PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 75, Number 2, July 1979

FINITARY CODINGS AND WEAK BERNOULLI PARTITIONS[1]

A. DEL JUNCO AND M. RAHE

ABSTRACT. An example is constructed of a two state stationary stochastic process (T, P) whose time zero partition P is weak Bernoulli under the shift, but which cannot be the image under a finitary coding of any independent process. A sufficient condition for a partition to be weak Bernoulli is developed, based on the rate of convergence of the conditional entropy $h(P|P_{-j}^{-1})$ to $h(P, T)$.

1. Introduction. Let Z denote the set of integers. For $i = 1, 2$, where A_i is a finite or countably infinite set, $\mathfrak{B}_i$ is the extension to sequence space of the σ-algebra on A_i generated by the partition into individual points, and μ_i is a normalized shift invariant measure on $(A_i^Z, \mathfrak{B}_i)$, let T_i be the shift on the sequence space $(A_i^Z, \mathfrak{B}_i, \mu_i)$. There is a natural correspondence between the sets in the partition at time zero of the sequence space and the elements of A_i; and by a small abuse of notation we shall also use A_i to indicate this partition. By a coding $\phi: A_1^Z \to A_2^Z$ we mean a measurable map from a subset of measure one of A_1^Z to A_2^Z which takes μ_1 to μ_2 and which commutes with T_1 and T_2. We say that ϕ is finitary if for μ_1-a.e. $(x_i) \in A_1^Z$ there exists a pair (s, t) of integers such that whenever $(x_s, \ldots, x_t) = (\hat{x}_s, \ldots, \hat{x}_t)$ and whenever ϕ is defined, then $(\phi(x))_0 = (\phi(\hat{x}))_0$.

For $\mathcal{P}$ and $\mathcal{Q}$ partitions of such a sequence space, we write $\mathcal{P} \perp_\varepsilon \mathcal{Q}$ if $\sum_{P \in \mathcal{P}, Q \in \mathcal{Q}} |\mu(P \cap Q) - \mu(P)\mu(Q)| < \varepsilon$. A partition $\mathcal{P}$ is said to be weak Bernoulli under a shift T if for each $\varepsilon > 0$ there is an $L = L(\varepsilon)$ such that $\bigvee_{j=-n}^{0} T^j\mathcal{P} \perp_\varepsilon \bigvee_{j=L}^{L+n} T^j\mathcal{P}$ for all $n \geqslant 0$. We say that the process $(A^Z, \mathfrak{B}, \mu)$ is weak Bernoulli if the generator A is.

It is not hard to see (for completeness we sketch a proof at the end of this article) that whenever $\phi: A_1^Z \to A_2^Z$ has finite expected time to code and whenever the partition A_1 is weak Bernoulli under T_1, then $A_2 = \phi(A_1)$ is also weak Bernoulli under T_2. In particular, if the sequence $\{T_1^i(A_1) \text{ for } i \in Z\}$ is independent, then A_2 must be weak Bernoulli under T_2. On the other hand, as this article shows, not every process $(A_2^Z, \mathfrak{B}_2, \mu_2)$ for which A_2 is weak Bernoulli under the shift can be a finitary image of an independent process, even if we allow an infinite time to code. This negative statement contrasts with recent positive results in [1] which show that every weak Bernoulli

Received by the editors January 10, 1978.
AMS (MOS) subject classifications (1970). Primary 28A65.
[1]Research supported in part by NRC Grant.

process in which the generator is Markovian can be finitarily coded from any independent process of strictly higher entropy.

We point out that it is already known from unpublished work of D. Ornstein [3] that there are examples of very weak Bernoulli processes which are not finitarily codable from any independent process. His argument runs as follows: If 0 is a state in such a finitary image, then for all i sufficiently large, in each generic string the set of times at which 0 occurs contains an arithmetic progression with separation i, where the number of terms in the progression is arbitrarily large. Ornstein's idea was to construct a partition in a process with very weak Bernoulli generator by means of Rohlin towers such that for all i, arbitrarily long arithmetic progressions of times of occurrence of 0 with separation i were impossible. The resulting partition is very weak Bernoulli since every partition has that property if the generator of the process has it. However, since it is not true that every partition in a process with weak Bernoulli generator is again weak Bernoulli [4], we must proceed by a different method.

Our example consists of a process (T, P) on $\{0, 1\}^Z$ where the time zero partition $P = \{P_0, P_1\}$ is weak Bernoulli under the shift T, yet where for each i there is a value k_i such that whenever the state zero occurs at times which form an arithmetic progression with separation i, then the progression contains at most k_i terms. By the observation of Ornstein, such a process cannot be a finitary image of an independent process. In constructing this example we shall utilize a criterion involving the rate at which the conditional entropy $h(P|P_{-j}^{-1}) = h(P| \bigvee_{-j}^{-1} T^iP)$ converges to $h(P, T)$. This information will be sufficient to imply that P is weak Bernoulli under T, provided (T, P) is mixing.

2. Construction. The construction of (T, P) proceeds in stages. To begin, we construct an infinite sequence of transformations $\bar{\bar{T}}_i$, each with a two set generating partition $\bar{\bar{P}}_i = \{\bar{\bar{P}}_i^0, \bar{\bar{P}}_i^1\}$. Next we construct the infinite direct product $\bar{T} = \prod_{i=1}^{\infty} \bar{\bar{T}}_i$ and a generating partition $\bar{P} = \prod_{i=1}^{\infty} \bar{\bar{P}}_i$, which will be countable by construction. The process $(\bar{T}, \bar{P})$ will be mixing and will have $\sum_{j=1}^{\infty} [h(\bar{P}|\bar{P}_{-j}^{-1}) - h(\bar{P}, \bar{T})] < 1$. These two conditions will then guarantee that $\bar{P}$ is weak Bernoulli under $\bar{T}$.

The final step of the construction gives the process (T, P) by a finite coding of $(\bar{T}, \bar{P})$: if all the components of $\bar{P}$ are zero, P will be zero; otherwise P will be one. The partition will be nontrivial; and as the image under a finite coding of a weak Bernoulli process, (T, P) will again be weak Bernoulli. Moreover, (T, P) will inherit any limitations on the length of arithmetic progressions that any of the $(\bar{\bar{T}}_i, \bar{\bar{P}}_i)$ may possess.

We now proceed with the details.

PROPOSITION. *We can construct an infinite sequence of transformations* $\bar{\bar{T}}_i$ *for* $i = 1, 2, \ldots,$ *each with a two set partition* $\bar{\bar{P}}_i = \{\bar{\bar{P}}_i^0, \bar{\bar{P}}_i^1\}$, *such that the* *processes* $(\bar{\bar{T}}_i, \bar{\bar{P}}_i)$ *have the following properties*:

(1) $\mu(\bar{\bar{P}}_i^1)$ *is sufficiently small that* $h(\bar{\bar{P}}_i) < 2^{-i}$,

(2) $\sum_{j=1}^{\infty} [h(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)_{-j}^{-1}) - h(\bar{\bar{P}}_i, \bar{\bar{T}}_i)] < 2^{-i}$,

(3) *there is a value* k_i *such that no arithmetic progression with separation* i *of times at which* 0 *occurs will contain more than* k_i *terms, and*

(4) *the process* $(\bar{\bar{T}}_i, \bar{\bar{P}}_i)$ *is mixing*.

PROOF. Since $\lim_{x \to 0+} x \log x = 0$, we can find values α, $1 - \alpha$ such that $h(\alpha, 1 - \alpha) = -\alpha \log \alpha - (1 - \alpha)\log(1 - \alpha) < 2^{-i}$. We then consider an independent process (S, R) on $\{0, 1\}^Z$ with distribution $(\alpha, 1 - \alpha)$. The distribution of $(\bar{\bar{P}}_i^0, \bar{\bar{P}}_i^1)$ will be chosen close to $(\alpha, 1 - \alpha)$ and this choice will satisfy the first requirement.

We begin with the collection of all strings of length $m \geqslant 3i$ from the independent process (S, R) just selected, and we define a new process which will travel only through strings with i consecutive 1's. In so doing we must avoid not only those strings which fail to have i consecutive 1's, but also those that cannot be forced to lead to a string with i consecutive 1's. The following paragraphs describe the selection of such strings and transitions between them.

Let N_m be the class of m-strings having a run of at least $2i - 1$ consecutive outputs of 1, where the run begins in other than the leftmost position. Let F_m be the class of m-strings which are not in N_m and which begin on the left with k consecutive 1's and end on the right with j consecutive 1's, where $j + k \geqslant 2i - 1$. We may characterize the set $G_m = N_m \cup F_m$ as the collection of strings that have a "cyclic run" of length at least $2i - 1$ of 1's. Note that all strings in G_m have a noncyclic run of at least i outputs of 1.

For $t \in G_m$ denote by t^j, $j = 0, 1$, the string formed by deleting the leftmost digit of t and appending the digit j on the extreme right. We may now define a stochastic matrix M_m as follows: $M_m(t, s) = \alpha$ if $t \in N_m$ and $s = t^0$; $M_m(t, s) = 1 - \alpha$ if $t \in N_m$ and $s = t^1$; $M_m(t, s) = 1$ if $t \in F_m$ and $s = t^1$; and $M_m(s, t) = 0$ for other $t, s \in G_m$. It is easy to see that there exists a unique stationary probability measure ν_m on strings in G_m for which M_m is the matrix of transition probabilites of a mixing 1-step Markov process with state space G_m. For each m by using this 1-step Markov process on G_m we may define a stationary m-step mixing Markov process $(\bar{\bar{T}}_{m,i}, \bar{\bar{P}}_{m,i})$ on $\{0, 1\}^Z$ by setting $\mu\{\bar{\bar{T}}_{m,i}^m(\omega) \in \bar{\bar{P}}^j_{m,i}|\omega \in A\} = M_m(t, t^j)$ and $\mu(A) = \nu_m(t)$, where A is the finite cylinder set in $\bigvee_0^{m-1}\bar{\bar{T}}_{m,i}^i\bar{\bar{P}}_{m,i}$ indexed by the m-string t.

Denote by ν the measure on m-strings given by the $(\alpha, 1 - \alpha)$ distribution in (S, R). If we denote by ν' the probability measure $\nu(G_m)^{-1}\nu$ restricted to G_m, then we claim that $\nu_m = \nu'$. This is virtually obvious since we must verify that $\nu' M_m = \nu'$ for only two cases. Suppose that for $s \in G_m$ there exists only one $t \in G_m$ with $M_m(t, s) > 0$. Observe that $t \notin P_m$ so $M_m(t, s) = 1$ and we have that $\nu'(t)M_m(t, s) = \nu'(t)$. On the other hand we observe that s and t both have the same number of 1's so $\nu'(t) = \nu'(s)$. Similarly, if for $s \in G_m$

there exist distinct $t, \hat{t} \in G_m$ with $M_m(t, s) \geq 0$ and $M_m(\hat{t}, s) \geq 0$, we may verify that

$$\nu'(t)M_m(t, s) + \nu'(\hat{t})M_m(\hat{t}, s) = \nu'(s).$$

Thus for each m we have an explicit expression for the distribution of $\bigvee_0^{m-1} \bar{\bar{T}}_{m,i}^i(\bar{\bar{P}}_{m,i})$. Moreover, an easy probabilistic argument shows that $1 - \nu(G_m)$ goes to zero exponentially fast, so that $m[1 - \nu(G_m)]$ also goes to zero as m becomes large. Thus we calculate as follows:

$$\begin{aligned}
&\sum_{j=1}^{\infty} \left[h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)_{-j}^{-1}\right) - h\left(\bar{\bar{P}}_i, \bar{\bar{T}}_i\right)\right] \\
&\quad = \sum_{j=1}^{m} \left[h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)_{-j}^{-1}\right) - h(\bar{\bar{P}}_i, \bar{\bar{T}}_i)\right] \\
&\quad \leq \sum_{j=1}^{m} \left[h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)^{-1}\right) - h(\bar{\bar{P}}_i, \bar{\bar{T}}_i)\right] \\
&\quad = m\left[h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)^{-1}\right) - h(\bar{\bar{P}}_i, \bar{\bar{T}}_i)\right] \\
&\quad = m\left[h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)^{-1}\right) - h\left(R|(R)^{-1}\right)\right] + m\left[h(R, S) - h(\bar{\bar{P}}_i, \bar{\bar{T}}_i)\right] \\
&\quad \leq m\left|h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)^{-1}\right) - h\left(R|(R)^{-1}\right)\right| + m\left|h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)_{-m}^{-1}\right) - h\left(R|(R)_{-m}^{-1}\right)\right| \\
&\quad \leq 4h(\alpha, 1 - \alpha)m\left[1 - \nu(G_m)\right]
\end{aligned}$$

and this product goes to zero as m becomes large. We now choose an appropriate value for m and set $(\bar{\bar{T}}_i, \bar{\bar{P}}_i) = (\bar{\bar{T}}_{m,i}, \bar{\bar{P}}_{m,i})$. Thus conditions (2) and (4) are satisfied and condition (3) follows with $k_i = m/i$ from the fact that each string in G_m has a run of at least i consecutive 1's.

Next, given the sequence $(\bar{\bar{T}}_i, \bar{\bar{P}}_i)$ we construct in the standard manner the infinite direct product process $(\bar{T}, \bar{P})$ with $\bar{T} = \prod_{i=1}^{\infty} \bar{\bar{T}}_i$ and $\bar{P} = \prod_{i=1}^{\infty} \bar{\bar{P}}_i$. We observe that since $\sum_{i=1}^{\infty} h(\bar{\bar{P}}_i) < 1$ we also have that $\sum_{i=1}^{\infty} \mu(\bar{\bar{P}}_i^1) < 1$, so that by the Borel-Cantelli lemma the time zero output of the ith component process is zero for all except a finite number of i's, except perhaps on a set of measure zero. Thus $\bar{P}$ is actually a countable partition. Moreover, sets in the product σ-algebra are arbitrarily well approximated by finite cylinder sets in $\prod_{i=1}^{l} (\bar{\bar{P}}_i)_{-l}^{l}$ for sufficiently large l so the process $(\bar{T}, \bar{P})$ is mixing. Further

$$h(\bar{P}) = \lim_{l\to\infty} h\left(\prod_1^l \bar{\bar{P}}_i\right) = \lim_{l\to\infty} \sum_{i=1}^{l} h(\bar{\bar{P}}_i) < 1,$$

so that

$$h\left(\bar{P}|\bar{P}_{-j}^{-1}\right) = \sum_{i=1}^{\infty} h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)_{-j}^{-1}\right) \quad \text{for each } j$$

and

$$h(\bar{P}, \bar{T}) = \sum_{i=1}^{\infty} h(\bar{\bar{P}}_i, \bar{\bar{T}}_i).$$

Thus

$$\sum_{j=1}^{\infty}\left[h\left(\bar{P}|\bar{P}_{-j}^{-1}\right)-h(\bar{P},\bar{T})\right]$$
$$=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\left[h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)_{-j}^{-1}\right)-h(\bar{\bar{P}}_i,\bar{\bar{T}}_i)\right]$$
$$=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\left[h\left(\bar{\bar{P}}_i|(\bar{\bar{P}}_i)_{-j}^{-1}\right)-h(\bar{\bar{P}}_i,\bar{\bar{T}}_i)\right]<\sum_{i=1}^{\infty}2^{-i}<1.$$

3. Sufficient condition for weak Bernoulli behavior. We now prove that $\bar{P}$ is weak Bernoulli under $\bar{T}$ in a manner which is similar to the way that one shows an n-step mixing Markov process is weak Bernoulli.

THEOREM. *If $(\bar{T},\bar{P})$ is mixing and if $\sum_{j=1}^{\infty}[h(\bar{P}|\bar{P}_{-j}^{-1})-h(\bar{P},\bar{T})]$ is finite, then $(\bar{T},\bar{P})$ is weak Bernoulli.*

PROOF. Given $\varepsilon>0$ there exists $\delta>0$ such that for any partitions $\mathcal{P}$ and $\mathcal{Q}$, $h(\mathcal{P})-h(\mathcal{P}|\mathcal{Q})<\delta$ implies $\mathcal{P}\perp_\varepsilon\mathcal{Q}$. By the finite sum hypothesis, given $\delta/3$ we can choose N such that for all positive integers M, K we have that

$$0\leqslant h\left(\bar{P}_{N+1}^{N+M}|\bar{P}_0^N\right)-h\left(\bar{P}_{N+1}^{N+M}|\bar{P}_{-K}^N\right)$$
$$\leqslant h\left(\bar{P}_{N+1}^{N+M}|\bar{P}_0^N\right)-h\left(\bar{P}_{N+1}^{N+M}|\bar{P}_{-\infty}^N\right)$$
$$=\sum_{j=N+1}^{N+M}\left[h\left(\bar{P}^j|\bar{P}_0^{j-1}\right)-h\left(\bar{P}^j|\bar{P}_{-\infty}^{j-1}\right)\right]$$
$$=\sum_{j=N+1}^{N+M}\left[h\left(\bar{P}|\bar{P}_{-j}^{-1}\right)-h\left(\bar{P}|\bar{P}_{-\infty}^{-1}\right)\right]<\delta/3.$$

By the mixing hypothesis, having chosen N we may now choose L such that a second inequality will hold; namely, $0\leqslant h(\bar{P}_L^{L+N})-h(\bar{P}_L^{L+N}|\bar{P}_0^N)<\delta/3$. The first inequality now shows that

$$h\left(\bar{P}_{L+N+1}^{L+M}|\bar{P}_L^{L+N}\right)-h\left(\bar{P}_{L+N+1}^{L+M}|\bar{P}_L^{L+N}\vee\bar{P}_0^N\vee\bar{P}_{-K}^{-1}\right)<\delta/3$$

and also that

$$h\left(\bar{P}_L^{L+N}|\bar{P}_0^N\right)-h\left(\bar{P}_L^{L+N}|\bar{P}_0^N\vee\bar{P}_{-K}^{-1}\right)<\delta/3.$$

Thus

$$0\leqslant h\left(\bar{P}_{L+N+1}^{L+M}\vee\bar{P}_L^{L+N}\right)-h\left(\bar{P}_{L+N+1}^{L+M}\vee\bar{P}_L^{L+N}|\bar{P}_0^N\vee\bar{P}_{-K}^{-1}\right)$$
$$=h\left(\bar{P}_{L+N+1}^{L+M}\vee\bar{P}_L^{L+N}\right)-h\left(\bar{P}_{L+N+1}^{L+M}|\bar{P}_L^{L+N}\vee\bar{P}_0^N\vee\bar{P}_{-K}^{-1}\right)$$
$$-h\left(\bar{P}_L^{L+N}|\bar{P}_0^N\vee\bar{P}_{-K}^{-1}\right)$$
$$\leqslant h\left(\bar{P}_{L+N+1}^{L+M}\vee\bar{P}_L^{L+N}\right)-h\left(\bar{P}_{L+N+1}^{L+M}|\bar{P}_L^{L+N}\right)-h\left(\bar{P}_L^{L+N}\right)+\delta=\delta,$$

and we are done.

4. (T, P) is not finitarily codable. The process (T, P) inherits the properties of each $(\overline{\overline{T}}_i, \overline{\overline{P}}_i)$: any arithmetic progression of times at which 0 occurs with separation i can have at most k_i terms. If (T, P) were the image of a finitary code from an independent process, there would be some finite string from the independent process which would code to 0. This string would recur with positive probability in arbitrarily long arithmetic progressions with separation i whenever i is larger than the length of the string. Hence (T, P) cannot be a finitary image of an independent process.

5. Images under finite expected time to code. A coding $\phi: A_1^Z \to A_2^Z$ is finitary if for μ_1-a.e. $(x_i) \in A_1^Z$ there exists a pair (s, t) of integers assigned to (x_i) such that whenever $(x_s, \ldots, x_t) = (\hat{x}_s, \ldots, \hat{x}_t)$ and whenever ϕ is defined then $(\phi(x))_0 = (\phi(\hat{x}))_0$. Hence we may define a random variable T, called the time to code, by $T(x_i) = \min\{\max(|s|, |t|): (s, t) \text{ assigned to } (x_i)\}$.

Following [2] we define the term boundedly coded. Let $\mathcal{P}$ and $\mathcal{Q}$ be two partitions of a measure space (X, μ) with automorphism τ. We say that $\mathcal{P}$ is boundedly coded by $\mathcal{Q}$ with respect to τ if for every $\varepsilon > 0$ there is a $K = K(\varepsilon)$ with $\bigvee_{j=0}^{n} \tau^j \mathcal{P} \subseteq_\varepsilon \bigvee_{j=-K}^{n+K} \tau^j \mathcal{Q}$ for all n. It is shown in [2] that if $\mathcal{P}$ is boundedly coded by $\mathcal{Q}$ and $\mathcal{Q}$ is weak Bernoulli, then $\mathcal{P}$ is also weak Bernoulli.

LEMMA. *Suppose that ϕ: $(A^Z, \tau) \to (B^Z, \sigma)$ is a finitary coding with finite expected time to code. Then $\phi^{-1}(B)$ is boundedly coded by A.*

PROOF. Since $\infty > \sum_{t=1}^{\infty} t\mu\{T = t\} = \sum_{t=1}^{\infty} \mu\{T \geq t\}$, for $\varepsilon > 0$ we can choose K such that $\sum_{t=K}^{\infty} \mu\{T \geq t\} < \varepsilon/2$. Now for $0 \leq k \leq [n/2]$ we observe that $\tau^k(\phi^{-1}B) \subseteq_{\mu\{T \geq K+k\}} \bigvee_{-K}^{K+n} \tau^j A$ and $\tau^{n-k}(\phi^{-1}B) \subseteq_{\mu\{T \geq K+k\}} \bigvee_{-K}^{K+n} \tau^j A$ so that we have at last $\bigvee_0^n \tau^j(\phi^{-1}B) \subseteq_\varepsilon \bigvee_{-K}^{K+n} \tau^j A$ for all $n \geq 0$.

REFERENCES

1. M. A. Akcoglu, A. del Junco and M. Rahe, *Finitary codes between Markov processes* (preprint).

2. R. Bowen, *Smooth partitions of Anosov diffeomorphisms are weak Bernoulli*, Israel J. Math. **21** (1975), 95–100.

3. D. S. Ornstein, personal communication.

4. M. Smorodinsky, *A partition on a Bernoulli shift which is not weak Bernoulli*, Math. Systems Theory **5** (1971), 201–203.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, TORONTO, CANADA M5S 1A1

Current address (A. del Junco): Department of Mathematics, Ohio State University, Columbus, Ohio 43210

Current address (M. Rahe): Department of Mathematics, Texas A&M University, College Station, Texas 77843