Can. J. Math., Vol. XXX, No. 3, 1978, pp. 531–548

RELATIVELY FINITELY DETERMINED IMPLIES RELATIVELY VERY WEAK BERNOULLI

M. RAHE

1. Introduction. In the isomorphism theory of Bernoulli shifts, the $\bar{d}$ or Hamming distance plays an important role. Two finite state stationary processes are close in the $\bar{d}$ metric if, on the average, strings from one process can be matched with strings from the other process so that their outputs agree with high frequency. Since such agreement must hold in the arbitrarily distant future, it might at first seem difficult to control the pairing. However in approximating processes with an independent generator (a coin flip, for instance) the $\bar{d}$ match is determined by only two parameters, the distribution at time zero and the entropy. The more general property of having the $\bar{d}$ distance from a process controlled by a joint distribution over a finite number of times and the entropy is called being *finitely determined* (FD) and implies isomorphism with a Bernoulli shift.

Unfortunately, with the exception of processes defined in terms of an independent generator, it is not obvious whether or not a given process is finitely determined. For this reason an alternative, more tractable criterion called *very weak Bernoulli* (VWB) was developed. This new property deals with a type of asymptotic independence of past and future events which implies finitely determined. Among other applications, VWB was used by Friedman and Ornstein to show that mixing Markov processes are finitely determined.

At the present there are a number of other metrics available which are different from but related to the $\bar{d}$. One of these, called $\bar{d}_H$, is used in the Thouvenot theory of isomorphism relative to or conditioned on a factor $\mathscr{H}$ generated by a finite partition H. The property of having the $\bar{d}_H$ approximation controlled by a finite number of parameters is called *conditionally finitely determined*. Corresponding to VWB is the property of being *conditionally very weak Bernoulli*.

Here, by the word *factor* we mean the action of a transformation T on an invariant sub-σ-algebra. The Thouvenot theory is concerned with how the factor $\mathscr{H}$ is embedded in a transformation. Viewed from the standpoint of the original process we are concerned with questions of the existence of isomorphisms of the original process which take a finite partition H onto a specified image partition H'. Viewed from the standpoint of a given fibre in the process (H, T), we are concerned with the behavior of the non-stationary process that the original process produces on that fibre.

The known results concerning the properties very weak Bernoulli condi-

Received February 21, 1977 and in revised form, November 14, 1977.

tioned on a factor and finitely determined conditioned on a factor are as follows:

1) any conditionally finitely determined process can be written as the direct product of a Bernoulli shift with the factor;

2) in a process which is finitely determined conditioned on a factor $\mathscr{H}$, any factor containing $\mathscr{H}$ is again finitely determined conditioned on $\mathscr{H}$; and

3) any conditionally very weak Bernoulli process is conditionally finitely determined.

In the standard isomorphism theory, there is a result due to Ornstein and Weiss which states that FD implies VWB, hence that the properties are equivalent. The purpose of this paper is to establish the corresponding result for the Thouvenot theory by showing that conditionally finitely determined (relative to some factor $\mathscr{H}$) implies conditionally very weak Bernoulli (relative to $\mathscr{H}$). (It is perhaps pertinent to point out that not only is this the corresponding result for the Thouvenot theory; but also, when $\mathscr{H}$ is the trivial factor, the result is the Ornstein-Weiss theorem itself.)

Although the extension of the Ornstein-Weiss result to the conditional case is not difficult, it is somewhat intricate. Therefore, to give the reader a framework within which to follow the conditional version, it is perhaps worthwhile to give an informal, simplified version of the original Ornstein-Weiss argument.

Given a stochastic process (T, P), the Ornstein-Weiss result states that the concurrent assumptions that (T, P) is not VWB and that (T, P) is FD yield a contradiction. One uses the assumption that (T, P) is not VWB to construct a sequence of processes (T_l, P_l) which converge to (T, P) in entropy and in finite distributions. If one assumes that (T, P) is FD, such convergence must also hold in the $\bar{d}$ sense. The contradiction arises by showing that there is a value $c > 0$ such that for all l one has $\bar{d}[(T_l, P_l), (T, P)] > c^2/100$.

To say that (T, P) is VWB means that for each $c > 0$ there is an l such that for all $m \geqq 0$, $\bar{d}(\{T^iP|A\}_1{}^l, \{T^iP\}_1{}^l) < c$ on a set $\mathscr{A}$ of atoms $A \in \bigvee_{-m}^{0} T^iP$, with $\mu(\mathscr{A}) > 1 - c$. By negating the definition of VWB, one gets a fixed value $c > 0$ such that for all l one has some $m \geqq 0$ so that $\bar{d}(\{T^iP|A\}_1{}^l, \{T^iP\}_1{}^l) \geqq c$ on a set $\tilde{\mathscr{A}}_{l,m}$ with $\mu(\tilde{\mathscr{A}}_{l,m}) \geqq c$, where $\tilde{\mathscr{A}}_{l,m}$ is a union of atoms $A \in \bigvee_{-m}^{0} T^iP$.

For each pair (l, A) where l is a positive integer and $A \in \bigvee_{-m}^{0} T^iP$ one can define a process $(T_l{}^A, P_l{}^A)$ in the following way. Consider l disjoint copies of A, stacked one on top of another in the standard tower construction. Partition the base of the tower according to $\bigvee_1^l T^iP|A$. Label the levels in the column above each atom in $\bigvee_1^l T^iP|A$ with the corresponding P name. The tower is the space on which $T_l{}^A$ operates; and $P_l{}^A$ is its partition according to the labelling. Points in the space are moved vertically up the tower by $T_l{}^A$; and the transformation induced by returns to the base, along with the partition on the base, gives an independent process with distribution $\bigvee_1^l T^iP|A$. For each l the process (T_l, P_l) is selected from the class $(T_l{}^A, P_l{}^A)$ by choosing an atom A_l. [Note that each pair (l, A) gives a distribution of certain l-strings and that the tower construction is merely a way of concatenating these strings to form infinite strings.]

As l grows, for each fixed n, the frequency distribution of n-strings in l-strings becomes arbitrarily close to the distribution $\bigvee_1^n T^i P$ on a set of measure arbitrarily close to one. As l grows, the average partition entropy $(1/l)E_A(\bigvee_1^l T^i P)$ becomes arbitrarily close to the entropy of (T, P) on a set of A with measure converging uniformly in m to one. [See below for more detail.] Since for all choices of l and m one has $\mu(\tilde{\mathscr{A}}_{l,m}) \geqq c$, for l large these sets have non-void intersection; and one can choose A_l such that $\bigvee_1^l T^i P|A_l$ will have arbitrarily good n-distribution, $(1/l)E(\bigvee_1^l T^i P|A_l)$ will be arbitrarily close to the entropy of (T, P), and $\bar{d}(\{T^i P|A_l\}_1{}^l, \{T^i P\}_1{}^l) \geqq c$. By means of Abramov's result relating the entropy of the process induced on the base of a tower to the entropy of the tower process, one can see that the second property forces (T_l, P_l) to converge to (T, P) in entropy; and the convergence in finite distributions is clear from the first condition. The third condition bounds the sequence of processes (T_l, P_l) away from (T, P) in $\bar{d}$.

[The statement at the beginning of the last paragraph about distribution follows from the ergodic theorem. The assertion about entropy requires a detailed computation which is merely encapsulated here. Since the average $\sum_A \mu(A)(1/l)E_A(\bigvee_1^l T^i P)$ of the values $(1/l)E_A(\bigvee_1^l T^i P)$ is close to the entropy of (T, P), one needs only to show that $(1/l)E_A(\bigvee_1^l T^i P) \leqq E(T, P) + \epsilon$ on a set of large measure, in order to get a similar lower bound for the values on a set of measure nearly as large. There are two methods to show that $(1/l)E_A(\bigvee_1^l T^i P) \leqq E(T, P) + \epsilon$. The first involves a counting argument. One observes that if an l-string in (T, P) has good frequency distribution of n-strings, then the same l-string must occur in the $(n-1)$-step Markov approximation to (T, P), since the Markov measure of each such string is bounded below, uniformly, by a constant greater than zero. Since the Markov measures are bounded below, one has an upper bound on the number of l-strings in the original (T, P) which have good distribution. Since for large l one has that most atoms A are mostly covered by l-strings with good distribution, and since $E_A(\bigvee_1^l T^i P)$ is bounded above by the logarithm of the number of l-strings, one gets $(1/l)E_A(\bigvee_1^l T^i P) \leqq E(T, P) + \epsilon$.

The second method to show this inequality involves the fact that a finitely determined transformation is a K-automorphism. This approach is used in the body of this article and is explained in detail there.]

The subtlety of the Ornstein-Weiss result lies in the method of showing that the sequence of processes (T_l, P_l) remains bounded away from (T, P) in the $\bar{d}$ sense. [Note that it is not trivial to show the easier result that the distribution of l-strings from (T_l, P_l) does not match well in $\bar{d}$ with the distribution of l-strings from (T, P). While it is true that the processes (T_l, P_l) were constructed using distributions of l-strings which were chosen specifically not to match $\bigvee_1^l T^i P$, the distribution $\bigvee_1^l (T_l)^i P_l$ is not necessarily the same as

$\bigvee_1^l T^iP|A_l$, which is the distribution conditioned on the base of the tower.] If one had good $\bar{d}$ matching between (T_l, P_l) and (T, P) there would be a pair of generic strings, one from each process, so that each output agreed with high frequency with the corresponding output in the other string. By cutting the double string into segments of length l, based on where the P_l-l-strings begin, one gets a matching of l-strings, one from each process, by looking across to the other string in each segment. By examining the frequency with which l-strings are matched by these segments from the pair of generic strings, one gets a joint distribution of l-strings for a $\bar{d}$ match between (T_l, P_l) and (T, P). If the generic strings match very well, the distributions of l-strings will match almost as well.

Since the locations of the segments were defined by the beginnings of P_l-l-strings, the distribution of l-strings from (T_l, P_l) given by the segments is the distribution on the base of the tower which was used to define (T_l, P_l); namely, $\bigvee_1^l T^iP|A_l$. [Note that $(T_l)^l$ is not ergodic, so that the distribution can depend on where segments occur.] On the other hand, the distribution of l-strings from (T, P) given by the segments is $\bigvee_1^l T^iP$, since T^i is ergodic if (T, P) is FD. Thus if $\bar{d}[(T, P), (T_l, P_l)] < c^2/100$, one has $\bar{d}(\{T^iP|A_l\}_1{}^l, \{T^iP\}_1{}^l) < 2c/10$; but by construction $\bar{d}(\{T^iP|A_l\}_1{}^l, \{T^iP\}_1{}^l) \geqq c$, which is a contradiction.

In the light of this simplified version of the Ornstein-Weiss argument, one can more easily observe the complications that arise when one requires that the processes (T_l, P_l) be mixing. Under the new construction [which we omit], generic strings from (T_l, P_l) no longer consist exclusively of concatenated l-strings chosen from a distribution on some past atom A. Now the strings are occasionally connected by a string of one or more outputs of a special state, denoted by zero. Occurrences of zero are determined by a zero entropy mixing process. In the simplified argument presented above, when showing that (T_l, P_l) and (T, P) were not close in $\bar{d}$, it was possible to divide the pair of generic strings into segments of length exactly l. In the mixing case, segments are cut at intervals which are only approximately l outputs apart, where the exact location is determined by the beginnings of P_l-l-strings, which are in turn governed by outputs from the zero entropy process. Whereas in the simplified version it was easy to see that the distribution of l-strings in (T, P) as determined by the beginnings of segments was $\bigvee_1^l T^iP$, with mixing it is necessary to apply a result of Pinsker to show that the distribution is $\bigvee_1^l T^iP$. Pinsker's result states that when K-automorphisms and zero entropy processes are factors of a given process, they are always orthogonal. Since (T, P) is FD, it is a K-automorphism; hence it is orthogonal to the zero entropy process that determines segements, when both are embedded in the ergodic process which gives pairs of generic strings for $\bar{d}$ matching of (T, P) and (T_l, P_l).

It is, of course, not necessary to use processes (T_l, P_l) which are mixing in order to have convergence in $\bar{d}$ to a finitely determined process (T, P) : ergodicity is sufficient. Thus the result due to Pinsker is not an inherent part of the basic Ornstein-Weiss result, although it is used in their paper. However, in dealing with the extension of Ornstein-Weiss to the Thouvenot theory, Pinsker's result, or rather a conditionalized version of it, seems to occur more naturally. The factor (T, H) with which one works may be badly behaved, say T^l is non-ergodic, for each $l \geqq 2$, when restricted to the factor. An initial difficulty in the extension is how to construct processes $(T_l, P_l \vee H)$ which contain a copy of the factor (T, H) and which are ergodic, using no properties of (T, H) other than ergodicity. [Note for instance that concatenation of $(P \vee H)$-l-strings will not work.] The construction technique chosen to deal with possible lack of mixing properties of (T, H) leads one inexorably to the difficulties inherent in the mixing version of the Ornstein-Weiss proof, where in the $\bar{d}$ matching the segments are determined by a zero entropy process; but this time the arguments have a conditional flavor.

2. Preliminaries. We assume familiarity with the basic definitions of the $\bar{d}$ metric; and we shall use the notation of [**3**].

Throughout this paper T will be an ergodic, invertible, measure-preserving transformation of $(X, \mathscr{B}, \mu)$, P a partition of X into v-sets, and H a finite partition of X. Then (T, P) and (T, H) define, in the usual way, finite valued stationary stochastic processes.

The following definition of conditionally finitely determined is from [**1**, p. 181].

Definition. Let T be an ergodic transformation on X, and H and P two finite partitions of X. We say that the partition P is *H-conditionally finitely determined* $(H - FD)$ if for every $\epsilon > 0$ there exists $\delta > 0$ and a positive integer n such that for every ergodic transformation T' on a space Y, with H' and P' two finite partitions of Y, the following conditions

(1) $d\left(\bigvee_0^m T'^i H'\right) = d\left(\bigvee_0^m T^i H\right)$ for every m

(2) $d\left(\bigvee_0^n T'^i (P' \vee H'), \bigvee_0^n T^i (P \vee H)\right) < \delta$

(3) $|E(P \vee H, T) - E(P' \vee H', T')| < \delta$

imply that there exists a Lebesgue space Z and, for each integer $p > 0$, sequences of partitions of Z, $\{H_i\}$, $\{P_i\}$, $\{P_i'\}$ $0 \leqq i \leqq p$ such that

(4) $d\left(\bigvee_0^p T^i (P \vee H)\right) = d\left(\bigvee_0^p (P_i \vee H_i)\right)$

(5) $d\left(\bigvee_0^p T'^i (P' \vee H')\right) = d\left(\bigvee_0^p (P_i' \vee H_i)\right)$

(6) $|P_i - P_i'| < \epsilon.$

We shall say that a transformation T is *finitely determined conditioned on the* *factor* $\bigvee_{-\infty}^{+\infty} T^iH$ generated by H under T if there is a finite partition P such that $P \vee H$ generates under T and P is finitely determined conditioned on H.

We shall also use Corollary 5.1 from [**1**, p. 199].

COROLLARY 5.1. *Let (X, T) be an ergodic dynamical system and B and H two partitions of X satisfying*

$$(1)\ X = \bigvee_{-\infty}^{+\infty} T^i(B \vee H)$$

$$(2)\ \bigvee_{-\infty}^{+\infty} T^iB \perp \bigvee_{-\infty}^{+\infty} T^iH$$

(3) *The T^iB, for i integral, are independent.*

Let $\bigvee_{-\infty}^{+\infty} T^iP$ be a factor of T containing the factor $\bigvee_{-\infty}^{+\infty} T^iH$. Then if $E(P, T) = E(H, T)$ we have in fact that $\bigvee_{-\infty}^{+\infty} T^iP = \bigvee_{-\infty}^{+\infty} T^iH$.

This result states that if P is finitely determined conditioned on the factor $\bigvee_{-\infty}^{+\infty} T^iH$, then the factor is maximal in its entropy class relative to T. As shown in the appendix to this paper, for a factor to be maximal in its entropy class relative to T is tantamount to saying that T is a K-automorphism conditioned on the factor. Thus if P is $H - FD$, P is $H - K$.

We now define the term *conditionally very weak Bernoulli.*

Definition. Given an ergodic transformation T with finite partitions P and H, we say that P is $H - VWB$ if given $c > 0$ there exists l such that for all $m > 0$ and all k sufficiently large (with values of l, m, k integers) there is a collection D_k of atoms $Q \in \bigvee_{-k}^{k} T^iH$ with $\mu(D_k) > 1 - c$ and on each atom $Q \in D_k$ there is a collection $F_{Q,m}$ of atoms $C \in (\bigvee_{-m}^{0} T^iP) \cap Q$ with $\mu_Q(F_{Q,m}) > 1 - c$ for which $\bar{d}(\{T^iP|Q\}_1{}^l, \{T^iP|Q \cap C\}_1{}^l) \leqq c$.

Upon occasion we shall use the notation $E(P; T, H)$, the entropy of P relative to T and H. We define $E(P; T, H) = (P \vee H, T) - E(H, T) = E(P|\bigvee_{-\infty}^{-1} T^iP \bigvee_{-\infty}^{+\infty} T^iH)$. Moreover, the notation $E_Q(\bigvee_1^l T^iP)$ will indicate the entropy of the partition $\bigvee_1^l T^iP$ computed using the conditional measure $\mu_Q(\cdot) = \mu(\cdot \cap Q)/\mu(Q)$.

3. Relatively finitely determined implies relatively very weak Bernoulli.

THEOREM. *Given a transformation T with generating partition $P \vee H$ where P is $H - FD$, then P is $H - VWB$.*

Proof. We will show that the concurrent assumptions P is not $H - VWB$ and P is $H - FD$ lead to a contradiction. The idea of the proof is to use not $H - VWB$ but $H - FD$ to construct a sequence of processes $(T_n, P_n \vee H)$ so that

(1) (T_n, H) has the same joint distributions as (T, H), which we denote by $(T, H) \sim (T_n, H)$;

https://doi.org/10.4153/CJM-1978-048-7 Published online by Cambridge University Press

(2) the $(T_n, P_n \vee H)$ approximate $(T, P \vee H)$ arbitrarily well in entropy and in finite joint distributions,

(3) but the $(T_n, P_n \vee H)$ remain bounded away from $(T, P \vee H)$ in $\bar{d}_H$.

Such a result contradicts $H - FD$.

For the construction we begin by choosing a sequence $\epsilon_n \to 0$. For each n the following proposition then gives us a source of long P strings which we concatenate in a special way to form generic strings for the (T_n, P_n) process.

PROPOSITION. *There exists $c > 0$ such that integers l_n, k_n, and m_n may be chosen in such a way that on a set $\tilde{D}_{k_n}$ of atoms $Q \in \bigvee_{-k_n}^{k_n} T^iH$ with $\mu(\tilde{D}_{k_n}) > c/2$ we can find at least one atom $C \in \bigvee_{-m_n}^{0} T^iP$ for which*

$$(1) \quad \left\| \text{dist } (P \vee H)\text{-}n\text{-blocks in } \bigvee_1^{l_n} T^i(P \vee H) \mid Q \cap C \right.$$
$$\left. - \text{ dist } (P \vee H)\text{-}n\text{-blocks on } X \right\| \leqq \epsilon_n$$

$$(2) \quad \left| \frac{1}{l_n} E_{Q \cap C}\left(\bigvee_1^{l_n} T^iP \right) - E(P; T, H) \right| \leqq \epsilon_n$$

(3) *for* $A \in \bigvee_1^{l_n} T^iP$,

$$\sum_A |\mu(A|Q) - \mu(A|\tilde{Q})| < c/10$$

for all smaller atoms $\tilde{Q}$ (i.e., longer H-strings) with $\tilde{Q} \subseteq Q$.

(4) $\bar{d}(\{T^iP|Q \cap C\}_1^{l_n}, \{T^iP|Q\}_1^{l_n}) \geqq c.$

Remark. We will characterize atoms Q according to whether or not there is an atom C in the past of the (T, P) process so that the $\bar{d}$ distance is at least c on $Q \cap C$. Those atoms having such a C will be called "bad"; and the others, "good". The idea of the proof is that the statement "not $H - VWB$" gives a set of fixed measure c of "bad" atoms Q. However by choosing large l_n and k_n we can guarantee good behavior of distribution and entropy on a set of atoms Q of measure at least $1 - \epsilon_n$. Hence we will eventually get an overlap of measure at least $c/2$ of "bad" atoms Q with good entropy and distribution on the particular $Q \cap C$ which makes Q "bad".

Proof. The order of selection of l_n, k_n, and m_n is as follows:

1) In order to have the distribution of $(P \vee H)$-n-blocks at least ϵ_n-good on most of the space, we use the ergodic theorem. This requirement places a minimum value on l_n and k_n.

2) In order to have good entropy on most atoms of the space, we wait until the entropy of the P partition "sets up" relative to T, say at time L, then go out much further along the (T, H) process so that beyond K all the information in the (T, H) process is subtracted off when we condition on the H-strings. Next, having chosen a value for L, we get a separation N, after which the

assumption $H - K$ will give ϵ-independence of past and future P-strings when conditioned on long H-strings. Thus we now have a second set of minimum values for l_n and k_n:

i) we need l_n so large that N and L are small compared to it;

ii) we need k_n so large that no point of the $P - l_n$-string is closer than K to the end of the H-string, and

iii) we need k_n large enough for ϵ-independence to hold for the values L and N.

3) We now choose l_n larger than either minimum value. However, since later in the proof of the theorem we may wish to look at conditioning by extremely long strings, we now place a new requirement on k_n; namely, that for our choice of l_n the distribution of $P - l_n$-strings will not change much. The martingale convergence theorem then gives us a new minimum on k_n.

4) Finally we go to the statement "not $H - VWB$" to get the final minimum value on k_n for our choice of l_n. We now choose k_n larger than any of the minimum values so that for our choice of l_n, k_n the statement "not $H - VWB$" then gives us an m_n.

The details of statements (1) and (3) of the proposition are clear and are omitted. We include, however, a more careful proof of statements (4) and (2)

First, the assumption that P is not $H - VWB$ implies that there exists. some $c > 0$ such that for every l there are k sufficiently large compared with l and some m so that if $\tilde{D}_k$ is the set of atoms $Q \in \bigvee_{-k}^{k} T^iH$ on each of which there is a collection $\tilde{F}_{Q,m}$ of atoms $C \in (\bigvee_{-m}^{0} T^iP) \cap Q$ with $\mu_Q(\tilde{F}_{Q,m}) > c$ for which

$$\bar{d}(\{T^iP|Q\}_1{}^l, \{T^iP|Q \cap C\}_1{}^l) > c, \text{ then } \mu(\tilde{D}_k) > c.$$

This result is simply the negation of the definition and provides the basis for (4).

Next, we deal with statement (2), whose proof is developed in the following sequence of lemmas. In the first two lemmas, we use the assumption that P is $H - K$ to establish an upper bound: $(1/l)E_{Q \cap C}(\bigvee_1^l T^iP) \leqq E(P; T, H) + \epsilon$ for most of the space. The third lemma merely observes that such an upper bound on most of the space implies a lower bound for most of the space, as well.

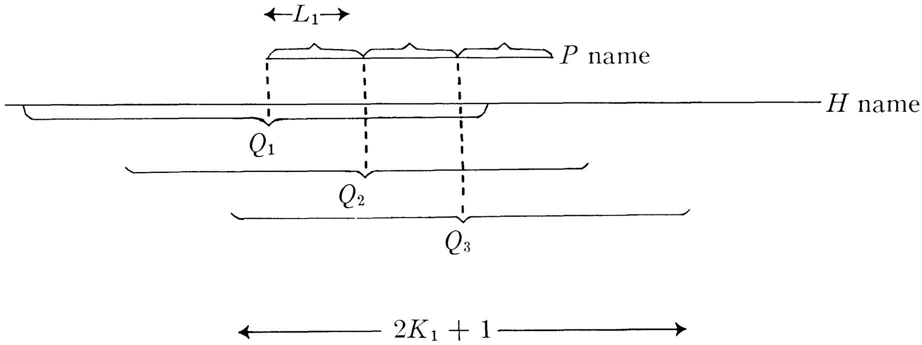
LEMMA 1. *Given $\epsilon > 0$ there exist L, a set F with $\mu(F) > 1 - \epsilon$ and a K such that for all $\tilde{\tilde{Q}} \in \bigvee_{-k}^{k} T^iH$ with $k > K$ and $\tilde{\tilde{Q}} \cap F \neq \emptyset$ we have*

$$\frac{1}{L} E_{\tilde{\tilde{Q}}}\left(\bigvee_1^L T^iP\right) \leqq E(P; T, H) + \epsilon.$$

Proof. Choose L_1 and K_1 such that

$$\frac{1}{L_1} E\left(\bigvee_1^{L_1} T^iP \;\middle|\; \bigvee_{-K_1}^{K_1} T^iH\right) \leqq E(P; T, H) + \epsilon/4.$$

(Assure that the martingale convergence theorem controls $P - L_1$-string distribution on $k - H$-strings, $k \geqq K_1$). The conditionnl entropy is a weighted average of entropies conditioned on individual H-strings. If we break long strings of P-outputs into shorter pieces of length L_1, the ergodic theorem then tells us that the individual $H - (2K_1 + 1)$-strings on which the pieces sit will give the same weighted average as before.



We thus get an $L \gg L_1$, a set F with $\mu(F) > 1 - \epsilon$ and a $K \gg K_1$ such that

$$\frac{L_1}{L} E_Q\left(\bigvee_1^L T^iP\right) \leqq \frac{1}{L}\sum_{q=0}^{L-L_1} E_{\tilde{Q}}\left(\bigvee_{q+1}^{q+L_1} T^iP\right) + \frac{L_1\epsilon}{4}$$

$$\leqq \sum_{Q\in\bigvee_{-K}^{K} T^iH} [\text{relative frequency in } \tilde{Q} \text{ name of } H\text{-}(2K_1+1)$$

$$\text{-block corresponding to } Q]\cdot\left[E_Q\left(\bigvee_1^{L_1}\right)T^iP + \frac{L_1\epsilon}{4}\right]$$

$$+ L_1\epsilon/4 \leqq L_1[E(P;T,H) + \epsilon].$$

LEMMA 2. *Let P be $H - K$. Then given $\epsilon > 0$ there is a set F with $\mu(F) > 1 - \epsilon^2$ and a value $\bar{L}$ such that for all $l > \bar{L}$ there is a $\bar{K}$ such that for all m and all j, $k > \bar{K}$, $\tilde{Q} \cap C \in \bigvee_{-j}^{k} T^iH \bigvee_{-m}^{0} T^iP$ with $\tilde{Q} \cap C \cap F \neq \emptyset$ implies that we have*

$$\frac{1}{l} E_{\tilde{Q}\cap C}\left(\bigvee_1^l T^iP\right) \leqq E(P;T,H) + \epsilon(2 + \log v)$$

Proof. Use the previous lemma to choose L and K such that

$$\frac{1}{L} E_Q\left(\bigvee_1^L T^iP\right) \leqq E(P;T,H) + \epsilon^2/3.$$

We may also assure that the martingale convergence theorem controls distribution of $P - L$-strings on long H-strings. For fixed L we use the assumption P is $H - K$ to find a separation N so that we can remove the conditioning by atoms C in the remote past to get

$$\frac{1}{L} E_{Q\cap C}\left(\bigvee_1^L T^iP\right) \leqq \frac{1}{L} E_Q\left(\bigvee_1^L T^iP\right) + \epsilon \leqq E(P;T,H) + 2\epsilon.$$

If we break long strings of P outputs from time zero to time l into pieces of length L, and if we then translate back to time zero, the net effect is to move the conditioning atom C into the remote past. Take $\bar{L} \gg L$ and $l > \bar{L}$.

Denote by G the collection of atoms $\tilde{Q} \cap C \in \bigvee_{-j}^{k} T^i H \bigvee_{-m}^{0} T^i P$ for which $(1/l)$ [number of s, $1 \leqq s \leqq l$, such that $T^{-s}(\tilde{Q} \cap C)$ is well behaved] $> 1 - \epsilon$. Then $\mu(G) > 1 - \epsilon$ and for $\tilde{Q} \cap C \cap G$ we have that, letting

$$\alpha(s) = [(l - N - s)/L],$$

$$\begin{aligned}\frac{L}{l} E_{\tilde{Q} \cap C}\left(\bigvee_{1}^{l} T^i P\right) &\leqq \frac{1}{l} \sum_{s=0}^{L-1} \left[E_{\tilde{Q} \cap C}\left(\bigvee_{1}^{N+s} T^i P\right) + \sum_{q=0}^{\alpha-1} E_{Q \cap C}\left(\bigvee_{N+s+Lq+1}^{N+s+Lq+L} T^i P\right) \right. \\ &\qquad \left. + E_{\tilde{Q} \cap C}\left(\bigvee_{\alpha L+N+s}^{l} T^i P\right)\right] \\ &\leqq L[E(P; T, H) + \epsilon(2 + \log v)].\end{aligned}$$

LEMMA 3. *Given $\epsilon > 0$ there is a set F with*

$$\mu(F) > 1 - [\epsilon(2 + \log v)] - [\epsilon(2 + \log v + \epsilon \log v)]^{1/2},$$

an L and for each $l > L$ a K such that for all m and all $Q \cap C \in \bigvee_{-m}^{0} T^i P \bigvee_{-k}^{k} T^i H$ with $Q \cap C \cap F \neq \emptyset$ we have that

$$\left| \frac{1}{l} E_{Q \cap C}\left(\bigvee_{1}^{l} T^i P\right) - E(P; T, H) \right| < [\epsilon(2 + \log v + \epsilon \log v)]^{1/2}.$$

Proof. Given $\epsilon > 0$ we can use the previous lemma to find a set F_1 with $\mu(F_1) > 1 - \epsilon^2$ and an L such that for each $l > L$ there is a K for which

$$\frac{1}{l} E_{Q \cap C}\left(\bigvee_{1}^{l} T^i P\right) \leqq E(P; T, H) + \epsilon(2 + \log v)$$

when $Q \cap C \in \bigvee_{-m}^{0} T^i P \bigvee_{-j}^{k} T^i H$ with j, $k > K$ and $Q \cap C \cap F_1 \neq \emptyset$.

For $Q \cap C \cap F_1 = \emptyset$ we have the universal bound

$$\frac{1}{l} E_{Q \cap C}\left(\bigvee_{1}^{l} T^i P\right) \leqq \log v.$$

We now use the fact that given an upper bound for a set of numbers and given their average, we can determine a lower bound for the same set.

Since

$$\begin{aligned}E(P; T, H)' \leqq \frac{1}{l} E\left(\bigvee_{1}^{l} T^i P \,\middle|\, \bigvee_{-m}^{0} T^i P \bigvee_{-j}^{k} T^i H\right) \\ = \sum \mu(Q \cap C) \cdot \frac{1}{l} E_{Q \cap C}\left(\bigvee_{1}^{l} T^i P\right),\end{aligned}$$

and since the right hand side can exceed $E(P; T, H)$ by at most $2\epsilon + \epsilon \log v + \epsilon^2 \log v$, we can have at most a set of $Q \cap C$ atoms of measure $[\epsilon(2 + \log v + \epsilon \log v)]^{1/2}$ on which

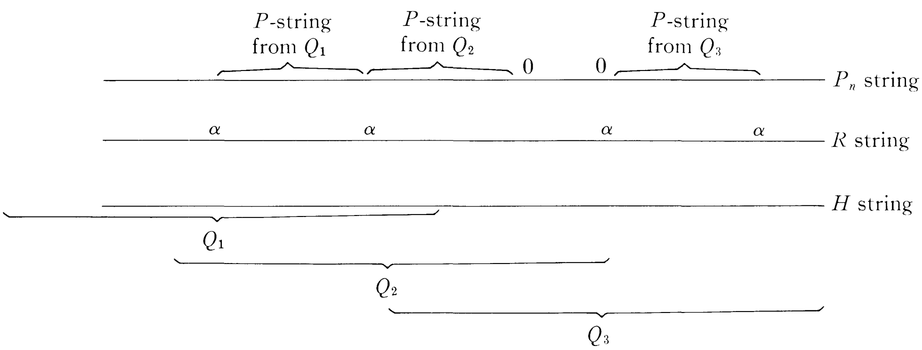
$$\frac{1}{l} E_{Q \cap C}\left(\bigvee_{1}^{l} T^i P\right) < E(P; T, H) - [\epsilon(2 + \log v + \epsilon \log v)]^{1/2}.$$

4. The construction.

Description of the construction. For each ϵ_n the proposition gives us a distribution of $P - l_n$-strings associated with "bad" atoms $Q \in \tilde{D}_{k_n}$ by looking at $\bigvee_1^{l_n} T^i P | Q \cap C$, where C is the past atom for which $Q \cap C$ has good distribution and entropy but bad $\bar{d}$ match. Similarly, for all except ϵ_n of the other atoms Q, even though there may not be an atom C with bad $\bar{d}$ match on $Q \cap C$, we may still choose some $Q \cap C$ with good entropy and distribution and use the resulting distribution $\bigvee_1^{l_n} T^i P | Q \cap C$. For the remaining set of Q atoms with total measure at most ϵ_n, we choose $Q \cap C$ arbitrarily. We wish to use $P - l_n$-strings from these distributions to lay along a generic H-string in a special way. The resulting partition will be denoted by P_n and the measures of each atom will be determined by limiting frequencies in the string.

We begin with a zero entropy mixing process S. Then construct a Rohlin tower of height l_n with residual error 2^{-l_n}. If we designate the base of the tower by α and the rest of the space by β we get a partition R_n and a zero entropy mixing process (S, R_n) which prints out α's some l_n units apart, except for an occasional longer gap between α's.

By considering the transformation T' which is T restricted to the factor generated by H, we can construct the direct product $(S, R_n) \times (T', H)$. Since the product of a mixing process and an ergodic one is again ergodic, we can consider a generic string from the $(S, R_n) \times (T', H)$ process. To construct the $(T_n, P_n \vee H)$ process we examine the double string and look for an α output. An α must occur because $\mu(\{S = \alpha\}) = (1/l_n)(1 - 2^{-l_n})$. Next, we examine the H-process output for k_n locations on either side of the α. This determines a Q atom and an associated distribution of $P - l_n$-strings. We select an l_n-string at random according to the distribution for Q and lay it as a third string over the output of the $(S, R_n) \times (T', H)$ process, starting over the α output and going to the right. Occasionally there will be a gap longer than l_n between the α's, and we fill gaps between l_n-strings with a special symbol 0. Thus the partition P_n will have $v + 1$ atoms. We now continue the procedure at the next α with the corresponding new atom Q.



https://doi.org/10.4153/CJM-1978-048-7 Published online by Cambridge University Press

We must now show that the processes $(T_n, P_n \vee H)$ so constructed are ergodic, close in distribution and entropy to $(T, P \vee H)$, but remain bounded away from $(T, P \vee H)$ by $c^2/100$ in $\bar{d}_H$ in contradiction to the assumption that $(T, P \vee H)$ is $H - FD$.

We remark for future use that the $(T_n, P_n \vee H)$ process contains a copy of the (S, R_n) process. By examining the locations of the special symbol 0 in P_n, which has probability 2^{-l_n} and hence must always occur in a generic string, we can discern where the α's were.

Ergodicity. To show that T_n is ergodic, we note that for any finite cylinder sets $A, B \in \bigvee_a^b T_n{}^i(P_n \vee H \vee R_n)$ we have

$$\lim_{m\to\infty} \frac{1}{m} \sum_{u=0}^{m-1} \mu(A \cap T_n{}^{-u}(B)) = \mu(A_P|A_R \cap A_H)\mu(A_R)\mu(A_H)$$
$$\cdot\, \mu(B_P|B_R \cap B_H)\mu(B_R)\mu(B_H) = \mu(A)\mu(B)$$

since P_n depends only on the R_n and H close below it and since $(T_n, H) \sim (T, H)$ is ergodic.

Distribution. To show that the distribution of strings of length n from the process $(T_n, P_n \vee H)$ is close to the distribution of strings of length n from the process $(T, P \vee H)$, we make estimates in the following cases:

1) for a $P_n - n$-string A with no zeros in the name, and
2) for a $P_n - n$-string containing zeros.

The basic idea is that the distribution of n-strings is good inside l_n-strings; and since l_n is large compared with n, the end effects are negligible.

Entropy. While the description of the construction in terms of strings is useful for checking distribution, the entropy of the $(T_n, P_n \vee H)$ processes is best examined by viewing them as towers over ergodic transformations on the base. We can then apply a theorem due to Abramov which states that

$$E(T_n) = E(\tilde{T}_n) \Big/ \int r_M d\mu_M = E(\tilde{T}_n)\mu(M),$$

where $\tilde{T}_n$ is the transformation induced on the base M of the tower and r_M is the return time to M under T_n.

The construction of $(T_n, P_n \vee H)$ is effected as follows: First we look at $(S, R_n) \times (T', H)$ as before. Then look at the transformation τ which $(S, R_n) \times (T', H)$ induces on $N = \{S = \alpha\} \times X_H$, where X_H is the space where T' acts. Now the set N can be partitioned into atoms Q_j according to the $H - (2k_n + 1)$-name centered at the time where $s = \alpha$. We then construct a skew product $\tilde{T}_n$ on the set $M \times [0, 1]$ in such a way that $\tilde{T}_n$ induced on the set $Q_j \times [0, 1]$ is the transformation $(\tau)_{Q_j} \times B_j$, where $(\tau)_{Q_j}$ is the transformation induced by τ on Q_j and B_j is a Bernoulli shift. Here the B_j are chosen to be independent of each other and of τ, with distribution given by $\bigvee_1^{l_n} T^i P | Q_j \cap C$.

https://doi.org/10.4153/CJM-1978-048-7 Published online by Cambridge University Press

The transformation $\tilde{T}_n$ is then extended over the base M to a transformation T_n on the tower by placing the appropriate $P_n \vee H$ strings above it. The length of these strings is, of course, governed by the return time to $\{S = \alpha\} \times X_H$ under $S \times T'$.

We may now examine the entropy of $\tilde{T}_n$. Observe that $S \times T'$ has entropy equal to $E(H, T)$, so that

$$E(\tau) = E(H, T)/\mu(\{S = \alpha\} \times X_H) = E(H, T)l_n(1 + 2^{-l_n}).$$

Also, if

$$\tilde{P}_{n,j} = \left\{\bigvee_0^{l_n} (T_n)^i P_n | Q_j \times [0, 1)], (N - Q_j) \times [0, 1]\right\}$$

and

$$\tilde{P}_n = \bigvee_j \tilde{P}_{n,j}; \quad \tilde{H}_n = \bigvee_0^{r_M} (T_n)^i H,$$

we have that

$$E(\tilde{T}_n) = E(\tilde{P}_n \vee \tilde{H}_n, \tilde{T}_n) = E(\tilde{H}_n, \tilde{T}_n) + E\left(\tilde{P}_n \,\middle|\, \bigvee_{-\infty}^{-1} (\tilde{T}_n)^i \tilde{P}_n \bigvee_{-\infty}^{+\infty} (\tilde{T}_n)^i \tilde{H}_n\right)$$

$$= E(\tilde{H}_n, \tilde{T}_n) + E\left(\tilde{P}_n \,\middle|\, \bigvee_{-\infty}^{-1} (\tilde{T}_n)^i \tilde{P}_n \vee G\right)$$

where $G \subseteq \bigvee_{-\infty}^{+\infty} (\tilde{T}_n)^i \tilde{H}_n$ is the partition according to $(2k_n + 1)$ -H-names centered at the time when $S = \alpha$

$$= E(\tilde{H}_n, \tilde{T}_n) + \sum_{Q_j} E_{Q_j}\left(\tilde{P}_n \,\middle|\, \bigvee_{-\infty}^{-1} (\tilde{T}_n)^i \tilde{P}_n\right) \cdot \mu(Q_j)$$

$$= E(\tau) + \sum_{Q_j} E_{Q_j}\left(\tilde{P}_{n,j} \,\middle|\, \bigvee_{-\infty}^{-1} (\tilde{T}_n)^i \tilde{P}_{n,j}\right) \cdot \mu(Q_j)$$

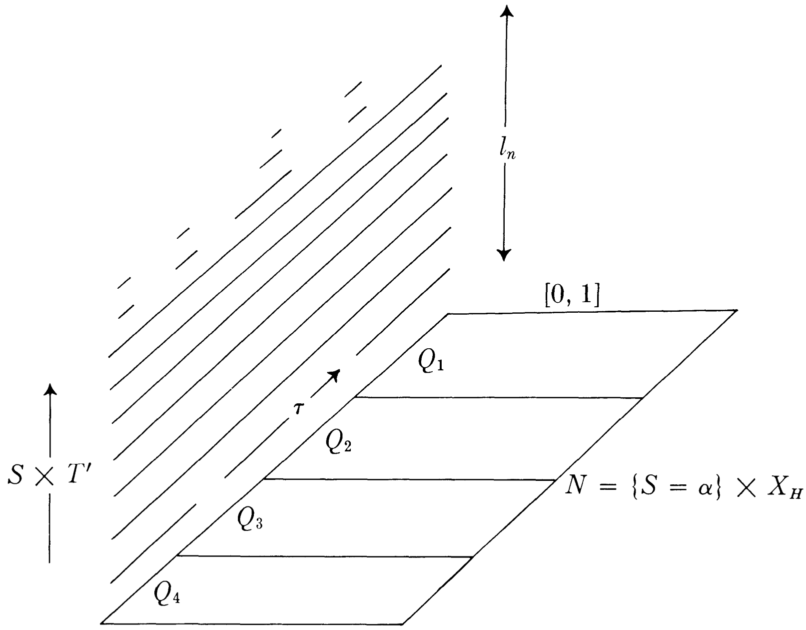
$$= E(\tau) + \sum_{Q_j} E(B_j)\mu(Q_j), \quad \text{where } Q_j \in \bigvee_{-k_n}^{k_n} T^i H.$$

But by construction we have that

i) $E(B_j) = E_{Q_j \cap C}(\bigvee_1^{l_n} T^i P)$, which is near $l_n E(P; T, H)$ for most Q_j, and
ii) $E(B_j) \leqq l_n \log v$ for a set of Q_j of measure ϵ_n on which we do not have better control of entropy.

Thus for large n, $E(\tilde{T}_n)$ will be near $l_n E(H, T) + l_n E(P; H, T) = l_n E(P \vee H, T)$.

But the return time to M under $\tilde{T}_n$ is constructed to be the same as the return time to N under $S \times T'$, hence is $l_n(1 + 2^{-l_n})$. Thus for large n we have $E(P_n \vee H, T)$ will be near $E(P \vee H, T)$.



5. The $\bar{d}_H$ matching. In order to deal with the question of the $\bar{d}_H$ distance between $(P \vee H, T)$ and $(P_n \vee H, T)$ we need the following lemma on the disjointness of $H - K$ and zero entropy processes.

LEMMA. *Let P, H, R, Q be finite partitions for a transformation T with*

$$\bigvee_{-\infty}^{+\infty} T^iQ \subseteq \bigvee_{-\infty}^{+\infty} T^iR; \quad E(Q; T, H) = E(R; T, H); \quad (P \vee H, T) \text{ is } H - K;$$

and

$$\bigvee_{-\infty}^{+\infty} T^iP \perp \bigvee_{-\infty}^{+\infty} T^iQ \quad \textit{given} \bigvee_{-\infty}^{+\infty} T^iH,$$

then

$$\bigvee_{-\infty}^{+\infty} T^iP \perp \bigvee_{-\infty}^{+\infty} T^iR \quad \textit{given} \bigvee_{-\infty}^{+\infty} T^iH.$$

In particular, we take Q to be the trivial partition and $E(R, T) = 0$ to get

$$\bigvee_{-\infty}^{+\infty} T^iP \perp \bigvee_{-\infty}^{+\infty} T^iR \quad \textit{given} \bigvee_{-\infty}^{+\infty} T^iH.$$

Proof. We shall show first that $P \perp R$ given $\bigvee_{-\infty}^{+\infty} T^iH$. Then replacing P by $\bigvee_{-n}^{n} T^iP$ and R by $\bigvee_{-n}^{n} T^iR$, the hypotheses remain in force so $\bigvee_{-n}^{n} T^iP \perp \bigvee_{-n}^{n} T^iR$ given $\bigvee_{-\infty}^{+\infty} T^iH$ for all n, from which we get our result.

Suppose then that P and R are not independent given $\bigvee_{-\infty}^{+\infty} T^iH$. Hence

$$E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} T^iH\right) - E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} T^iH \vee R\right) \geqq c > 0.$$

Since $(P \vee H, T)$ is $H - K$, we can find an n such that

$$E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} T^i H\right) - E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} T^i H \vee T^{-n}P \vee T^{-2n}P \vee \ldots\right) \leqq c/2.$$

Now set $\bar{T} = T^n$, $\bigvee_0^{n-1} T^i R = \bar{R}$, $\bigvee_0^{n-1} T^i Q = \bar{Q}$, $\bigvee_0^{n-1} T^i H = \bar{H}$ and the hypotheses remain in force with $\bar{T}$, $\bar{R}$, $\bar{Q}$, $\bar{H}$ replacing T, R, Q, H. Also,

$$E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right) - E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H} \vee \bar{R}\right)$$
$$\geqq E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right) - E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H} \vee R\right) \geqq c > 0.$$

Further,

$$E\left(P \vee \bar{Q} \,\middle|\, \bigvee_{-\infty}^{-1} \bar{T}^i (P \vee \bar{Q}) \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right)$$
$$= E\left(P \,\middle|\, \bigvee_{-\infty}^{-1} \bar{T}^i P \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right) + E\left(\bar{Q} \,\middle|\, \bigvee_{-\infty}^{-1} \bar{T}^i \bar{Q} \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right)$$
$$\geqq E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right) - c/2 + E(\bar{Q}; \bar{T}, \bar{H}).$$

On the other hand,

$$E\left(P \vee \bar{Q} \,\middle|\, \bigvee_{-\infty}^{-1} \bar{T}^i (P \vee \bar{Q}) \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right) = E(P \vee \bar{Q}; \bar{T}, \bar{H})$$
$$\leqq E(P \vee \bar{R}; \bar{T}, \bar{H}) = E\left(P \vee \bar{R} \,\middle|\, \bigvee_{-\infty}^{-1} \bar{T}^i (P \vee \bar{R}) \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right)$$
$$= E\left(P \,\middle|\, \bar{R} \bigvee_{-\infty}^{-1} \bar{T}^i (P \vee \bar{R}) \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right) + E\left(\bar{R} \,\middle|\, \bigvee_{-\infty}^{-1} \bar{T}^i (P \vee \bar{R}) \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right)$$
$$\leqq E\left(P \,\middle|\, \bar{R} \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right) + E\left(\bar{R} \,\middle|\, \bigvee_{-\infty}^{-1} \bar{T}^i \bar{R} \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right)$$
$$\leqq E\left(P \,\middle|\, \bigvee_{-\infty}^{+\infty} \bar{T}^i \bar{H}\right) - c + E(\bar{R}; \bar{T}, \bar{H}).$$

But $E(\bar{R};\, \bar{T}, \bar{H}) = nE(R; T, H) = nE(Q; T, H) = E(\bar{Q};\, \bar{T}, \bar{H})$ so the result is impossible, since $c > 0$.

$\bar{d}_H$ *matching*. Since by hypothesis $(P \vee H, T)$ is $H - FD$ and $(P_n \vee H, T) \to (P \vee H, T)$ in distribution and entropy, for all sufficiently large n we should have $\bar{d}_H[(P \vee H, T), (P_n \vee H, T)] < c^2/100$. We now show that the processes cannot be matched that well.

If $\bar{d}_H[(P \vee H, T), (P_n \vee H, T)] < c^2/100$ we can construct an ergodic process U with associated measure ν acting on a space Z and partitions $\bar{P}$, $\bar{\bar{P}}$, $\bar{H}$ of Z with $\sum_a \nu((\bar{P}_a - \bar{\bar{P}}_a) \cup (\bar{\bar{P}}_a - \bar{P}_a)) < c^2/100$, where the sum is over each set in the partitions, and where $(U, \bar{P} \vee \bar{H}) \sim (T, P \vee H)$ and $(U, \bar{\bar{P}} \vee \bar{H}) \sim (T_n, P_n \vee H)$.

Given a typical cylinder set $Q_j \in \bigvee_{-j}^{j} T^i H$, we can match the distribution $(\bigvee_1^{l_n} (T_n)^i P_n | Q_j \cap \{S = \alpha\})$ and the distribution $(\bigvee_1^{l_n} T^i P | Q_j \cap \{S = \alpha\})$ well in the $\bar{d}$ sense. To match them we merely scan a generic string of the U process while looking for the Q_j string with α in the correct (middle) position relative to it. Each time we encounter the Q_j, α combination we match the $P_n - l_n$-string and the $P - l_n$-string which lie above and to the right of α. The limiting frequencies of matched pairs of l_n-strings give a probability measure and we will have $\bar{d}[\{(T_n)^i P_n | Q_j \cap \{S = \alpha\}\}_1^{l_n}, \{T^i P | Q\}_1^{l_n}] < c/10$ for all except a set of Q_j of measure less than $c/10$.

Now by the martingale convergence theorem and the independence property of the preceding lemma applied to the joint process U, there is a set F with $\mu(F) > 1 - c/10$ and a J so that whenever $j > J$, $Q \in \bigvee_{-j}^{j} T^i H$, and $Q \cap \{S = \alpha\} \cap F \neq \emptyset$ we have that

$$\bar{d}[\{T^i P | Q \cap \{S = \alpha\}\}_1^{l_n}, \{T^i P | Q\}_1^{l_n}] < c/10.$$

Note that it is possible to have $J > k_n$ but, by part (3) of the proof of the proposition, if $\tilde{Q} \in \bigvee_{-k_n}^{k_n} T^i H$ is one of the "bad" H-strings which we used to lay in $P_n - l_n$-strings, we have

$$\bar{d}[\{T^i P | \tilde{Q}\}_1^{l_n}, \{T^i P | Q\}_1^{l_n}] < c/10,$$

where $Q \subseteq \tilde{Q}$ and $Q \in \bigvee_{-J}^{J} T^i H$. Thus we have that:

$$\text{distribution} \bigvee_1^{l_n} T^i P | \tilde{Q} \cap C$$

$$= \text{distribution} \bigvee_1^{l_n} (T_n)^i P_n | \tilde{Q} \cap \{S = \alpha\}$$

$$= \text{distribution} \bigvee_1^{l_n} (T_n)^i P_n | Q_j \cap \{S = \alpha\}$$

$\updownarrow$ $c/10$ close in $\bar{d}$ by limiting frequency in process U

$$\text{distribution} \bigvee_1^{l_n} T^i P | Q_j \cap \{S = \alpha\}$$

$\updownarrow$ $c/10$ close in $\bar{d}$ by independence of $H - K$ and zero entropy processes

$$\text{distribution} \bigvee_1^{l_n} T^i P | Q_j$$

$\updownarrow$ $c/10$ close in $\bar{d}$ by part (3) in proof of the proposition

$$\text{distribution} \bigvee_1^{l_n} T^i P | \tilde{Q}.$$

This is clearly impossible since the $\bar{d}$ distance must be at least c. Hence we have a contradiction and the theorem is proved.

Appendix. In the preceding argument we used the property that a process is H-conditionally a K-automorphism on two occasions: once to find an upper bound for the entropy of the distribution of $P - l$-strings and again to show that the $\bar{d}_H$-match cannot be good. The first use is avoidable by a counting argument similar to that of Weiss; the other seems inherent to the method of proof. Since a discussion of the $H - K$ property is not available in print, we include a short summary of its various characterizations.

Proofs are left to the reader (see [**4**] and modify results by inserting the invariant σ-algebra $\bigvee_{-\infty}^{+\infty} T^iH$).

Definition. Given an ergodic transformation T and a finite partition H, we say that T is a *K-automorphism relative to or conditioned on* $\bigvee_{-\infty}^{\infty} T^iH \quad (H - K)$ if

$$\bigcap_{n=1}^{\infty} \left(\bigvee_{-\infty}^{-n} T^iP \bigvee_{-\infty}^{+\infty} T^iH \right) = \bigvee_{-\infty}^{+\infty} T^iH$$

for every finite partition P.

LEMMA 1. *Let P, Q, H be finite partitions. Then*

$$\lim_{j\to\infty} E\left(P \,\middle|\, \bigvee_{-\infty}^{-1} T^iP \bigvee_{-\infty}^{-j} T^iQ \bigvee_{-\infty}^{+\infty} T^iH \right) = E\left(P \,\middle|\, \bigvee_{-\infty}^{-1} T^iP \bigvee_{+\infty}^{+\infty} T^iH \right).$$

LEMMA 2. *Let P, H be finite partitions such that $P \vee H$ generates under T. Let Q be a finite partition. Then*

$$\bigcap_{n=1}^{\infty} \left(\bigvee_{-\infty}^{-n} T^iQ \bigvee_{-\infty}^{+\infty} T^iH \right) \subseteq \bigcap_{n=1}^{\infty} \left(\bigvee_{-\infty}^{-n} T^iP \bigvee_{-\infty}^{+\infty} T^iH \right).$$

LEMMA 3. *Let F be a finite partition. Then*

$$E\left(F \,\middle|\, \bigvee_{-\infty}^{-1} T^iF \bigvee_{-\infty}^{+\infty} T^iH \right) = 0$$

if and only if there exists a finite partition Q such that

$$F \subseteq \bigcap_{n=1}^{\infty} \left(\bigvee_{-\infty}^{-n} T^iQ \bigvee_{-\infty}^{+\infty} T^iH \right).$$

Definition. T is *of completely positive entropy conditioned on* $\bigvee_{-\infty}^{+\infty} T^iH$ if and only if for every finite partition $F \not\subseteq \bigvee_{-\infty}^{+\infty} T^iH$ we have

$$E\left(F \,\middle|\, \bigvee_{-\infty}^{-1} T^iF \bigvee_{-\infty}^{+\infty} T^iH \right) > 0.$$

LEMMA 4. *T is $H - K$ if and only if T is of completely positive entropy conditioned on $\bigvee_{-\infty}^{+\infty} T^iH$.*

https://doi.org/10.4153/CJM-1978-048-7 Published online by Cambridge University Press

LEMMA 5. *If T is $H - K$ then T^l is $H^l - K$ with $H^l = \bigvee_0^{l-1} T^iH$, for l a nonzero integer. If T is $H - K$ then factors of T containing $\bigvee_{-\infty}^{+\infty} T^iH$ are $H - K$. If T is $H - K$ and S is a root of T for which $\bigvee_{-\infty}^{+\infty} T^iH$ is a factor, then S is $H - K$.*

Definition. We say that H *is maximal in its entropy class relative to* T if $\bigvee_{-\infty}^{+\infty} T^iH \subseteq \bigvee_{-\infty}^{+\infty} T^iF$ and $E(H, T) = E(F, T)$ imply that $\bigvee_{-\infty}^{+\infty} T^iH = \bigvee_{-\infty}^{+\infty} T^iF$, for every finite partition F.

LEMMA 6. *T is of completely positive entropy conditioned on $\bigvee_{-\infty}^{+\infty} T^iH$ if and only if H is maximal in its entropy class relative to T.*

LEMMA 7. *If P is $H - FD$, then P is $H - K$.*

Proof. By Thouvenot's Corollary 5.1 if P is $H - FD$ then H is maximal in its entropy class relative to T.

Definition. T is *K-mixing conditioned on* $\bigvee_{-\infty}^{+\infty} T^iH$ if for any set A and finite partition P we have

$$\lim_{n\to\infty}\lim_{k\to\infty}\lim_{m\to\infty}\left|\mu\left(A \,\middle|\, \bigvee_{-k}^{k} T^iH\right) - \mu\left(A \,\middle|\, \bigvee_{-n-m}^{-n} T^iP \bigvee_{-k}^{k} T^iH\right)\right| = 0 \text{ a.e. and } L^1.$$

LEMMA 8. *T is $H - K$ if and only if T is K-mixing conditioned on $\bigvee_{-\infty}^{+\infty} T^iH$.*

LEMMA 9. *Suppose that T is $H - K$, Then given ϵ and l there exists N, and for that N a number $K(N)$ and a set $F(N)$ with $\mu(F) > 1 - \epsilon$ such that for all atoms $Q \in \bigvee_{-j}^{j} T^iH$ with $j, k > K(N)$ and $Q \subseteq F$ we have that $\bigvee_{-\infty}^{-N} T^iP$ is ϵ-independent of $\bigvee_0^l T^iP$ given Q.*

REFERENCES

1. Jean-Paul Thouvenot, *Quelques propriétés des systèmes dynamiques qui se décomposent en un produit de deux systèmes dont l'un est un schéma de Bernoulli*, Israel J. Math. *21* (1975), 177–207.
2. D. S. Ornstein and B. Weiss, *Finitely determined implies very weak Bernoulli*, Israel J. Math. *17* (1974), 94–104.
3. D. S. Ornstein, *Ergodic theory, randomness, and dynamical systems* (Yale Mathematical Monographs *5*, New Haven, 1974).
4. M. Smorodinsky, *Ergodic theory, entropy*, Lecture Notes in Mathematics *214* (Springer-Verlag, New York, 1971).

University of Toronto,
Toronto, Ontario