10

Sonderabdruck aus
ARCHIV DER MATHEMATIK
Vol. VIII, 1957 BIRKHÄUSER VERLAG, BASEL UND STUTTGART Fasc. 5

Matrix Norms applied to Weakly Ergodic Markov Chains

By J. L. Mott in Edinburgh and Hans Schneider in Belfast

Recently Hajnal [2] and Mott [4] have considered conditions for a non-homogeneous finite Markov chain to be *weakly ergodic*, that is, for the probability distribution to tend to independence of the initial distribution as the number of trials increases. This case is to be distinguished from the more particular one of a chain which is *strongly ergodic*, where the probability distribution tends not only to independence of the initial distribution but also to a limit. Thus in the case of weak ergodicity, if P_1, P_2, ... are the successive matrices of the chain and

$$P^{(n)} = P_1 P_2 \ldots P_n ,$$

there exists for each given n a stochastic matrix $Q^{(n)}$ such that the difference $P^{(n)} - Q^{(n)}$ tends to the limit zero as n increases. Because of the tendency to independence of the initial distribution the rows of $P^{(n)}$ tend to be the same, so that here $Q^{(n)}$ is a stochastic matrix with identical rows: such a matrix we call *stable*. In this note we shall use a result on matrix norms to prove another condition sufficient for weak ergodicity.

Let $\mathfrak{A}$ be a subalgebra of the algebra of real matrices of some fixed order k over the real numbers. A (multiplicative) norm on $\mathfrak{A}$ is a mapping ν of $\mathfrak{A}$ into the non-negative numbers satisfying

(i) $\nu(A) > 0 \quad \text{if } A \neq 0$;
(ii) $\nu(\lambda A) = |\lambda| \, \nu(A) \quad \text{if } \lambda \text{ is real}$;
(iii) $\nu(A + B) \leqq \nu(A) + \nu(B)$;
(iv) $\nu(A B) \leqq \nu(A) \, \nu(B)$.

It is known that if $A_1, A_2, \ldots$ is a sequence of matrices in $\mathfrak{A}$, then $\lim\limits_{n\to\infty} \nu(A_n) = 0$ implies that $\lim\limits_{n\to\infty} A_n = 0$. This theorem is an immediate consequence of a result proved by Ostrowski [5] in the course of the proof of his Theorem 4; it is also a special case of Theorem 2 of Bourbaki [1], §2.3. It has been used extensively by Householder [3], and others.

Now take $\mathfrak{A}$ as the subalgebra generated by the stochastic matrices of order k. Clearly A belongs to $\mathfrak{A}$ if and only if all rows of A have the same sum. We now prove the

Theorem. *Let ν be a norm on $\mathfrak{A}$, and $T_1, T_2, \ldots$ be some sequence of stable stochastic matrices. Then the chain of stochastic matrices $P_1, P_2, \ldots$ is weakly ergodic if*

$$\lim_{n\to\infty} \prod_{i=1}^{n} \nu\,(P_i - T_i\, P_i) = 0 .$$

Proof. If P is stochastic and T is stable then $PT = T$ and TP is stable. Hence in particular $T_1 P^{(n)}$ is stable and so, by the result on matrix norms quoted above, it is enough to prove that $\lim_{n\to\infty} \nu(P^{(n)} - T_1 P^{(n)}) = 0$. But

$$\begin{aligned}(P_1 - T_1 P_1)(P_2 - T_2 P_2) &= P_1 P_2 - T_1 P_1 P_2 - P_1 T_2 P_2 + T_1 P_1 T_2 P_2 = \\ &= P_1 P_2 - T_1 P_1 P_2\,,\end{aligned}$$

and so by induction it follows that

$$\prod_{i=1}^{n} (P_i - T_i P_i) = P^{(n)} - T_1 P^{(n)}\,.$$

The proof now follows from the hypothesis of the theorem since $P_i - T_i P_i$ belongs to $\mathfrak{A}$ and

$$\prod_{i=1}^{n} \nu(P_i - T_i P_i) \geqq \nu\left\{\prod_{i=1}^{n} (P_i - T_i P_i)\right\} = \nu(P^{(n)} - T_1 P^{(n)})\,.$$

The usefulness of our theorem lies in the wide variety of norms available. In addition, it is clear that the rows of TP are weighted means of the rows of P, and consequently we should normally choose our T_i corresponding to each given P_i so as to make the elements of $P_i - T_i P_i$ as small as possible.

As a simple example consider a homogeneous chain whose matrix of transition probabilities is P where

$$P = \begin{bmatrix} \cdot & 0{\cdot}7 & 0{\cdot}1 & 0{\cdot}2 \\ 0{\cdot}4 & \cdot & 0{\cdot}3 & 0{\cdot}3 \\ 0{\cdot}6 & 0{\cdot}1 & \cdot & 0{\cdot}3 \\ 0{\cdot}4 & 0{\cdot}2 & 0{\cdot}4 & \cdot \end{bmatrix}.$$

For the norm here use the column norm; $\nu(A) = \max_j \sum_i |a_{ij}|$, where $A = [a_{ij}]$. Then clearly, whatever the choice of T, the least possible values of $\sum_i |a_{ij}|$ in our case of $A = P - TP$ are 0.6, 0.8, 0.6, 0.4 for $j = 1, 2, 3, 4$ respectively; so that, with this choice of norm, the least value of $\nu(P - TP)$ is 0.8. This value is achieved, for example, by T with each row [0.2, 0.4, 0.2, 0.2]. With each row of T the simpler $\frac{1}{4}$ [1, 1, 1, 1] we have $\nu(P - TP) = 0.9$, but this value of the norm, although not the best possible, is sufficient to show that the chain is weakly ergodic. (In fact, the chain, being homogeneous, is also strongly ergodic.) We note that the ergodicity of this chain, and of analogous non-homogeneous chains, can not be demonstrated by use of the sufficient conditions given by HAJNAL [2] and MOTT [4].

The following example illustrates some further points. Let

$$Q = \begin{bmatrix} \cdot & \frac{1}{2} & \frac{1}{2} & \cdot \\ \cdot & \frac{1}{2} & \frac{1}{2} & \cdot \\ \frac{1}{2} & \cdot & \cdot & \frac{1}{2} \\ \frac{1}{2} & \cdot & \cdot & \frac{1}{2} \end{bmatrix} \quad \text{and } R = \begin{bmatrix} \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \cdot & \cdot \\ \frac{1}{2} & \frac{1}{2} & \cdot & \cdot \\ \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Then the homogeneous chains with P and R respectively as their matrices of transition probabilities are each strongly ergodic. But

$$QR = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \cdot & \cdot \\ \frac{1}{2} & \frac{1}{2} & \cdot & \cdot \\ \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \\ \cdot & \cdot & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

and, since $(QR)^n = QR$ which is not stable, the homogeneous chain with matrix of transition probabilities QR is not weakly ergodic. Thus for all stable stochastic matrices V and all norms ν we have from our theorem that $\nu(QR - VQR) \geqq 1$. It follows that

$$\nu(Q - TQ)\,\nu(R - SR) \geqq \nu(QR - TQR) \geqq 1$$

for all stable stochastic matrices T, S and all norms ν. The equality is attained by taking $S = T$ and each row of T as $[1, \cdot, \cdot, \cdot]$, and for norm the column norm as above; for then $\nu(Q - TQ) = \nu(R - TR) = 1$.

Now consider the non-homogeneous chain defined by the sequence $\{Q_1, R, Q_2, R, Q_3, R, \ldots\}$ where Q_n is as Q above but with the first row of Q replaced by

$$\left[\alpha_n, \frac{1}{2} - \alpha_n, \frac{1}{2} - \alpha_n, \alpha_n\right], \qquad 0 \leqq \alpha_n \leqq \frac{1}{2}.$$

Then $\nu(Q_n - TQ_n) = 1 - \alpha_n$ so that the chain is weakly ergodic provided that $\sum_{r=1}^{n} \alpha_r$ diverges as $n \to \infty$.

References

[1] N. BOURBAKI, Éléments de Mathématique, I, Livre V. Espaces vectoriels topologiques, Chapitre I. Paris 1953.

[2] J. HAJNAL, The ergodic properties of non-homogeneous finite Markov chains. Proc. Cambridge philos. Soc. 52, 67—77 (1956).

[3] A. S. HOUSEHOLDER, On the convergence of matrix iterations. Oak Ridge National Laboratory-1883, Physics (1955).

[4] J. L. MOTT, Conditions for the ergodicity of non-homogeneous finite Markov chains. Proc. roy. Soc. Edinburgh, Sect. A 64, 369—380 (1957).

[5] A. OSTROWSKI, Über Normen von Matrizen. Math. Z. 63, 2—18 (1955).

Eingegangen am 3. 10. 1957