The Annals of Probability
1983, Vol. 11, No. 3, 789–794

ROTATIONAL REPRESENTATIONS OF STOCHASTIC MATRICES

BY STEVE ALPERN

London School of Economics

Let $\{S_i\}$, $i = 1, \cdots, n$, be a partition of the circle into sets S_i each consisting of a finite union of arcs. Let f be a rotation of the circle and let u denote Lebesgue measure. Then the matrix P defined by $p_{ij} = u(S_i \cap f^{-1}S_j)/u(S_i)$ is stochastic. We prove (and improve) a conjecture of Joel E. Cohen asserting that every irreducible stochastic matrix arises from a construction of this type.

1. Introduction. In a recent note Cohen [2] has suggested the following geometric construction as a means of representing stochastic matrices. (A matrix is stochastic if it is square, nonnegative, and all its row sums are one.)

Let u denote Lebesgue measure on $X = [0, 1)$. Let $S = \{S_i\}_{i=1}^{n}$ be a partition of X into nonnull sets S_i which are each finite unions of intervals. Let $f = f_t$ be the u-preserving transformation of X onto itself defined by $f(x) = x + t \pmod 1$. This construction defines a stochastic matrix $P = (p_{ij})$ by

$$p_{ij} = u(S_i \cap f^{-1}(S_j))/u(S_i). \tag{1}$$

In such a case we will say that the matrix P has a rotational representation, or that P is represented by (t, S). Cohen has shown that every 2×2 irreducible stochastic matrix has a rotational representation. (A stochastic matrix P is irreducible if for any row i and column $j \neq i$ there is a positive integer k, which may depend on i and j, such that the i, j element of P^k is not zero. If k doesn't depend on i and j then P is called primitive.) Cohen conjectured that every irreducible $n \times n$ stochastic matrix, $n > 2$, has a rotational representation. In this paper we will prove (and improve) Cohen's conjecture.

It is easy to see that not every matrix P with a rotational representation is irreducible. The simplest counter example is the identity matrix, which is represented by $(0, S)$ for any partition S. However there is a simple property which P must have, namely recurrence. We will say that P is recurrent if it is stochastic and has a (strictly) positive invariant distribution, i.e., if there is a probability vector $v > 0$ satisfying $vP = v$. It is known that P is recurrent if and only if it has no transient states (see Doob [3], page 183). If P is represented by (t, S) then $(u(S_1), \cdots, u(S_n))$ is invariant by (1) and positive by assumption on S, so P is recurrent. Our main result shows that the converse of this statement is also true.

THEOREM 1. *A stochastic matrix is recurrent if and only if it has a rotational representation. Furthermore suppose that P is an $n \times n$ recurrent matrix and v is any positive invariant distribution. Then P is represented by (t, S) where*
(i) $(u(S_1), \cdots, u(S_n)) = v$ *and*
(ii) $t = 1/n!$

Theorem 1 complements the following special case of a previous result of the author [1]: Every primitive stochastic matrix can be represented as in (1) for any given irrational t and some partition S—not necessarily composed of unions of intervals.

The proof of Theorem 1 is based on a constructive finite algorithm for finding (t, S) given P and v. For expository reasons we will precede the proof by an illustration of the algorithm, as applied to a specific 3×3 matrix. This example will be treated in Section

Received March 1982.
AMS 1980 *subject classification.* Primary 15A51; secondary 28A65, 60J10.
Key words and phrases. Measure-preserving transformations, ergodic theory, mapping of the unit interval, Markov chain.

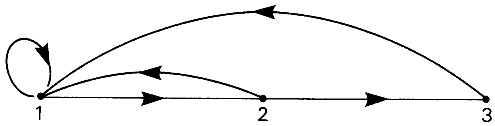

FIG. 1

two. Section three is devoted to proving a result (Lemma 1) which guarantees that a crucial construction of the algorithm can always be carried out. This result is perhaps well known but I know of no reference. In section four we complete the proof of Theorem 1. A final section (five) is then devoted to an analysis of the complexity of rotational representations, in the following sense. Theorem 1 guarantees the existence of a least number $b = b(n)$ such that every $n \times n$ recurrent matrix can be represented by (t, S) where each S_i has no more than b component intervals. In addition to the conjecture mentioned above, Cohen [2] also conjectured that $b(n) = n - 1$. This conjecture is false. We cannot determine any $b(n)$ exactly, although we can establish (Theorem 2) that $\exp(\alpha n^{1/2}) \le b(n) \le \exp(\beta n)$ for some positive constants α and β.

I wish to thank my colleague Nick Logothetis for several useful discussions on these matters.

2. An example of the algorithm. In this section we will compute a rotational representation for the matrix

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}$$

with invariant distribution $v = (4/7, 2/7, 1/7)$. If P is represented by (t, S) with $(u(S_1), u(S_2), u(S_3)) = v$ then the distribution of the partition $\{S_{ij}\}$, $i, j = 1, \cdots, n$ defined by $S_{ij} = S_i \cap f_t^{-1}(S_j)$ will be given by $r_{ij} = u(S_{ij}) = p_{ij}u(S_i) = p_{ij}v_i$. In our example this defines a matrix $R = (r_{ij})$ by

$$R = \begin{bmatrix} 2/7 & 2/7 & 0 \\ 1/7 & 0 & 1/7 \\ 1/7 & 0 & 0 \end{bmatrix}.$$

R induces the following directed graph on three vertices as shown in Figure 1.

Intuititively this graph is composed of three (not distinct) cycles, [1], [1, 2], and [1, 2, 3]. This intuition motivates the following decomposition of R as a convex combination of three "cycle" matrices, corresponding to [1], [1, 2] and [1, 2, 3].

$$R = 2/7 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + 2/7 \begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + 3/7 \begin{bmatrix} 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \\ 1/3 & 0 & 0 \end{bmatrix}. \tag{2}$$

The general existence of such a decomposition will be proved in the next section. Let M be a common multiple of the cycle lengths and let $t = 1/M$. In our example the cycle lengths are 1, 2, 3 so we can take $M = 6$ and $t = 1/6$. Partition the interval $[0, t)$ into three subintervals A_1, A_2, A_3, with relative distribution given by the coefficients of (2). In our example we have $A_1 = [0, 2/42)$, $A_2 = [2/42, 4/42)$ and $A_3 = [4/42, 1/6)$. For $k = 1, \cdots, 3$ and $l = 1, \cdots, 6$ define intervals $A_{kl} = A_k + (l - 1)/6$ which partition $[0, 1)$ as drawn below.

$$|A_{11}|A_{21}|A_{31}|A_{12}|A_{22}|A_{32}|A_{13}|A_{23}|A_{33}|A_{14}|A_{24}|A_{34}|A_{15}|A_{25}|A_{35}|A_{16}|A_{26}|A_{36}|$$

This partition (A_{kl}) is more easily grasped in "stacked" form, in which $f(x) = x + 1/6 \pmod 1$ lies immediately above x, except for the top level, which is mapped five levels down to the bottom, as shown in Diagram 1.

Diagram 1.

	A_{25}	
A_{14}		
		A_{32}
A_{11}	A_{21}	A_{31}

As can be seen from the diagram, the "columns" defined by $U_k = \cup_{l=1}^{6} A_{kl}$, $k = 1, \cdots, 3$ are invariant under f. We now define the partition $S = (S_1, S_2, S_3)$ by assigning one of the labels 1, 2, 3 to each interval A_{kL} and letting S_i be the union of all intervals with label i. We do the labeling separately on each invariant set U_k. Let's start with U_2, which corresponds to the cycle [1, 2]. Starting with either 1 or 2 (we'll take 2) on the bottom level (A_{21}) of U_2 we label all the levels of U_2 as in Diagram 2.

Diagram 2.

1
2
1
2
1
2

Similarly we label the levels of U_1 and U_3 according to the cycles [1] and [1, 2, 3] respectively, obtaining Diagram 3.

Diagram 3.

1	1	3
1	2	2
1	1	1
1	2	3
1	1	2
1	2	1

We now demonstrate that if S is the partition given by Diagram 3 then the distribution of the partition $\{S_{ij}\}$ is given by the matrix R. Observe that an interval A_{kl} belongs to S_{ij} if and only if it has label i and the interval above it (or five below it, if it's on top) has label j. By construction, the relative (conditional) distribution of $\{S_{ij}\}$ on U_1 (the left column) is given by

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Similarly the relative distribution of $\{S_{ij}\}$ on U_2 and U_3 is given by

$$\begin{bmatrix} 0 & ½ & 0 \\ ½ & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \\ 1/3 & 0 & 0 \end{bmatrix}$$

respectively. Since by construction we have $u(U_1) = 2/7$, $u(U_2) = 2/7$, and $u(U_3) = 3/7$, the absolute distribution of $\{S_{ij}\}$ is given by R because of (2). It follows that $p_{ij} = r_{ij}/v_i$ is given by (1), so that P is represented by $(1/6, S)$ with $(u(S_1), u(S_2), u(S_3)) = v$.

3. Decomposition into cycles. In this section we give the theoretical justification of line (2) of the algorithm presented in the last section. Observe that the matrix R of (2) satisfies the following two properties.

(3) $$\sum_{i=1}^{n} r_{ik} = \sum_{j=1}^{n} r_{kj} \quad \text{for each} \quad k = 1, \cdots, n$$

(4) $$\sum_{i,j=1}^{n} r_{ij} = 1.$$

We will call any $n \times n$ nonnegative matrix satisfying (3) and (4) "equi-summed". An important class of equi-summed matrices is the class of "cycle" matrices, as illustrated by the three matrices on the right hand side in (2). If $[a_1, a_2, \cdots, a_m]$ is a sequence of distinct integers chosen from $1, \cdots, n$, then we define the cycle matrix of $[a_1, \cdots, a_m]$ as the $n \times n$ matrix C given by $c_{a_1a_2} = c_{a_2a_3} = \cdots = c_{a_{m-1}a_m} = c_{a_ma_1} = 1/m$ and $c_{ij} = 0$ otherwise. We say that m is the length of C.

LEMMA 1. *Every equi-summed matrix is a convex combination of cycle matrices. Furthermore, if R is any equisummed $n \times n$ matrix then for some N, $N \leq n^2 - n + 1$, some probability vector $(z_1, \cdots, z_N)$ and some cycle matrices C^k, $k = 1, \cdots, N$, we have*

$$R = \sum_{k=1}^{N} z_k C^k.$$

PROOF. A short proof may be based on the observation that the extreme points of the compact convex set of equi-summed matrices are the cycle matrices. Since the set of $n \times n$ equi-summed matrices has dimension $n^2 - n$ (one may specify all but the last row) the estimate on N follows from Carathéodory's extension of Minkowksi's result on extreme points.

We prefer however to give a combinatorial proof which constructs the z_k and C^k (and incidentally verifies the above assertion about extreme points). For this proof it is necessary to work with the larger class of matrices which satisfy (3)—but not necessarily (4). For the proof we will call such matrices "normal" and consider the following sublemma: "A normal matrix with no more than w nonzero entries is a nonnegative combination of cycle matrices." We prove the sublemma by induction on w. If w is zero then the result is obvious since we may take $z_1 = 0$ and any cycle C^1. Suppose the sublemma is true for w and let R be a normal matrix with $w + 1$ nonzero entries. Since $w + 1$ is at least one, there must exist indices a_1 and a_2 such that $r_{a_1a_2}$ is positive. By property (3), or the normality of R, there is an a_3 such that $r_{a_2a_3}$ is positive. (If not, the a_2 column has positive sum while the a_2 row has zero sum, contradicting (3) for $k = a_2$.) Continuing in this manner we obtain a sequence $a_1, a_2, \cdots$ with $r_{a_ia_{i+1}}$ positive. Since the a_i are chosen from a finite set there must be repetitions. Renumber the a_i so that $a_1, \cdots a_m$ are distinct and $a_{m+1} = a_1$. Let C be the cycle matrix of $[a_1, \cdots, a_m]$ and let $z = m.\min\{r_{ij}: c_{ij} = 1/m\}$. Then $R' = R - zC$ is a normal matrix with no more that w nonzero entries. Hence by the inductive hypothesis R' is a nonnegative combination of cycle matrices, and so is $R = R' + zC$. The sublemma is thus established for all w. Now suppose R is equi-summed, so that (4) is satisfied also. Since R is in particular normal, we can write $R = \sum z_k C^k$ with nonnegative z_k. However (4) ensures that the z_k sum to one.

4. Proof of Theorem 1. Let P and v be as stated in the Theorem. Define $R = (r_{ij})$ by $r_{ij} = p_{ij}v_i$. Observe that $\sum_{i=1}^{k} r_{ik} = \sum_{i=1}^{n} v_i p_{ik} = v_k$ because $vP = v$ by assumption. Also observe that $\sum_{j=1}^{n} r_{kj} = \sum_{j=1}^{n} v_k p_{kj} = v_k \sum_{j=1}^{n} p_{kj} = v_k$ because P is stochastic. Thus R satisfies (3). Next compute $\sum_{i,j=1}^{n} r_{ij} = \sum_{i,j=1}^{n} v_i p_{ij} = \sum_{i=1}^{n} v_i \sum_{j=1}^{n} p_{ij} = 1$, demonstrating that

R satisfies (4). Hence R is equi-summed and Lemma 1 ensures that we can write

$$R = \sum_{k=1}^{N} z_k C^k \tag{5}$$

where z is a probability vector and the C^k are cycle matrices. Let m_k denote the length of the cycle matrix C_k and let M be any multiple of $m_1, m_2, \cdots, m_N$. In particular we may take $M = n!$ (since $m_k \leq n$) or $M =$ l.c.m. $(m_1, \cdots, m_N)$, where l.c.m. stands for least common multiple. Let $t = 1/M$ and let f denote f_t. Let $\{A_k\}$, $k = 1, \cdots, N$ be a partition of $A = [0, 1/M)$ into N subintervals with relative distribution $(z_1, \cdots, z_N)$. That is, $u(A_k)/u(A) = z_k$. Define $A_{kl} = f^{l-1}(A_k)$ and $U_k = \cup_{l=1}^{M} A_{kl}$ for $k = 1, \cdots, N$ and $l = 1, \cdots, M$. We now define the partition $S = \{S_i\}_{i=1}^{n}$ by $S_i = \cup_{h(k,l)=i} A_{kl}$ where h is the following labeling of the intervals A_{kl}. Fix k and suppose C^k is the cycle matrix based on $[a_1, \cdots, a_m]$, where $m = m_k$. Define $h(k, 1) = a_1, h(k, 2) = a_2, \cdots, h(k, m) = a_m, h(k, m+1) = a_1, \cdots, h(k, M) = a_m$. The fact that the last label is a_m follows from the choice of M as a multiple of m. This process defines S on each invariant set U_k so that $u(S_{ij} \cap U_k)/u(U_k) = c_{ij}^k$ where $S_{ij} = S_i \cap f^{-1}(S_j)$ and c_{ij}^k *is the* i, j entry of C^k. The measure of U_k is z_k, because $u(A_{kl}) = u(A_k)$, $u(A_k)/1/M) = z_k$, and $u(U_k) = Mu(A_{kl})$. Finally we compute

$$u(S_{ij}) = \sum_{k=1}^{N} u(U_k) u(S_{ij}/U_k) = \sum_{k=1}^{N} z_k c_{ij}^k = r_{ij}$$

by (5), and

$$u(S_i \cap f^{-1}(S_j))/u(S_i) = u(S_{ij})/u(S_i) = r_{ij}/v_i = p_{ij}. \tag{6}$$

Thus we have shown that $(1/M, S)$ is a rotational representation of P with $u(S_i) = \sum_{j=1}^{n} r_{ij} = v_i$, and that we may choose $M = n!$.

5. Complexity of the representation. We say that a partition $S = \{S_i\}_{i=1}^{n}$ has type L if the number of components of S_i is less than or equal to L, $i = 1, \cdots, n$. Let $b = b(n)$ be the least integer such that every $n \times n$ recurrent matrix has a representation of type b, that is, a representation (t, S) where S is of type b. To obtain a lower bound on $b(n)$ we will need the following:

LEMMA 2. *Let c_k, $k = 1, \cdots, r$ be positive integers and let $n = 1 + c_1 + c_2 + \cdots + c_r$. Let $Q = Q(c_1, \cdots, c_r)$ be an $n \times n$ permutation matrix with cycles of lengths* $1, c_1, c_2, \cdots, c_r$. *Then if Q is represented by (t, S) the type of S is at least* l.c.m. $(c_1, \cdots, c_r)$.

PROOF. Let 1 be the label of the 1-cycle of Q so that $q_{11} = 1$. Then if f denotes f_t the set S_1 is invariant under f. It follows from Weyl's well-known result (see Halmos [4]) that t is rational (irrational rotations are ergodic—have no nontrivial invariant sets). Let $t = p/q$ in lowest terms, so that every point in $[0, 1)$ has f-period q. The invariant set S_1 consequently consists of at least q intervals and hence the type of S is at least q. To estimate q from below, observe that if a point x belongs to S_i where the index i belongs to a Q-cycle of length c_k, then the f-period of x must be a multiple of c_k. But the f-period of every x is q, so q is a multiple of c_k. Hence $q \geq$ l.c.m.$(c_1, \cdots, c_r)$.

THEOREM 2. *There exist positive constants α and β such that for all n,* $\exp(\alpha n^{1/2}) < b(n) < \exp(\beta n)$.

PROOF. We shall need the following notation and estimates ((7) and (8)) which can be found on pages 89–91 of Landau [5]. Let p_k denote the kth prime ($p_1 = 2$) and let $\pi(n)$ denote the number of primes less than or equal to n. A partial result in the direction of the famous Prime Number Theorem (due to Chebyshev) asserts the existence of a positive constant β_1 such that

$$\pi(n) \leq \beta_1/\log n. \tag{7}$$

Let $d(k, n)$ be the largest integer power d such that $p_k^d \leq n$. Then for any even n,

(8) $$2^{n/2} \leq \text{l.c.m.}(2, 3, \cdots, n) = \prod_{k=1}^{\pi(n)} p_k^{d(k,n)} \leq n^{\pi(n)}.$$

We can now proceed with the proof proper, beginning with the upper bound. The algorithm presented in the proof of Theorem 1 represents any recurrent $n \times n$ matrix by (t, S) where S is composed of intervals A_{kl}, $k = 1, \cdots, N$, $l = 1, \cdots, M$, where $N \leq n^2 - n + 1$ and $M = \text{l.c.m.}(1, 2, 3, \cdots, n)$. Consequently we have that

(9) $$b(n) \leq NM \leq n^2 \text{l.c.m.}(1, 2, \cdots, n).$$

If we combine (9) with (8) and take logs we get,

(10) $$\begin{aligned} \log b(n) &\leq 2 \log n + \pi(n) \log n \\ &\leq \log n \, (2 + \beta_1 n/\log n) \quad &&\text{(by (7))} \\ &\leq \log n \, (\beta n/\log n) &&\text{for some } \beta > \beta_1 \\ &= \beta n &&\text{yielding the upper bound.} \end{aligned}$$

To obtain the lower bound, fix any even m and define $c_k = p_k^{d(k,m)}$ for $k = 1, \cdots, \pi(m)$. Let $n = n_m = 1 + \sum_{k=1}^{\pi(m)} c_k \leq m^2$. Then apply Lemma 2 to the permutation matrix $Q = Q(c_1, \cdots, c_{\pi(m)})$, obtaining

(11) $$b(n_m) \geq \text{l.c.m.}(c_1, \cdots, c_{\pi(m)}) = \prod_{k=1}^{\pi(m)} p_k^{d(k,m)} \geq 2^{m/2} \quad \text{by (8).}$$

Since $n_m \leq m^2$ and $b(n)$ *is nondecreasing,* (11) *implies*

(12) $$b(m^2) \geq 2^{m/2} \quad \text{and hence}$$

(13) $$b(m) \geq 2^{m^{1/2}/2} \quad \text{or}$$

(14) $$b(m) \geq \exp(\alpha m^{1/2}) \quad \text{where} \quad \alpha = \log 2^{1/2}.$$

REFERENCES

[1] ALPERN, S. (1979). Generic properties of measure preserving homeomorphisms. Ergodic Theory, Proceedings, Oberwohlfach 1978. *Springer Lecture Notes in Mathematics* **729** 16–27.

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

[3] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[4] HALMOS, P. (1956). *Lectures on Ergodic Theory.* Chelsea, New York.

[5] LANDAU, E. (1958). *Elementary Number Theory.* Chelsea, New York.

THE LONDON SCHOOL OF ECONOMICS
AND POLITICAL SCIENCE
UNIVERSITY OF LONDON
HOUGHTON STREET
LONDON WC2A 2AE
ENGLAND