Entropy rate of product of independent processes

Joanna Kułaga-Przymus[1] · Michał D. Lemańczyk[2]

Abstract

We study the multiplicative version of the classical Furstenberg's filtering problem, where instead of the sum $\mathbf{X} + \mathbf{Y}$ one considers the product $\mathbf{X} \cdot \mathbf{Y}$ ($\mathbf{X}$ and $\mathbf{Y}$ are bilateral, real, finitely-valued, stationary independent processes, $\mathbf{Y}$ is taking values in $\{0, 1\}$). We provide formulas for $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y})$. As a consequence, we show that if $\mathbf{H}(\mathbf{X}) > \mathbf{H}(\mathbf{Y}) = 0$ and $\mathbf{X} \amalg \mathbf{Y}$, then $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) < \mathbf{H}(\mathbf{X})$ (and thus $\mathbf{X}$ cannot be filtered out from $\mathbf{X} \cdot \mathbf{Y}$) whenever $\mathbf{X}$ is not bilaterally deterministic, $\mathbf{Y}$ is ergodic and $\mathbf{Y}$ first return to 1 can take arbitrarily long with positive probability. On the other hand, if almost surely $\mathbf{Y}$ visits 1 along an infinite arithmetic progression of a fixed difference (with possibly some more visits in between) then we can find $\mathbf{X}$ that is not bilaterally deterministic and such that $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) = \mathbf{H}(\mathbf{X})$. As a consequence, a $\mathscr{B}$-free system (X_η, S) is proximal if and only if there is always an entropy drop $h(\kappa * \nu_\eta) < h(\kappa)$ for any κ corresponding to a non-bilaterally deterministic process of positive entropy. These results partly settle some open problems on invariant measures for $\mathscr{B}$-free systems.

Keywords Entropy rate · Independent processes · Bilaterally deterministic processes · $\mathscr{B}$-free systems · Invariant measures · Filtering

Mathematics Subject Classification 37A35 · 28D20 · 60G10 · 37B10 · 60G35

Contents

1 Background and main results

In this paper we concentrate on two seemingly unrelated areas:

(A) multiplicative version of the classical Furstenberg's problem on defiltering a noisy signal,

(B) open questions related to invariant measures for so-called $\mathscr{B}$-free systems.

We will now give some background on both, (A) and (B). Then we will present the main technical result and its consequences. Finally, since the paper is a mixture of probabilistic and ergodic tools, we present in a separate section a dictionary allowing for a simultaneous use of both. The remainder of the paper is devoted to the proofs and examples illustrating our results. In the appendix we give some more detailed comments on $\mathscr{B}$-free systems that can be of an independent interest.

1.1 Furstenberg's filtering problem

The classical Furstenberg's filtering problem from the celebrated paper [17] concerns two stationary real processes: $\mathbf{X}$ (the emitted signal) and $\mathbf{Y}$ (the noise), with $\mathbf{X} \amalg \mathbf{Y}$, and the following question is asked:

Question 1 ([17]) When is $\mathbf{X}$ measurable with respect to σ-algebra generated by $\mathbf{X} + \mathbf{Y}$? In other words, when it is possible to recover $\mathbf{X}$ from the received signal $\mathbf{X} + \mathbf{Y}$?

In order to address this problem, Furstenberg [17] introduced the notion of disjointness of dynamical systems, which even today remains one of the central concepts

in ergodic theory. Recall that measure-theoretic dynamical systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ are *disjoint* if the product measure $\mu \otimes \nu$ is the only $(T \times S)$-invariant measure, projecting as μ and ν onto the first and second coordinates, respectively.[1] Recall also that each measure-theoretic dynamical system $(X, \mathcal{B}, \mu, T)$ yields a family of bilateral, real, stationary processes in the following way: for any measurable function $f : X \to \mathbb{R}$, the process $\mathbf{X} = (f \circ T^i)_{i \in \mathbb{Z}}$ is stationary. In particular, each measurable partition of X into finitely many pieces yields a finitely-valued stationary process. On the other hand, each real stationary process $\mathbf{X}$ yields a (symbolic) measure-theoretic dynamical system by taking the left shift S on the product space $\mathbb{R}^{\mathbb{Z}}$, with the invariant measure given by the distribution of $\mathbf{X}$ (if the state space of $\mathbf{X}$ is smaller than $\mathbb{R}$, we can consider the left shift S on the appropriate smaller product space). A crucial basic observation is that whenever the family of functions $\{f \circ T^i : i \in \mathbb{Z}\}$ generates $\mathcal{B}$ then the resulting symbolic (measure-theoretic) dynamical system is isomorphic to $(X, \mathcal{B}, \mu, T)$. Last, but not least, we say that processes $\mathbf{X}$ and $\mathbf{Y}$ are *absolutely independent*, whenever the resulting dynamical systems are disjoint. Furstenberg showed that absolute independence is a sufficient condition, under which one has the positive answer to Question 1:

Theorem 1.1 ([17]) *Suppose that* $\mathbf{X}$ *and* $\mathbf{Y}$ *are integrable and that* $\mathbf{X}$ *is absolutely independent from* $\mathbf{Y}$. *Then* $\mathbf{X}$ *is measurable with respect to* σ-*algebra generated by* $\mathbf{X} + \mathbf{Y}$.

Garbit [19] showed that the integrability assumption can be dropped and the assertion of Theorem 1.1 still holds.

We are interested in the following modification of Question 1: instead of the sum of processes $\mathbf{X}$ and $\mathbf{Y}$, we consider their product

$$\mathbf{M} := \mathbf{X} \cdot \mathbf{Y} = (X_i \cdot Y_i)_{i \in \mathbb{Z}}.$$

Notice that if $\mathbf{X}$ and $\mathbf{Y}$ take only positive values, we can define processes $\log \mathbf{X}$ and $\log \mathbf{Y}$. Since $\log \mathbf{M} = \log \mathbf{X} + \log \mathbf{Y}$, by the result of Garbit, $\mathbf{X}$ can be recovered from $\mathbf{M}$ whenever $\mathbf{X}$ and $\mathbf{Y}$ are disjoint. Therefore, it is natural to ask whether the same conclusion as in Theorem 1.1 holds for processes that admit zero as a value. The simplest instance of this is when the state space, e.g., of $\mathbf{Y}$, equals $\{0, 1\}$. One can think of $\mathbf{M}$ as of the original signal $\mathbf{X}$, where some of the information was lost (due to $Y_i = 0$), instead of just being perturbed (by adding Y_i to X_i). Thus, we deal with the following problem:

Question 2 Let $\mathbf{X}$ and $\mathbf{Y}$ be bilateral, real, finitely-valued, stationary processes, with $Y_i \in \{0, 1\}$. Suppose that $\mathbf{X} \sqcup \mathbf{Y}$. Is it possible to recover $\mathbf{X}$ and / or $\mathbf{Y}$ from $\mathbf{M}$?

A similar (in fact, much more general) problem of retrieving a lost signal was studied by Furstenberg, Peres and Weiss in [18]. Let $\mathbf{X}^{(i)} = \left(X_i^{(U_i)} \right)_{i \in \mathbb{Z}}$, where $i \in \mathbb{N}$, be a family of processes and $\mathbf{U}$ be an $\mathbb{N}$-valued process. Suppose that all these processes

---

[1] The ergodicity of $(X, \mathcal{B}, \mu, T)$ or $(Y, \mathcal{C}, \nu, S)$ is a necessary condition for disjointness.

are stationary and define

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

(informally, $\mathbf{U}$ chooses among the family of processes).

Question 3 Is it possible to recover $\mathbf{U}$ from $\mathbf{X}^{(\mathbf{U})}$?

In order to answer this question the authors of [18] introduce the notion of double disjointness. We say that process $\mathbf{Y}$ is *doubly disjoint* from $\mathbf{X}$ if every self-joining of $\mathbf{Y}$ is absolutely disjoint from $\mathbf{X}$. In other words if $(\mathbf{X}', \mathbf{Y}', \mathbf{Y}'')$ is a stationary process such that $\mathbf{X}' \sim \mathbf{X}$ and $\mathbf{Y}', \mathbf{Y}'' \sim \mathbf{Y}$ then $\mathbf{X}' \perp\!\!\!\perp (\mathbf{Y}', \mathbf{Y}'')$. The most basic example of doubly disjoint processes arises when $\mathbf{Y}$ is of zero entropy rate (then every self-joining of $\mathbf{Y}$ has zero entropy) and $\mathbf{X}$ has trivial tail-σ-field (let us add that, in fact, if $\mathbf{Y}$ is doubly disjoint from $\mathbf{X}$ then **necessarily $\mathbf{H}(\mathbf{Y}) = 0$ and $\mathbf{X}$ is ergodic**). (For the definition of entropy rate, see (1.2) below.) Now, the main result of [18] 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 doubly disjoint 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 2 (and retrieve $\mathbf{Y}$ from $\mathbf{M}$). Consider two processes $\mathbf{X}^{(i)}$, for $i \in \{0, 1\}$, where

$$X_j^{(i)} = i X_j \tag{1.1}$$

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 doubly disjoint from $\mathbf{X}$. Since the role of $\mathbf{X}$ and $\mathbf{Y}$ is here not symmetric (and $\mathbf{M}$ and $\mathbf{Y}$ do not determine $\mathbf{X}$ unlike when one studies the sum $\mathbf{X} + \mathbf{Y}$), it is interesting to ask whether one can also retrieve $\mathbf{X}$. To stay compatible with the notion of double disjointness, we will assume that $\mathbf{H}(\mathbf{X}) > \mathbf{H}(\mathbf{Y}) = 0$. Then, clearly, a necessary condition for having the positive answer to Question 2 is that $\mathbf{H}(\mathbf{M}) = \mathbf{H}(\mathbf{X})$. Having this in mind, we will deal with the following three more specific problems:

Question 4 (A) Is there a general formula for the entropy rate $\mathbf{H}(\mathbf{M})$ of $\mathbf{M} = \mathbf{X} \cdot \mathbf{Y}$?
(B) Do we always have $\mathbf{H}(\mathbf{M}) > 0$ whenever $\mathbf{H}(\mathbf{X}) > 0$?
(C) Can we have $\mathbf{H}(\mathbf{M}) = \mathbf{H}(\mathbf{X})$ with $\mathbf{H}(\mathbf{X}) > 0$?

Remark 1.2 Notice that the answers to Question 1 in [17] and to Question 3 in [18] depend only on the properties of the underlying dynamical systems corresponding to $\mathbf{X}$ and $\mathbf{Y}$. In this paper the situation will be different and the ability to defilter $\mathbf{X}$ from $\mathbf{M}$ will highly depend on the properties of the stochastic processes under consideration, cf. Example 2.16.

1.2 Invariant measures for $\mathscr{B}$-free systems

Question 4 is a generalization of some questions asked in [28] in the context of $\mathscr{B}$-free systems. For $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$, consider the corresponding sets of multiples and $\mathscr{B}$-free

integers:

$$\mathcal{M}_{\mathscr{B}} := \bigcup_{b \in \mathscr{B}} b\mathbb{Z} \text{ and } \mathcal{F}_{\mathscr{B}} := \mathbb{Z} \setminus \mathcal{M}_{\mathscr{B}}.$$

Such sets were studied already in the 30's from the number-theoretic viewpoint (see, e.g. [2, 5, 7–9, 14]). The most prominent example of $\mathcal{F}_{\mathscr{B}}$ is the set of square-free integers (with $\mathscr{B}$ being the set of squares of all primes). The dynamical approach to $\mathscr{B}$-free sets was initiated by Sarnak [35] who 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 $\{-1, 0, 1\}^{\mathbb{Z}}$.[2] For an arbitrary $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$, let X_η be the orbit closure of $\eta = \mathbb{1}_{\mathcal{F}_{\mathscr{B}}} \in \{0, 1\}^{\mathbb{Z}}$ under the left shift, i.e. we deal with a *subshift* of $(\{0, 1\}^{\mathbb{Z}}, S)$.[3] We say that (X_η, S) is a *$\mathscr{B}$-free system*. In the so-called Erdös case (when the elements of $\mathscr{B}$ are pairwise coprime, $\mathscr{B}$ is infinite and $\sum_{b \in \mathscr{B}} 1/b < \infty$), X_η is *hereditary*: for $y \leqslant x$ coordinatewise, with $x \in X_\eta$ and $y \in \{0, 1\}^{\mathbb{Z}}$, we have $y \in X_\eta$. In other words, $X_\eta = M(X_\eta \times \{0, 1\}^{\mathbb{Z}})$, where M stands for the coordinatewise multiplication of sequences. For a general $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$, X_η may no longer be hereditary and we consider its hereditary closure $\widetilde{X}_\eta := M(X_\eta \times \{0, 1\}^{\mathbb{Z}})$ instead. Usually, one assumes at least the *primitivity* of $\mathscr{B}$ (i.e. $b \nmid b'$ for $b \neq b'$ in $\mathscr{B}$).

Given a *topological dynamical system* (X, T), i.e. a homeomorphism T acting on a compact metric space X, let $\mathcal{B}$ be the σ-algebra of Borel subsets of X. By $\mathcal{M}(X, T)$ we will denote the set of all probability Borel T-invariant measures on X and $\mathcal{M}^e(X, T)$ will stand for the subset of ergodic measures. Each choice of $\mu \in \mathcal{M}(X, T)$ results in a *measure-theoretic dynamical system*, i.e. a 4-tuple $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B}, \mu)$ is a standard probability Borel space, with an automorphism T. We often skip $\mathcal{B}$ and write (X, μ, T). Recall also that $x \in X$ is said to be *generic* for $\mu \in \mathcal{M}(X, T)$, whenever $\lim_{N \to \infty} \frac{1}{N} \sum_{n \leqslant N} \delta_{T^n x} = \mu$ in the weak topology. If the convergence takes place only along a subsequence $(N_k)_{k \geqslant 1}$ then we say that x is *quasi-generic* for μ. Each measure μ resulting this way yields a measure theoretic dynamical system $(X, \mathcal{B}, \mu, T)$.

A central role in the theory of $\mathscr{B}$-free systems is played by the so-called *Mirsky measure*, denoted by ν_η. In the Erdös case, η is a generic point for ν_η (in general, η is quasi-generic along some natural sequence (N_k)), see [12]. It was shown in [12, 28] that all invariant measures for $\widetilde{X}_\eta$ are of the following special form:

Theorem 1.3 (cf. Sect. 1) *For any $\nu \in \mathcal{M}(\widetilde{X}_\eta, S)$, there exists $\rho \in \mathcal{M}(X_\eta \times \{0, 1\}^{\mathbb{Z}}, S \times S)$ such that $\rho|_{X_\eta} = \nu_\eta$ and $M_*(\rho) = \nu$.*[4]

---

[2] Recall that $\boldsymbol{\mu}(n) = (-1)^k$ if n is a product of k distinct primes, $\boldsymbol{\mu}(1) = 1$ and $\boldsymbol{\mu}(n) = 0$ otherwise; $\boldsymbol{\mu}^2$ is the characteristic function of the set of square-free integers.

[3] More generally, given a finite alphabet $\mathcal{X}$, we define S to be the left shift on $\mathcal{X}^{\mathbb{Z}}$, i.e. $S((x_i)_{i \in \mathbb{Z}}) = (y_i)_{i \in \mathbb{Z}}$, where $y_i = x_{i+1}$, $i \in \mathbb{Z}$. Each closed S-invariant subset of $\mathcal{X}^{\mathbb{Z}}$ is called a *subshift*.

[4] By $\rho|_{X_\eta}$, we denote the projection of ρ onto the first coordinate X_η; $M_*(\rho)$ stands for the image of ρ via M. We will use similar notation later on, too.

Recall that given a measure-theoretic dynamical system $(X, \mathcal{B}, \mu, T)$, any T-invariant sub-σ-algebra $\mathcal{A} \subset \mathcal{B}$ is called a *factor* of $(X, \mathcal{B}, \mu, T)$.[5] Notice that given ν and ρ as in Theorem 1.3, $(\widetilde{X}_\eta, \nu, S)$ is a factor of $(X_\eta \times \{0, 1\}^{\mathbb{Z}}, \rho, S \times S)$.

The measure-theoretic entropy of $(X, \mathcal{B}, \mu, T)$ will be denoted by $h_\mu(T, \mathcal{B})$. If no confusion arises, we will also write $h(\mu, T)$ or even $h(\mu)$. If $\mathbf{X}$ is a finitely-valued stationary process determining (X, μ, T) (as described in Sect. 1.1) then $\mathbf{H}(\mathbf{X}) = h(\mu)$.

The Mirsky measure ν_η is of zero entropy. Moreover, it was shown in [28] in the Erdös case that (X_η, S) is intrinsically ergodic (it has exactly one measure realizing the topological entropy). Its measure of maximal entropy equals $M_*(\nu_\eta \otimes B_{1/2, 1/2})$, where $B_{1/2, 1/2}$ stands for the Bernoulli measure on $\{0, 1\}^{\mathbb{Z}}$ of entropy $\log 2$. These results were extended in [12] to a general $\mathscr{B}$ (one needs to replace X_η with $\widetilde{X}_\eta$). In the Erdös case, the topological entropy of (X_η, S) is equal $\prod_{i \geqslant 1} (1 - 1/b_i)$[6] (in general, the topological entropy of $(\widetilde{X}_\eta, S)$ equals $\overline{d}(\mathcal{F}_{\mathscr{B}})$ [12]).[7] This led to the study of *product type measures* (or *multiplicative convolutions*):

$$\nu_\eta * \kappa := M_*(\nu_\eta \otimes \kappa).$$

In particular, it was proved that

$$0 < h(\nu_\eta * B_{1/2, 1/2}) < h(B_{1/2, 1/2}).$$

Moreover, it was shown that for each value $0 \leqslant h \leqslant \prod_{i \geqslant 1} (1 - 1/b_i)$ there is an ergodic measure κ satisfying $h(X_\eta, \nu_\eta * \kappa) = h$. However, some fundamental questions related to such measures were left open – they turn out to be a special instance of Question 4 (see Question 1 in [28]):

Question 5 (A) Is there a general formula for the entropy $h(\nu_\eta * \kappa)$ of $\nu_\eta * \kappa$?
(B) Do we always have $h(\nu_\eta * \kappa) > 0$ whenever $h(\kappa) > 0$?
(C) Can we have $h(\nu_\eta * \kappa) = h(\kappa)$ with $h(\kappa) > 0$?

1.3 Main technical result

Our main tool used to answer Questions 4 and 5 is concerned with the entropy rate of stationary processes. Before we can formulate it, we need some definitions and notation that will be used througout the whole paper.

All random variables and processes will be defined on a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Sometimes, we will 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 $\mathbb{P}(A) > 0$. In particular, $\mathbb{E}_A$ will stand for the expectation taken with respect to $\mathbb{P}_A$. For convenience sake, we will write A, B instead of $A \cap B$ for any $A, B \in \mathcal{F}$: for example, $\mathbb{E}_{A, B}$ stands

---

[5] Equivalently, if $\pi : (X, \mathcal{B}, \mu, T) \to (Z, \mathcal{D}, \rho, R)$ intertwines the actions of T and R, then R is called a factor of T (as $\mathcal{A} = \pi^{-1}(\mathcal{D}) \subset \mathcal{B}$ is T-invariant).

[6] In the special case of the square-free system the result was proved by Peckner [33].

[7] By $\overline{d}$ we denote the upper asymptotic density.

for $\mathbb{E}_{A \cap B}$. A central role will be played by the **Shannon entropy** of a random variable X, denoted by us by $\mathbf{H}(X)$. Although we will recall basic definitions and properties related to $\mathbf{H}(X)$, some well-known facts will be taken for granted (all of them can be found in [22]). All random processes will be **bilateral** and **real**. Usually, they will be also **finitely-valued** and **stationary**, however sometimes we will need auxiliary countably-valued, non-stationary processes. Recall that a process $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ is stationary if $(X_i)_{i \in \mathbb{Z}}$ has the same distribution as $(X_{i+1})_{i \in \mathbb{Z}}$ and finitely-valued if, for every $i \in \mathbb{Z}$, $X_i \in \mathcal{X}$, with $|\mathcal{X}| < \infty$.

Let now X, Y be random variables taking values in finite state spaces $\mathcal{X}$, $\mathcal{Y}$ respectively and fix $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. We put $\mathbf{H}_A(X) = -\sum_{x \in \mathcal{X}} \mathbb{P}_A(X = x) \log_2 \mathbb{P}_A(X = x)$. Moreover, $\mathbf{H}_A(X \mid Y) = \sum_{y \in \mathcal{Y}} \mathbb{P}_A(Y = y) \mathbf{H}_{Y=y,A}(X)$ will stand for the *conditional Shannon entropy* of X with respect to Y. When $\mathbb{P}(A) = 1$, we will omit subscript A and write $\mathbf{H}(X)$ and $\mathbf{H}(X \mid Y)$, respectively.

To shorten the notation, we will use the following convention. For a subset $A = \{i_1, \ldots, i_n\} \subset \mathbb{Z}$ with $i_1 < i_2 < \cdots < i_n$ and a process $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$, we will write

$$X_A = \left(X_{i_1}, X_{i_2}, \ldots, X_{i_n} \right).$$

Moreover, for any $k \leqslant \ell$ in $\mathbb{Z}$, we define *integer intervals*:

$$[k, \infty) := \{k, k+1, \ldots\}, \quad (-\infty, \ell] := \{\ell, \ell-1, \ldots\}, \quad [k, \ell] := \{k, k+1, \ldots, \ell\}.$$

For example, $X_{[0,n]} = (X_0, \ldots, X_n)$ for $n \in \mathbb{N}$. It is natural and convenient to interpret $[k, \ell]$ as $\varnothing$ if $\ell < k$, $\mathbf{H}(X_\varnothing) = 0$ and $\mathbf{H}(X \mid Y_\varnothing) = \mathbf{H}(X)$.

Consider now two random processes $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ and $\mathbf{Y} = (Y_i)_{i \in \mathbb{Z}}$ such that $(\mathbf{X}, \mathbf{Y}) := ((X_i, Y_i))_{i \in \mathbb{Z}}$ is stationary. Then

$$\mathbf{H}(\mathbf{X}) = \lim_{n \to \infty} \frac{1}{n} \mathbf{H}(X_{[0,n-1]}), \quad \mathbf{H}(\mathbf{X} \mid \mathbf{Y}) = \lim_{n \to \infty} \frac{1}{n} \mathbf{H}(X_{[0,n-1]} \mid Y_{[0,n-1]}) \quad (1.2)$$

will denote, respectively, the *entropy rate* of $\mathbf{X}$ and the *relative entropy rate* of $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ with respect to $\mathbf{Y} = (Y_i)_{i \in \mathbb{Z}}$. By the stationarity of $\mathbf{X}$, $\mathbf{H}(\mathbf{X}) = \lim_{n \to \infty} \mathbf{H}(X_0 \mid X_{[-n,-1]})$. Note that both limits in (1.2) exist due to the subadditivity of appropriate sequences.

Remark 1.4 Sometimes it is convenient to extend the classical definition of the conditional entropy, $\mathbf{H}(X \mid Y)$, to $\mathbf{H}(X \mid \mathcal{G})$, where X is a finitely-valued random variable and $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-algebra (see [20], Chapter 14, for a precise construction and proofs). This extension is justified by the following facts. If $\mathcal{G} = \sigma(Y)$ then $\mathbf{H}(X \mid \sigma(Y)) = \mathbf{H}(X \mid Y)$ for any random variable Y.[8] If $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ are sub-σ-algebras then $\mathbf{H}(X \mid \mathcal{G}) \leqslant \mathbf{H}(X \mid \mathcal{H})$. Moreover, 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. Thus, for example, it makes sense to write $\mathbf{H}(\mathbf{X}) = \mathbf{H}(X_0 \mid X_{(-\infty,-1]}) = \lim_{n \to \infty} \mathbf{H}(X_0 \mid X_{(-n,-1]})$. The chain rule is still valid, namely, if X and Y are finitely-valued then

---

[8] Recall that $\sigma(Y)$ stand for the smallest σ-algebra making Y measurable.

$$\mathbf{H}(X, Y \mid \mathcal{G}) = \mathbf{H}(X \mid \mathcal{G}) + \mathbf{H}(Y \mid \sigma(\mathcal{G}, \sigma(X))). \tag{1.3}$$

Furthermore, $\mathbf{H}(X \mid \mathcal{G}) = 0$ if and only if X is $\mathcal{G}$-measurable and $\mathbf{H}(X \mid \mathcal{G}) = \mathbf{H}(X)$ if and only if X is independent of $\mathcal{G}$.

Remark 1.5 We will often omit some technicalities concerning events of zero probability. First, we tacitly assume that $\mathcal{F}$ is complete (i.e. all subevents of zero-measure events are measurable). Secondly, when considering sub-σ-fields associated with random processes, we think of them as of **measure-σ-algebras** (intuitively, we look at them "up to events of probability zero"). Given sub-σ-fields $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$, sometimes we will write

$$\mathcal{G} \overset{\mathbb{P}}{\subset} \mathcal{H}$$

to stress that for every $G \in \mathcal{G}$ there is $H \in \mathcal{H}$ such that $\mathbb{P}(G \triangle H) = 0$ but not necessarily $\mathcal{G} \subset \mathcal{H}$ (with obvious modifications for $\overset{\mathbb{P}}{\supset}$ and $\overset{\mathbb{P}}{=}$). However, in most cases, we will skip such considerations, cf. the last sentence of the previous remark.

Given processes $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ and $\mathbf{Y} = (Y_i)_{i \in \mathbb{Z}}$, we will be interested in the entropy rate $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y})$ of their product $\mathbf{X} \cdot \mathbf{Y} = (X_i \cdot Y_i)_{i \in \mathbb{Z}}$. Our **standing assumptions** (unless stated otherwise) will be that:

(i) $\mathbf{X}$ is finitely-valued, $\mathbf{Y}$ is binary ($Y_i \in \{0, 1\}$ for $i \in \mathbb{Z}$) and $\mathbb{P}(Y_0 = 1) > 0$,
(ii) $\mathbf{X} \amalg \mathbf{Y}$, i.e. $\mathbf{X}$ and $\mathbf{Y}$ are independent.

Notice that by the independence of $\mathbf{X}$ and $\mathbf{Y}$, process $(\mathbf{X}, \mathbf{Y})$ is stationary. Moreover, $\mathbf{X} \cdot \mathbf{Y}$ is a factor of $(\mathbf{X}, \mathbf{Y})$.[9] The quantity $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y})$ turns out to be easier to deal with than $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y})$. A particular emphasis will be put on the case when $\mathbf{H}(\mathbf{Y}) = 0$ in which $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) = \mathbf{H}(\mathbf{X} \cdot \mathbf{Y})$[10] and $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) \leqslant \mathbf{H}(\mathbf{X})$.[11]

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 1:

$$R_i = \begin{cases} \inf\{j \geqslant 0 : Y_j = 1\}, & i = 0, \\ \inf\{j \geqslant R_{i-1} : Y_j = 1\}, & i \geqslant 1, \\ \sup\{j < R_{i+1} : Y_j = 1\}, & i \leqslant -1. \end{cases} \tag{1.4}$$

Note that, in general, $\mathbf{R}$ can be countably-valued. If $\mathbf{Y}$ is ergodic then it visits 1 infinitely often, both in the future and in the past and, thus, $\mathbf{R}$ is well-defined almost everywhere. However, we don't need to assume the ergodicity of $\mathbf{Y}$ to be able to speak of $\mathbf{R}$ and we will just assume that:

---

[9] Cf. Remark 1.23.

[10] To see this, it suffices to notice that given a process $\mathbf{Z}$, we have $\frac{1}{n}\mathbf{H}(Z_{[0,n]} \mid Y_{[0,n]}) \leqslant \frac{1}{n}\mathbf{H}(Z_{[0,n]}) = \frac{1}{n}\mathbf{H}(Z_{[0,n]} \mid Y_{[0,n]}) + \frac{1}{n}\mathbf{H}(Y_{[0,n]})$. In particular, we can take $\mathbf{Z} = \mathbf{X} \cdot \mathbf{Y}$.

[11] Using the fact that factors cannot increase entropy and by the subadditivity of entropy rate, we have $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) \leqslant \mathbf{H}((\mathbf{X}, \mathbf{Y})) \leqslant \mathbf{H}(\mathbf{X}) + \mathbf{H}(\mathbf{Y})$.

(iii) **Y** is such that the definition of **R** makes sense.

Whenever (i), (ii) and (iii) hold, we will say that the pair $(\mathbf{X}, \mathbf{Y})$ is *good*. If **Y** is binary, with $\mathbb{P}(Y_0 = 1) > 0$ and such that (iii) holds, we will say that **Y** is *good*.

Remark 1.6 We will use lowercase letters to denote *realizations* of the corresponding random processes (denoted by uppercase letters). Recall that $\mathbf{x} = (x_i)_{i \in \mathbb{Z}}$ is a realization of $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ if there exists $\omega \in \Omega$ such that $x_i = X_i(\omega)$ for all $i \in \mathbb{Z}$. Moreover, we will tacitly assume that ω belongs to some "good" subset of Ω of probability 1. For example, for **R**, our standing assumption will be that ω realizing **r** belongs to the set where **Y** visits 1 infinitely often in both directions. In general, if some property of a process **X** has probability 1, then realization **x** inherits it. For example, if we consider **Y** under $\mathbb{P}_{Y_0=1}$ then every realization **y** will satisfy $y_0 = 1$.

The main technical result contains entropy formulas for *good* processes.

Theorem 1.7 (answer to Question 4(A)) *Let* $\mathbf{X} = (X_n)_{n \in \mathbb{Z}}$, $\mathbf{Y} = (Y_n)_{n \in \mathbb{Z}}$ *be a pair of mutually independent stationary processes, where* **X** *is finitely valued and* **Y** *is binary and such that* $\mathbb{P}(Y_0 = 1) > 0$. *Assume also that* **Y** *is such that the definition of the corresponding return process* **R** *to state* 1 *makes sense. Then*

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

If additionally **Y** *is ergodic then*

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

Remark 1.8 The above expectations are to be understood in the following way:

- we compute $\mathbf{H}(X_0 | X_{\{r_{-1}, r_{-2}, \cdots\}})$ or $\mathbf{H}(X_{[1,r_1-1]} | X_{(-\infty,0]}, X_{\{r_1, r_2, \cdots\}})$ for all realizations $\mathbf{r} = (r_i)_{i \in \mathbb{Z}}$ thus obtaining a function $f(\mathbf{r})$ depending on $\mathbf{r}$;
- we find $\mathbb{E}_{Y_0=1} f(\mathbf{R})$.

1.4 Consequences of the main technical result

Clearly, Theorem 1.7 gives an answer to Questions 4(A) and 5(A). We will say now how it is related to Questions 4(B), 5 (B), 4(C) and 5(C). The details and longer proofs are included in Sect. 2.3.

1.4.1 Answer to Questions 4(B) and 5(B)

Notice first that

$$\mathbf{H}(\mathbf{X}) \leqslant \mathbf{H}(X_0 | X_{\{r_{-1}, r_{-2}, \cdots\}}) \leqslant \mathbf{H}(X_0)$$

for each choice of negative integers $\cdots < r_{-2} < r_{-1} < 0$. Therefore, by Theorem 1.7 (A), we obtain immediately the following:

Corollary 1.9 (positive answer to Question 4(B)) *Suppose that* $(\mathbf{X}, \mathbf{Y})$ *is* good, *i.e.* $\mathbf{X} = (X_n)_{n \in \mathbb{Z}}$, $\mathbf{Y} = (Y_n)_{n \in \mathbb{Z}}$ *is a pair of mutually independent stationary processes,*

where $\mathbf{X}$ *is finitely valued and* $\mathbf{Y}$ *is binary, such that* $\mathbb{P}(Y_0 = 1) > 0$ *and the definition of the corresponding return process* $\mathbf{R}$ *to state* 1 *makes sense. Assume additionally that* $\mathbf{H}(\mathbf{Y}) = 0$. *Then*

$$\mathbb{P}(Y_0 = 1)\, \mathbf{H}(\mathbf{X}) \leqslant \mathbf{H}(\mathbf{M}) \leqslant \mathbb{P}(Y_0 = 1)\, \mathbf{H}(X_0).$$

In particular,

$$\mathbf{H}(\mathbf{M}) > 0 \text{ whenever } \mathbf{H}(\mathbf{X}) > 0. \tag{1.5}$$

Remark 1.10 The lower bound in Corollary 1.9 is attained for exchangeable processes (see Proposition 2.5), whereas the upper bound is attained for i.i.d. processes. If $\mathbf{X}$ is a Markov chain (which is not i.i.d.), both inequalities are strict, see Sect. 2.1.2.

Remark 1.11 (positive answer to Question 5(B)) Implication (1.5) means, in particular, that the answer to Question 5(B) is positive, whenever $\nu_\eta \neq \delta_{(\dots 0,0,0\dots)}$. In Sect. 1 we present an alternative ergodic-theoretic approach to this problem. The proof presented therein is much shorter, on the other hand it addresses directly Question 5(B), without providing any explicit formulas.

Remark 1.12 If one drops the assumption that $\mathbf{X} \amalg \mathbf{Y}$ then the situation changes completely and one can get $\mathbf{H}(\mathbf{M}) = 0$ (with $\mathbf{H}(\mathbf{X}) > 0$ and $\mathbb{P}(Y_0 = 1) > 0$). To see how far this can go, consider

$$\mathbf{X} = \mathbf{Z} \cdot \mathbf{W} \text{ and } \mathbf{Y} = \mathbf{1} - \mathbf{W} = (1 - W_i)_{i \in \mathbb{Z}},$$

where

$$\mathbf{Z} \amalg \mathbf{W}, \mathbf{H}(\mathbf{W}) = 0 \text{ and } \mathbb{P}(W_0 = 0) \cdot \mathbb{P}(W_0 = 1) > 0.$$

Then $\mathbf{M}$ is a trivial zero process, in particular, we have $\mathbf{H}(\mathbf{M}) = 0$. On the other hand, by Corollary 1.9, $\mathbf{H}(\mathbf{X}) = \mathbf{H}(\mathbf{Z} \cdot \mathbf{W}) > 0 = \mathbf{H}(\mathbf{W}) = \mathbf{H}(\mathbf{Y})$. Cf. also Sect. 1 for more examples of ergodic-theoretic flavour.

1.4.2 Answer to Questions 4(C) and 5(C)

Answers to Questions 4(C) and 5(C) are more complex and they are related to the notion of a bilaterally deterministic process.

Definition 1.13 We say that a stationary process $\mathbf{Z} = (Z_i)_{i \in \mathbb{Z}}$ is *bilaterally deterministic* if, for all $k \in \mathbb{N}$,

$$\mathbf{H}(Z_{[0,k]} \mid Z_{(-\infty,-1]}, Z_{[k+1,\infty)}) = 0.$$

Remark 1.14 The notion of a bilaterally deterministic process was introduced by Ornstein and Weiss [32], in terms of the following (double) tail sigma-algebra:

$$\mathcal{T}_d := \bigcap_{n \geqslant 1} \sigma\left(Z_{(-\infty,-n]}, Z_{[n,\infty)}\right).$$

Notice that the following conditions are equivalent:

- **Z** is bilaterally deterministic,
- $Z_{[-k,k]} \in \mathcal{T}_d$ for each $k \geqslant 1$,
- $\sigma(\mathbf{Z}) = \mathcal{T}_d$.

Indeed, e.g., if **Z** is bilaterally deterministic then $\mathbf{H}\left(Z_{[0,k]} \mid Z_{(-\infty,-\ell]}, Z_{[k+1+m,\infty)}\right) = 0$ for any $k, \ell, m \in \mathbb{N}$ and by taking $\ell, m \to \infty$, we easily obtain $Z_{[-k,k]} \in \mathcal{T}_d$ for each $k \geqslant 1$. Cf. also Remark 1.4. Informally, "given the far past and the distant future, the present can be reconstructed" [32].

Remark 1.15 Given a stationary finitely-valued process **Z**, let

$$\mathcal{T}_p := \bigcap_{n \geqslant 1} \sigma\left(Z_{(-\infty,-n]}\right), \qquad \mathcal{T}_f := \bigcap_{n \geqslant 1} \sigma\left(Z_{[n,\infty)}\right)$$

denote, respectively, the tail σ-algebra corresponding to the past and to the future. By a celebrated result of Pinsker [34], $\mathcal{T}_p \overset{\mathbb{P}}{=} \mathcal{T}_f \overset{\mathbb{P}}{=} \Pi$, where Π denotes the Pinsker σ-algebra (i.e., the largest zero entropy sub-σ-algebra). Thus, the following conditions are equivalent (cf. Remark 1.14):

- $\mathbf{H}(\mathbf{Z}) = 0$,
- $Z_{[-k,k]} \in \mathcal{T}_p$ for each $k \geqslant 1$,
- $\sigma(\mathbf{Z}) = \mathcal{T}_p$.

A direct consequence of Remark 1.14 and Remark 1.15 is the following observation:

Corollary 1.16 *Suppose that* $\mathbf{H}(\mathbf{Z}) > 0$. *Then* **Z** *is not bilaterally deterministic whenever* $\mathcal{T}_d = \mathcal{T}_p$. *In particular, this happens if* $\mathcal{T}_d$ *is trivial.*

Notice that from this point of view, stationary processes can be split into three pairwise disjoint classes:

(a) of zero entropy rate (they are automatically bilaterally deterministic),
(b) of positive entropy rate that are bilaterally deterministic,
(c) of positive entropy rate but not bilaterally deterministic.

Class (c) includes the following positive entropy rate processes:

- exchangeable processes,
- Markov chains,
- weakly Bernoulli processes (here $\mathcal{T}_d$ is trivial),

for more details, see Sect.on 2.1. Theorem 1.7 allows us to "compare" a large subclass of processes from class (a) with processes from class (c), see Corollaries 1.17 and 1.19 below. In particular, the zero entropy class that we have in mind contains **all** $\mathscr{B}$-free systems (considered with the Mirsky measure), cf. Proposition 1.20 and Corollary 1.21. We leave as an open problem to find answers to analogous questions on the relations between class (a) with class (b).

Notice that

$$
\begin{aligned}
\mathbb{E}_{Y_0=1} &\mathbf{H}\left(X_{[1,r_1-1]} \mid X_{(-\infty,0]}, X_{\{r_1,r_2,\dots\}}\right)_{|r_i=R_i} \\
&\geqslant \mathbb{E}_{Y_0=1}\mathbf{H}\left(X_{[1,r_1-1]} \mid X_{(-\infty,0]}, X_{[r_1,\infty)}\right)_{|r_1=R_1} \\
&= \sum_{k\geqslant 1} \mathbb{P}_{Y_0=1}(R_1=k+1)\mathbf{H}\left(X_{[1,k]} \mid X_{(-\infty,0),X_{[k+1,\infty)}}\right).
\end{aligned}
$$

Moreover, if $\mathbf{X}$ fails to be bilaterally deterministic, then, for all k sufficiently large, we have

$$
\mathbf{H}\left(X_{[1,k]} \mid X_{(-\infty,0]}, X_{[k+1,\infty)}\right) > 0. \tag{1.6}
$$

Thus, using Theorem 1.7(B), we obtain immediately the following:

Corollary 1.17 *(answer to Question 4(C)) Suppose that* $(\mathbf{X}, \mathbf{Y})$ *is* good *and* $\mathbf{Y}$ *is ergodic of zero entropy rate (i.e.* $\mathbf{X}$ *and* $\mathbf{Y}$ *is a pair of mutually independent stationary processes,* $\mathbf{X}$ *is finitely-valued,* $\mathbf{Y}$ *is binary and ergodic, with* $\mathbf{H}(\mathbf{Y}) = 0$*). If additionally*

$$
\mathbb{P}\left(R_1 = k\right) > 0 \text{ for infinitely many } k \in \mathbb{N} \tag{1.7}
$$

and $\mathbf{X}$ *is not bilaterally deterministic then* $\mathbf{H}\left(\mathbf{M}\right) < \mathbf{H}\left(\mathbf{X}\right)$.

Remark 1.18 In fact, if we know more about $\mathbf{X}$ than just (1.6) then the assumption that $\mathbb{P}\left(R_1 = k\right) > 0$ for infinitely many $k \in \mathbb{N}$ can be relaxed and we can still have $\mathbf{H}\left(\mathbf{M}\right) < \mathbf{H}\left(\mathbf{X}\right)$. E.g. if $\mathbf{X}$ is Bernoulli then we will always have $\mathbf{H}\left(\mathbf{M}\right) < \mathbf{H}\left(\mathbf{X}\right)$ whenever $(\mathbf{X}, \mathbf{Y})$ is good ($\mathbf{X}, \mathbf{Y}$ is a pair of mutually independent stationary processes, $\mathbf{X}$ is finitely-valued, $\mathbf{Y}$ is binary and such that the definition of the corresponding return process $\mathbf{R}$ to state 1 makes sense) and $\mathbf{Y}$ is of zero entropy rate.

A natural question arises what happens when (1.7) fails to hold. Suppose that our processes are of dynamical origin and the underlying dynamical system is a transitive symbolic dynamical system. Namely, take $\mathbf{w} \in \{0, 1\}^{\mathbb{Z}}$ such that the support of $\mathbf{w}$ is unbounded both from below and from above, and suppose that $\mathbf{w}$ is quasi-generic along some subsequence for an invariant zero entropy measure ν. Let Y be the orbit closure of $\mathbf{w}$ under the left shift S and let $\mathbf{Y} \sim \nu$ be the corresponding stationary process. Clearly,

$$(1.7) \implies \text{the support of } \mathbf{w} \text{ does not contain a two-sided infinite arithmetic progression.}$$

It turns out that if we assume that the support of $\mathbf{w}$ does contain a two-sided (infinite) arithmetic progression then one can obtain a complementary result to Corollary 1.17:

Corollary 1.19 *Let* **Y** *be a good, ergodic process (i.e.* **Y** *is a stationary binary ergodic process, with* $\mathbb{P}(Y_0 = 1) > 0$*) of zero entropy rate. Assume additionally that there exists* $L \geqslant 1$ *such that for a.e. realization* **y***, the corresponding return time sequence* **r** *contains an arithmetic progression of difference* L*. Then there exists a stationary binary process* **X** *that is not bilaterally deterministic and such that* $\mathbf{H}(\mathbf{M}) = \mathbf{H}(\mathbf{X})$*.*

Let us turn now to the interpretation of Corollaries 1.17 and 1.19 from the point of view of $\mathscr{B}$-free systems. Recall that a topological dynamical system (T, X) is said to be *proximal* whenever for any $x, y \in X$ there exists $n_k \to \infty$ such that $d(T^{n_k}x, T^{n_k}y) \to 0$. It turns out that in $\mathscr{B}$-free setting we have the following dychotomy:

Proposition 1.20 *Let* $\mathscr{B} \subset \mathbb{N}$ *and let* η *be the characteristic function of the corresponding* $\mathscr{B}$*-free set. Then exactly one of the following holds:*

- (X_η, S) *is proximal and then for infinitely many* $k \geqslant 1$ *the block of the form* $10 \ldots 01$ *(with* k *zeros between the 1's) is of positive Mirsky measure* ν_η*,*
- (X_η, S) *is not proximal and then* η *contains a two-sided infinite arithmetic progression.*

As a direct consequence of Corollaries 1.17, 1.19 and Proposition 1.20, for $\mathscr{B}$-free systems we have the following result:

Corollary 1.21 *Let* $\mathscr{B} \subset \mathbb{N}$*. Then* (X_η, S) *is proximal if and only if for any* **X** *that is not bilaterally deterministic, such that* $\mathbf{X} \sqcup \mathbf{Y}$*, we have* $\mathbf{H}(\mathbf{M}) < \mathbf{H}(\mathbf{X})$*.*

Finally, let us remark that **X** in Corollary 1.19 can be chosen to be very weakly Bernoulli (i.e. as a dynamical system, isomorphic to a Bernoulli process [31]) (compare Example 2.16 below and Remark 1.18). That is, for **Y** as in Corollary 1.19, we can find a measure-theoretic dynamical system $(X, \mathcal{B}, \mu, T)$ with two stochastic representations **X** and **X**′ (both not bilaterally deterministic!) such that $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) < \mathbf{H}(\mathbf{X}) = \mathbf{H}(\mathbf{X}') = \mathbf{H}(\mathbf{X}' \cdot \mathbf{Y})$. More than that, in some cases **X**′ can be retrieved from $\mathbf{X}' \cdot \mathbf{Y}$. This matches well with the fact that the notion of a bilaterally deterministic process is not stable under taking various process representation of a given dynamical system [32]. It makes the situation completely different from the one in [18], where the results are purely ergodic-theoretic.

1.5 Dictionary between ergodic theory and probability theory

In our paper, both ergodic-theoretic and stochastic questions and tools are often intertwined. Let us now give some samples of ergodic-theory results translated into the language of stochastic processes. Our basic object is an ergodic-theoretic dynamical system $(\mathcal{X}^{\mathbb{Z}}, \mu, S)$, where S stands, as usual, for the left shift, together with a subset $A \subset X$ satisfying $\mu(A) > 0$. Recall that for $x \in A$, the *first return time* n_A is defined as $n_A(x) = \inf\{n \geqslant 1 \mid S^n x \in A\}$ and the corresponding *induced transformation* as $S_A(x) = S^{n_A(x)}(x)$, with the corresponding conditional measure μ_A being invariant under S_A.

Fix now a stationary process $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, with distribution μ, i.e. $\mathbf{X} \sim \mu$. This is a stochastic counterpart of $(\mathcal{X}^{\mathbb{Z}}, \mu, S)$, cf. also Sect. 1.1. Left shift S

naturally acts on processes by $S\mathbf{X} = (X_{i+1})_{i\in\mathbb{Z}}$. In particular,

$$S_*(\mu) = \mu \text{ precisely if } S\mathbf{X} \sim \mathbf{X}.$$

Similarly, μ_A corresponds to the distribution of $\mathbf{X}$ **under** $\mathbb{P}_{\mathbf{X}\in A}$. To see how one should interpret S_A in terms of stochastic processes, let $R_A = \inf\{n \geqslant 1 \mid S^n\mathbf{X} \in A\}$ be the first return time, defined on $\mathbf{X} \in A$, cf. (1.4). Now, we set $S_A\mathbf{X} = (X_{i+R_A})_{i\in\mathbb{Z}}$ and one can easily check that

$$S_A\mu_A = \mu_A \text{ precisely if } S_A\mathbf{X} \sim \mathbf{X} \text{ under } \mathbb{P}_{\mathbf{X}\in A}.$$

Finally, recall that $h(\mu) = \mathbf{H}(\mathbf{X})$.

Let us present a summary of some classical ergodic theorems (formulated for $(\mathcal{X}^{\mathbb{Z}}, \mu, S)$), with their counterparts for random processes.

	Ergodic	Probabilistic
Ergodicity of μ[1]	$\frac{1}{n}\sum_{i=0}^{n-1} S^i f \to \int f \, d\mu$	$\frac{1}{n}\sum_{i=0}^{n-1} f(S^i\mathbf{X}) \to \mathbb{E}f(\mathbf{X})$
Poincaré Rec.	$\mu_A\left(\{x : S^k x \in A \text{ i.o.}\}\right) = 1$	$\mathbb{P}_{\mathbf{X}\in A}(S^k\mathbf{X} \in A \text{ i.o.}) = 1$
Kac's Lemma	$\int_A n_A d\mu_A = 1$	$\mathbb{P}(\mathbf{X} \in A)\mathbb{E}_{\mathbf{X}\in A} R_A = 1$
Invariance of μ_A	$S_A\mu_A = \mu_A$	$S_A\mathbf{X} \sim \mathbf{X}$, under $\mathbb{P}_{\mathbf{X}\in A}$
Ergodicity of μ_A	$\frac{1}{n}\sum_{i=0}^{n-1} S^i f \to \int f \, d\mu_A$	$\frac{1}{n}\sum_{i=0}^{n-1} f(S_A^i\mathbf{X}) \to \mathbb{E}_{\mathbf{X}\in A}f(\mathbf{X})$
Maker's ET[2]	$\frac{1}{n}\sum_{i=0}^{n-1} S^i f_{n-i} \to \int f \, d\mu$	$\frac{1}{n}\sum_{i=0}^{n-1} f_{n-i}(S^i\mathbf{X}) \to \mathbb{E}f(\mathbf{X})$

[1] Here, in fact, we state Birkhoff ergodic theorem under the assumption that μ is ergodic
[2] ET stands for "ergodic theorem"

We owe the reader a word of explanation concerning the abbreviations in the table above. The convergence of ergodic averages is always meant a.e. / a.s. with respect to the appropriate underlying measure (μ or μ_A / $\mathbb{P}$ or $\mathbb{P}_{\mathbf{X}\in A}$). Also, we tacitly assume that all required assumptions are satisfied, e.g. functions appearing in ergodic averages are integrable with respect to the underlying measure. Finally, let us give some details concerning Maker's ergodic theorem [29] which will play a central role in the proof of Theorem 1.7 (A). We recall it now (in the ergodic-theoretic language, i.e. as in [25], under the extra assumption that T is ergodic).

Theorem 1.22 (*Maker's ergodic theorem*) *Let (X, μ, T) be an ergodic measure-theoretic dynamical system. Let $f \in L_1(\mu)$ and $f_n \to f$ μ-a.e. Suppose that $\sup_n |f_n| \in L_1(\mu)$. Then*

$$\frac{1}{n}\sum_{i=0}^{n-1} T^i f_{n-i} \to \mathbb{E}_\mu f \text{ a.e.}$$

Let us now return to our general setting, with standing assumptions (i) and (ii) on **X** and **Y**. Consider the *inter-arrival process* $\mathbf{T} = (T_i)_{i \in \mathbb{Z}}$, where

$$T_i = R_i - R_{i-1} \tag{1.8}$$

and the *return-process* **R** is as in (1.4). Thus, T_i tells us how much time elapses between $(i-1)$'th and i'th visit of **Y** to the state 1.

Remark 1.23 (Factor of a random process) Recall that whenever **Y** is ergodic, the return process **R** and thus also **T** is well-defined. Moreover, **T** can be regarded as a **factor** of **Y** in the ergodic-theoretic sense. More precisely, by the very definition of **T**, there is a natural measurable function $\pi \colon \left(\{0, 1\}^{\mathbb{Z}}, S_{[1]}, \mathcal{L}(\mathbf{X} | \mathbb{P}_{Y_0=1})\right) \to \left(\mathbb{Z}^{\mathbb{Z}}, S, \mathcal{L}(\mathbf{T} | \mathbb{P}_{Y_0=1})\right)$ such that $\pi(\mathbf{X}) = \mathbf{T}$ almost surely, where $\mathcal{L}(\cdot \mid \cdot)$ stands for the "distribution of $\cdot$ under $\cdot$", $[1] = \{\mathbf{y} \mid y_0 = 1\}$ and $S_{[1]}$ is the corresponding induced shift operator (cf. the beginning of this section). Clearly, $\pi S_{[1]} = S\pi$. In particular, since $\mathbf{Y} \sim S_{[1]}\mathbf{Y}$ and **Y** is ergodic (under $\mathbb{P}_{Y_0=1}$), we get that **T** is stationary and ergodic (under $\mathbb{P}_{Y_0=1}$) as well.

As a consequence of the above remark, we can apply Maker's ergodic theorem to **T**, which results in the following corollary:

Corollary 1.24 *Suppose that* $\sup_{i \in \mathbb{N}} g_i(\mathbf{T}) \in L_1(\mathbb{P}_{Y_0=1})$ *and* $g_i \xrightarrow{\mathbb{P}_{Y_0=1} \ a.s.} g$. *Then,* $\mathbb{P}_{Y_0=1}$ *a.s.,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} g_{n-i}\left((T_{i+j})_{j \in \mathbb{Z}}\right) = \mathbb{E}_{Y_0=1} g(\mathbf{T}).$$

2 Examples, comments and proofs

2.1 Examples of non-bilaterally deterministic processes

In the subsections below we tacitly assume that $(\mathbf{X}, \mathbf{Y})$ is *good*, i.e. **X**, **Y** is a pair of mutually independent stationary processes, where **X** is finitely-valued, **Y** is binary, with $\mathbb{P}(Y_0 = 1) > 0$ and such that the definition of the corresponding return process **R** to state 1 makes sense.

2.1.1 Exchangeable processes

Definition 2.1 We say that a process **X** is *exchangeable* if for any $n \in \mathbb{N}$ and distinct times $i_1, i_2, \ldots, i_n$,

$$\left(X_{i_1}, X_{i_2}, \ldots, X_{i_n}\right) \sim (X_1, X_2, \ldots, X_n).$$

In other words, the distribution of **X** is invariant under finite permutations.

Remark 2.2 Let $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ be exchangeable. By a celebrated result of de Finetti [10] (cf. also [24]), this is equivalent to $\mathbf{X}$ being a convex combination of i.i.d. processes. Thus, there exists a random variable Θ such that, conditionally on Θ, $\mathbf{X}$ is i.i.d. Note that this ensures that $\mathbf{H}(\mathbf{X}) > 0$, unless $X_i = f_i(\Theta)$ for some Borel functions f_i. Indeed,

$$\mathbf{H}(X_1, \ldots, X_n) \geqslant \mathbf{H}(X_1, \ldots, X_n)[\Theta] = \sum_{i=1}^{n} \mathbf{H}(X_i \mid \Theta) = n\mathbf{H}(X_1 \mid \Theta),$$

which gives $\mathbf{H}(\mathbf{X}) \geqslant \mathbf{H}(X_1 \mid \Theta)$. Therefore, $\mathbf{H}(\mathbf{X}) = 0$ implies $X_i = f_i(\Theta)$.

Remark 2.3 Olshen in [30] showed that if $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ is exchangeable then

$$\mathcal{I} = \mathcal{E} = \mathcal{T}_d = \mathcal{T}_f = \mathcal{T}_p,$$

(as measure-algebras), where $\mathcal{I}, \mathcal{E}$ denote the σ-algebra of shift-invariant and finite permutation invariant sets respectively and $\mathcal{T}_d, \mathcal{T}_f, \mathcal{T}_p$ are double, future, past tails respectively.

As an immediate consequence of Remark 2.3 and Corollary 1.16, we obtain the following:

Corollary 2.4 *Suppose that* $\mathbf{X}$ *is exchangeable. Then* $\mathbf{H}(\mathbf{X}) > 0$ *if and only if* $\mathbf{X}$ *is not bilaterally deterministic.*

Proposition 2.5 *Suppose that* $\mathbf{X}$ *is exchangeable. Then* $\mathbf{H}(\mathbf{M} \mid \mathbf{Y}) = \mathbb{P}(Y_0 = 1)\mathbf{H}(\mathbf{X})$.

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

$$\mathbf{H}(X_0 \mid X_{\{r_{-1}, r_{-2}, \ldots\}}) = \mathbf{H}(X_0 \mid X_{\{-1, -2, \ldots\}}) = \mathbf{H}(\mathbf{X})$$

It remains to use Theorem 1.7 (A).   □

2.1.2 Markov chains

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 $X_{[i+1, \infty)}$. Colloquially, given present, the past and the future are independent. This immediately leads to the following corollary of Theorem 1.7 (A):

Corollary 2.6 *If* $\mathbf{X}$ *is a Markov chain (and* $(\mathbf{X}, \mathbf{Y})$ *is good) then*

$$\mathbf{H}(\mathbf{M}) = \mathbb{P}(Y_0 = 1) \sum_{k=1}^{\infty} \mathbb{P}_{Y_0} = 1(R_1 = k)\mathbf{H}(X_k \mid X_o).$$

Remark 2.7 Corollary 2.6 easily extends to the case of k-Markov chains but for simplicity sake we decided to present it for $k = 1$.

Remark 2.8 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 [15], Chapter XV, Section 6, Theorem 3, page 392) that we can uniquely decompose the state space $\mathcal{X}$ into disjoint union

$$\mathcal{X} = C \sqcup D_1 \sqcup D_2 \sqcup \cdots \sqcup D_k, \tag{2.1}$$

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 [15], Chapter XV, Section 7, Criterion, page 395), if π is a stationary measure then necessarily $\pi(C) = 0$.

Now suppose that a bilateral, finitely-valued Markov chain $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ is stationary (thus, $C = \emptyset$ in (2.1)). Fix $1 \leqslant j \leqslant k$ 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 [6], 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 [3], 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 [3] 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

$$\mathcal{T}_d(\mathbf{X}) = \mathcal{T}_p(\mathbf{X}) = \mathcal{T}_f(\mathbf{X}) = \sigma\left\{\{X_0 \in D_{j,\ell}\} \mid 1 \leqslant j \leqslant k, 0 \leqslant \ell \leqslant p_j\right\}. \tag{2.2}$$

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 (2.2). As a consequence of (2.2), we obtain the following:

Corollary 2.9 *Suppose that $\mathbf{X}$ is a stationary finitely-valued Markov-chain. Then $\mathbf{H}(\mathbf{X}) > 0$ if and only if $\mathbf{X}$ is not bilaterally deterministic.*

Remark 2.10 Since $\mathbf{H}(\mathbf{X}) = \mathbf{H}(X_1 \mid X_0) = \mathbf{H}(X_{i+1} \mid X_i)$, it follows that $\mathbf{H}(\mathbf{X}) = 0$ if and only if, for every $i \in \mathbb{Z}$, $X_i = f_i(X_0)$ for some functions f_i. It is not hard to see that if for every $x \in \mathcal{X}$, $\mathbb{P}(X_0 = x) > 0$, then every f_i must be a bijection on $\mathcal{X}$. Moreover, by the stationarity of $\mathbf{X}$, for $f_1(x) = y$, we get

$$\begin{aligned}
\mathbb{P}(X_0 = x) &= \mathbb{P}(X_0 = x, f_1(x) = y) = \mathbb{P}(X_0 = x, f_1(X_0) = y) \\
&= \mathbb{P}(f_i(X_0) = x, f_{i+1}(X_0) = y) = \mathbb{P}(X_0 = f_i^{-1}(x)) \mathbb{1}_{f_{i+1}\left(f_i^{-1}(x)\right) = f_1(x)}.
\end{aligned}$$

Thus, necessarily, $f_{i+1}(z) = f_1(f_i(z))$. Consequently, if we set $f := f_1$ then $f_i = f^{\circ i}$. Moreover, f must be such that, for all x, $\mathbb{P}(X_0 = x) = \mathbb{P}(X_0 = f(x))$.

Therefore, if **X** is bilateral, finitely-valued, stationary Markov chain, with $\mathbb{P}(X_0 = x) > 0$ for all $x \in \mathcal{X}$, then the following are equivalent:

- **X** is bilaterally deterministic;
- there exist a bijection $f : \mathcal{X} \to \mathcal{X}$, such that $X_i = f^{\circ i}(X_0)$ and for all $x \in \mathcal{X}$, $\mathbb{P}(X_0 = x) = \mathbb{P}(X_0 = f(x))$.

2.1.3 Weakly Bernoulli processes

Weakly Bernoulli processes were introduced by Friedman and Ornstein [16] and belong to the classics of ergodic theory. Equivalently, one speaks of finitely determined processes. Recall that any process **X** that is weakly Bernoulli is also very weakly Bernoulli (i.e. as a dynamical system, it is isomorphic to a Bernoulli process [31]). In particular, $\mathbf{H}(\mathbf{X}) > 0$. We refer the reader, e.g., to [36] for more information on the subject.

Suppose now that **X** is weakly Bernoulli. Then $\mathcal{T}_d$ is trivial (see, e.g., Proposition 5.17 in [4]). Therefore, as an immediate consequence of Corollary 1.16, we obtain the following:

Corollary 2.11 *Suppose that* **X** *is weakly Bernoulli. Then* **X** *is not bilaterally deterministic.*

In fact, the results in [4] are formulated in a different language. One more notion, equivalent to the weak Bernoulli property, is absolute regularity. It first appeared in works of Volkonskii and Rozanov [37, 38] who, in turn, attribute it to Kolmogorov. Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ be sub-σ-algebras and let

$$\beta(\mathcal{A}, \mathcal{B}) := \sup \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|,$$

where the supremum is taken over all (finite) partitions $\{A_1, \ldots, A_I\}, \{B_1, \ldots, B_J\}$ of Ω, with $A_i \in \mathcal{A}, B_j \in \mathcal{B}$ for each i, j. Now, given a process **X**, for $-\infty \leqslant J < L \leqslant \infty$, we define the σ-algebra

$$\mathcal{F}_J^L := \sigma(X_k : J \leqslant k \leqslant L).$$

Then, for each $n \geqslant 1$, we define the following β-dependence coefficients:

$$\beta(n) := \sup_{j \in \mathbb{Z}} \beta(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty).$$

We say that **X** is *absolutely regular* (or β-*mixing*) if $\beta(n) \to 0$ as $n \to \infty$.

Berbee, in [1], studied β-dependence coefficients for stationary ergodic processes. He showed that

$$\lim_{n \to \infty} \beta(n) = \beta = 1 - \frac{1}{p} \text{ for some } p \in \mathbb{N} \cup \{\infty\}.$$

Moreover, he proved that if $\beta < 1$ then $\mathcal{T}_d = \mathcal{T}_p$. As a consequence of his result and of Corollary 1.16, we have:

Corollary 2.12 *Suppose that* $\mathbf{X}$ *is a stationary ergodic process with* $\beta < 1$*. Then* $\mathbf{X}$ *is not bilaterally deterministic.*

2.2 Proof of the main technical result (Theorem 1.7)

2.2.1 Part (A)

By the chain rule (cf. (1.3)), we have

$$\mathbf{H}\left(M_{[0,n]} \mid Y_{[0,n]}\right) = \sum_{k=0}^{n} \mathbf{H}\left(M_k \mid Y_{[0,n]}, M_{[0,k)}\right) =: \sum_{k=0}^{n} H_{k,n}. \tag{2.3}$$

Fix $0 \leqslant k \leqslant n$. Since $M_k = X_k \cdot Y_k$ and $\mathbf{X} \amalg \mathbf{Y}$, we easily get that conditionally on $\left(Y_{[0,k]}, M_{[0,k)}\right)$, M_k is independent of $Y_{[k+1,n]}$. In other words,

$$H_{k,n} = H_k = \mathbf{H}\left(M_k \mid Y_{[0,k]}, M_{[0,k)}\right).$$

Now, using the definition of Shannon conditional entropy, the fact that on the event $Y_k = 0$, we have $M_k \equiv 0$, whereas on $Y_k = 1$, we have $M_k = X_k$ and the stationarity of the $(\mathbf{X}, \mathbf{Y})$, we get

$$H_k = \mathbb{P}\left(Y_k = 1\right)\mathbf{H}_{Y_k=1}\left(X_k \mid Y_{[0,k)}, M_{[0,k)}\right)$$
$$= \mathbb{P}\left(Y_0 = 1\right)\mathbf{H}_{Y_0=1}\left(X_0 \mid Y_{[-k,0)}, M_{[-k,0)}\right).$$

Moreover, if $Y = Y_{[-k,0)}$, $M = M_{[-k,0)}$, $y = y_{[-k,0)}$, $m = m_{[-k,0)}$, $s_{-k} = \sum_{i=-k}^{-1} y_i$, $r_{-s_{-k}} < \cdots < r_{-1}$ are such that $y_{r_{-i}} = 1$, then

$$\mathbb{P}_{Y_0=1}\left(Y = y, M = m\right) = \begin{cases} \mathbb{P}_{Y_0=1}\left(Y = y\right)\mathbb{P}\left(X_{\{r_{-1},\ldots,r_{-s_{-k}}\}} = m_{\{r_{-1},\ldots,r_{-s_{-k}}\}}\right), & s_{-k} > 0, \\ \mathbb{P}_{Y_0=1}(Y = y), & s_{-k} = 0, \end{cases}$$

whenever $m \leqslant y$ coordinatewise (otherwise, we get zero). This implies that

$$H_k = \mathbb{P}(Y_0 = 1)\,\mathbb{P}_{Y_0=1}\left(S_{-k} = 0\right)\mathbf{H}\left(X_0\right)$$
$$+ \mathbb{P}\left(Y_0 = 1\right)\mathbb{E}_{Y_0=1}\mathbb{1}_{S_{-k}>0}\mathbf{H}\left(X_0 \mid X_{\{r_{-1},\ldots,r_{-s_{-k}}\}}\right)|_{r_{-i}=R_{-i},s_{-k}=S_{-k}}.$$

Since $\mathbf{Y}$ visits 1 a.s. infinitely many times (in the past),

$$\mathbb{P}_{Y_0=1}\left(S_{-k} = 0\right) \to 0 \text{ as } k \to \infty.$$

Moreover, $\mathbb{P}_{Y_0=1}$ a.s., we have $\mathbb{1}_{S_{-k}>0} \to 1$ and

$$\mathbf{H}\left(X_0 \mid X_{\{r_{-1},\ldots,r_{-s_{-k}}\}}\right)|_{r_{-i}=R_{-i},s_{-k}=S_{-k}} \to \mathbf{H}\left(X_0 \mid X_{\{r_{-1},r_{-2},\ldots,\}}\right)|_{r_{-i}=R_{-i}}.$$

Thus, by the bounded convergence theorem, we get that

$$H_k \to \mathbb{P}(Y_0 = 1)\mathbb{E}_{Y_0=1}\mathbf{H}(X_0 \mid X_{\{r_{-1}, r_{-2}, \ldots,\}})|_{r_{-i}=R_{-i}},$$

which, by (2.3), concludes the proof of Theorem 1.7 (A).

2.2.2 Part (B)

First, we will prove a technical lemma.

Lemma 2.13 *We have*

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

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

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

Moreover, if $s_n = \sum_{i=0}^{n} y_i > 0$ then

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

whenever $m_{[0,n]} \leqslant y_{[0,n]}$ coordinatewise (otherwise, we get zero). Hence,

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

which results in

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

This completes the proof. □

Notice now that

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

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

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

More precisely, if we show that

$$\lim_{n\to\infty} H(n, \mathbf{R}) = \mathbb{P}(A_0)\mathbb{E}_{A_0}\mathbf{H}\left(X_{[r_0+1,r_1-1]} \mid X_{(-\infty,r_0]}, X_{\{r_1,r_2,\dots\}}\right)|_{r_i=R_i} \quad (2.4)$$

holds a.e. then by the bounded convergence theorem (as $H(n, \mathbf{R}) \leqslant \mathbf{H}(X_0)$) we will have

$$\lim_{n\to\infty} \mathbb{E}\mathbb{1}_{S_n>0}H(n, \mathbf{R}) = \mathbb{P}(A_0)\mathbb{E}_{A_0}\mathbf{H}\left(X_{[r_0+1,r_1-1]} \mid X_{(-\infty,r_0]}, X_{\{r_1,r_2,\dots\}}\right)|_{r_i=R_i}$$

since $\lim_{n\to\infty} \mathbb{1}_{S_n>0} = 1$ a.e. by the ergodicity of $\mathbf{Y}$.

Let

$$A_i = [Y_0 = \dots = Y_{i-1} = 0, Y_i = 1] \text{ for } i \geqslant 0$$

(in particular, $A_0 = [Y_0 = 1]$).

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,

$$\frac{1}{n}\Sigma_1(n) \leqslant \frac{1}{n}\mathbf{H}\left(X_{[0,r_0-1]}\right) \leqslant \frac{r_0}{n}H(X_0) \to 0 \quad (2.5)$$

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

$$\frac{\Sigma_3(n)}{n} \leqslant \frac{n - r_{s_n-1}}{n}H(X_0) = \left(1 - \frac{r_{s_n-1}}{s_{r_{s_n-1}}} \cdot \frac{s_n}{n}\right)H(X_0) \to 0. \quad (2.6)$$

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

$$\frac{1}{n}\Sigma_2(s_n - 1) = \frac{s_n}{n}\frac{1}{s_n}\Sigma_2(s_n - 1). \quad (2.7)$$

Because of $\frac{s_n}{n} \to \mathbb{P}(Y_0 = 1)$, it suffices to show that $\mathbb{P}_{A_0}$-a.e. we have

$$\lim_{n\to\infty} \frac{1}{n}\Sigma_2(n) = \mathbb{E}_{A_0}\mathbf{H}\left(X_{[r_0+1,r_1-1]} \mid X_{(-\infty,r_0]}, X_{\{r_1,r_2,\dots\}}\right). \quad (2.8)$$

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

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

$$= \sum_{i=0}^{n-1} \mathbf{H}\left(X_{[1, t_{i+1}-1]} \mid X_{[-r_i,0]}, X_{\{t_{i+1},\dots,t_{i+1}+\cdots+t_n\}}\right).$$

We would like to apply Maker's ergodic theorem 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 $r_i \geqslant k$ and for every $j \in \mathbb{N}$, we have

$$H_{\infty,j}(t_{i+1}, t_{i+2}, \dots) \leqslant \mathbf{H}\left(X_{[1, t_{i+1}-1]} \mid X_{[-r_i,0]}, X_{\{t_{i+1},\dots,t_{i+1}+\cdots t_{i+j}\}}\right)$$
$$\leqslant H_{k,j}(t_{i+1}, t_{i+2}, \dots), \tag{2.9}$$

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

$$H_{k,j}(t_1, t_2, \dots) \xrightarrow{j \to \infty} H_k(t_1, t_2, \dots) := \mathbf{H}\left(X_{[1, t_1-1]} \mid X_{(-k,0]}, X_{\{t_1, t_1+t_2,\dots\}}\right)$$
$$= \mathbf{H}\left(X_{[r_0+1, r_1-1]} \mid X_{(-k,r_0]}, X_{\{r_1, r_2,\dots\}}\right).$$

By the entropy chain rule and Kac's lemma,

$$\sup_{k,j \in \mathbb{N}} H_{k,j}(T_{[1,\infty)}) \leqslant \mathbf{H}(X_0)T_1 \in L_1(\mathbb{P}_{A_0}). \tag{2.10}$$

Therefore, Maker's ergodic theorem implies that, for every $k \in \mathbb{N} \cup \{\infty\}$, $\mathbb{P}_{A_0}$ a.s., we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} H_{k,n-i}(t_{i+1}, t_{i+2}, \dots) \to \mathbb{E}_{A_0} H_k(T_1, T_2, \dots). \tag{2.11}$$

Using (2.9), it follows from the definition of Σ_2 (and the chain rule) that

$$\frac{1}{n} \sum_{i=0}^{n-1} H_{\infty,n-i}(t_{i+1}, t_{i+2}, \dots) \leqslant \frac{1}{n} \Sigma_2(n)$$

$$\leqslant \frac{t_1 + \cdots + t_k}{n} H(X_0) + \frac{1}{n} \sum_{i=k}^{n-1} H_{k,n-i}(t_{i+1}, t_{i+2}, \dots) \tag{2.12}$$

$$\leqslant \frac{t_1 + \cdots + t_k}{n} H(X_0) + \frac{1}{n} \sum_{i=0}^{n-1} H_{k,n-i}(t_{i+1}, t_{i+2}, \dots),$$

with $\frac{t_1+\cdots+t_k}{n} H(X_0) \xrightarrow{n\to\infty} 0$. Thus, due to (2.11),

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

Notice that $H_k \to H_\infty$ as $k \to \infty$. Hence, combining (2.10) and the bounded convergence theorem, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \Sigma_2(n) = \mathbb{E}_{A_0} H_\infty (T_1, T_2, \ldots) \tag{2.13}$$

$\mathbb{P}_{A_0}$ a.s. which is exactly (2.8) under $\mathbb{P}_{A_0}$.

It remains to show (2.8) under $\mathbb{P}_{A_i}$ for $i \geqslant 1$. However, it is a direct consequence of the above and the following lemma:

Lemma 2.14 *Suppose that we have a sequence of measurable functions $(f_n)_{n\geqslant 1}$ depending on $(T_n)_{n\geqslant 1}$ and a measurable function f depending on $\mathbf{Y}$ such that*

$$f_n((T_n)_{n\geqslant 1}) \to f(\mathbf{Y}) \tag{2.14}$$

$\mathbb{P}_{A_0}$*-a.e. Then (2.14) holds also $\mathbb{P}_{A_i}$-a.e. for each $i \geqslant 1$.*

Proof For the sake of simplicity, we assume that $\mathbf{Y}$ is a cannonical process. Let $B_0 \subset A_0$ be the set where (2.14) holds. We claim that $B_i := A_i \cap S^{-i} B_0$ is of full measure $\mathbb{P}_{A_i}$ and (2.14) holds on B_i. Indeed, since $S^i A_i \subset A_0$, we have

$$\mathbb{P}_{A_i}(A_i \setminus B_i) = \frac{1}{\mathbb{P}(A_i)} \mathbb{P}(A_i \setminus S^{-i} B_0) = \frac{1}{\mathbb{P}(A_i)} \mathbb{P}(S^i A_i \setminus B_0) \leqslant \frac{1}{\mathbb{P}(A_i)} \mathbb{P}(A_0 \setminus B_0) = 0.$$

Moreover, if $\mathbf{y} \in B_i$ then $S^i \mathbf{y} \in S^i A_i \cap B_0 \subset A_0 \cap B_0 = B_0$. Since $\mathbf{y} \in A_i$, it follows immediately that $T_n(\mathbf{y}) = T_n(S^i \mathbf{y})$ for all $n \geqslant 1$ which completes the proof. $\qquad\square$

2.3 General setting: proof of Corollary 1.19 and related examples

In this section we will study a certain class of *good* $(\mathbf{X}, \mathbf{Y})$ with no entropy drop. We begin by the proof of Corollary 1.19.

Proof of Corollary 1.19 Let $L \geqslant 1$ be such that supp $\mathbf{y} \supset L\mathbb{Z} + a$ for some a and for a.e. realization $\mathbf{y}$ of $\mathbf{Y}$. Let $(X, \mathcal{B}, \mu, T)$ be a measure-theoretic dynamical system with entropy less than $\frac{1}{L} \log 2$ and take a measurable partition $X = J \cup J^c$ that is generating for the map T^L. Let Y be the orbit closure of $\mathbf{y}$ in $\{0, 1\}^{\mathbb{Z}}$ under the left shift.

Process $\mathbf{M}$ corresponds to coding of points in $(X \times Y, T \times S)$ with respect to $J \times C$ (with $C = [1] \subset Y$) and its complement. Using Theorem 1.7 (B), we obtain

$$\mathbf{H}(\mathbf{M}) = \mathbf{H}(\mathbf{X}) - \mathbb{P}(A_0) \mathbb{E}_{A_0} \mathbf{H}(X_{[r_0+1, r_1-1]} \mid X_{(-\infty, r_0]}, X_{\{r_1, r_2, \ldots\}})|_{r_i=R_i} = \mathbf{H}(\mathbf{X}).$$

(a.e. **r** contains a two-sided infinite arithmetic progression with difference L, the partition $\{J, J^c\}$ is generic for T^L and thus the conditional entropy in the above formula is equal to zero). $\qquad\square$

It would be interesting to know if in the above example **X** can be recovered from **M**. Let us see now that this can be the case when **Y** arises from the rotation on two points $\{0, 1\}$. We will look at it both from the probabilitic and ergodic-theoretic perspective.

Example 2.15 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 (relabelling) 1-1 function $F \colon \{0, 1\}^2 \to \{0, 1, 2, 3\}$ and put

$$X_i = F(\xi_i, \xi_{i+1}), \qquad \mathbf{Y} \sim \frac{1}{2}(\delta_a + \delta_{Sa}),$$

where $a := (\dots 0, 1, 0, 1, \overset{\circ}{0}, 1, 0, 1 \dots)$, S stands for the left shift and $\mathbf{X} \amalg \mathbf{Y}$. Since **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

$$\mathbb{E}_{Y_0=1}\mathbf{H}\,(X_0 \mid X_{\{r_{-1}, r_{-2}, \cdots\}})|_{r_i=R_i} = \mathbf{H}\,(X_0 \mid X_{-2}) = \mathbf{H}\,(X_0) = 2\log 2.$$

Clearly, for every $j \in \mathbb{Z}$, $(X_i)_{i \leqslant j} \amalg (X_i)_{i \geqslant j+2}$ yielding

$$\frac{1}{n} f(y_{[0,n]}) = \frac{1}{n}\mathbf{H}\,(X_{r_1}, \dots, X_{r_m}) = \frac{m}{n}\mathbf{H}\,(X_0) \to \frac{1}{2}\mathbf{H}\,(X_0).$$

Thus, by Theorem 1.7 (A), $\mathbf{H}\,(\mathbf{M}) = \frac{1}{2}\mathbf{H}\,(X_0) = \frac{1}{2}2\log 2 = \log 2 = \mathbf{H}\,(\mathbf{X})$. In fact, notice that since F is 1-1, knowing all even (resp. all odd) coordinates of a realization **x** of **X** determines its **all** coordinates. In other words, **M** contains full information about **X**.

We will now see how to use ergodic-theoretic approach to modify the above idea so that $X_i \in \{0, 1\}$ and keep the property $\mathbf{H}\,(\mathbf{M}) = \mathbf{H}\,(\mathbf{X})$ and the ability to recover **X** from **M**.

Example 2.16 Let $(X, \mathcal{B}, \mu, T)$ be an ergodic automorphism, with $h(\mu) \in (0, \log 2)$ and let S be the rotation on $Y = \{0, 1\}$, with the unique invariant measure denoted by ν. Let $\{J, J^c\}$ be a (measurable) generating partition of X for T (the existence of such a partition follows by Krieger's finite generator theorem [27]) and let $C := \{1\} \subset Y$ We consider the following stationary process:

$$\mathbf{X} = (\mathbb{1}_J \circ T^i)_{i \in \mathbb{Z}} \text{ and } \mathbf{Y} = (\mathbb{1}_C \circ S^i)_{i \in \mathbb{Z}}.$$

Then $\mathbf{M} := \mathbf{X} \cdot \mathbf{Y}$ corresponds to coding of points in the dynamical system $(X \times Y, T \times S)$ with respect to $J \times C$ and its complement:

$$(x, 1) \mapsto (\dots, \mathbb{1}_J(x), 0, \mathbb{1}_J(T^2 x), 0, \dots);$$
$$(x, 0) \mapsto (\dots, 0, \mathbb{1}_J(T x), 0, \mathbb{1}_J(T^3 x), \dots).$$

Equivalently, $\mathbf{M}$ corresponds to the dynamical system that is a tower of height two above the factor of T^2 corresponding to the partition $\{J, J^c\}$.

Assume now additionally that $h(T) < \frac{1}{2} \log 2$ and the partition $\{J, J^c\}$ is generating for T^2 (e.g. T can be a Bernoulli automorphism, with entropy less than $\frac{1}{2} \log 2$). Then $\mathbf{M}$ corresponds to a tower of height two above T^2, denoted by R, and given by

$$R(x, 0) = (x, 1), \quad R(x, 1) = (T^2 x, 0).$$

Notice that R is isomorphic to $T \times S$ via the map Φ given by

$$\Phi(x, 0) = (x, 0), \quad \Phi(x, 1) = (T x, 1)$$

(we easily check that $\Phi \circ R = (T \times S) \circ \Phi$). It follows that

$$\mathbf{H}(\mathbf{M}) = h(\mu \otimes \nu) = h(\mu) = \mathbf{H}(\mathbf{X}) > 0. \tag{2.15}$$

In fact, since Φ is an isomorphism, one can filter out $\mathbf{X}$ from $\mathbf{M}$.

2.4 $\mathscr{B}$-free systems: proof of Proposition 1.20

Let $\mathscr{B} \subset \mathbb{N}$, let $\eta = \mathbb{1}_{\mathcal{F}_{\mathscr{B}}}$ and let (X_η, S) be the corresponding $\mathscr{B}$-free system, with the underlying Mirsky measure ν_η. Recall that:

$$h(\widetilde{X}_\eta, S) = \overline{d}(\mathcal{F}_{\mathscr{B}}) = \nu_\eta(1),$$

so $\nu_\eta \neq \delta_{(\dots,0,0,0,\dots)}$ is equivalent to $h(\widetilde{X}_\eta, S) > 0$. Thus, $\nu_\eta \neq \delta_{(\dots,0,0,0,\dots)}$ is necessary and sufficient for the existence of κ with $h(\nu_\eta * \kappa) > 0$.

Proof of Proposition 1.20 It was shown in Theorem 3.7 in [12] that the following are equivalent:

- (X_η, S) is proximal,
- $\mathscr{B}$ contains an infinite pairwise coprime subset,
- the support of η does not contain a two-sided infinite arithmetic progression.

Thus, in order to complete the proof of Proposition 1.20, we need to show that in the proximal case, for infinitely many $k \geqslant 1$ the block of the form $10 \dots 01$ (with k zeros between the 1's) is of positive Mirsky measure ν_η. An important notion in the theory

of $\mathscr{B}$-free systems is that of *tautness* [23], defined in terms of the logarithmic density of sets of multiples. We say that $\mathscr{B}$ is taut if for any $b \in \mathscr{B}$, we have

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

where $\delta(A) = \lim_{N \to \infty} \frac{1}{\log N} \sum_{n \leqslant N} \frac{1}{n} \mathbf{1}_A(n)$ for any $A \subset \mathbb{Z}$. It was proved in [12] (see Theorem C and Lemma 4.11 therein) that given any $\mathscr{B}$, there exists a taut set $\mathscr{B}'$ such that $\mathcal{M}_{\mathscr{B}'} \subset \mathcal{M}_{\mathscr{B}}$ and $\nu_{\eta'} = \nu_{\eta}$. Keller [26] proved that the Mirsky measure of any taut set has full support. Therefore, whenver $\nu_{\eta} = \nu_{\eta'} \neq \delta_{(\dots,0,0,0,\dots)}$ then any block of the form $10 \dots 01$ appearing in η' (and there are infinitely many such blocks as we exclude the Dirac measure at $(\dots, 0, 0, 0, \dots)$!) is in fact of positive ν_{η}-measure. $\qquad \square$

A Ergodic theory viewpoint

A.1 Direct answer to Question 5(B)

Let us first recall the remaining necessary notions from ergodic theory and theory of joinings (for more information, we refer the reader, e.g., to [11, 13, 20]). Given two measure-preserving transformations $(X_i, \mathcal{B}_i, \mu_i, T_i)$, $i = 1, 2$, any $\rho \in \mathcal{M}(X_1 \times X_2, T_1 \times T_2)$ that projects onto μ_1 and μ_2 onto the first and second coordinate, respectively, is called a joining of T_1 and T_2. The set of joinings is always non-empty (it contains the product measure). If $T_1 = T_2$, we speak of self-joinings. The diagonal self-joining of $(X, \mathcal{B}, \mu, T)$ is determined by $\triangle(A \times B) = \mu(A \cap B)$ for $A, B \in \mathcal{B}$. If $(Z, \mathcal{D}, \rho, R)$ is a common factor of T_1 and T_2, then also the set of joinings of T_1 and T_2 that project onto the diagonal self-joining of the common factor is non-empty (it contains the so-called relatively independent extension over the common factor, see [20]).

Proposition A.1 *Assume that* $\nu, \kappa \in \mathcal{M}^e(\{0, 1\}^{\mathbb{Z}}, S)$ *satisfy* $h(\nu) = 0$ *with* $\nu \neq \delta_{(\dots 0,0,0\dots)}$ *and* $h(\kappa) > 0$. *Then* $h(\nu * \kappa) > 0$.

Proof Consider $(\{0, 1\}^{\mathbb{Z}} \times \{0, 1\}^{\mathbb{Z}}, \nu \otimes \kappa, S \times S)$ and denote by $\Pi(\kappa) \subset \mathcal{B}$ the Pinsker σ-algebra of κ. Recall that for (X_i, μ_i, T_i), $i = 1, 2$, we have the corresponding relation between the Pinsker σ-algebras: $\Pi(X_1 \times X_2, \mu_1 \otimes \mu_2, T_1 \times T_2) = \Pi(T_1) \otimes \Pi(T_2)$,

see, e.g. [21]. It follows that

$$\Pi(\nu \otimes \kappa) = \mathcal{B} \otimes \Pi(\kappa). \tag{A.1}$$

Let $C := \{x \in \{0, 1\}^{\mathbb{Z}} : x_0 = 1\}$ and suppose that $h(\nu * \kappa) = 0$, i.e. $\Pi(\nu * \kappa) = \mathcal{B}$. Therefore, additionally using (A.1), we obtain

$$M^{-1}(\mathcal{B}) = M^{-1}(\Pi(\nu * \kappa)) \subset \Pi(\nu \otimes \kappa) = \mathcal{B} \otimes \Pi(\kappa)$$

and it follows that

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

(even though $C \times C = M^{-1}C$ is an equality between sets, we think of it up to sets of measure zero, cf. also Remark 1.5). Hence, for C on the second coordinate in $C \times C$, we have $C \in \Pi(\kappa)$.[12] Since $\{C, C^c\}$ is a generating partition, $\Pi(\kappa) = \mathcal{B}$ (modulo κ) and it follows immediately that $h(\kappa) = 0$.   □

A.2 Simple proof of Theorem 1.3

We begin this section by the following simple but general observation (it overlaps with Theorem 1.3 for uniquely ergodic $\mathcal{B}$-free systems):

Proposition A.2 *Suppose that (Y, S) is a uniquely ergodic subshift of $\{0, 1\}^{\mathbb{Z}}$. Let $\widetilde{Y} = M(Y \times \{0, 1\}^{\mathbb{Z}})$ be the hereditary closure of Y. Then, for any $\nu \in \mathcal{M}^e(\widetilde{Y}, S)$, there exists $\rho \in \mathcal{M}^e(Y \times \{0, 1\}^{\mathbb{Z}}, S \times S)$ such that $M_*(\rho) = \nu$.*

Proof Let $z \in \widetilde{Y}$ be a generic point for ν. Then there exists $y \in Y$ such that $z \leqslant y$. Moreover, y is generic for the unique S-invariant measure on Y. Let $x \in \{0, 1\}^{\mathbb{Z}}$ be such that $M(y, x) = z$. Notice that (y, x) is quasi-generic for some measure $\rho \in \mathcal{M}(Y \times \{0, 1\}^{\mathbb{Z}}, S \times S)$. Moreover, $M_*(\rho) = \nu$ follows directly from the equality $M(y, x) = z$. 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).   □

Remark A.3 The original proof of Theorem 1.3 is much more involved than what we present below. However, it includes much more information about the structure of invariant measures for $(\widetilde{X}_\eta, S)$. E.g. it serves as a tool to prove that $(\widetilde{X}_\eta, S)$ is intrinsically ergodic [12, 28]. Cf. also Remark A.7.

Let now $\mathcal{B} = \{b_k : k \geqslant 1\} \subset \mathbb{N} \setminus \{1\}$ and, for each $K \geqslant 1$, let $\mathcal{B}_K := \{b_1, \ldots, b_K\}$. Set $\eta := \mathbb{1}_{\mathcal{F}_{\mathcal{B}}}$ and $\eta_K := \mathbb{1}_{\mathcal{F}_{\mathcal{B}_K}}$. The Mirsky measure ν_{η_K} (considered on $\widetilde{X}_{\eta_K}$) is the

---

[12] Let H_1, H_2 be Hilbert spaces and let $G_2 \subset H_2$ be a closed subspace. Suppose that $f \otimes g \in H_1 \otimes G_2$, with $f \neq 0$. Let $g = g_0 + g_0'$, with $g_0 \in G_2$ and $g_0' \in G_2^{\perp}$. It follows that $f \otimes g_0' \in H_1 \otimes G_2$. But, on the other hand, we can approximate $f \otimes g_0'$ by tensors of the form $\sum_n \alpha_n f_k \otimes h_k$ with $h_k \in G_2$ which are all orthogonal with $f \otimes g_0'$. This means that $g_0' = 0$ and, thus, we have $g \in G_2$.

purely atomic measure given by the periodic point η_K. Moreover, η_K is a generic point for ν_{η_K} while η is quasi-generic for ν_η, see [12]. Recall also the following classical result of Davenport and Erdös:

Theorem A.4 ([7, 8]) *For any $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$,*

$$\underline{d}(\mathcal{M}_\mathscr{B}) = \lim_{n \to \infty} d(\mathcal{M}_{\mathscr{B}_K}),$$

($\underline{d}$ and d denote the lower density and usual asymptotic density, respectively).[13]

Lemma A.5 *For each $\mathscr{B} = \{b_k : k \geqslant 1\} \subset \mathbb{N} \setminus \{1\}$, $\nu_{\eta_K} \to \nu_\eta$ weakly, as $K \to \infty$.*

Proof It suffices to show that

$$\int f \, d\nu_{\eta_K} \to \int f \, d\nu_\eta$$

for functions f on $\{0, 1\}^\mathbb{Z}$ depending on a finite number (say, L) of coordinates. Let $(N_k)_{k \geqslant 1}$ be an increasing sequence such that $\lim_{k \to \infty} \frac{1}{N_k} |\mathcal{M}_\mathscr{B} \cap [1, N_k]| = \underline{d}(\mathcal{M}_\mathscr{B})$. We then have

$$\left| \int f \, d\nu_{\eta_K} - \int f \, d\nu_\eta \right| = \lim_{k \to \infty} \left| \int f \, d\frac{1}{N_k} \sum_{n \leqslant N_k} \delta_{S^n \eta_K} - \int f \, d\frac{1}{N_k} \sum_{n \leqslant N_k} \delta_{S^n \eta} \right|$$

$$\leqslant \lim_{k \to \infty} \frac{1}{N_k} \sum_{n \leqslant N_k} \left| f(S^n \eta_K) - f(S^n \eta) \right|$$

$$\leqslant 2\|f\| \cdot (2L - 1) \cdot \lim_{k \to \infty} \frac{1}{N_k} \sum_{n \leqslant N_k} |\{1 \leqslant n \leqslant N_k : \eta_K(n) \neq \eta(n)\}|$$

$$= 2\|f\| \cdot (2L - 1) \cdot \left| \underline{d}(\mathcal{M}_\mathscr{B}) - d\left(\mathcal{M}_{\mathscr{B}_K}\right) \right| \to 0 \text{ as } K \to \infty,$$

where the convergence follows from Theorem A.4. $\square$

Proof of Theorem 1.3 Take $\nu \in \mathcal{M}^e(\widetilde{X}_\eta, S)$. Since for $K \geqslant 1$, we have $\eta \leqslant \eta_K$ (coordinatewise), it follows that $\widetilde{X}_{\eta_K} \supset \widetilde{X}_\eta$, whence $\nu \in \mathcal{M}^e(\widetilde{X}_{\eta_K}, S)$. Let $u_K \in \widetilde{X}_{\eta_K}$ be a generic point for ν. Since $u_K \in \widetilde{X}_{\eta_K}, u_K \leqslant S^i \eta_K$ for some i (because we consider the hereditary closure of a periodic sequence). In other words, we have $u_K = S^i \eta_K \cdot y_K$ for some $y_K \in \{0, 1\}^\mathbb{Z}$. We may assume without loss of generality that $i = 0$ (since $S^{-i} u_K$ and u_K are generic for the same measure). Now, (η_K, y_K) is quasi-generic for a measure ρ_K defined on $X_{\eta_K} \times \{0, 1\}^\mathbb{Z}$. Note that its projection $\rho_K|_{X_{\eta_K}}$ onto the first coordinate satisfies $\rho_K|_{X_{\eta_K}} = \nu_{\eta_K}$. Moreover,

$$M_*(\rho_K) = \nu$$

---

[13] For subset $A \subset \mathbb{Z}$ symmetric with respect to 0, we have $d(A) = \lim_{N \to \infty} \frac{1}{N} |A \cap [1, N]| = \lim_{N \to \infty} \frac{1}{2N} |A \cap [-N, N]|$; an analogous relation holds for $\underline{d}$.

as $u_K = \eta_K \cdot y_K = M(\eta_K, y_K)$ is (quasi-)generic for $M_*(\rho_K)$ and generic for ν. Passing to a subsequence, if necessary, $\rho_K \to \rho$ (a measure on $X_\eta \times \{0, 1\}^{\mathbb{Z}}$). Therefore, we have

$$\nu = M_*(\rho_K) \to M_*(\rho),$$

so $\nu = M_*(\rho)$. Moreover,

$$\nu_{\eta_K} = \rho_K|_{X_\eta} \to \rho|_{X_\eta}$$

so $\rho|_{X_\eta} = \nu_\eta$, in view of Lemma A.5. $\qquad\square$

A.3 Example related to Question 5(B)

This section is related to Remark 1.12: it turns out that after relaxing the independence assumption (ii), there might be plenty of joint distributions of $(\mathbf{X}, \mathbf{Y})$ such that $\mathbf{H}(\mathbf{X}) > \mathbf{H}(\mathbf{M}) = 0$, with $\mathbf{X}$ and $\mathbf{Y}$, satisfying (i). More precisely, one can prove the following ergodic-theoretic result on $\mathscr{B}$-free systems:

Theorem A.6 *For any $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$, there exists $\rho \in \mathcal{M}(X_\eta \times \{0, 1\}^{\mathbb{Z}}, S \times S)$ with $\rho|_{X_\eta} = \nu_\eta$, such that $h(\rho, S \times S) > 0$ and $h(M_*(\rho), S) = 0$.*

Remark A.7 The proof of Theorem A.6 is quite technical and it is beyond the scope of this paper, as we put emphasis on the "independent case". It will be published elsewhere. We present it below in the simplest possible case, i.e. for $\mathscr{B} = \{2\}$. Then $X_\eta = \{a, b\}$, where $a := (\ldots 0, 1, 0, 1, \underset{\circ}{0}, 1, 0, 1 \ldots)$ and $b := Sa$, where S is the left shift on $\{0, 1\}^{\mathbb{Z}}$. Our approach is ergodic-theoretic and draws heavily on the description of invariant measures for $(\widetilde{X}_\eta, S)$ from [12, 28]. The notation is also related to the one in [12, 28].

Proposition A.8 *There exists $\rho \in \mathcal{M}(\{a, b\} \times \{0, 1\}^{\mathbb{Z}}, S \times S)$ such that $h(\rho, S \times S) > 0$, whereas $h(M_*(\rho), S) = 0$.*

Define $\overline{\Psi} \colon \{a, b\} \times \{0, 1\}^{\mathbb{Z}} \to \{a, b\} \times \{0, 1\}^{\mathbb{Z}} \times \{0, 1\}^{\mathbb{Z}}$ in the following way:

$\overline{\Psi}(a, x) = (a, \widehat{x}_a, \widetilde{x}_a)$, where $\widehat{x}_a = (\ldots, \underset{\circ}{x_1}, x_3, x_5, \ldots)$ and $\widetilde{x}_a = (\ldots, \underset{\circ}{x_0}, x_2, x_4, \ldots)$,

$\overline{\Psi}(b, x) = (b, \widehat{x}_b, \widetilde{x}_b)$, where $\widehat{x}_b = (\ldots, \underset{\circ}{x_0}, x_2, x_4, \ldots)$ and $\widetilde{x}_b = (\ldots, \underset{\circ}{x_1}, x_3, x_5, \ldots)$.

Clearly, $\widetilde{x}_c = \widehat{x}_{Sc}$ for $c \in \{a, b\}$. Notice that one can interpret $\widehat{x}_c$ as "survivors", i.e. these coordinates of x that "survive" after applying M to (c, x), $c \in \{a, b\}$. Similarly, $\widetilde{x}_c, c \in \{a, b\}$ can be seen as "victims", i.e. the coordinates of x that are "killed" after applying M to (c, x). Moreover, $\overline{\Psi}$ is a homeomorphism.

Lemma A.9 *We have $\overline{\Psi} \circ (S \times S) = \overline{S} \circ \overline{\Psi}$, where*

$$\overline{S}(a, y, z) = (Sa, y, Sz) \text{ and } \overline{S}(b, y, z) = (Sb, Sy, z).$$

Proof Direct calculation. $\qquad\square$
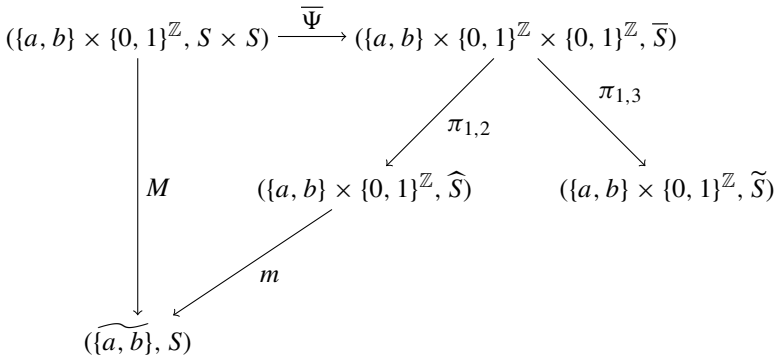
Denote the restriction of $\overline{S}$ to the first to coordinates by $\widehat{S}$ and to the first and third coordinate by $\widetilde{S}$. All these maps are homeomorphisms. Thus, $\overline{S}$ can be viewed as a (topological) joining of $\widehat{S}$ and $\widetilde{S}$. Both, $\widehat{S}$ and $\widetilde{S}$ act on $\{a, b\} \times \{0, 1\}^{\mathbb{Z}}$; in this joining, the first coodinates of $\widehat{S}$ and $\widetilde{S}$ are glued diagonally. Each choice of an invariant measure for $\overline{S}$ yields a joining (in ergodic theoretic sense) of the corresponding projections for $\widehat{S}$ and $\widetilde{S}$.

Lemma A.10 *We have* $M = m \circ \pi_{1,2} \circ \overline{\Psi}$, *where* $\pi_{1,2} \colon \{a, b\} \times \{0, 1\}^{\mathbb{Z}} \times \{0, 1\}^{\mathbb{Z}} \to \{a, b\} \times \{0, 1\}^{\mathbb{Z}}$ *stands for the projection onto the first two coordinates and* $m \colon \{a, b\} \times \{0, 1\}^{\mathbb{Z}} \to \widetilde{\{a, b\}} = M(\{a, b\} \times \{0, 1\})$ *is given by* $m(a, y) = (\ldots, \underset{\circ}{0}, y_0, 0, y_1, 0, y_2, \ldots), m(b, y) = (\ldots, y_0, 0, y_1, 0, y_2, 0, \ldots)$.

Proof Direct calculation. □

We can summarize the above in the following commuting diagram ($\pi_{1,3}$ stands for the projection onto the first and third coordinate):

$$
\begin{array}{ccc}
(\{a, b\} \times \{0, 1\}^{\mathbb{Z}}, S \times S) & \xrightarrow{\;\overline{\Psi}\;} & (\{a, b\} \times \{0, 1\}^{\mathbb{Z}} \times \{0, 1\}^{\mathbb{Z}}, \overline{S})
\end{array}
$$

with maps M, $\pi_{1,2}$, $\pi_{1,3}$, m to $(\{a, b\} \times \{0, 1\}^{\mathbb{Z}}, \widehat{S})$, $(\{a, b\} \times \{0, 1\}^{\mathbb{Z}}, \widetilde{S})$ and $(\widetilde{\{a, b\}}, S)$.

Proof of Proposition A.8 We start with any $\widehat{\kappa} \in \mathcal{M}^e(\{a, b\} \times \{0, 1\}^{\mathbb{Z}}, \widehat{S})$ and $\widetilde{\kappa} \in \mathcal{M}^e(\{a, b\} \times \{0, 1\}^{\mathbb{Z}}, \widetilde{S})$. The projection of both $\widehat{\kappa}$ and $\widetilde{\kappa}$ onto the first coordinate is the unique S-invariant measure on $\{a, b\}$, i.e. equals $\frac{1}{2}(\delta_a + \delta_b)$. Note that this is nothing but ν_η corresponding to $\mathscr{B} = \{2\}$. Therefore, we can "glue" these coordinates together diagonally and find $\kappa \in \mathcal{M}(\{a, b\} \times \{0, 1\}^{\mathbb{Z}} \times \{0, 1\}^{\mathbb{Z}}, \overline{S})$ such that

$$(\pi_{1,2})_*(\kappa) = \widehat{\kappa} \text{ and } (\pi_{1,3})_*(\kappa) = \widetilde{\kappa}$$

(for instance, one can take so-called relatively independent extension of the diagonal joining of the first coordinates).

Now, suppose additionally that $0 = h(\widehat{S}, \widehat{\kappa}) < h(\widetilde{S}, \widetilde{\kappa})$ (e.g., one can take $\widehat{\kappa} = \nu_\eta \otimes \delta_{(\ldots, 0, 0, 0, \ldots)}$ and $\widetilde{\kappa} = \nu_\eta \otimes B_{1/2, 1/2}$). Then

$$h(\widetilde{S}, \widetilde{\kappa}) \leqslant h(\overline{S}, \kappa) \leqslant h(\widetilde{S}, \widetilde{\kappa}) + h(\widehat{S}, \widehat{\kappa}) = h(\widetilde{S}, \widetilde{\kappa}),$$

where the first inequality follows from the fact that $(S, \widetilde{\kappa})$ is a factor of $(\overline{S}, \kappa)$, the second one is a direct consequence of $(\overline{S}, \kappa)$ being a joining of $(S, \widetilde{\kappa})$ and $(S, \widehat{\kappa})$, and

we use $h(\widehat{\kappa}) = 0$. Thus,

$$h(\overline{S}, \kappa) = h(\widetilde{S}, \widetilde{\kappa}) > 0.$$

Moreover,

$$h(S, m_*(\widehat{\kappa})) \leqslant h(\widehat{S}, \widehat{\kappa}) = 0, \text{ whence } h(S, m_*(\widehat{\kappa})) = 0.$$

Let $\rho := (\overline{\Psi}^{-1})_*(\kappa)$. We obtain

$$h(\rho, S \times S) = h(\overline{S}, \kappa) > 0.$$

Moreover, $M_*(\rho) = (m \circ \pi_{1,2} \circ \overline{\Psi})_*(\rho) = (m \circ \pi_{1,2})_*(\kappa) = m_*(\widehat{\kappa})$ and it follows immediately that

$$h(M_*(\rho), S) = 0.$$

$\square$

Remark A.11 Notice that in the above proof $\widehat{\kappa}$ and $\widetilde{\kappa}$ is arbitrary, the only additional assumption was concerned with their entropy. This (together with the fact that $\overline{\Psi}$ is a homeomorphism) indicates that the set of measures ρ satisfying the assertion of Proposition A.8 is very rich.

References

1. Berbee, H.: Periodicity and absolute regularity. Israel J. Math. **55**, 289–304 (1986)
2. Besicovitch, A.S.: On the density of certain sequences of integers. Math. Ann. **110**, 336–341 (1935)
3. Blackwell, D., Freedman, D.: The tail σ-field of a Markov chain and a theorem of Orey. The Annals of Mathematical Statistics **35**, 1291–1295 (1964)
4. Bradley, R.C.: Introduction to strong mixing conditions, vol. 1. Kendrick Press, Heber City, UT (2007)
5. CHOWLA, S.: *On abundant numbers*, J. Indian Math. Soc., New Ser., 1 (1934), pp. 41–44
6. CHUNG, K. L.: *Markov Chains with Stationary Transition Probabilities: 2d Ed*, Springer, 1967
7. DAVENPORT, H.: *Über numeri abundantes*, Sitzungsber.Preuss.Akad.Wiss., (1933), pp. 830–837
8. Davenport, H., Erdös, P.: On sequences of positive integers. Acta Arithmetica **2**, 147–151 (1936)
9. Davenport, H., Erdös, P.: *On sequences of positive integers*, J. Indian Math. Soc. (N.S.), 15: 19–24 (1951)
10. DE FINETTI, B.: *Funzione caratteristica di un fenomeno aleatorio*, Atti della R. Accademia Nazionale dei Lincei, Ser. 6. Memorie, Classe di Scienze Fisiche, Matematiche e Naturali 4, (1931), pp. 251–299
11. Downarowicz, T.: Entropy in Dynamical Systems. New Mathematical Monographs, vol. 18. Cambridge University Press, Cambridge (2011)
12. Dymek, A., Kasjan, S., Kułaga-Przymus, J., Lemańczyk, M.: $\mathscr{B}$-free sets and dynamics. Trans. Amer. Math. Soc. **370**, 5425–5489 (2018)
13. Einsiedler, M., Ward, T.: Ergodic theory with a view towards number theory. Graduate Texts in Mathematics, vol. 259. Springer-Verlag, London Ltd, London (2011)
14. Erdös, P.: On the Density of the Abundant Numbers. J. London Math. Soc. **9**, 278–282 (1934)
15. FELLER, W.: *An introduction to probability theory and its applications*, vol. 1, John Wiley & Sons, 2008
16. Friedman, N.A., Ornstein, D.S.: On isomorphism of weak Bernoulli transformations. Advances in Math. **5**(1970), 365–394 (1970)

17. Furstenberg, H.: Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. Math. Systems Theory **1**, 1–49 (1967)
18. FURSTENBERG, H., PERES, Y., WEISS, B.: *Perfect filtering and double disjointness*, Annales de l'I.H.P. Probabilités et statistiques, 31 (1995), pp. 453–465
19. Garbit, R.: A note on Furstenberg's filtering problem. Israel J. Math. **182**, 333–336 (2011)
20. Glasner, E.: Ergodic theory via joinings. Mathematical Surveys and Monographs, vol. 101. American Mathematical Society, Providence, RI (2003)
21. Glasner, E., Thouvenot, J.-P., Weiss, B.: Entropy theory without a past. Ergodic Theory and Dynamical Systems **20**, 1355–1370 (2000)
22. GRAY, R. M.: *Entropy and information theory*, Springer, New York, second ed., 2011
23. Hall, R.R.: Sets of multiples. Cambridge Tracts in Mathematics, vol. 118. Cambridge University Press, Cambridge (1996)
24. Hewitt, E., Savage, L.J.: Symmetric measures on Cartesian products. Trans. Amer. Math. Soc. **80**, 470–501 (1955)
25. HOCHMAN, M.: *Lectures on dynamical systems and entropy*. http://math.huji.ac.il/~mhochman/courses/dynamics2014/notes.5.pdf
26. Keller, G.: Tautness of sets of multiples and applications to $\mathcal{B}$-free systems. Studia Math. **247**, 205–216 (2019)
27. Krieger, W.: On entropy and generators of measure-preserving transformations. Trans. Amer. Math. Soc. **149**, 453–464 (1970)
28. KUAGA-PRZYMUS, J., LEMAŃCZYK, M., WEISS, B.: On invariant measures for $\mathcal{B}$-free systems, Proc. Lond. Math. Soc. (3), 110 (2015), pp. 1435–1474
29. Maker, P.T.: The ergodic theorem for a sequence of functions. Duke Mathematical Journal **6**, 27–30 (1940)
30. Olshen, R.A.: The coincidence of measure algebras under an exchangeable probability. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **18**, 153–158 (1971)
31. Ornstein, D.S., Weiss, B.: Finitely determined implies very weak Bernoulli. Israel J. Math. **17**, 94–104 (1974)
32. Ornstein, D.S., Weiss, B.: Every transformation is bilaterally deterministic. Israel Journal of Mathematics **21**, 154–158 (1975)
33. Peckner, R.: Uniqueness of the measure of maximal entropy for the squarefree flow. Israel J. Math. **210**, 335–357 (2015)
34. Pinsker, M.S.: Dynamical systems with completely positive or zero entropy. Soviet Math. Dokl. **1**, 937–938 (1960)
35. SARNAK, P.: *Three lectures on the Möbius function, randomness and dynamics*. http://publications.ias.edu/sarnak/
36. SHIELDS, P.: *The theory of Bernoulli shifts*, The University of Chicago Press, Chicago, Ill.-London, 1973. Chicago Lectures in Mathematics
37. VOLKONSKI, V. A., ROZANOV, Y. A.: *Some limit theorems for random functions. I*, Theor. Probability Appl., 4 (1959), pp. 178–197
38. VOLKONSKI, V. A., ROZANOV, Y. A.: *Some limit theorems for random functions. II*, Teor. Verojatnost. i Primenen., 6 (1961), pp. 202–215