A RELATIONSHIP BETWEEN ARBITRARY POSITIVE MATRICES AND DOUBLY STOCHASTIC MATRICES

By Richard Sinkhorn

University of Houston

1. Introduction. Suppose one observes n transitions of a Markov chain with N states and stochastic matrix $P = (p_{ij})$. The usual estimate of p_{ij} is $t_{ij} = a_{ij}/\lambda_i$ where a_{ij} is the number of transitions from i to j which are observed, and $\lambda_i = \sum_j a_{ij}$. (Cf. [1].) This amounts to a normalization of the rows of $A = (a_{ij})$, and can be expressed as a matrix equation $T = D_1A$ where $T = (t_{ij})$ and $D_1 = \text{diag}[\lambda_1^{-1}, \cdots, \lambda_N^{-1}]$.

If it is known that the stochastic matrix P is in fact doubly stochastic, (i.e., $\sum_i p_{ij} = 1$), what then is a good estimate of T? The maximum likelihood equations are difficult to solve. One estimate which has been used (for example, by Welch [4]) is to alternately normalize the rows and columns of A, in the belief that this iterative process converges to a doubly stochastic matrix, T, which might be, in some sense, a good estimate.

It is not the intent of this paper to obtain properties of this estimate but only to examine the mechanics of the iteration itself. In the next section we shall study this in detail and show that it is always convergent if the matrix A is strictly positive (i.e., $a_{ij} > 0$ for all i, j), and in fact that there exist diagonal matrices D_1 and D_2 (unique up to a scalar factor) with positive diagonals such that $T = D_1AD_2$. T is the only doubly stochastic matrix expressable in this form for a given strictly positive A.

For completeness we shall include a corollary to this result due to Marcus and Newman [3] which states that if A is symmetric and has positive entries, then there exists a diagonal matrix D with positive main diagonal entries such that DAD is doubly stochastic.

Finally in the last section we shall show by example that convergence need not occur at all if some $a_{ij} = 0$, or even if it does there need exist no associated diagonal matrices D_1 and D_2 as in the strictly positive case. Even the apparently natural artifice of replacing the zero entries by "small" functions $a_{ij}(\epsilon)$ of a parameter ϵ, getting $T(\epsilon)$ and letting $\epsilon \to 0$ leads to difficulties.

2. The alternating iteration for positive matrices.

Theorem 1. *To a given strictly positive $N \times N$ matrix A there corresponds exactly one doubly stochastic matrix T_A which can be expressed in the form $T_A = D_1AD_2$ where D_1 and D_2 are diagonal matrices with positive diagonals. The matrices D_1 and D_2 are themselves unique up to a scalar factor.*

Proof. We shall establish only the uniqueness part here. The existence will be demonstrated constructively in the proof of Theorem 2.

If there exist two different pairs of diagonal matrices D_1, D_2 and C_1, C_2 such

Received 17 December 1962; revised 25 November 1963.