ISRAEL JOURNAL OF MATHEMATICS, Vol. 24, Nos. 3–4, 1976

ON ATTAINING $\bar{d}$

BY

DANIEL J. RUDOLPH AND GIDEON SCHWARZ

ABSTRACT

This paper examines the restrictions that may be put on joint distributions of two or more stationary stochastic processes, and still attain $\bar{d}$, or approach it.

1. Introduction

When $\bar{d}$ was first defined by Ornstein [2], he showed that for two stochastic processes $\bar{d}$ is attained by a stationary joint distribution. Furthermore, this distribution can be chosen to be ergodic, if the given processes are ergodic. Here we study the possibility of extending this result in two different directions. One is the question of attaining $\bar{d}$ simultaneously for a collection of overlapping pairs of processes. The other direction is the question whether ergodicity can be replaced by a stronger degree of mixing.

For the first question, the possibility of simultaneously attaining $\bar{d}$ between consecutive entries of a sequence of processes, recently proved by Shields and Thouvenot [4], is shown to be an easy consequence of a well-known existence theorem on Markov processes. Then a simple construction is used to show that simultaneous attainment is no longer generally possible, if the family of pairs for which $\bar{d}$ is to be attained contains a loop when it is regarded as a graph.

For the second question, no stronger degree of mixing can be expected from a joint distribution without assuming the same of both given processes, since they are factors of the joint process, and factors inherit mixing properties. Here we show that even the slightest strengthening cannot hold in general: we construct a pair of Bernoulli processes for which $\bar{d}$ cannot be attained even by a totally ergodic joint distribution, and each of the two processes is nothing worse than a two-point extension of a two-shift that can be obtained from a four-state Markov chain by lumping together two of its states.

On the other hand, some of the machinery of Ornstein's Isomorphism Theorem is used to show that for a pair of processes, one of which is Bernoulli

Received December 15, 1975

and the other one ergodic and of entropy at least as large, $\bar{d}$ is approached arbitrarily closely by joint processes that not only share the mixing properties of the second process, but are in fact isomorphic to it.

2. Simultaneous attainment

PROPOSITION 1. *A sequence of stationary stochastic processes can be realized simultaneously on one space, so that $\bar{d}$ between consecutive members of the sequence is attained.*

PROOF. Let $X(n,i)$ be the ith random variable of the nth process in the sequence. For each n, regard $X(n,\cdot)$ as a (sequence-valued) random variable, disregarding temporarily its process-nature. Each such variable has a given distribution (on sequence space). Whenever we are also given a joint distribution of $X(n,\cdot)$ and $X(n+1,\cdot)$ for each n that is compatible with the distributions of the individual $X(n,\cdot)$, there exists a joint distribution of the entire sequence $X(1,\cdot)$, $X(2,\cdot),\cdots$ that is compatible with the given joint distributions of consecutive $X(n,\cdot)$: simply let the sequence be the (not necessarily stationary) Markov process whose transition probabilities are the conditional distributions of $X(n+1,\cdot)$ given $X(n,\cdot)$ that are determined by their joint distribution. The existence of this Markov process is a classical theorem of Ionescu–Tulcea [1]. Now, recalling the fact that each $X(n,\cdot)$ is a stationary stochastic process, one easily sees that if all the given joint distributions are also stationary, that is, invariant under the shift of the index i, so will be the joint distribution of all the $X(n,i)$ in the Markov process. Q.E.D.

The proof that no similar result can hold for attaining $\bar{d}$ between neighbours in a closed loop of three or more processes is based on the following combinatorial considerations. For fixed $n \geqq 3$, let A_k be, for $k = 1,2,\cdots,n$, the set of numbers from 1 to n with the exception of k. Any two different A_k have exactly $n-2$ elements in common. For $k = 1,2,\cdots,n-1$, the mapping f_k of A_{k+1} onto A_k that leaves their common elements fixed sends k to $k+1$. The composition $F = f_1(f_2(\cdots f_{n-1}(\cdot))\cdots)$ then maps A_n onto A_1 by sending each k to $k+1$, and this mapping leaves no point fixed.

Now let $X(1)$ be any stationary stochastic process with states in A_1 whose distribution is invariant under permutations of the states (for example, the Bernoulli shift based on the equidistribution on A_1). For $k = 2,\cdots,n$, let $X(k)$ be a copy of this process, where the states were relabelled by the elements of A_k.

Any two different X_k have $n-2$ states in common, so the distance between their zero-time partitions is $1/(n-1)$, which for n at least 3 is less than 1. But this distance is clearly also the $\bar{d}$-distance between the processes. The unique joint distribution of $X(k)$ and $X(k+1)$ that attains $\bar{d}$ is given by putting $X(k) = f_k(X(k+1))$, but this implies $X(1) = F(X(n))$, so that $X(1)$ and $X(n)$ are completely mismatched when all other pairs of neighbours in the loop $X(1)$, $X(2), \cdots, X(n-1)$, $X(n)$, $X(1)$ are matched to attain their $\bar{d}$-distance.

This counterexample can be joined to a slight generalization of the proposition preceding it, to yield: *For a graph G, the absence of closed loops is a necessary and sufficient condition for the simultaneous attainability of $\bar{d}$ on all edges, for any assignment of stationary stochastic processes to the nodes of the graph.*

3. Two Bernoulli processes with no totally ergodic $\bar{d}$-joining

First, let $Y = (Y_n)$ be a coin-tossing process, with its states labelled 1 and 2. A second coin tossing process X is defined by putting $X_n = 0$ when $Y_{n+1} = Y_n$, and $X_n = 1$ otherwise. The process (X, Y) is a Markov chain with transition matrix

$$\begin{array}{c|cccc} & 01 & 02 & 11 & 12 \\ \hline 01 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 02 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 11 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 12 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{array}$$

and stationary measure $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Being trivially isomorphic to Y, the process (X, Y) is isomorphic to the two-shift (on the other hand, when viewed from X, it is a two-point extension of a two-shift).

Next, define a process (X, Z) by letting Z be a third coin tossing process, whose joint distribution with X is the same as that of Y with $(1 - X_n)$. In other words, $Z_{n+1} = Z_n$ whenever $X_n = 1$. Although only the joint distributions of X with Y, and of X with Z have thus been defined, it can be shown that there is a unique stationary process (X, Y, Z) compatible with the joint distributions that were defined. In fact, it is an (eight-state) Markov chain. We shall, however, not use this uniqueness, since what follows would hold for any distribution of the triple process that projects on (X, Y) and on (X, Z) as defined above. In any case, (X, Y, Z) is periodic, since exactly one of the equalities $Y_{n+1} = Y_n$ and $Z_{n+1} = Z_n$ holds for each n, so that the parity of $Y_n + Z_n$ changes at each step.

Finally, put $R_n = X_nY_n$ and $S_n = -X_nZ_n$. The processes R and S can be thought of as resulting from (X, Y) and (X, Z), respectively, by erasing Y_n and Z_n whenever $X_n = 0$ and then relabelling the states 1 and 2 of the second process with the corresponding negative numbers. Neither R nor S are Markov chains, but as factors of Bernoulli processes they are themselves Bernoulli. Since each of the erased Y_n and Z_n can be retrieved almost surely from R and S, respectively, by proceeding along the sequence till the nearest nonzero entry, R and S are improper factors, that is, they generate (X, Y) and (X, Z).

Now R and S share no states but 0, and $R_n = S_n$ if and only if $R_n = S_n = X_n = 0$, which happens with probability $\frac{1}{2}$, so the joint distribution of R and S that was forced by the joint distribution of R and X, and of S and X, does attain $\bar{d}(R, S)$. Conversely, any stationary joint distribution of a copy of R and a copy of S that is to attain $\bar{d}$ must match all zeros perfectly. Defining $X_n = 0$ whenever $R_n = S_n = 0$ and $X_n = 1$ otherwise, the triple process (X, R, S) will have the same distribution as before, and so will *a fortiori* the pair (R, S).

Thus there is a unique process (R, S) that attains $\bar{d}(R, S)$, and since it is an improper factor of (X, Y, Z), it also has a factor of period two, and is not totally ergodic.

4. Bernoulli joint distributions will approach $\bar{d}$

Although $\bar{d}$ will not be attained by a Bernoulli joint distribution of the two processes in the preceding example, $\bar{d}$ will be approached arbitrarily closely. In fact, *if X is a Bernoulli process, and Y an ergodic process of entropy at least that of X, and $\varepsilon > 0$, there is a joint distribution of X and Y that makes X a factor of Y, and approaches $\bar{d}(X, Y)$ within ε (thus the joint distribution shares all mixing properties of Y).*

Since this is a modification of Sinai's Theorem, we proceed to formulate it in the language that will enable us to modify Ornstein's proof of Sinai's Theorem [2] to yield the result required here:

PROPOSITION 2. *Let $(T, \hat{P})$ be ergodic and $(\bar{T}, \bar{P})$ finitely determined of smaller or equal entropy. For $\varepsilon > 0$, there is a Q_ε that is $(T, \hat{P})$-measurable, such that (T, Q_ε) is a copy of $(\bar{T}, \bar{P})$, and the partition-distance $|\hat{P}, Q_\varepsilon| \leqq \bar{d}((T, \hat{P}), (\bar{T}, \bar{P})) + \varepsilon$.*

The proposition is an easy consequence of two lemmas. The first is Ornstein's Fundamental Lemma, which can be found on p. 84 of P. Shields [3]. We adopt his notation. The second is a modification of another lemma of Ornstein [2], and a sketch of a proof follows the statement.

LEMMA. *Let $(T, \hat{P})$ and $(\bar{T}, \bar{P})$ be ergodic, $H(T, \hat{P}) \geqq H(\bar{T}, \bar{P}) > 0$. Given ε, n and δ there is a $P \in \vee_{-\infty}^{\infty} T^i(\hat{P})$ so that:*

i) $|d(\vee_0^n T^i(P)) - d(\vee_0^n \bar{T}^i(\bar{P}))| \leqq \delta$,

ii) $|H(T, P) - H(\bar{T}, \bar{P})| \leqq \delta$, and

iii) $|\hat{P}, P| \leqq \bar{d}((T, \hat{P}), (\bar{T}, \bar{P})) + \varepsilon$.

We can assume that both processes appear as factors of an ergodic joint process $(U, \hat{P} \vee \bar{P})$ where $|\hat{P}, \bar{P}| = \bar{d}((T, \hat{P}), (\bar{T}, \bar{P}))$.

Conditions i), ii) and iii) can all be achieved by a construction on a tall enough Rohlin tower. That is to say, we build a tower with levels that are $(U, \hat{P})$ measurable. In this tower we assign to each $U, \hat{P}$ column-name a U, P-name that is the $U, \bar{P}$-name of some point in that $U, \hat{P}$-column. If the tower is tall enough, on all but a set of $\hat{P}$ columns of measure less than $\delta/3$ we can choose the $\bar{P}$-name we assign, to satisfy within $\delta/3$ the ergodic theorem on sets in $\vee_{-n}^{n} U^i(\bar{P})$. This will give us condition i) on P.

Also, if the tower is tall enough, on all but a set of $\hat{P}$-columns of measure at most $\varepsilon/3$ we can choose the point in the column whose $\bar{P}$-name is assigned to the column, so that its $\hat{P} \vee \bar{P}$-name up the column satisfies, within $\varepsilon/3$, the ergodic theorem on sets in $\hat{P} \vee \bar{P}$. This will guarantee condition iii) as the set where $\hat{P}$ and P differ will have size at most $\bar{d}((T, \hat{P}), (\bar{T}, \bar{P})) + \varepsilon$.

We get condition ii) in the standard manner by using an auxiliary independent process $(U, \bar{R})$ independent of $(U, \bar{P})$ with $H(U, \bar{P} \vee \bar{R})$ slightly larger than $H(U, \hat{P})$. The Marriage Lemma is then used, so that assigned $\bar{P} \vee \bar{R}$ columns are exactly the $\hat{P}$-columns. Continuity of entropy in the partition metric finishes the argument.

This establishes Proposition 2. Since the example in the preceding section shows that generally ε cannot be replaced by zero, the Q_ε cannot be made to converge when ε tends to zero.

Donald Ornstein's suggestion, that some of the methods used in the second part of Section 2 may yield the result in Section 3, is one of several ideas he generously shared with us.

REFERENCES

1. C. Ionescu-Tulcea, *Mesures dans les espaces produits*, Atti Accad. Naz. Lincei Rend. **7** (1949), 208–211.

2. D. S. Ornstein, *Ergodic Theory, Randomness and Dynamical Systems*, Yale Univ. Press, New Haven, 1973.

3. P. Shields, *The Theory of Bernoulli Shifts*, Univ. Chicago Press, Chicago, 1973.

4. P. Shields and J. P. Thouvenot, *Entropy zero x Bernoulli processes are closed in the $\bar{d}$-metric*, Ann. Prob. **3** (1975), 732–736.

THE INSTITUTE FOR ADVANCED STUDIES
THE HEBREW UNIVERSITY OF JERUSALEM
JERUSALEM, ISRAEL