SYMMETRIZABLE MARKOV MATRICES

By H. P. Kramer

Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

Introduction. Suppose that the evolution of the state probabilities $p_i(t)$ of a Markov process is governed by the system of differential equations

$$\frac{dp_i}{dt} = \sum_{j=1}^{N} Q_{ij} p_j(t), \qquad i = 1, \cdots, N, \tag{1}$$

where Q_{ij} represents the transition rate from state S_j to state S_i ([1], p. 235). In many applications one is interested primarily in knowing the equilibrium state probabilities π_i defined by $\pi_i = \lim_{t\to\infty} p_i(t)$, which, if they exist, can be obtained by solving the system of homogeneous linear equations

$$\sum_{j=1}^{N} Q_{ij}\pi_j = 0, \qquad i = 1, 2, \cdots, N. \tag{2}$$

While (2) can be solved in principle (and in practice if N is not too large), the solution in general does not fulfill the ultimate desideratum of being susceptible to representation as a simple function of the transition rates Q_{ij}. If the states are simply ordered and a transition from a given state S_i can occur only to a neighboring state S_{i-1} or S_{i+1} then the equilibrium probability π_i satisfies the following simple formula

$$\pi_i = \frac{\lambda_1 \lambda_2 \cdots \lambda_i}{\mu_2 \mu_3 \cdots \mu_{i+1}} \pi_1, \tag{3}$$

where λ_k is the transition rate from state S_k to state S_{k+1} and μ_k, the transition rate from state S_{k+1} to state S_k, and π_1 is chosen so that

$$\sum_{i=1}^{N} \pi_i = 1. \tag{4}$$

Processes of this sort, with simply ordered sets of states, are called birth and death processes. The discussion of Section 1 below deals with a class of Markov matrices Q which includes the set of birth and death matrices and allows representations analogous to (3) for the equilibrium probabilities. In Section 2 it is shown that all the matrices in this class have non-positive characteristic values and in consequence of this fact, the difference between the state probability $p_i(t)$ and the corresponding equilibrium probability π_i is majorized by a function of t and the set of initial state probabilities $p_i(0)$. In Section 3 the foregoing theory is illustrated by an anisotropic random walk.

The following notions and notations are used. Let Q be an $N \times N$ matrix. The graph $G(Q)$ associated with Q consists of vertices $V_1, V_2, \cdots, V_N$ and of

Received April 10, 1958.

directed line segments joining certain of the vertices. The directed line segment $\overline{V_jV_i}$ connecting V_j to V_i belongs to $G(Q)$ if, and only if, $Q_{ij} \neq 0$. If $k_1, k_2, \cdots, k_s$ is a sequence of indices, the product of matrix elements $Q_{k_1k_2} Q_{k_2k_3} \cdots Q_{k_{s-1}k_s}$ shall be denoted by the symbol $[k_1k_2 \cdots k_s]$. V_i shall be said to be connected to V_j if there exists a sequence of indices $k_1, k_2, \cdots, k_s$ such that $[jk_1k_2 \cdots k_si] \neq 0$ If μ is a sequence of N positive numbers, $l_2(\mu)$ shall refer to an inner product space of sequences of N complex numbers with the inner product given by

$$(x, y) = \sum_{i=1}^{N} x_i \bar{y}_i \mu_i .$$

Every $N \times N$ matrix represents an operator on $l_2(\mu)$. A matrix A is called $l_2(\mu)$-symmetric if $(Ax, y) = (x, Ay)$ for every $x \in l_2(\mu)$ and every $y \in l_2(\mu)$.

1. The equilibrium probabilities. Let μ be a sequence of N positive numbers. It is clear that a matrix A is $l_2(\mu)$-symmetric if, and only if, for every pair of indices i and j,

$$A_{ij}\mu_i = A_{ji}\mu_j . \tag{5}$$

What conditions must a matrix A satisfy in order to be $l_2(\mu)$-symmetric for some sequence μ? The answer is that the matrix is characterized by the "loop condition", i.e., the product of matrix elements around a closed loop is independent of the direction in which the loop is traversed. This statement is made precise in

THEOREM 1. *A matrix A with non-negative off-diagonal elements A_{ij} such that $A_{ij} \neq 0$ implies $A_{ji} \neq 0$ is $l_2(\mu)$-symmetric for some sequence μ of N positive numbers if, and only if, for every index i and every sequence of indices $k_1, k_2, \cdots, k_s$,*

$$[ik_1k_2 \cdots k_si] = [ik_sk_{s-1} \cdots k_1i]. \tag{6}$$

PROOF. Suppose A is $l_2(\mu)$-symmetric. A straight-forward induction based on (5) shows that (6) holds.

Now suppose that (6) holds for every sequence of indices $(k_1, k_2, \cdots k_s)$ and every index i. The proof shall be based on the construction of a μ sequence of N positive numbers such that A is $l_2(\mu)$-symmetric.

By hypothesis, if $\overline{V_iV_j}$ is in $G(A)$, then so is $\overline{V_jV_i}$. Thus $G(A)$ is the union of graphs $G(A) = G_1 + G_2 + \cdots + G_m$, which are pairwise disconnected, each one separately being, however, a connected graph.[1] Let G_i be a particular one of these subgraphs and V_i one of its vertices. Let $\mu_i > 0$ and for every $V_j \in G_i$ define

$$\mu_j = \frac{[ik_1k_2 \cdots k_sj]}{[jk_sk_{s-1} \cdots k_1i]} \mu_i , \tag{7}$$

where the chain of vertices in G_i with indices $k_1, k_2 \cdots, k_s$ connects i to j. It is clear that (6) guarantees the uniqueness of μ_j once μ_i has been chosen. For, sup-

[1] A graph G is called connected if a) in case the line $\overline{V_mV_n} \in G$ then also $\overline{V_nV_m} \in G$ and b) every pair of vertices is connected by at least one chain of lines.

pose $k_1', k_2', \cdots, k_t'$ were another chain connecting i to j. Let

$$\mu_j' = \frac{[ik_1'k_2' \cdots k_t'j]}{[jk_t'k_{t-1}' \cdots k_1'i]} \mu_i .$$

(6) requires that

$$[ik_1k_2 \cdots k_sj][jk_t'k_{t-1}' \cdots k_1'i] = [ik_1' \cdots k_t'j][jk_s \cdots k_1i]$$

and thus, $\mu_j = \mu_j'$. The procedure is repeated for all G_i. For the resulting sequence μ it is apparent that all $\mu_i > 0$ and that (5) is satisfied.

COROLLARY 1. *If A has non-negative off-diagonal elements and the graph $G(A)$ consists of only one connected set and* (6) *is satisfied, then there exists a unique sequence μ of positive numbers such that A is $l_2(\mu)$-symmetric and*

$$\mu_1^{-1} + \mu_2^{-1} + \cdots + \mu_N^{-1} = 1. \tag{8}$$

In addition to possessing non-negative off-diagonal elements, a Markov matrix Q has the property that

$$Q_{ii} = -\sum_{j \neq i} Q_{ji} \tag{9}$$

and thus that the sum of all of the elements in a column vanishes. A consequence of this fact is

THEOREM 2. *If a Markov matrix Q is $l_2(\mu)$-symmetric for some sequence μ satisfying* (8) *and its graph is connected, then its equilibrium probabilities π_i are given by $\pi_i = \mu_i^{-1}$.*

PROOF. By hypothesis (5) holds: $Q_{ij}\,\mu_j^{-1} = Q_{ji}\,\mu_i^{-1}$. Summation of both members with respect to i and use of (9) complete the proof. The preceding results are summarized in

THEOREM 3. *If the evolution of state probabilities of a Markov process is governed by* (1) *and transition is possible between any pair of states in one or more jumps and for every i and every set (k_α) of indices, $[i\, k_1k_2 \cdots k_s\, i] = [i\, k_sk_{s-1} \cdots k_1\, i]$,* and $Q_{ij} \neq 0$ implies $Q_{ji} \neq 0$, *then the equilibrium probability π_f is given by*

$$\pi_f = [fs\; s-1 \cdots 1i]\pi_i/[i12 \cdots sf], \tag{10}$$

where the initial state has been chosen arbitrarily and the set of states $S_1, S_2, \cdots, S_s$ has been picked so that the product in the denominator does not vanish. Finally, of course, (10) *together with $\sum_{k=1}^{N} \pi_k^{-1} = 1$ prescribes a unique set of equilibrium probabilities.*

2. The approach to equilibrium. If a transition rate matrix Q is $l_2(\mu)$-symmetric then it represents a symmetric operator on $l_2(\mu)$. But every characteristic value of Q is of the form $\lambda(x) = (Qx, x)/(x, x)$ for some sequence $x \in l_2(\mu)$. Therefore $\lambda(x) \leqq 0$ implies that all characteristic values of Q are non-positive. We have

$$(Qx, y) = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} \overline{(y_i\mu_i - y_j\mu_j)}\,(Q_{ij}x_j - Q_{ji}x_i), \tag{11}$$

but $Q_{ij}\mu_i = Q_{ji}\mu_j$. Therefore,

$$(Qx, y) = -1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} Q_{ji}/\mu_i \overline{(y_i \mu_i - y_j \mu_j)} (x_i \mu_i - x_j \mu_j),$$

and thus $\lambda(x) \leqq 0$. The result is stated in

THEOREM 4. *If a Markov matrix is $l_2(\mu)$-symmetric then all of its characteristic values are non-positive. Zero is the largest characteristic value.*

The importance of the characteristic values of Q lies in the fact that they control the approach to equilibrium. This fact is elucidated in

THEOREM 5. *Let Q be an $l_2(\mu)$-symmetric Markov matrix with connected graph. Let π be the sequence of equilibrium probabilities. Let $p(t)$ be the solution of* (1) *with initial value $p(0)$. If $-\lambda_M$ and $-\lambda_m$ are the largest and smallest negative characteristic values of Q then*

$$\exp(-\lambda_m t) \| p(0) - \pi \| \leqq \| p(t) - \pi \| \leqq \exp(-\lambda_M t) \| p(0) - \pi \|$$

PROOF. Letting $r(t) = p(t) - \pi$, (1) implies that

$$\frac{d \| r \|^2}{dt} = 2 \left(r, \frac{dr}{dt} \right) = 2(r, Qr).$$

Since $p(t) = a \pi + q$, where $(q, \pi) = 0$, we have

$$a = (p(t), \pi) = \textstyle\sum_{k=1}^{N} p_k(t) \ \pi_k \pi_k^{-1} = \sum_{k=1}^{N} p_k(t) = 1,$$

and thus $r(t)$ is orthogonal to π. Hence, on identifying characteristic values by their variational properties (see, for example, [2], p. 230),

$$-\lambda_m = 2 \min_{q \perp \pi} (q, Qq)/\| q \|^2 \leqq \frac{d \| \pi \|^2}{dt} \Big/ \| \pi \|^2 \leqq 2 \min_{q \perp \pi} (q, Qq)/\| q \|^2 = -\lambda_M$$

Integration and identification of $r(t)$ and the constant of integration complete the proof.

3. A homogeneous anisotropic random walk on a finite lattice. Suppose a particle moves about among the points of a finite 3-dimensional lattice and that the conditional probability that at time $t + \Delta t$ it is at point $P' = (x_1', x_2', x_3')$ having been at $P = (x_1, x_2, x_3)$ at time t is given by:

M_E	for	$x_1' = x_1 + 1$	$x_2' = x_2$	$x_3' = x_3$
M_W	"	$x_1' = x_1 - 1$	$x_2' = x_2$	$x_3' = x_3$
M_N	"	$x_1' = x_1$	$x_2' = x_2 + 1$	$x_3' = x_3$
M_S	"	$x_1' = x_1$	$x_2' = x_2 - 1$	$x_3' = x_3$
M_U	"	$x_1' = x_1$	$x_2' = x_2$	$x_3' = x_3 + 1$
M_D	"	$x_1' = x_1$	$x_2' = x_2$	$x_3' = x_3 - 1$
0		otherwise.		

Transition rates to points other than those of the lattice vanish. The resulting process is $l_2(\mu)$-symmetric and thus if π_p denotes the equilibrium probability of the particle's being at $P = (x_1, x_2, x_3)$ and π_0 that of being at the origin, then

$$\pi_p = \left(\frac{M_E}{M_W}\right)^{x_1} \left(\frac{M_N}{M_S}\right)^{x_2} \left(\frac{M_Z}{M_N}\right)^{x_3} \pi_0 .$$

It is a pleasure to acknowledge the help provided through stimulating conversations with my colleagues J. Tukey and V. E. Beneš.

(*Added in proofs.*) There is considerable overlap between the work reported here and the results (particularly Theorem VIII) of the Research Announcement, "Integral Representations for Markov Transition Probabilities," by D. G. Kendall, *Bulletin A.M.S.* **64** (1958), 358–362. The notion of symmetrizability was suggested to the author, as it seems to have been to D. G. Kendall (who calls it reversability), by the spectral decomposition of Birth and Death Markov matrices effected by Kac [(1)], Ledermann and Renter [7], and Karlin and McGregor [2]. (References are to the bibliography of the above cited announcement.)

REFERENCES

[1] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, New York, 1952.
[2] F. Riesz et B. Sz. Nagy *Lecons d'Analyse Fonctionelle*, Académie des Sciences de Hongrie, Budapest, 1952.