ISRAEL JOURNAL OF MATHEMATICS, Vol. 21, Nos. 2–3, 1975

FACTORS OF BERNOULLI SHIFTS

BY

DONALD S. ORNSTEIN*

ABSTRACT

An example is constructed of a proper factor of a Bernoulli shift, that cannot be increased without increasing its entropy, and still has no independent complement. The construction mirrors, in a sense, that of a K-automorphism that is not a Bernoulli shift.

Introduction

By a factor of T we mean a sub-sigma-algebra invariant under T. It is known (see [1]) that every factor of a Bernoulli shift is a Bernoulli shift (i.e., if T has an independent generator, then so does every factor of T). The next step in the study of factors is to see how a factor sits in a Bernoulli shift. (We say that two factors sit the same way if there is an automorphism of the Bernoulli shift taking one factor onto the other.) There is a way of looking at the factors of a Bernoulli shift so that their classification mirrors the classification of transformations. If a factor $\mathscr{A}$ of T, has the same entropy as T, then we say that the action of T relative to $\mathscr{A}$ is analogous to a 0-entropy transformation. If $\mathscr{A}$ cannot be increased without increasing its entropy (we then call $\mathscr{A}$ maximal relative to its entropy or simply maximal), we will say that T relative to $\mathscr{A}$ is analogous to a K-automorphism. (A K-automorphism has no factors of 0-entropy.) Another possibility is that $\mathscr{A}$ splits off. (By this we mean that T is the direct product of $\mathscr{A}$ and another orthogonal factor, or, stated in another way, there is a factor $\mathscr{B}$ such that any set in $\mathscr{A}$ is orthogonal to any set in $\mathscr{B}$ and $\mathscr{A} \vee \mathscr{B}$ generate. As we already noted, if T is Bernoulli then so is T acting on $\mathscr{B}$.) If $\mathscr{A}$ splits off then we say that T relative to $\mathscr{A}$ is analogous to a Bernoulli shift.

* This research was supported in part by National Science Research Grant NSF GP 33581X.
Received January 14, 1975

It can be seen from [6] or Lemma 2 in [5] that if $\mathcal{A}$ splits off then $\mathcal{A}$ is maximal. (This is analogous to the fact that Bernoulli implies K.)

The purpose of this paper is to prove the following:

THEOREM. *There exists a factor of a Bernoulli shift that is maximal but does not split off.*

This theorem is analogous to the theorem that there is a K-automorphism that is not Bernoulli. The analogy is a strong one, since our example will be formed by taking a skew product with a K-automorphism that is not Bernoulli.

I think that it would be interesting to see how far these analogies go and to what extent the classification of transformations is mirrored by the factors of a Bernoulli shift.

So far the positive results about Bernoulli shifts and the negative results about K-automorphisms have been fairly separate, and it might be of some interest to note that the study of factors of a Bernoulli shift combines both of these areas.

There are some beautiful and deep results in the positive direction about when a factor splits off, due to Jean Paul Thouvenot. He gets splittings from "relativised" isomorphism theorems, and it is from him that I got the idea of looking at a transformation "relative to a factor".

1. Description of the example

Our example will consist of a certain skew product with a K-automorphism that is not Bernoulli. Instead of describing this example here, we will assume that the reader is familiar with the example in [1]. We will refer to that transformations as $\tilde{T}$ and to the special partition described there as $\tilde{P}$.

The other part of the skew product will consist of a Bernoulli shift T_B and a generator $P = \{P_0, P_1, P_e, P_f\}$. T_B, P will be described below but first we will show how to put $\tilde{T}, \tilde{P}$ and T_B, P together.

Let X be the space on which T_B acts and let Y be the space on which $\tilde{T}$ acts. T will act on $X \times Y$ as follows: $T(x, y) = (T_B x, \tilde{T} y)$ if $x \in P_0$ and $T(x, y) = (T_B x, y)$ if $x \notin P_0$.

With a small abuse of notation we will let P also denote the partition of $X \times Y$, that partitions according to the atom of P containing x. $\bigvee_{-\infty}^{\infty} T^i P$ is the factor that is maximal but does not split off.

It will also be convenient to define $\bar{P} = \{\bar{P}_0, \bar{P}_1, \bar{P}_e, \bar{P}_f, \bar{P}_r\}$ as follows: $\bar{P}_r = (P_1 \cup P_e \cup P_f) \times Y$. On $P_0 \times Y$, $\bar{P}$ will coincide with $X \times \tilde{P}$. (Thus $\bar{P}_t = P_0 \times \tilde{P}_t$, for $t = 0, 1, e, f$.)

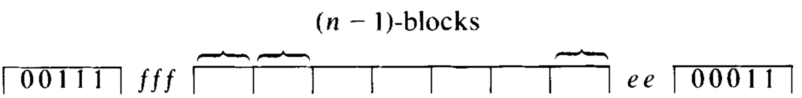
$P \vee \bar{P}$ will be a convenient generator for T.

We will now describe T_B, P. We will do this informally, in terms of the P-names of points. This can be made rigorous by a stacking construction in exactly the same way that the informal description of $\tilde{T}, \tilde{P}$ in terms of $\tilde{P}$ names is made rigorous. It will be fairly obvious how to proceed and we will spare the reader the details.

P will have four atoms, P_0, P_1, P_e, P_f and the names of points will have a block structure similar to that of $\tilde{P}, \tilde{T}$. There will be several kinds of n-blocks but they will all have the same length. The n-blocks will be formed as follows: There will be a fixed number, l_n, of consecutive $(n-1)$-blocks and the type of $(n-1)$-block in each position will be independent of the type of $(n-1)$-block in any other position. In front of these $(n-1)$-blocks we will see a string of consecutive f, and at the end of the $(n-1)$-blocks a string of consecutive e. The number of f plus the number of e will be a fixed number $f(n)$, and each of these possibilities will occur with the same probability and independently of anything else.

In front of the string of f's there will be a string of 0 and 1 of a fixed length $g(n)$. After the string of e there will be another string of 0 and 1 of length $g(n)$. The total number of 0's in both strings will be $g(n)$, both strings will begin with consecutive 0's and end with consecutive 1's. Each possibility will have the same probability and will be independent of anything else.

Here is a picture of an n-block:



To get things going we could define a 1-block to be simply a string of 0.

Choice of $f(n)$, $g(n)$ and $l(n)$

$f(n)$ will be large compared to the length of an $(n-1)$-block and small compared to the length of an n-block. We will want:

$$\text{length of } (n-1)\text{-block}/f(n) \to 0 \text{ and } \sum f(n)/\text{length of } n\text{-block} < \infty.$$

Let $h(n)$ denote the number of 0's in an n-block. Let $\tilde{h}(n)$ denote the length of a $\tilde{P}$-n-block. $l(n)$ will be chosen so that $\tilde{h}(n) < h(n) < \tilde{h}(n+1)$, and furthermore $h(n)/\tilde{h}(n+1) \to 0$ and $\tilde{h}(n)/h(n) \to 0$.

$g(n)$ will be chosen so that $\tilde{h}(n) < g(n) < h(n)$, and furthermore $\tilde{h}(n)/g(n) \to 0$ and $\sum g(n)/h(n) < \infty$.

It is easy to see (because of the growth rate of $\tilde{h}(n)$) that $f(n)$, $g(n)$ and $l(n)$ can be chosen so that they will satisfy the above conditions.

3. Proof that T is Bernoulli

We will prove that T is Bernoulli by checking that $T, P \vee \bar{P}$ is VWB. We will start by stating what has to be checked in the terminology of stationary processes. A sequence of partitions $\{Q_i\}_1^n$ can be thought of as a measure on sequences of length n, since each point in the measure space has a Q name. To check that two measures on sequences of length n are $\bar{d}$ close we must find a measure preserving correspondence between the sequences under the 2 measures, such that most sequences correspond to a sequence that differs in a small percentage of places.

To check VWB we must show that: given ε there is an n^1 such that for all m, we can ignore a collection of sequences 0 to $-m$ of measure less than ε and for any two remaining past sequences (0 to $-m$) we have that the conditional measure on sequences from 1 to n^1 are $\bar{d}$ close.

We will start by checking that T_B, P is VWB.

We are given ε and the first thing we must do is chose n^1. Instead of doing this now we will note as we go along that we need certain quantities to be small compared to ε, and that they will be small if n^1 is large enough.

Now note that we can condition on any refinement of the past, instead of the past, in the test for VWB. It will be convenient to condition on past sequences (0 to $-m$) together with the terms of the future (1 to n^1) that lie outside a complete n-block. Let A^n and B^n be two such refined past atoms. Conditioning on A^n (or B^n) fixes the positions of the n-blocks between 0 to n^1, but puts no restrictions on which n-block can appear. The n-block appearing in one position is thus independent of the n-blocks appearing in other positions. We will call the future strings (from 0 to n^1) conditioned by A^n (or B^n) A^n (or B^n) strings.

We must show that the $\bar{d}$ distance between A^n and B^n strings is $< \varepsilon$, ignoring ε of the $A^n \cup B^n$.

The first step will be to group the A^n and B^n strings according to the number of f at the beginning of each n-block. (Fixing the f is the same as fixing everything outside of the $(n-1)$-blocks. Our groups are thus A^{n-1} strings.) Each group will have the same measure and we will show that (A) we can set up a $1-1$ correspondence between the groups of A^n and B^n strings so that $> \frac{1}{4}$ of the $(n-1)$-block coincide. (More precisely: the positions of $(n-1)$-blocks are the same for all strings in a group. For each A^n group look at the fraction of $(n-1)$-blocks such that the corresponding B^n group has an $(n-1)$-block in exactly the same position. Now average over all A^n groups. This average will be $> \frac{1}{4}$.)

We can see (A) as follows: If n^1 is large enough, we can choose n so that we need only consider A^n and B^n strings such that all but $\frac{1}{10}\varepsilon n^1$ of their terms lie in a complete n-block.

Furthermore, we can make a $1-1$ correspondence between the (positions of) n-blocks in A^n and B^n strings so that all but $\frac{1}{100}\varepsilon$ of the pairs overlap in $>\frac{1}{4}$ of the length of an n-block.

Fixing attention on the first pair of positions, we can pair off the n-blocks that can fill those positions according to the number of f at their beginning. Because $f(n)$ is large compared to the length of an $(n-1)$-block, we can do this pairing so that for all but $\frac{1}{100}\varepsilon$ of the pairs, a, b, we have that if the positions of an $(n-1)$-block in a and an $(n-1)$-block in b intersect, then the positions are the same.

Because the n-blocks appearing in any two positions are independent, we can line up all of the pairs simultaneously getting (A).

We will get that the A^n and B^n strings are $\bar{d}$ close by repeated applications of (A). Let A^{n-1} and B^{n-1} be a pair of corresponding A^n, B^n groups as in (A). We are in the same situation as before, except that $>\frac{1}{4}$ of the $(n-1)$-blocks line up exactly. We can group the A^{n-1} and B^{n-1} strings according to the positions of the $(n-2)$-blocks, so that the $(n-2)$-blocks coming from $(n-1)$-blocks that are lined up remain lined up and $>\frac{1}{4}$ of the rest of the $(n-2)$-blocks now line up.

Repeated application of this process means that we can group the A^n and B^n strings according to the position of the $(n-k)$-blocks, and make a correspondence between these groups so that $>(1-(\frac{3}{4})^k)$ of the $(n-k)$-blocks line up. Since all of the $(n-k)$-blocks have the same probability and occur independently, we can extend our correspondence to strings so that in corresponding strings the $(n-k)$-blocks that line up are identical. Furthermore, if n^1 is large enough most of the terms in an n^1 string can be assumed to lie in $(n-k)$-blocks. This proves that the A^n and B^n strings are $\bar{d}$ close.

We will now check VWB for $T, P \vee \bar{P}$. Each x, y now has a pair of names, a P and $\bar{P}$ name. For each x, y we will define the $\tilde{P}$ name of x, y as follows: Take those terms in the $\bar{P}$ name of x that correspond to a 0 in the P name (those terms $\neq r$) and order them consecutively, taking as the first term the atoms of $\bar{P}$ containing the first $T^i x$, $0 \leqq i$, such that $T^i x \in P_0$. The $\tilde{P}$ name of x, y depends only on y and will be a typical name for the K-automorphism that is not Bernoulli. Note that the $\tilde{P}$ name does not shift properly with T.

Now let A and B denote two past $(-1-m)$ $P \vee \bar{P}$ strings. We must show (for most choices of A and B) that the $\bar{d}$ distance between the future $P \vee \bar{P}$ strings (0 to n') conditioned on A and B is small.

Our previous argument still applies (because the $\tilde{P}$ name is independent of the P-name) so that (changing $n-k$ to n) we have that we can group the A and B strings according to the positions of the P-n-blocks and form a measure preserving correspondence between the A and B groups, so that most of the P-n-blocks in corresponding groups line up. Let A_n and B_n denote two such corresponding groups.

Let us call the past or future $\tilde{P}$ name the part of the $\tilde{P}$ name corresponding to the $\bar{P}$ name from -1 to $-m$ or 0 to n'.

Let us further group the A_n and B_n strings according to the part of the future $\tilde{P}$ name that does not lie in a complete $\tilde{P}$-n-block. Let $\bar{A}_n$ (or $\bar{B}_n$) denote one such A_n (or B_n) group.

We must show that for most $\bar{A}_n$ strings and $\bar{B}_n$ strings their $\bar{d}$ distance is small. (We will do this by lining up the $\bar{P}$ names in the n-blocks that line up.)

If a is a P-n-block in the name of (x, y), define $\tilde{a}$ to be the part of the $\tilde{P}$ name of (x, y) corresponding to the 0's in the $(n-1)$-blocks in a. (If we change the number of 0's at the beginning of a without changing the $\tilde{P}$ name of x, y – i.e., keep y fixed—then we shift $\tilde{a}$.)

We will check the $\bar{d}$ distance in two steps. The first step will be to show that (1) we can group the $\bar{A}_n$ and $\bar{B}_n$ strings according to the number of 0's at the beginning of each P-n-block, and pair these groups so that if a and b are P-n-blocks in strings in corresponding groups, occupying the same position, then (for most positions and most a, b) one of the $\tilde{P}$-n-blocks in $\tilde{a}$ and $\tilde{b}$ have the same position. (Because of the varying lengths of s between $\tilde{P}$-n-blocks we will not be able to line up more than one pair. However, the other pairs will not line up too badly.)

We can see (1) as follows: because the total number of 0's in a P-n-block is large compared to the length of a $\tilde{P}$-n-block and small compared to the length of a $\tilde{P}$-$(n+1)$-block, we get that most $\tilde{a}$ (and $\tilde{b}$) consist of a large number of $\tilde{P}$-n-blocks all belonging to the same $\tilde{P}$-$(n+1)$-block (The measure of the exceptions can be made arbitrarily small by taking n' and n large enough). Furthermore, the number of possible 0's at the beginning of a P-n-block is large compared to the length of a $\tilde{P}$-n-block. Fix a position that is occupied by a P-n-block in both $\bar{A}_n$ and $\bar{B}_n$ strings. We can group the $\bar{A}_n$ and $\bar{B}_n$ strings according to the number of 0's at the beginning of P-n-blocks in this position, and pair these groups so that for most pairs, if a and b denote the P-n-blocks in that position, then one $\tilde{P}$-n-block in $\tilde{a}$ and $\tilde{b}$ have the same position. Because of the independence of the P-n-blocks in different positions we can continue this process to prove (1).

The second step goes as follows: Let $\bar{A}'_n$ and $\bar{B}'_n$ denote two corresponding groups of $\bar{A}_n$ and $\bar{B}_n$ strings produced in (1). Let a and b be P-n-blocks in an $\bar{A}_n$ and $\bar{B}_n$ string, occupying the same position. For most a and b, $\tilde{a}$ and $\tilde{b}$ are mostly made up of $\tilde{P}$-n-blocks belonging to the same $\tilde{P}$-$(n+1)$-block, and one $\tilde{P}$-n-block in $\tilde{a}$ has the same position as a $\tilde{P}$-n-block in $\tilde{b}$. This implies that for any $\tilde{P}$-n-block in $\tilde{a}$ there is a P-$(n-1)$-block in $\tilde{b}$ whose position differs by less than the length of a $\tilde{P}-(n-1)$ block (see [1]). We can also assume that corresponding $\tilde{P}$-n-blocks are the same by a further grouping according to the $\tilde{P}$-names. Let a_1 and b_1 denote corresponding P-$(n-1)$-blocks in a and b. Then for most a_1(and b_1), $\tilde{a}_1$(and $\tilde{b}_1$) lies entirely in one $\tilde{P}$-n-block in $\tilde{a}$ (or $\tilde{b}$). The positions of these $\tilde{P}$-n-blocks (in $\tilde{a}$ and $\tilde{b}$) differ by less than the length of a $\tilde{P}$-$(n-1)$-block. The number of possible 0's at the beginning of a_1 (or b_1) is large compared to the length of a $\tilde{P}$-$(n-1)$-block. We can therefore pair the possible a_1 and b_1 in this position according to the number of 0's at the beginning, so that if a_1 corresponds to b_1 then they differ only by the initial number of 0's and for most pairs $\tilde{a}_1$ is the same as $\tilde{b}_1$. We can do this independently for each position of a P-$(n-1)$-block. This gives the desired result.

4. Proof that $\bigvee_{-\infty}^{\infty} T^iP$ does not split off

We will now prove that $\bigvee_{-\infty}^{\infty} T^iP$ does not split off.

We will argue by contradiction. We will suppose that there is some partition Q such that $\bigvee_{-\infty}^{\infty} T^iQ \perp \bigvee_{-\infty}^{\infty} T^iP$ and $P \vee Q$ generate. This implies that for any ε there is a K such that $\bar{P} \overset{\varepsilon}{\subset} \bigvee_{-K}^{K} T^i(P \vee Q)$. In process terminology: we can find a finite code which applied to the $P \vee Q$ name of x, y gives the $P \vee \bar{P}$ name, except for ε errors.

Let a_i denote a collection of $P \vee Q$ string such that the P part consists of two consecutive P-n-blocks, fixed, except for the number of 0 at the end of the first block and at the beginning of the second, while the Q part is arbitrary but fixed. (Thus, as i changes, the only thing that changes is the number of 0's between the two P-n-blocks. For convenience we can set i equal to the above number of 0's.)

Recall that there was an $\alpha > 0$ associated with $\tilde{P}, \tilde{T}$ (either n-blocks in 2 names lined up better than n^3, or the fraction of disagreement was $> \alpha$).

Now, suppose that a_i occurs in the $P \vee Q$ name of some x, y. Let b denote the string of corresponding terms in the $P \vee \bar{P}$ name. We will say that a_i codes badly if either the finite code applied to a_i differs from b in a fraction of places exceeding α, or if the $\tilde{P}$-n-blocks† in b do not all lie in the same $\tilde{P}$-$(n+1)$-block. The probability that a_i codes badly can be made arbitrarily small by taking K and n large enough.

(A) Suppose that for some i, a_i appears in the name of some x, y and does not code badly. If $|k| > n^3$ then a_{i+k} must code badly wherever it occurs. (A) follows from basic properties of $\tilde{P}$ names: If a string is such that more than $\frac{9}{10}$ of its term lies in $\tilde{P}$-n-blocks all belonging to the same $\tilde{P}$-$(n+1)$-block, and there are > 10 $\tilde{P}$-n-blocks, and if we modify the string by inserting $> n^3$ terms near the middle ($> \frac{1}{4}$ of the terms in the string lie on each side of the insertion), then it is impossible for both the original and modified string to α-occur in a $\tilde{P}$-$(n+1)$-block. See [1].

The number of possible 0's at the beginning and end of a P-n-block will be large compared to n^3 (for large n) and all of these possibilities occur with the same probability: Hence all of the a_i occur with the same probability and for most i, a_i and a_{i+n^3} are both defined. Hence for n-large enough at least $\frac{1}{4}$ of the a_i code badly, giving a contradiction.

5. Proof that $\vee_{-\infty}^{\infty} T^iQ$ is maximal

Let Q be a partition such that $Q \not\subset \vee_{-\infty}^{\infty} T^iP$. We will show that $H(P \vee Q, T) > H(P, T)$.

If R is a partition we will let R_x denote the partition of Y that R induces on $x \times Y$. It is enough to show that there is an $\alpha > 0$ and

$$\text{(A)} \lim (1/n) H(\vee_1^n T^iQ)_x > \alpha \text{ for a.e. } x.$$

It is easy to see that we can find a set $F \subset X$ of non zero measure and a partition $\bar{Q}$ of Y, $H(\bar{Q}) > 0$, such that for all x in F there is a partition $R(x)$ of Y, $R(x) \vee Q_x \supset \bar{Q}$ and $H(R(x)) < \frac{1}{2} H(\bar{Q}, \tilde{T})$. We could also assume that $F \subset P_0$ (since we could replace Q by any translate of Q).

In order to get (A) we first note that we have (B) $\{(T^iQ)_x\}_{i=1}^{i=n}$ contains more than $\frac{1}{2}m(F) \cdot n$ partitions of the form $\tilde{T}^iQ_{x_i}$, where all of the i are distinct and $x_i \in F$.

† We use here the notation of the previous section.

We get (B) as follows: suppose $T^{-i}x \in F$. Let $r(i)$ be the number of j, $-i \leqq j \leqq 0$ such that $T^j(x) \in P_0$. Then $(T^iQ)_x = \bar{T}^{r(i)}(Q_{T^{-i}x})$. The ergodic theorem now tells us that the fraction of $-i$ such that $T^{-i}x \in F$ tends to $m(F)$.

We get (A) from (B) as follows: Let S be the set of i in (B), and let $|S|$ denote the number of i in S. Then

$$H\left[\bigvee_{i \in S} \bar{T}^i(Q_{x_i} \vee R(x_i))\right] > H\left(\bigvee_{i \in S} \bar{T}^i\bar{Q}\right) \geqq |S| \cdot H(\bar{Q}, \bar{T}).$$

We also have

$$H\left(\bigvee_{i \in S} \bar{T}^i R(x_i)\right) < |S| \cdot \tfrac{1}{2} \cdot H(\bar{Q}, \bar{T}).$$

Therefore,

$$H\left(\bigvee_{i \in S} \bar{T}^i Q_{x_i}\right) > \tfrac{1}{2}|S| H(\bar{Q}, \bar{T}) > \tfrac{1}{4} m(F) \cdot n,$$

which gives (A).

REFERENCES

1. D. Ornstein, *Ergodic Theory, Randomness, and Dynamical Systems*, Yale U. Press, 1974.

2. J.-P. Thouvenot, *Quelques propriétés des systèmes dynamiques qui se décomposent en un produit de deux systèmes dont l'un est un schema de Bernoulli.* Israel J. Math. **21** (1975), 177–207.

3. J.-P. Thouvenot, *Une classe de systèmes pour lesquels la conjecture de Pinsker est vraie*, Israel J. Math. **21** (1975), 208–214.

4. J.-P. Thouvenot and P. C. Shields, *Entropy zero × Bernoulli processes are closed in the $\bar{d}$-metric*, to appear.

5. D. S. Ornstein and B. Weiss, *Finitely determined implies very weak Bernoulli*, Israel J. Math. **17** (1974), 94–104.

6. K. Berg, *Convolution of invariant measures, maximal entropy*, Math. Systems Theory **3** (1969), 146–151.

STANFORD UNIVERSITY
STANFORD, CALIFORNIA, U.S.A.