ISRAEL JOURNAL OF MATHEMATICS **164** (2008), 265–284
DOI: 10.1007/s11856-008-0029-y

ON PROCESSES WHICH CANNOT BE DISTINGUISHED BY FINITE OBSERVATION*

BY

Yonatan Gutman

Departmemt of Mathematics, Edmond J. Safra Campus
Hebrew University of Jerusalem, Jerusalem 91904, Israel
e-mail: gyonatan@math.huji.ac.il

AND

Michael Hochman

Department of Mathematics, Fine Hall,
Washington Rd. Princeton, NJ 08544, USA
e-mail: hochman@math.princeton.edu

ABSTRACT

A function J defined on a family $\mathcal{C}$ of stationary processes is finitely observable if there is a sequence of functions s_n such that $s_n(x_1, \ldots, x_n) \to J(\mathcal{X})$ in probability for every process $\mathcal{X} =(x_n) \in \mathcal{C}$. Recently, Ornstein and Weiss proved the striking result that if $\mathcal{C}$ is the class of aperiodic ergodic finite valued processes, then the only finitely observable isomorphism invariant defined on $\mathcal{C}$ is entropy [8]. We sharpen this in several ways. Our main result is that if $\mathcal{X} \to \mathcal{Y}$ is a zero-entropy extension of finite entropy ergodic systems and $\mathcal{C}$ is the family of processes arising from generating partitions of $\mathcal{X}$ and $\mathcal{Y}$, then every finitely observable function on $\mathcal{C}$ is constant. This implies Ornstein and Weiss' result, and extends it to many other families of processes, e.g., it follows that there are no nontrivial finitely observable isomorphism invariants for processes arising from the class of Kronecker systems, the class of mild mixing zero entropy systems, or the class of strong mixing zero entropy systems. It also follows that for the class of processes arising from irrational rotations, every finitely observable isomorphism invariant must be constant for rotations belonging to a set of full Lebesgue measure.

* This research was supported by the Israel Science Foundation (grant No. 1333/04)
Received August 13, 2006 and in revised form November 5, 2006.

1. Introduction

Let $(x_n)_{n=-\infty}^{\infty}$ be an aperiodic ergodic process taking on finitely many values (without loss of generality the values are in $\mathbb{N}$). Up to isomorphism there is a unique measure preserving system, $\mathcal{X} = (X, \mathcal{B}, \mu, T)$, such that (x_n) arises from a generating partition $\mathcal{P} = (P_i)$ of an $\mathcal{X}$. The question we are interested in is, what can one learn about the underlying system $\mathcal{X}$ by observing a sample path (x_n)?

In principle, the answer is “everything”, since by the ergodic theorem a typical sample path of $(x_n)_{n=1}^{\infty}$ determines all finite distributions of the process and this determines $\mathcal{X}$. However, a more realistic scenario is one in which at each time step another output of the process is revealed, i.e. at time n we have observed the finite sequence $x_1, \ldots, x_n$, and we are asked to make a guess about the nature of $\mathcal{X}$ based on this data. These guesses should converge as $n \to \infty$. We call a scheme for producing such a sequence of guesses an observation scheme. To be precise

Definition 1.1: An **observation scheme** (or scheme for short) is a metric space Δ and a sequence of functions $s_n : \mathbb{N}^n \to \Delta$. An observation scheme is said to converge for a family of processes $\mathcal{C}$ if $\lim_{n\to\infty} s_n(x_1, \ldots, x_n)$ exists in probability for every process $(x_n) \in \mathcal{C}$. A function $J : \mathcal{C} \to \Delta$ is **finitely observable** if there is an observation scheme (s_n) which converges to $J((x_n))$ for every $(x_n) \in \mathcal{C}$.

Note that the larger a family of processes is, the harder it is for a scheme to converge for every member of the family, hence large families have fewer finitely observable functions.

Nonetheless, there are many observation schemes (s_n) for which the sequence $s_1(x_1),\ s_2(x_1, x_2),\ s_3(x_1, x_2, x_3), \ldots$ converges in probability, or even almost surely, for *every* ergodic process (x_n). For example, if $s_n(x_1, \ldots, x_n)$ counts the frequencies of 1's appearing in $x_1, \ldots, x_n$, then by the ergodic theorem $\lim_{n\to\infty} s_n(x_1 \ldots x_n)$ exists almost surely and equals the probability of the symbol 1 in the process (x_n). This example and others like it show that some things about a process can be calculated from finite observations; but these are generally not isomorphism invariants, and so tell us nothing about the underlying dynamical system.

We will say that a process (x_n) **arises** from a measure preserving system $\mathcal{X}$ if it is defined by a generating partition of $\mathcal{X}$. For processes (x_n) and (y_n) etc. we denote by $\mathcal{X}$ and $\mathcal{Y}$ respectively the measure preserving system determined by them (and from which they arise). Write $(x_n) \cong (y_n)$ or $\mathcal{X} \cong \mathcal{Y}$ to indicate that $\mathcal{X}$ and $\mathcal{Y}$ are isomorphic as dynamical systems. We are interested in families of processes $\mathcal{C}$ which are closed under isomorphism, that is, they will have the property that if $(x_n) \in \mathcal{C}$ and $(y_n) \cong (x_n)$ then $(y_n) \in \mathcal{C}$. Such a family is called **saturated**. Usually we will specify $\mathcal{C}$ by some property of the underlying systems, e.g., $\mathcal{C}$ might be the family of all processes arising from an irrational rotation. In this case we would say for brevity that $\mathcal{C}$ is the class of irrational rotations.

Definition 1.2: Let $\mathcal{C}$ be a saturated family of processes, Δ a metric space and $J : \mathcal{C} \to \Delta$. Then J is an **isomorphism invariant** for $\mathcal{C}$ (or invariant for short) if for every $(x_n), (y_n) \in \mathcal{C}$,

$$(x_n) \cong (y_n) \;\Rightarrow\; J((x_n)) = J((y_n))$$

and J is a **complete** invariant for $\mathcal{C}$ if the reverse implication also holds. When J is an invariant we write $J(\mathcal{X})$ instead of $J((x_n))$.

For quite some time it has been known that the entropy, $h((x_n)) = h(\mathcal{X})$, of a process is finitely observable in the class of all ergodic processes. The earliest observation scheme for entropy is due to D. Bailey [1]. A number of simpler schemes have been developed, such as the Lempel–Ziv compression algorithm [12] and the Ornstein–Weiss estimators [6, 7].

D. Ornstein and B. Weiss recently proved a striking converse to this: Every finitely observable invariant for the class of all ergodic processes is a continuous function of entropy [8]. They also showed that there are no finitely observable invariants except entropy for any class which contains the Bernoulli processes, for the class of zero entropy processes or for the class of zero entropy weak mixing processes.

However, their techniques do not settle what is finitely observable in several other interesting classes of systems. Ornstein and Weiss have asked if there exists a complete finitely observable invariant for the class of irrational rotations (translations by an irrational on the group $\mathbb{R}/\mathbb{Z}$); this is not implausible, since for this class there is a complete invariant for isomorphism, namely the spectrum, or equivalently the modulus of rotation (up to sign and mod1). We remark that in the classes which Ornstein and Weiss studied there are no known complete invariants, finitely observable or not, with the exception of the class of Bernoulli systems, in which entropy is itself a complete invariant.

In an attempt to get a handle on this problem, we came up with the following, which is interesting in its own right:

THEOREM: *Suppose $\mathcal{X} \to \mathcal{Y}$ is a zero entropy extension of finite entropy dynamical systems, that is $h(\mathcal{X}) = h(\mathcal{Y})$. Let $\mathcal{C}$ be the class of processes arising from $\mathcal{X}, \mathcal{Y}$. Then every finitely observable invariant for $\mathcal{C}$ is constant.*

This allows us reclaim the results of Ornstein and Weiss, and to settle the following problems.

THEOREM: *If J is a finitely observable invariant on one of the following classes:*

(1) *the Kronecker systems (the class of systems with pure point spectrum),*
(2) *the zero entropy mild mixing processes,*
(3) *the zero entropy strong mixing processes,*

Then J is constant.

For the class of irrational rotations we obtain a slightly weaker result.

THEOREM: *For every finitely observable invariant J on the class of irrational rotations, there is a Borel set $\Theta \subseteq [0,1)$ of full Lebesgue measure such that J assigns the same value to processes arising from rotations by angles in Θ. In particular, there is no complete finitely observable invariant for irrational rotations.*

The rest of the paper is organized as follows. Section 2 presents some definitions and background. In section 3 we prove the theorem about zero-entropy extensions. Section 4 contains proofs of the other results, and in section 5 we mention some open problems.

ACKNOWLEDGEMENT. This paper was written during the authors' graduate studies. We would like to thank our adviser Professor Benjamin Weiss for his encouragement and good advice.

2. Preliminaries

For general background on ergodic theory we refer to [3, 9, 11].

2.1. Dynamical systems, partitions and processes. By an aperiodic ergodic system $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ we mean that $(X, \mathcal{B}, \mu)$ is a standard probability space, $T : X \to X$ is an invertible measure preserving map which acts ergodically, and the set of periodic points is of measure zero. A measure preserving systems $\mathcal{Y} = (Y, \mathcal{C}, \nu, S)$ is a **factor** of the system $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ if there is a measure-preserving map $f : X \to Y$ defined almost everywhere satisfying $Sf = fT$. If there is such a map which is also invertible and bi-measurable then $\mathcal{X}$,$\mathcal{Y}$ are **isomorphic**.

A partition $\mathcal{P}$ of X is a finite ordered collection of pairwise disjoint measurable sets $(P_i)_{i=1}^{|\mathcal{P}|}$ whose union is X (up to measure zero). If $\mathcal{P}, \mathcal{Q}$ are partitions of X then the partition $\mathcal{P} \vee \mathcal{Q} = (P_i \cap Q_j)_{(i,j)}$ is the **join** of P, Q (order the pairs (i, j) lexicographically); the join of finitely many partitions is defined similarly. Write $T^n\mathcal{P} = (T^n P_i)$.

A partition $\mathcal{P}$ of X **generates** $\mathcal{X}$ if $\bigvee_{n=-\infty}^{\infty} T^n\mathcal{P} = \mathcal{B}$ up to measure zero, where $\bigvee_{n=-\infty}^{\infty} T^n\mathcal{P}$ is the σ-algebra generated by the collection

$$\bigcup_N \bigvee_{n=-N}^{N} T^n\mathcal{P}.$$

For a partition $\mathcal{P} = (P_i)_{i\in\mathbb{N}}$ and $\omega \in X$ we write $\mathcal{P}(\omega)$ for the index of the set in $\mathcal{P}$ that contains ω. A partition $\mathcal{P}$ determines a stationary ergodic process (x_n) with values in $\mathbb{N}$ by

$$x_n(\omega) = \mathcal{P}(T^n\omega).$$

We say that $x_i(\omega), x_{i+1}(\omega), \ldots, x_j(\omega)$ is the **itinerary** of ω (with respect to $\mathcal{P}$) from time i to time j. The itinerary of ω from time 0 to time $N-1$ is called the $(\mathcal{P}, N)$-**name** of ω. If $\mathcal{P}$ is a generating partition for $\mathcal{X}$, then the system $\mathcal{X}$ and the partition $\mathcal{P}$ are determined, up to isomorphism, by the process (x_n). We say that this process **arises** from $\mathcal{P}$ if $\mathcal{P}$ generates $\mathcal{X}$.

The space of ordered partitions of X into n sets comes with a metric $\rho = \rho_n$ defined by

$$\rho(\mathcal{P}, \mathcal{Q}) = \sum_{i=1}^{n} \mu(P_i \triangle Q_i),$$

for $\mathcal{P} = (P_1, \ldots, P_n)$ and $\mathcal{Q} = (Q_1, \ldots, Q_n)$ (here Δ denotes symmetric difference). The metric ρ_n is complete; note, however, that if $\mathcal{P}_i \to \mathcal{P}$ in ρ_n it may happen that some of the members of $\mathcal{P}$ are empty.

It is easy to check that if $\rho(\mathcal{P},\mathcal{Q})<\varepsilon$, then $\rho(\bigvee_{n=1}^N T^n\mathcal{P},\bigvee_{n=1}^N T^n\mathcal{Q})<N\varepsilon$. It follows that if $\mathcal{P}_k\to\mathcal{P}$ in ρ and $(x_n^{(k)}),(x_n)$ denote the processes arising from $\mathcal{P}^k,\mathcal{P}$ respectively, then the sequence of processes $(x_n^{(k)})_{n=-\infty}^{\infty}$ converges to $(x_n)_{n=-\infty}^{\infty}$ in probability.

Given a partition $\mathcal{P}$ of X into r sets and an integer N we may consider the distribution that μ induces on $\{1,\ldots,r\}^N$, where the measure of a word $w\in\{1,\ldots,r\}^N$ is the measure of the set of points whose $(\mathcal{P},N)$-name is w, or in other words $\mu(\bigcap_{n=1}^N T^{-n}P_{w(n)})$. We refer to this as the distribution of N-names determined by $\mathcal{P}$.

Since a distribution on N-names is just a r^N-dimensional probability vector, we can compare these distributions using e.g. the ℓ^1 metric. When we talk of closeness of N-name distributions, we will mean it in this sense. Note that if $\mathcal{P},\mathcal{Q}$ are partitions and $\rho(\mathcal{P},\mathcal{Q})<\varepsilon$ then the distance between the N-name distributions associated with $\mathcal{P}$ and $\mathcal{Q}$ is at most $N\varepsilon$.

2.2. Entropy. Let $\mathcal{X}=(X,\mathcal{B},\mu,T)$ be an invertible ergodic measure preserving system and $\mathcal{P}=(P_i)$ a partition. The **entropy** of a partition $\mathcal{P}$ is

$$H(\mathcal{P})=-\sum_i \mu(P_i)\log\mu(P_i).$$

(all logarithms are to base 2 unless specified otherwise). $H(\mathcal{P})$ is non-negative and finite (define $0\log 0=0$). The entropy of the system $\mathcal{X}$ with respect to $\mathcal{P}$ (equivalently, the entropy of the process arising from $\mathcal{P}$) is

$$h(\mathcal{X},\mathcal{P})=\lim_{n\to\infty}\frac{1}{n}H(\mathcal{P}\vee T\mathcal{P}\vee\cdots\vee T^{n-1}\mathcal{P}).$$

The limit above can be shown to exist. The entropy of $\mathcal{X}$ is

$$h(\mathcal{X})=\sup\{h(\mathcal{X},\mathcal{P})\,:\,\mathcal{P}\text{ a finite partition of }X\}.$$

If $\mathcal{P}$ is a finite generating partition then $h(\mathcal{X})=h(\mathcal{X},\mathcal{P})$, but the relation $h(\mathcal{X})=h(\mathcal{X},\mathcal{P})$ is not in itself enough to guarantee that $\mathcal{P}$ generates. However, the Krieger generator theorem [5] guarantees that if $h(\mathcal{X})<\log k$ for an integer k, then there exists a generating partition $\mathcal{P}=(P_1,\ldots,P_k)$ of $\mathcal{X}$ into k sets.

In the space of partitions of X into n sets, the entropy is continuous in the metric ρ_n: that is, for a partition $\mathcal{P}$, for every $\delta>0$ there is an $\varepsilon>0$ such that if $\rho(\mathcal{P},\mathcal{Q})<\delta$ then $|h(\mathcal{X},\mathcal{P})-h(\mathcal{X},\mathcal{Q})|<\varepsilon$.

The main fact about entropy we use is the following classical theorem:

THEOREM 2.1 (Shannon–McMillan–Breiman theorem): *For any finite partition $\mathcal{P}$ of $\mathcal{X}$ and almost every $x \in X$,*

$$\frac{1}{n} \log \mu\Big(\bigcap_{i=0}^{n-1} \mathcal{P}(T^{-i}x) \Big) \to h(\mathcal{X}, \mathcal{P}).$$

A proof can be found in [10, p. 55].

Denote

$$\mu(u) = \mu(\{x \in X \,:\, \text{the } (\mathcal{P}, n)\text{-name of } x \text{ is } u\}).$$

With this notation the Shannon–McMillan–Breiman theorem states that

$$\frac{1}{n} \log \mu(x_1, \ldots, x_n) \to h(\mathcal{X}, \mathcal{P})$$

almost surely, where (x_n) is the process arising from $\mathcal{P}$.

Also, for partitions $\mathcal{P}, \mathcal{Q}$ and $(u, v) \in \mathbb{N}^n \times \mathbb{N}^n$, we say that (u, v) is the $(\mathcal{P} \times \mathcal{Q}, n)$ name of a point $\omega \in X$ if u is the $(\mathcal{P}, n)$-name of ω and v is the $(\mathcal{Q}, n)$-name of ω. This is just another way of talking about the partition $\mathcal{P} \vee \mathcal{Q}$. Denote

$$\mu(v|u) = \frac{\mu(\{x \in X \,:\, \text{the } (\mathcal{P} \times \mathcal{Q}, n)\text{-name of } x \text{ is } (u, v)\})}{\mu(\{x \in X \,:\, \text{the } (\mathcal{P}, n)\text{-name of } x \text{ is } u\})} \,.$$

We will actually use the following "relative" version of the Shannon–McMillan–Breiman theorem.

THEOREM 2.2 (Relative Shannon–McMillan–Breiman): *Let $\mathcal{P}, \mathcal{Q}$ be partitions of $\mathcal{X}$ with entropies $h(\mathcal{X}, \mathcal{P}) = s \leq t = h(\mathcal{X}, \mathcal{Q})$. For every $\varepsilon > 0$ there are collections of words $A_n \subseteq \mathbb{N}^n \times \mathbb{N}^n$ for $n = 1, 2, 3, \ldots$ such that*

(1) $\#\{u \in \mathbb{N}^n \,:\, (u, v) \in A_n$ *for some* $v\} < 2^{(s+\varepsilon)n}$ *for every* n.
(2) $\#\{v \in \mathbb{N}^n \,:\, (u, v) \in A_n\} < 2^{(t-s+\varepsilon)n}$ *for every* n.
(3) *For almost every point $x \in X$ the $(\mathcal{P} \times \mathcal{Q}, n)$-name of x is in A_n for all sufficiently large n.*

Proof. Define

$$A_n = \{(u, v) \in \mathbb{N}^n \times \mathbb{N}^n \,:\, \mu(u) > 2^{-(s+\varepsilon)n} \text{ and } \mu(v|u) > 2^{-(t-s+\varepsilon)n}\}.$$

The fact that for almost every $x \in X$ the $(\mathcal{P} \times \mathcal{Q}, n)$-name of x is eventually in A_n follows from the Shannon–McMillan–Breiman theorem, once applied to the partition $\mathcal{P}$ and once to the partition $\mathcal{P} \times \mathcal{Q}$. The estimates on the size of the u's represented in A_n and the v's associated to a given u in A_n follow easily from the definition since the mass of the u's and the mass of the v's relative to a given u must add to at most 1. ∎

2.3. Towers. A **tower of height** n in $\mathcal{X}$ is a set of the form

$$B \cup TB \cup T^2B \cup \cdots \cup T^{n-1}B \subseteq X$$

such that the sets T^iB are measurable and pairwise disjoint for $i = 0, \dots, n-1$. The set B is called the **base** of the tower, and the set T^iB is called the i-th **level** of the tower.

Given a partition $\mathcal{P} = (P_i)$ and a tower $\bigcup_{i=0}^{n-1} T^iB$, we can partition the base B into disjoint (possibly empty) sets B_w indexed by words $w \in \mathbb{N}^n$, such that

$$B_u = \{\omega \in B : u \text{ is the } (\mathcal{P}, n) - \text{name of } \omega\}.$$

This partitions the tower into disjoint subtowers $\bigcup_{i=0}^{n-1} T^iB_u$ whose base is B_u; these subtowers are called **columns**. Each level T^iB_u is contained entirely in the element $P_{u(i)}$ of $\mathcal{P}$. Rephrasing, if (x_n) is the process associated with $\mathcal{P}$ then for $\omega \in B_u$ the first n outputs $(x_1(\omega), \dots, x_{n-1}(\omega))$ of the process are equal to $u = (u_1, \dots, u_{n-1})$.

We will need two tower lemmas.

Lemma 2.3 (Kakutani towers lemma): *Let B be a set of positive measure and N an integer. Then the space X can be partitioned into countably many pairwise disjoint towers all of height no less than N, all of whose bases are subsets of B.*

Proof. Since $\mathcal{X}$ is aperiodic we can choose a set $B' \subseteq B$ of positive measure such that if $x \in B'$ then $T^ix \notin B'$ for $1 \le i < N$. Partition the base B according to the first return time to B', i.e., let

$$B^{(n)} = \{x \in B' : n \text{ is the first positive integer such that } T^nx \in B'\}.$$

Then for each $n \ge N$ we have a tower $B^{(n)} \cup TB^{(n)} \cup \cdots \cup T^{(n-1)}B^{(n)}$, these towers are pairwise disjoint, and their union fills X because their union is invariant. ∎

A stronger result is a version of the Rohlin lemma whose proof can be found in [9]

Lemma 2.4 (Strong Rohlin lemma): *Let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a partition of X and $\varepsilon > 0$. Then for every N there is a tower $B \cup TB \cup \cdots \cup T^{N-1}B$ of height N whose complement is of measure at most ε and such that the partition*

$\mathcal{Q} = \{B \cap P_1, \ldots, B \cap P_k\}$ *induced on* B *by* $\mathcal{P}$ *has the same distribution relative to* B *as* $\mathcal{P}$ *has relative to* X.

COROLLARY 2.5: *Given* $A \subseteq X$ *with* $\mu(A) > 1 - \varepsilon$ *and any* N, *there is a tower* $B \cup TB \cup \cdots \cup T^{N-1}B$ *in* X *filling all but* 2ε *of the space and with* $B \subseteq A$.

Proof. Let $C \cup TC \cup \cdots \cup T^{N-1}C$ be the tower provided by the strong Rohlin lemma with respect to the partition $\{A, X \setminus A\}$ and set $B = C \cap A$. ∎

2.4. APPROXIMATION METHODS FOR PARTITIONS. Often, a generating partition with some property is constructed by approximation, that is, a sequence of partitions is defined satisfying more and more of our requirements and which converge in ρ to a partition with the properties we want. Below we outline some of the tools we use for such constructions.

If $\mathcal{A}$ is a partition or an algebra of measurable sets and B is a measurable set then we write $B \subseteq_\varepsilon \mathcal{A}$ to indicate that there is a set $A \in \mathcal{A}$ such that $\mu(A \triangle B) < \varepsilon$. Clearly $B \in \mathcal{A}$ (up to measure zero) if and only if $B \subseteq_\varepsilon \mathcal{A}$ for every $\varepsilon > 0$. For a partition $\mathcal{P}$ we write $\mathcal{P} \subseteq_\varepsilon \mathcal{A}$ if $P_i \subseteq_\varepsilon \mathcal{A}$ for every $P_i \in \mathcal{P}$.

Let $\mathcal{P}$ be a generating partition for $\mathcal{X}$ and suppose that $\mathcal{Q}$ is a partition such that, for every $\varepsilon > 0$, there is an N such that $\mathcal{P} \subseteq_\varepsilon \bigvee_{n=-N}^{N} T^n Q$. It follows that $P \subseteq \bigvee_{n=-\infty}^{\infty} T^n \mathcal{Q}$, and since $\bigvee_{n=-\infty}^{\infty} T^n \mathcal{Q}$ is T-invariant, $\mathcal{B} = \bigvee_{n=-\infty}^{\infty} T^n \mathcal{P} \subseteq \bigvee_{n=-\infty}^{\infty} T^n \mathcal{Q}$. Thus $\mathcal{Q}$ generates.

Suppose $\mathcal{P}, \mathcal{Q}$ are partitions of X into n elements and $A \subseteq_\varepsilon \mathcal{P}$. Then if $\rho(\mathcal{P}, \mathcal{Q}) < \delta$ we have $A \subseteq_{\varepsilon+\delta} \mathcal{Q}$. Thus if $A \subseteq_\varepsilon \bigvee_{n=1}^{N} T^n \mathcal{P}$ and $\rho(\mathcal{P}, \mathcal{Q}) < \delta$ then $A \subseteq_{\varepsilon+N\delta} \bigvee_{n=1}^{N} T^n \mathcal{Q}$.

These observations are essentially the proof of the following lemma, see also [9, p. 79].

LEMMA 2.6: *Let* $(\mathcal{P}_k)_{k=1}^{\infty}$ *be a sequence of partitions of* X *and* $\mathcal{Q}$ *a partition of* X. *Suppose that* $\rho(\mathcal{P}_{k-1}, \mathcal{P}_k) < \varepsilon(k)$ *and* $\mathcal{Q} \subset_{\varepsilon(k)} \bigvee_{j=-N(k)}^{N(k)} T^{-j} \mathcal{P}_k$ *for some sequences* $\varepsilon(k) > 0$ *and* $N(k) \in \mathbb{N}$ *which satisfy* $\sum_{k=1}^{\infty} \varepsilon(k) < \infty$ *and* $N(k) \cdot \sum_{j=k+1}^{\infty} \varepsilon(j) \to 0$ *as* $k \to \infty$. *Then* $(\mathcal{P}_k)$ *converges to a partition* $\mathcal{P}$ *and* $Q \subseteq \bigvee_{j=-\infty}^{\infty} T^{-j} P$.

The following theorem shows that in order to change a partition $\mathcal{P}$ into a generating partition, you need to perturb $\mathcal{P}$ by an amount of the same order as the difference $h(\mathcal{X}) - h(\mathcal{P})$. This result is not new but, lacking a reference, we include a proof for completeness.

THEOREM 2.7 (Entropy and generating partitions): *Let $h \geq 0$ and k be an integer with $\log k > h$. Let $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ be an aperiodic ergodic system with entropy h and let $\mathcal{P} = (P_1, \ldots, P_k)$ be a partition of $\mathcal{X}$ with $h(\mathcal{X}, \mathcal{P}) = h'$ (so $h' \leq h$). Then for every $\delta > 0$ there is a generating partition $\mathcal{P}' = (P'_1, \ldots, P'_k)$ of $\mathcal{X}$ such that $\rho(\mathcal{P}, \mathcal{P}') < \delta + \frac{h-h'}{\log k - h}$. In particular, the generating partitions are dense in the ρ-metric among the partitions of maximal entropy.*

Remark: The parameter δ was introduced only in order to deal with the case that $h = h'$.

Proof. Let $\delta > 0$ be given. Fix a very small $\varepsilon > 0$ which will determined later. Fix a generating partition $\mathcal{Q}$ of size k, and for $n = 1, 2, 3, \ldots$ let $A_n \subseteq \mathbb{N}^n \times \mathbb{N}^n$ be as in Theorem 2.2 for the partitions $\mathcal{P}, \mathcal{Q}$ and parameter ε. Let $N \geq 1/\varepsilon$ be large enough so that the the set X_0 of ω's whose $(\mathcal{P} \times \mathcal{Q}, n)$-name in A_n for all $n \geq N$ has positive measure. Applying Lemma 2.3 we can partition the space X into disjoint towers of height at least N/ε whose bases are contained in X_0, that is, for each $n \geq N/\varepsilon$ we get disjoint towers $B^{(n)} \cup TB^{(n)} \cup \cdots \cup T^{n-1}B^{(n)}$ of height n with $B^{(n)} \subseteq X_0$, and the union of these towers has full measure. Partition the bases $B^{(n)}$ according to A_n, so for a word $(u, v) \in A_n$ the set $B^{(n)}_{u,v}$ consists of points whose $(\mathcal{P} \times \mathcal{Q}, n)$-name is (u, v).

We construct a partition $\mathcal{P}'$ by modifying the labels of some levels of the columns $B^{(n)}_{u,v}$. The construction proceeds in three stages.

MARKING THE BASE: Fix $m = 1/\varepsilon$ (for simplicity we ignore rounding errors and treat m as an integer, and adopt a similar philosophy later as well). Label the lower $2m$ levels of the column $B^{(n)}_{u,v}$ (i.e., the levels indexed 0 to $2m - 1$) with 1's and mark levels $2m, 3m, \ldots, [n/m]m$ with 0's.

The result of this procedure is that given any point $\omega \in \bigcup_{i=0}^{n-1} T^i B^{(n)}$ the base of the column can be identified as the largest index $i \in \{-n, -n+1, \ldots, 0\}$ such that the $(\mathcal{P}', 2m)$-name of $T^i\omega$ consists of all 1's. Thus given the $\mathcal{P}'$ itinerary of ω from time $-n$ to n, we can reconstruct the $\mathcal{P}$-name of the column to which ω belongs. We will preserve this property in the following steps, hence with probability 1 given the $\mathcal{P}'$ itinerary of a point from time $-\infty$ to $-\infty$ we can determine the n corresponding to the column the point belongs to, and the $\mathcal{P}'$-name of that column.

CODING THE $\mathcal{Q}$-ITINERARY INTO $\mathcal{P}'$: Denote $A_n(u) = \{v : (u, v) \in A_n\} \subseteq \mathbb{N}^n$. Fix $(u, v) \in A_n$ and enumerate $A_n(u) = \{v_1, \ldots, v_r\}$ in a way depending only on u; by assumption $|A_n(u)| < 2^{(h-h'+\varepsilon)n}$. We modify the column over $B^{(n)}_{u,v}$ so as to record the index i for which $v = v_i$. We do this by writing the base-k representation of i near the bottom of the column. To be precise, we record the base-k digits of i starting at level $2m+1$ and writing consecutively in blocks of $m-1$, skipping levels of height 0 mod m so as not to overwrite what we did in the previous stage. Since there are at most $2^{(h-h'+\varepsilon)n}$ possible values for i we need to overwrite $n(h-h'+\varepsilon)\log_k 2$ levels of the column.

The result of this procedure is that if we know both the $(\mathcal{P}, n)$-name (the word u) and the $(\mathcal{P}', n)$-name of a point in the base $B^{(n)}$, we can deduce its $(\mathcal{Q}, n)$-name (the word v) by extracting the index i coded just above the base marker in the $(\mathcal{P}', n)$ name, and looking at the i-th word in the list $A_n(u)$.

Re-coding the $\mathcal{P}$-itinerary: Fix again $(u, v) \in A_n$. The $\mathcal{P}$-name of the column $B^{(n)}_{u,v}$ has been partly destroyed by the previous steps. We will fix this by overwriting still more of the $\mathcal{P}$-name, starting where we stopped at the previous stage, skipping levels which are at height 0 mod m, and stopping at some height $M = M(n)$ which we will determine. This gives us $M - (2m + n/m + n(h-h'+\varepsilon)\log_k 2)$ symbols in which to store information. In this space we want to record the portion of the name u which has been overwritten in all three stages (including the current stage). This consists of the first M symbols of u plus at most n/m additional levels overwritten in the first stage. Assuming as we may that $M > \varepsilon n \geq N$, we know that the number of possibilities for the first M symbols of u is bounded by $2^{(h'+\varepsilon)M}$ so using the k symbols at our disposal we need $M(h'+\varepsilon)\log_k 2$ symbols in order to record it, plus another n/m symbols to record what was erased in the first stage. Thus we require of M that in addition to $\varepsilon n < M < n$ it satisfies the inequalities

$$M - (2m + n/m + n(h-h'+\varepsilon)\log_k 2) \geq M(h'+\varepsilon)\log_k 2 + n/m$$

or equivalently

$$M \geq \frac{((h-h'+\varepsilon)\log_k 2 + 2(1/m + m/n))n}{1-(h'+\varepsilon)\log_k 2}.$$

Since $h' \leq h < \log k$, $n/m = \varepsilon n$ and $m = \big(m/n\big)n = \big(1/(\varepsilon n)\big)n \leq \big(1/N\big)n \leq \varepsilon n$, when ε is small enough it suffices that

$$M \quad \geq \quad \frac{((h-h'+\varepsilon)\log_k 2 + 4\varepsilon)}{1-(h'+\varepsilon)\log_k 2}n.$$

Denote the coefficient of n in expression on the right hand side by $C(\varepsilon)$. Note that $C(\varepsilon) \to \frac{h-h'}{\log k - h'}$ as $\varepsilon \to 0$ and $0 \le C(\varepsilon) < 1$. Thus if we choose $\varepsilon > 0$ small enough (in a manner depending only on h, h' and k) we can set $M = \max\{\varepsilon, C(\varepsilon)\} \cdot n$ and M will satisfy all the requirements, including $\varepsilon n \le M \le n$.

The results of this procedure is that given the $(\mathcal{P}', n)$-name of a point in the base of the tower column $B_{u,v}^{(n)}$, we can reconstruct its $(\mathcal{P}, n)$-name by looking at the data written in this step, and hence by the previous step its $(\mathcal{Q}, n)$ name. Together with the previous stages, this means that for any point in X if we know the entire $\mathcal{P}'$ itinerary we now can determine the column it is in and the $\mathcal{P}'$-name of that column, and hence $\mathcal{Q}(\omega)$. This means that $\mathcal{P}'$ generates.

It remains to estimate how much $\mathcal{P}$ has changed. We have modified $M + n/m$ levels of each column $B_{u,v}^{(n)}$, or a $(C(\varepsilon) + \varepsilon)$-fraction of the mass of that column. Summing over all columns, this is the fraction of X that has changed. For $\varepsilon > 0$ sufficiently small, this is less than $\delta + \frac{h-h'}{\log k - h'}$, implying that $\rho(\mathcal{P}, \mathcal{P}') < \delta + \frac{h-h'}{\log k - h'}$. This completes the proof. ∎

3. Zero-entropy extensions

This section is dedicated to proving our main theorem, theorem 3.1. Before going into the details, we would like to say a few words about the relation of this theorem to the work of Ornstein and Weiss in [8], where it was shown that entropy is the only finitely observable invariant in some saturated classes of processes. Their proof used a diagonalization argument: Assuming to the contrary that for some class $\mathcal{C}$ there exists a finitely observable invariant finer than entropy, choose two non-isomorphic processes $(x_n), (y_n) \in \mathcal{C}$ with the same entropy h. A third process (z_n) is then constructed, for which the observation scheme does not converge. This is done by inductively defining the N-block distributions for the process (z_n) for a sequence of rapidly increasing Ns, where at each step copying lemmas are used to make (z_n) look at different time scales as though they come from $\mathcal{X}$ or $\mathcal{Y}$. However, in order to obtain a contradiction it must be ensured that $(z_n) \in \mathcal{C}$, since otherwise the observation scheme is not expected to converge. With some care one can ensure that (z_n) is Bernoulli if $h > 0$, or weak mixing and deterministic if $h = 0$, but other properties, such as pure point spectrum or non-Bernoullicity in positive entropy, are harder to build into (z_n).

Our results derive from the observation that when (x_n) is a zero-entropy extension of (y_n), one can control the isomorphism class of the diagonal process (z_n) and, in fact, it can be made isomorphic to (y_n).

THEOREM 3.1: *Suppose $\mathcal{X} \to \mathcal{Y}$ is a zero entropy extension of finite entropy dynamical systems. Let $\mathcal{C}$ be the family of processes arising from $\mathcal{X}$ and $\mathcal{Y}$. Then every finitely observable invariant for $\mathcal{C}$ is constant.*

Proof. We identify $\mathcal{Y}$ with the sub-σ-algebra of $\mathcal{X}$ which is the pull-back of the σ-algebra of $\mathcal{Y}$ through the factor map. Let $r \in \mathbb{N}$ with $\log r > h(\mathcal{X})$; all partitions in the sequel are partitions into r sets.

To simplify notation we assume that (s_n) is an observation scheme whose range is $\mathbb{R}$; there is no loss of generality here since given some other range, we can always compose with continuous functions from the range to $\mathbb{R}$. Suppose that there are $\xi, \eta \in \mathbb{R}$ such that for every pair of processes $(x_n), (y_n)$ arising from $\mathcal{X}, \mathcal{Y}$ respectively and generating them,

$$\begin{aligned} \lim s_n(x_1 \dots x_n) &= \xi \quad \text{in probability} \\ \lim s_n(y_1 \dots y_n) &= \eta \quad \text{in probability.} \end{aligned}$$

We must show that $\eta = \xi$. In order to do this will construct a generating partition $\mathcal{P}^*$ of $\mathcal{Y}$ and a sequence $N(k)$ such that $s_{N(k)}(y_1^*, \dots, y_{N(k)}^*) \to \xi$ in probability (here (y_n^*) is the process arising from $\mathcal{P}^*$). This suffices because by assumption, $\lim s_n(y_1^*, \dots, y_n^*) \to \eta$, so $\eta = \xi$.

The partition $\mathcal{P}^*$ will be obtained as the limit of a sequence of generating partitions $\mathcal{P}^{(k)}$ of $\mathcal{Y}$, which will be constructed inductively. The induction step is provided by the following lemma.

LEMMA 3.2: *Let $\mathcal{X} \to \mathcal{Y}$ be a zero-entropy extension of finite entropy dynamical systems. Let $\mathcal{C}$ be the class of processes arising from $\mathcal{X}$ and $\mathcal{Y}$. Let $J : \mathcal{C} \to \mathbb{R}$ be a finitely observable invariant computed by a sequence of functions s_n, and denote $J(\mathcal{X}) = \xi$. Then for any generating partition $\mathcal{P}$ of $\mathcal{Y}$, and any $\varepsilon > 0$, there is a generating partition $\overline{\mathcal{P}}$ of $\mathcal{Y}$ with $\rho(\mathcal{P}, \overline{\mathcal{P}}) < \varepsilon$, and an integer N so that*

$$P(|s_N(\overline{y}_1, \dots, \overline{y}_N) - \xi| < \varepsilon) > 1 - \varepsilon,$$

where $(\overline{y}_n)$ is the process arising from $\overline{\mathcal{P}}$.

Before proving the lemma let us show how it is used to prove the theorem. We construct a sequence $\mathcal{P}^{(k)}$ of generating partitions of $\mathcal{Y}$ and associated processes $(y_n^{(k)})$, starting with an arbitrary generating partition $\mathcal{P}^{(0)}$ provided by the Krieger generator theorem.

At the induction step, given $\mathcal{P}^{(k-1)}$ we construct $\mathcal{P}^{(k)}$ using the lemma; we choose the parameter $\varepsilon = \varepsilon(k) < 1/k$ in the lemma to be very small with respect to the previous stages of the construction (see below). Thus we have

$$\rho(\mathcal{P}^{(k-1)}, \mathcal{P}^{(k)}) < \varepsilon(k). \tag{3.1}$$

From the lemma we also get an integer $N(k)$ such that

$$P\Big(\Big|s_{N(k)}\Big(y_1^{(k)}, \ldots, y_{N(k)}^{(k)}\Big) - \xi\Big| < 1/k\Big) > 1 - 1/k, \tag{3.2}$$

and since $\mathcal{P}^{(k)}$ generates $\mathcal{Y}$ there is an integer $L(k)$ such that

$$\mathcal{P}^{(0)} \subseteq_{1/k} \bigvee_{i=-L(k)}^{L(k)} T^i\mathcal{P}^{(k)}. \tag{3.3}$$

During the construction we are free to choose the $\varepsilon(k)$ as small as we like. First of all we choose them so that $\sum \varepsilon(k) < \infty$. Since the metric $\rho = \rho_r$ is complete (or using the Borel–Cantelli lemma) this guarantees that $\mathcal{P}^{(k)}$ converges to a partition $\mathcal{P}^*$ of $\mathcal{Y}$, with associated process (y_n^*). Second, note that $\rho(\mathcal{P}^*, \mathcal{P}^{(k-1)}) \leq \sum_{m=k}^{\infty} \varepsilon(m)$. Thus, at the beginning of step k of the construction, when $\mathcal{P}^{(k-1)}$ is given, we may choose a parameter $\delta(k) > 0$ depending on all the data defined so far and ensure that $\rho(\mathcal{P}^*, \mathcal{P}^{(k-1)}) < \delta(k)$ by requiring $\varepsilon(m) \leq 2^{-m}\delta(k)$ for every $m \geq k$. The point is that the conditions (3.2) and (3.3) remain true for any partition (and associated process) sufficiently close to $\mathcal{P}^{(k)}$, and hence a prudent choice of $\delta(k)$ implies that they hold for $\mathcal{P}^*$ and (y_n^*), that is,

$$\forall m\; P(|s_{N(m)}(y_1^*, \ldots, y_{N(m)}^*) - \xi| < 1/m) > 1 - 1/m$$

and

$$\forall k\; \mathcal{P}^{(0)} \subseteq_{1/k} \bigvee_{i=-L(k)}^{L(k)} T^i\mathcal{P}^*.$$

The first of these implies $\lim_{k\to\infty} s_{N(k)}(y_1^*, \ldots, y_{N(k)}^*) = \xi$ in probability, and the second that $\mathcal{P}^{(0)} \subseteq \bigvee_{i=-\infty}^{\infty} T^i\mathcal{P}^*$, so $\mathcal{P}^*$ generates $\mathcal{Y}$. This completes the proof of Theorem 3.1 given Lemma 3.2 ∎

Proof of Lemma 3.2. First we present a sketch of the proof, and then give the details. Since $\mathcal{P}$ generates $\mathcal{Y}$ it has full entropy, which by assumption is equal to the entropy of $\mathcal{X}$. Therefore, we can find a generating partition $\mathcal{Q}$ for $\mathcal{X}$ with $\rho(\mathcal{P},\mathcal{Q}) < \varepsilon/2$. Let (x_n) be the process determined by $\mathcal{Q}$; then $s_n(x_1,\ldots,x_n) \to \xi$ in probability, so we can choose an N such that

$$P(|s_N(x_1,\ldots,x_N) - \xi| < \varepsilon) > 1-\varepsilon.$$

Since $\mathcal{P}$ and $\mathcal{Q}$ are both partitions of $\mathcal{X}$ we get a joining of the $\mathcal{P}$- and $\mathcal{Q}$-processes. Choose now a $\delta > 0$ and a suitably large K. Now working in $\mathcal{Y}$ again, we can construct a partition $\mathcal{R}$ whose joint K-block distribution with $\mathcal{P}$ is within δ of the joint K-block distribution of $\mathcal{P},\mathcal{Q}$. Thus (assuming we chose K large enough), the order of magnitude of $\rho(\mathcal{P},\mathcal{R})$ will be of the order of $\rho(\mathcal{P},\mathcal{Q})+\delta$, the N-block distribution of the $\mathcal{R}$-process will be within δ of the N-block distribution of the $\mathcal{Q}$-process, and the entropy of the $\mathcal{R}$-process is δ-close to $h(\mathcal{Y})$. though $\mathcal{R}$ does not necessarily generate $\mathcal{Y}$ we need only make an additional small correction to get a generating partition $\overline{\mathcal{P}}$ for $\mathcal{Y}$, and we can arrange that this does not disturb the N-block distributions very much.

Now for the details.

Choosing $\mathcal{Q}$: Since $h(\mathcal{X},\mathcal{P}) = h(\mathcal{Y}) = h(\mathcal{X})$, by theorem 2.7 we can find a generating partition $\mathcal{Q}$ for $\mathcal{X}$ with

$$\rho(\mathcal{P},\mathcal{Q}) < \varepsilon/2.$$

Choosing N and δ: Denote by (x_n) the process arising from $\mathcal{Q}$. Then $s_n(x_1,\ldots,x_n) \to \xi$ in probability, so there is an integer N such that

$$\mu(|s_N(x_1,\ldots,x_N) - \xi| < \varepsilon) > 1-\varepsilon.$$

Note that condition above is a property of the N-block distribution of (x_n). Thus there is a $\delta \in (0,\varepsilon/2)$ with the property that if (z_n) is a process arising from a partition $\mathcal{R}$ and the N-block distribution induced by $\mathcal{R}$ is within δ in L^1 of the N-block distribution of $\mathcal{Q}$, then $\mu(|s_N(z_1,\ldots,z_N) - \xi| < \varepsilon) > 1-\varepsilon$. Note also that if $\mathcal{R},\mathcal{R}'$ are two partitions of $\mathcal{Y}$ and if $\rho(\mathcal{R},\mathcal{R}') < \delta/N$ then the N-bock distributions of the processes arising from $\mathcal{R},\mathcal{R}'$ differ by at most δ.

Choosing α,β and M: Invoking Theorem 2.7, choose $\alpha > 0$ such that if $\mathcal{R}$ is a partition of $\mathcal{Y}$ with entropy $h-\alpha$ then there is a generating partition $\mathcal{R}'$ of

$\mathcal{Y}$ with $\rho(\mathcal{R},\mathcal{R}') < \delta/2N$. Let $\beta > 0$ be such that for any partition $\mathcal{S}$ of $\mathcal{Y}$, if $\mathcal{P} \subseteq_\beta \mathcal{S}$ then $h(\mathcal{S}) > h - \alpha$. We may assume that $\beta < \delta/N$.

Since $\mathcal{Q}$ generates $\mathcal{X}$ and $\mathcal{P}$ is measurable in $\mathcal{X}$ there is an $M > N$ such that

$$\mathcal{P} \subseteq_{\beta/2} \bigvee_{i=-M}^{M} T^i \mathcal{Q}.$$

Note that this property depends only on the distribution of $(\mathcal{P}\times\mathcal{Q}, 2M+1)$-names, and if $\mathcal{R}$ is a partition of $\mathcal{Y}$ such that the distribution of $(\mathcal{P}\times\mathcal{Q}, 2M+1)$-names is within τ of the distribution $(\mathcal{P}\times\mathcal{R}, 2M+1)$-names (in $\ell^1(\mathbb{R}^{2M+1})$) then $\mathcal{P} \subseteq_{\beta/2+\tau} \bigvee_{i=-M}^{M} T^i\mathcal{R}$.

Choosing L, B and $\mathcal{R}$: Fix an integer L with $\max\{M,N\}/L < \beta/8$ and choose a tower $B \cup TB \cup \cdots \cup T^L B$ of height L in $\mathcal{Y}$, filling all but $\beta/4$ of the space. We will define a partition $\mathcal{R}$ of $\mathcal{Y}$ by modifying $\mathcal{P}$ at some of the points in the tower.

Let (B_u) be the partition of the base B according to $(\mathcal{P}, L)$-names. This partition is measurable in $\mathcal{Y}$. We can further partition each B_u according to the $(\mathcal{Q}, L)$-names as $B_u = \bigcup_v B_{u,v}$. The $B_{u,v}$'s are measurable in $\mathcal{X}$ but may not be measurable in $\mathcal{Y}$. However, since $\mathcal{Y}$ is non-atomic we can partition the sets B_u into sets $B'_{u,v}$ in $\mathcal{Y}$ such that $\mu(B'_{u,v}) = \mu(B_{u,v})$. For each $B'_{u,v}$, modify the column over $B'_{u,v}$ so that it is labeled by v (instead of u). Call the resulting partition $\mathcal{R}$.

Since

$$\rho(\mathcal{P},\mathcal{R}) = 2\mu(\{x \in X \,:\, \mathcal{P}(x) \neq \mathcal{R}(x)\}),$$

and on the tower $\bigcup_{i=0}^{L-1} T^i B$ we have

$$\mu\left\{x \in \bigcup_{i=0}^{L-1} T^i B \,:\, \mathcal{P}(x) \neq \mathcal{R}(x)\right\} = \mu\left\{x \in \bigcup_{i=0}^{L-1} T^i B \,:\, \mathcal{P}(x) \neq \mathcal{Q}(x)\right\},$$

and the tower fills all but $\beta/4$ of the mass, it follows that

$$\rho(\mathcal{P},\mathcal{R}) \leq \rho(\mathcal{P},\mathcal{Q}) + \beta/4 < \varepsilon/2 + \beta/4.$$

Choosing $\overline{\mathcal{P}}$: Consider now the difference between the distributions of $(\mathcal{P}\times\mathcal{Q}, 2M+1)$-names and the distributions of $(\mathcal{P}\times\mathcal{R}, 2M+1)$-names. The only difference between them is incurred at the top and bottom M levels of the tower, which have total mass $< 2M/L < \beta/4$, and the exceptional set outside the tower whose mass is $< \beta/4$. Therefore, the distributions of $(\mathcal{P}\times\mathcal{Q}, 2M+1)$- and $(\mathcal{P}\times\mathcal{R}, 2M+1)$-names differ by at most $\tau = \beta/2$ so

$$\mathcal{P} \subseteq_{\beta/2+\beta/2} \bigvee_{i=-M}^{M} T^i\mathcal{R}.$$

Since the entropy of $\bigvee_{i=-M}^{M} T^i\mathcal{R}$ is the same as the entropy of $\mathcal{R}$, we conclude by the choice of β that $\mathcal{R}$ has entropy $> h - \alpha$. We can therefore choose a generating partition $\overline{\mathcal{P}}$ of $\mathcal{Y}$ with $\rho(\overline{\mathcal{P}}, \mathcal{R}) < \delta/2N$. We conclude that

$$\rho(\mathcal{P}, \overline{\mathcal{P}}) < \rho(\mathcal{P}, \mathcal{R}) + \rho(\mathcal{R}, \overline{\mathcal{P}}) < \varepsilon/2 + \beta/4 + \delta/(2N) < \varepsilon.$$

Finally, note that from the construction of $\mathcal{R}$, the N-block distribution is the same as the N-block distribution of Q except for an error introduced by the top N levels of the tower, which have mass $< \beta/4$, and the exceptional set also of measure $\beta/4$, which means that the N-block distribution of $\mathcal{R}$ and $\mathcal{Q}$ differ by less than $\delta/2$. Since $\rho(\mathcal{R}, \overline{\mathcal{P}}) < \delta/2N$ we see that the N-block distributions of the $\mathcal{R}$-process and the $\overline{\mathcal{P}}$-process differ by at most $\delta/2$, so the N-block distributions of the $\overline{\mathcal{P}}$-process and the $\mathcal{Q}$-process differ by at most δ; by the definition of δ this implies

$$\mu(|s_N(\overline{y}_1 \dots \overline{y}_N) - \xi| < \varepsilon) > 1 - \varepsilon,$$

where $(\overline{y}_n)$ is the process defined by $\overline{\mathcal{P}}$.

This completes the proof. ∎

4. Some Applications

An immediate consequence of theorem 3.1 is

Proposition 4.1: *Let $\mathcal{C}$ be a saturated class of processes with entropy h. Suppose that every $\mathcal{X}, \mathcal{Y} \in \mathcal{C}$ either have a common factor or a common extension in $\mathcal{C}$. Then every finitely observable invariant is constant on $\mathcal{C}$.*

Proof. If $\mathcal{X}$ and $\mathcal{Y}$ have a common factor $\mathcal{Z}$, then no scheme can distinguish $\mathcal{X}$ and $\mathcal{Z}$, and no scheme can distinguish $\mathcal{Y}$ and $\mathcal{Z}$; so every scheme must give the same value to $\mathcal{X}$ and $\mathcal{Y}$. The case of a common extension is similar. ∎

We turn now to some specific classes of processes. We begin by recovering some of the results of [8] using the techniques of the last section.

PROPOSITION 4.2 ([8]): *There are no nontrivial finitely observable invariants for the class of zero entropy systems or for the class of zero entropy weakly mixing processes.*

Proof. Any zero-entropy ergodic systems $\mathcal{X}$ and $\mathcal{Y}$ have an ergodic zero entropy joining (take a typical ergodic component of $\mathcal{X} \times \mathcal{Y}$), and if $\mathcal{X}$ and $\mathcal{Y}$ are zero entropy weakly mixing systems then so is the joining $\mathcal{X} \times \mathcal{Y}$. ∎

PROPOSITION 4.3 ([8]): *If $\mathcal{C}$ is a saturated family of processes which contains the Bernoulli processes (e.g. $\mathcal{C}$ = all aperiodic finite valued ergodic processes) then entropy is the only finitely observable invariant.*

Proof. For $h \geq 0$ let $\mathcal{C}_h = \{\mathcal{X} \in \mathcal{C} : h(\mathcal{X}) = h\}$. We must show that every finitely observable invariant scheme on $\mathcal{C}$ is constant on each $\mathcal{C}_h$. For $h = 0$ this is the previous proposition. For $h > 0$, we use Sinai's theorem, which states that every $\mathcal{X}, \mathcal{Y} \in \mathcal{C}_h$ have Bernoulli factors with entropy h. By Ornstein's isomorphism theorem, these factors are isomorphic. Since the Bernoulli processes are in $\mathcal{C}$ we conclude that every $\mathcal{X}, \mathcal{Y} \in \mathcal{C}_h$ have a common factor in $\mathcal{C}_h$, so every scheme is constant on $\mathcal{C}_h$. ∎

THEOREM 4.4:

(1) *Every finitely observable invariant for the class of Kronecker systems is constant*
(2) *Every finitely observable invariant for the class of mildly mixing zero entropy systems is constant.*
(3) *Every finitely observable invariant for the class of strong mixing zero entropy systems is constant.*

Proof. Again, we need only note that in these classes every two systems have a joining in the same class. ∎

An elementary class of systems is the class $\mathcal{R}$ of irrational rotations. A delicate and perplexing question is whether there exist nonconstant finitely observable invariants on this class.

To fix notation, let $([0,1), \mathcal{B}, \lambda)$ be the probability space of the unit interval with Lebesgue measure. For $\alpha \in [0,1) \setminus \mathbb{Q}$ let $\mathcal{X}_\alpha = ([0,1), \mathcal{B}, \lambda, T_\alpha)$ where $T_\alpha : [0,1) \to [0,1)$ is translation by α, that is, $T_\alpha(x) = x + \alpha (\mathrm{mod} 1)$. Let $\mathcal{R} = \bigcup \{\mathcal{X}_\alpha : \alpha \in [0,1) \setminus \mathbb{Q}\}$ be these systems (note that $\mathcal{X}_\alpha \cong \mathcal{X}_{-\alpha}$). Thus an invariant $J : \mathcal{R} \to \Delta$ induces a map $\widetilde{J} : [0,1) \setminus \mathbb{Q} \to \Delta$ by $\widetilde{J}(\alpha) = J(\mathcal{X}_\alpha)$.

LEMMA 4.5: *If J is a finitely observable invariant on $\mathcal{R}$, then $\widetilde{J}$ is Lebesgue measurable.*

Proof. We may assume that $\Delta = \mathbb{R}$ by composing continuous real-valued functions on s_n. Let (s_n) be an observation scheme which calculates J. Fix the partition $\mathcal{P} = ([0,1/2),[1/2,1))$ of the interval into two equal halves, and note that $\mathcal{P}$ generates for every $\mathcal{X}_\alpha \in \mathcal{R}$. Thus denoting by $(x_k^{(\alpha)})$ the process arising from $\mathcal{P}$ and the system $\mathcal{X}_\alpha$, we have

$$\widetilde{J}(\alpha) = J(\mathcal{X}_\alpha) = \lim_{n\to\infty} s_n(x_1^{(\alpha)},\ldots,x_n^{(\alpha)}),$$

where the limit exists in probability and is constant λ-a.e. in $\mathcal{X}_\alpha$.

Define $f_n : [0,1)\times[0,1) \to \Delta$ by

$$f_n(\alpha,\omega) = s_n(x_1^{(\alpha)}(\omega),\ldots,x_n^{(\alpha)}(\omega))$$

and $f : [0,1)\times[0,1) \to \Delta$ by

$$f(\alpha,y) = \widetilde{J}(\alpha).$$

To show that $\widetilde{J}$ is measurable it suffices to show that f is measurable, and, in fact, the f_n are measurable with respect to the product σ-algebra and since f_n converges in probability on every fiber $\{\alpha\}\times[0,1)$ (with respect to λ), and the limit is the constant function $J(\alpha)$, it follows that f_n converges to f in probability on $[0,1)\times[0,1)$ with respect to $\lambda\times\lambda$. ∎

THEOREM 4.6: *Let $J : \mathbb{R} \to \Delta$ be a finitely observable invariant for $\mathcal{R}$. Then $\widetilde{J}$ is constant on a set of full measure. In particular, no finitely observable invariant on $\mathcal{R}$ is complete.*

Proof. If $\alpha,\beta \in [0,1)\backslash\mathbb{Q}$ are rationally dependent then $\gamma = m\alpha = n\beta \in \mathbb{R}\backslash\mathbb{Q}$ for some $m,n \in \mathbb{N}$. Thus $\mathcal{R}_\gamma$ is a factor both of $\mathcal{R}_\alpha$ and of $\mathcal{R}_\beta$, so $J(\mathbb{R}_\alpha) = J(\mathbb{R}_\beta)$. We conclude that $\widetilde{J}$ is a Lebesgue-measurable function on $[0,1)\setminus\mathbb{Q}$ which is invariant to the action of multiplication by 2 mod 1. Any such map is constant on a set of full measure, because $x \mapsto 2x \mod 1$ is ergodic with respect to Lebesgue measure. ∎

5. Remarks and problems

Let us mention two problems which we have not been able to resolve:

PROBLEM: Is every finitely observable scheme on $\mathcal{R}$ constant?

Let $\mathcal{K}$ be the class of non-Bernoulli K-processes. Another problem is:

PROBLEM: Are there any finitely observable invariants on $\mathcal{K}$ finer than entropy?

It has been known for some time that there are no complete Borel invariants on this space (when it is topologized in a natural way — see Feldman's paper [2]). It also follows from work of Hoffman [4] that there exist nonisomorphic K-systems $\mathcal{X}, \mathcal{Y}$ of the same entropy such that $\mathcal{X} \to \mathcal{Y}$ is an extension. This implies by Proposition 4.1 that there are no complete finitely observable invariants on $\mathcal{K}$; but this is not new in view of Feldman's work.

If it were true that every two processes $\mathcal{X}, \mathcal{Y} \in \mathcal{K}$ had a common zero-entropy non-Bernoulli K-extension, then Proposition 4.1 would imply that there are no finitely observable invariants but entropy on $\mathcal{K}$. However, the existence of such a joining has been an open question for some time.

References

[1] D. Bailey, *Sequential Schemes for Clasifying and Predicting Ergodic Processes*, Stanford University, Ph.D. Dissertation, 1976.

[2] J. Feldman, *Borel structures and invariants for measurable transformations*, Proceedings of the American Mathematical Society **46** (1974), 383–394.

[3] P.R. Halmos, *Lectures on Ergodic Theory*, Chelsea Publishing Co., New York, 1960.

[4] C. Hoffman, *A K counterexample machine*, Transactions of the American Mathematical Society **351** (1999), 4263–4280.

[5] W. Krieger, *On entropy and generators of measure-preserving transformations*, Transactions of the American Mathematical Society **149** (1970), 453–464.

[6] D. Ornstein and B. Weiss, *How sampling reveals a process*, The Annals of Probability **18** (1990), 905–930.

[7] D. Ornstein and B. Weiss, *Entropy and data compression schemes*, Institute of Electrical and Electronics Engineers. Transactions on Information Theory **39** (1993), 78–83.

[8] D. Ornstein and B. Weiss, *Entropy is the only finitely observable invariant*, The Journal of Modern Dynamics **1** (2007), 93–105.

[9] P. Shields, *The Theory of Bernoulli Shifts*, The University of Chicago Press, Chicago, Ill.-London, 1973.

[10] P.C. Shields, *The Ergodic Theory of Discrete Sample Paths*, Graduate Studies in Mathematics, vol. 13, American Mathematical Society, Providence, RI, 1996.

[11] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982.

[12] J. Ziv and A. Lempel, *A universal algorithm for sequential data compression*, Institute of Electrical and Electronics Engineers. Transactions on Information Theory **IT-23** (1977), 337–343.