ON THE ASYMPTOTIC PROBABILITY DISTRIBUTION FOR CERTAIN MARKOFF PROCESSES

By WALTER LEDERMANN

Communicated by M. S. Bartlett

Received 16 March 1950; *and in revised form* 26 April 1950

1. *Introduction.* The type of Markoff process which is considered in this paper* corresponds to a system capable of n states, the time being regarded as a continuously varying parameter. At any instant t the probability distribution is represented by the vector

$$p(t) = \begin{pmatrix} p^{(1)}(t) \\ \cdots \\ p^{(n)}(t) \end{pmatrix}, \quad \sum_{i=1}^{n} p^{(i)}(t) = 1. \tag{1}$$

We assume that the probability of transition from state j to state i $(i \neq j)$ in time Δt is given by

$$a_{ij}\Delta t + o(\Delta t),$$

where a_{ij} is a constant. This implies a relation between the components of $p(t+\Delta t)$ and those of $p(t)$. For, by the elementary laws of the probability calculus, we must have

$$p^{(1)}(t+\Delta t) = [1-(a_{21}+\ldots+a_{n1})\,\Delta t + o(\Delta t)]\,p^{(1)}(t) + \{a_{12}p^{(2)}(t)+\ldots+a_{1n}p^{(n)}(t)\}\,\Delta t + o(\Delta t),$$

where the expression in the square brackets denotes the probability that the system does not move from the state 1 to any other state during the interval Δt. In the case $\Delta t \to 0$, which we shall henceforth consider, we obtain the differential equation

$$\frac{d}{dt}p^{(1)}(t) = -\{(a_{21}+\ldots+a_{n1})\,p^{(1)}(t) - a_{12}p^{(2)}(t) - \ldots - a_{1n}p^{(n)}(t)\}.$$

Thus the vector $p(t)$ satisfies the equation

$$\frac{d}{dt}p(t) = -Ap(t), \tag{2}$$

where A is a constant matrix of the form

$$A = \begin{pmatrix} a_{11} & -a_{12} & -a_{13} & \ldots & -a_{1n} \\ -a_{21} & a_{22} & -a_{23} & \ldots & -a_{2n} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ -a_{n1} & -a_{n2} & -a_{n3} & \ldots & a_{nn} \end{pmatrix}, \tag{3}$$

in which

$$a_{ij} \geqslant 0 \quad (i \neq j), \qquad a_{jj} = \sum_{i=1}^{n}{}' a_{ij}. \tag{4}$$

Here as elsewhere the accent at the summation symbol indicates that the term which would correspond to equal suffixes is to be omitted.

* I am indebted to Prof. M. S. Bartlett for drawing my attention to this problem. I should like to thank him and also Mr J. E. Moyal and Mr D. G. Kendall for many valuable suggestions and for the advice they gave me during the writing of this paper.

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

A matrix A of this type will be called a *transition matrix*.

It is known that the solution of (2) may formally be written

$$p(t) = \left\{ \sum_{k=0}^{\infty} \frac{1}{k!} (-tA)^k \right\} p_0 = e^{-tA} p_0, \tag{5}$$

where

$$p_0 = p(0) \tag{6}$$

is the initial probability distribution. Our problem is then to find the limiting distribution

$$p_\infty = \lim_{t \to \infty} e^{-tA} p_0. \tag{7}$$

When zero is a simple latent root of A, the evaluation of the limit presents no great difficulties and has in fact been carried out by several authors*, the most interesting feature being that the result is independent of the initial conditions. But when zero is a multiple latent root, the existence of the limit is assured only after it has been proved that the corresponding elementary divisors are linear (§ 4), and in this case the final distribution does depend on the initial distribution. However, it will be shown that the n states are divided into certain sets of states such that the ratio of the asymptotic probabilities of two states of the same set is still independent of the initial conditions.

The problem with which the present paper deals is only a very special instance of a stochastic process, and it is to be expected that some of my conclusions follow from the general and far-reaching results that various writers have obtained in this important branch of probability theory. Indeed, W. Doeblin (2) has stated results which are virtually equivalent to establishing the nature of our final solution (§ 6). But I have not come across in the literature any method for constructing the asymptotic distribution by a simple algebraical process, and it is perhaps also worth while to derive the existence theorems for this elementary case without invoking the much deeper results of the general theory (see Fréchet (3) and a forthcoming paper of D. G. Kendall.

2. *The canonical form of A.* According to the classical theory of canonical matrices† there exists a non-singular matrix H such that

$$HAH^{-1} = \begin{pmatrix} C_1(\alpha_1) & & & \\ & C_2(\alpha_2) & & \\ & & \ddots & \\ & & & C_k(\alpha_k) \end{pmatrix} = L, \tag{8}$$

where $C_1, C_2, \ldots, C_k$ are the *simple classical submatrices* corresponding to the elementary divisors of $\lambda I - A$.

A typical submatrix, of order f, say, is

$$C(\alpha) = \begin{pmatrix} \alpha & 1 & 0 & \ldots & 0 & 0 \\ 0 & \alpha & 1 & \ldots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & \alpha & 1 \\ 0 & 0 & 0 & \ldots & 0 & \alpha \end{pmatrix}_f.$$

* See, for example, B. J. Prendiville (9). A full account is given by M. Fréchet ((3), pp. 201 ff.).

† H. W. Turnbull and A. C. Aitken ((11), Chapter VI).

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

It is only when all elementary divisors are linear, i.e. when f is always equal to unity, that A can be transformed into diagonal form. For in that case $C(\alpha)$ reduces to a single number (α). Of course, non-linear elementary divisors can occur only if A has multiple latent roots.

By applying a result of A. Brauer (1) we can show* that the *latent roots of A have either positive real parts or else are zero.* For if α is a latent root, the equation $x'A = \alpha x'$, i.e. the n equations

$$x_j a_{jj} - \sum_i{}' x_i a_{ij} = \alpha x_j \quad (j = 1, 2, \ldots, n) \tag{9}$$

have a non-trivial solution $x_1, x_2, \ldots, x_n$. Let

$$|x_M| = \max(|x_1|, |x_2|, \ldots, |x_n|).$$

Putting $j = M$ in (9), we have

$$(\alpha - a_{MM})x_M = -\sum_i{}' x_i a_{iM},$$

$$|\alpha - a_{MM}||x_M| \leqslant |x_M| \sum_i{}' a_{iM} = |x_M| a_{MM}$$

by (4). Thus

$$|\alpha - a_{MM}| \leqslant a_{MM},$$

which means that, in the complex plane, α lies within or on the circumference of the circle with centre a_{MM} $(\geqslant 0)$ and radius a_{MM}. Hence

$$\text{either} \quad \alpha = 0, \quad \text{or} \quad \Re(\alpha) > 0. \tag{10}$$

In particular, there are no non-zero pure imaginary latent roots, and the possibility of recurring solutions is therefore ruled out.

3. *Evaluation of* $\lim e^{-tA}$. Evidently

$$\lim_{t\to\infty} e^{-tA}$$

exists if and only if

$$\lim e^{-tC(\alpha)}$$

exists for each simple submatrix of the canonical form (8). Let

$$C(\alpha) = \alpha I + J, \tag{11}$$

where I $(= I_f)$ is the unit matrix of order f and

$$J = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix}_f. \tag{12}$$

Since

$$J^f = 0,$$

we have

$$e^{-tC(\alpha)} = e^{-\alpha t I} e^{-tJ} = e^{-\alpha t} \sum_{m=0}^{f-1} \frac{(-tJ)^m}{m!}. \tag{13}$$

* Following a remark by D. G. Kendall; see also O. Taussky ((10), p. 675).

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

The last matrix factor is a polynomial of degree $f-1$ in t. We shall have to distinguish two cases

(i) if $$\Re(\alpha) > 0, \quad \lim_{t\to\infty} e^{-tC(\alpha)} = 0,$$

since the exponential factor tends to zero more rapidly than the polynomial tends to infinity;

(ii) if $$\alpha = 0, \quad e^{-tC(0)} = \sum_{m=0}^{f-1} \frac{(-tJ)^m}{m!} \tag{14}$$

and the limit, as $t\to\infty$, does not exist unless $f = 1$, when

$$C(0) = 0$$

and $$e^{-tC(0)} = (1).$$

Thus the limit exists if and only if all elementary divisors which correspond to the latent root zero are of degree unity.

Suppose now that this condition is fulfilled and that zero is a ν-fold root of

$$|\lambda I - A| = 0.$$

Then if in the canonical form A all submatrices that belong to the latent root zero are collected at the beginning, we can write

$$HAH^{-1} = \begin{pmatrix} 0_\nu & 0 \\ 0 & M \end{pmatrix}, \tag{15}$$

where the matrix in the top left-hand corner is the zero matrix with ν rows and columns, and where

$$M = \begin{pmatrix} C_{\nu+1}(\alpha_{\nu+1}) & & \\ & \ddots & \\ & & C_k(\alpha_k) \end{pmatrix}$$

comprises all submatrices which correspond to non-zero latent roots, whose real parts are therefore positive.

Let H be decomposed into rows and H^{-1} $(= G)$ into columns, thus

$$H = \begin{pmatrix} h_1' \\ h_2' \\ \vdots \\ h_n' \end{pmatrix}, \quad H^{-1} = G = (g_1, g_2, \ldots, g_n),$$

where $$h_i' g_j = \delta_{ij} \quad (i,j = 1, 2, \ldots, n).$$

On writing (15) in either of the forms

$$HA = \begin{pmatrix} 0_\nu & 0 \\ 0 & M \end{pmatrix} H, \quad AH^{-1} = H^{-1} \begin{pmatrix} 0_\nu & 0 \\ 0 & M \end{pmatrix},$$

we see that $$h_1' A = h_2' A = \ldots = h_\nu' A = 0, \tag{16}$$

$$Ag_1 = Ag_2 = \ldots = Ag_\nu = 0. \tag{17}$$

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

Also $$\lim_{t\to\infty} e^{-tM} = 0, \quad \lim_{t\to\infty} e^{-t0_\nu} = I_\nu,$$

and hence $$\lim_{t\to\infty} e^{-tA} = H^{-1}\begin{pmatrix} I_\nu & 0 \\ 0 & 0 \end{pmatrix} H,$$

$$\lim_{t\to\infty} e^{-tA} = g_1 h_1' + g_2 h_2' + \ldots + g_\nu h_\nu'. \tag{18}$$

Therefore, by (7), $$p_\infty = g_1(h_1' p_0) + g_2(h_2' p_0) + \ldots + g_\nu(h_\nu' p_0),$$

or more briefly $$p_\infty = \omega_1 g_1 + \omega_2 g_2 + \ldots + \omega_\nu g_\nu, \tag{19}$$

where $$\omega_l = h_l p_0 \quad (l = 1, 2, \ldots, \nu) \tag{20}$$

are scalars which depend on the initial distribution, while the vectors $g_1, g_2, \ldots, g_\nu$ depend only on A, but not on p_0.

It is convenient to express the condition for the existence of an asymptotic distribution in a different form.

THEOREM I. *Let A be a transition matrix which has zero as a ν-fold latent root. Then*

$$\lim_{t\to\infty} e^{-tA}$$

exists if and only if A is of rank $n-\nu$.

Proof. We have already seen that when the limit exists, the canonical form of A is given by (15). The matrix is clearly of rank $n-\nu$, and hence so is A itself.

In order to prove the converse we assume that $r(A) = n-\nu$ and remark that by (8)

$$r(A) = \sum_{m=1}^{k} r(\mathfrak{C}_m(\alpha_m)),$$

where $r(X)$ denotes the rank of X. Now if $C(\alpha)$ is a submatrix of degree f corresponding to a latent root α, then

$$r(C(\alpha)) = f-1, \quad \text{if} \quad \alpha = 0,$$
$$r(C(\alpha)) = f, \quad \text{if} \quad \alpha \neq 0.$$

Suppose now that the first μ submatrices $C(\alpha)$ correspond to the latent root zero, while the remaining $k-\mu$ submatrices correspond to non-zero latent roots. Denoting the order of $C_j(\alpha_j)$ by f_j, we have

$$n = \sum_{j=1}^{n} f_j,$$

and since $C_j(0)$ corresponds to a f-fold zero root,

$$\nu = \sum_{j=1}^{\mu} f_j.$$

On the other hand, $$r(A) = \sum_{j=1}^{\mu} (f_j - 1) + \sum_{i=\mu+1}^{k} f_i = n-\mu.$$

Thus our assumption about $r(A)$ implies that

$$\nu = \mu,$$

and since $f_j \geqslant 1$, $$f_1 = f_2 = \ldots = f_\mu = 1.$$

Thus the condition for the existence of the limit is satisfied (see p. 584), if $r(A) = n-\nu$. This completes the proof.

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

We shall eventually prove that the condition expressed in Theorem I is automatically fulfilled by every transition matrix. But it may be of some interest to point out that the elementary divisors which correspond to a non-zero latent root need not be linear. For example, if

$$A = \begin{pmatrix} 4 & -1 & -2 \\ -2 & 3 & -1 \\ -2 & -2 & 3 \end{pmatrix},$$

we have

$$|\lambda I - A| = \lambda(\lambda - 5)^2.$$

Hence 5 is a double latent root, but the rank of

$$5I - A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 1 \\ 2 & 2 & 2 \end{pmatrix}$$

is two. This implies that the canonical form of A is

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{pmatrix}.$$

4. *The rank of A.* Our object is to prove that

$$r(A) = n - \nu,$$

where ν is the multiplicity of the root $\lambda = 0$ of the equation

$$|\lambda I - A| = 0.$$

The following remark of Minkowski (6) is useful in determining the rank of A.

LEMMA. *Let $u_1, u_2, \ldots, u_n$ be non-negative quantities and let*

$$D(u) = \begin{vmatrix} a_{11}+u_1 & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & a_{22}+u_2 & \ldots & -a_{2n} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ -a_{n1} & -a_{n2} & \ldots & a_{nn}+u_n \end{vmatrix},$$

where

$$a_{ij} \geqslant 0 \quad (i \neq j), \qquad \sum_{i=1}^{n}{}' a_{ij} = a_{jj}. \tag{4}$$

Then

$$D(u) \geqslant 0.$$

Proof. It will be convenient to have a concise notation for the principal minors of $D(u)$. Thus if

$$l_1 < l_2 < \ldots < l_k$$

are k integers out of the set $1, 2, \ldots, n$, we denote by

$$[l_1, l_2, \ldots, l_k]_u \tag{21}$$

the principal minor formed from $D(u)$ by selecting the rows and columns with indices (21). The corresponding minor of $D(0)$, in which $u_1 = u_2 = \ldots = u_n = 0$ will be denoted by

$$[l_1, l_2, \ldots, l_k]_0.$$

The lemma is proved by induction with respect to n. When $n = 1$, the proposition is trivial. We note that, generally, a principal minor of $D(u)$ has the same characteristic property as $D(u)$ itself, namely, that all elements off the main diagonal are negative

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

or zero, but that the sum of the elements in each column is non-negative. Hence we may assume that, if $1 \leqslant k \leqslant n-1$,

$$[l_1, l_2, \ldots, l_k]_u > 0 \quad \text{if} \quad u \geqslant 0.$$

Expanding $D(u)$ with respect to the u's we obtain

$$D(u) = u_1 u_2 \ldots u_n + \Sigma u_1 \ldots u_{n-1}[n]_0 + \Sigma u_1 \ldots u_{n-2}[n-1, n]_0 + \ldots + \Sigma u_1[2, 3, \ldots, n]_0 + [1, 2, \ldots, n]_0 = 0,$$

where only the leading term in each sum is written out. Since

$$[1, 2, \ldots, n]_0 = |A| = 0,$$

and since by hypothesis,

$$u_i \geqslant 0, \quad [l_1, l_2, \ldots, l_k]_0 \geqslant 0 \quad (k < n),$$

the result follows at once. It is worth pointing out that $D(u) = 0$ only if each term in the expansion vanishes separately.

It will be convenient to interpret the rank of A as the number of linearly independent columns, and we shall therefore study the conditions under which a set of columns of A is linearly dependent. It is clear that any set of m ($\leqslant n$) columns of A contains exactly one m-rowed minor that is a principal minor of A. Hence we may speak of *the principal minor of a set of m columns.* It is a remarkable property of transition matrices that the linear dependence of any of their columns is completely decided by the vanishing of their principal minor of greatest order.

THEOREM II. *A set of m columns of A is linearly dependent if and only if their m-rowed principal minor vanishes.*

This theorem is an immediate consequence of the following more explicit result.

THEOREM III. *If m columns of A have the property that their principal m-rowed minor vanishes while no principal minor of lower order is zero, then all elements which do not form part of the m-rowed principal minor, are zero.*

Proof. We use induction with respect to m. When $m = 1$, we have to consider a single column of A, say the first, whose diagonal element vanishes. Thus we have

$$\sum_{i=1}^{n}{}' a_{i1} = 0.$$

Since $a_{ij} \geqslant 0$ it follows that $a_{21} = a_{31} = \ldots = a_{n1} = 0$,

i.e. the column is, in fact, the zero vector.

In dealing with the general case we may assume that the m columns in question are the first m columns of A. Let

$$s_j = \sum_{i=1}^{m}{}' a_{ij} \quad (j = 1, 2, \ldots, m), \tag{22}$$

and

$$u_j = a_{m+1,j} + a_{m+2,j} + \ldots + a_{nj}, \tag{23}$$

so that

$$a_{jj} = s_j + u_j \quad (j = 1, 2, \ldots, m). \tag{24}$$

Our hypothesis is now that

$$D(u) = \begin{vmatrix} s_1 + u_1 & -a_{12} & \ldots & -a_{1m} \\ -a_{21} & s_2 + u_2 & \ldots & -a_{2m} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ -a_{m1} & -a_{m2} & \ldots & s_m + u_m \end{vmatrix} = 0, \tag{25}$$

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

while no principal minor of $D(u)$, other than $D(u)$ itself, is zero. We have to prove that

$$a_{ij} = 0 \quad (j = 1, 2, \ldots, m;\ i = m+1, m+2, \ldots, n), \tag{26}$$

or, alternatively, that

$$u_1 = u_2 = \ldots = u_m = 0. \tag{27}$$

We use the same notation for the principal minors of $D(u)$ as on p. 586. Expanding (25) in terms of the u's we obtain

$$0 = u_1 u_2 \ldots u_m + \Sigma u_1 \ldots u_{m-1}[m]_0 + \Sigma u_1 \ldots u_{m-2}[m-1, m]_0 + \ldots + \Sigma u_1 [2, 3, \ldots, m]_0 + [1, 2, \ldots, m]_0. \tag{28}$$

As before

$$[1, 2, \ldots, m]_0 = 0,$$

since in this determinant all column sums are zero. As no term on the right-hand side of (28) is negative, we infer that all terms must vanish. In particular

$$u_1 u_2 \ldots u_m = 0.$$

It may be assumed that the columns are numbered in such a way that

$$u_m = 0. \tag{29}$$

But in that case

$$s_m = [m]_0 \neq 0. \tag{30}$$

For, if not, we should have

$$u_m + s_m = 0,$$

which would mean that a certain principal minor of $D(u)$ of order 1 vanishes, contrary to our hypothesis. The expansion (28) now reads

$$0 = u_1 u_2 \ldots u_{m-1}[m]_0 + \Sigma u_1 \ldots u_{m-2}[m-1, m]_0 + \ldots,$$

where again each term vanishes. Hence

$$u_1 u_2 \ldots u_{m-1}[m]_0 = 0, \tag{31}$$

whence by (30)

$$u_1 u_2 \ldots u_{m-1} = 0,$$

and after suitably arranging the columns

$$u_{m-1} = 0. \tag{32}$$

On the other hand, we now have

$$\begin{vmatrix} s_{m-1} + u_{m-1} & -a_{m-1,m} \\ -a_{m,m-1} & s_m + u_m \end{vmatrix} = \begin{vmatrix} s_{m-1} & -a_{m-1,m} \\ a_{m.m-1} & -s_m \end{vmatrix},$$

i.e.

$$[m-1, m]_u = [m-1, m]_0,$$

and as the quantity on the left does not vanish by hypothesis,

$$[m-1, m]_0 \neq 0.$$

Next we consider the term

$$u_1 u_2 \ldots u_{m-2}[m-1, m]_0 = 0.$$

It follows that

$$u_1 u_2 \ldots u_{m-2} = 0$$

and

$$u_{m-2} = 0,$$

say. Continuing this argument we eventually obtain (27). This completes the proof of Theorem III and hence establishes Theorem II. We especially note the following result.

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

COROLLARY. *If all principal minors of A vanish, then all m-rowed minors of A vanish.*

It is now easy to show that the condition of Theorem I is always fulfilled in the case of a transition matrix. For if zero is a ν-fold latent root of A, the last ν coefficients in the polynomial

$$|\lambda I - A| = \lambda^n - c_1\lambda^{n-1} + c_2\lambda^{n-2} - \ldots + (-1)^n c_n$$

must vanish, thus

$$c_n = c_{n-1} = \ldots = c_{n-\nu+1} = 0, \tag{33}$$

while

$$c_{n-\nu} \neq 0. \tag{34}$$

It is well known that

$$c_r = \textit{sum of all principal minors of order } r. \tag{35}$$

As by the Lemma (p. 586) all principal minors of $|A|$ are non-negative, it follows from (33) that all principal minors of order $n-\nu+1$ are zero. Hence, by the Corollary above, all minors of order $n-\nu+1$ vanish, while at least one (principal) minor of order $n-\nu$ is non-zero. This proves that A is of rank $n-\nu$.

5. *A normal form for a transition matrix.* It is clear that the underlying statistical problem is not altered if the n states of the system are renumbered. This amounts to replacing A by

$$PAP', \tag{36}$$

where P is a permutation matrix. The set of principal minors is left invariant by this operation. Our task is to determine a normal form for A under the transformation (36).

It is convenient to introduce the following concept:

DEFINITION. *m (>1) columns of A are said to form a minimal dependent set if they are linearly dependent, but if no proper subset of them is linearly dependent. A single column is said to form a minimal dependent set if and only if it is the zero vector.*

It is an immediate consequence of Theorem II that m columns form a minimal dependent set if and only if their m-rowed principal minor vanishes while no principal minor of lower order is zero.

As $|A| = 0$, the n columns of A are certainly linearly dependent. If no proper subset of them is linearly dependent we shall consider A to be in normal form.

In the contrary case, let m_1 be the least number of columns that form a minimal dependent set. (These columns need not be uniquely determined.) By permuting the rows and columns of A we can arrange that such a set of columns consists of the first m_1 columns and that their principal m_1-rowed minor is the leading m_1-rowed minor. The matrix obtained in this way will again be denoted by A. It follows from Theorem III that the first m_1 columns are of the form

$$\begin{pmatrix} A_1 \\ 0 \end{pmatrix},$$

where the determinant of A_1 is their m_1-rowed principal minor. We note that the column-sums of A_1 are zero, i.e. that A_1 is annihilated on the left by the row-vector

$$(1, 1, \ldots, 1)_{m_1}.$$

Of the remaining $n-m_1$ rows of A we select a further minimal dependent set of m_2 columns, where

$$1 \leqslant m_1 \leqslant m_2. \tag{37}$$

It may be assumed that the rows and columns of A have been arranged in such a way that this second set of columns follows immediately upon the first. Hence by a further

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

application of Theorem III we now see that the first m_1+m_2 columns appear in the shape

$$\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \\ 0 & 0 \end{pmatrix},$$

where the determinant of A_2 is the principal minor of the second minimal set. Continuing this process we find that A can be transformed into one or the other of the following normal forms*:

First type:

$$PAP' = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_\nu \end{pmatrix}. \tag{38}$$

Second type:

$$PAP' = \begin{pmatrix} A_1 & 0 & \dots & 0 & -B_1 \\ 0 & A_2 & \dots & 0 & -B_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_\nu & -B_\nu \\ 0 & 0 & \dots & 0 & B \end{pmatrix}. \tag{39}$$

The first type corresponds to the case in which the n columns of A are divided into exactly ν minimal dependent sets, there being no columns that do not belong to any such set. The submatrix

$$A_k \quad (k = 1, 2, \dots, \nu)$$

is a transition matrix of order m_k. As none of its principal minors vanishes, it is of rank $m-1$ and hence has only a simple zero latent root, i.e.

$$|\lambda I - A_k|_{m_k} = \lambda \psi_k(\lambda), \tag{40}$$

where

$$\psi_k(0) \neq 0.$$

Since $|\lambda I - A|$ is the product of the functions (40) and is, on the other hand, divisible by exactly the νth power of λ, it follows that ν is also equal to the number of submatrices in the normal form (38).

In the second type of canonical form the ν minimal dependent sets do not exhaust the n columns of A, but there is a residue of

$$q = n - \sum_{k=1}^{\nu} m_k \tag{41}$$

linearly independent columns, whose principal q-rowed minor is then necessarily non-zero, thus

$$|B| \neq 0. \tag{42}$$

6. *The asymptotic distribution.* Assuming now that A is in normal form we can describe in more detail the asymptotic distribution

$$p_\infty = \omega_1 g_1 + \omega_2 g_2 + \dots + \omega_\nu g_\nu \tag{19}$$

derived in § 3. Corresponding to the partitioning of the normal form of A, the n states of the system may be divided into ν subsets of states, answering to the submatrices

* An analysis decomposition in the case of discontinuous time was given by R. von Mises ((7), pp. 533–49). See also Fréchet ((3), pp. 187–98.

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

$A_1, A_2, \ldots, A_\nu, B$ and containing $m_1, m_2, \ldots, m_k, q$ states respectively. It will be sufficient to discuss the second type of normal form, as the first type can be obtained from it by obvious modifications ($q = 0$).

The vectors g in (19) are solutions of

$$Ax = 0. \tag{43}$$

On writing

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_\nu \\ y \end{pmatrix},$$

where x_k and y_k are column vectors of orders m_k and q respectively, we find that (43) is equivalent to

$$\left.\begin{aligned} A_k x_k - B_k y &= 0 \\ By &= 0 \end{aligned}\right\} \quad (k = 1, 2, \ldots, \nu).$$

On account of (42) we must have $y = 0$

and hence $$A_k x_k = 0 \quad (k = 1, 2, \ldots, \nu).$$

As A_k is of rank $m - 1$, the vector x_k is uniquely determined, apart from a scalar factor. In the next section it will be shown that the components of x_k can always be taken to be non-negative and such that their sum is equal to unity. When the scalar factor has been determined in this manner we shall put

$$x_k = f_k \quad (k = 1, 2, \ldots, \nu). \tag{44}$$

We can now write down a complete set of ν linearly independent solutions of (43), namely,

$$g_1 = \begin{pmatrix} f_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad g_2 = \begin{pmatrix} 0 \\ f_2 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad g_\nu = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ f_\nu \end{pmatrix}.$$

The corresponding solutions of

$$h'A = 0,$$

which form the first rows of

$$H(= G^{-1})$$

(see p. 584), may be written

$$h_1 = \begin{pmatrix} e_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad h_2 = \begin{pmatrix} 0 \\ e_2 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad h_\nu = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ e_\nu \end{pmatrix},$$

where

$$e_k = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{m_k}.$$

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

Hence $$\omega_k = h'_k p_0$$

is equal to the sum of those components of p_0 which correspond to the kth subset of states, i.e. *ω_k is the total probability that the system should, at time $t = 0$, be in one of the states of the kth subset.*

The asymptotic distribution (19) is finally given by

$$p_\infty = \begin{pmatrix} \omega_1 f_1 \\ \omega_2 f_2 \\ \vdots \\ w_\nu f_\nu \\ 0 \end{pmatrix}. \tag{46}$$

We see that, unless $\nu = 1$, it can no longer be asserted that the ultimate distribution is independent of the initial conditions, which enter into the ω's (but not into the f's). However, even when $\nu > 1$, the ratio of the asymptotic values of the probabilities of any two states is independent of the initial conditions, provided the two states belong to the same subset. From the structure of the normal form (39) it is to be expected that the last q elements of p_∞ are zero, for the corresponding set of q states has the property that the system can only move out of this set but not into it.

7. *The cofactors of a Minkowski determinant.* We conclude this paper with a formal proof of the assertion made on p. 591 that the components of the vector f_k which satisfies the equation

$$A_k f_k = 0 \tag{47}$$

can always be taken to be non-negative. As the rank of the m_k-rowed matrix A_k is equal to $m_k - 1$, the vector f_k is proportional to any column of the adjoint of A_k. It is evidently sufficient to prove the following theorem, which is related to certain results established by Perron (8), Frobenius (4) and Gersgorin (5):

THEOREM IV*. *If* $a_{ij} \geqslant 0$ $(i \neq j)$ *and*

$$b_j \geqslant a_{1j} + \ldots + a_{j-1,j} + a_{j+1,j} + \ldots + a_{nj} \quad (j = 1, 2, \ldots, n)$$

then all cofactors of

$$D = \begin{vmatrix} b_1 & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & b_2 & \ldots & -a_{2n} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ -a_{n1} & -a_{n2} & \ldots & b_n \end{vmatrix}$$

are non-negative.

Proof. We denote by D_{ij} the determinant of order $n-1$ obtained from D by deleting the ith row and the jth column. The cofactor of the (i,j)th element of D is then

$$A_{ij} = (-1)^{i+j} D_{ij}. \tag{48}$$

The theorem is proved by induction with respect to n.

* I am indebted to the referee for suggesting an improvement in the enunciation of this theorem.

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

When $n = 2$, the cofactors of $\begin{vmatrix} b_1 & -a_{12} \\ -a_{21} & b_2 \end{vmatrix}$

are the elements of the matrix $\begin{pmatrix} b_2 & a_{21} \\ a_{12} & b_1 \end{pmatrix}$

which are all non-negative.

Assume now that $n \geqslant 3$. All principal minors of D are non-negative by the Lemma on p. 586. Hence $A_{ii} \geqslant 0 \quad (i = 1, 2, \ldots, n).$

It only remains to prove that $A_{ij} \geqslant 0$, when $i \neq j$. We begin by considering

$$A_{12} = -D_{12} = -\begin{vmatrix} -a_{21} & -a_{23} & \ldots & -a_{2n} \\ -a_{31} & b_3 & \ldots & -a_{3n} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ -a_{n1} & -a_{n3} & \ldots & b_n \end{vmatrix}.$$

Expanding with respect to the first row and column we find that

$$A_{12} = a_{21}C + \sum_{i,j=3}^{n} C_{ij} a_{i1} a_{2j},$$

where $$C = \begin{vmatrix} b_3 & \ldots & -a_{3n} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ -a_{n3} & \ldots & b_n \end{vmatrix},$$

and where the C_{ij} are the cofactors of C. By the inductive hypothesis, C and C_{ij} are non-negative, whence

$$\frac{\partial D}{\partial(-a_{12})} = A_{12} \geqslant 0. \tag{49}$$

In order to deal with the general case of any cofactor A_{ij} $(i \neq j)$, we simultaneously permute the rows and columns of D in such a way that i goes into 1 and j goes into 2. Let the modified determinant be

$$D' = \begin{vmatrix} b'_1 & \ldots & a'_{1n} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ a'_{n1} & \ldots & b'_n \end{vmatrix}.$$

We have, of course, $D = D', \quad a'_{12} = a_{ij},$

and by the result just obtained $\dfrac{\partial D'}{\partial(-a'_{12})} \geqslant 0.$

Hence $\dfrac{\partial D}{\partial(-a_{ij})} = A_{ij} \geqslant 0.$

This concludes the proof of the theorem.

REFERENCES

(1) Brauer, A. Limits for the characteristic roots of a matrix. *Duke Math. J.* 13 (1946), 387–395.

(2) Doeblin, W. Sur l'équation matricielle $A^{(t+s)} = A^{(t)} A^{(s)}$ et ses applications aux probabilités en chaîne. *Bull. Sci. Math.* 62 (1938), 21–32 and 64 (1940), 35–37.

(3) Fréchet, M. *Méthode des fonctions arbitraires* (*Traité du Calcul des Probabilités et de ses Applications*, ed. by E. Borel), tome 1, fasc. 3, livre 2 (Paris 1938).

(4) Frobenius, G. Über Matrizen aus positiven Elementen. *S.B. Preuss. Akad. Wiss.* (1908), pp. 471–476.

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press

(5) GERSGORIN, S. Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk. S.S.S.R.* 7 (1931), 749–754.
(6) MINKOWSKI, H. Zur Theorie der Einheiten in algebraischen Zahlkörpern. *Göttingen Nachr. ges. Wiss* (1900), 90–93; *Gesammelte Abhandlungen*, 1 (Leipzig, 1911), 316–319.
(7) MISES, R. VON. *Wahrscheinlichkeitsrechnung* (Leipzig, 1931).
(8) PERRON, O. *Algebra II*, 2nd ed. (Berlin, 1933), 35–41.
(9) PRENDIVILLE, B. J. Contribution to discussion: Symposium on stochastic processes. *J. Roy. Statist Soc.* B, 11 (1949) (in the Press).
(10) TAUSSKY, OLGA. A recurring theorem on determinants. *Am. Math. Mon.* 56 (1949), 672–676.
(11) TURNBULL, H. W. and AITKEN, A. C. *Introduction to the theory of canonical matrices* (Glasgow, 1932).

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MANCHESTER

https://doi.org/10.1017/S0305004100026141 Published online by Cambridge University Press