The Annals of Probability
1981, Vol. 9, No. 5, 899–901

A GEOMETRIC REPRESENTATION OF A STOCHASTIC MATRIX: THEOREM AND CONJECTURE[1]

BY JOEL E. COHEN

The Rockefeller University

An irreducible stochastic matrix may be constructed by partitioning a line of unit length into a finite number of intervals, shifting the line to the right (mod 1) by a small amount, and defining transition probabilities in ters of the overlaps among the intervals before and after the shift. It is proved that every 2×2 irreducible stochastic matrix arises from this construction. Does every $n \times n$ irreducible stochastic matrix arise this way?

A stochastic matrix P is an $n \times n$ matrix with nonnegative real elements p_{ij} such that every row sum is 1. We assume $1 < n < \infty$. We shall describe a simple geometric construction, based on partitioning and mapping the unit interval, that produces an irreducible stochastic matrix. A matrix P is irreducible if and only if, for any row i and any column $j \neq i$, there exists a positive integer k, which may depend on i and j, such that the i, j element of P^k is not zero. Whenever such a k exists, it may be chosen to be less than n. We shall show that every irreducible 2×2 stochastic matrix arises from such a construction. We have neither a proof nor a counterexample to the conjecture that every irreducible $n \times n$ stochastic matrix, $n > 2$, also arises from this construction. We offer the conjecture as an open problem.

A theorem of Alpern (1978) gives a geometric representation of every primitive stochastic matrix. A stochastic matrix P is primitive if, for some positive integer k, every element of P^k is positive. Our conjectured geometric representation would simplify certain special cases of his theorem, which we describe in more detail later.

In addition, our conjecture, if true, would reveal an attractively simple structure in irreducible stochastic matrices.

We now describe a geometric construction that produces an irreducible stochastic matrix.

Let X be the unit interval $[0, 1)$. Partition X into n sets $J_1, \cdots, J_n$ each of positive Lebesgue measure. Suppose each J_i is the union of $n - 1$ intervals $I_{ik} = [x_{ik}^L, x_{ik}^R)$, where $[a, b) = \emptyset$ if $b \leq a$. The superscripts ^{L} and ^{R} identify the left and right endpoints. We will explain later why we choose $n - 1$ intervals, rather than some other number, to compose each J_i. In case $n = 2$, $J_i = I_{i1}$, $i = 1, 2$ and we may take $J_1 = [0, x_1)$, $J_2 = [x_1, 1)$ for some x_1 in $(0, 1)$. For any t and x in X, let

$$f_t(x) = (x + t)(\text{mod } 1),$$

that is, $f_t(x)$ is the fractional part of $x + t$. f_t preserves Lebesgue measure, which we denote by m: $m(I_{ik}) = x_{ik}^R - x_{ik}^L$.

Define

$$p_{ij} = m(f_t(J_i) \cap J_j)/m(J_i). \tag{1}$$

Since f_t is measure-preserving, (1) is the same as

Received June 9, 1980; revised September 16, 1980.

[1] This work was partially supported by N.S.F. grant DEB 80-11026.

AMS 1970 *subject classifications.* Primary 15A51; secondary 28A65, 60J10.

Key words and phrases. Measure-preserving transformation, ergodic theory, mapping of the unit interval, Markov chain.

$$p_{ij} = m(f_t(J_i) \cap J_j)/m(f_t(J_i)).$$

Since $m \geq 0$ and $m(J_i) > 0$ by assumption, p_{ij} is well defined and non-negative. Also $\sum_{j=1}^{n} p_{ij} = m(f_t(J_i) \cap [\cup_j J_j])/m(J_i) = m(f_t(J_i))/m(J_i) = 1$. So $P = (p_{ij})$ is stochastic. In case $n = 2$ and $0 < t < 1$, P is irreducible.

Let π be a row n-vector with $\pi_i = m(J_i)$, $i = 1, \cdots, n$. Then the jth element of πP is $\Sigma_i \pi_i p_{ij} = \Sigma_i m(f_t(J_i) \cap J_j) = m(J_j) = \pi_j$. Hence $\pi P = \pi$. Thus we can construct a stochastic matrix P with an arbitrary invariant distribution $\pi > 0$ by choosing J_i such that $\pi_i = m(J_i)$.

P can be chosen arbitrarily close to the identity matrix $I = (\delta_{ij})$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$, because $\lim_{t \to 0} m(f_t(J_i) \cap J_j) = \delta_{ij} m(J_i)$.

Now let M be an irreducible 2×2 stochastic matrix. M is irreducible if and only if both elements off the main diagonal are not zero. There exists a positive row vector v such that $vM = v$ (Seneta, 1973). Assume $v_1 + v_2 = 1$. It may be checked that

$$v = (m_{21}/(m_{12} + m_{21}),\ m_{12}/(m_{12} + m_{21})). \tag{2}$$

We will show how to find J_1, J_2 and $t > 0$ such that $m_{ij} = p_{ij}$ where p_{ij} is given by (1). Since v is the invariant distribution of M and of the desired P, it is natural, in the light of the above, to let $J_1 = [0, v_1)$, $J_2 = [v_1, 1)$.

Let

$$t = m_{12} m_{21}/(m_{12} + m_{21}). \tag{3}$$

Since M is irreducible, $t > 0$. Comparing (2) and (3) shows that $t \leq v_i$, $i = 1, 2$, because $m_{ij} \leq 1$, $i \neq j$.

Now $f_t(J_1) \cap J_1 = [t, v_1 + t) \cap [0, v_1) = [t, v_1)$ so $m(f_t(J_1) \cap J_1) = v_1 - t$ and by (1),

$$p_{11} = (v_1 - t)/v_1. \tag{4}$$

Substituting (2) and (3) into the right side of (4) yields $p_{11} = m_{11}$ as desired. It follows that $p_{12} = m_{12}$. Since $t > 0$, $p_{11} < 1$ and $p_{12} > 0$.

Similarly, $f_t(J_2) \cap J_2 = ([v_1 + t, 1) \cup [0, t)) \cap [v_1, 1) = [v_1 + t, 1) \cup [v_1, t)$. But $[v_1, t) = \emptyset$ because $t \leq v_1$. Thus $m(f_t(J_2) \cap J_2) = 1 - v_1 - t = v_2 - t < v_2$. By (1), using (2) and (3) as before,

$$p_{22} = (v_2 - t)/v_2 = m_{22}.$$

We have shown that for any irreducible stochastic 2×2 matrix M, there exists a representation of the form (1), with $0 < t < 1$. Since M and the representation (1) each have exactly 2 degrees of freedom (e.g., m_{12} and m_{21} in M and t and v_1 in (1)), the representation is unique, except possibly for a cyclic permutation of J_1 and J_2.

What happens if a 2×2 stochastic matrix M is reducible? Then at least one off-diagonal element, say m_{12}, is 0. If we seek a representation in the form (1), then $p_{12} = 0$ implies $t = 0$, however J_1 and J_2 are chosen. Now if $m_{21} \neq 0$, then no representation (1) is possible because $f_0(J_2) = J_2$ has no intersection with J_1. If $m_{21} = 0$, then a representation (1) exists for every partition of X into intervals J_1, J_2 of positive length because $P = M = I$. Thus a reducible stochastic matrix M has either no representation or an infinity of representations of the form (1).

To see why we suppose each J_i is the union of $n - 1$ intervals, consider the positive $n \times n$ matrix M with $m_{ii} = 1 - nd$, $m_{ij} = d$, $i, j = 1, \cdots, n$, $i \neq j$, for $0 < nd < 1$ and $n > 2$. Take d very small. Since the diagonal elements of M then approach 1, $t \pmod 1$ must also become very small. Now suppose J_1, say, consisted of $n - 2$ or fewer intervals I_{1k}. Then the $n - 2$ or fewer intervals that lie just to the right (mod 1) of each I_{1k} could belong to at most $n - 2$ different J_i, $i \neq 1$. Hence for small enough t, there would have to be at least one $p_{1k} = 0$, $k \neq 1$, according to (1). Thus if J_1 consists of fewer than $n - 1$ intervals, not every positive stochastic matrix can be represented by (1). We conjecture that J_i need contain no more than $n - 1$ intervals because we know no argument why more should be required.

For every irreducible stochastic matrix of order 3×3 or larger, does a representation in the form (1) exist? If a representation exists, under what conditions, if any, is it unique (up to cyclic permutations of the set of all I_{ik} in X)?

A special case of a result due to Alpern (1978, page 18) may be stated as follows. Let P be a primitive stochastic matrix. Let g be an invertible aperiodic m-preserving transformation of X. (For example, every f_t with t irrational is such a g.) Then there exists a partition $H_1, \cdots, H_n$ of X such that $m(g(H_i) \cap H_j)/m(H_i) = p_{ij}$. This result is stronger than our conjecture in guaranteeing that for every g, a partition $\{H_i\}$ exists that represents a given P. This result is weaker than our conjecture in assuming P primitive and in permitting H_i to be any measurable subset of X, rather than only a union of $n-1$ intervals like J_i.

The definition of $f_t(x)$ suggests a stochastic process that mimics the marginal frequencies and one-step transition probabilities of a stationary Markov chain specified by (1). Let U_0 be uniformly distributed on $[0, 1)$ and let, for every positive integer k, $U_k = (U_{k-1} + t)(\text{mod } 1)$. Let $Z_k = i$ if $U_k \in J_i$, $k = 0, 1, 2, \cdots$ Then $P[Z_k = i] = \pi_i$ and $P[Z_{k+1} = j \mid Z_k = i] = p_{ij}$, $k = 0, 1, \cdots$ A referee asks: is $\{Z_k\}$ Markovian? To see that the answer is no, in general, let

$$M = \begin{pmatrix} 1/5 & 4/5 \\ 2/5 & 3/5 \end{pmatrix},$$

with $\pi = (1/3, 2/3)$ and $t = 4/15$. Then $P[Z_0 = 1, Z_1 = 2] = 4/15$ and $P[Z_0 = 1, Z_1 = 2, Z_2 = 1] = 0$ so $P[Z_2 = 1 \mid Z_1 = 2, Z_0 = 1] = 0$. However, $P[Z_0 = 2, Z_1 = 2] = 6/15$ and $P[Z_0 = 2, Z_1 = 2, Z_2 = 1] = 4/15$ so $P[Z_2 = 1 \mid Z_1 = 2, Z_0 = 2] = 2/3$.

REFERENCES

[1] ALPERN, S. (1978). Generic properties of measure preserving homeomorphisms. M. Denker and K. Jacobs, eds., *Ergodic Theory*, Proceedings, Oberwolfach. *Lecture Notes in Mathematics* **729** 16–27. Springer, Berlin.

[2] SENETA, E. (1973). *Non-negative Matrices*. George Allen and Unwin, London, page 20.

THE ROCKEFELLER UNIVERSITY
1230 YORK AVENUE
NEW YORK, NEW YORK 10021