FINITELY DETERMINED IMPLIES VERY WEAK BERNOULLI

BY

DONALD S. ORNSTEIN AND BENJAMIN WEISS

ABSTRACT

It is shown that if a process is finitely determined then it is very weak Bernoulli (VWB). Combined with known results this says that a process is isomorphic to a Bernoulli shift if and only if it satisfies an asymptotic independence condition, namely that of being VWB.

1.

Two properties of finite state stochastic processes have played an important role in the isomorphism theory of Bernoulli-processes, namely finitely determined (FD) and very weak Bernoulli (VWB), (see [4], [5]). The main results concerning these properties are: (i) any FD process is isomorphic to a Bernoulli-shift (B-shift), (ii) any factor of a B-shift is FD, (iii) any VWB process is FD and hence, from (i) isomorphic to a B-shift. In practice, while proving positive results, it has turned out that (iii) is most useful. The main result of this work establishes the fact that failing to prove VWB for a process, or more precisely, proving that a process is not VWB is tantamount to proving that a process is not isomorphic to a B-shift. We prove in fact that FD implies VWB, so that from (ii) it follows that any factor of a B-shift is VWB. For concrete systems such as the geodesic flow, this means that while initially one can only prove that smooth partitions are VWB, when enough of them have been shown to have the asymptotic independence of VWB, it then follows that all partitions are VWB.

We assume familiarity with the basic definitions of FD, VWB, $\bar{d}$-metric and

Received October 10, 1973

shall use the notation of [4]. In Section 2 we prove several lemmas which are not really new, but should be viewed as part of the folklore, while Section 3 contains the main result.

2.

The first lemma is closely related to the Shannon-McMillan theorem for n-step, Markov chains, and is undoubtedly well known to information theorists. Let $V = \{1, 2, \cdots, v\}$ be a fixed alphabet; elements of V^k will be called either k-blocks or k-strings. The number of occurrences of an n-block w, in an l-string $a = (a_1, a_2, \cdots, a_l)$, $n \leqq l$, is denoted by

$$f_a(w) = \left|\{i: (a_i, a_{i+1}, \cdots, a_{i+n-1}) = w\}\right|. \tag{1}$$

The *frequency distribution* on n-blocks, defined by an s-string a, is the measure on V^n given by

$$\mu_a(w) = f_s(w)/(l - n + 1). \tag{2}$$

Now suppose that μ_n is a measure on V^n derived from the joint distribution of a stationary, ergodic, V-valued stochastic process $\{x_n\}$. Letting $H(x/y)$ denote the conditional entropy of the partition defined by x, given the partition defined by y, we have:

LEMMA 1. *Let N_l be the number of l-strings, a, such that*

$$\left|\mu_a(w) - \mu_n(w)\right| \leqq \varepsilon, \quad \textit{all } w \in V^n. \tag{3}$$

Then for large l we have

$$\frac{\log N_l}{l} = H(x_n \,|\, x_{n-1}, x_{n-2}, \cdots, x_1) + O(\varepsilon). \tag{4}$$

PROOF. Let $(y_1, y_2, \cdots)$ be a stationary $(n-1)$-step Markov chain such that the distribution of $(y_1, y_2, \cdots, y_n)$ equals that of $(x_1, x_2, \cdots, x_n)$. If we let v_l denote the distribution of $(y_1, y_2, \cdots, y_l)$ then we compute the v_l-measure on an l-string $a = (a_1, a_2, \cdots, a_l)$ as follows:

$$\begin{aligned} v_l(a_1, a_2, \cdots, a_l) &= \mu_{n-1}(a_1, a_2, \cdots, a_{n-1})\mu_n(a_n \,|\, a_1, \cdots, a_{n-1}) \\ &\qquad \mu_n(a_{n+1} \,|\, a_2, \cdots, a_n), \cdots, \mu_n(a_l) a_{l-n+1}, \cdots, a_{l-1}) \\ &= \mu_{n-1}(a_1, a_2, \cdots, a_{n-1}) \prod_{w \in V^n} [\mu_n(w_n \,|\, w_1, \cdots, w_{n-1})]^{f_a(w)} . \end{aligned} \tag{5}$$

Now if a satisfies (3), taking logarithms we have

$$\begin{aligned}
&\log \nu_l(a_1,\cdots,a_l)\\
&= \log \mu_{n-1}(a_1,\cdots,a_{n-1}) + \sum_w f_a(w)\log\mu_n(w_n\,|\,w_{n-1},\cdots,w_1)\\
(6)\quad &= \log \mu_{n-1}(a_1,\cdots,a_{n-1}) + (l-n+1)\sum_w \mu_a(w)\log\mu_n(w_n\,|\,w_{n-1},\cdots,w_1)\\
&\geqq \log \mu_{n-1}(a_1,\cdots,a_{n-1}) + (l-n+1)\cdot\varepsilon \sum_w \log\mu_n(w_n\,|\,w_{n-1},\cdots,w_1)\\
&\quad - (l-n+1)\cdot H(x_n\,|\,x_1,\cdots,x_{n-1}).
\end{aligned}$$

Exponentiating (6) and using the obvious fact that

$$\sum_a \nu_l(a_1,\cdots,a_l) = 1 \tag{7}$$

we obtain the upper bound

$$\log N_l \leqq (l-n+1)\cdot H(x_n\,|\,x_{n-1},\cdots,x_1) + O(l\cdot\varepsilon). \tag{8}$$

The fact that there are enough l-strings that satisfy (3) follows from the mean ergodic theorem, which says that most l-strings satisfy (3). Then as in (6), we get an upper bound for the measure of these typical l-strings and thus obtain

$$\log N_l \geqq (l-n+1)\cdot H(x_n\,|\,x_{n-1},\cdots,x_1) - O(\varepsilon). \tag{9}$$

Together with (8) this implies (4). ∎

The discerning reader will notice that all this lemma has done is, by an explicit calculation, identify the typical l-strings in the Shannon-McMillan theorem. The next two lemmas are related to the ideas developed in [2], [3], and [6] and should be considered as proofs of known results that are not easy to locate.

Recall that an element $x=(\cdots\xi_{-1},\xi_0,\xi_1,\cdots)$ of $X=V^{\mathbf{Z}}$ is said to be a *generic point* for a stationary V-valued stochastic process which is given by a measure μ on X, if

$$\lim_{n\to\infty} \mu_{a_n}(w) = \mu_k(w), \quad \text{all } w\in V^k, \quad k = 1,2,\cdots \tag{10}$$

where $a_n = (\xi_{-n},\xi_{-n+1},\cdots,\xi_0,\cdots,\xi_n)$. Using the standard notation of δ_z for the point mass concentrated at z, (10) can be rewritten as follows:

$$\frac{1}{2n+1}\sum_{-n}^{n}\delta_{\sigma^i x} \xrightarrow{\text{weak}^*} \mu \tag{11}$$

where $\sigma\colon X\to X$ is the shift transformation. The lemma we need in Section 3, gives a sufficient condition for a sequence $\{n_j\}$ to have the property that sampling x along n_j will not affect the asymptotic frequency of k-blocks. But first we prove a result that is due, in essence, to Pinsker (see the reference in [2]). We prove a more general version than we actually need here since it costs very little extra effort. An equivalent result may be found in [1].

LEMMA 2. *Let P, Q, R be finite entropy partitions for T, a measure preserving transformation, that satisfy:*

(12) $$\bigvee_{-\infty}^{\infty} T^iP \supset \bigvee_{-\infty}^{\infty} T^iQ$$

(13) $$h(T,P) = h(T,Q)$$

(14) $$(T,R) \text{ is a } K\text{-automorphism}$$

(15) $$\bigvee_{-\infty}^{\infty} T^iR \text{ is independent of } \bigvee_{-\infty}^{\infty} T^iQ$$

then

(16) $$\bigvee_{-\infty}^{\infty} T^iR \text{ is independent of } \bigvee_{-\infty}^{\infty} T^iP.$$

In particular, if Q is the trivial partition and $h(T,P) = 0$, *then* (12)–(15) *hold as soon as* (14) *does and we have that K-automorphisms are always disjoint from zero entropy processes.*

PROOF. We shall show first that R is independent of P. Replacing then, R by $\bigvee_{-n}^{n} T^iR$ and P by $\bigvee_{-n}^{n} T^iP$ note that the hypotheses remain in force so we obtain, by repeating the argument, that $\bigvee_{-n}^{n} T R$ is independent of $\bigvee_{-n}^{n} T^iP$ for all n whence (16). Suppose then that R and P are not independent, that is,

(17) $$H(R) - H(R \mid P) \geqq c > 0.$$

By (14) we can find an n such that

(18) $$H(R) - H(R \mid T^nR \vee T^{2n}R \vee T^{3n}R \vee \cdots) \leqq \tfrac{1}{2}c.$$

Set now $\bar{T} = T^n$, $\bigvee_0^{n-1} T^iP = \bar{P}$, $\bigvee_0^{n-1} T^iQ = \bar{Q}$; then (12), (13), (15) and (17) remain in force with $\bar{T}$, $\bar{P}$, $\bar{Q}$ replacing T, P, and Q. We calculate now as follows:

(19) $$H(R \vee \bar{Q} \mid \bigvee_1^{\infty} \bar{T}^i(R^i \vee \bar{Q})) = H(R \mid \bigvee_1^{\infty} \bar{T}^iR) + h(\bar{T},\bar{Q}) \geqq H(R) + h(\bar{T},\bar{Q}) - \tfrac{1}{2}c$$

by (15) and (18). On the other hand, by (12), we have

(20) $$H\Big(R \vee \bar{Q} \mid \bigvee_1^{\infty} \bar{T}^i(R \wedge \bar{Q})\Big)$$

$$\leqq H\left(R \vee \bar{P} \,\middle|\, \bigvee_{1}^{\infty} \bar{T}^{i}(R \vee \bar{P})\right)$$

$$= H\left(R \,\middle|\, \bar{P} \vee \bigvee_{1}^{\infty} \bar{T}^{i}(R \vee \bar{P})\right) + H\left(\bar{P} \,\middle|\, \bigvee_{1}^{\infty} \bar{T}^{i}(R \vee \bar{P})\right)$$

$$\leqq H(R \mid \bar{P}) + H\left(\bar{P} \,\middle|\, \bigvee_{1}^{\infty} \bar{T}^{i}\bar{P}\right) \leqq H(R) - c + h(\bar{T},\bar{P}).$$

But (19) and (20) together with (13) for $\bar{T},\bar{P},\bar{Q}$ are impossible with $c > 0$. ∎

LEMMA 3. *Suppose that* $y = (\cdots\eta_{-1},\eta_0,\eta_i,\cdots) \in Y = \{0,1\}^{\mathbf{Z}}$ *is a generic point for a process with zero entropy defined by* ν *on* Y, *and* $d_n = |\{i : |i| \leqq n, \eta_i = 1\}|$ *satisfies* $\lim_{n\to\infty} d_n/n > 0$. *Then for any generic points* x *of a* V*-valued* K*-automorphism given by* μ *on* X *we have for all* k*-blocks,* $w, k = 1,2,\cdots$

(21) $$\lim_{n\to\infty} |\{i : |i| \leqq n, \eta_i = 1, (\xi_i,\xi_{i+1},\cdots,\xi_{i+k-1}) = w\}| d_n^{-1} = \mu_k(w).$$

PROOF. Consider the joint sequence

$$u = (y,x) = (\cdots(\eta_{-1},\xi_{-1}),(\eta_0,\xi_0),(\eta_1,\xi_1),\cdots) \in (\{0,1\} \times V)^{\mathbf{Z}} = U,$$

and let $\sigma : U \to U$ denote the shift on U. Let λ be any weak* limit of the measures $1/(2n+1) \sum_{i=-n}^{n} \delta_{\sigma^i u}$. Since y,x are generic for ν and μ respectively we have that $\pi_Y\lambda = \nu$, $\pi_X\lambda = \mu$ where $\pi_Y : U \to Y$, $\pi_X : U \to X$ are the projections of U on the first and second coordinates. Letting $\mathscr{B}_X, \mathscr{B}_Y$ denote the Borel σ-algebras on X, Y, since ν has zero entropy so does σ on $\pi_y^{-1}(\mathscr{B}_Y)$. Similarly σ is a K-automorphism on $\pi^{-1}(\mathscr{B}_X)$ and thus by Lemma 2, $\pi_x^{-1}(\mathscr{B}_X)$ is independent of $\pi_y^{-1}(\mathscr{B}_Y)$, and it follows that $\lambda = \nu \times \mu$ the product measure defined by ν and μ. But since this is true for any weak* limit it follows that $1/(2n+1) \sum_{i=-n}^{n} \delta_{\sigma^i u}$ converges weak* to $\nu \times \mu$. Using this fact for the indicator function of the event

$$\{y_1 = 1, (x_1,x_2,\cdots,x_k) = w\}$$

we obtain

(22) $$\lim_{n\to\infty} |\{i : |i| \leqq n, \eta_i = 1, (\xi_i,\xi_{i+1},\cdots\xi_{i+k-1}) = w\}| / 2n+1 = \nu\{y_1 = 1\}\cdot\mu_k(w).$$

But $\lim_{n\to\infty} d_n/(2n+1) = \nu\{y_1 = 1\}$, since $(\cdots\eta_{-1},\eta_0,\eta_1,\cdots)$ is a generic point for ν, and thus (21) follows from (22). ∎

3.

Let T be an ergodic, invertible, measure-preserving transformation of $(X,\mathscr{B},\mu)$ and P a partition into v-sets. Then (T,P) defines in the usual way a V-valued stationary stochastic process. This section is devoted to proving the following theorem.

THEOREM. *If* (T, P) *is* FD *then* (T, P) is VWB.

The proof will be by contradiction, that is, we shall show that the assumptions (T, P) not VWB and (T, P) FD lead to a contradiction. In doing this we may as well assume that (T, P) is a K-automorphism, since FD even implies that (T, P) is isomorphic to a B-shift. The idea of the proof is to use the fact that (T, P) is not VWB to construct a sequence of processes that approximates (T, P) in entropy and finite distributions but whose $\bar{d}$-distance from (T, P) remains greater than some fixed positive quantity. The processes that we construct will be mixing, although this is not really necessary to contradict FD. In order to construct these processes we will need the following.

LEMMA. *There exists* $c > 0$, *such that* ε_n, k_n, *and* l_n *can be chosen, and we can find at least one atom* $A \in \bigvee_0^{k_n} T^i P$ *with the properties*:

$$\left| H\left(\bigvee_1^{l_n} T^{-i}P \,\middle|\, A \right) - l_n \cdot h(T, P) \right| \leqq l_n \cdot \varepsilon_n$$

$$\left\| \text{dist}\,(n\text{-blocks in dist}\left(\bigvee_1^{l_n} T^{-i}P \,\middle|\, A \right) \; - \mu_n \right\|_1 \leqq \varepsilon_n$$

$$\bar{d}(\{T^{-i}P \mid A\}_1^{l_n}, \{T^{-i}P\}_1^{l_n}) \geqq c$$

where $\varepsilon_n \to 0$ as $n \to \infty$.

PROOF OF LEMMA. (i) The fact that (T, P) is not VWB means that for some $c > 0$ and all l, and k large compared to l if $\mathscr{A}$ denotes the set of atoms A of $\bigvee_0^k T^i P$ such that

(1) $$\bar{d}(\{T^{-i}P \mid A\}_1^l, \{T^{-i}P\}_1^l) \geqq c,$$

then

(2) $$\mu(\mathscr{A}) \geqq c.$$

(ii) Given δ_1 we can find l^1 so that if $l \geqq l^1$ and $\mathscr{A}_1$ denotes the set of atoms B of $\bigvee_1^l T^{-i}P$ for which

(3) $$|\mu_a(w) - \mu_n(w)| \leqq \delta_1, \text{ all } w \in v^n,$$

where a is the l-string defined by B, that is, $B = \bigcap_{i=1}^l T^{-i}P_{a_i}$, then

(4) $$\mu(\mathscr{A}_1) \geqq 1 - \delta_1.$$

(This follows from the ergodic theorem.)

Because of (2), if δ_1 is small enough most atoms $A \in \mathscr{A}$ are covered almost enterely by atoms of $\mathscr{A}_1$. For such atoms the distribution of n-blocks given by dist $(\bigvee_1^l T^{-i}P \mid A)$ is close to the (T,P) distribution of n-blocks.

(iii) We now want to find atoms $A \in \mathscr{A}$ for which $H(\bigvee_1^l T^{-i}P \mid A)$ is close to $l \cdot h(T,P)$. To this end, note that

$$H\left(\bigvee_1^l T^{-i}P \,\Big|\, \bigvee_0^k T^iP\right) \geqq l \cdot h(T,P). \tag{5}$$

Now since

$$H\left(\bigvee_1^l T^{-i}P \,\Big|\, \bigvee_0^k T^iP\right) = \sum_{A \in \bigvee_0^k T^iP} H\left(\bigvee_1^l T^{-i}P \,\Big|\, A\right)\mu(A), \tag{6}$$

if we could show for any δ_2 that k, δ_1 and $l > l^1$ could be chosen so that for all but δ_2 atoms $A \in \bigvee_0^k T^iP$

$$\frac{1}{l}H\left(\bigvee_1^l T^{-i}P \,\Big|\, A\right) \leqq h(T,P) + \delta_2, \tag{7}$$

then (5), (6), and (7) combined would give (8):

For any δ, we could choose k, δ_1, and $l > l^1$ so that

$$\left| H\left(\bigvee_1^l T^{-i}P \,\Big|\, A\right) - l \cdot h(T,P)\right| / l < \delta \tag{8}$$

for all but δ of the A in $\mathscr{A}$ which is what we need.

To establish (7), note that if $\mathscr{A}_2$ consists of those atoms A of $\bigvee_0^k T^iP$ such that

$$\mu(A \cap \mathscr{A}_1)/\mu(A) \geqq 1 - (\delta_1)^{\frac{1}{2}} \tag{9}$$

then (4) implies

$$\mu(\mathscr{A}_2) \geqq 1 - (\delta_1)^{\frac{1}{2}}. \tag{10}$$

Apply Lemma 1 with $\varepsilon = \delta_1$ to find $l^{(2)}$ so that for $l \geqq l^{(2)}$

$$\log N_l \leqq l \cdot H(T^nP \mid T^{n-1}P \vee \cdots \vee P) + (\delta_1)^{\frac{1}{2}}, \tag{11}$$

and suppose that n is large enough so that

$$\left| H(P \mid T^{-1}P \vee \cdots \vee T^{-n+1}P) - h(T,P)\right| \leqq (\delta_1)^{\frac{1}{2}}. \tag{12}$$

Then for $l \geqq l^1, l^{(2)}$ an elementary computation yields

(13) $$H\left(\bigvee_1^l T^{-i}P \,\middle|\, A\right) \leqq l\cdot h(T,P) + l(2+\log v)\cdot(\delta_1) - H((\delta_1)^{\frac{1}{2}},\ 1-(\delta_1)^{\frac{1}{2}})$$

for all $A \in \mathscr{A}_2$. If

$$(2+\log v)\cdot(\delta_1)^{\frac{1}{2}} + H(\delta_1)^{\frac{1}{2}},\ 1-(\delta_1)^{\frac{1}{2}} < \delta_2$$

we obtain (7).

An alternative way to derive an estimate similar to (7) would be to use the fact that (T,P) is a K-automorphism and choose l large enough for M' to be chosen such that if $M = in$, $M' < in \leqq l$, then $\bigvee_{M+1}^{M+n} T^{-i}P$ will be ε-independent of $\bigvee_0^k T^iP$ and M'/l small.

(8) follows from (5), (6), and (7) by a standard argument which will we now give Let $\mathscr{A}_3$ denote the set of atoms A of $\bigvee_o^k T^iP$ for which

(14) $$H\left(\bigvee_1^l T^{-i}P \,\middle|\, A\right) \leqq l\cdot h(T,P) - l\cdot 3(\delta_2)^{\frac{1}{2}}$$

(15) $$l\cdot h(T,P) \leqq \sum_A H\left(\bigvee_1^l T^{-i}P \,\middle|\, A\right)\mu(A)$$

$$= \sum_{A\in\mathscr{A}_3} + \sum_{A\notin\mathscr{A}_3} \leqq \sum_{A\in\mathscr{A}_3} + \sum_{\substack{A\notin\mathscr{A}_3\\ A\in\mathscr{A}}} + \sum_{A\notin\mathscr{A}_2}$$

$$\leqq \mu(\mathscr{A}_3)\cdot[l\cdot h(T,P) - 3\cdot l(\delta_2)^{\frac{1}{2}}] + (1-\mu(\mathscr{A}_3))[l\cdot h(T,P) + l\cdot\delta_2 + (\delta_1)^{\frac{1}{2}}\cdot l\ \log v$$

whence

(16) $\mu(\mathscr{A}_3) \leqq (\delta_2)^{\frac{1}{2}}$. (Note that $(\delta_1)^{\frac{1}{2}} l \log v < l\delta_2$ by our choice of δ_1).

Combining (10), (13), (14), and (16) we obtain finally that if $\mathscr{A}_4$ denotes the atoms $A \in \bigvee_0^k T^iP$ for which

(17) $$\left| H\left(\bigvee_l^l T^{-i}P \,\middle|\, A\right) - l\cdot h(T,P)\right| \leqq l\cdot 3\cdot(\delta_2)^{\frac{1}{2}}$$

then

(18) $$\mu(\mathscr{A}_4) \geqq 1 - 2(\delta_2)^{\frac{1}{2}}.$$

This and steps (1) and (2) prove our lemma.

In order to complete the proof of the theorem, it remains to construct a mixing process using the distribution given by the lemma which we shall denote henceforth by $\lambda_l = \text{dist}(\bigvee_1^l T^{-i}P \,|\, A)$.

The idea here is simply to put down l-strings chosen independently with distribution λ_l. Doing this in the obvious way will lead to a process whose lth power may not be ergodic. So we modify the construction by a random spacing of these l-blocks To this end let S be a mixing, zero entropy, measure-preserving transformation of Y. By the Rokhlin-Kakutani tower theorem there is a partition Q of the space Y into sets $(Q_1, Q_2, \cdots, Q_l, Q_{l+1})$ such that $TQ_i = Q_{i+1}, i = 1, \cdots, l-1$ and the measure of Q_{l+1} is $\leqq 2^{-l}$. Let $\bar{S}$ denote the transformation induced by S on Q_1 and let $R = \{R_l, R_{l+1}, \cdots\}$ be the partition of Q_1 defined by the return time to Q_1, that is, $x \in R_m$ means that m is the least positive integer for which $S^m x \in Q_1$. Naturally most of Q_1 is filled with R_l. Now let T_n be the direct product of a B-shift on V^l with distribution λ_l and $\bar{S}$, and build a tower over T_n as follows (see Figure 1): for each atom B in the generating partition of the B-shift

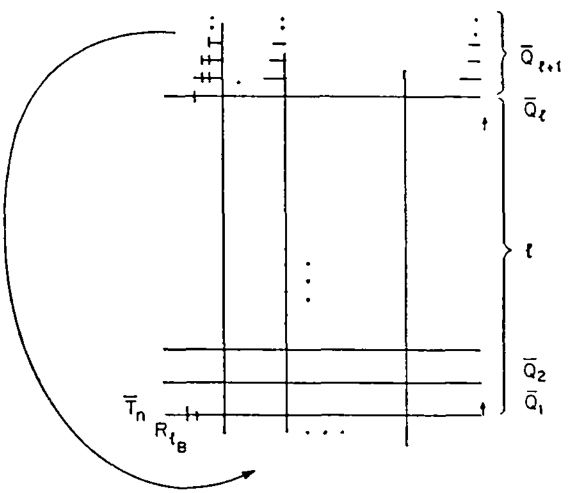

Fig. 1.

part of T_n build l-full levels, and then one level more above $R_{l+1} \cap B$, two additional levels above $R_{l+2} \cap B$, and in general k additional levels above $R_{l+k} \cap B$. Denote the tower transformation by T_n and let m be the normalized measure; observe that the partition $\bar{Q} = (\bar{Q}_1, \cdots, \bar{Q}_{l+1})$ for T_n into the l full levels and the remainder is such that $(T_n, \bar{Q})$ is isomorphic to (S, Q). It will be convenient at this point to add another symbol for P_n. This need cause no trouble, since one can always augment P by adding a set with zero measure. Define, now, P_n by assigning the jth level over the atom B to $P^n_{a_j}$ if B corresponds to $(a_1, a_2, \cdots, a_l) \in V^l$, and combining the other levels, that is, $\bar{Q}_{l+1}$, to the extra symbol 0.

Now the fact is clear that T_n is mixing. Indeed since $\bar{Q} \vee P_n$ is a generator it suffices to show that for any atoms C, D in a finite span $\bigvee_R^k T_n{}^i(Q \vee P_n)$ we have

$$\lim_{k\to\infty} m(T_n^i C \cap D) = m(C)\cdot m(D). \tag{19}$$

To check (19), note that for i sufficiently large, on most of the space the B-shift part of C and D is independent, while the $\bar{Q}$ part is isomorphic to a mixing transformation (S,Q), independent of the B-shift part.

It remains to show that (T_n, P_n) has the requisite properties to contradict FD. First of all we have

$$\left| \operatorname{dist} \bigvee_1^n T_n^i P_n - \operatorname{dist} \bigvee_1^n T^i P \right| \leqq \varepsilon_n + n + 1/l \tag{20}$$

by (20) and the fact that $m(\bar{Q}_{l+1}) \leqq 2^{-l}$. This is seen by computing the distribution of n-blocks of P_n along columns of the tower that defined T_n. Next, since (S,Q) has zero entropy, T_n induced on $\bar{Q}_1$ has the entropy of the B-shift part, namely $H(\lambda_l)$. Then by Abramov's theorem we have

$$h(T_n, P_n) = m(\bar{Q}_1)\cdot H(\lambda_l) \tag{21}$$

which, by (19), yields

$$\left| h(T_n, P_n) - h(T,P) \right| \leqq \varepsilon_n + 1/l. \tag{22}$$

Finally, if $\bar{d}\{(T_n, P_n), (T,P)\}$ were less than $c/2$, then we could find a generic point for (T,P), say $x \in V^{\mathbb{Z}}$, and a generic point for (T_n, P_n), say $y \in (V \cup \{0\})^{\mathbb{Z}}$, with $x = (\xi_i)$, $y = (\eta_i)$ and

$$\lim_{k\to\infty} \left|\{i : |i| \leqq k,\ \xi_i \neq \eta_i\}\right|/2k+1 < c/2. \tag{23}$$

Now clearly the zeros in y are separated by l-strings which have the frequency of λ_l, since y is generic for T_n. Furthermore, the starting points of these l-strings form a generic sequence for a zero entropy process obtained in an obvious way as a factor of $(T_n, \bar{Q})$. Thus by Lemma 2, the corresponding l-strings in x have the distribution of μ_l, that is, the distribution of the (T,P)-process. Now any weak limit of these joint l-strings would produce a measure on pairs of l-strings whose first coordinate would have the distribution λ_l and the second μ_l. Furthermore the average discrepancy would be less than or equal to $c/2 + 1/l$ which for large l contradicts our lemma. This contradiction completes the proof of the theorem. ■

References

1. K. Berg, *Convolution of invariant measures, maximal entropy*, Math. Systems Theory **3** (1969), 146–151.
2. H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Systems Theory **1** (1967), 1–49.
3. T. Kamae, *Subsequences of normal sequences*, Israel J. Math. **16** (1973), 121–149.
4. D. Ornstein, *Ergodic theory, randomness, and dynamical systems.*
5. D. Ornstein and B. Weiss, *Geodesic flows are Bernoullian*, Israel J. Math. **14** (1973), 189–198.
6. B. Weiss, *Normal sequences as collectives*, Proc. Symp. on Topological Dynamics & Ergodic Theory, Univ. of Kentucky, 1971.

Department of Mathematics
Stanford University
Palo Alto, California, U. S. A.

and

Department of Mathematics
The Hebrew University of Jerusalem
Jerusalem, Israel