The Annals of Probability
1989, Vol. 17, No. 4, 1658–1663

CODING A STATIONARY PROCESS TO ONE WITH PRESCRIBED MARGINALS[1]

BY S. ALPERN AND V. S. PRASAD

London School of Economics and University of Lowell

In this paper we consider the problem of coding a given stationary stochastic process to another with a prescribed marginal distribution. This problem after reformulation is solved by proving the following theorem. Let $(M, \mathscr{A}, \mu)$ be a Lebesgue probability space and let σ be an antiperiodic bimeasurable μ-preserving automorphism of M. Let $\mathbf{N}$ be the set of nonnegative integers. Suppose that $(p_{i,j}: i, j \in \mathbf{N})$ are the transition probabilities of a positive recurrent, aperiodic, irreducible Markov chain with state space $\mathbf{N}$ and that $\pi = (\pi_i)$, $i \in \mathbf{N}$, is the unique positive invariant distribution $\pi_j = \sum_{i \in \mathbf{N}} \pi_i p_{i,j}$. Then there is a partition $\mathbf{P} = \{P_i\}_{i \in \mathbf{N}}$ of M such that for all $i, j \in \mathbf{N}$, $\mu(P_i \cap \sigma^{-1}P_j) = \mu(P_i)p_{i,j} = \pi_i p_{i,j}$.

1. Introduction. This paper is about the problem of coding a given stationary stochastic process to another with a prescribed marginal distribution. We reformulate the problem as one for an abstract dynamical system and give a simple solution by using a result from ergodic theory stating that an antiperiodic dynamical system has a set with a given distribution of return times.

Suppose that the given stationary stochastic process $\{X_n\}_{n=-\infty}^{\infty}$ takes values in a measure space $(S, \mathscr{S})$. Consider the associated dynamical system $(M, \mathscr{A}, \mu, \sigma)$, where $M = S^\infty$ the doubly infinite sequence space, $\mathscr{A} = \mathscr{S}^\infty$, the product σ-field, and μ is the joint distribution of $\{X_n\}$. Since $\{X_n\}$ is stationary, μ is preserved by the left shift transformation σ defined on S^∞ by $(\sigma(s))_n = s_{n+1}$. We wish to code the process $\{X_n\}$ onto a stationary process $\{Y_n\}$ which takes values on a denumerable set that we take to be the nonnegative integers $\mathbf{N}$, so that for a given positive integer m, the joint distribution of $(Y_0, Y_1, \dots, Y_{m-1})$ is a previously specified marginal distribution on $\mathbf{N}^m$. Without loss of generality we may take $m = 2$ and consider given transition probabilities $p_{i,j} = \text{Prob}(Y_{n+1} = j | Y_n = i)$, with invariant distribution $\pi_i = \text{Prob}(Y_n = i)$. By a coding function we mean a measurable function $f: S^\infty \to \mathbf{N}$ so that $Y_j = f(\sigma^j(X_i: -\infty < i < \infty))$ for each j. Each coding function f corresponds to a measurable partition $\mathbf{P} = \{P_i\}_{i \in \mathbf{N}}$ of $M = S^\infty$ with $P_i = f^{-1}\{i\}$ and conversely. Hence if the process $\{X_n\}$ is aperiodic and the given distribution on $\{Y_n\}$ is "mixing," then the possibility of coding is guaranteed by the following theorem.

THEOREM 1. *Let $(M, \mathscr{A}, \mu)$ be a Lebesgue probability space and let σ be an antiperiodic bimeasurable μ-preserving automorphism of M. Let $\mathbf{N}$ be the set of*

Received March 1988; revised October 1988.

[1]The authors were supported by NATO grant 245/89. The second author was also supported by an NSERC grant.

AMS 1980 *subject classifications*. Primary 28D05, 60G10, 60J10.

Key words and phrases. Stationary stochastic process, Markov transitions, coding, partitions, dynamical system, measure preserving transformation, mixing.

nonnegative integers. Suppose that $(p_{i,j}: i, j \in \mathbf{N})$ *are the transition probabilities of a positive recurrent, aperiodic, irreducible Markov chain with state space* $\mathbf{N}$ *and that* $\pi = (\pi_i)$, $i \in \mathbf{N}$, *is the unique positive invariant distribution* $\pi_j = \sum_{i \in \mathbf{N}} \pi_i p_{i,j}$. *Then there is a partition* $\mathbf{P} = \{P_i\}_{i \in \mathbf{N}}$ *of* M *such that for all* $i, j \in \mathbf{N}$,

$$(*) \qquad \mu\left(P_i \cap \sigma^{-1} P_j\right) = \mu(P_i) p_{i,j} = \pi_i p_{i,j}.$$

If we consider *finite* state spaces, rather than denumerable ones, then the following is known. Theorem 1 was obtained independently by Alpern (1979) and Kieffer (1980). The connection with coding, as described above, was established by Kieffer (1980) and the interested reader is referred there for further details. As observed by Kieffer for finite state spaces and ergodic σ, Theorem 1 is a consequence of a deep result of Grillenberger and Krengel (1976). Kieffer also proved a "universal" version of Theorem 1 in which there is one partition $\mathbf{P}$ which satisfies $(*)$ for all antiperiodic σ-invariant measures μ. Cohen (1981) considered the following variation of Theorem 1. The $(p_{i,j}: 1 \le i, j \le n)$ are given and it is required to find a circle rotation σ and a circle partition $\mathbf{P}$ consisting of intervals satisfying $(*)$ [see also Alpern (1983) and Haigh (1985)].

In addition to covering the denumerable case, our proof of Theorem 1 also simplifies the finite case. Our method of coding is based on return times to some fixed state.

2. Proof of Theorem 1. Theorem 1 will be an easy consequence of the following result in ergodic theory due to Alpern (1981). The ideas used in the proof of Theorem 1 are illustrated by an example in Section 3 which the reader may wish to consult.

APPROXIMATE CONJUGACY THEOREM [Theorem 4 of Alpern (1981)]. *Let* τ *and* σ *be bimeasurable measure preserving automorphisms of the Lebesgue probability spaces* $(M_1, \mathscr{A}_1, \mu_1)$ *and* $(M_2, \mathscr{A}_2, \mu_2)$, *respectively. Assume* τ *is weakly mixing and* σ *is antiperiodic. Let* $F \in \mathscr{A}_1$ *with* $\mu_1(F) < 1$ *be given. Then there is a bimeasurable bijection*

$$\psi: (M_1, \mathscr{A}_1, \mu_1) \to (M_2, \mathscr{A}_2, \mu_2)$$

with $\mu_2 \psi = \mu_1$ *and such that* $\hat{\sigma}(x) = \psi^{-1} \sigma \psi(x) = \tau(x)$ *for* μ_1 *a.e.* $x \in F$.

PROOF OF THEOREM 1. The Markov chain $p_{i,j}$ may be represented in the usual way as a dynamical system $(\mathbf{N}^\infty, \mathscr{B}, \nu, \tau)$. Here $\mathbf{N}^\infty$ is the set of bilateral sequences $\omega = (\omega_n)_{n \in \mathbf{Z}}$ with $\omega_n \in \mathbf{N}$, $\mathscr{B}$ is the product σ-field generated by the subsets of $\mathbf{N}$ and τ is the left shift given by $\omega' = \tau\omega$ where $\omega'_n = \omega_{n+1}$ for all n. Let $\mathbf{Q} = \{Q_i\}_{i \in \mathbf{N}}$ be the time 0 partition of $\mathbf{N}^\infty$ given by

$$Q_i = \{\omega: \omega_0 = i\}.$$

The measure ν is defined on cylinder sets by

$$\nu\left(Q_{i_0} \cap \tau^{-1}Q_{i_1} \cap \cdots \cap \tau^{-n}Q_{i_n}\right) = \pi_{i_0} \prod_{k=1}^{n} p_{i_{k-1}, i_k}.$$

It is well known that if the Markov chain ($p_{i,j}$: $i, j \in \mathbf{N}$) is positive recurrent, irreducible and aperiodic, then $(\mathbf{N}^\infty, \mathscr{B}, \nu, \tau)$ is mixing. Hence we may apply the approximate conjugacy theorem to the measure preserving system $(\mathbf{N}^\infty, \mathscr{B}, \nu, \tau)$ and $F = \sim \tau^{-1}Q_0 = \{\omega\colon\ \omega_1 \neq 0\}$, where 0 is some index in $\mathbf{N}$. By the approximate conjugacy theorem there is an isomorphism,

$$\psi\colon (\mathbf{N}^\infty, \mathscr{B}, \nu) \to (M, \mathscr{A}, \mu)$$

with $\mu\psi = \nu$, and such that $\psi^{-1}\sigma\psi = \tau$ on F. We claim that the partition $\mathbf{P} = \{P_i\}_{i \in \mathbf{N}}$ defined by

$$P_i = \psi Q_i$$

satisfies the requirements of Theorem 1. Now fix any $i, j \in \mathbf{N}$ with $j \neq 0$. We will show that condition (*) of Theorem 1 holds for such pairs. It will then follow trivially that (*) also holds when $j = 0$. Since $j \neq 0$, $\tau^{-1}Q_j \subset F$, so

$$\begin{aligned} \psi^{-1}\sigma\psi(\omega) &= \tau(\omega) && \text{for all } \omega \in \tau^{-1}Q_j, \\ \psi^{-1}\sigma^{-1}\psi(\omega) &= \tau^{-1}(\omega) && \text{for all } \omega \in Q_j, \\ \sigma^{-1}(\zeta) &= \psi\tau^{-1}\psi^{-1}(\zeta) && \text{for all } \zeta \in \psi Q_j = P_j. \end{aligned}$$

Therefore, $\sigma^{-1}P_j = \psi\tau^{-1}\psi^{-1}P_j$. Thus

$$\begin{aligned} \mu\left(P_i \cap \sigma^{-1}P_j\right) &= \mu\left(P_i \cap \psi\tau^{-1}\psi^{-1}P_j\right) \\ &= \mu\left(\psi Q_i \cap \psi\tau^{-1}Q_j\right) \\ &= \mu\psi\left(Q_i \cap \tau^{-1}Q_j\right) \\ &= \nu\left(Q_i \cap \tau^{-1}Q_j\right) \\ &= \pi_i p_{i,j}. \end{aligned}$$

Actually the process $(M, \mathscr{A}, \mu, \mathbf{P})$ [this is the process which lists for each time n, and each $\omega \in M$ the subscript of the element of $\mathbf{P}$ to which $\sigma^n(\omega)$ belongs] assigns the correct Markov measure to all words without a 0 in the interior. Thus we have that the distribution of $\mathbf{P} \vee \sigma^{-1}\mathbf{P}$ is what the Markov chain prescribes. This completes the proof of Theorem 1. □

3. Example. We include for the reader unfamiliar with ergodic theory the following example which illustrates the technique used in the proof. Consider the following simple Markov chain on three states $\{0, 1, 2\}$ and transition

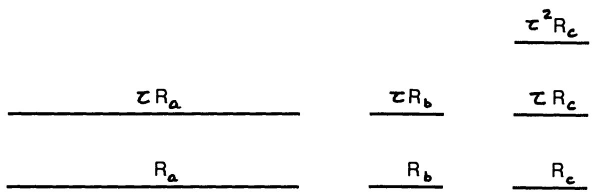

FIG. 1. *Names. The transformation acts by mapping each point to the point directly above it; otherwise it is mapped to some point on the bottom level.*

probabilities given by the matrix

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix} \end{array}.$$

The stationary invariant measure is given by $\pi_0 = 6/13$, $\pi_1 = 2/13$, $\pi_2 = 5/13$. Suppose we start from the state 0 and consider all possible future paths (and their lengths) from 0 until the first return back to the state 0. Since the Markov chain is aperiodic, the lengths of these paths form a relatively prime set of integers. In this case there are exactly three allowable paths (words) between successive 0's, given by two paths of length 2, $a = 020$, $b = 010$, and one of length 3, given by $c = 0120$. The set $Q_0 = \{\omega \in \mathbf{N}^\infty \colon \omega_0 = 0\}$ is partitioned into three sets,

$$R_a = \{\omega \in \mathbf{N}^\infty \colon \omega_0 = 0, \omega_1 = 2, \omega_2 = 0\},$$
$$R_b = \{\omega \in \mathbf{N}^\infty \colon \omega_0 = 0, \omega_1 = 1, \omega_2 = 0\},$$
$$R_c = \{\omega \in \mathbf{N}^\infty \colon \omega_0 = 0, \omega_1 = 1, \omega_2 = 2, \omega_3 = 0\}.$$

The set F given in the previous proof consists of the sets $R_a \cup R_b \cup R_c \cup \tau R_c$.

Consider the three diagrams (Figures 1–3), all of the same tower picture for τ, which each illustrate different information about the return times to the set Q_0.

The set Q_1 is the union of all sets with the label 1. In this case it consists of the two sets τR_b, τR_c. The set $Q_2 = \tau R_a \cup \tau^2 R_c$, i.e., all sets labelled 2 belong to

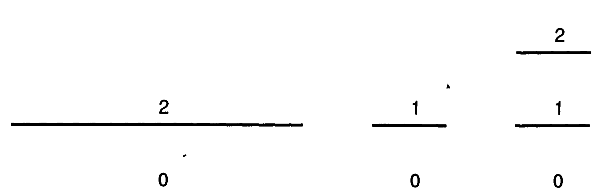

FIG. 2. *Labels. Each level of this tower consists of points with the same* 0*th coordinate. The label is that common value.*

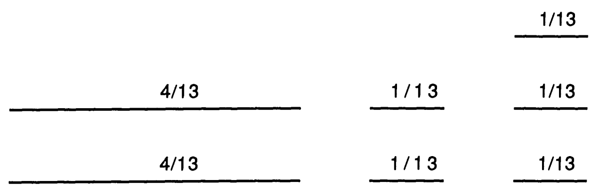

FIG. 3. *Measures. The measures of the various levels of the tower.*

Q_2. Note that τ satisfies $(*)$ with the partition $\mathbf{Q} = \{Q_0, Q_1, Q_2\}$. For example, $\nu\{\omega\colon \omega_{n+1} = 2|\omega_n = 1\} = \nu(\tau R_c)/(\nu(\tau R_c) + \nu(\tau R_b)) = \frac{1}{2}$.

Any transformation agreeing with τ on F will also satisfy $(*)$ with the same partition. Actually all we need to know is that the transformation σ has the same distribution of return times to some set B (called the base of the tower) as τ does to the set Q_0 (which is the base of the tower for the shift τ). Once we have such a tower for σ, we can subdivide the column of height 2 for the σ tower into two columns, in the same way as for the τ tower. Then the various levels of the σ tower are assigned labels in the same way as the labels occur in the τ tower.

The measure of the two subcolumns of height 2 for the σ tower is determined by looking at the corresponding columns in the tower for τ.

This will yield a partition for σ satisfying $(*)$. In this example we had only a finite number of return times to the set Q_0. Even when there may be an infinite number of different return times, the following theorem (which implies the approximate conjugacy theorem) guarantees that we can find a set having those return times with a prescribed distribution.

MULTIPLE TOWER THEOREM [Theorem 1 of Alpern (1981)]. *Let σ be an antiperiodic bimeasurable μ-preserving automorphism of $(M, \mathscr{A}, \mu)$ and let $\phi = (\phi_k)$, $k = 1, 2, \ldots, \infty$, be a probability vector such that the k's with $\phi_k > 0$ are relatively prime. Then there is a partition $\{P_{k,i}\}$, $k = 1, 2, \ldots, \infty$, $i = 1, 2, \ldots, k$, of M satisfying* (i) $P_{k,i} = \sigma^{i-1}(P_{k,1})$ *and* (ii) $\mu(\bigcup_{i=1}^{k} P_{k,i}) = \phi_k$.

Note that we would apply this theorem to the numbers

$\phi_k =$ Prob(the Markov chain returns to 0 in exactly k steps|the chain starts at 0).

The set $B = \bigcup_{k=1}^{\infty} P_{k,1}$ is the base of the tower for σ and the distribution of return times to B is the same as the distribution of return times of the Markov chain to the state 0.

Note added in proof. We have recently shown [Alpern and Prasad (1989)] that if instead of assuming that σ preserves μ in Theorem 1, one assumes only that σ preserves μ null sets, then for any $\varepsilon > 0$ the weaker conclusion that

$$1 - \varepsilon < \frac{\mu\left(P_i \cap \sigma^{-1} P_j\right)}{\mu(P_i) p_{i,j}} < 1 + \varepsilon$$

holds for all i, j.

REFERENCES

ALPERN, S. (1979). Generic properties of measure preserving homeomorphisms. *Ergodic Theory, Proceedings, Oberwolfach 1978. Lecture Notes in Math.* **729** 16–27. Springer, New York.

ALPERN, S. (1981). Return times and conjugates of an antiperiodic transformation. *Ergodic Theory Dynamical Systems* **1** 135–143.

ALPERN, S. (1983). Rotational representations of stochastic matrices. *Ann. Probab.* **11** 789–794.

ALPERN, S. and PRASAD, V. S. (1989). Return times for nonsingular measurable transformations. Unpublished.

COHEN, J. E. (1981). A geometric representation of stochastic matrices; theorem and conjecture. *Ann. Probab.* **9** 899–901.

GRILLENBERGER, C. and KRENGEL, U. (1976). On marginal distributions and isomorphisms of stationary processes. *Math. Z.* **149** 131–154.

HAIGH, J. (1985). Rotational representations of stochastic matrices. *Ann. Probab.* **13** 1024–1027.

KIEFFER, J. C. (1980). On coding a stationary process to achieve a given marginal distribution. *Ann. Probab.* **8** 131–141.

DEPARTMENT OF MATHEMATICS
LONDON SCHOOL OF ECONOMICS
HOUGHTON STREET
LONDON WC2A 2AE
ENGLAND

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LOWELL
LOWELL, MASSACHUSETTS 01854