \qquad \mathbf{X} \sim \mathbf{X}^{(\mathbf{A},\mathbf{k})}.$$

Remark A.1.1. For the sake of convenience, we reserve symbol $\tau_{\mathbf{A}}$ for the first return time to $\mathbf{A}$ only for the process denoted by letter $\mathbf{X}$. Furthermore, $S_{\mathbf{A}}\mathbf{X}$ is well defined only on the set $\{\tau_{\mathbf{A}} < \infty\}$. Moreover, there is a slight difference between $\tau_{\mathbf{A}}$ and $R_0^{(\mathbf{A})}$ (recall (A.1.1)). The first one is a return time whereas the latter is a hitting time. More precisely, $\tau_{\mathbf{A}}$ coincides with $R_0 = R_0^{(\mathbf{A})}$ on the set $\{R_0 > 0\}$ and with $R_1^{(\mathbf{A})}$ on $\{R_0 = 0\}$. In order to better grasp the action of $S_{\mathbf{A}}$ on processes, note that for example for any $k \in \mathbb{Z}$, $\mathbf{X}^{(\mathbf{A},k+1)} = S_{\mathbf{A}}\mathbf{X}^{(\mathbf{A},k)}$ (on the set where both processes are well-defined).

Remark A.1.2 (Inter-arrival process as a factor)**.** Note that inter-arrival process $\mathbf{T} = \mathbf{T}^{(\mathbf{A})}$ can be regarded as a factor of $\mathbf{X}$. Indeed, clearly there is a natural function $\pi\colon \mathcal{X}^{\mathbb{Z}} \to \mathbb{N}^{\mathbb{Z}}$ such that $\mathbf{T} \overset{\mathbb{P}_{\mathbf{X}\in\mathbf{A}}}{=} \pi(\mathbf{X})$. Moreover, by the very definition of π,

$$\pi \circ S_{\mathbf{A}}(\mathbf{X}) = S\pi(\mathbf{X}) = S\mathbf{T}. \tag{A.1.7}$$

A.2 Summary of basic facts from the ergodic theory.

As usual, let $\mathbf{X} \sim \mu$, and denote by $\mathcal{I}_\mu$ or $\mathcal{I}_{\mathbf{X}}$ the σ-field of μ invariant sets, that is

$$\mathcal{I}_\mu = \mathcal{I}_{\mathbf{X}} = \left\{\mathbf{A} \subset \mathcal{X}^{\mathbb{Z}} \mid S\mathbf{A} \overset{\mu}{=} \mathbf{A}\right\}. \tag{A.2.1}$$

Recall that μ is called **ergodic** if $\mathcal{I}_\mu$ is trivial, in the sense of measure algebras, that is $\mathbf{A} \in \mathcal{I}_\mu$ implies $\mu(\mathbf{A})(1-\mu(\mathbf{A})) = 0$. We say that $\mathbf{X}$ is S-ergodic or $S_{\mathbf{A}}$-ergodic if so is the corresponding dynamical system $\mathcal{Q} = (\mathcal{X}^T, S, \mathcal{B}, \mu)$ or $\mathcal{Q}_{\mathbf{A}}$ from (A.1.4) respectively.

Remark A.2.1. It will turn out in a moment that $\mathbf{X}$ is S-ergodic (under $\mathbb{P}$) iff $\mathbf{X}$ is $S_{\mathbf{A}}$-ergodic (under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$).

Let us give some very simple observation which is tacitly used throughout this chapter.

Remark A.2.2 (Push forward of a conditional mean value)**.** Suppose that $X\colon \Omega \to \mathcal{X}$ is a random variable with $X \sim \mu$. Let $\mathcal{E}$ be a σ-field on $\mathcal{X}$ and $f\colon \mathcal{X} \to \mathcal{Y}$. Then

$$\mathbb{E}\left(f(X)|X^{-1}\mathcal{E}\right) \overset{\mathbb{P}}{=} \mathbb{E}_\mu\left(f|\mathcal{E}\right)(X).$$

Indeed, if $E \in \mathcal{E}$ then

$$\mathbb{E}\mathbb{E}_\mu\left(f|\mathcal{E}\right)(X)\mathbb{1}_{X\in E} = \int \mathbb{E}_\mu\left(f|\mathcal{E}\right)\mathbb{1}_E d\mathbb{P}_X = \int f\mathbb{1}_E dP_X = \mathbb{E}f(X)\mathbb{1}_{X\in E} = \mathbb{E}\mathbb{E}\left(f(X)|X^{-1}\mathcal{E}\right)\mathbb{1}_{X\in E}.$$

Now we list some basic ergodic facts translated into the language of stationary random processes. There are discussed in details (including most proofs) in Section A.3.

Recurrence

1. **Poincaré recurrence lemma (see Lemma A.3.2).** If $\mathbf{X}$ is stationary then for any $\mathbf{A}$ such that $\mathbb{P}(\mathbf{X} \in \mathbf{A}) > 0$,
$$\mathbb{P}_{\mathbf{X}\in\mathbf{A}}(S^n\mathbf{X} \in \mathbf{A} \ \ i.o.) = 1.$$
In particular all random variables $R_k^{(\mathbf{A})}$ and $T_k^{(\mathbf{A})}$ are $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$ a.s. finite (for $k \in \mathbb{Z}$) and all processes $\mathbf{X}^{(\mathbf{A},k)}$ are well-defined on whole set $\{\mathbf{X} \in \mathbf{A}\}$.

2. **Kac's lemma (see Lemma A.3.3).** Let $\mathbf{X} \sim \mu$ be a stationary process and $\mathbf{A}$ be such that $\mathbb{P}(\mathbf{X} \in \mathbf{A}) > 0$. Assume that a set $\mathbf{B} \in \mathcal{I}_\mu$ satisfies $\mathbb{P}(\mathbf{X} \in \mathbf{B}) > 0$ and $\mathbb{P}_{\mathbf{X}\in\mathbf{B}}(\tau_\mathbf{A} < \infty) = 1$. Then $\mathbf{X}$ is stationary under $\mathbb{P}_{\mathbf{X}\in\mathbf{B}}$ and for any f such that $f_+ \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{B}})$ or $f_- \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{B}})$, we have
$$\mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(\mathbf{X}) = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} \mathbb{1}_{\mathbf{X}\in A} \sum_{i=0}^{\tau_\mathbf{A}-1} f(S^i\mathbf{X}).$$
In particular, we recover the classical version of Kac's lemma, that is
$$\mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \tau_\mathbf{A} = 1/\mathbb{P}_{\mathbf{X}\in\mathbf{B}}(\mathbf{X} \in \mathbf{A}).$$
Furthermore, $\mathbb{P}_{\mathbf{X}\in\mathbf{B}}(\tau_\mathbf{A} = k) = \frac{1}{\mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}}\tau_\mathbf{A}} \mathbb{P}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}}(\tau_\mathbf{A} \geq k)$ and
$$1 + \mathbb{P}_{\mathbf{X}\in\mathbf{B}}(\mathbf{X} \notin \mathbf{A}) \mathbb{E}_{\mathbf{X}\notin\mathbf{A},\,\mathbf{X}\in\mathbf{B}} \tau_\mathbf{A} = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} \tau_\mathbf{A} = \frac{1}{2} \mathbb{P}_{\mathbf{X}\in\mathbf{B}}(\mathbf{X} \in \mathbf{A}) \mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \tau_\mathbf{A}^2 + \frac{1}{2}.$$

Induced process

1. **Induced process, $S\mathbf{X} \sim \mathbf{X}$ implies $S_\mathbf{A}\mathbf{X} \sim \mathbf{X}$ under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$ (see Lemma A.3.5 and Corollary A.3.6).** Suppose that $\mathbf{X}$ is stationary. Then for any $k \in \mathbb{Z}$, $\mathbf{X}^{(\mathbf{A},k)} \sim \mathbf{X}$ **under** $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$. In particular, **under** $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$, the inter-arrival process $\mathbf{T}^{(\mathbf{A})}$ is stationary (recall (A.1.3) and (A.1.7)).

   The next fact reverses this observation.

2. **Induced process, $S_\mathbf{A}\mathbf{X} \overset{\mathbb{Q}}{\sim} \mathbf{X}$ implies $S\mathbf{X} \overset{\mathbb{P}}{\sim} \mathbf{X}$ (see Lemma A.3.7).** Suppose that $\mathbf{X}$ is a **canonical process**. Let us denote $S_\mathbf{A}\mathbf{X}$ by $\mathbf{X}^{(\mathbf{A})}$. Suppose that under $\mathbb{Q}$ we have $\mathbf{X} \sim \mathbf{X}^{(\mathbf{A})}$ and $\mathbb{E}_\mathbb{Q}\eta_\mathbf{A} < \infty$, where $\eta_\mathbf{A} = \inf\{n \geq 1 \mid S^n\mathbf{X}^{(\mathbf{A})} \in \mathbf{A}\}$. Let
$$\mathbb{P} = \frac{1}{\mathbb{E}_\mathbb{Q}\eta_\mathbf{A}} \mathbb{E}_\mathbb{Q} \sum_{i=0}^{\eta_\mathbf{A}-1} S^i \delta_{\mathbf{X}^{(\mathbf{A})}}. \tag{A.2.2}$$
Then $\mathbf{X}$ is stationary with respect to $\mathbb{P}$. Note that the construction given by (A.2.2) can be treated as the inverse to $\mathbb{P} \to \mathbb{P}_{\mathbf{X}\in\mathbf{A}}$ because if $\mathbb{Q} = \mathbb{P}_{\mathbf{X}\in\mathbf{A}}$ then (A.2.2) retrieves $\mathbb{P}$. For the details we refer to Remark A.3.9.

3. **Ergodicity of the randomly shifted process (see [27], Lemma 2.43.)** If $\mathbf{X}$ is stationary and ergodic then $\mathbf{X}$ is $S_\mathbf{A}$-stationary and $S_\mathbf{A}$-ergodic under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$ (and so is the inter-arrival process $\mathbf{T} = \mathbf{T}^{(\mathbf{A})}$ due to (A.1.7)).

Ergodic theorems

1. **Birkhoff's ergodic theorem (see Theorem 2.30 in [27]).** Suppose that $\mathbf{X}$ is stationary. Then for any f such that $f_+(\mathbf{X}) \in L_1(\mathbb{P})$ or $f_-(\mathbf{X}) \in L_1(\mathbb{P})$ we have
$$\frac{1}{n}\sum_{i=0}^{n-1} f\left(S^i\mathbf{X}\right) \xrightarrow[L_1(\mathbb{P})]{\mathbb{P}\ a.s.} \mathbb{E}\left(f(\mathbf{X}) \,|\mathbf{X}^{-1}\mathcal{I}_\mu\right).$$
Furthermore, for any $\mathbf{A}$ such that $\mathbb{P}(\mathbf{X} \in \mathbf{A}) > 0$ (recall that in such case $\mathbf{X} \sim S_\mathbf{A}\mathbf{X}$ under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$) and f such that $f_+(\mathbf{X}) \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{A}})$ or $f_-(\mathbf{X}) \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{A}})$ we have
$$\frac{1}{n}\sum_{i=0}^{n-1} f\left(\mathbf{X}^{(\mathbf{A},i)}\right) \xrightarrow[L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{A}})]{\mathbb{P}_{\mathbf{X}\in\mathbf{A}}\ a.s.} \mathbb{E}_{\mathbf{X}\in\mathbf{A}}\left(f(\mathbf{X}) \,|\mathbf{X}^{-1}\mathcal{I}_{\mu_\mathbf{A}}\right). \tag{A.2.3}$$

Since $\mathbf{T}$ is stationary (see Corollary A.3.6), it follows that for any f such that $f_+(\mathbf{T}) \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{A}})$ or $f_-(\mathbf{T}) \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{A}})$ we have

$$\frac{1}{n}\sum_{i=0}^{n-1} f\left(S^i\mathbf{T}\right) \xrightarrow[L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{A}})]{\mathbb{P}_{\mathbf{X}\in\mathbf{A}}\ a.s.} \mathbb{E}_{\mathbf{X}\in\mathbf{A}}\left(f\left(\mathbf{T}\right)|\mathbf{T}^{-1}\mathcal{I}_{\mathbf{T}}\right). \tag{A.2.4}$$

2. **Maker's ergodic theorem (see [71]).** Suppose that $\mathbf{X}$ is stationary. Then for any sequence (f_i) such that $f_i(\mathbf{X}) \xrightarrow{a.s.} f(\mathbf{X})$ and $\sup_i |f_i(\mathbf{X})| \in L_1(\mathbb{P})$, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f_i\left(S^i\mathbf{X}\right) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} f_{n-i}\left(S^i\mathbf{X}\right) = \mathbb{E}\left(f\left(\mathbf{X}\right)|\mathbf{X}^{-1}\mathcal{I}_{\mu}\right), \tag{A.2.5}$$

where the limit exists almost surely and in $L_1(\mathbb{P})$. Moreover, the obvious modifications of (A.2.3) and (A.2.4) (where f is replaced by the sequence of f_i's) hold.

A.3 Proofs

In this part, unless stated otherwise, $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ is a **stationary bilateral process** with $X_i \in \mathcal{X}$, $\mathbf{A}$ is such that $\mathbb{P}(\mathbf{X}\in\mathbf{A}) > 0$ and μ stands for the distribution of $\mathbf{X}$ under $\mathbb{P}$. For any events $A_i \in \mathcal{F}$ where $i \in N \subset \mathbb{Z}$ and $|N| = \infty$, the abbreviation "$\{A_i,\ N\text{-i.o.}\}$" denotes the event $\bigcap_{i\in N}\bigcup_{k\geq i,\ k\in N} A_k$ that is the event in which infinitely many A_i's, for $i \in N$, occurred simultaneously.

At the beginning, let us give a simple remark which will be used in the upcoming proofs.

Remark A.3.1. We have $S^{-k}\{\mathbf{X}\in\mathbf{A}_0,\dots,S^n\mathbf{X}\in\mathbf{A}_n\} = \left\{S^k\mathbf{X}\in\mathbf{A}_0,\dots,S^{n+k}\mathbf{X}\in\mathbf{A}_n\right\}$ and $\{S^k\mathbf{X}\in\mathbf{A}\} = \{\mathbf{X}\in S^{-k}\mathbf{A}\}$ for any $k\in\mathbb{Z}$ and sets $\mathbf{A},\mathbf{A}_i$. In particular, if $S\mathbf{B} \stackrel{\mu}{=} \mathbf{B}$ then for any $k\in\mathbb{Z}$, $\left\{S^k\mathbf{X}\in\mathbf{B}\right\} \stackrel{\mathbb{P}}{=} \{\mathbf{X}\in\mathbf{B}\}$.

Lemma A.3.2 (Poincaré Recurrence lemma)**.** *We have*

$$\mathbb{P}_{\mathbf{X}\in\mathbf{A}}\left(S^i\mathbf{X}\in\mathbf{A},\ \ \mathbb{N}-i.o.\right) = 1. \tag{A.3.1}$$

Proof. Since the events, $B_i = \left\{S^i\mathbf{X}\in\mathbf{A}, S^{i+1}\mathbf{X}\notin\mathbf{A}, S^{i+2}\mathbf{X}\notin\mathbf{A},\dots\right\} = S^{-i}B_0$ are pairwise disjoint, thus, by the stationarity of $\mathbf{X}$, we must have $\mathbb{P}(B_i) = 0$. In particular $\mathbb{P}(B_0) = 0$ implies

$$\mathbb{P}\left(B_0^c\right) = \mathbb{P}_{\mathbf{X}\in\mathbf{A}}\left(\bigcup_{i\geq 1} S^i\mathbf{X}\in\mathbf{A}\right) = 1.$$

It remains to use $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}\left(\bigcap_{i\geq 0} B_i^c\right) = 1$. ∎

Lemma A.3.3 (Kac's lemma)**.** *Assume that $\mathbf{B}$ is such that $\mathbb{P}(\mathbf{X}\in\mathbf{B}) > 0$, $\mathbb{P}_{\mathbf{X}\in\mathbf{B}}(\tau_{\mathbf{A}} < \infty) = 1$ and $\mathbf{B}\in\mathcal{I}_{\mathbf{X}}$, where $\mathcal{I}_{\mathbf{X}}$ denotes the invariant σ-field (recall (A.2.1)). Then $\mathbf{X}$ under $\mathbb{P}_{\mathbf{X}\in\mathbf{B}}$ is stationary and for any f such that $f_+ \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{B}})$ or $f_- \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{B}})$ we have*

$$\mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(\mathbf{X}) = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} \mathbb{1}_{\mathbf{X}\in\mathbf{A}} \sum_{i=0}^{\tau_{\mathbf{A}}-1} f(S^i\mathbf{X}). \tag{A.3.2}$$

Proof. We have

$$\mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(\mathbf{X}) = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(\mathbf{X})\mathbb{1}_{\mathbf{X}\in\mathbf{A}} + \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(\mathbf{X})\mathbb{1}_{\mathbf{X}\in\mathbf{A}^c}.$$

By the stationarity and the shift-invariance of $\mathbf{B}$, we get

$$\begin{aligned}&\mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(\mathbf{X})\mathbb{1}_{\mathbf{X}\in\mathbf{A}^c} = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S\mathbf{X})\mathbb{1}_{S\mathbf{X}\in\mathbf{A}^c} = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S\mathbf{X})\mathbb{1}_{S\mathbf{X}\in\mathbf{A}^c,\mathbf{X}\in\mathbf{A}} + \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S\mathbf{X})\mathbb{1}_{S\mathbf{X}\in\mathbf{A}^c,\mathbf{X}\in\mathbf{A}^c}\\ &= \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S\mathbf{X})\mathbb{1}_{\tau_{\mathbf{A}}>1,\mathbf{X}\in\mathbf{A}} + \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S\mathbf{X})\mathbb{1}_{S\mathbf{X}\in\mathbf{A}^c,\mathbf{X}\in\mathbf{A}^c}.\end{aligned}$$

Similarly,

$$\mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S\mathbf{X})\mathbb{1}_{S\mathbf{X}\in\mathbf{A}^c,\mathbf{X}\in\mathbf{A}^c} = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S^2\mathbf{X})\mathbb{1}_{\tau_{\mathbf{A}}>2,\mathbf{X}\in\mathbf{A}} + \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S^2\mathbf{X})\mathbb{1}_{\mathbf{X},S\mathbf{X},S^2\mathbf{X}\in\mathbf{A}^c}.$$

Therefore, inductively, we get that for any $n \in \mathbb{N}$,

$$\mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(\mathbf{X}) = \sum_{i=0}^{n} \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S^i\mathbf{X}) \mathbb{1}_{\tau_{\mathbf{A}}>i, \mathbf{X}\in\mathbf{A}} + \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S^n\mathbf{X}) \mathbb{1}_{\mathbf{X}, S\mathbf{X}, \ldots, S^n\mathbf{X}\in\mathbf{A}^c}. \tag{A.3.3}$$

Firstly, we show our claim for $f \geq 0$ and $\|f\|_\infty < \infty$. Note that

$$\lim_{n\to\infty} \sum_{i=0}^{n} \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S^i\mathbf{X}) \mathbb{1}_{\tau_{\mathbf{A}}>i, \mathbf{X}\in\mathbf{A}} = \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} \mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S^i\mathbf{X}) \mathbb{1}_{\tau_{\mathbf{A}}=j, \mathbf{X}\in\mathbf{A}} = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} \sum_{i=0}^{\tau_{\mathbf{A}}-1} f(S^i\mathbf{X}) \mathbb{1}_{\mathbf{X}\in\mathbf{A}}$$

and

$$\mathbb{E}_{\mathbf{X}\in\mathbf{B}} f(S^n\mathbf{X}) \mathbb{1}_{\mathbf{X}, S\mathbf{X}, \ldots, S^n\mathbf{X}\in\mathbf{A}^c} \leq \|f\|_\infty \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X}, S\mathbf{X}, \ldots, S^n\mathbf{X} \in \mathbf{A}^c) = \|f\|_\infty \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\tau_{\mathbf{A}} > n) \to 0.$$

Now the case of $f \geq 0$ follows from considering $f \wedge n$ and the monotone convergence theorem, whereas the general one, from the decomposition $f = f_+ - f_-$. ■

Remark A.3.4. This version of Kac lemma is slightly less known, though, the idea of the proof is exactly the same as in the classical case where $f = 1$ is a constant function and $\mathbf{B}$ is the whole space. Moreover, it can be treated as a translator between systems $\mathcal{Q}$ and $\mathcal{Q}_{\mathbf{A}}$. Here we list some useful consequences of this version of Kac's lemma.

- For $f = 1$, (A.3.2) reduces to

$$\mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \tau_{\mathbf{A}} = 1/\mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}). \tag{A.3.4}$$

- Recall (A.1.5) and that by the very definition $\tau_{\mathbf{A}} = n_{\mathbf{A}}(\mathbf{X})$. We have

$$1 + \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \notin \mathbf{A}) \mathbb{E}_{\mathbf{X}\notin\mathbf{A},\, \mathbf{X}\in\mathbf{B}} \tau_{\mathbf{A}} = \mathbb{E}_{\mathbf{X}\in\mathbf{B}} \tau_{\mathbf{A}} = \frac{1}{2} \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}) \mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \tau_{\mathbf{A}}^2 + \frac{1}{2}. \tag{A.3.5}$$

In particular,

$$\tau_A \in L_1(\mathbb{P}_{\mathbf{X}\in\mathbf{B}}) \quad \Leftrightarrow \quad \tau_A \in L_2(\mathbb{P}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}}).$$

Indeed, the left hand side equality of (A.3.5) follows from the splitting of the integral according to sets $\{\mathbf{X} \in \mathbf{A}\}$ and $\{\mathbf{X} \in \mathbf{A}^c\}$ and use of (A.3.4). The other equality is a consequence of an application of (A.3.2) for $f = n_{\mathbf{A}} \geq 0$, namely,

$$\mathbb{E}_{\mathbf{X}\in\mathbf{B}} \tau_{\mathbf{A}} \overset{(A.3.2)}{=} \mathbb{P}(\mathbb{1}_{\mathbf{X}\in\mathbf{A}}) \mathbb{E}_{\mathbf{X}\in\mathbf{B}} \sum_{i=0}^{\tau_{\mathbf{A}}-1} (\tau_{\mathbf{A}} - i) = \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}) \mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \frac{\tau_{\mathbf{A}}(\tau_{\mathbf{A}} + 1)}{2}.$$

It remains to use (A.3.4).

Similarly, one can obtain that for any $j \geq 2$,

$$\mathbb{E}_{\mathbf{X}\in\mathbf{B}} [R_j - R_{j-1}] \leq \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}) \mathbb{E}_{\mathbf{X}\in\mathbf{B}} \tau_{\mathbf{A}}^2 = 2\mathbb{E}_{\mathbf{X}\in\mathbf{B}} \tau_{\mathbf{A}} - 1.$$

Indeed, using (A.3.2), Schwarz's inequality, the stationarity of sequence $R_j - R_{j-1}$ (under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$) and (A.3.5) we get

$$\begin{aligned}
&\mathbb{E}_{\mathbf{X}\in\mathbf{B}} [R_j - R_{j-1}] \overset{(A.3.2)}{=} \mathbb{E}_{\mathbf{X}\in\mathbf{B}} \mathbb{1}_{\mathbf{X}\in\mathbf{A}} \sum_{i=0}^{\tau_{\mathbf{A}}-1} [R_j - R_{j-1}] = \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}) \mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \tau_{\mathbf{A}} (R_j - R_{j-1}) \\
&\leq \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}) \sqrt{\mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \tau_{\mathbf{A}}^2} \sqrt{\mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} (R_j - R_{j-1})^2} \\
&= \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}) \mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \tau_{\mathbf{A}}^2 \overset{(A.3.5)}{=} 2\mathbb{E}_{\mathbf{X}\in\mathbf{B}} \tau_{\mathbf{A}} - 1.
\end{aligned}$$

- We can easily identify the distribution of $\tau_{\mathbf{A}}$ **under** $\mathbb{P}_{\mathbf{X}\in\mathbf{B}}$. Namely,

$$\mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\tau_{\mathbf{A}} = k) = \frac{1}{\mathbb{E}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} \tau_{\mathbf{A}}} \mathbb{P}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} (\tau_{\mathbf{A}} \geq k).$$

To see this, for any fixed $k \in \mathbb{N}$ consider $f(\mathbf{x}) = \mathbb{1}_{n_{\mathbf{A}}(\mathbf{x})=k}$ and then use (A.3.2). Indeed, if $0 \leq i \leq \tau_{\mathbf{A}} - 1$ and $\tau_{\mathbf{A}} \geq k$ then $f(S^i\mathbf{X}) = \mathbb{1}_{n_{\mathbf{A}}(S^i\mathbf{X})=k} = \mathbb{1}_{\tau_{\mathbf{A}}=i+k}$ and

$$\mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\tau_{\mathbf{A}} = k) = \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}) \mathbb{E}_{\mathbf{X}\in\mathbf{B}\cap\mathbf{A}} \mathbb{1}_{\tau_{\mathbf{A}}\geq k} \sum_{i=0}^{\tau_{\mathbf{A}}-1} \mathbb{1}_{\tau_{\mathbf{A}}=i+k} = \mathbb{P}_{\mathbf{X}\in\mathbf{B}} (\mathbf{X} \in \mathbf{A}) \mathbb{P}_{\mathbf{X}\in\mathbf{A}\cap\mathbf{B}} (\tau_{\mathbf{A}} \geq k).$$

Lemma A.3.5. *Let $\mathbf{A}_0 \subset \mathbf{A}$. Then for any $n \in \mathbb{N}$,*

$$\mathbb{P}_{\mathbf{X}\in\mathbf{A}}\left(\mathbf{X} \in \mathbf{A}_0, \ldots, S^n\mathbf{X} \in \mathbf{A}_n\right) = \mathbb{P}_{\mathbf{X}\in\mathbf{A}}\left(S^{\tau_\mathbf{A}}\mathbf{X} \in \mathbf{A}_0, \ldots, S^{\tau_\mathbf{A}+n}\mathbf{X} \in \mathbf{A}_n\right). \tag{A.3.6}$$

*In other words, $S_\mathbf{A}\mathbf{X} \sim \mathbf{X}$ **under** $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$.*

Proof. We do a similar trick as in the Kac lemma. Since $\mathbf{X}$ is stationary

$$\mathbb{P}\left(\mathbf{X} \in \mathbf{A}_0, \ldots, S^n\mathbf{X} \in \mathbf{A}_n\right) = \mathbb{P}\left(\tau_\mathbf{A} = 1, \mathbf{X} \in \mathbf{A}, S^1\mathbf{X} \in \mathbf{A}_0, \ldots, S^{n+1}\mathbf{X} \in \mathbf{A}_n\right)$$
$$+ \mathbb{P}\left(\mathbf{X} \notin \mathbf{A}, S^1\mathbf{X} \in \mathbf{A}_0, \ldots, S^{n+1}\mathbf{X} \in \mathbf{A}_n\right).$$

Now, we repeat this argument to get inductively that

$$\mathbb{P}\left(\mathbf{X} \in \mathbf{A}_0, \ldots, S^n\mathbf{X} \in \mathbf{A}_n\right) = \sum_{k=1}^{N} \mathbb{P}\left(\tau_\mathbf{A} = k, \mathbf{X} \in \mathbf{A}, S^k\mathbf{X} \in \mathbf{A}_0, \ldots, S^{n+k}\mathbf{X} \in \mathbf{A}_n\right)$$
$$+ \mathbb{P}\left(\mathbf{X} \notin \mathbf{A}, \ldots, S^{N-1}\mathbf{X} \notin \mathbf{A}, S^N\mathbf{X} \in \mathbf{A}_0, \ldots, S^{n+N}\mathbf{X} \in \mathbf{A}_n\right).$$

Note that, if $N \to \infty$, then the first term converges to $\mathbb{P}\left(\mathbf{X} \in \mathbf{A}, S^{\tau_\mathbf{A}}\mathbf{X} \in \mathbf{A}_0, \ldots, S^{\tau_\mathbf{A}+n}\mathbf{X} \in \mathbf{A}_n\right)$. Therefore, it remains to show that the second term vanishes as $N \to \infty$. However, this immediately follows from the observation that sets

$$A_N = \left\{\mathbf{X} \notin \mathbf{A}, \ldots, S^{N-1}\mathbf{X} \notin \mathbf{A}, S^N\mathbf{X} \in \mathbf{A}_0, \ldots, S^{n+N}\mathbf{X} \in \mathbf{A}_n\right\} \tag{A.3.7}$$

are pairwise disjoint and $\sum_{N\in\mathbb{N}} \mathbb{P}\left(A_N\right) \le 1$.

■

Since the inter-arrival process $\mathbf{T}^{(\mathbf{A})}$ is a function of $\mathbf{X}$ (say $\mathbf{T}^{(\mathbf{A})} = f(\mathbf{X})$) and $S\mathbf{T} = f(S_\mathbf{A}\mathbf{X})$ (in other words $\mathbf{T}^{(\mathbf{A})}$ is a factor of $\mathbf{X}$) we immediately get the following.

Corollary A.3.6 (Stationarity of inter-arrival times)**.** *Under $\mathbb{P}_\mathbf{A}$ the inter-arrival process $\mathbf{T}^{(\mathbf{A})}$ is stationary.*

We have shown that if $\mathbf{X}$ is stationary then under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$, $\mathbf{X} \sim S_\mathbf{A}\mathbf{X}$. Now we reverse this observation.

Lemma A.3.7. *Suppose that $\mathbf{X}$ is a* **canonical** *process. Let us denote $S_\mathbf{A}\mathbf{X}$ by $\mathbf{X}^{(\mathbf{A})}$. Suppose that under $\mathbb{Q}$, $\mathbf{X} \sim \mathbf{X}^{(\mathbf{A})}$ with $\mathbb{E}_\mathbb{Q}\eta_\mathbf{A} < \infty$, where $\eta_\mathbf{A} = \inf\{n \ge 1 \mid S^n\mathbf{X}^{(\mathbf{A})} \in \mathbf{A}\}$. Let*

$$\mathbb{P} = \frac{1}{\mathbb{E}_\mathbb{Q}\eta_\mathbf{A}}\mathbb{E}_\mathbb{Q} \sum_{i=0}^{\eta_\mathbf{A}-1} S^i\delta_{\mathbf{X}^{(\mathbf{A})}} \tag{A.3.8}$$

Then $\mathbf{X}$ is stationary with respect to $\mathbb{P}$.

Remark A.3.8. If $\mathbf{X}$ is not canonical then (A.3.8) can be rephrased as

$$\mathbb{P}\left(\mathbf{X} \in \mathbf{F}\right) = \frac{1}{\mathbb{E}_\mathbb{Q}\eta_\mathbf{A}}\mathbb{E}_\mathbb{Q} \sum_{i=0}^{\eta_\mathbf{A}-1} \mathbb{1}_{S^i\mathbf{X}^\mathbf{A} \in \mathbf{F}} \tag{A.3.9}$$

and Lemma A.3.7 asserts that if $\mathbf{X}$ has a distribution given by (A.3.9) then $\mathbf{X}$ is stationary.

Proof. We have

$$\mathbb{P} - S\mathbb{P} = \frac{1}{\mathbb{E}_\mathbb{Q}\eta_\mathbf{A}}\mathbb{E}_\mathbb{Q}\left[\delta_{\mathbf{X}^{(\mathbf{A})}} - S^{\eta_\mathbf{A}}\delta_{\mathbf{X}^{(\mathbf{A})}}\right] = \frac{1}{\mathbb{E}_\mathbb{Q}\eta_\mathbf{A}}\left[\mathcal{L}_\mathbb{Q}\left(\mathbf{X}^{(\mathbf{A})}\right) - \mathcal{L}_\mathbb{Q}\left(S^{\eta_\mathbf{A}}\mathbf{X}^{(\mathbf{A})}\right)\right].$$

It remains to notice that the assumption $\mathbf{X} \overset{\mathbb{Q}}{\sim} \mathbf{X}^{(\mathbf{A})}$ implies $\mathbf{X}^{(\mathbf{A})} \overset{\mathbb{Q}}{\sim} S^{\eta_\mathbf{A}}\mathbf{X}^{(\mathbf{A})}$. ■

Remark A.3.9. The changes of underlying measures proposed by Lemmas A.3.5 and A.3.7 may be treated as a reverse to each other. More precisely, if $\mathbf{X}$ is stationary under $\mathbb{P}$ then $S_\mathbf{A}\mathbf{X}$ is stationary under $\mathbb{Q} = \mathbb{P}_{\mathbf{X}\in\mathbf{A}}$. Note that in this case Lemma A.3.7 transforms such $\mathbb{Q}$ back to $\mathbb{P}$. Indeed, this is a consequence of the Kac's lemma (see (A.3.2)). Conversely, if $S_\mathbf{A}\mathbf{X}$ is stationary under $\mathbb{Q}$ and $\mathbb{P}$ is as in Lemma A.3.7 then $\mathbb{P}_{\mathbf{X}\in\mathbf{A}} = \mathbb{Q}$. This follows from $\mathbb{Q}\left(\mathbf{X}^{(\mathbf{A})} \in \mathbf{A}\right) = 1$ and

$$\mathbb{P}\left(\mathbf{X} \in \mathbf{A} \cap \mathbf{F}\right) = \frac{1}{\mathbb{E}_\mathbb{Q}\eta_\mathbf{A}}\mathbb{E}_\mathbb{Q} \sum_{i=0}^{\eta_\mathbf{A}-1} \mathbb{1}_{\mathbf{A}\cap\mathbf{F}}\left(S^i\mathbf{X}^\mathbf{A}\right) = \frac{1}{\mathbb{E}_\mathbb{Q}\eta_\mathbf{A}}\mathbb{Q}\left(\mathbf{X}^\mathbf{A} \in \mathbf{A} \cap \mathbf{F}\right) = \frac{1}{\mathbb{E}_\mathbb{Q}\eta_\mathbf{A}}\mathbb{Q}\left(\mathbf{X}^\mathbf{A} \in \mathbf{F}\right).$$

The following lemma is standard and thus we omit its proof (which can be found for example in [27], Lemma 2.43).

Lemma A.3.10. *If $\mathbf{X}$ is stationary and ergodic under $\mathbb{P}$ then $\mathbf{X}$ is $S_\mathbf{A}$-stationary and $S_\mathbf{A}$-ergodic under $\mathbb{P}_{\mathbf{X}\in\mathbf{A}}$.*

Appendix B

Tail σ-fields

Our studies concerning tail σ-algebras are motivated by the fact they naturally appear in entropy problems. For example, for any zero entropy process the one-sided tail σ-algebra explains the whole process (recall the end of Section 3.1.2). Moreover, as we have already seen in Theorem 3.2.12, the double sided tail σ-fields is a crucial part of criterion for the problem of retrieving a lost signal.

Fix some stationary process $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ such that $X_i \in \mathcal{X}$, with $|\mathcal{X}| < \infty$. Recall that the tail σ-fields are defined as

$$\mathcal{T}_{past}(\mathbf{X}) = \bigcap_{n\geq 0} \sigma\left(X_{(-\infty,-n]}\right), \quad \mathcal{T}_{future}(\mathbf{X}) = \bigcap_{n\geq 0} \sigma\left(X_{[n,\infty)}\right), \quad \mathcal{T}_{double}(\mathbf{X}) = \bigcap_{n\geq 0} \sigma\left(X_{(-\infty,-n]}, X_{[n,\infty)}\right).$$

When the whole process $\mathbf{X}$ is explained by $\mathcal{T}_{future}$ or $\mathcal{T}_{past}$ ($\mathcal{T}_{double}$ respectively) then we say that $\mathbf{X}$ is ***deterministic*** (***bilaterally-deterministic*** respectively). In other words, $\mathbf{X}$ is deterministic iff $\mathbf{H}(\mathbf{X}) = 0$.

B.1 Pinsker's algebra

Let T be an ergodic endomorphism of a standard probability Borel space $(X, \mathcal{B}, \mu)$. If T is (is not) an automorphism, we will speak of invertible (non-invertible) case.

Remark B.1.1. In what follows we will extensively use some properties of Shannon's entropy with respect to a measurable (at most countable) partition $\mathcal{A}$ of X,

$$\mathbf{H}(\mathcal{A}) = \sum_{A\in\mathcal{A}} -\mu(A)\log_2 \mu(A) \tag{B.1.1}$$

and Kolmogorov-Sinai entropy

$$\mathbf{H}(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n}\mathbf{H}\left(\mathcal{A}_0^{n-1}\right), \tag{B.1.2}$$

where for any $i \leq j$, $\mathcal{A}_i^j = \bigvee_{k=i}^{j} T^{-k}\mathcal{A}$. Since these objects are very closely related to those of Shannon's entropy of a random variable (see Section 3.1.1) and entropy rate of a process (see Section 3.1.2), respectively, we take for granted that the reader is familiar with these notions. If not, as an introduction to this subject, we recommend the second part of Glasner's book (see [45]).

Recall that with any set A we can associate the binary partition

$$\mathcal{P}_A = \{A, A^c\}.$$

Morever, the ***Pinsker σ-algebra*** is given by

$$\Pi(T) = \Pi = \{A \in \mathcal{B} \mid \mathbf{H}(T, \mathcal{P}_A) = 0\}.$$

Furthermore, the ***tail σ-fields (associated to some partition $\mathcal{A}$)*** are defined as

$$\mathcal{T}_{past}(T, \mathcal{A}) = \bigcap_{n\geq 0} \mathcal{A}_{-\infty}^{-n}, \qquad \mathcal{T}_{future}(T, \mathcal{A}) = \bigcap_{n\geq 0} \mathcal{A}_n^{\infty},$$

where for any $i \leq j$, $\mathcal{A}_i^j = \bigvee_{k=i}^{j} T^{-k}\mathcal{A}$.

Let us list some basic properties of Π.

- Π is T-invariant, countably μ-generated σ-algebra (see Proposition B.1.3).
- $\Pi(T^k) = \Pi(T)$ for any $k \geq 1$. If T is invertible then one can take $k \in \mathbb{Z}\backslash\{0\}$ (see Proposition B.1.4).
- If $\mathcal{A}$ is a countable partition such that $\mathbf{H}(\mathcal{A}) < \infty$ and T is invertible then $\mathcal{A} \in \Pi$ iff $\mathbf{H}(T, \mathcal{A}) = 0$ (see Proposition B.1.6).
- If T is invertible then $\Pi(T) = \bigvee_{\mathcal{A}\in\mathscr{P}_{fin}} \mathcal{T}_{past}(\mathcal{A}, T)$, where $\mathscr{P}_{fin}$ stands for the family of finite partitions (see Proposition B.1.7).
- If T is invertible and $\mathcal{A}$ is a **countable generating partition** for T then in the sense of measure algebras
$$\mathcal{T}_{past}(\mathcal{A}, T) = \mathcal{T}_{future}(\mathcal{A}, T) = \Pi(T) \tag{B.1.3}$$
(see Proposition B.1.8).
- For any ergodic systems (X, μ) and (Y, ν) we have
$$\Pi(X \times Y, \mu \otimes \nu) = \Pi(X, \mu) \otimes \Pi(Y, \nu) \tag{B.1.4}$$
(unlike for the other facts, we do not provide the proof of this one and refer to Theorem 18.13 in [45]).

Remark B.1.2. Let us show how Π can be used to analyse stationary random processes. An application of (B.1.3) to stationary countably-valued process $\mathbf{X}$ and generating partition $\mathcal{P} = \{[x] \mid x \in \mathcal{X}\}$ (for any $x \in \mathcal{X}$, $[x] = \{\mathbf{x} \in \mathscr{X} \mid x_0 = x\}$ stands for the corresponding cylinder set) immediately yields a non-trivial result:

$$\mathcal{T}_{past}(\mathbf{X}) \stackrel{\mathbb{P}}{=} \mathcal{T}_{future}(\mathbf{X}) \tag{B.1.5}$$

as soon as $\mathbf{H}(X_0) < \infty$. Furthermore, by (B.1.4), (B.1.5),

$$\mathcal{T}_{past}((\mathbf{X}, \mathbf{Y})) = \mathcal{T}_{past}(\mathbf{X}) \otimes \mathcal{T}_{past}(\mathbf{Y})$$

as soon as $\mathbf{X}$ and $\mathbf{Y}$ are stationary countably valued processes such that $\mathbf{H}(X_0) + \mathbf{H}(Y_0) < \infty$ and $\mathbf{X} \amalg \mathbf{Y}$.

Proposition B.1.3. *The family $\Pi = \Pi(T)$ is a T-invariant μ-countably generated σ-algebra.*

Proof. The T-invariance follows from $T^{-1}\mathcal{P}_A = \mathcal{P}_{T^{-1}A}$ and

$$\mathbf{H}(T, \mathcal{P}_{T^{-1}A}) = \lim_{n\to\infty} \frac{1}{n}\mathbf{H}\left((\mathcal{P}_{T^{-1}A})_0^{n-1}\right) = \lim_{n\to\infty} \frac{1}{n}\mathbf{H}\left((\mathcal{P}_A)_1^n\right) = \mathbf{H}(T, \mathcal{P}_A),$$

where we have used

$$\mathbf{H}\left((\mathcal{P}_A)_0^n\right) - \mathbf{H}(\mathcal{P}_A) \leq \mathbf{H}\left((\mathcal{P}_A)_1^n\right) \leq \mathbf{H}\left((\mathcal{P}_A)_0^n\right).$$

In order to see that Π is μ-countably generated, recall that $\mathbf{H}(T, \mathcal{A}) - \mathbf{H}(T, \mathcal{B}) \leq \mathbf{H}(\mathcal{A} \mid \mathcal{B})$. It follows that $\mathbf{H}(T, \cdot)$ is continuous (with respect to μ-symetric difference metric) on the space of 2-partitions. This space can be treated as a closed subspace of $L_1(\mu)$. Thus, the result follows from the separability of $L_1(\mu)$.

At last, it is clear that if $A \in \Pi$ then $A^c \in \Pi$ and $\mathcal{X} \in \Pi$. Let $A_i \in \Pi$ and $A = \bigcup_i A_i$. Note that $\mathcal{P}_A \in \mathcal{A} = \bigvee_i \mathcal{P}_{A_i}$ and hence $\mathbf{H}(T, \mathcal{P}_A) \leq \mathbf{H}(T, \mathcal{A}) \leq \sum_i \mathbf{H}(T, \mathcal{P}_{A_i}) = 0$. ■

Proposition B.1.4. *For any $k \in \mathbb{Z}$, $k \neq 0$, $\Pi(T^k) = \Pi(T)$.*

Proof. It is a consequence of the fact that for arbitrary finite partition $\mathcal{A}$, $\mathbf{H}\left(T^k, \mathcal{A}\right) = k\mathbf{H}(T, \mathcal{A})$ for $k \geq 1$ (if T is invertible then $\mathbf{H}\left(T^k, \mathcal{A}\right) = |k|\mathbf{H}(T, \mathcal{A})$ for $k \neq 0$). ■

Remark B.1.5. Slightly informally, given a countable partition $\mathcal{A}$ and a σ-field $\mathcal{G}$ we write $\mathcal{A} \in \mathcal{G}$ if every element of $\mathcal{A}$ is $\mathcal{G}$-measurable (this notation is consistent with one used for the random variables).

Proposition B.1.6. *Let T be invertible. Then for any countable partition $\mathcal{A}$ satisfying $\mathbf{H}(\mathcal{A}) < \infty$, $\mathcal{A} \in \Pi$ iff $\mathbf{H}(T, \mathcal{A}) = 0$. In particular, if $\mathcal{A} \in \Pi$ then $\mathcal{A} \in \mathcal{A}_{-\infty}^k$ for all $k \in \mathbb{Z}$.*

Proof. Let $\mathcal{A} = \{A_1, A_2, \ldots\}$. If $\mathcal{A} \in \Pi$ then $\mathbf{H}(T, \mathcal{A}) \leq \sum_i \mathbf{H}(T, \mathcal{P}_{A_i}) = 0$. Conversely, if $\mathbf{H}(T, \mathcal{A}) = 0$ then $\mathbf{H}(T, \mathcal{P}_{A_i}) \leq \mathbf{H}(T, \mathcal{A}) = 0$ and thus $A_i \in \Pi$ for all i.

In order to get $\mathcal{A} \in \mathcal{A}_{-\infty}^k$ for any $k \leq -1$, we proceed inductively (the proof for $k \geq 1$ is analogous). Clearly, $\mathcal{A} \in \mathcal{A}_{-\infty}^{-1}$. On the other hand, $\mathcal{A} \in \mathcal{A}_{-\infty}^{-1}$ implies $T\mathcal{A} \in \mathcal{A}_{-\infty}^{-2}$. Thus, $\mathcal{A} \in \mathcal{A}_{-\infty}^{-2} \vee T\mathcal{A} = \mathcal{A}_{-\infty}^{-2}$. ■

Proposition B.1.7. *If T is invertible then*

$$\Pi(T) = \bigvee_{\mathcal{A}\in\mathcal{P}_{fin}} \mathcal{T}_{past}(T, \mathcal{A}). \tag{B.1.6}$$

Proof. We already know that if $\mathcal{A} \in \Pi$ then $\mathcal{A} \in \mathcal{A}_{-\infty}^{k}$ for all $k \in \mathbb{Z}$ (Proposition B.1.6). In particular, $\mathcal{A} \in \mathcal{T}_{past}(T, \mathcal{A})$.

Conversely, if $\mathcal{B} \in \mathcal{T}_{past} = \mathcal{T}_{past}(T, \mathcal{A})$ then $\mathcal{B} \subset \mathcal{B}_{-\infty}^{\infty} \in \mathcal{T}_{past} \subset \mathcal{A}_{-\infty}^{-1}$. Therefore, on the one hand

$$\mathbf{H}\left(\mathcal{A}, \mathcal{B} \mid \mathcal{A}_{-\infty}^{-1}\right) = \mathbf{H}\left(\mathcal{B} \mid \mathcal{A}_{-\infty}^{-1}\right) + \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}\right) = \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}\right)$$

and on the other

$$\begin{aligned} \mathbf{H}\left(\mathcal{A}, \mathcal{B} \mid \mathcal{A}_{-\infty}^{-1}\right) &= \mathbf{H}\left(\mathcal{A}, \mathcal{B} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}_{-\infty}^{-1}\right) = \mathbf{H}\left(T, \mathcal{A} \vee \mathcal{B}\right) = \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}_{-\infty}^{\infty}\right) + \mathbf{H}\left(\mathcal{B} \mid \mathcal{B}_{-\infty}^{-1}\right) \\ &= \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}\right) + \mathbf{H}\left(\mathcal{B} \mid \mathcal{B}_{-\infty}^{-1}\right), \end{aligned}$$

which, combined together, give $\mathbf{H}\left(\mathcal{B} \mid \mathcal{B}_{-\infty}^{-1}\right) = 0$ or, equivalently, $\mathcal{B} \in \Pi$. ∎

Proposition B.1.8. *Let T be invertible. If $\mathcal{A}$ is a countable generating partition for T such that $\mathbf{H}(\mathcal{A}) < \infty$ then*

$$\mathcal{T}_{past}(T, \mathcal{A}) = \mathcal{T}_{future}(T, \mathcal{A}) = \Pi(T). \tag{B.1.7}$$

Proof. Since $\Pi(T) = \Pi(T^{-1})$, it is enough to show that for example $\mathcal{T}_{past} = \mathcal{T}_{past}(T, \mathcal{A}) = \Pi(T)$.

Firstly, we show $\mathcal{T}_{past} \subset \Pi(T)$. Take $\mathcal{B} \in \mathcal{T}_{past}(T, \mathcal{A})$. Using the fact that $\mathcal{A}$ is generating we get

$$\mathbf{H}(T, \mathcal{A}) = \mathbf{H}(T, \mathcal{A} \vee \mathcal{B}) = \mathbf{H}\left(T^{-1}, \mathcal{A} \vee \mathcal{B}\right) = \mathbf{H}\left(T^{-1}, \mathcal{B}\right) + \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}_{-\infty}^{\infty}\right).$$

Since for any $k \in \mathbb{Z}$, $T^{k}\mathcal{B} \in \mathcal{T}_{past}(T, \mathcal{A})$ and $\mathcal{T}_{past} \subset \mathcal{A}_{-\infty}^{-1}$,

$$\mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}_{-\infty}^{\infty}\right) = \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}\right) = \mathbf{H}(T, \mathcal{A})$$

which results in $\mathbf{H}(T, \mathcal{A}) = \mathbf{H}\left(T^{-1}, \mathcal{B}\right) + \mathbf{H}(T, \mathcal{A})$. Thus, $\mathbf{H}(T, \mathcal{B}) = 0$.

Conversely, let $\mathcal{B} \in \Pi$. Then

$$\mathbf{H}(T, \mathcal{A}) = \mathbf{H}(T, \mathcal{A} \vee \mathcal{B}) = \mathbf{H}\left(T^{-1}, \mathcal{A} \vee \mathcal{B}\right) = \mathbf{H}\left(T^{-1}, \mathcal{B}\right) + \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}_{-\infty}^{\infty}\right).$$

However,

$$\mathbf{H}\left(T^{-1}, \mathcal{B}\right) = \mathbf{H}(T, \mathcal{B}) = 0, \qquad \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}_{-\infty}^{\infty}\right) \leq \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}\right)$$

which for any $k \in \mathbb{Z}$ gives

$$\mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}\right) = \mathbf{H}(T, \mathcal{A}) = \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}_{-\infty}^{\infty}\right) \leq \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, T^{k+1}\mathcal{B}\right) \leq \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}\right). \tag{B.1.8}$$

Thus, all inequalities must be equalities. Now, we will show that (B.1.8) implies

$$\mathbf{H}\left(\mathcal{B} \mid \mathcal{A}_{-\infty}^{k}\right) = \mathbf{H}\left(\mathcal{B} \mid \mathcal{A}_{-\infty}^{k+1}\right), \qquad \forall_{k \in \mathbb{Z}}. \tag{B.1.9}$$

Before we prove (B.1.9) let us present how (B.1.9) concludes the proof. Since $\mathcal{A}$ is generating, taking $k \to \infty$ in (B.1.9) gives that for $k \in \mathbb{Z}$, $\mathbf{H}\left(\mathcal{B} \mid \mathcal{A}_{-\infty}^{k}\right) = 0$. Now taking $k \to -\infty$ results in $\mathbf{H}\left(\mathcal{B} \mid \mathcal{T}_{past}(T, \mathcal{A})\right) = 0$, that is $\mathcal{B} \in \mathcal{T}_{past}(T, \mathcal{A})$.

Hence it remains to show (B.1.9). Since $\mathcal{A}_{-\infty}^{\infty}$ is countably generated, for any $k \in \mathbb{Z}$

$$\mathbf{H}\left(\mathcal{B}, T^{-k-1}\mathcal{A} \mid \mathcal{A}_{-\infty}^{k}\right) = \mathbf{H}\left(\mathcal{B} \mid \mathcal{A}_{-\infty}^{k}\right) + \mathbf{H}\left(T^{-k-1}\mathcal{A} \mid \mathcal{B}, \mathcal{A}_{-\infty}^{k}\right)$$

and

$$\mathbf{H}\left(\mathcal{B}, T^{-k-1}\mathcal{A} \mid \mathcal{A}_{-\infty}^{k}\right) = \mathbf{H}\left(T^{-k-1}\mathcal{A} \mid \mathcal{A}_{-\infty}^{k}\right) + \mathbf{H}\left(\mathcal{B} \mid \mathcal{A}_{-\infty}^{k+1}\right).$$

Therefore (B.1.9) holds iff

$$\mathbf{H}\left(T^{-k-1}\mathcal{A} \mid \mathcal{B}, \mathcal{A}_{-\infty}^{k}\right) = \mathbf{H}\left(T^{-k-1}\mathcal{A} \mid \mathcal{A}_{-\infty}^{k}\right), \qquad \forall_{k \in \mathbb{Z}}.$$

Note that $\mathbf{H}\left(T^{-k-1}\mathcal{A} \mid \mathcal{B}, \mathcal{A}_{-\infty}^{k}\right) = \mathbf{H}\left(\mathcal{A} \mid T^{k+1}\mathcal{B}, \mathcal{A}_{-\infty}^{-1}\right)$ and $\mathbf{H}\left(T^{-k-1}\mathcal{A} \mid \mathcal{A}_{-\infty}^{k}\right) = \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}\right)$. Now, it is enough to notice that (B.1.8) implies

$$\mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, \mathcal{B}_{-\infty}^{\infty}\right) = \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}, T^{k+1}\mathcal{B}\right) = \mathbf{H}\left(\mathcal{A} \mid \mathcal{A}_{-\infty}^{-1}\right), \qquad \forall_{k \in \mathbb{Z}}.$$

∎

B.2 General relations

In this section we return to our standard setting, that is we consider a stationary finitely valued process $\mathbf{X}$ with the corresponding subshift $(\mathscr{X}, S)$.

Clearly, we always have $\mathcal{T}_{past} = \mathcal{T}_{past}(\mathbf{X}), \mathcal{T}_{future} = \mathcal{T}_{future}(\mathbf{X}) \subset \mathcal{T}_{double} = \mathcal{T}_{double}(\mathbf{X})$. Are there any relations between the σ-fields of shift invariant sets $\mathcal{I} = \{\mathbf{A} \subset \mathscr{X} \mid S\mathbf{A} = \mathbf{A}\}$ and these tail σ-algebras? It turns out that if $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ is unilateral then we always have $\mathcal{I} \subset \mathcal{T}_{future}(\mathbf{X})$. In the bilateral case, things get a little more complicated but one can still show that $\mathcal{I} \subset \mathcal{T}_{future}(\mathbf{X})$ in the sense of measure algebras. These facts have a nice immediate corollary. If one of the tail σ-fields is trivial then so is $\mathcal{I}$ and thus $\mathbf{X}$ is ergodic. In fact, with a little more effort, one can show that in case of trivial tail σ-field, $\mathbf{X}$ is mixing (i.e. the corresponding dynamical system is mixing in the ergodic setting). The converse fact is not true in general, that is the fact that mixing (in the ergodic theoretic sense) implies the triviality of tail σ-algebras. The following natural questions arise. What kind of mixing ensures that $\mathcal{T}_{past}, \mathcal{T}_{future}$ or even $\mathcal{T}_{double}$ are trivial? What are condition under which $\mathcal{T}_{past} \overset{\mathbb{P}}{=} \mathcal{T}_{future} \overset{\mathbb{P}}{=} \mathcal{T}_{double}$ in the sense of measure algebras? When $\mathcal{T}_{past}, \mathcal{T}_{future} \subset \mathcal{T}_{double}$ is strict?

Firstly, let us note that there is a "mixing" criterion which is equivalent to the triviality of $\mathcal{T}_{past}$ or $\mathcal{T}_{future}$.

Lemma B.2.1 (When "past" or "future" tail is trivial)**.** *The past tail sigma algebra is trivial iff for all $B \in \mathcal{F}$,*

$$\lim_{n\to-\infty} \sup_{A_n\in\mathcal{F}^n} \left|\mathbb{P}\left(A_n \cap B\right) - \mathbb{P}\left(A_n\right)\mathbb{P}\left(B\right)\right| = 0, \tag{B.2.1}$$

where $\mathcal{F}^n = \sigma\left(X_{(-\infty,n]}\right)$.

Proof. Suppose that past sigma tail is trivial. Let us denote $\mathbb{1}_{A_n} - \mathbb{P}\left(A_n\right)$ and $\mathbb{1}_B - \mathbb{P}\left(B\right)$ by X_n and Y respectively. Then

$$\left|\mathbb{P}\left(A_n \cap B\right) - \mathbb{P}\left(A_n\right)\mathbb{P}\left(B\right)\right| = \mathbb{E}X_nY = \mathbb{E}X_n\mathbb{E}\left(Y|\mathcal{F}^n\right) \leq \mathbb{E}\left|\mathbb{E}\left(Y|\mathcal{F}^n\right)\right| \xrightarrow{n\to-\infty} \mathbb{E}\left|\mathbb{E}\left(Y|\mathcal{T}_p\right)\right| = 0.$$

Suppose that past sigma tail is not trivial that is $B \in \mathcal{T}_p$, $\mathbb{P}\left(B\right) \in (0,1)$. But then taking $A = B$ gives

$$\sup_{A\in\mathcal{F}^n} \left|\mathbb{P}\left(A, B\right) - \mathbb{P}\left(A\right)\mathbb{P}\left(B\right)\right| \geq \left|\mathbb{P}\left(B\right) - \mathbb{P}\left(B\right)^2\right| > 0.$$

∎

Remark B.2.2. In the language of ergodic theory, a process which has trivial "single" tail σ-algebra $\mathcal{T}_{past}$ (equivalently satisfy (B.2.1)) is called ***K-mixing***.

In the next section we present some examples illustrating the complicated relations which can arise between tail σ-fields and other natural σ-algebras associated with processes.

B.3 Examples

B.3.1 Exchangeable processes

Recall that $\mathbf{X} = (X_i)_{i\in T}$ is ***exchangeable*** if for any distinct $\{i_0, \ldots, i_n\} \subset T$ we have $X_{\{i_0,\ldots,i_n\}} \sim X_{[0,n]}$. By a celebrated result of de Finetti [39] (cf. also [50]), this condition is equivalent to $\mathbf{X}$ being a convex combination of i.i.d. processes. In other words, $\mathbf{X}$ is exchangeable iff there exists a random variable Θ such that, conditionally on Θ, $\mathbf{X}$ is an i.i.d. process. Furthermore, if we assume that $\mathcal{X}$ is Polish space then the following fact holds: if $\mathcal{H}$ is a σ-algebra conditionally on which X_i are i.i.d then essentially $\mathcal{H} = \sigma(\Theta)$ (in sense of measure algebras). For this fact see the only theorem in [81]. One can say more about tail-σ fields. Olshen in [82] showed that if $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ is exchangeable then

$$\mathcal{I} = \mathcal{E} = \mathcal{T}_{double} = \mathcal{T}_{future} = \mathcal{T}_{past}, \tag{B.3.1}$$

(as measure-algebras), where $\mathcal{E}$ denotes the σ-algebra of finite permutation invariant sets. If $\mathbf{X}$ is unilateral then one still has $\mathcal{I} = \mathcal{E} = \mathcal{T}_{future}$.

B.3.2 Markov chains

Let $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ be a finitely-valued Markov chain, $X_i \in \mathcal{X}$. It is well-known (see [37], Chapter XV, Section 6, Theorem 3, page 392) that we can uniquely decompose the state space $\mathcal{X}$ into disjoint union

$$\mathcal{X} = C \sqcup D_1 \sqcup D_2 \sqcup \cdots \sqcup D_k, \tag{B.3.2}$$

where C is the set of transient states and D_i are closed sets. If $\mathbf{X}$ starts in D_j (i.e. $X_0 \in D_j$) then it remains in D_j forever. If $X_0 \in C$ then $\mathbf{X}$ stays in C for finite time and jumps to some D_j (and never leaves D_j afterwards). Moreover (see [37], Chapter XV, Section 7, Criterion, page 395), if π is a stationary measure then necessarily $\pi(C) = 0$.

Remark B.3.1. In this part, for brevity's sake, sometimes we shorten $\mathcal{T}_{double}(\mathbf{X}), \mathcal{T}_{past}(\mathbf{X}), \mathcal{T}_{future}(\mathbf{X})$ to $\mathcal{T}_d(\mathbf{X}), \mathcal{T}_p(\mathbf{X}), \mathcal{T}_f(\mathbf{X})$, respectively.

Now, suppose that a bilateral, finitely-valued Markov chain $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ is stationary (thus, $C = \varnothing$ in (B.3.2)). In that case we will show that $\mathcal{T}_{double}(\mathbf{X}) = \mathcal{T}_{past}(\mathbf{X}) = \mathcal{T}_{future}(\mathbf{X})$.

Fix $1 \le j \le k$ and let $\mathbf{X}_{D_j}$ stand for $\mathbf{X}$ conditioned on $X_0 \in D_j$. By the definition of D_j, process $\mathbf{X}_{D_j}$ is an irreducible (equivalently, ergodic), stationary Markov chain. Now, let p_j be the period of $\mathbf{X}_{D_j}$. Then D_j can be decomposed into p_j disjoint sets (see [23], Chapter 1, Section 3, Theorem 4)

$$D_j = D_{j,0} \sqcup \cdots \sqcup D_{j,p_j-1}$$

such that $\mathbb{P}\left(X_1 \in D_{j,(\ell+1) \bmod p_j} \mid X_0 \in D_{j,\ell}\right) = 1$. Using Corollary 2 from [16], we get that

$$\mathcal{T}_d\left(\mathbf{X}_{D_j}\right) = \mathcal{T}_p\left(\mathbf{X}_{D_j}\right) = \mathcal{T}_f\left(\mathbf{X}_{D_j}\right) = \sigma\left\{\{X_0 \in D_{j,0}\}, \{X_0 \in D_{j,1}\}, \ldots, \{X_0 \in D_{j,p_j-1}\}\right\}.$$

Note that Corollary 2 from [16] is stated only for $\mathcal{T}_f$ but a perusal of the proofs of Theorem 1 and Corollaries 1 and 2 therein gives the same result for $\mathcal{T}_d$. Thus, $\mathbf{X}$, conditionally on $X_0 \in D_{j,l}$, has trivial tail σ-algebras. This immediately leads to

$$\mathcal{T}_d(\mathbf{X}) = \mathcal{T}_p(\mathbf{X}) = \mathcal{T}_f(\mathbf{X}) = \sigma\left\{\{X_0 \in D_{j,\ell}\} \mid 1 \le j \le k, 0 \le \ell \le p_j\right\}. \tag{B.3.3}$$

Indeed, if for example $A \in \mathcal{T}_d(\mathbf{X})$ then, for all j, ℓ, $\mathbb{P}(A \mid X_0 \in D_{j,\ell}) \in \{0,1\}$ which yields (B.3.3).

B.3.3 α-mixing processes

Recall that for any σ-fields $\mathcal{A}$ and $\mathcal{B}$, we define ***α-mixing coefficient*** as

$$\alpha(\mathcal{A}, \mathcal{B}, \mathbb{P}) = 2 \sup_{A\in\mathcal{A},\, B\in\mathcal{B}} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|.$$

More intuitively, one can show that $\alpha(\mathcal{A}, \mathcal{B}, \mathbb{P}) = \sup_{\|X\|_\infty, \|Y\|_\infty \le 1} |\mathrm{Cov}(X,Y)|$ (see (1.12a) in [88]). Moreover, for any process $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ we define its n'th α mixing coefficient as

$$\alpha_n = \sup_{k\in\mathbb{Z}} \alpha\left(\sigma\left(X_{(-\infty,k]}\right), \sigma\left(X_{[k+n,\infty)}\right)\right).$$

If $\mathbf{X}$ is stationary, this definition simplifies to $\alpha_n = \alpha\left(\sigma\left(X_{(-\infty,0]}\right), \sigma\left(X_{[n,\infty)}\right)\right)$. We say that $\mathbf{X}$ is ***α-mixing* (or *strongly mixing*)** if $\alpha_n \xrightarrow{n\to\infty} 0$. One can show that if $\mathbf{X}$ (not necessarily stationary) is α-mixing then the single-sided tail σ-algebras $\mathcal{T}_{future}$ and $\mathcal{T}_{past}$ are trivial. However, even if $\mathbf{X}$ is strongly mixing, $\mathcal{T}_{double}$ can be non-trivial. Even more, in [19] one can find a construction of a strictly stationary, finite-state, strongly mixing, bilaterally deterministic $\mathbf{X}$.

B.3.4 β-mixing processes (weak Bernoulli processes)

For any σ-fields $\mathcal{A}$ and $\mathcal{B}$, we define ***β-mixing coefficient*** as

$$\beta(\mathcal{A}, \mathcal{B}, \mathbb{P}) = \frac{1}{2} \sup_{\mathcal{A}_{fin}\subset\mathcal{A}, \mathcal{B}_{fin}\subset\mathcal{B}} \sum_{A\in\mathcal{A}_{fin}, B\in\mathcal{B}_{fin}} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|, \tag{B.3.4}$$

where $\mathcal{A}_{fin}$ and $\mathcal{B}_{fin}$ stand for finite partitions. One can show that $\alpha(\mathcal{A}, \mathcal{B}, \mathbb{P}) \le 2\beta(\mathcal{A}, \mathcal{B}, \mathbb{P})$. Moreover, for any process $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ we define its n'th β mixing coefficient as

$$\beta_n = \sup_{k\in\mathbb{Z}} \beta\left(\sigma\left(X_{(-\infty,k]}\right), \sigma\left(X_{[k+n,\infty)}\right)\right).$$

If $\mathbf{X}$ is stationary, this definitions simplifies to $\beta_n = \beta\left(\sigma\left(X_{(-\infty,0]}\right), \sigma\left(X_{[n,\infty)}\right)\right)$. We say that $\mathbf{X}$ is ***β-mixing* (or *absolutely regular*)** if $\beta_n \xrightarrow{n\to\infty} 0$. One can show that for finitely-valued stationary processes $\mathbf{X}$, $\mathbf{X}$ is absolutely mixing iff $\mathbf{X}$ is weak Bernoulli (see [18], equation (2.4) and surroundings). Furthermore, one can show that if $\mathbf{X}$ (not necessarily stationary) is β-mixing then necessarily $\mathcal{T}_{double}$ is trivial (and thus so are the one-sided tail σ-fields). In fact, Berbee in [9] showed more. He introduced a notion of period for general random process $\mathbf{X}$ and showed the following theorem.

Theorem B.3.2. *Let $\mathbf{X}$ be a stationary, ergodic process. If $\beta_n < 1$ for some n then $\mathbf{X}$ has a finite period p and*

$$\beta_n \to 1 - \frac{1}{p}. \tag{B.3.5}$$

Moreover, $\mathcal{T}_{double} = \mathcal{T}_{past} = \mathcal{T}_{future} = \mathcal{I}_p$ (in the sense of measure algebras), where $\mathcal{I}_p$ is the S^p-invariant σ-field, is partitioned by $\{S^i\mathbf{X} \in E\}$, for $i \in [0,p)$, into atoms that are S^p-invariant. Furthermore, for each $1 \le i \le p$, process $\mathbf{X}$ conditioned on $S^i\mathbf{X} \in E$ is absolutely regular.

Appendix C

Besicovitch and Prokhorov metrics

This summary of basic facts concerning the Besicovitch pseudo-distance is based on [63]. Unlike in some other parts of this thesis, we assume here that $(\mathcal{X}, d)$ is a compact (not necessarily finite) metric space. Every such metric d determines the ***Besicovitch pseudo metric***

$$d_B \colon \mathcal{X}^{\mathbb{N}} \times \mathcal{X}^{\mathbb{N}} \to \mathbb{R}_+, \qquad d_B(\mathbf{x}, \mathbf{y}) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} d(x_i, y_i).$$

Moreover, recall that the ***upper density*** of a set $N \subset \mathbb{N}$ is given by

$$\overline{d}(N) = \limsup_{n\to\infty} \frac{1}{n} \left| N \cap [0, n-1] \right|.$$

The following example is connected with $\mathscr{B}$-free systems and the convergence of periodic approximations (cf. Section 4.2.3).

Example C.0.1. Assume that $\mathcal{X} = \{0, 1\}$ is equipped with the Hamming distance $d(x, y) = \mathbb{1}_{x \neq y}$. Let $\mathbf{x}^{(n)} \in \mathcal{X}^{\mathbb{Z}}$ be a sequence of binary sequences such that $\mathbf{x}^{(n)} \leq \mathbf{x}^{(n-1)}$ (coordinatewise) for all n and $d^{(n)} = \limsup\limits_{k\to\infty} \frac{1}{k} \#_1 \left(x^{(n)}_{[0,k-1]} \right) \xrightarrow{n\to\infty} d = \liminf\limits_{k\to\infty} \frac{1}{k} \#_1 \left(x_{[0,k-1]} \right)$. Then

$$d_B(\mathbf{x}^{(n)}, \mathbf{x}) \to 0.$$

Indeed, it is enough to notice that due to the assumptions

$$\frac{1}{k} \sum_{i=0}^{k-1} \mathbb{1}_{x_i^{(n)} \neq x_i} = \frac{1}{k} \sum_{i=0}^{k-1} \mathbb{1}_{x_i^{(n)} = 1} - \frac{1}{k} \sum_{i=0}^{k-1} \mathbb{1}_{x_i = 1}.$$

Therefore,

$$d_B(\mathbf{x}^{(n)}, \mathbf{x}) \leq d^{(n)} - d \to 0.$$

Let us now introduce a metric strongly connected with d_B, namely,

$$d_P \colon \mathcal{X}^{\mathbb{N}} \times \mathcal{X}^{\mathbb{N}} \to \mathbb{R}_+, \qquad d_P(\mathbf{x}, \mathbf{y}) = \inf \left\{ \delta > 0 \mid \overline{d} \left(\{ i \in \mathbb{N} \mid d(x_i, y_i) \geq \delta \} \right) \leq \delta \right\}. \tag{C.0.1}$$

Remark C.0.2. We use index P to express the resemblance to Prokhorov(-Lévy) metric on measures.

We have the following relations between d_B and d_P.

Lemma C.0.3. *Let $(\mathcal{X}, d)$ be a compact space. Then*

$$d_P^2(\mathbf{x}, \mathbf{y}) \leq d_B(\mathbf{x}, \mathbf{y}) \leq d_P(\mathbf{x}, \mathbf{y}) \left[1 + \|\mathbf{x} - \mathbf{y}\|_\infty \right] \leq d_P(\mathbf{x}, \mathbf{y}) \left[1 + diam(\mathcal{X}) \right],$$

where $\|\mathbf{x} - \mathbf{y}\|_\infty = \sup_{i\in\mathbb{N}} d(x_i, y_i)$ stands for the supremum "norm" and $diam(\mathcal{X}) = \sup_{x,y} d(x, y)$ for the diameter of $\mathcal{X}$.

Proof. Let $d_P(\mathbf{x}, \mathbf{y}) < \delta$. Then, by the very definition of d_P,

$$\overline{d} \left(C_\delta^c \right) = \limsup_{n\to\infty} \frac{1}{n} \left| [0, n] \cap C_\delta^c \right| \leq \delta, \qquad C_\delta = \{ i \in \mathbb{N} \mid d(x_i, y_i) < \delta \},$$

where $C_\delta^c = \mathbb{N} \backslash C_\delta$. Thus,

$$d_B(\mathbf{x}, \mathbf{y}) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} d(x_i, y_i) \leq \limsup_{n\to\infty} \frac{1}{n} \sum_{i\in[0,n]\cap C_\delta} d(x_i, y_i) + \|\mathbf{x}-\mathbf{y}\|_\infty \overline{d}\,(C_\delta^c) \leq \delta\,[1+\|\mathbf{x}-\mathbf{y}\|_\infty]\,.$$

Conversely, if $d_B(\mathbf{x}, \mathbf{y}) < \delta$ then for sufficiently big n, $\frac{1}{n}\sum_{i=0}^{n-1} d(x_i, y_i) < \delta$. Thus,

$$\left|\left\{0 \leq i \leq n \mid d(x_i, y_i) \geq \sqrt{\delta}\right\}\right| < n\sqrt{\delta},$$

which implies that $\overline{d}\left(\left\{i \in \mathbb{N} \mid d(x_i, y_i) \geq \sqrt{\delta}\right\}\right) \leq \sqrt{\delta}$ and $d_P(\mathbf{x}, \mathbf{y}) \leq \sqrt{\delta}$. ■

Lemma C.0.3 enables us to obtain a continuity property for limits of **empirical measures**. More precisely, recall that, given $\mathbf{x} \in \mathcal{X}^{\mathbb{N}}$, the ***family of empirical measures*** is given by

$$\delta_{\mathbf{x},n} = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{S^i\mathbf{x}}, \tag{C.0.2}$$

where $n \in \mathbb{N}$. Moreover, we define the ***set of limits of empirical measures***:

$$\mathcal{M}(\mathbf{x}) = \left\{\mu \in \mathcal{M}_{\mathfrak{X}} \mid \exists_{(n_k)} \quad \delta_{\mathbf{x},n_k} = \frac{1}{n_k}\sum_{i=0}^{n_k-1} \delta_{S^i\mathbf{x}} \Rightarrow \mu\right\}. \tag{C.0.3}$$

Now, we would like to say that if $d_B(\mathbf{x}, \mathbf{y})$ is small then $\mathcal{M}(\mathbf{x})$ is close to $\mathcal{M}(\mathbf{y})$. To do so formally, recall that with each metric space $(\mathcal{X}, d)$ we can associate the ***Hausdorff distance*** between sets given by

$$d_H(A, B) = \inf\left\{\delta > 0 \mid A \subset B^\delta,\ B \subset A^\delta\right\}, \qquad A^\delta = \{x \in \mathcal{X} \mid d(x, A) < \delta\}\,. \tag{C.0.4}$$

Moreover, the ***Prokhorov-Lévy metric on the space of probability Borel measures on*** $\mathcal{X}$ is defined as

$$d_P(\mu, \nu) = \inf\left\{\delta > 0 \mid \forall_{A-Borel} \quad \mu(A) \leq \nu(A^\delta) + \delta,\ \nu(A) \leq \mu(A^\delta) + \delta\right\}. \tag{C.0.5}$$

It is well-known that if $(\mathcal{X}, d)$ is separable then d_P is equivalent to the weak convergence topology (for general metric space $d_P(\mu_n, \mu) \to 0$ implies $\mu_n \Rightarrow \mu$) (see [15], Section "The Prohorov metric").

Beware of the difference between $d_P(\mu, \nu)$ and $d_P(\mathbf{x}, \mathbf{y})$. The latter is given by (C.0.1). Now, we are ready to state the continuity property.

Theorem C.0.4. *Let $(\mathcal{X}, d)$ be a compact metric space. Then for any $\mu \in \mathcal{M}(\mathbf{x})$ and $\nu \in \mathcal{M}(\mathbf{y})$ such that μ and ν are generated on some common subsequence (n_k),*

$$d_P(\mu, \nu) \leq d_P(\mathbf{x}, \mathbf{y}) \leq \sqrt{d_B(\mathbf{x}, \mathbf{y})}. \tag{C.0.6}$$

In particular,

$$d_{PH}(\mathcal{M}(\mathbf{x}), \mathcal{M}(\mathbf{y})) \leq d_P(\mathbf{x}, \mathbf{y}), \tag{C.0.7}$$

where d_{PH} stands for the Hausdorff distance (cf. (C.0.4)) induced by the Prokhorov metric (C.0.5).

Proof. The second inequality in (C.0.6) is a content of Lemma C.0.3. To see the first inequality in (C.0.6), assume that $d_P(\mathbf{x}, \mathbf{y}) < \delta$ for some $\delta > 0$. Then, for all sufficiently large n we have

$$\frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{d(x_i,y_i)\geq\delta} < \delta.$$

Hence, for such n's and any Borel set $B \subset \mathcal{X}$

$$\delta_{\mathbf{x},n}(B) = \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{x_i\in B} \leq \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{1}_{y_i\in B^\delta} + \delta = \delta_{\mathbf{y},n}(B^\delta) + \delta.$$

Exchanging the role of $\mathbf{x}$ and $\mathbf{y}$ we obtain (after taking $\delta \to d_P(\mathbf{x}, \mathbf{y})$)

$$d_P(\delta_{\mathbf{x},n}, \delta_{\mathbf{y},n}) \leq d_P(\mathbf{x}, \mathbf{y}) \tag{C.0.8}$$

for all $n \geq n(\mathbf{x}, \mathbf{y})$. Since $\mu \in \mathcal{M}(\mathbf{x})$ and $\nu \in \mathcal{M}(\mathbf{y})$ are generated on a common subsequence and the Prokhorov metric is equivalent to the weak convergence, (C.0.8) yields (C.0.6). Now, (C.0.7) immediately follows from (C.0.8). ■

Combining Example C.0.1 with Theorem C.0.4, we immediately get the following corollary.

Corollary C.0.5. *Assume that $\mathcal{X} = \{0,1\}$ is equipped with the Hamming distance $d(x,y) = \mathbb{1}_{x \neq y}$. Let $\mathbf{x}^{(n)} \in \mathcal{X}^{\mathbb{Z}}$ be a sequence of binary sequences such that $\mathbf{x}^{(n)} \geq \mathbf{x}^{(n-1)}$ for all $n \in \mathbb{N}$ and $\mathbf{x}^{(n)} \searrow \mathbf{x}^{(\infty)}$ coordinatewise. If $\mu_n \in \mathcal{M}(\mathbf{x}^{(n)})$ for all $n \in \mathbb{N} \cup \{\infty\}$ are generated along the same subsequence and $\lim_{k\to\infty} \frac{1}{k} \#_1 \left(x^{(n)}_{[0,k-1]} \right) \xrightarrow{n\to\infty} \lim_{k\to\infty} \frac{1}{k} \#_1 \left(x_{[0,k-1]} \right)$ then*

$$\mu_n \Rightarrow \mu_\infty.$$

Appendix D

Orlicz exponential norm

At the beginning recall the definition of the exponential Orlicz quasi-norm. For any random variable X and $\alpha > 0$ we define

$$\|X\|_{\psi_\alpha} = \inf\Big\{c > 0 \mid \mathbb{E}\exp\left(\frac{|X|^\alpha}{c^\alpha}\right) \leq 2\Big\}. \tag{D.0.1}$$

Note that if $\alpha \geq 1$ then $\|\cdot\|_{\psi_\alpha}$ is a norm whereas for $0 < \alpha < 1$, $\|\cdot\|_{\psi_\alpha}$ is only a quasi-norm. More precisely, we have the following version of the triangle inequality (see Lemma 3.7 in [5]).

Lemma D.0.1 (Triangle inequality for $\alpha \leq 1$)**.** *Fix $0 < \alpha \leq 1$. Then for any random variables X, Y we have*

$$\|X+Y\|_{\psi_\alpha} \leq \left(\|X\|_{\psi_\alpha}^\alpha + \|Y\|_{\psi_\alpha}^\alpha\right)^{1/\alpha} \leq 2^{1/\alpha-1}\left(\|X\|_{\psi_\alpha} + \|Y\|_{\psi_\alpha}\right).$$

Now, we present a moment estimation for random variables with bounded exponential moment.

Lemma D.0.2. *If Y is non negative random variable such that $\mathbb{E}\exp(Y) \leq 2$ then for any $\beta > 0$ we have*

$$\mathbb{E}Y^\beta \leq 2\Gamma(\beta+1).$$

Furthermore, if $\beta \in \mathbb{N}$ then one can replace the constant 2 with 1.

Proof. If β is a natural number then the claim follows from Taylor's expansion of $\exp(x)$. The general case is obtained by Markov's inequality, namely

$$\mathbb{E}Y^\beta = \int_0^\infty \mathbb{P}\left(Y^\beta \geq t\right) dt = \int_0^\infty \mathbb{P}\left(e^Y \geq e t^{\frac{1}{\beta}}\right) dt \leq \int_0^\infty 2e^{-t^{\frac{1}{\beta}}} dt = 2\int_0^\infty e^{-s}\beta s^{\beta-1} ds = 2\beta\Gamma(\beta).$$

∎

The next lemma allows us to pass from the ψ_α-norm of a random variable to the norm of its conditional expectation.

Lemma D.0.3 (Orlicz's norm of Conditional Mean Value)**.** *Let $0 < \alpha \leq 1$. Assume that a random variable X satisfies $\|X\|_{\psi_\alpha} < \infty$. Moreover, let $\mathcal{F}$ be some sigma field. Then*

$$\|\mathbb{E}(X|\mathcal{F})\|_{\psi_\alpha} \leq \left(1 + \frac{\log\left(\alpha\exp\left(\frac{1-\alpha}{\alpha}\right)\right)}{\log(2)}\right)^{\frac{1}{\alpha}} \|X\|_{\psi_\alpha} \leq \left(\frac{2}{\alpha}\right)^{\frac{1}{\alpha}} \|X\|_{\psi_\alpha}.$$

Proof. Set $\varphi_\alpha(x) = \exp(x^\alpha)$ for $x \geq 0$ and notice that φ_α is concave on $(0, x_\alpha)$ and convex on (x_α, ∞), where $x_\alpha = \left(\frac{1-\alpha}{\alpha}\right)^{1/\alpha}$. Define Ψ_α to be a smallest convex function greater than or equal to φ_α which is equal to φ_α on (x_α, ∞), that is

$$\Psi_\alpha(x) = \begin{cases} \alpha\exp\left(\dfrac{1-\alpha}{\alpha}\right)(x x_\alpha^{\alpha-1} + 1), & \text{if } 0 \leq x \leq x_\alpha, \\ \varphi_\alpha(x), & \text{if } x \geq x_\alpha. \end{cases}$$

Clearly, Ψ_α is a convex function on $\mathbb{R}_+$ and it is easy to see that $\varphi_\alpha \leq \Psi_\alpha \leq \alpha\exp\left(\frac{1-\alpha}{\alpha}\right)\varphi_\alpha$. Using these properties, Jensen's inequality and the definition of the Orlicz norm, we get

$$\mathbb{E}\varphi_\alpha\left(\frac{|\mathbb{E}(X|\mathcal{F})|}{\|X\|_{\psi_\alpha}}\right) \leq \mathbb{E}\Psi_\alpha\left(\frac{|\mathbb{E}(X|\mathcal{F})|}{\|X\|_{\psi_\alpha}}\right) \leq \mathbb{E}\Psi_\alpha\left(\frac{|X|}{\|X\|_{\psi_\alpha}}\right) \leq 2\alpha\exp\left(\frac{1-\alpha}{\alpha}\right).$$

Put $c_\alpha = \left(1 + \frac{\log\left(\alpha\exp\left(\frac{1-\alpha}{\alpha}\right)\right)}{\log(2)}\right)^{\frac{1}{\alpha}} \geq 1$ and note that due to Jensen's inequality

$$\mathbb{E}\varphi_\alpha\left(\frac{|\mathbb{E}\,(X|\mathcal{F})|}{c_\alpha \|X\|_{\psi_\alpha}}\right) \leq \left(\mathbb{E}\varphi_\alpha\left(\frac{|\mathbb{E}\,(X|\mathcal{F})|}{\|X\|_{\psi_\alpha}}\right)\right)^{\frac{1}{c_\alpha^\alpha}} \leq 2,$$

which completes the proof. ■

Now we give two concentration inequalities which are valid for random variables with finite Orlicz norm. The first one is an easy consequence of the Markov inequality, therefore we omit the proof.

Lemma D.0.4. *For any random variable X with $\|X\|_{\psi_\alpha} < \infty$ and $t > 0$,*

$$\mathbb{P}\left(|X| \geq t\right) \leq 2\exp\left(-\frac{t^\alpha}{\|X\|_{\psi_\alpha}^\alpha}\right).$$

Lemma D.0.5 (Tail inequality for conditional mean value)**.** *Let $0 < \alpha \leq 1$. Assume that a random variable X satisfies $\|X\|_{\psi_\alpha} < \infty$. Moreover, let $\mathcal{F}$ be some sigma field. Then for any $t \geq \left(\frac{2}{\alpha}\right)^{1/\alpha} \|X\|_{\psi_\alpha}$,*

$$\mathbb{P}\left(|\mathbb{E}(X|\mathcal{F})| > t\right) \leq 6\exp\left(-\frac{t^\alpha}{2\|X\|_{\psi_\alpha}^\alpha}\right).$$

Proof. Fix $c > \|X\|_{\psi_\alpha}$ and $t \geq \left(\frac{2}{\alpha}\right)^{1/\alpha} c$. Then in particular we have $\alpha\left(\frac{t}{c}\right)^\alpha \geq 2$. Using the Markov and Jensen inequalities along with $\Gamma(x) \leq x^x/e^{x-1}$ ([69], Thm. 1) and Lemma D.0.2 with $Y = (|X|/c)^\alpha$, $\beta = t^\alpha/c^\alpha$, we get

$$\begin{aligned}
&\mathbb{P}\left(|\mathbb{E}(X|\mathcal{F})| > t\right) \leq \mathbb{P}\left(|\mathbb{E}(X|\mathcal{F})|^{\alpha\frac{t^\alpha}{c^\alpha}} > t^{\alpha\frac{t^\alpha}{c^\alpha}}\right) \leq t^{-\alpha\left(\frac{t}{c}\right)^\alpha}\mathbb{E}\,|\mathbb{E}(X|\mathcal{F})|^{\alpha\left(\frac{t}{c}\right)^\alpha} \leq t^{-\alpha\left(\frac{t}{c}\right)^\alpha}\mathbb{E}\,|X|^{\alpha\left(\frac{t}{c}\right)^\alpha} \\
&= (t/c)^{-\alpha\left(\frac{t}{c}\right)^\alpha}\mathbb{E}\,|X/c|^{\alpha\left(\frac{t}{c}\right)^\alpha} \leq 2e\,(t/c)^\alpha\exp\left(-\,(t/c)^\alpha\right) \leq 2e\exp\left(-(1/2)\,(t/c)^\alpha\right),
\end{aligned}$$

where in the last inequality we used the estimate $xe^{-x} \leq e^{-\frac{x}{2}}$ which is valid for all $x \in \mathbb{R}$. Now, it is enough to take limit $c \to \|X\|_{\psi_\alpha}$ and notice that $2e \leq 6$. ■

Appendix E

Markov-like properties of the split chain

Let $P(\cdot,\cdot)$ be a transition kernel. Recall that the split chain $\overline{\mathbf{X}} = (X_i, Y_i)$ was defined via (for the details and intuition see Section 6.1.11)

$$\mathbb{P}\left(Y_{km} = i, X_{[km+1,(k+1)m]} \in \bigtimes_{i=1}^{m} A_i \mid \mathcal{F}^{\mathbf{X}}_{km}, \mathcal{F}^{\mathbf{Y}}_{km-m}, X_{km} = x\right) = \mathbb{P}\left(Y_0 = i, X_{[1,m]} \in \bigtimes_{i=1}^{m} A_i \mid X_0 = x\right)$$
$$= \int_{A_1} \cdots \int_{A_m} r(x, x_m, i) P(x_{m-1}, dx_m) P(x_{m-2}, dx_{m-1}) \dots P(x, dx_1), \tag{E.0.1}$$

where

$$r(x,y,i) = \begin{cases} \mathbb{1}_{x\in C}\, r(x,y), & \text{if } i = 1, \\ 1 - \mathbb{1}_{x\in C}\, r(x,y), & \text{if } i = 0, \end{cases}, \qquad r(x,y) = \frac{\delta\nu(dy)}{P^m(x,dy)} \tag{E.0.2}$$

and for any process $\mathbf{Z} = (Z_i)_{i\in N}$, $\mathcal{F}^{\mathbf{Z}} = \left(\mathcal{F}^{\mathbf{Z}}_i\right)_{i\in N}$ stands for the ***natural filtration associated with*** $\mathbf{Z}$, that is

$$\mathcal{F}^{\mathbf{Z}}_i = \sigma\left(Z_0^i\right).$$

Moreover, for any $k, i \in \mathbb{N}$ such that $km < i < (k+1)m$ we put

$$Y_i = Y_{km}. \tag{E.0.3}$$

Remark E.0.1. Recall that such definition (E.0.1) of $\overline{\mathbf{X}}$ ensures that the first coordinate $\mathbf{X}$ forms a Markov chain with transition kernel $P(\cdot,\cdot)$. However, for $m > 1$ it may happen that $\overline{\mathbf{X}}$ is not a Markov chain.

In this section we present how such a definition of $\overline{\mathbf{X}}$ implies a Markov-like property of $\overline{\mathbf{X}}$ (see Lemma E.1.3), the Markov property of m-block process

$$\left(\overline{X}_{[im,im+m-1]}\right)_{i\in\mathbb{N}}$$

(see Lemma E.2.1) and then the Markov property of random blocks (recall (6.1.22))

$$\mathbf{\Xi} = (\Xi_i)_{i\geq 0}, \quad \Xi_i = X_{[\tau_{i-1}+m, \tau_i+m-1]},$$

(see Lemma E.3.2). In particular, we justify the formulas we provided in Section 6.1.11.

Let us add that in this section we use extensively the Dynkin $\pi - \lambda$ lemma. For the exact formulation of it we refer to the Lemma 4.10 in [48]. Furthermore, let us warn the reader that this part is very technical and we assume a good knowledge of standard tools and arguments from the probability field. In order to avoid lengthy writings we use an additional notation, namely, for arbitrary integers k, l such that $k \leq l$ and a sequence $\mathbf{x} = (x_i)_{i\in\mathbb{Z}}$,

$$x_k^l = (x_k, x_{k+1}, \dots, x_l)$$

with similar convention for random processes (we do not use this convention for sets to not perplex the reader; A_i^j is just too ambiguous; however, $A_{[i,j]} = \bigtimes_{k=i}^{j} A_k$ is used frequently).

Moreover, for clarity's sake, we omit measurability details, in particular, we tacitly assume that products spaces are equipped with the product σ-fields, similarly, all functions we consider are bounded and measurable (with respect to appropriate underlying σ-field). If need be, one can easily extend all below facts to the case of integrable functions.

E.1 Markov-like property of the split chain

In this section we explain how to generalize (E.0.1) to the arbitary functions.

Lemma E.1.1. *Let $L = \{0,1\}$ be a "level" space and $k \in \mathbb{N}$. For any bounded measurable real function $F : L \times \mathcal{X}^m \to \mathbb{R}$,*

$$\begin{aligned}&\mathbb{E}\left(F(Y_{km}, X_{km+1}^{(k+1)m}) \mid \mathcal{F}_{km}^{\mathbf{X}}, \mathcal{F}_{km-m}^{\mathbf{Y}}\right) = \mathbb{E}\left(F(Y_{km}, X_{km+1}^{(k+1)m}) \mid X_{km}\right)\\ &= \left(\int_{\mathcal{X}}\right)^m \int_L F(y_0, x_1^m) r(X_{km}, x_m, dy_0) P(x_{m-1}, dx_m) P(x_{m-2}, dx_{m-1}) \dots P(X_{km}, dx_1),\end{aligned} \tag{E.1.1}$$

where r is given by (E.0.2).

Remark E.1.2. Note that due to the definition of function r (recall (E.0.2)), $i \to r(x, y, i)$ can be treated as a probability measure.

Proof. We use the standard argument of approximation. Firstly, notice that for functions F of the form

$$\mathbb{1}_{y_0 = i,\ x_1 \in A_1,\ \dots,\ x_m \in A_m}$$

Lemma E.1.1 reduces to the very definition given in (E.0.1). It follows that Lemma E.1.1 is valid for all functions of the form $(B \subset L)$

$$\mathbb{1}_{y_0 \in B,\ x_1 \in A_1,\ \dots,\ x_m \in A_m}. \tag{E.1.2}$$

Furthermore, by the Dynkin $\pi - \lambda$ lemma, we conclude that Lemma E.1.1 holds for all functions F of the form

$$\mathbb{1}_D, \tag{E.1.3}$$

where $D \subset 2^L \times \mathcal{X}^m$. Now, by the linearity (in F) of terms in Lemma E.1.1 we obtain that Lemma E.1.1 is true for linear combinations of functions of the form (E.1.3). It remains to use the approximation argument to get that Lemma E.1.1 holds for all non-negative and then for all bounded functions F. ∎

Now, using induction (on number of coordinates the F below depends on), the definition (E.0.3) and the Dynkin $\pi - \lambda$ lemma we can generalize previous lemma to the functions depending on infinitely many coordinates.

Lemma E.1.3 (Markov-like property of the split chain). *For any $k \in \mathbb{N}$ and measurable bounded function $F : (L^m \times \mathcal{X}^m)^{\mathbb{N}} \to \mathbb{R}$,*

$$\mathbb{E}\left(F\left((Y_{lm}^{m(l+1)-1}, X_{lm+1}^{m(l+1)})_{l \geq k}\right) | \mathcal{F}_{km}^{X}, \mathcal{F}_{km-m}^{Y}\right) = \mathbb{E}\left(F\left((Y_{lm}^{m(l+1)-1}, X_{lm+1}^{m(l+1)})_{l \geq k}\right) | X_{km}\right) \tag{E.1.4}$$

Proof. Let us only sketch the inductive step. The induction is on $n \in \mathbb{N}$ in the number of arguments for F, that is $F\colon (L^m \times \mathcal{X}^m)^n \to \mathbb{R}$. If $n = 1$ then we are in the setting of the previous lemma (recall (E.0.3)). For $n = 2$ we proceed as follows. Firstly, we consider function F of the form

$$F\left((y_{km}^{n(k+1)-1}, x_{km+1}^{m(k+1)}), (y_{m(k+1)}^{m(k+2)-1}, x_{m(k+1)+1}^{m(k+2)})\right) = G\left(y_{km}^{m(k+1)-1}, x_{km+1}^{m(k+1)}\right) H\left(y_{m(k+1)}^{m(k+2)-1}, x_{m(k+1)+1}^{m(k+2)}\right)$$

where $G, H : L^m \times \mathcal{X}^m \to \mathbb{R}$. For such F, the tower property of the conditional mean value implies that

$$\begin{aligned}&\mathbb{E}\left(F\left((Y_{lm}^{m(l+1)-1}, X_{lm+1}^{m(l+1)})_{k+1 \geq l \geq k}\right) | \mathcal{F}_{km}^{\mathbf{X}}, \mathcal{F}_{km-m}^{\mathbf{Y}}\right)\\ &= \mathbb{E}\left(G\left(Y_{km}^{m(k+1)-1}, X_{km+1}^{m(k+1)}\right) \mathbb{E}\left(H\left(Y_{m(k+1)}^{m(k+2)-1}, X_{m(k+1)+1}^{m(k+2)}\right) | \mathcal{F}_{(k+1)m}^{\mathbf{X}}, \mathcal{F}_{(k+1)m-m}^{\mathbf{Y}}\right) | \mathcal{F}_{km}^{\mathbf{X}}, \mathcal{F}_{km-m}^{\mathbf{Y}}\right).\end{aligned}$$

Now, by Lemma E.1.1, the inner conditional mean value is a function of $X_{m(k+1)}$. Another application of Lemma E.1.1 (now to the external mean value) implies that

$$\mathbb{E}\left(F\left((Y_{lm}^{m(l+1)-1}, X_{lm+1}^{m(l+1)})_{k+1 \geq l \geq k}\right) | \mathcal{F}_{km}^{\mathbf{X}}, \mathcal{F}_{km-m}^{\mathbf{Y}}\right)$$

is a function of X_{km}. This combined with Dynkin's $\pi - \lambda$ lemma concludes the inductive step (the argument for the geanearal n is analogous).

Now it is enough to apply once more π-λ Dynkin's lemma to obtain (E.1.4) for F depending on infinitely many coordinates. The proof is completed. ∎

E.2 Markov property of the vectorized split chain

In this part we show that a ***vectorized split chain*** $\mathbf{V} = (V_k)_{k\in\mathbb{N}}$, where

$$V_k = (X_{km}, Y_{km}, X_{km+1}, Y_{km+1}, \dots, X_{km+m-1}, Y_{km+m-1}) \in (\mathcal{X} \times \{0,1\})^m \tag{E.2.1}$$

is in fact a Markov chain.

Lemma E.2.1 (Markov property of m-blocks)**.** *For any measurable bounded function* $F : [(\mathcal{X} \times \{0,1\})^m]^{\mathbb{N}} \to \mathbb{R}$,

$$\mathbb{E}\left(F\left(V_{[k,\infty)}\right) \mid V_{[0,k)}\right) = \mathbb{E}\left(F\left(V_{[k,\infty)}\right) \mid V_{k-1}\right) = \mathbb{E}\left(F\left(V_{[k,\infty)}\right) \mid X_{mk-m}, X_{mk-1}, Y_{km-m}\right). \tag{E.2.2}$$

Proof. Notice that due to the Markov-like property of the split chain from Lemma E.1.3,

$$\begin{aligned}&\mathbb{E}\left(F\left(V_{[k,\infty)}\right) \mid V_{[0,k)}\right) = \mathbb{E}\left(\mathbb{E}\left(F\left(V_{[k,\infty)}\right) \mid X_{km}, V_{[0,k)}\right) | V_{[0,k)}\right) = \mathbb{E}\left(\mathbb{E}\left(F\left(V_{[k,\infty)}\right) \mid X_{km}\right) | V_{[0,k)}\right)\\ &= \mathbb{E}\left(G\left(X_{km}\right) | V_{[0,k)}\right),\end{aligned}$$

for some measurable function G. Therefore in order to prove (E.2.2) it is enough to show that for any bounded measurable function $G : \mathcal{X} \to \mathbb{R}$,

$$\mathbb{E}\left(G\left(X_{km}\right) | V_{[0,k)}\right) = \mathbb{E}\left(G\left(X_{km}\right) | X_{km-1},\ X_{km-m}, Y_{km-m}\right). \tag{E.2.3}$$

To this end for $i \in \mathbb{N}$ consider $\boldsymbol{A}_i = A_{im} \times B_i \times A_{im+1} \times B_i \times \cdots \times A_{im+m-1} \times B_i$ where $B_i \subset \{0,1\}$ and A_i are measurable subsets of $\mathcal{X}$. Recall that for brevity's sake we write $\boldsymbol{A}_{[i,j]}$ instead of $\boldsymbol{A}_i \times \boldsymbol{A}_{i+1} \times \cdots \boldsymbol{A}_j$ and similarly for $A_{[i,j]}$. Now,

$$\begin{aligned}&\mathbb{E}G(X_{km})\mathbb{1}_{V_{[0,k-1]}\in\boldsymbol{A}_{[0,k-1]}}\\ &= \mathbb{E}\bigg[\mathbb{1}_{V_{[0,k-2]}\in\boldsymbol{A}_{[0,k-2]}}\mathbb{E}\left(G(X_{km})\mathbb{1}_{Y_{km-m}\in B_{k-1}}\mathbb{1}_{X_{mk-m+1}^{mk-1}\in A_{[mk-m+1,mk-1]}}|\mathcal{F}^X_{km-m},\ \mathcal{F}^Y_{km-2m}\right)\bigg].\end{aligned} \tag{E.2.4}$$

Thus, using Lemma E.1.1 we obtain

$$\begin{aligned}&\mathbb{E}\left(G(X_{km})\mathbb{1}_{Y_{km-m}\in B_{k-1}}\mathbb{1}_{X_{[mk-m+1,mk-1]}\in A_{[mk-m+1,mk-1]}}|\mathcal{F}^X_{km-m},\ \mathcal{F}^Y_{km-2m}\right)\\ &= \left(\int_{\mathcal{X}}\right)^{m-1}\mathbb{1}_{x_1^{m-1}\in A_{[m(k-1)+1,mk-1]}}\bigg[\int_{\mathcal{X}}\int_L G(x_m)\mathbb{1}_{i\in B_{k-1}}\\ &\qquad r(X_{km-m}, x_m, di)P(x_{m-1}, dx_m)\bigg]P(x_{m-2}, dx_{m-1})\dots P(X_{km-m}, dx_1).\end{aligned} \tag{E.2.5}$$

Define

$$H(x_{km-m}, x_{m-1}, i) = \begin{cases}\dfrac{\int_{\mathcal{X}} G(x_m)r(x_{km-m}, x_m, i)P(x_{m-1}, dx_m)}{\int_{\mathcal{X}} r(x_{km-m}, x_m, i)P(x_{m-1}, dx_m)}, & x_{km-m} \in C\\ \displaystyle\int_{\mathcal{X}} G(x_m)P(x_{m-1}, dx_m), & x_{km-m} \notin C.\end{cases}$$

(it will turn out in a moment that in view of our aim (E.2.6) it is not important how we define H when $\int_{\mathcal{X}} r(x_{km-m}, x_m, i)P(x_{m-1}, dx_m) = 0$; in particular, here and later on we omit this case in our considerations). Then one can check that

$$\begin{aligned}&\int_{\mathcal{X}}\int_L G(x_m)\mathbb{1}_{i\in B_{k-1}}r(X_{km-m}, x_m, di)P(x_{m-1}, dx_m)\\ &= \int_{\mathcal{X}}\int_L H(X_{km-m}, x_{m-1}, i)\mathbb{1}_{i\in B_{k-1}}r(X_{km-m}, x_m, di)P(x_{m-1}, dx_m).\end{aligned} \tag{E.2.6}$$

Indeed, if $B_{k-1} = \{1\}$ then

$$\int_{\mathcal{X}}\int_L G(x_m)\mathbb{1}_{i\in B_{k-1}}r(X_{km-m}, x_m, di)P(x_{m-1}, dx_m) = \mathbb{1}_{X_{km-m}\in C}\int_{\mathcal{X}} G(x_m)\ r(X_{km-m}, x_m)P(x_{m-1}, dx_m)$$

and

$$\begin{aligned}&\int_{\mathcal{X}}\int_L H(X_{km-m}, x_{m-1}, i)\mathbb{1}_{i\in B_{k-1}}r(X_{km-m}, x_m, di)P(x_{m-1}, dx_m)\\ &= H(X_{km-m}, x_{m-1}, 1)\mathbb{1}_{X_{km-m}\in C}\int_{\mathcal{X}}\ r(X_{km-m}, x_m)P(x_{m-1}, dx_m)\end{aligned}$$

and it is enough to use the very definition of H in the case of $x_{km-m} \in C$. If $B_{k-1} = \{0\}$ then

$$\begin{aligned}
&\int_{\mathcal{X}} \int_L G(x_m) \mathbb{1}_{i\in B_{k-1}} r(X_{km-m}, x_m, di) P(x_{m-1}, dx_m) \\
&= \int_{\mathcal{X}} G(x_m) P(x_{m-1}, dx_m) - \mathbb{1}_{X_{km-m}\in C} \int_{\mathcal{X}} G(x_m)\, r(X_{km-m}, x_m) P(x_{m-1}, dx_m)
\end{aligned}$$

and

$$\begin{aligned}
I &:= \int_{\mathcal{X}} \int_L H(X_{km-m}, x_{m-1}, i) \mathbb{1}_{i\in B_{k-1}} r(X_{km-m}, x_m, di) P(x_{m-1}, dx_m) \\
&= H(X_{km-m}, x_{m-1}, 0) \left[1 - \mathbb{1}_{X_{km-m}\in C} \int_{\mathcal{X}} r(X_{km-m}, x_m) P(x_{m-1}, dx_m)\right] \\
&= H(X_{km-m}, x_{m-1}, 0) \mathbb{1}_{X_{km-m}\notin C} \\
&\quad + H(X_{km-m}, x_{m-1}, 0) \mathbb{1}_{X_{km-m}\in C} \left[1 - \int_{\mathcal{X}} \mathbb{1}_{X_{km-m}\in C}\, r(X_{km-m}, x_m) P(x_{m-1}, dx_m)\right].
\end{aligned}$$

Using the definition of H (and (E.0.2)) we obtain

$$\begin{aligned}
I &= \int_{\mathcal{X}} G(x_m) P(x_{m-1}, dx_m) \mathbb{1}_{X_{km-m}\notin C} \\
&\quad + \frac{\int_{\mathcal{X}} G(x_m) r(X_{km-m}, x_m, 0) P(x_{m-1}, dx_m)}{\int_{\mathcal{X}} r(X_{km-m}, x_m, 0) P(x_{m-1}, dx_m)} \mathbb{1}_{X_{km-m}\in C} \left[1 - \int_{\mathcal{X}} \mathbb{1}_{X_{km-m}\in C}\, r(X_{km-m}, x_m) P(x_{m-1}, dx_m)\right] \\
&= \mathbb{1}_{X_{km-m}\notin C} \int_{\mathcal{X}} G(x_m) P(x_{m-1}, dx_m) + \mathbb{1}_{X_{km-m}\in C} \int_{\mathcal{X}} G(x_m) r(X_{km-m}, x_m, 0) P(x_{m-1}, dx_m).
\end{aligned}$$

It remains to expand $r(X_{km-m}, x_m, 0)$ and simplify expressions.

Now, the repetition of arguments used for H in place of G (in backward manner; roughly, we proceed as follows: (E.2.6) allows us to "replace" G by H in (E.2.5) which leads to a version of (E.2.4) with G substituted by H) yields

$$\mathbb{E} G(X_{km}) \mathbb{1}_{V_{[0,k-1]}\in \boldsymbol{A}_{[0,k-1]}} = \mathbb{E} H(X_{km-m}, X_{km}, Y_{km-m}) \mathbb{1}_{V_{[0,k-1]}\in \boldsymbol{A}_{[0,k-1]}}.$$

■

Remark E.2.2 (Strong Markov property of m-blocks)**.** By standard arguments (in the area of stochastic processes) Lemma E.2.1 immediately implies that for any stopping time τ (with respect to natural filtration $\mathcal{F}^{\mathbf{V}}$), the ***strong Markov property holds***, namely

$$\mathbb{E}\left(F\left(V_{\tau+1}^{\infty}\right) | \mathcal{F}_{\tau}^{\mathbf{V}}\right) = \mathbb{E}\left(F\left(V_{\tau+1}^{\infty}\right) | V_{\tau}\right) = \mathbb{E}\left(F\left(V_{\tau+1}^{\infty}\right) \mid X_{m\tau+m-1}, X_{m\tau}, Y_{\tau m}\right). \tag{E.2.7}$$

E.3 Markov property of the random block process

Recall that the regeneration times τ_i are defined in the following way. For convenience's sake $\tau_{-1} = -m$ and for $i \geq 0$,

$$\tau_i = \min\{k > \tau_{i-1} \mid Y_k = 1,\ m|k\}. \tag{E.3.1}$$

Furthermore, ***the random block process*** is given by

$$\boldsymbol{\Xi} = (\Xi_i)_{i\geq 0}, \quad \Xi_i = X_{[\tau_{i-1}+m, \tau_i+m-1]}. \tag{E.3.2}$$

In this part we show how the strong Markov property of the vectorized split chain (recall (E.2.7)) implies that process $\boldsymbol{\Xi}$ is Markov. To this end let

$$S = \bigcup_{n\geq 1} \mathcal{X}^{nm}.$$

Remark E.3.1. In the definition of S the union should be treated as a disjoint one. In other words we can think about S as about $\bigtimes_{n\geq 1} \mathcal{X}^{nm} \times \{n\}$. Furthermore, a set $A \in S$ is measurable iff $A \cap \mathcal{X}^{nm}$ is measurable for any $n \in \mathbb{N}$.

Now, we have the following fact.

Lemma E.3.2 (Markov property of random blocks)**.** *For any $i \geq 1$ and measurable bounded function $F\colon S^{\mathbb{N}} \to \mathbb{R}$,*

$$\mathbb{E}(F(\Xi_{i+1}^{\infty}) | \Xi_0^i) = \mathbb{E}(F(\Xi_{i+1}^{\infty}) | \Xi_i) = \mathbb{E}(F(\Xi_{i+1}^{\infty}) | X_{\tau_i+m-1}, X_{\tau_i}). \tag{E.3.3}$$

Remark E.3.3. Note that due to the stationarity of $(\Xi_i)_{i\in\mathbb{N}\setminus\{0\}}$, there exists $G : \mathcal{X}^2 \to \mathbb{R}$ such that

$$\mathbb{E}(F(\Xi_{i+1},\Xi_{i+2},\ldots)|\Xi_0,\Xi_1,\ldots,\Xi_i) = \mathbb{E}(F(\Xi_{i+1},\Xi_{i+2},\ldots)|X_{\tau_i+m-1},X_{\tau_i}) = G(X_{\tau_i+m-1},X_{\tau_i}),$$

holds for all $i \geq 1$. In other words, $(\Xi_{i+1},\Xi_{i+2},\ldots)$ depends on $\Xi_0,\Xi_1,\ldots,\Xi_i$ only through the starting and ending point of the last block of length m in Ξ_i.

Proof. Clearly, by the Dynkin lemma and standard approximation techniques (recall the proof of Lemma E.1.1), it is enough to show that

$$\mathbb{E}(F(\Xi_{i+1},\Xi_{i+2},\ldots)|\Xi_0,\Xi_1,\ldots,\Xi_i) = \mathbb{E}(F(\Xi_{i+1},\Xi_{i+2},\ldots)|X_{\tau_i+m-1},X_{\tau_i}),$$

is valid for F of the form

$$F(\boldsymbol{x}_{i+1},\boldsymbol{x}_{i+2},\ldots) = \mathbb{1}_{\boldsymbol{x}_{i+1}\in\boldsymbol{A}_{i+1}}\mathbb{1}_{\boldsymbol{x}_{i+2}\in\boldsymbol{A}_{i+2}}\cdots\mathbb{1}_{\boldsymbol{x}_{i+n}\in\boldsymbol{A}_{i+n}}.$$

where for any $k \in \mathbb{N}$

$$\boldsymbol{A}_k = A_k^1 \times A_k^2 \times \cdots \times A_k^{a_k} \in \mathcal{B}^{a_k}.$$

and strictly positive $a_k \in \mathbb{N}$ are chosen is such a way that $m|a_k$.

To this end notice that for any $i, j \in \mathbb{N}$, $i \geq 0$,

$$\mathbb{1}_{\Xi_i\in\boldsymbol{A}_i}\mathbb{1}_{\Xi_{i+1}\in\boldsymbol{A}_{i+1}}\cdots\mathbb{1}_{\Xi_{i+j}\in\boldsymbol{A}_{i+j}} = \left(\prod_{k=\tau_{i-1}+m}^{\tau_{i+j}+m-1}\mathbb{1}_{X_k\in A_{b(k)}^{c(k)}}\right)\left(\prod_{k=0}^{j}\mathbb{1}_{\tau_{k+i}-\tau_{k+i-1}=a_{k+i}}\right) =: G_j\left(V_{\tau_{i-1}/m+1}^{\tau_{i+j}/m}\right),$$

where for any $k \in \mathbb{N}$, $b(k)$ and $c(k) \leq a_{b(k)}$ and functions G_j are uniquely determined and $\mathbf{V}$ is the vectorized split chain as in (E.2.1). Clearly,

$$G_j\left(V_{\tau_{i-1}/m+1}^{\tau_{i+j}/m}\right) \in \mathcal{F}_{\tau_{i+j}/m}^{\mathbf{V}}.$$

Thus, (for brevity's sake let $\sigma_i = \tau_i/m$) using the strong Markov property of the vectorized split chain $\mathbf{V}$ (see (E.2.7)), we obtain

$$\begin{aligned}\mathbb{E}F\left(\Xi_{i+1}^{\infty}\right)\mathbb{1}_{\Xi_0^i\in\mathbf{A}_{[0,i]}} &= \mathbb{E}G_{n-1}\left(V_{\sigma_i+1}^{\sigma_{i+n+1}}\right)G_i\left(V_0^{\sigma_i}\right) = \mathbb{E}G_i\left(V_0^{\sigma_i}\right)\mathbb{E}\left(G_{n-1}\left(V_{\sigma_i+1}^{\sigma_{i+n+1}}\right)|\mathcal{F}_{\sigma_i}^{\mathbf{V}}\right)\\ &\overset{(E.2.7)}{=} \mathbb{E}G_i\left(V_0^{\sigma_i}\right)\mathbb{E}\left(G_{n-1}\left(V_{\sigma_i+1}^{\sigma_{i+n+1}}\right)|X_{\tau_i+m-1},X_{\tau_i},Y_{\tau_i}\right) = \mathbb{E}\tilde{F}\left(X_{\tau_i+m-1},X_{\tau_i},Y_{\tau_i}\right)\mathbb{1}_{\Xi_0^i\in\mathbf{A}_{[0,i]}},\end{aligned}$$

for some measurable function $\tilde{F} : \mathcal{X}^2 \times \{0,1\} \to \mathbb{R}$. It is enough to recall that by the very definition, $Y_{\tau_i} = 1$. The proof is concluded. ∎

Bibliography

[1] J. Aaronson, D. Gilat, and M. Keane. On the structure of 1-dependent Markov chains. *Journal of Theoretical Probability*, 5(3):545–561, 1992.

[2] E. H. El Abdalaoui, J. Kułaga-Przymus, M. Lemańczyk, and T. de la Rue. The Chowla and the Sarnak conjectures from ergodic theory point of view. *Discrete Contin. Dyn. Syst.*, 37(6):2899–2944, 2017.

[3] E. H. El Abdalaoui, M. Lemańczyk, and T. de la Rue. A dynamical point of view on the set of $\mathscr{B}$-free integers. *Int. Math. Res. Not. IMRN*, pages 7258–7286, 2015.

[4] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.

[5] R. Adamczak and W. Bednorz. Exponential concentration inequalities for additive functionals of Markov chains. *ESAIM Probab. Stat.*, 19:440–481, 2015.

[6] K. B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.*, 245:493–501, 1978.

[7] M. Baake and U. Grimm. *Aperiodic order. Vol. 1*, volume 149 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2013. A mathematical invitation, with a foreword by Roger Penrose.

[8] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1B):700–738, 2005.

[9] H. Berbee. Periodicity and absolute regularity. *Israel J. Math.*, 55(3):289–304, 1986.

[10] V. Bergelson and T. Downarowicz. Large sets of integers and hierarchy of mixing properties of measure preserving systems. *Colloquium Mathematicum*, (1):117–150, 2008.

[11] P. Bertail and G. Ciołek. New Bernstein and Hoeffding type inequalities for regenerative Markov chains. *ALEA Lat. Am. J. Probab. Math. Stat.*, 16(1):259–277, 2019.

[12] P. Bertail and S. Clémencon. Regenerative block bootstrap for Markov chains. *Bernoulli*, 12(4):689–712, 2006.

[13] A. Besicovitch. On the density of certain sequences of integers. *Mathematische Annalen*, 110(1):336–341, 1935.

[14] A. S. Besicovitch. On the density of certain sequences of integers. *Math. Ann.*, 110(1):336–341, 1935.

[15] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999.

[16] D. Blackwell and D. Freedman. The tail σ-field of a Markov chain and a theorem of Orey. *The Annals of Mathematical Statistics*, 35(3):1291–1295, 1964.

[17] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, revised edition, 2008. With a preface by David Ruelle, Edited by Jean-René Chazottes.

[18] R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.*, 2:107–144, 2005. Update of, and a supplement to, the 1986 original.

[19] R. M. Burton, M. Denker, and M. Smorodinsky. Finite state bilaterally deterministic strongly mixing processes. *Israel Journal of Mathematics*, 95(1):115–133, 1996.

[20] J.-R. Chazottes and G. Keller. Pressure and equilibrium states in ergodic theory. *arXiv preprint arXiv:0804.2562*, 2008.

[21] X. Chen. Limit theorems for functionals of ergodic Markov chains with general state space. *Mem. Amer. Math. Soc.*, 139(664):xiv+203, 1999.

[22] S. Chowla. On abundant numbers. *J. Indian Math. Soc., New Ser.*, 1:41–44, 1934.

[23] K. L. Chung. *Markov Chains with Stationary Transition Probabilities: 2d Ed.* Springer, 1967.

[24] S. J. M. Clémencon. Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. *Statist. Probab. Lett.*, 55(3):227–238, 2001.

[25] V. Climenhaga and D. J. Thompson. Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors. *Israel J. Math.*, 192(2):785–817, 2012.

[26] V. Climenhaga and D. J. Thompson. Intrinsic ergodicity via obstruction entropies. *Ergodic Theory Dynam. Systems*, 34(6):1816–1831, 2014.

[27] S. Dani. Ergodic Theory: with a view towards Number Theory. By Manfred Einsiedler Thomas Ward. Springer Verlag, london, 2011. *Ergodic Theory and Dynamical Systems*, 32(3):1157–1160, 2012.

[28] H. Davenport. Über numeri abundantes. *Sitzungsber.Preuss.Akad.Wiss.*, pages 830–837, 1933.

[29] H. Davenport and P. Erdös. On sequences of positive integers. *Acta Arithmetica*, 2(1):147–151, 1936.

[30] J. Dedecker and S. Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electron. Commun. Probab.*, 20:no. 64, 12, 2015.

[31] R. Douc, A. Guillin, and E. Moulines. Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(2):239–257, 2008.

[32] T. Downarowicz. *Entropy in dynamical systems*, volume 18. Cambridge University Press, 2011.

[33] A. Dymek. Automorphisms of Toeplitz $\mathscr{B}$-free systems. *Bull. Pol. Acad. Sci. Math.*, 65(2):139–152, 2017.

[34] A. Dymek, S. Kasjan, J. Kułaga-Przymus, and M. Lemańczyk. $\mathscr{B}$-free sets and dynamics. *Trans. Amer. Math. Soc.*, 370:5425–5489, 2018.

[35] P. Erdös. On the Density of the Abundant Numbers. *J. London Math. Soc.*, 9(4):278–282, 1934.

[36] X. Fan, I. Grama, Q. Liu, et al. Exponential inequalities for martingales with applications. *Electronic Journal of Probability*, 20, 2015.

[37] W. Feller. *An introduction to probability theory and its applications*, volume 1. John Wiley & Sons, 2008.

[38] S. Ferenczi, J. Kułaga-Przymus, and M. Lemańczyk. Sarnak's conjecture: what's new. In *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, pages 163–235. Springer, 2018.

[39] B. de Finetti. Funzione caratteristica di un fenomeno aleatorio. *Atti della R. Accademia Nazionale dei Lincei, Ser. 6. Memorie, Classe di Scienze Fisiche, Matematiche e Naturali 4*, pages 251–299, 1931.

[40] N. A. Friedman and D. Ornstein. On isomorphism of weak Bernoulli transformations. *Advances in mathematics*, 5(3):365–394, 1970.

[41] H. Furstenberg. Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Systems Theory*, 1:1–49, 1967.

[42] H. Furstenberg, Y. Peres, and B. Weiss. Perfect filtering and double disjointness. In *Annales de l'IHP Probabilités et statistiques*, volume 31, pages 453–465, 1995.

[43] F. Gao, A. Guillin, and L. Wu. Bernstein-type concentration inequalities for symmetric Markov processes. *Theory Probab. Appl.*, 58(3):358–382, 2014.

[44] R. Garbit. A note on Furstenberg's filtering problem. *Israel J. Math.*, 182:333–336, 2011.

[45] E. Glasner. *Ergodic theory via joinings.* American Mathematical Soc., 2003.

[46] E. Glasner, J.-P. Thouvenot, and B. Weiss. Entropy theory without a past. *Ergodic Theory and Dynamical Systems*, 20(5):1355–1370, 2000.

[47] R. M. Gray. *Entropy and information theory.* Springer, New York, second edition, 2011.

[48] A. H. Guide. *Infinite dimensional analysis.* Springer, 2006.

[49] R. R. Hall. *Sets of multiples*, volume 118 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 1996.

[50] E. Hewitt and L. J. Savage. Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.*, 80:470–501, 1955.

[51] A. Holroyd and T. Liggett. Finitely dependent coloring. *Forum of Mathematics, Pi*, 4, 02 2014.

[52] C. Houdré and R. Restrepo. A probabilistic approach to the asymptotics of the length of the longest alternating subsequence. *Electron. J. Combin.*, 17(1):Research Paper 168, 19, 2010.

[53] S. Janson. On degenerate sums of m-dependent variables. *J. Appl. Probab.*, 52(4):1146–1155, 2015.

[54] B. Jiang, Q. Sun, and J. Fan. Bernstein's inequality for general Markov chains. *arXiv preprint arXiv:1805.10721*, 2018.

[55] A. Kechris. *Classical descriptive set theory*, volume 156. Springer Science & Business Media, 2012.

[56] G. Keller. Tautness of sets of multiples and applications to $\mathcal{B}$-free systems. *Studia Math.*, 247(2):205–216, 2019.

[57] G. Keller and C. Richard. Dynamics on the graph of the torus parametrization. *Ergodic Theory Dynam. Systems*, 38(3):1048–1085, 2018.

[58] D. Kerr and H. Li. Independence in topological and C^*-dynamics. *Math. Ann.*, 338(4):869–926, 2007.

[59] A. Kłopotowski and J. B. Robertson. Some New Examples of Pairwise Independent Random Variables. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 72–88, 1999.

[60] J. Kułaga-Przymus and M. Lemańczyk. Sarnak's conjecture from the ergodic theory point of view. *arXiv preprint arXiv:2009.04757 (to appear in Encyclopedia for Complexity and System Science)*, 2020.

[61] J. Kułaga-Przymus, M. Lemańczyk, and B. Weiss. On invariant measures for $\mathscr{B}$-free systems. *Proc. Lond. Math. Soc. (3)*, 110(6):1435–1474, 2015.

[62] D. Kwietniak. Topological entropy and distributional chaos in hereditary shifts with applications to spacing shifts and beta shifts. *Discrete Contin. Dyn. Syst.*, 33(6):2451–2467, 2013.

[63] D. Kwietniak, M. Łącka, and P. Oprocha. Generic points for dynamical systems with average shadowing. *Monatshefte für Mathematik*, 183(4):625–648, 2017.

[64] O. E. Lanford. Entropy and equilibrium states in classical statistical mechanics. *Lecture Notes in Physics*, pages 1–113, 1975.

[65] K. Lau and A. Zame. On weak mixing of cascades. *Math. Systems Theory*, 6:307–311, 1972/73.

[66] M. Lemańczyk. General Bernstein-Like Inequality for Additive Functionals of Markov Chains. *Journal of Theoritical Probability*, 2020.

[67] P. Lezaud. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8(3):849–867, 1998.

[68] P. Lezaud. Chernoff and Berry-Esséen inequalities for Markov processes. *ESAIM Probab. Statist.*, 5:183–201, 2001.

[69] X. Li and C. Chen. Inequalities for the gamma function. *JIPAM. J. Inequal. Pure Appl. Math.*, 8(1):Article 28, 3, 2007.

[70] D. Lind and B. Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.

[71] P. T. Maker. The ergodic theorem for a sequence of functions. *Duke Mathematical Journal*, 6(1):27–30, 1940.

[72] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *annals of Mathematics*, pages 1015–1056, 2016.

[73] F. Matúš. On two–block–factor sequences and one–dependence. *Proceedings of the American Mathematical Society*, 124(4):1237–1242, 1996.

[74] F. Merlevède, M. Peligrad, and E. Rio. *High dimensional probability V: the Luminy volume*, volume 5. Institute of Mathematical Statistics, Beachwood, OH, 2009. Including papers from the 5th International Conference (HDP V) held in Luminy, May 26–30, 2008.

[75] F. Merlevède, M. Peligrad, and E. Rio. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields*, 151(3-4):435–474, 2011.

[76] F. Merlevède and E. Rio. Strong approximation for additive functionals of geometrically ergodic Markov chains. *Electronic Journal of Probability*, 20, 2015.

[77] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.

[78] L. Mirsky. Note on an asymptotic formula connected with r-free integers. *The Quarterly Journal of Mathematics*, (1):178–182, 1947.

[79] L. Mirsky. Arithmetical pattern problems relating to divisibility by rth powers. *Proceedings of the London Mathematical Society*, 2(1):497–508, 1948.

[80] E. Nummelin. *General irreducible Markov chains and nonnegative operators*, volume 83 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 1984.

[81] R. Olshen. A note on exchangeable sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 28(4):317–321, 1974.

[82] R. A. Olshen. The coincidence of measure algebras under an exchangeable probability. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 18:153–158, 1971.

[83] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.*, 20:no. 79, 32, 2015.

[84] R. Pavlov. On controlled specification and uniqueness of the equilibrium state in expansive systems. *Nonlinearity*, 32(7):2441–2466, 2019.

[85] R. Peckner. Uniqueness of the measure of maximal entropy for the squarefree flow. *Israel J. Math.*, 210(1):335–357, 2015.

[86] O. Ramaré. Chowla's Conjecture: From the Liouville Function to the Moebius Function. In *Ergodic Theory and Dynamical Systems in their Interactions with Arithmetics and Combinatorics*, pages 317–323. Springer, 2018.

[87] A. Rényi. Representations for real numbers and their ergodic properties. *Acta Mathematica Academiae Scientiarum Hungarica*, 8(3):477–493, 1957.

[88] E. Rio. Inequalities and limit theorems for weakly dependent sequences. 2013.

[89] V. A. Rokhlin. New progress in the theory of transformations with invariant measure. *RuMaS*, 15(4):1–22, 1960.

[90] D. Ruelle. *Statistical mechanics.* The Mathematical Gazette, 1999. Rigorous results, Reprint of the 1989 edition.

[91] P.-M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *The Annals of Probability*, 28(1):416–461, 2000.

[92] P. Sarnak. Three lectures on the Möbius function, randomness and dynamics. unpublished, 2010.

[93] P. C. Shields. *The ergodic theory of discrete sample paths*, volume 13. American Mathematical Soc., 1996.

[94] J. G. Sinaĭ. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972.

[95] B. Stanley. Bounded density shifts. *Ergodic Theory Dynam. Systems*, 33(6):1891–1928, 2013.

[96] T. Tao. The Chowla conjecture and the Sarnak conjecture. https://terrytao.wordpress.com/2012/10/14/the-chowla-conjecture-and-the-sarnak-conjecture/.

[97] V. De Valk. Hilbert space representations of m-dependent processes. *The Annals of Probability*, pages 1550–1570, 1993.

[98] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[99] P. Walters. A variational principle for the pressure of continuous transformations. *American Journal of Mathematics*, 97(4):937–971, 1975.

[100] P. Walters. *An introduction to ergodic theory*, volume 79. Springer Science & Business Media, 2000.

[101] B. Weiss. Subshifts of finite type and sofic systems. *Monatsh. Math.*, 77:462–474, 1973.

[102] O. Wintenberger. Deviation inequalities for sums of weakly dependent time series. *Electron. Commun. Probab.*, 15:489–503, 2010.