CHAPTER 3

Inhomogeneous Products of Non-negative Matrices

In a number of important applications the asymptotic behaviour as $r \to \infty$ of one of the

$$\text{Forward Products:} \quad T_{p,r} = \{t_{ij}^{(p,r)}\} = H_{p+1}H_{p+2}\cdots H_{p+r}$$

$$\text{Backward Products:} \quad U_{p,r} = \{u_{ij}^{(p,r)}\} = H_{p+r}\cdots H_{p+2}H_{p+1}$$

and its dependence on p is of interest, where $\{H_k, k = 1, 2, \ldots\}$ is a set of $(n \times n)$ matrices satisfying $H_k \geq 0$. We shall write $H_k = \{h_{ij}(k)\}$, $i, j = 1, \ldots, n$. The kinds of asymptotic behaviour of interest are weak ergodicity and strong ergodicity, and a commonly used tool is a contraction coefficient (coefficient of ergodicity). We shall develop the general theory in this chapter. The topic of inhomogeneous products of (row) stochastic matrices has special features, and is for the most part deferred to Chapter 4.

3.1 Birkhoff's Contraction Coefficient: Generalities

Definition 3.1. An $(n \times n)$ matrix $T \geq 0$ is said to be row-allowable if it has at least one positive entry in each row. It is said to be column-allowable if T' is row-allowable. It is said to be allowable if it is both row and column allowable.

In order to introduce Birkhoff's contraction coefficient which will serve as a fundamental tool in this chapter, we need to introduce the quantity,

defined for any two vectors $x' = (x_1, \ldots, x_n) > \mathbf{0}'$, $y' = (y_1, \ldots, y_n) > \mathbf{0}'$, by

$$d(x', y') = \ln \left[\frac{\max_i (x_i/y_i)}{\min_i (x_i/y_i)}\right] = \max_{i,j} \ln \left(\frac{x_i y_j}{x_j y_i}\right).$$

This function has, on the set of $(1 \times n)$ positive vectors, the properties of a metric or distance, with the notable exception that $d(x', y') = 0$ if and only if $x = \lambda y$ for some $\lambda > 0$. (Exercise 3.1). It is a pseudo-metric giving the "projective distance" between $x' > \mathbf{0}'$ and $y' > \mathbf{0}'$. Henceforth we assume as usual that all vectors are of length n and all matrices $(n \times n)$, unless otherwise stated.

It follows that if $x', y' > \mathbf{0}'$ and T is column-allowable, then $x'T, y'T > \mathbf{0}'$; the essence of the contraction property is in the inequality

$$d(x'T, y'T) \leq d(x', y'), \tag{3.1}$$

which we shall establish by recourse to the averaging (contraction) properties of row stochastic matrices in a manner similar to that already employed repeatedly in our Section 2.6.

Lemma 3.1. *If $x, y > \mathbf{0}$ and $\hat{x} = Tx$, $\hat{y} = Ty$, where $T = \{t_{ij}\} \geq 0$ is row-allowable, then*

$$\max_i \left(\frac{\hat{x}_i}{\hat{y}_i}\right) \leq \max_i \left(\frac{x_i}{y_i}\right), \qquad \min_i \left(\frac{\hat{x}_i}{\hat{y}_i}\right) \geq \min_i \left(\frac{x_i}{y_i}\right). \tag{3.2}$$

PROOF:

$$\frac{\hat{x}_i}{\hat{y}_i} = \frac{\sum_j t_{ij} x_j}{\sum_k t_{ik} y_k} = \sum_j p_{ij} \frac{x_j}{y_j}$$

where $p_{ij} = t_{ij} y_j / \sum_k t_{ik} y_k$ is the (i, j) element of a stochastic matrix P. In particular $\sum_j p_{ij} = 1$, so (3.2) follows [so (3.1) follows for a column allowable T, from the definition of $d(\cdot, \cdot)$]. □

We may sharpen the above result by recourse to

Theorem 3.1. *Let $w = \{w_i\}$ be an arbitrary vector and $P = \{p_{ij}\}$ a stochastic matrix. If $z = Pw$, $z = \{z_i\}$, then for any two indices h, h'*

$$z_h - z_{h'} \leq \frac{1}{2} \sum_j |p_{hj} - p_{h'j}| \left\{\max_j w_j - \min_j w_j\right\} \tag{3.3}$$

and

$$\left\{\max_j z_j - \min_j z_j\right\} \leq \tau_1(P)\left\{\max_j w_j - \min_j w_j\right\};$$

or equivalently to the last,

$$\max_{h, h'} |z_h - z_{h'}| \le \tau_1(P)\left\{\max_{j, j'} |w_j - w_{j'}|\right\} \tag{3.4}$$

where

$$\tau_1(P) = \frac{1}{2} \max_{i, j} \sum_{s=1}^{n} |p_{is} - p_{js}| = 1 - \min_{i, j} \sum_{s=1}^{n} \min (p_{is}, p_{js}).$$

Proof: $z_h - z_{h'} = \sum_j u_j w_j$, where $u_j = p_{hj} - p_{h'j}$ (since we are considering h and h' arbitrary but fixed). Let j' denote the indices for which $u_j \ge 0$, and j'' those for which $u_j < 0$, noting that $\sum_j u_j = 1$ (and bearing in mind that the set of j'''s will be empty only if $\boldsymbol{u} = \mathbf{0}$). Put

$$\theta = \sum_{j'} u_{j'} = \sum_{j'} |u_{j'}| = -\sum_{j''} u_{j''} = \sum_{j''} |u_{j''}| = \frac{1}{2} \sum_j |u_j|$$
$$= \frac{1}{2} \sum_j |p_{hj} - p_{h'j}|.$$

Then

$$z_h - z_{h'} = \theta \left\{ \frac{\sum_{j'} |u_{j'}| w_{j'}}{\sum_{j'} |u_{j'}|} - \frac{\sum_{j''} |u_{j''}| w_{j''}}{\sum_{j''} |u_{j''}|} \right\}$$
$$\le \theta \left\{ \max_j w_j - \min_j w_j \right\}$$
$$\le \tau_1(P) \left\{ \max_j w_j - \min_j w_j \right\}.$$

The alternative expression for $\tau_1(P)$ is given as part (equation (2.14)) of Theorem 2.10. □

The preceding result is of most interest in the situation where $\tau_1(P) < 1$ (it is obvious that $0 \le \tau_1(P) \le 1$), and indeed this condition is also that under which the spectral bounding result of Theorem 2.10 becomes of interest for a stochastic matrix. It is clear from the alternative expression for $\tau_1(P)$ that $\tau_1(P) < 1$ if and only if no two rows of P are orthogonal (or, alternatively, any two rows intersect in the sense of having at least one positive element in a coincident position). Such stochastic matrices have been called *scrambling*; we extend this definition to arbitrary non-negative T.

Definition 3.2. A row-allowable matrix $T \ge 0$ is called scrambling if any two rows have at least one positive element in a coincident position.

Lemma 3.2. *If T' is scrambling and $\boldsymbol{x}, \boldsymbol{y} > \mathbf{0}$, then*

$$d(x'T, y'T) < d(x', y'). \tag{3.5}$$

Proof. Referring to the proof of Lemma 3.1, and T' replacing T, we see that the stochastic matrix $P = \{p_{ij}\}$ defined therein is scrambling, so by Theorem 3.1, strict inequality obtains in at least one of the inequalities in (3.2), whence the result follows from the definition of $d(x', y')$. □

Corollary. (3.5) *holds if T has a positive row.*

It follows in view of (3.1) for a column-allowable matrix T and the fact that $d(x', y') = 0$ if and only if $\boldsymbol{x} = \lambda \boldsymbol{y}$ for some positive $\lambda > 0$, that we may define a quantity $\tau_B(T)$ by

$$\tau_B(T) = \sup_{\substack{x, y > 0 \\ x \ne \lambda y}} \frac{d(x'T, y'T)}{d(x', y')}$$

which must then satisfy

$$0 \le \tau_B(T) \le 1. \tag{3.6}$$

Clearly, if T_1 and T_2 are both column-allowable then so is $T_1 T_2$ and for $\boldsymbol{x}, \boldsymbol{y} > \mathbf{0}$, it follows (from Exercise 3.1) that

$$d(x'T_1 T_2, y'T_1 T_2) \le \tau_B(T_2)\, d(x'T_1, y'T_1) \le \tau_B(T_2)\tau_B(T_1)\, d(x', y')$$

whence

$$\tau_B(T_1 T_2) \le \tau_B(T_1)\tau_B(T_2). \tag{3.7}$$

$\tau_B(\cdot)$ is Birkhoff's contraction coefficient (or: coefficient of ergodicity), and properties (3.6) and (3.7) are fundamental to our development of the theory of inhomogeneous products. In view of relation (3.7), we see that if from a sequence $\{H_k\}$ of column-allowable matrices we select the matrices $H_{p+1}, \ldots, H_{p+r}$ and form their product in *any order* and call this product $H_{p,r}$, then still

$$\tau_B(H_{p,r}) \le \prod_{k=p+1}^{p+r} \tau_B(H_k). \tag{3.8}$$

A matrix T will be contractive if $\tau_B(T) < 1$, and, clearly, from (3.8) and (3.6) the significance of a matrix T for which $\tau_B(T) = 0$ is of central significance. We remark that if T is of rank 1 as well as column-allowable, i.e. is of the form $T = \boldsymbol{w}\boldsymbol{v}' = \{w_i v_j\}$ where $\boldsymbol{v} > \mathbf{0}$; $\boldsymbol{w} \ge \mathbf{0}$, $\ne \mathbf{0}$, then from Exercise 3.1

$$\tau_B(T) = \sup_{\substack{x, y > 0 \\ x \ne \lambda y}} \frac{d(x'wv', y'wv')}{d(x', y')} = 0$$

since $d((x'w)v', (y'w)v') = d(v', v') = 0$. To develop further the use of $\tau_B(T)$ we require its explicit form for a column-allowable $T = \{t_{ij}\}$ in terms of the entries of such a matrix. An explicit form is difficult to obtain, and we defer an elementary, but long, derivation to Section 3.4. The form for an *allowable* T is

$$\tau_B(T) = \{1 - [\phi(T)]^{1/2}\}/\{1 + [\phi(T)]^{1/2}\}$$

where

$$\phi(T) = \min_{i,j,k,l} \frac{t_{ik}t_{jl}}{t_{jk}t_{il}} \quad \text{if } T > 0;$$
$$= 0 \quad \text{if } T \not> 0.$$

From this it is clear that given T is allowable, $\tau_B(T) = 0$ if and only if T is of rank 1, i.e. $T = wv'$, $w, v > \mathbf{0}$.

Definition 3.3. The products $H_{p,r} = \{h_{ij}^{(p,r)}\}$ formed from the allowable matrices $H_{p+1}, H_{p+2}, \ldots, H_{p+r}$ multiplied in some specified order for each $p \geq 0$, $r \geq 1$, are said to be weakly ergodic if there exist positive matrices $S_{p,r} = \{s_{ij}^{(p,r)}\}$ $(p \geq 0, r \geq 1)$ each of rank 1 such that for any fixed p, as $r \to \infty$

$$h_{ij}^{(p,r)}/s_{ij}^{(p,r)} \to 1 \quad \text{for all } i, j. \tag{3.9}$$

Lemma 3.3. *The products $H_{p,r}$ are weakly ergodic if and only if for all $p \geq 0$ as $r \to \infty$*

$$\tau_B(H_{p,r}) \to 0. \tag{3.10}$$

PROOF: From the explicit form of $\tau_B(T)$, which implies continuity with $T > 0$, (3.9) evidently implies (3.10). Conversely, define the rank 1 matrices

$$H_{p,r}\mathbf{1}\mathbf{1}'H_{p,r}/\mathbf{1}'H_{p,r}\mathbf{1}$$

(since (3.10) is assumed to hold, since $H_{p,r}$ is allowable, $H_{p,r} > 0$ for sufficiently large r, from the explicit form of $\tau_B(\cdot)$). Then

$$\frac{h_{ij}^{(p,r)}}{s_{ij}^{(p,r)}} = h_{ij}^{(p,r)} \Big/ \left\{ \sum_{k,s} \frac{h_{ik}^{(p,r)}h_{sj}^{(p,r)}}{h_{sk}^{(p,r)}h_{ij}^{(p,r)}} \cdot \frac{h_{sk}^{(p,r)}h_{ij}^{(p,r)}}{\sum_{k,s} h_{ks}^{(p,r)}} \right\}$$
$$\to 1$$

by (3.10), since $\phi(H_{p,r}) \to 1$. □

Lemma 3.3 together with relation (3.7) indicates the power of the coefficient of ergodicity $\tau_B(\cdot)$ as a tool in the study of weak ergodicity of arbitrary products of allowable non-negative matrices. In Lemma 3.4 we shall see that for forward products $T_{p,r}$ the general notion of weak ergodicity defined above coincides with the usual notion for the setting, when $T_{p,r} > 0$ for $r \geq r_0(p)$. We have included Lemma 3.2 because this provides a means of

approaching the problem of weak ergodicity of forward products *without* the requirement that $T_{p,r} > 0$ for $r \geq r_0(p)$ (so that $\tau_B(T_{p,r}) < 1$), although we shall not pursue this topic for products of not necessarily stochastic matrices. Theorem 3.1 which is here used only for the proof of Lemma 3.2 achieves its full force within the setting of products of stochastic matrices.

3.2 Results on Weak Ergodicity

We shall focus in this section on forward products $T_{p,r} = H_{p+1}H_{p+2} \cdots H_{p+r}$ and backward products

$$U_{p,r} = H_{p+r} \cdots H_{p+2}H_{p+1}$$

as $r \to \infty$ since these are the cases of usual interest.

Lemma[1] **3.4.** *If $H_{p,r} = H_{p+1}H_{p+2} \cdots H_{p+r}$, i.e. $H_{p,r}$ is the forward product $T_{p,r} = \{t_{ij}^{(p,r)}\}$ in the previous notation, and all H_k are allowable, then $\tau_B(T_{p,r}) \to 0$ as $r \to \infty$ for each $p \geq 0$ if and only if the following conditions both hold:*

(*a*) $T_{p,r} > 0, \quad r \geq r_0(p)$;

(*b*) $t_{ik}^{(p,r)}/t_{jk}^{(p,r)} \to W_{ij}^{(p)} > 0 \tag{3.11}$

for all i, j, p, k where the limit is independent of k (i.e. the rows of $T_{p,r}$ tend to proportionality *as $r \to \infty$).*

PROOF: The implication: (3.11) ⇒ (3.10) is obvious since under (3.11) clearly $\phi(T_{p,r}) \to 1$. Assume (3.10) obtains; then clearly $T_{p,r} > 0$ for sufficiently large r ($r \geq r_0(p)$, say). Now consider i and j fixed and note that

$$\frac{t_{ik}^{(p,r+1)}}{t_{jk}^{(p,r+1)}} = \sum_s \frac{d_{ks}^{(p,r)}t_{is}^{(p,r)}}{t_{js}^{(p,r)}}$$

where $d_{ks}^{(p,r)} = t_{js}^{(p,r)}h_{sk}(p+r+1)/t_{jk}^{(p,r+1)}$ is the k, s element of a (row) stochastic matrix with strictly positive entries and so a scrambling matrix. Hence by Lemma 3.1

$$\max_k \left(\frac{t_{ik}^{(p,r)}}{t_{jk}^{(p,r)}}\right) \quad \text{is non-increasing with } r;$$

$$\min_k \left(\frac{t_{ik}^{(p,r)}}{t_{jk}^{(p,r)}}\right) \quad \text{is non-decreasing with } r.$$

[1] For the analogous result for backward products see Exercise 3.3. For column-proportionality of forward products see Exercise 3.7.

Since $\tau_B(T_{p,r}) \to 0$, $\phi(T_{p,r}) \to 1$, so as $r \to \infty$

$$\frac{t_{ik}^{(p,r)}}{t_{jk}^{(p,r)}} \frac{t_{js}^{(p,r)}}{t_{is}^{(p,r)}} \to 1$$

for all i, j, k, s, so the two monotone quantities as $r \to \infty$ have the same positive limit, which is independent of k and may be denoted by $W_{ij}^{(p)}$. □

Theorem 3.2. *For a sequence $\{H_k\}$, $k = 1, 2, \ldots$ of non-negative allowable matrices if $H_{p,r} = T_{p,r}$ or $H_{p,r} = U_{p,r}$ then weak ergodicity obtains if and only if there is a strictly increasing sequence of positive integers $\{k_s\}$, $s = 0, 1, 2, \ldots$ such that*

$$\sum_{s=0}^{\infty} [\phi(H_{k_s, k_{s+1}-k_s})]^{1/2} = \infty. \tag{3.12}$$

PROOF: Suppose $H_{p,r} = T_{p,r}$, $p \geq 0$, $r \geq 1$. Take $p = 0$ for simplicity to prove sufficiency of (3.12) and large r (for arbitrary p the argument will be similar).

$$T_{0,r} = T_{0,k_0} T_{k_0, k_1 - k_0} T_{k_1, k_2 - k_1} \cdots T_{k_{t-1}, k_t - k_{t-1}} T^*$$

for some allowable $T^* \geq 0$ where k_t is the nearest member of the sequence $\{k_s\}$ not greater than r, so $t \to \infty$ as $r \to \infty$. Then by (3.7)

$$\tau_B(T_{0,r}) \leq \prod_{s=0}^{t-1} \tau_B(T_{k_s, k_{s+1}-k_s}),$$

and as $r \to \infty$ the right hand side $\to 0$ if and only if

$$\sum_{s=0}^{\infty} \{1 - \tau_B(T_{k_s, k_{s+1}-k_s})\} = \infty.$$

From the definition of $\phi(\cdot)$, $0 \leq \phi(\cdot) \leq 1$, and taking into account the explicit form of $\tau_B(\cdot)$, the divergence of the sum is implied by (3.12). Hence (3.12) is sufficient for weak ergodicity (of forward products).

If we assume weak ergodicity then by Lemma 3.3 $\tau_B(T_{p,r}) \to 0$ as $r \to \infty$, $p \geq 0$. Let $1 > \delta > 0$ be fixed. Then define the sequence $\{k_s\}$ recursively by choosing k_0 arbitrarily, and k_{s+1} once k_s has been determined so that

$$\tau_B(T_{k_s, k_{s+1}-k_s}) \leq \delta.$$

Then

$$[\phi(T_{k_s, k_{s+1}-k_s})]^{1/2} = \left\{\frac{1 - \tau_B(T_{k_s, k_{s+1}-k_s})}{1 + \tau_B(T_{k_s, k_{s+1}-k_s})}\right\}$$
$$\geq (1 - \delta)/(1 + \delta) > 0$$

which implies (3.12) for this sequence.

The proof for backwards products $U_{p,r}$ is analogous. □

Corollary. *If for a sequence $\{k_s\}$, $s \geq 0$, of positive integers such that $k_{s+1} - k_s = g$ (constant),*

$$\phi(T_{k_s, k_{s+1}-k_s}) \geq \varepsilon^2$$

then $\tau_B(T_{p,r}) \to 0$ as $r \to \infty$, $p \geq 0$, at a geometric rate, e.g. in the case $p = 0$:

$$\tau_B(T_{0,r}) \leq \{(1 - \varepsilon)/(1 + \varepsilon)\}^{-(k_0/g)-1}\{(1 - \varepsilon)/(1 + \varepsilon)\}^{r/g}$$

for r sufficiently large. Analogous results hold for backwards products.

Theorem 3.3. *If for the sequence of non-negative allowable matrices $H_k = \{h_{ij}(k)\}$, $k \geq 1$, (i) $H_{p,r_0} > 0$ for $p \geq 0$ where r_0 (≥ 1) is some fixed integer* independent *of p; and (ii):*

$$\min_{i,j}{}^{+} h_{ij}(k) \Big/ \max_{i,j} h_{ij}(k) \geq \gamma > 0 \tag{3.13}$$

(where $\min^+$ refers to the minimum of the positive elements and γ is independent of k), then if $H_{p,r} = T_{p,r}$ (or $= U_{p,r}$), $p \geq 0$, $r \geq 1$, weak ergodicity (at a geometric rate) obtains.

PROOF: From the structure of $\tau_B(\cdot)$ and its dependence on $\phi(\cdot)$ it is evident that the value of $\tau_B(H_{p,r})$ is unchanged if each H_k is multiplied by some positive scalar (each scalar dependent on k). Since by Lemma 3.3 weak ergodicity is dependent only on such values, we may assume without loss of generality in place of (3.13) that

$$0 < \gamma \leq \min_{i,j}{}^{+} h_{ij}(k), \qquad \max_{i,j} h_{ij}(k) \leq 1. \tag{3.14}$$

It follows, since $H_{p,r_0} > 0$, $p \geq 0$, that

$$\gamma^{r_0} \mathbf{1}\mathbf{1}' \leq H_{p,r_0} \leq n^{r_0 - 1} \mathbf{1}\mathbf{1}' \tag{3.15}$$

so $\phi(H_{p,r_0}) \geq (\gamma^{r_0}/n^{r_0-1})^2 = \varepsilon^2$, *say*, $p \geq 0$. We may now apply the Corollary to Theorem 3.2, with $g = r_0$, and $k_0 = 1$, say. □

One may, finally, obtain a result of the nature of Theorem 3.2 and its Corollary for products $H_{p,r}$ formed in arbitrary manner from $H_{p+1}, H_{p+2}, \ldots, H_{p+r}$.

Theorem 3.2′ *For a sequence $\{H_k\}$, $k \geq 1$ of non-negative allowable matrices, if*

$$\sum_{k=1}^{\infty} [\phi(H_k)]^{1/2} = \infty \tag{3.16}$$

then the products $H_{p,r}$ are weakly ergodic.

PROOF: $\tau_B(H_{p,r}) \leq \prod_{k=p+1}^{p+r} \tau_B(H_k) \to 0$ as $r \to \infty$ as in the proof of Theorem 3.2. □

Corollary. *The condition* (3.16) *may be replaced by the condition*

$$\sum_{k=1}^{\infty} \frac{\min_{i,j} h_{ij}(k)}{\max_{i,j} h_{ij}(k)} = \infty.$$

Bibliography and Discussion to §§3.1–3.2

The crucial averaging property of a stochastic matrix, already used in §2.6 and mentioned in its discussion, manifests itself here in Lemma 3.1, and, in more refined manner, in Theorem 3.1. Both these results occur in Markov (1906), the first of Markov's papers to deal with Markov chains (with a finite but arbitrary number of states). Hostinsky (1931, p. 15) calls these results the *Théorème fondamentale sur la limite de la probabilité*, and elsewhere: . . . *l'importante méthode de moyennes successives employée par Markoff* . . . and they were taken up by Fréchet (1938, pp. 25–30), but their potential, in the context of inhomogeneous products of stochastic matrices, was not fully realized until the work of Hajnal (1958). We shall develop these themes further in the more appropriate setting of Chapter 4 where the notion of a scrambling matrix is used extensively.

The Corollary to Lemma 3.2 is due to Cohen (1979a); the lemma itself is a direct consequence of Theorem 3.1. Apart from the results mentioned so far, the development of §§3.1–3.2 largely follows Hajnal (1976). Cohen's subject matter, following on from Hajnal, is the study of $d(x'T_{0,r}, y'T_{0,r})$ as $r \to \infty$ where the matrices $\{H_k\}$ are column allowable inasmuch as each is supposed to have a positive row.

The origins and chief application of the notion of weak ergodicity of forward products $T_{p,r}$, $p \geq 0, r \geq 1$ is in the context of *demography*. A simple demographic model for the evolution of the age structure of a human population, regarded as consisting of n age groups (each consisting of the same number of years), over a set of "time points" $r = 0, 1, 2, \ldots$, (spaced apart by the same time interval as successive age groups) may be described as follows. If $\boldsymbol{\mu}_r$ is an $(n \times 1)$ vector whose components give numbers in various age groups at time r, then

$$\boldsymbol{\mu}'_{r+1} = \boldsymbol{\mu}'_r H_{r+1}, \qquad r = 0, 1, 2, \ldots$$

where H_{r+1} is a known matrix of non-negative entries, depending (in general) on time r, and expressing mortality-fertility conditions at that time. Indeed, each of the matrices H_k, $k \geq 1$, has the same form (has the same incidence matrix); specifically, it is of the form:

$$\begin{bmatrix} b_1 & s_1 & 0 & \cdot & \cdot & \cdot & 0 \\ b_2 & 0 & s_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & & 0 & \ddots & & & \\ \cdot & & & & & & \\ b_{n-1} & & & & & 0 & s_{n-1} \\ b_n & 0 & 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix}$$

which is known as a "Leslie matrix" in this context (and as a "Renewal-type matrix" more generally). Here s_i is the proportion of survivors from age group i to age group $(i + 1)$ in the next time-step, and b_i is the number of contributions to age group 1 per individual in age group i. If we assume all $b_i, s_i > 0$ for all k, then the matrix set H_k, $k \geq 1$, has a rather special structure (in the sense of a common incidence matrix), with, moreover, a positive diagonal entry; see Exercise 3.11; and it is usual to assume that the (coincident) positive entries are bounded away from zero and infinity. Thus the conditions of Theorem 3.3 are certainly satisfied, and its conclusion enables us to make certain inferences about

$$\boldsymbol{\mu}'_r = \boldsymbol{\mu}'_0 H_1 H_2 \cdots H_r = \boldsymbol{\mu}'_0 T_{0,r}$$

where $\boldsymbol{\mu}_0$ $(\geq \mathbf{0}, \neq \mathbf{0})$ represents the initial age structure of the population, and H_k, $k \geq 1$, the history of mortality-fertility pressures over time. Thus consider two different initial population structures: $\boldsymbol{\alpha} = \{\alpha_i\}$, $\boldsymbol{\beta} = \{\beta_i\}$ subjected to the same "history" H_k, $k \geq 1$. Thus from Lemma 3.4, as $r \to \infty$ (first dividing numerator and denominator by $t_{qk}^{(0,r)}$):

$$\frac{\sum_{i=1}^{n} \alpha_i t_{ik}^{(0,r)}}{\sum_{j=1}^{n} \beta_j t_{jk}^{(0,r)}} \to \frac{\sum_{i=1}^{n} \alpha_i W_{iq}^{(0)}}{\sum_{j=1}^{n} \beta_j W_{jq}^{(0)}} = \frac{\sum_{i=1}^{n} \alpha_i w_i^{(0)}}{\sum_{j=1}^{n} \beta_j w_j^{(0)}}$$

(the last step being a consequence of Exercise 3.5), the limit value being independent of k. This independence is called the *weak* ergodicity property, since the $\boldsymbol{\mu}_k$ arising from different $\boldsymbol{\mu}_0$ tend to *proportionality* for large k. If we focus attention on the *age-distribution*, which at time r gives the proportions in the various age groups viz. $\boldsymbol{\mu}'_r/\boldsymbol{\mu}'_r\mathbf{1}$, then this conclusion may be reinterpreted as saying that *the age-distributions tend to coincidence for large r, but the common age structure may still tend to evolve with r*. To see this note as $r \to \infty$

$$\sum_{i=1}^{n} \alpha_i t_{ik}^{(0,r)} \Big/ \sum_{i=1}^{n} \sum_{s=1}^{n} \alpha_i t_{is}^{(0,r)} \sim \left\{ t_{qk}^{(0,r)} \sum_{i=1}^{n} \frac{\alpha_i w_i^{(0)}}{w_q^{(0)}} \right\} \Big/ \left\{ \left(\sum_{s=1}^{n} t_{qs}^{(0,r)} \right) \left(\sum_{i=1}^{n} \frac{\alpha_i w_i^{(0)}}{w_q^{(0)}} \right) \right\}$$

$$\sim t_{qk}^{(0,r)} \Big/ \sum_{s=1}^{n} t_{qs}^{(0,r)}$$

for any fixed $q = 1, \ldots, n$ and all $k = 1, \ldots, n$, and hence is independent of $\boldsymbol{\alpha}$.

Strong ergodicity, discussed in §3.3, is the situation where the common age-distribution tends to constancy as $r \to \infty$.

We may thus call a combination of Theorem 3.3 with Lemma 3.4, the "Weak Ergodicity Theorem". A variant in the demographic literature is generally called the Coale-Lopez theorem, since a proof in this setting was provided by Lopez (1961) along the lines of Exercises 3.9–3.10. This original proof was written under the influence of the approach to the basic ergodic theorem for Markov chains with primitive transition matrix occurring in the book of Kemeny and Snell (1960, pp. 69–70). The present approach (of the text) is much more streamlined, but, insofar as it depends on the contraction form of $\tau_B(\cdot)$ needs to be established for completeness of exposition. The theory of weak ergodicity in a demographic setting begins with the work of Bernardelli (1941), Lewis (1942), and Leslie (1945). For the theory in this setting we refer the reader to Pollard (1973, Chapter 4); see also Cohen (1979b) for a survey and extensive references.

Of interest also is an economic interpretation of the same model (that is, where all H_k are of renewal-type as discussed above), within a paper (Feldstein and Rothschild, 1974, esp. §2), which served as a motivation for the important work of Golubitsky, Keeler and Rothschild (1975) on which much of our §3.3 depends. The vector $\boldsymbol{\mu}_r$ represents, at time r, the amount of capital of each "age" $i = 1, 2, \ldots, n$, where goods last (precisely) n years. A machine, say, provides s_i as much service in its $(i+1)$th year as it did in its ith. Define $b \equiv b(r+1)$ (the "expansion coefficient") as the ratio of gross investment in year r (for year $r+1$) to capital stock in year r. Then each H_k, $k \geq 1$, has the renewal form with each b_i being $b(k+1)$ in H_{k+1}, and all s_i, $i = 1, \ldots, n-1$ being constant with k.

Exercises on §§3.1–3.2

3.1. Show for positive $(1 \times n)$ vectors x', y', z' that for the projective distance $d(\cdot, \cdot)$ defined for positive vectors:

(i) $d(x', y') \geq 0$;
(ii) $d(x', y') = d(y', x')$;
(iii) $d(x', y') \leq d(x', z') + d(z', y')$ [Triangle Inequality];
(iv) $d(x', y') = 0$ if and only if $x = \lambda y$ for some $\lambda > 0$;
(v) $d(x', y') = d(\alpha x', \beta y')$ for any two positive scalars α, β.

3.2. Show that on the set $D^+ = \{x;\ x > \mathbf{0};\ x'\mathbf{1} = 1\}$ the projective distance is a metric. Further, show that if $d(\cdot, \cdot)$ is *any* metric on D^+, and $P \in S^+$, the set of column-allowable stochastic matrices, and $\tau(P)$ is defined by

$$\tau(P) = \sup_{\substack{x,\, y \in D^+ \\ x \neq y}} \frac{d(x'P, y'P)}{d(x', y')}$$

then

(i) $\tau(P_1 P_2) \leq \tau(P_1)\tau(P_2)$, $P_1, P_2 \in S^+$;
(ii) $\tau(P) = 0$ for $P \in S^+$ if and only if $P = \mathbf{1}v'$, $v \in D^+$.
(iii) $\tau(P) = 1$ for $P = I$, the unit matrix.

(Seneta, 1979)

3.3. Check that if T is allowable, then $\tau_B(T) = \tau_B(T')$ and hence state the analogue of Lemma 3.4 for backwards products $U_{p,r} = H_{p+r} \cdots H_{p+1} H_p$.

3.4. If $H_{p,r} = \{h_{ij}^{(p,r)}\}$ is an arbitrary product formed from the allowable matrices $H_{p+1}, H_{p+2}, \ldots, H_{p+r}$ in some specified order, show that the products are weakly ergodic if and only if

(a) $H_{p,r} > 0$, $r \geq r_0(p)$;
(b) $h_{ik}^{(p,r)}/h_{jk}^{(p,r)} \sim W_{ij}^{(p,r)}$, $h_{ki}^{(p,r)}/h_{kj}^{(p,r)} \sim V_{ij}^{(p,r)}$ as $r \to \infty$ for all i, j, p, k, where $V_{ij}^{(p,r)}$, $W_{ij}^{(p,r)}$ are independent of k.

[See also Exercises 3.6–3.7.]

3.5. Show that in Lemma 3.4, $W_{i,j}^{(p)} = \lim_{r \to \infty} t_{ik}^{(p,r)}/t_{jk}^{(p,r)}$ may be written as $w_i^{(p)}/w_j^{(p)}$ for some $\boldsymbol{w}^{(p)} = \{w_i^{(p)}\}$, where $\boldsymbol{w}^{(p)} > \mathbf{0}$, $(\boldsymbol{w}^{(p)})'\mathbf{1} = 1$.

3.6. Proceeding along the lines of Lemma 3.4 show that if $T_{p,r} > 0$, $r \geq r_0(p)$, then

$$\max_j \left(\frac{t_{jq}^{(p,r)}}{t_{js}^{(p,r)}} \right), \qquad \min_j \left(\frac{t_{jq}^{(p,r)}}{t_{js}^{(p,r)}} \right)$$

are, respectively, non-increasing and non-decreasing with r. Hence show that if $\tau_B(T_{p,r}) \to 0$, $r \to \infty$, for an allowable sequence H_k, $k \geq 1$, then both sequences have the same positive limit which (is independent of j and) may be denoted by $V_{qs}^{(p)}$.

3.7. Show that in Exercise 3.6 the limit may be written in the form $V_{qs}^{(p)} = v_q^{(p)}/v_s^{(p)}$ for some $\boldsymbol{v}^{(p)} = \{v_i^{(p)}\}$, where $\boldsymbol{v}^{(p)} > \mathbf{0}$, $(\boldsymbol{v}^{(p)})'\mathbf{1} = 1$. Combine this result with that of Exercise 3.5 to show that if $\tau_B(T_{p,r}) \to 0$, $r \to \infty$, then

$$t_{ij}^{(p,r)}/t_{qs}^{(p,r)} \to w_i^{(p)} v_j^{(p)} / w_q^{(p)} v_s^{(p)} > 0.$$

3.8. Show that for a stochastic matrix $P = \{p_{ij}\}$, $i, j = 1, \ldots, n$

$$\tau_1(P) \leq 1 - \sum_{s=1}^{n} \min_i p_{is} \leq 1 - n\varepsilon$$

where $\varepsilon = \min_{i,j} p_{ij}$. [Thus if $P > 0$, $1 - n\varepsilon < 1$ and there is a simpler bound for $\tau_1(P)$.]

3.9. Suppose H_k, $k \geq 1$, are all positive and satisfy condition (3.13). Using the stochastic matrix with (k, s) entry

$$d_{ks}^{(p,r)} = t_{js}^{(p,r)} \frac{h_{sk}(p+r+1)}{t_{jk}^{(p,r+1)}}$$

$$= h_{sk}(p+r+1) \Big/ \sum_q \left(t_{jq}^{(p,r)}/t_{js}^{(p,r)}\right) h_{qk}(p+r+1)$$

of Lemma 3.4, and Exercise 3.6, show that $\min_{k,s} d_{ks}^{(r)} \geq \gamma^2 n^{-1}$, and hence via Theorem 3.1 and Exercise 3.8 that

$$\left| t_{ik}^{(p,r)}/t_{jk}^{(p,r)} - W_{ij}^{(p)} \right| \leq K_p (1 - \gamma^2)^r, \qquad r \geq 1,$$

for some $W_{ij}^{(p)} > 0$ independent of k, where K_p, $\infty > K_p \geq 0$ is a constant independent of i, j, k.

3.10. Under the conditions of Theorem 3.3 show [without recourse to $\phi(\cdot)$ or $\tau_B(\cdot)$] by using the results of Exercise 3.9 and the weak monotonicity argument in Lemma 3.4 that there are $W_{ij}^{(p)}$ independent of k such that for all i, j, k, p

$$\left| t_{ik}^{(p,r)}/t_{jk}^{(p,r)} - W_{ij}^{(p)} \right| \leq C_p \left(1 - (\gamma^{r_0}/n^{r_0 - 1})^2\right)^{r/r_0}, \qquad r \geq 1,$$

where C_p, $0 \leq C_p < \infty$ is a constant independent of i, j, k. [Note that this form of "geometric rate" of ergodicity differs somewhat in nature from that asserted for this situation by the Corollary to Theorem 3.2.]

(Lopez, 1961; Seneta, 1973)

3.11. If all H_k, $k \geq 1$, have the same incidence matrix which is irreducible and has at least one positive diagonal entry, show that $H_{p,r} > 0$ for $p \geq 0$, $r \geq 2(n-1)$.

3.3 Strong Ergodicity for Forward Products

Definition 3.4. The forward products $T_{p,r} = \{t_{ij}^{(p,r)}\}$ formed from a sequence of row-allowable matrices H_k, $k \geq 1$, are said to be strongly ergodic if for each i, j, p.

$$\frac{t_{ij}^{(p,r)}}{\sum_{s=1}^{n} t_{is}^{(p,r)}} \xrightarrow{r\to\infty} v_j^{(p)} \tag{3.17}$$

independently of i.

Lemma 3.5. *If strong ergodicity obtains, the limit vector $\boldsymbol{v}'_p = \{v_j^{(p)}\}$ of (3.17) is independent of $p \geq 0$.*

PROOF: For any $\boldsymbol{x} \geq \boldsymbol{0}$, $\neq \boldsymbol{0}$, it follows from (3.17) that as $r \to \infty$

$$\boldsymbol{x}'T_{p,r}/\boldsymbol{x}'T_{p,r}\mathbf{1} \to \boldsymbol{v}'_p,$$

whence

$$\boldsymbol{x}'H_{p+1}T_{p+1,r}/\boldsymbol{x}'H_{p+1}T_{p+1,r} \to \boldsymbol{v}'_p.$$

But $\boldsymbol{x}'H_{p+1} \geq \boldsymbol{0}'$, $\neq \boldsymbol{0}'$ (since $\boldsymbol{x}'H_{p+1}\mathbf{1} > 0$ from the row allowability of H_{p+1}) so the limit of the left-hand side is also $\boldsymbol{v}'_{p+1}$, Hence all $\boldsymbol{v}_p$ have a common value, say $\boldsymbol{v}$. Moreover $\boldsymbol{v} \geq \boldsymbol{0}$, $\boldsymbol{v}'\mathbf{1} = 1$, so $\boldsymbol{v}$ is a *probability* vector. □

Definition 3.5. A sequence H_k, $k \geq 1$, of row-allowable matrices is said to be asymptotically homogeneous (with respect to $\boldsymbol{D}$) if there exists a probability vector $\boldsymbol{D}$ (i.e. $\boldsymbol{D} \geq \boldsymbol{0}$, $\boldsymbol{D}'\mathbf{1} = 1$) such that

$$\boldsymbol{D}'H_k/\boldsymbol{D}'H_k\mathbf{1} \xrightarrow{k\to\infty} \boldsymbol{D}'.$$

For the sequel we shall repeatedly make the compactness assumption (3.13) which in the present context (insofar as we consider ratios) may again without loss of generality, be replaced by (3.14). We restate this condition here for convenience

(C) $\qquad 0 < \gamma \leq \min_{i,j}{}^{+} h_{ij}(k), \qquad \max_{i,j} h_{ij}(k) \leq 1$

and call it condition (C). □

Lemma 3.6. *Strong ergodicity of $T_{p,r}$, $p \geq 0$, $r \geq 1$, with limit vector $\boldsymbol{v}$, and condition (C) on the sequence H_k, $k \geq 1$, implies asymptotic homogeneity (with respect to $\boldsymbol{v}$) of the sequence H_k, $k \geq 1$.*

PROOF: Let f_i be the vector with unity in the ith position and zeros elsewhere. Then

$$\frac{f'_iT_{p,r+1}}{f'_iT_{p,r+1}\mathbf{1}} = \left(\frac{f'_iT_{p,r}}{f'_iT_{p,r}\mathbf{1}}\right)\left(\frac{H_{p+r+1}}{\boldsymbol{v}'H_{p+r+1}\mathbf{1}}\right)\rho(r,p,i) \tag{3.18}$$

where $\rho(r, p, i)$ is the scalar given by

$$\rho(r,p,i) = (f'_iT_{p,r}\mathbf{1})(\boldsymbol{v}'H_{p+r+1}\mathbf{1})/f'_iT_{p,r+1}\mathbf{1}.$$

Multiplying (3.18) from the right by $\mathbf{1}$ yields

$$1 = (f'_iT_{p,r}/f'_iT_{p,r}\mathbf{1})(H_{p+r+1}\mathbf{1}/\boldsymbol{v}'H_{p+r+1}\mathbf{1})\rho(r,p,i);$$

and we may write

$$f'_iT_{p,r}/f'_iT_{p,r}\mathbf{1} = \boldsymbol{v}' + \boldsymbol{E}'(r,p,i)$$

where $\boldsymbol{E}(r, p, i) \to \boldsymbol{0}$ as $r \to \infty$ by strong ergodicity. By (C)

$$\rho(r,p,i) \xrightarrow{r\to\infty} 1.$$

Applying this to (3.18) and using similar reasoning, as $r \to \infty$

$$\boldsymbol{v}' \leftarrow \boldsymbol{v}'H_{p+r+1}/\boldsymbol{v}'H_{p+r+1}\mathbf{1} \qquad \text{as required.}$$ □

Theorem 3.4. *If all H_k, $k \geq 1$, are irreducible,*[1] *and condition (C) is satisfied, then asymptotic homogeneity of H_k (with respect to a probability vector $\boldsymbol{D}$) is equivalent to*

$$\boldsymbol{e}_k \xrightarrow{k\to\infty} \boldsymbol{e} \tag{3.19}$$

where $\boldsymbol{e}'_k$ is the positive left Perron-Frobenius eigenvector of H_k normed so that it is a probability vector ($\boldsymbol{e}'_k\mathbf{1} = 1$) and $\boldsymbol{e}$ is a limit vector. In the event that either (equivalent) condition holds, $\boldsymbol{D} = \boldsymbol{e} > \boldsymbol{0}$.

PROOF: Let us assume that the prior conditions and (3.19) hold, Then since by definition of $\boldsymbol{e}_k$,

$$\boldsymbol{e}'_k = \boldsymbol{e}'_kH_k/\boldsymbol{e}'_kH_k\mathbf{1}$$

it follows by condition (C) that as $k \to \infty$

$$\boldsymbol{e}' \leftarrow \boldsymbol{e}'H_k/\boldsymbol{e}'H_k\mathbf{1}$$

so asymptotic homogeneity, with respect to the (necessarily probability) vector $\boldsymbol{e}'$ obtains.

Conversely assume the prior conditions hold and H_k, $k \geq 1$, is asymptotically homogeneous with respect to a probability vector $\boldsymbol{D}'$. Since the set of probability vectors is closed and bounded in R^n, it contains all its limit

[1] Any irreducible matrix is, clearly, allowable.

points; let e be a limit point of a convergent subsequence $\{e_{k_i}\}$ of $\{e_k\}$, so $e_{k_i} \to e$, $i \to \infty$. Then

$$e'_{k_i} = e'_{k_i} H_{k_i} / e'_{k_i} H_{k_i} \mathbf{1}, \qquad i \geq 1. \tag{3.20}$$

Now, by condition (C)

$$\gamma \mathscr{I}_k \leq H_k \leq \mathscr{I}_k, \qquad k \geq 1,$$

where $\mathscr{I}_k$ is the incidence matrix of H_k, and the $\mathscr{I}_k$ are all members of the finite set of all irreducible incidence matrices $\mathscr{I}(j)$, $j = 1, \ldots, t$. Further, the set $[\gamma\mathscr{I}(j), \mathscr{I}(j)] = \{T;\ \gamma\mathscr{I}(j) \leq T \leq \mathscr{I}(j)\}$ is a closed bounded set of R^{n^2}, whence so is

$$Q = \bigcup_{j=1}^{t} [\gamma\mathscr{I}(j), \mathscr{I}(j)]$$

(which contains only irreducible matrices satisfying (C)). Hence referring to (3.20) and taking a subsequence of $\{k_i\}$, $i \geq 1$, if necessary, $H_{k_i} \to H \in Q$, $i \to \infty$, so that

$$e' = e'H/e'H\mathbf{1}.$$

From asymptotic homogeneity, on the other hand

$$D' = D'H/D'H\mathbf{1}.$$

Since H is irreducible, it is readily seen that both e' and D' must be the unique probability-normed left Perron–Frobenius eigenvector of H, so $D = e > \mathbf{0}$, and, further, the sequence $\{e_k\}$ has a unique limit point D, whence $e_k \to D = e$. □

Corollary. *Under the prior conditions of Theorem* 3.4, *if strong ergodicity with limit vector* v *holds, then* (3.19) *holds with* $e = v$. [Follows from Lemma 3.6.]

Lemma 3.7. *Suppose* $y > \mathbf{0}$ *and the sequence* $\{x_m\}$, $m \geq 1$, $x_m > 0$ *each* m, *are probability vectors* (*i.e.* $y'\mathbf{1} = x'_m\mathbf{1} = 1$). *Then as* $m \to \infty$

$$d(x'_m, y') \to 0 \Leftrightarrow x'_m \to y' \qquad (m \to \infty).$$

PROOF: $x'_m \to y' \Rightarrow d(x'_m, y') \to 0$ follows from the explicit form of $d(\cdot, \cdot)$. Conversely suppose $d(x'_m, y') \to 0$. Writing $x'_m = \{x_i^{(m)}\}$, we have from the form of $d(\cdot, \cdot)$ that $y_i x_j^{(m)}/x_i^{(m)} y_j \to 1$, $m \to \infty$ i.e.

$$\ln(y_i/x_i^{(m)}) + \ln(x_j^{(m)}/y_j) \to 0.$$

Now, since the set of all $(1 \times n)$ probability vectors is bounded and closed, there is a subsequence $\{m_k\}$ of the integers such that $x^{(m_k)} \to z$, where z, being a probability vector, has at least one positive entry, say z_{i_0}. Putting $i = i_0$ and $m = m_k$ above, it follows that for any $j = 1, \ldots, n$, $x_j^{(m_k)} \to z_j$, and that $\ln(y_j/z_j) = C = \text{const}$. Thus $y_j = (\exp C) z_j$ and since y' and z' are both probability vectors, $C = 0$, and $y = z$. Hence any limit point of x_m in the sense of pointwise convergence is y, so $x_m \to y$, $m \to \infty$. □

In the following theorem we introduce a new condition, (3.21), which is related to that of the Corollary to Theorem 3.2 and implies the same geometric convergence result for weak ergodicity.

Theorem 3.5. *Assume all* H_k, $k \geq 1$, *are irreducible and satisfy condition* (C); *and*

$$\tau_B(T_{p,r}) \leq \beta < 1 \tag{3.21}$$

for all $r \geq t$ (*for some* $t \geq 1$), *uniformly in* $p \geq 0$. *Then asymptotic homogeneity of* H_k, $k \geq 1$, *is necessary and sufficient for strong ergodicity of the* $T_{p,r}$.

PROOF: (Necessity.) Given strong ergodicity and condition (C) asymptotic homogeneity follows from Lemma 3.6. [Note that neither irreducibility nor (3.21) are needed for this.] (Sufficiency.) We shall only prove strong ergodicity of $T_{p,r}$ in the case $p = 0$ since the argument is invariant under shift of starting point. Consider the behaviour as $r \to \infty$ of the probability vectors $\bar{v}'_r = v'_r/v'_r\mathbf{1}$ where $v'_r = x'T_{0,r}$, $r \geq 1$, for arbitrary fixed $x = v_0 \geq \mathbf{0}, \neq \mathbf{0}$. From Theorem 3.4, $e_k \to e > \mathbf{0}$, so, from Lemma 3.7 it follows that there is an $r_0(\varepsilon) \geq t$ such that $d(e'_r, e') < \varepsilon$ for $r \geq r_0(\varepsilon)$: consider such an r. Then taking into account that for $a \geq 0$, $v_{a+k} > \mathbf{0}$ for any $k \geq t$ since by (3.21) $T_{0,a+k} > 0$, and the properties of $d(\cdot, \cdot)$ [see Exercise 3.1]

$$\begin{aligned} d(\bar{v}'_{r+t}, e') &= d(v'_{r+t}, e') = d(v'_r T_{r,t}, e') \\ &\leq d(v'_r T_{r,t}, e'T_{r,t}) + d(e'T_{r,t}, e'_{r+1} T_{r,t}) + d(e'_{r+1} T_{r,t}, e') \\ &\leq \beta\, d(v'_r, e') + \beta\, d(e', e'_{r+1}) + d(e'_{r+1} T_{r,t}, e') \\ &\leq \beta\, d(\bar{v}'_r, e') + \varepsilon + d(e'_{r+1} T_{r,t}, e') \end{aligned}$$

the β (< 1) arising from (3.21) and the definition of $\tau_B(\cdot)$. Now focussing on the term on the extreme right, since $e'_{r+1} H_{r+1} = e'_{r+1} \rho(H_{r+1})$

$$\begin{aligned} d(e'_{r+1} T_{r,t}, e') &= d(e'_{r+1} T_{r+1,t-1}, e') \\ &\leq d(e'_{r+1} T_{r+1,t-1}, e'T_{r+1,t-1}) + d(e'T_{r+1,t-1}, e'_{r+2} T_{r+1,t-1}) \\ &\quad + d(e'_{r+2} T_{r+1,t-1}, e') \\ &\leq 2\varepsilon + d(e'_{r+2} T_{r+1,t-1}, e'); \\ &\leq 2\varepsilon(t-1) + d(e'_{r+t} T_{r+t-1,1}, e') \\ &\leq \varepsilon(2t-1) \end{aligned}$$

since $T_{r+t-1,1} = H_{r+t}$. Thus from (3.21) for $r \geq r_0(\varepsilon)$,

$$d(\bar{v}'_{r+t}, e') \leq \beta\, d(\bar{v}'_r, e') + 2t\varepsilon,$$

whence

$$d(\bar{v}'_{a+(r+s)t}, e') \leq \beta^s\, d(\bar{v}'_{a+rt}, e') + 2t\varepsilon\left(\frac{1-\beta^s}{1-\beta}\right)$$

so letting $s \to \infty$, $r \to \infty$ yields

$$\lim_{k\to\infty} d(\bar{v}'_{a+kt}, e') = 0 \qquad \text{for arbitrary } a \geq 0.$$

From Lemma 3.7

$$\lim_{k\to\infty} \bar{v}'_{a+kt} = e', \qquad \text{especially for } a = 0, \ldots, t-1.$$

Hence

$$\lim_{r\to\infty} \bar{v}_r = e. \qquad \square$$

Corollary. *If: (i) all H_k, $k \geq 1$, are allowable; (ii) (3.21) holds; and (iii) $e_k \xrightarrow{k\to\infty} e$ for some sequence of left probability eigenvectors $\{e'_k\}$, $k \geq 0$, and for some limit vector $e' > 0'$, then strong ergodicity obtains.*

The following results, culminating in Theorem 3.7, seek to elucidate the nature of the crucial assumption (3.21) be demonstrating within Theorems 3.6 and 3.7 situations which in essence imply it.

Theorem 3.6. *If each H_k, $k \geq 1$, is row-allowable and $H_k \to H$ (elementwise) as $k \to \infty$, where H is primitive, then strong ergodicity obtains, and the limit vector v' is the probability-normed left Perron-Frobenius eigenvector of H.*

Proof: (Again we only prove ergodicity of $T_{p,r}$ in the case $p = 0$.) Let k_0 be such that for $k > k_0$, H_k has positive entries in at least the same positions as H. Let $j_0 \geq 1$ be such that $H^{j_0} > 0$ (recall that H is primitive). Then for $p \geq 0, j \geq j_0$

$$T_{p+k_0,\,j} > 0, \qquad T_{k,\,j} \xrightarrow{k+\infty} H^j.$$

Then for $r \geq 2j_0 + k_0$, and $p \geq 0$

$$T_{p,r} = T_{p,k_0} T_{p+k_0,\,r-j_0-k_0} T_{p+r-j_0,\,j_0} > 0 \tag{3.22}$$

since $T_{p+k_0,\,r-j_0-k_0} > 0$, $T_{p+r-j_0,\,j_0} > 0$ and T_{p,k_0} is row-allowable. In view of property (3.7) of $\tau_B(\cdot)$ and (3.22)

$$\tau_B(T_{p,r}) \leq \tau_B(T_{p+r-j_0,\,j_0}) \tag{3.23}$$

for $r \geq 2j_0 + k_0$. Now $\tau_B(T_{k,j_0}) \to \tau_B(H^{j_0})$ as $k \to \infty$, so for $k \geq \alpha_0$, $\tau_B(T_{k,j_0}) \leq \beta < 1$, since $\tau_B(H^{j_0}) < 1$. Thus for $r \geq 2j_0 + k_0 + \alpha_0$, $= t$ say, from (3.23)

$$\tau_B(T_{p,r}) \leq \beta < 1$$

uniformly in $p \geq 0$. This is condition (3.21).

The proof of sufficiency of Theorem 3.5 is now applicable in the manner encapsuled in the Corollary to that theorem, since H_k, $k \geq t$, are certainly allowable, once we prove that $e_k \to e > 0$, where e'_k is the probability-normed left Perron-Frobenius eigenvalue of the primitive matrix H_k $(k \geq k_0)$, and e' is that of H. We have that

$$e'_k = e'_k H_k / e'_k H_k \mathbf{1}.$$

Let e^* be a limit point of $\{e_k\}$, so for some subsequence $\{k_i\}$ of the integers $e_{k_i} \to e^*$, where e^* must be a probability vector (the set of $(n \times 1)$ probability vectors is bounded and closed). Since $H_{k_i} \to H$, we have

$$(e^*)' = (e^*)'H/(e^*)'H\mathbf{1}$$

so $(e^*)'$ is the unique probability-normed left Perron-Frobenius eigenvector, e' $(> 0')$ of H. Hence $e_k \to e$. [This part has followed the proof of Theorem 3.4.] $\square$

We now denote by M_j the class of non-negative matrices T such that for some k (and hence for all larger k), T_k has its jth column positive. Clearly $\bigcap_{j=1}^{n} M_j$ is the set of all primitive matrices.

We also write $A \sim B$ for two non-negative matrices A and B if they have the *same incidence matrix*, i.e. have zero elements and positive elements in the same positions, so that the "pattern" is the same.

Lemma[1] 3.8. *If A is row-allowable and $AB \sim A$ for a matrix $B \in M_j$ then A has its jth column strictly positive.*

Proof: Since $AB^k \sim A$ for all $k \geq 1$, and AB^k has its jth column positive for some k, A has its jth column positive. $\square$

Corollary. *If B is positive then $A > 0$.*

Lemma 3.9. *If $T_{p,r}$ is primitive, $p \geq 0$, $r \geq 1$, then $T_{p,r} > 0$ for $r \geq t$ where t is the number of primitive incidence matrices.*

Proof: For a fixed p, there are some a, b satisfying $1 \leq a < b \leq t + 1$, such that

$$H_{p+1}H_{p+2}\cdots H_{p+a}H_{p+a+1}\cdots H_{p+b} \sim H_{p+1}H_{p+2}\cdots H_{p+a}$$

since the number of distinct primitive incidence matrices is t. Hence

$$T_{p,a}T_{p+a,\,b-a} \sim T_{p,a}.$$

By the Corollary to Lemma 3.8, $T_{p,a} > 0$, so $T_{p,r} > 0$, $r \geq t$. $\square$

Theorem 3.7. *If $T_{p,r}$, $p \geq 0$, $r \geq 1$, is primitive, and condition (C) holds, asymptotic homogeneity is necessary and sufficient for strong ergodicity.*

[1] We shall use the full force of this lemma only in Chapter 4.

PROOF: Clearly all H_k are primitive, so irreducible; and condition (C) is satisfied. Moreover for $r \geq t$ where t has the meaning of Lemma 3.9

$$\tau_B(T_{p,r}) \leq \tau_B(T_{p,t})$$

by the property (3.7) of $\tau_B(\cdot)$. From condition (C) {analogously to (3.15), since, by Lemma 3.9, $T_{p,t} > 0$}

$$\gamma^t \mathbf{1}\mathbf{1}' \leq T_{p,t} \leq n^{t-1}\mathbf{1}\mathbf{1}'.$$

Since $\tau_B(A)$ clearly varies continuously with $A > 0$, if A varies over the compact set $\gamma^t \mathbf{1}\mathbf{1}' \leq A \leq n^{t-1}\mathbf{1}\mathbf{1}'$, the sup of $\tau_B(A)$, say β, over such A is obtained for some A^* in the set. Thus $A^* > 0$ and $\beta = \tau_B(A^*) < 1$ whence for all $p \geq 0$, $r \geq t$,

$$\tau_B(T_{p,r}) \leq \beta < 1.$$

We can now invoke Theorem 3.5 to obtain the conclusion of that theorem. □

We conclude this section by touching on some results relating to *uniform* strong ergodicity.

Lemma 3.10. *If all H_k, $k \geq 1$, are allowable and* (3.21) *obtains, then*

$$d(\bar{v}_r', \bar{w}_r') \leq K\beta^{r/t}, \qquad r \geq 2t,$$

where $\bar{w}_r' = w_r'/w_r'\mathbf{1}$, $r \geq 1$, with $w_r' = y' T_{0,r}$, $r \geq 1$, for arbitrary $y \equiv w_0 \geq \mathbf{0}$, $\neq \mathbf{0}$, and $\bar{v}_r'$ as in the proof of Theorem 3.5, *with K independent of w_0 and v_0.*

PROOF: Writing $r = a + t + st$, where $a = 0, \ldots, t-1$, $s \geq 1$, with a and s depending on r $(\geq 2t)$, we have

$$d(\bar{v}_r', \bar{w}_r') = d(\bar{v}_{a+t}' T_{a+t,st}, \bar{w}_{a+t}' T_{a+t,st})$$
$$\leq \tau_B(T_{a+t,st})\, d(\bar{v}_{a+t}', \bar{w}_{a+t}')$$

by definition of $\tau_B(\cdot)$;

$$\leq \left\{\prod_{k=1}^{s} \tau_B(T_{a+kt,t})\right\} d(\bar{v}_{a+t}, \bar{w}_{a+t})$$

by (3.7);

$$\leq \beta^s\, d(\bar{v}_{a+t}', \bar{w}_{a+t}') = \beta^{r/t}\{\beta^{-(a/t)-1}\, d(\bar{v}_{a+t}', \bar{w}_{a+t}')\}.$$

Now $\{\beta^{-(a/t)-1}\, d(\bar{v}_0' T_{0,a+t}, \bar{w}_0' T_{0,a+t})\}$ for fixed a is evidently well-defined and continuous in $\bar{v}_0$, $\bar{w}_0'$ (since $T_{0,a+t} > 0$) and these are probability vectors thus varying over a *compact* set. The sup is thus attained and finite; and the final result follows by taking the maximum over $a = 0, \ldots, t-1$. □

Theorem 3.8. *Suppose $\mathscr{A}$ is any set of primitive matrices satisfying condition* (C). *Suppose $e'(H)$ is the left probability Perron–Frobenius eigenvector of*

$H \in \mathscr{A}$. Then for $x \geq \mathbf{0}$, $\neq \mathbf{0}$, if $\bar{v}_0(H) = x/x'\mathbf{1}$ and $\bar{v}_r'(H) = x_r' H^r/x' H^r \mathbf{1}$, $d(\bar{v}_r'(H), e'(H)) = d(x'H^r, e'(H)) \leq K\beta^{r/t}$ for $r \geq 2t$, where $t = n^2 - 2n + 2$, $K > 0$, $0 \leq \beta < 1$, both independent of H and x.

PROOF: For *any* $H \in \mathscr{A}$, $H^r > 0$ for $r \geq t$ by Theorem 2.9, and, from condition (C),

$$\gamma^t \mathbf{1}\mathbf{1}' \leq H^t \leq n^{t-1}\mathbf{1}\mathbf{1}'.$$

Thus for $r \geq t$, by (3.7)

$$\tau_B(H^r) \leq \tau_B(H^t) \leq \beta < 1$$

where β is *independent* of $H \in \mathscr{A}$ being the sup of $\tau_B(A)$ as A varies over the compact set $\gamma^t \mathbf{1}\mathbf{1}' \leq A \leq n^{t-1}\mathbf{1}\mathbf{1}$ (as in Theorem 3.7). Following Lemma 3.10, with $w_0 = e(H)$, we have for $r \geq 2t$

$$d(x'H^r, e'(H)) \leq \beta^{r/t}\{\beta^{-(a/t)-1}\, d(x'H^{a+t}, e'(H))\}$$

and the result follows by taking sup over

$$\{\beta^{-(a/t)-1}\, d(x'H^{a+t}, e'(H))\}$$

as x' and H vary over their respective compact sets (see the proof of Theorem 3.4), and then taking the maximum over $a = 0, 1, 2, \ldots, t-1$. □

Bibliography and Discussion to §3.3

The development of this section in large measure follows Seneta and Sheridan (1981), and owes much to Golubitsky, Keeler and Rothschild (1975, §3). Theorem 3.6 {given with a long direct proof as Theorem 3.5 in Seneta (1973c)}, together with Exercise 3.15, is similar in statement to a peripheral result given by Joffe and Spitzer (1966, pp. 416–417). Lemmas 3.8 and 3.9 have their origins in the work of Sarymsakov (1953a; summary), Sarymsakov and Mustafin (1957), and Wolfowitz (1963). Theorem 3.8 is akin to a result of Buchanan and Parlett (1966); see also Seneta (1973c, §3.3).

The results of §§3.2–3.3, with the exception of Theorem 3.8, may be regarded as attempts to generalize Theorem 1.2 (of Chapter 1) for *powers* of a non-negative matrix to *inhomogeneous products* of non-negative matrices. A great deal of such theory was first developed, also with the aid of "coefficients of ergodicity", for the special situation where all H_k are *stochastic*, in the context of inhomogeneous Markov chains. We shall take up this situation in the next chapter, where, owing to the stochasticity restriction, an easier development is possible. The presentation of the present chapter has, however, been considerably influenced by the framework and concepts of the stochastic situation, and the reader will notice close parallels in the development of the theory. Theorem 3.8 touches marginally on the concept of "ergodic sets": see Hajnal (1976).

In relation to the demographic setting discussed following §3.2, as already noted, the property of *strong* ergodicity for forward products relates to the situation where the common age structure after a long time k ("weak ergodicity") will tend to remain constant as k increases.

Exercises on §3.3

3.12. Show that $\bigcup_{j=1}^{n} M_j = G_1$, the class of $(n \times n)$ nonnegative matrices whose index set contains a single essential class of indices, which is aperiodic. [Recall that M_j is that class of $(n \times n)$ non-negative matrices T such that, for some k, T^k has its jth column positive.]

3.13. Show that if T is scrambling [Definition 3.2], then $T \in G_1$ [as defined in Exercise 3.12]. Construct an example to show that a $T \in G_1$ is not necessarily scrambling.

3.14. By generalizing the proof of sufficiency for (the Corollary of) Theorem 3.5 by leaving $\tau_B(T_{r,t})$ in place of β, show that if:

(*i*) all H_k, $k \geq 1$, are allowable and $T_{p,r} > 0$ for all $r \geq t$ (for some $t \geq 1$);

(*ii*) $\sum_{j=1}^{s-1} \prod_{k=s-j}^{s-1} \tau_B(T_{p+kt,t}) < L < \infty$

uniformly for all $s \geq 2$ and $p \geq 0$; and

(*iii*) $e_k \xrightarrow{k\to\infty} e$ for some sequence of left probability eigenvectors $\{e_k'\}$, $k \geq 0$, and for some limit vector $e' > \mathbf{0}$,

then strong ergocicity obtains. (Seneta & Sheridan, 1981)

3.15. Taking note of the technique of Lemma 3.10, show that under the conditions of the Corollary to Theorem 3.5 (and hence certainly under the conditions of Theorem 3.5),

$$d(\bar{v}_r', e') \to 0$$

uniformly with respect to $v_0 \equiv x \geq \mathbf{0}, \neq \mathbf{0}$.

3.4 Birkhoff's Contraction Coefficient: Derivation of Explicit Form

In this section we show that if $d(x', y')$ is the projective distance between $x' = \{x_i\}$, $y' = \{y_i\} > \mathbf{0}'$, i.e.

$$d(x', y') = \max_{i,j} \ln \left(\frac{x_i y_j}{x_j y_i} \right),$$

then for an *allowable* $T = \{t_{ij}\}$

$$\tau_B(T) \overset{\text{def}}{=} \sup_{\substack{x, y > 0 \\ x \neq \lambda y}} \frac{d(x'T, y'T)}{d(x', y')} = \left\{ \frac{1 - [\phi(T)]^{1/2}}{1 + [\phi(T)]^{1/2}} \right\}$$

where

$$\phi(T) = \min_{i,j,k,l} \frac{t_{ik} t_{jl}}{t_{jk} t_{il}} \quad \text{if } T > 0;$$
$$= 0 \quad \text{if } T \not> 0.$$

To this end we first define two auxiliary quantities, max (x, y), min (x, y).

For any

$$x = \{x_i\} \in R_n, \qquad y = \{y_i\} \geq \mathbf{0}, \qquad \neq \mathbf{0},$$

$$\max\left(\frac{x}{y}\right) = \max_i \left(\frac{x_i}{y_i}\right), \qquad \min\left(\frac{x}{y}\right) = \min_i \left(\frac{x_i}{y_i}\right)$$

where $(x_i/y_i) = \infty$ if for some i, $x_i > 0$ and $y_i = 0$; and $(x_i/y_i) = -\infty$ if for some i, $x_i < 0$ and $y_i = 0$; $(0/0) = 0$.

The following results list certain properties of max $(\cdot, \cdot)$, and min $(\cdot, \cdot)$ necessary for the sequel and serve also to introduce the quantities osc (x/y) and $\theta(x, y)$.

Lemma 3.11.

(*i*) $\max [(x + y)/z] \leq \max (x/z) + \max (y/z)$
$\min [(x + y)/z] \geq \min (x/z) + \min (y/z)$
for any $x, y \in R_n$; $z \geq \mathbf{0}, \neq \mathbf{0}$.

(*ii*) $\max (-x/y) = -\min (x/y)$
$\min(-x/y) = -\max (x/y)$
for any $x \in R_n$; $y > \mathbf{0}, \neq \mathbf{0}$.

(*iii*) $\min (x/y) \leq \max (x/y)$, $x \in R_n$; $y \geq \mathbf{0}$, $\neq \mathbf{0}$
$0 \leq \min (x/y) \leq \max (x/y)$, $x \geq \mathbf{0}$; $y \geq \mathbf{0}$; $\neq \mathbf{0}$.

(*iv*) *If* $\text{osc}\,(x/y) = \max (x/y) - \min (x/y)$, $x \in R_n$; $y \geq \mathbf{0}$, $\neq \mathbf{0}$ [*this is well-defined since* $\max (x, y) > -\infty$ *and* $\min (x, y) < \infty$], *then* $\infty \geq \text{osc}\,(x/y) \geq 0$, *and* $\text{osc}\,(x/y) = 0$ *if and only if*: $x = cy$ *for some* $c \in R$, *and in the case* $c \neq 0$, $y > 0$.

(*v*) $\max (\sigma x/\tau y) = (\sigma/\tau) \max (x/y)$,
$\min (\sigma x/\tau y) = (\sigma/\tau) \min (x/y)$
$x \in R_n$; $y \geq \mathbf{0}$, $\neq \mathbf{0}$; $\tau > 0$, $\sigma \geq 0$.

(*vi*) $\max [(x + cy)/y] = \max (x/y) + c$,
$\min [(x + cy)/y] = \min (x/y) + c$
$\text{osc}\,[(x + cy)/y] = \text{osc}\,(x/y)$,
$x \in R_n$; $y \geq \mathbf{0}$, $\neq \mathbf{0}$; $c \in R$.

(*vii*) $\max (x/y) = [\min (y/x)]^{-1}$, $\qquad x, y \geq \mathbf{0}, \quad \neq \mathbf{0}$.

(*viii*) *If* $x, y, z \geq \mathbf{0}, \neq \mathbf{0}$, *then*
$$\max(x/y) \leq \max(x/z) \cdot \max(z/y),$$
$$\min(x/y) \geq \min(x/z) \cdot \min(z/y).$$
(*ix*) $\max[x/(x+y)] = \max(x/y)/[1 + \max(x/y)] \leq 1,$
$\min[x/(x+y)] = \min(x/y)/[1 + \min(x/y)] \leq 1,$
$x, y \geq \mathbf{0}, \neq \mathbf{0}.$

PROOF: Exercise 3.16. □

Lemma 3.12. *Let* $x, y \geq \mathbf{0}$ *and define for such* x, y
$$\theta(x, y) = \max(x/y)/\min(x/y)$$
(*this is well-defined, since the denominator is finite*). *Then*

(*i*) $\theta(\alpha x, \beta y) = \theta(x, y), \quad \alpha, \beta > 0;$
(*ii*) $\theta(x, y) = \theta(y, x);$
(iii) $\infty \geq \theta(x, y) \geq 1$, *and* $\theta(x, y) = 1$ *if and only if* $x = cy > \mathbf{0}$ *for some* $c > 0$;
(*iv*) $\theta(x, z) \leq \theta(x, y)\theta(y, z) \quad$ if $z \geq \mathbf{0}, \neq \mathbf{0}$.

PROOF: (*i*) follows from Lemma 3.11(*v*). Lemma 3.11(*vii*) yields $\theta(x, y) = \max(x/y)\max(y/x), = \theta(y/x)$, which gives (*ii*). (*iii*) follows from Lemma 3.11(*iii*). (*iv*) follows from Lemma 3.11(*viii*). □

Lemma 3.13. *Suppose* A *is a matrix such that* $z > \mathbf{0} \Rightarrow z'A > \mathbf{0}'$. *Assume* $x \in R_n$, $y > \mathbf{0}$ *and* $0 \leq \operatorname{osc}(x/y) < \infty$. *Then for any* $\varepsilon > 0$,
$$\operatorname{osc}(x'A/y'A) = (\operatorname{osc}(x'/y') + 2\varepsilon)\omega(z_1'A, z_2'A)$$
where for $w, z > \mathbf{0}$
$$\omega(w, z) = \frac{\max(w/z) \cdot \max(z/w) - 1}{(\max(w/z) + 1) \cdot (\max(z/w) + 1)};$$
$$z_1 = x - (\min(x/y) - \varepsilon)y > \mathbf{0},$$
$$z_2 = (\max(x/y) + \varepsilon)y - x > \mathbf{0},$$
so that
$$z_1 + z_2 = y(\operatorname{osc}(x/y) + 2\varepsilon).$$

[*N.B.* We are adopting, for notational convenience, the convention that for any $x, y \in R_n$, and each f, $f(x, y) = f(x', y')$.]

PROOF:
$$\begin{aligned}\operatorname{osc}(x'A/y'A) &= \max(x'A/y'A) - \min(x'A/y'A)\\ &= \{\max([x'A - (\min(x'/y') - \varepsilon) \cdot y'A]/y'A) + \min(x'/y') - \varepsilon\}\\ &\quad - \{\min([x'A - (\min(x'/y') - \varepsilon) \cdot y'A]/y'A) + \min(x'/y') - \varepsilon\}\end{aligned}$$

by Lemma 3.11(*vi*);
$$\begin{aligned}&= ((\operatorname{osc}(x'/y') + 2\varepsilon) \cdot \max([x'A - (\min(x'/y') - \varepsilon) \cdot y'A]/\\ &\quad(\operatorname{osc}(x'/y') + 2\varepsilon) \cdot y'A) - ((\operatorname{osc}(x'/y') + 2\varepsilon)\\ &\quad\cdot \min([x'A - (\min(x'/y') - \varepsilon) \cdot y'A]/(\operatorname{osc}(x'/y') + 2\varepsilon)y'A)\end{aligned}$$
by Lemma 3.11(*v*);
$$\begin{aligned}&= (\operatorname{osc}(x'/y') + 2\varepsilon)\{\max(z_1'A/z_2'A)/(1 + \max(z_1'A/z_2'A))\\ &\quad - \min(z_1'A/z_2'A)/(1 + \min(z_1'A/z_2'A))\}\end{aligned}$$
by Lemma 3.11(*ix*);
$$\begin{aligned}&= (\operatorname{osc}(x'/y') + 2\varepsilon)\{\max(z_1'A/z_2'A)/(1 + \max(z_1'A/z_2'A))\\ &\quad - [1/(1 + \max(z_2'A/z_1'A))]\}\end{aligned}$$
by Lemma 3.11(*vii*);
$$= (\operatorname{osc}(x'/y') + 2\varepsilon)\omega(z_1'A, z_2'A).$$
□

The purpose of Lemma 3.13 was to establish a relation between $\operatorname{osc}(x'/y')$ and $\operatorname{osc}(x'A/y'A)$, which leads to the inequality in Theorem 3.9.

Theorem 3.9. *For* $x \neq \mathbf{0}$ *and* $y > \mathbf{0}$, *such that* $0 < \operatorname{osc}(x/y)$, and A such that $z > \mathbf{0} \Rightarrow z'A > \mathbf{0}'$
$$\operatorname{osc}(x'A/y'A)/\operatorname{osc}(x'/y') \leq (\theta^{1/2}(A) - 1)/(\theta^{1/2}(A) + 1)$$
where $\theta(A) = \sup \theta(w'A, z'A)$, *and this sup is over* $w, z > \mathbf{0}$. [*N.B. If* $\theta(A) = \infty$, *the right-hand side is to be interpreted as unity.*]

PROOF: Since both $\max(w'A/z'A)$, $\max(z'A/w'A) > 0$ (and necessarily finite)
$$\omega(w'A, z'A) = \frac{\max(w'A/z'A) \cdot \max(z'A/w'A) - 1}{[\max(w'A/z'A) \cdot \max(z'A/w'A) + \max(z'A/w'A) + \max(w'A/z'A) + 1]}$$
(from the definition of $\omega(\cdot, \cdot)$)
$$\leq \frac{\max(w'A/z'A) \cdot \max(z'A/w'A) - 1}{[\max(w'A/z'A) \cdot \max(z'A/w'A) + 2\{\max(z'A/w'A) \cdot \max(w'A/z'A)\}^{1/2} + 1]}$$
since if $a, b \geq 0$, $a^2 + b^2 \geq 2ab \geq 0$, and the numerator is ≥ 0 by Lemma 3.11(*iii*) and (*vii*);
$$= \left\{\max\left(\frac{w'A}{z'A}\right) \cdot \max\left(\frac{z'A}{w'A}\right) - 1\right\} \Big/ \left\{\left[\max\left(\frac{w'A}{z'A}\right)\max\left(\frac{z'A}{w'A}\right)\right]^{1/2} + 1\right\}^2$$
$$= \left\{\theta\left(\frac{w'A}{z'A}\right) - 1\right\} \Big/ \left\{\theta^{1/2}\left(\frac{w'A}{z'A}\right) + 1\right\}^2$$

by Lemma 3.11(*vii*) and the definition of $\theta(\cdot,\cdot)$;

$$= \{\theta^{1/2}(w'A/z'A) - 1\}/\{\theta^{1/2}(w'A/z'A) + 1\}.$$

Hence by Lemma 3.13, since $(\alpha - 1)/(\alpha + 1)$ is increasing with $\alpha > 0$,

$$\operatorname{osc}(x'A/y'A)/(\operatorname{osc}(x'/y') + 2\varepsilon) \le \{\theta^{1/2}(A) - 1\}/\{\theta^{1/2}(A) + 1\}.$$

Letting $\varepsilon \to 0+$ yields the result. □

Since it is seen without difficulty (Exercise 3.17), that for an $(n \times n)$ matrix A, $z > \mathbf{0} \Rightarrow z'A > \mathbf{0}$, if and only if A is non-negative and column-allowable, we henceforth use the usual notation for an $(n \times n)$ non-negative matrix, $T = \{t_{ij}\}$.

Lemma 3.14. *If $T > 0$, and f_i denoted the vector with the ith of its n entries unity, and the others zero, then, for all $k, l = 1, \ldots, n$*

$$\sup_{x \ge 0, \ne 0} \left(\frac{x'Tf_k}{x'Tf_l}\right) = \max_i \left(\frac{f_i'Tf_k}{f_i'Tf_l}\right).$$

PROOF: Write $x = \{x_i\} = \sum_i x_i f_i$. Then

$$\sup_{x \ge 0, \ne 0} \left(\frac{x'Tf_k}{x'Tf_l}\right) = \sup_{x \ge 0, \ne 0} \left(\sum_{i=1}^n x_i f_i'Tf_k \Big/ \sum_{i=1}^n x_i f_i'Tf_l\right).$$

Assume without loss of generality that

$$f_1'Tf_k/f_1'Tf_l \ge f_2'Tf_k/f_2'Tf_l \ge \cdots \ge f_n'Tf_k/f_n'Tf_l.$$

Now, if $a, b, c, d > 0$, then $a/b \ge c/d \Leftrightarrow a/b \ge (a + c)/(b + d)$, applying which to the immediately preceding (from the right) yields for any $x \ge \mathbf{0}$, $\ne \mathbf{0}$

$$\frac{f_1'Tf_k}{f_1'Tf_l} \ge \sum_{i=1}^n x_i f_i'Tf_k \Big/ \sum_{i=1}^n x_i f_i'Tf_l$$

with equality in the case $x = f_1$, which is as asserted. □

Corollary. *If $T > 0$,*

$$\sup_{x > 0} \left(\frac{x'Tf_k}{x'Tf_l}\right) = \max_i \left(\frac{f_i'Tf_k}{f_i'Tf_l}\right).$$

Theorem 3.10. *For the possible cases of column-allowable T:*

$$\theta(T) = \max_{i,j,k,l} \left(\frac{t_{ik}t_{jl}}{t_{il}t_{jk}}\right) \quad \text{if } T > 0,$$

$$= \infty \quad \text{if } T \not> 0,\ T \text{ allowable};$$

and if T is column-allowable but not row-allowable,

$$\theta(T) = \infty$$

if and only if there is a row containing both zero and positive elements.[1]

PROOF: Suppose $T > 0$, $w, z > \mathbf{0}$. Then

$$\frac{w'Tf_k}{z'Tf_k} \Big/ \frac{w'Tf_l}{z'Tf_l} = \frac{w'Tf_k \cdot z'Tf_l}{w'Tf_l \cdot z'Tf_k} \le \max_i \frac{f_i'Tf_k}{f_i'Tf_l} \cdot \max_j \frac{f_j'Tf_l}{f_j'Tf_k}$$

by the Corollary to Lemma 3.14;

$$= \max_i \left(\frac{t_{ik}}{t_{il}}\right) \cdot \max_j \left(\frac{t_{jl}}{t_{jk}}\right) = \max_{i,j} \left(\frac{t_{ik}t_{jl}}{t_{il}t_{jk}}\right) \le \max_{i,j,k,l} \left(\frac{t_{ik}t_{jl}}{t_{il}t_{jk}}\right) = \left(\frac{t_{i_0k_0}t_{j_0l_0}}{t_{i_0l_0}t_{j_0k_0}}\right)$$

for some i_0, j_0, k_0, l_0.

Hence

$$\theta(T) = \sup_{w, z > 0} \frac{\max_k (w'Tf_k/z'Tf_k)}{\min_l (w'Tf_l/z'Tf_l)} \le \left(\frac{t_{i_0k_0}t_{j_0l_0}}{t_{i_0l_0}t_{j_0k_0}}\right).$$

On the other hand

$$\theta(T) \ge \frac{w'Tf_{k_0}}{z'Tf_{k_0}} \Big/ \frac{w'Tf_{l_0}}{z'Tf_{l_0}}$$

where $w = \{w_i\} > \mathbf{0}$ with $w_{i_0} = (1 - \delta)$, $\delta > 0$, $w'\mathbf{1} = 1$ and $z = \{z_i\} > \mathbf{0}$ with $z_{j_0} = (1 - \delta)$, $\delta > 0$, $z'\mathbf{1} = 1$; and letting $\delta \to 0+$ yields the required result that

$$\theta(T) = (t_{i_0k_0}t_{j_0l_0}/t_{i_0l_0}t_{j_0k_0}).$$

If T has a row containing both positive and zero elements then for some j, $t_{jk} = 0$, $t_{jh} > 0$ for some k, h. Choose $w = \{w_i\} > \mathbf{0}$ so that $w_j = (1 - \delta)$, $\delta > 0$, $w'\mathbf{1} = 1$, and $z = \mathbf{1}$. Then

$$0 \le \min\left(\frac{w'T}{z'T}\right) \le \frac{\delta \sum_i t_{ik}}{\sum_i t_{ik}} \le \delta$$

[1] What happens when this fails is treated at the conclusion of the proof.

and

$$\max\left(\frac{w'T}{z'T}\right) = \max_s\left(\frac{\sum_i w_i t_{is}}{\sum_i t_{is}}\right) \geq \frac{(1-\delta)t_{jh}}{\sum_i t_{ih}}$$

so $\theta(w'T, z'T) \geq (1-\delta)t_{jh}/\delta$, and the result follows by letting $\delta \to 0+$.

The only case remaining is where all rows are either zero or strictly positive, and there is at least one of each. Then call the $m \times n$ matrix $(1 \leq m < n)$ formed from T by deleting the zero rows $A = \{a_{ij}\}$. By a preceding sequence of arguments

$$\theta(T) = \max_{i,j,k,l}\left(\frac{a_{ik}a_{jl}}{a_{il}a_{jk}}\right). \qquad \square$$

Theorem 3.11. *If T is column-allowable*

$$\sup \operatorname{osc}(x'T/y'T)/\operatorname{osc}(x'/y') = (\theta^{1/2}(T) - 1)/(\theta^{1/2}(T) + 1)$$

where the sup is over $x > \mathbf{0}$, $y > \mathbf{0}$ such that $x \neq cy$. (Interpret the right hand side as 1 if $\theta(T) = \infty$.)

PROOF: Since $\operatorname{osc}(x/y) = 0$ if and only if $x = cy$ (Lemma 3.11), we have $\operatorname{osc}(x/y) > 0$ and are within the framework of Theorem 3.9. If $\theta(T) = 1$, from Theorem 3.10 all rows of T are non-negative multiples of a single positive row, and so $\operatorname{osc}(x'T/y'T) = 0$ for all $x, y > \mathbf{0}$ (Lemma 3.12) and the proposition is established for this case. If $\theta(T) = \infty$, by Theorem 3.10 T has a row, say the jth, containing both positive and zero elements, and by Theorem 3.9

$$\operatorname{osc}(x'T/y'T)/\operatorname{osc}(x'/y') \leq 1$$

$x, y > \mathbf{0}$, $w \neq cz$. Suppose $t_{jk} = 0$ and $t_{jh} > 0$. Choose $y = \lambda f_j + \mathbf{1}$, $\lambda > 0$, and $x = \mathbf{1} - f_j$.

Then $\operatorname{osc}(x'/y') = 1$, while

$$\operatorname{osc}(x'T/y'T) \geq \left(\frac{x'Tf_k}{y'Tf_k}\right) - \left(\frac{x'Tf_h}{y'Tf_h}\right)$$

$$= \left(\frac{\sum_{s\neq j} t_{sk}}{(\lambda t_{jk} + \sum_s t_{sk})}\right) - \left(\frac{\sum_{s\neq j} t_{sh}}{(\lambda t_{jh} + \sum_s t_{sh})}\right)$$

and, since

$$\sum_{s\neq j} t_{sk} = \sum_s t_{sk} > 0,$$

$$= 1 - \left(\frac{\sum_{s\neq j} t_{sh}}{[(1+\lambda)t_{jh} + \sum_{s\neq j} t_{sh}]}\right)$$

and letting $\lambda \to \infty$ yields the result. In this argument $x \not> \mathbf{0}$, but an approximating argument (use $x = \mathbf{1} - (1-\delta)f_j$) will yield the result required.

We now turn to the remaining case $1 < \theta(T) < \infty$, and suppose first (see Theorem 3.10) that $T > 0$. Choose $\varepsilon > 0$ small enough so that $\theta^{1/2}(T) \times$

$(1-\varepsilon) > 1$. From Theorem 3.10, putting $S = \max_i (f_i'Tf_k/f_i'Tf_l)$ and $I = \min_j (f_j'Tf_k/f_j'Tf_l)$, it follows that $\theta(T) = \max_{k,l} S/I$, so there exist k, l such that $S/I > \theta(T)(1-\varepsilon)^2$, and we henceforth consider k and l fixed at these values. We can now find a $\delta > 0$ such that

$$(S-\delta)/(I+\delta) > \theta(T)(1-\varepsilon)^2 > 1.$$

Let

$$M = \{f_i\colon (f_i'Tf_k/f_i'Tf_l) > S - \delta\} \qquad (\neq \phi)$$

$$m = \{f_j\colon (f_j'Tf_k/f_j'Tf_l) < I + \delta\} \qquad (\neq \phi).$$

Clearly $M \cap m = \phi$, since $(S-\delta)/(I+\delta) > 1$. Put $F = \bigcup_{i=1}^n f_i$, $N = F - m$ and $\tilde{N} = F - m - M = N - M$; and if $B \subseteq F$ and $x = \{x_i\} > \mathbf{0}$, then

$$x_B = \sum_{f_i \in B} x_i f_i.$$

Take, along the lines of the preceding argument

$$y = \lambda\mathbf{1}_m + \mathbf{1}_M + \eta\mathbf{1}_{\tilde{N}}$$

$$x = y_N = \mathbf{1}_M + \eta\mathbf{1}_{\tilde{N}}$$

(while $y > \mathbf{0}$, x has some zero elements). Then $\operatorname{osc}(x'/y') = 1$, and we need focus only on $\operatorname{osc}(x'T/y'T)$:

$$\operatorname{osc}(x'T/y'T) \geq (x'Tf_k/y'Tf_k) - (x'Tf_l/y'Tf_l)$$

$$= (y_N'Tf_k/y'Tf_k) - (y_N'Tf_l/y'Tf_l)$$

from the choice of x.

Now

$$(y_m'Tf_k/y_m'Tf_l) = (\mathbf{1}_m'Tf_k/\mathbf{1}_m'Tf_l) < I + \delta,$$

$$(y_M'Tf_k/y_M'Tf_l) = (\mathbf{1}_M'Tf_k/\mathbf{1}_M'Tf_l) > S - \delta$$

since e.g. if $a/b, c/d > \alpha$ for $a, b, c, d, \alpha > 0$, then $(a+c)/(b+d) > \alpha$. Hence, since $N = M \cup \tilde{N}$, $M \cap \tilde{N} = \phi$

$$(y_N'Tf_k/y_N'Tf_l) = [(y_M'Tf_k + y_{\tilde{N}}'Tf_k)/(y_M'Tf_l + y_{\tilde{N}}'Tf_l)]$$

$$= [(\mathbf{1}_M'Tf_k + \eta\mathbf{1}_{\tilde{N}}'Tf_k)/(\mathbf{1}_M'Tf_l + \eta\mathbf{1}_{\tilde{N}}'Tf_l)];$$

$$> S - \delta$$

if η *is chosen sufficiently small*, and then fixed. Thus if we put $t = y_m'Tf_l/y_N'Tf_l$, $\bar{t} = y_m'Tf_k/y_N'Tf_k$, then

$$t/\bar{t} > (S-\delta)/(I+\delta)$$

$$> \theta(T)(1-\varepsilon)^2.$$

Now, since $F = N \cup m$ and $N \cap m = \phi$, and $N \supseteq M \neq \phi$, $m \neq \phi$,

$$\begin{aligned}(y'_N Tf_k/y'Tf_k) - (y'_N Tf_l/y'Tf_l) &= [y'_N Tf_k/(y'_N Tf_k + y'_m Tf_k)] - [y'_N Tf_l/(y'_N Tf_l + y'_m Tf_l)] \\ &= (1+\bar{t})^{-1} - (1+t)^{-1} \\ &> (1+\bar{t})^{-1} - [1 + \bar{t}\theta(T)(1-\varepsilon)^2]^{-1}\end{aligned}$$

in view of the inequality for $t/\bar{t}$;

$$= \frac{[\theta(T)(1-\varepsilon)^2 - 1]}{(1+\bar{t})(\bar{t}^{-1} + \theta(T)(1-\varepsilon)^2]}$$

after simplification.

Further,

$$\begin{aligned}\bar{t} &= y'_m Tf_k/(y'_N Tf_k) = y'_m Tf_k/(y'_M Tf_k + y'_{\bar{N}} Tf_k) \\ &= \lambda \mathbf{1}'_m Tf_k/(\mathbf{1}'_M Tf_k + \eta \mathbf{1}'_{\bar{N}} Tf_k)\end{aligned}$$

so, (η now being fixed) $\lambda > 0$ can still be chosen so that $\bar{t} = \{\theta^{1/2}(T) \times (1-\varepsilon)\}^{-1}$, and then fixed. From the above, for these choices of λ, η,

$$\begin{aligned}\operatorname{osc}(x'T/y'T) &\geq (y'_N Tf_k/y'T f_k) - (y'_N Tf_l/y'Tf_l) \\ &> (1-\bar{t})/(\bar{t}+1) = (\bar{t}^{-1} - 1)/(\bar{t}^{-1} + 1) \\ &= \frac{\theta^{1/2}(T)(1-\varepsilon) - 1}{\theta^{1/2}(T)(1-\varepsilon) + 1}.\end{aligned}$$

If we now replace x by $x + y$ on the left and use Lemma 3.11(vi), we see that

$$\sup_{\substack{x,\, y>0 \\ x \neq cy}} \operatorname{osc}(x'T/y'T) > \left\{\frac{\theta^{1/2}(T)(1-\varepsilon) - 1}{\theta^{1/2}(T)(1-\varepsilon) + 1}\right\}$$

and letting $\varepsilon \to 0+$, together with Theorem 3.9 yields the final result.

The remaining case for the theorem in that where T has only strictly positive and strictly zero rows, and at least one of each. This is tantamount to treating a rectangular matrix $A > 0$ as in Theorem 3.10, and is analogous to the treatment for $T > 0$. □

The following result finally yields the explicit form for $\tau_B(T)$.

Theorem 3.12. *If T is column-allowable,*

$$\tau_B(T) \stackrel{\text{def}}{=} \sup_{\substack{x,\, y>0 \\ x \neq \lambda y}} \frac{d(x'T, y'T)}{d(x', y')} = \left\{\frac{1 - \phi^{1/2}(T)}{1 + \phi^{1/2}(T)}\right\}$$

where $\phi(T) = \theta^{-1}(T)$, $\theta(T)$ having the value specified by Theorem 3.10.

PROOF: For any $c > 0$, since $x \neq \lambda y$, $x + cy \neq \lambda y$, so

$$d[(x+cy)'T, y'T]/d[(x+cy)', y'] \leq \tau_B(T).$$

Since the numerator of the left-hand side is

$$\ln\left(\frac{\max [(x+cy)'T/y'T]}{\min [(x+cy)'T/y'T]}\right), \qquad = \ln\left(\frac{\max (x'T/y'T) + c}{\min (x'T/y'T) + c}\right)$$

by Lemma 3.11(vi);

$$\begin{aligned}&= \ln[1 + c^{-1}\max(x'T/y'T)] \\ &\quad - \ln[1 + c^{-1}\min(x'T/y'T)],\end{aligned}$$

and similarly for the denominator, it follows by letting $c \to \infty$ that

$$\frac{\operatorname{osc}(x'T/y'T)}{\operatorname{osc}(x', y')} = \lim_{c\to\infty} \frac{d[(x+cy)'T, y'T]}{d[(x+cy)', y']} \leq \tau_B(T), \tag{3.24}$$

Next, we note from Theorem 3.9 that

$$\{\max(x'T/y'T) - \min(x'T/y'T)\} \leq \sigma(T)\{\max(x'/y') - \min(x'/y')\}$$

where we have put, for convenience, $\sigma(T) = [1 - \phi^{1/2}(T)]/[1 + \phi^{1/2}(T)]$; that is [by Lemma 3.11(vii)]

$$\frac{\max(y'T/x'T) - \min(y'T/x'T)}{\max(y'T/x'T)\min(y'T/x'T)} \leq \sigma(T)\left\{\frac{\max(y'/x') - \min(y'/x')}{\max(y'/x')\min(y'/x')}\right\}.$$

Replacing y by $ky + x$, $k > 0$, and using Lemma 3.11(vi) and (v)

$$\frac{\max(y'T/x'T) - \min(y'T/x'T)}{[1 + k\max(y'T/x'T)][1 + k\min(y'T/x'T)]} \leq \sigma(T)\left\{\frac{\max(y'/x') - \min(y'/x')}{[1 + k\max(y'/x')][1 + k\min(y'/x')]}\right\}.$$

Integrating both sides in the interval $(0, c)$, $c > 0$, over k, we obtain

$$\ln[1 + c\max(y'T/x'T)] - \ln[(1 + c\min(y'T/x'T)] \leq \sigma(T)\{\ln[1 + c\max(y'/x')] - \ln[1 + c\min(y'/x')]\}$$

i.e.

$$\frac{\ln\{[1 + c\max(y'T/x'T)]/[1 + c\min(y'T/x'T)]\}}{\ln\{[1 + c\max(y'/x')]/[1 + c\min(y'/x')]\}} \leq \sigma(T)$$

and letting $c \to \infty$ yields for $x, y > \mathbf{0}$, $x \neq \lambda y$, that

$$d(y'T, x'T)/d(y', x') \leq \sigma(T)$$

so that

$$\tau_B(T) \leq \sigma(T).$$

But, from (3.24) and Theorem 3.9

$$\sigma(T) \le \tau_B(T)$$

so the required follows. □

To conclude this section we consider the important case where $T = P'$ where P is stochastic (so $\mathbf{1}'T = \mathbf{1}'$ and T is certainly column allowable). This relates directly to the spectrum localization results mentioned in the Bibliography and Discussion §2.5 in relating $\tau_B(T)$ and $\tau_1(T)$ for a non-negative irreducible T, and, not surprisingly, relates to Theorem 3.1.

Theorem 3.13. *If $P = \{p_{ij}\}$ is a stochastic matrix, then $\tau_B(P') \ge \tau_1(P)$ where $\tau_1(P) = \frac{1}{2}\max_{i,j}\sum_{s=1}^{n} |p_{is} - p_{js}|$. In particular, if P is stochastic and allowable, then*

$$\tau_B(P) \ge \tau_1(P).$$

Proof. For $x, y > \mathbf{0}$, $x \ne \lambda y$, by Theorem 3.9 and Theorem 3.12,

$$\tau_B(P') \ge \operatorname{osc}(x'P'/y'P')/\operatorname{osc}(x'/y')$$

and in particular if $y = \mathbf{1}$, for $x > \mathbf{0}$, $x \ne \lambda\mathbf{1}$

$$\tau_B(P') \ge \frac{\operatorname{osc}(Px/\mathbf{1})}{\operatorname{osc}(x/\mathbf{1})} = \left(\frac{\max(Px/\mathbf{1}) - \min(Px/\mathbf{1})}{\max(x/\mathbf{1}) - \min(x/\mathbf{1})}\right)$$

Now, Theorem 3.1 states that certainly the right-hand side is always $\le \tau_1(P)$. We need to tighten this result by proving

$$\tau_1(P) = \sup_{\substack{x > \mathbf{0} \\ x \ne \lambda\mathbf{1}}} \frac{\operatorname{osc}(Px/\mathbf{1})}{\operatorname{osc}(x/\mathbf{1})} \tag{3.25}$$

We shall suppose $\tau_1(P) > 0$; otherwise the theorem is already established. Suppose i_0, j_0 are such that

$$\tau_1(P) = \frac{1}{2}\sum_{s=1}^{n} |p_{i_0 s} - p_{j_0 s}| = \sum_{s \in S} (p_{i_0 s} - p_{j_0 s})$$

where $S = \{s;\ p_{i_0 s} - p_{j_0 s} > 0\} \ne \phi$ and is a proper subset of $\{1, 2, \ldots, n\}$. Let $x = \mathbf{1}_S$; then

$$\tau_1(P) = f'_{i_0} P\mathbf{1}_S - f'_{j_0} P\mathbf{1}_S \le \frac{\operatorname{osc}(P\mathbf{1}_S/P\mathbf{1})}{\operatorname{osc}(\mathbf{1}_S/\mathbf{1})} = \frac{\operatorname{osc}\{P[\delta\mathbf{1} + (1-\delta)\mathbf{1}_S]/P\mathbf{1}\}}{\operatorname{osc}\{[\delta\mathbf{1} + (1-\delta)\mathbf{1}_S]/\mathbf{1}\}}$$

since $P\mathbf{1} = \mathbf{1}$, by Lemma 3.11(v) and (vi), noting that for small $\delta > 0$, $x(\delta) = \delta\mathbf{1} + (1-\delta)\mathbf{1}_S > \mathbf{0}$, $x(\delta) \ne \lambda\mathbf{1}$. Thus (3.25) is established.

The remaining portion of the theorem follows from the fact that for allowable P, $\tau_B(P) = \tau_B(P')$ (Exercise 3.3). □

Bibliography and Discussion to §3.4

The development of this section follows Bauer (1965) up to and including Theorem 3.9. The proof of Theorem 3.11 is essentially due to Hopf (1963). The arguments leading to the two inequalities which comprise the proof of Theorem 3.12 are respectively due to Ostrowski (1964) and Bushell (1973). Theorem 3.13, as already noted in the Bibliography and Discussion to §2.5, is due to Bauer, Deutsch and Stoer (1969). The evaluation of $\tau_B(T)$ was first carried out in a more abstract setting by Birkhoff (1957) [see also Birkhoff (1967)], whose proof relies heavily on projective geometry. The paper of Hopf (1963) was apparently written without knowledge of Birkhoff's earlier work. The section as a whole is based on the synthesis of Sheridan (1979, Chapter 2) of the various approaches from the various settings, carried out by her for the case when T is allowable.

Exercises on §3.4

3.16. Prove Lemma 3.11.

3.17. Suppose A is an $(n \times n)$ real matrix. Show that
(i) $z' \ge \mathbf{0}'$, $\ne \mathbf{0}' \Rightarrow z'A \ge \mathbf{0}'$, $\ne \mathbf{0}'$ if and only if A is non-negative and row-allowable;
(ii) $z' > \mathbf{0}' \Rightarrow z'A > \mathbf{0}'$ if and only if A is non-negative and column-allowable.

3.18. In view of Lemmas 3.11 and 3.12, and Exercise 3.17(i), attempt to develop the subsequent theory for row-allowable T, taking, for example, "sup" directly over $x, y \ge \mathbf{0}$, $\ne \mathbf{0}$ etc.

3.19. Suppose $A = \{a_{ij}\} \ge 0$ is $(m \times n)$ and column-allowable. Define $\theta(A)$ as in Theorem 3.9. Evaluate $\theta(A)$ as in Theorem 3.10, and investigate in general how far the theory of this section and §§3.1–3.2 can be developed for such rectangular A.

CHAPTER 4

Markov Chains and Finite Stochastic Matrices

Certain aspects of the theory of non-negative matrices are particularly important in connection with that class of simple stochastic processes known as Markov chains. The theory of finite Markov chains in part provides a useful illustration of the more widely applicable theory developed hitherto; and some of the theory of countable Markov chains, once developed, can be used as a starting point, as regards ideas, towards an analytical theory of infinite non-negative matrices (as we shall eventually do) which can then be developed without reference to probability notions.

In this chapter, after the introductory concepts, we shall confine ourselves to finite Markov chains, which is virtually tantamount to a study from a certain viewpoint of finite stochastic matrices. We have encountered the notion of a stochastic matrix, central in the subject-matter of this book, as early as §2.5. A number of the ideas on inhomogeneous products of finite non-negative matrices developed in Chapter 3 will also play a prominent role in the context of stochastic matrices. In the next chapter we shall pass to the study of countable Markov chains, which is thus tantamount to a study of stochastic matrices with countable index set, which of course will subsume the finite index set case. Thus this chapter in effect concludes an examination of finite non-negative matrices, and the next initiates our study of the countable case.

We are aware that the general reader may not be acquainted with the simple probabilistic concepts used to initiate the notions of these two chapters. Nevertheless, since much of the content of this chapter and the next is merely a study of the behaviour of stochastic matrices, we would encourage him to persist if he is interested in this last, skipping the probabilistic passages. Chapters 5 and 6 are almost free of probabilistic notions.

Nevertheless, Chapters 4 to 6 are largely intended as an analytical/matrix treatment of the theory of Markov chains, in accordance with the title of this book.

4.1 Markov Chains

Informally, Markov chains (MCs) serve as theoretical models for describing a "system" which can be in various "states", the fixed set of possible states being countable (i.e. finite, or denumerably infinite). The system "jumps" at unit time intervals from one state to another, and the probabilistic law according to which jumps occur is

> "If the system is in the ith state at time $k-1$, the next jump will take it to the jth state with probability $p_{ij}(k)$."

The set of transition probabilities $p_{ij}(k)$ is prescribed for all i, j, k and determines the probabilistic behavior of the system, once it is known how it starts off "at time 0".

A more formal description is as follows. We are given a countable set $\mathcal{S} = \{s_1, s_2, \cdots\}$ or, sometimes, more conveniently $\{s_0, s_1, s_2, \ldots\}$ which is known as the state space, and a sequence of random variables $\{X_k\}$, $k = 0, 1, 2, \ldots$ taking values in $\mathcal{S}$, and having the following *probability property*: if $x_0, x_1, \ldots, x_{k+1}$ are elements of $\mathcal{S}$, then

$$P(X_{k+1} = x_{k+1} \mid X_k = x_k, X_{k-1} = x_{k-1}, \ldots, X_0 = x_0)$$
$$= P(X_{k+1} = x_{k+1} \mid X_k = x_k)$$

if

$$P(X_k = x_k, \ldots, X_0 = x_0) > 0$$

(if $P(B) = 0$, $P(A \mid B)$ is undefined).

This property which expresses, roughly, that future probabilistic evolution of the process is determined once the *immediate past* is known, is the Markov property, and the stochastic process $\{X_k\}$ possessing it is called *a Markov chain.*

Moreover, we call the probability

$$P(X_{k+1} = s_j \mid X_k = s_i)$$

the transition probability from state s_i to state s_j, and write it succinctly as

$$p_{ij}(k+1), \; s_i, s_j \in \mathcal{S}, \qquad k = 0, 1, 2, \ldots$$

Now consider

$$P[X_0 = s_{i_0}, X_1 = s_{i_1}, \ldots, X_k = s_{i_k}].$$

Either *this is positive*, in which case, by repeated use of the Markov property and conditional probabilities it is in fact

$$P[X_k = s_{i_k} \mid X_{k-1} = s_{i_{k-1}}] \cdots P[X_1 = s_{i_1} \mid X_0 = s_{i_0}] P[X_0 = s_{i_0}]$$
$$= p_{i_{k-1}, i_k}(k) p_{i_{k-2}, i_{k-1}}(k-1) \cdots p_{i_0, i_1}(1) \Pi_{i_0}$$

where $\Pi_{i0} = P[X_0 = s_{i0}]$

or it is zero, in which case for some $0 \le r \le k$ (and we take such minimal r)

$$P[X_0 = s_{i_0}, X_1 = s_{i_1}, \ldots, X_r = s_{i_r}] = 0.$$

Considering the cases $r = 0$ and $r > 0$ separately, we see (repeating the above argument), that it is *nevertheless* true that

$$P[X_0 = s_{i_0}, X_1 = s_{i_1}, \ldots, X_k = s_{i_k}] = \Pi_{i_0} p_{i_0, i_1}(1) \cdots p_{i_{k-1}, i_k}(k)$$

since the product of the first $r + 1$ elements on the right is zero. Thus we see that the probability structure of any finite sequence of outcomes is *completely defined* by a knowledge of the *non-negative quantities*

$$p_{ij}(k);\ s_i, s_j \in \mathscr{S}, \qquad \Pi_i;\ s_i \in \mathscr{S}.$$

The set $\{\Pi_i\}$ of probabilities is called the *initial probability distribution* of the chain. We consider these quantities as specified, and denote the row vector of the initial distribution by $\boldsymbol{\Pi}_0'$.

Now, for fixed $k = 1, 2, \ldots$ the matrix

$$P_k = \{p_{ij}(k)\},\ s_i, s_j \in \mathscr{S}$$

is called the *transition matrix* of the MC at time k. It is clearly a square matrix with non-negative elements, and will be doubly infinite if $\mathscr{S}$ is denumerably infinite.

Moreover, its row sums (understood in the limiting sense in the denumerably infinite case) are unity, for

$$\sum_{j \in \mathscr{S}} p_{ij}(k) = \sum_{j \in \mathscr{S}} P[X_k = s_j \mid X_{k-1} = s_i]$$
$$= P[X_k \in \mathscr{S} \mid X_{k-1} = s_i]$$

by the addition of probabilities of disjoint sets;

$$= 1.$$

Thus the matrix P_k is *stochastic.*

Definition 4.1. If $P_1 = P_2 = \cdots = P_k = \cdots$ the Markov chain is said to have stationary transition probabilities or is said to be *homogeneous.* Otherwise it is *non-homogeneous* (or: *inhomogeneous*).

In the homogeneous case we shall refer to the common transition matrix as *the* transition matrix, and denote it by P.

Let us denote by $\boldsymbol{\Pi}_k'$ the row vector of the probability distribution of X_k; then it is easily seen from the expression for a single finite sequence of outcomes in terms of transition and initial probabilities that

$$\boldsymbol{\Pi}_k' = \boldsymbol{\Pi}_0' P_1 \cdots P_k$$

by summing (possibly in the limiting sense) over all sample paths for any fixed state at time k. In keeping with the notation of Chapter 3, we might now adopt the notation

$$T_{p,r} = P_{p+1} P_{p+2} \cdots P_{p+r}$$

and write

$$\boldsymbol{\Pi}_k' = \boldsymbol{\Pi}_0' T_{0,k}.$$

[We digress for a moment to stress that, even in the case of infinite transition matrices, the above products are well defined by the natural extension of the rule of matrix multiplication, and are themselves stochastic. For: let

$$P_\alpha = \{p_{ij}(\alpha)\} \qquad \text{and} \qquad P_\beta = \{p_{ij}(\beta)\}$$

be two infinite stochastic matrices defined on the index set $\{1, 2, \ldots\}$. Define their product $P_\alpha P_\beta$ as the matrix with i, j entry given by the (non-negative) number:

$$\sum_{k=1}^{\infty} p_{ik}(\alpha) p_{kj}(\beta).$$

This sum converges, since the summands are non-negative, and

$$\sum_{k=1}^{\infty} p_{ik}(\alpha) p_{kj}(\beta) \le \sum_{k=1}^{\infty} p_{ik}(\alpha) \le 1$$

since probabilities always take on values between 0 and 1. Further the ith row sum of the new matrix is

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} p_{ik}(\alpha) p_{kj}(\beta) = \sum_{k=1}^{\infty} p_{ik}(\alpha) \left(\sum_{j=1}^{\infty} p_{kj}(\beta) \right)$$
$$= \sum_{k=1}^{\infty} p_{ik}(\alpha) = 1$$

by stochasticity of both P_α and P_β. (The interchange of summations is justified by the non-negativity of the summands.)]

It is also easily seen that for $k > p$

$$\boldsymbol{\Pi}_k' = \boldsymbol{\Pi}_p' T_{p, k-p}.$$

We are now in a position to see why the theory of homogeneous chains is substantially simpler than that of non-homogeneous ones: for then

$$T_{p,k} = P^k$$

so we have only to deal with powers of the common transition matrix P, and further, the probabilistic evolution is *homogeneous in reference to any initial time point p.*

In the remaining section of this chapter we assume that we are dealing with finite ($n \times n$) *matrices as before,* so that the index set is $\{1, 2, \ldots, n\}$ as before (or perhaps, more conveniently, $\{0, 1, \ldots, n-1\}$).

Examples

(1) *Bernoulli scheme.* Consider a sequence of independent trials in each of which a certain event has fixed probability, p, of occurring (this outcome being called a "success") and therefore a probability $q = 1 - p$ of not occurring (this outcome being called a "failure"). We can in the usual way equate success with the number 1 and failure with the number 0; then $\mathscr{S} = \{0, 1\}$, and the transition matrix at any time k is

$$P = \begin{bmatrix} q & p \\ q & p \end{bmatrix}$$

so that we have here a homogeneous 2-state Markov chain. Notice that here the rows of the transition matrix are identical, which must in fact be so for any "Markov chain" where the random variables $\{X_k\}$ are independent.

(2) *Random walk between two barriers.* A particle may be at any of the points $0, 1, 2, 3, \ldots, s$ $(s \geq 1)$ on the x-axis. If it reaches point 0 it remains there with probability a and is reflected with probability $1 - a$ to state 1; if it reaches point s it remains there with probability b and is reflected to point $s - 1$ with probability $1 - b$. If at any instant the particle is at position i, $1 \leq i \leq s - 1$, then at the next time instant it will be at position $i + 1$ with probability p, or at $i - 1$ with probability $q = 1 - p$.

It is again easy to see that we have here a homogeneous Markov chain on the finite state set $\mathscr{S} = \{0, 1, 2, \ldots, s\}$ with transition matrix

$$P = \begin{bmatrix} a & 1-a & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 & 0 \\ & & & & \vdots & & & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1-b & b \end{bmatrix}; \qquad p + q = 1, \quad 0 < p < 1.$$

If $a = 0$, 0 is a *reflecting barrier*, if $a = 1$ it is an *absorbing barrier*, otherwise i.e. if $0 < a < 1$ it is an *elastic barrier*; and similarly for state s.

(3) *Random walk unrestricted to the right.* The situation is as above, except that there is no "barrier" on the right, i.e. $\mathscr{S} = \{0, 1, 2, 3, \ldots\}$ is denumerably infinite, and so is the transition matrix P.

(4) *Recurrent event.* Consider a "recurrent event", described as follows. A system has a variable lifetime, whose length (measured in discrete units) has probability distribution $\{f_i\}$, $i = 1, 2, \ldots$. When the system reaches age $i \geq 1$, it either continues to age, or "dies" and starts afresh from age 0. The movement of the system if its age is $i - 1$ units, $i \geq 2$ is thus to i, with (conditional) probability $(1 - f_1 - \cdots - f_i)/(1 - f_1 - \cdots - f_{i-1})$ or to age 0, with probability $f_i/(1 - f_1 - \cdots - f_{i-1})$. At age $i = 0$, it either reaches age 1 with probability $1 - f_1$, or dies with probability f_1.

We have here a homogeneous Markov chain on the state set $\mathscr{S} = \{0, 1, 2, \ldots\}$ describing the movement of the age of the system. The transition matrix is then the denumerably infinite one:

$$\begin{bmatrix} f_1 & 1-f_1 & 0 & 0 & 0 & \cdots \\ \dfrac{f_2}{1-f_1} & 0 & \dfrac{1-f_1-f_2}{1-f_1} & 0 & 0 & \cdots \\ \dfrac{f_3}{1-f_1-f_2} & 0 & 0 & \dfrac{1-f_1-f_2-f_3}{1-f_1-f_2} & 0 & \cdots \\ & & \vdots & & & \end{bmatrix}.$$

It is customary to specify only that $\sum_{i=1}^{\infty} f_i \leq 1$, thus allowing for the possibility of an infinite lifetime.

(5) *Pólya Urn scheme.* Imagine we have a white and b black balls in an urn. Let $a + b = N$. We draw a ball at random and before drawing the next ball we replace the one drawn, adding also s balls of the same colour.

Let us say that after r drawings the system is in state i, $i = 0, 1, 2, \ldots$ if i is the number of white balls obtained in the r drawings. Suppose we are in state i $(\leq r)$ after drawing number r. Thus $r - i$ black balls have been drawn to date, and the number of white balls in the urn is $a + is$, and the number of black is $b + (r - i)s$. Then at the next drawing we have movement to state $i + 1$ with probability

$$p_{i,\,i+1}(r + 1) = \frac{a + is}{N + rs}$$

and to state i with probability

$$p_{i,\,i}(r + 1) = \frac{b + (r - i)s}{N + rs} = 1 - p_{i,\,i+1}(r + 1).$$

Thus we have here a *non-homogeneous* Markov chain (if $s > 0$) with transition matrix P_k at "time" $k \equiv r + 1 \geq 1$ specified by

$$\begin{aligned} p_{ij}(k) &= \frac{a + is}{N + (k-1)s}, \qquad j = i + 1 \\ &= \frac{b + (k - 1 - i)s}{N + (k-1)s}, \qquad j = i \\ &= \text{otherwise}, \end{aligned}$$

where $\mathscr{S} = \{0, 1, 2, \ldots\}$.

N.B. This example is given here because it is a good illustration of a non-homogeneous chain; the non-homogeneity clearly occurring because of the addition of s balls of colour like the one drawn at each stage. Nevertheless, the reader should be careful to note that this example does not fit into the framework in which we have chosen to work in this chapter, since the matrix P_k is really *rectangular*, viz. $k \times (k+1)$ in this case, a situation which can occur with non-homogeneous chains, but which we omit from further theoretical consideration. Extension in both directions to make each P_k doubly infinite corresponding to the index set $\{0, 1, 2, \ldots\}$ is not necessarily a good idea, since matrix dimensions are equalized at the cost of zero rows (beyond the $(k-1)$th) thus destroying stochasticity.

4.2 Finite Homogeneous Markov Chains

Within this section we are in the framework of the bulk of the matrix theory developed hitherto.

It is customary in Markov chain theory to classify states and chains of various kinds. In this respect we shall remain totally consistent with the classification of Chapter 1.

Thus a chain will be said to be *irreducible*, and, further, *primitive* or *cyclic* (*imprimitive*) according to whether its transition matrix P is of this sort. Further, states of the set

$$\mathscr{S} = \{s_1, s_2, \ldots, s_n\}$$

(or $\{s_0, s_1, \ldots, s_{n-1}\}$) will be said to be *periodic, essential and inessential*, to *lead* one to another, to *communicate*, to form *essential and inessential classes* etc. according to the properties of the corresponding indices of the index set $\{1, 2, \ldots, n\}$ of the transition matrix.

In fact, as has been mentioned earlier, this terminology was introduced in Chapter 1 in accordance with Markov chain terminology. The reader examining the terminology in the present framework should now see the logic behind it.

Irreducible MCs

Suppose we consider an irreducible MC $\{X_k\}$ with (irreducible) transition matrix P. Then putting as usual $\mathbf{1}$ for the vector with unity in each position,

$$P\mathbf{1} = \mathbf{1}$$

by stochasticity of P; so that 1 is an eigenvalue and $\mathbf{1}$ a corresponding eigenvector. Now, since all row sums of P are equal and the Perron–Frobenius eigenvalue lies between the largest and the smallest, 1 is the Perron–Frobenius eigenvalue of P, and $\mathbf{1}$ may be taken as the corresponding right Perron–Frobenius eigenvector. Let $\boldsymbol{v}'$, normed so that $\boldsymbol{v}'\mathbf{1} = 1$, be the corresponding positive left eigenvector. Then we have that

$$\boldsymbol{v}'P = \boldsymbol{v}', \tag{4.1}$$

where $\boldsymbol{v}$ is the column vector of a probability distribution.

Definition 4.2. Any initial probability distribution $\boldsymbol{\Pi}_0$ is said to be *stationary*, if

$$\boldsymbol{\Pi}_0 = \boldsymbol{\Pi}_k, \qquad k = 1, 2, \ldots;$$

and a Markov chain with such an initial distribution is itself said to be stationary.

Theorem 4.1. *An irreducible MC has a unique stationary distribution given by the solution* $\boldsymbol{v}$ *of* $\boldsymbol{v}'P = \boldsymbol{v}'$, $\boldsymbol{v}'\mathbf{1} = 1$.

PROOF. Since

$$\boldsymbol{\Pi}'_{k+1} = \boldsymbol{\Pi}'_k P, \qquad k = 0, 1, 2, \ldots$$

it is easy to see by (4.1) that such $\boldsymbol{v}$ is a stationary distribution. Conversely, if $\boldsymbol{\Pi}_0$ is a stationary distribution

$$\boldsymbol{\Pi}'_0 = \boldsymbol{\Pi}'_0 P, \qquad \boldsymbol{\Pi}_0 \geq \mathbf{0}, \qquad \boldsymbol{\Pi}'_0 \mathbf{1} = 1$$

so that by uniqueness of the left Perron–Frobenius eigenvector of P, $\boldsymbol{\Pi}_0 = \boldsymbol{v}$. □

Theorem 4.2. (*Ergodic Theorem for primitive MCs*). *As* $k \to \infty$, *for a primitive MC*

$$P^k \to \mathbf{1}\boldsymbol{v}'$$

elementwise where $\boldsymbol{v}$ *is the unique stationary distribution of the MC; and the rate of approach to the limit is geometric.*

PROOF. In view of Theorem 4.1, and preceding remarks, this is just a restatement of Theorem 1.2 of Chapter 1 in the present more restricted framework. □

This theorem is extremely important in MC theory for it says that for a primitive MC at least, the probability distribution of X_k, viz. $\boldsymbol{\Pi}'_0 P^k \to \boldsymbol{v}'$, which is *independent* of $\boldsymbol{\Pi}_0$, and the rate of approach is very fast. Thus, after a relatively short time, past history becomes irrelevant, and the chain approaches a stationary regime.[1]

We see, in view of the Perron–Frobenius theory that the analytical (rather that probabilistic) reasons for this are (*i*) $r = 1$, (*ii*) $\boldsymbol{w} = \mathbf{1}$.

We leave here the theory of irreducible chains, which can be further developed without difficulty via the results of Chapter 1.

[1] See also Theorem 4.7 and its following notes.

Reducible Chains with Some Inessential States

We know from Lemma 1.1 of Chapter 1 that there is always *at least one* essential class associated with a finite MC. Let us assume P is in canonical form as in Chapter 1, §1.2, and that Q is the submatrix of P associated with transitions between the inessential states. We recall also that in P^k we have Q^k in the position of Q in P.

Theorem 4.3. $Q^k \to 0$ *elementwise as* $k \to \infty$, *geometrically fast.*

Proof. [We could here invoke the classical result of Oldenburger (1940); however we have tried to avoid this result in the present text, since we have nowhere proved it, and so we shall prove Theorem 4.3 directly. In actual fact, Theorem 4.3 can be used to some extent to replace the need of Oldenburger's result for reducible non-negative matrices.]

Any inessential state leads to an essential state.[1] Let the totality of essential indices of the chain be denoted by E, and of the inessential matrices by I.

We have then that

$$1 - \sum_{j \in I} p_{ij}^{(k)} = \sum_{j \in E} p_{ij}^{(k)} > 0$$

for some k, for any fixed $i \in I$, so that

$$\sum_{j \in I} p_{ij}^{(k)} < 1.$$

Now $\sum_{j \in I} p_{ij}^{(k)}$ is *non-increasing with* k, for

$$\sum_{j \in I} p_{ij}^{(k+1)} = \sum_{j \in I} \sum_{r \in I} p_{ir}^{(k)} p_{rj} \le \sum_{r \in I} p_{ir}^{(k)}.$$

Hence for $k \ge k_0(i)$ and some $k_0(i)$

$$\sum_{j \in I} p_{ij}^{(k)} < \theta(i) < 1$$

and since the number of indices in I is finite, we can say that for $k \ge k_0$, and $\theta < 1$, where k_0 and θ are independent of i,

$$\sum_{j \in I} p_{ij}^{(k)} < \theta < 1, \qquad \text{all } i \in I.$$

Therefore

$$\sum_{j \in I} p_{ij}^{(mk+k)} = \sum_{r \in I} p_{ir}^{(mk)} \sum_{j \in I} p_{rj}^{(k)} \le \theta \sum_{r \in I} p_{ir}^{(mk)}$$

[1] See Exercise 4.11.

for fixed $k \ge k_0$, and each $m \ge 0$ and $i \in I$. Hence

$$\sum_{j \in I} p_{ij}^{[k(m+1)]} \le \theta \sum_{r \in I} p_{ir}^{(mk)} \le \theta^{m+1} \to 0 \qquad \text{as } m \to \infty.$$

Hence a subsequence of

$$\sum_{j \in I} p_{ij}^{(k)}$$

approaches zero; but since this quantity is itself positive and monotone non-increasing with m, it has a limit also, and must have the same limit as the subsequence.

Hence
$$Q^k \mathbf{1} \to \mathbf{0} \qquad \text{as } k \to \infty,$$
and hence
$$Q^k \to 0.$$
□

Now, if the process $\{X_k\}$ passes to an essential state, it will stay forever after in the essential class which contains it. Thus the process cannot ever return to or pass the set I from the essential states, E. Hence if $\mathbf{\Pi}_0(I)$ is that subvector of the initial distribution vector which corresponds to the inessential states we have from the above theorem that

$$P[X_k \subset I] = \mathbf{\Pi}_0'(I) Q^k \mathbf{1} \to 0$$

as $k \to \infty$, which can be seen to imply, in view of the above discussion, that the process $\{X_k\}$ leaves the set I of states in a finite time with probability 1, i.e. the process is eventually "absorbed", with probability 1, into the set E of essential states.

Denote now by E_ρ a specific essential class, ($\bigcup_\rho E_\rho = E$), and let $x_{i\rho}$ be the probability that the process is eventually absorbed into E_ρ, so that

$$\sum_\rho x_{i\rho} = 1,$$

having started at state $i \in I$. Let $x_{i\rho}^{(1)}$ be the probability of absorption after precisely one step, i.e.

$$x_{i\rho}^{(1)} = \sum_{j \in E_\rho} p_{ij}, \qquad i \in I,$$

and let $\mathbf{x}_\rho$ and $\mathbf{x}_\rho^{(1)}$ denote the column vectors of these quantities over $i \in I$.

Theorem 4.4.

$$\mathbf{x}_\rho = [I - Q]^{-1} \mathbf{x}_\rho^{(1)}$$

Proof. First of all we note that since $Q^k \to 0$ as $k \to \infty$ by Theorem 4.3 $[I - Q]^{-1}$ exists by Lemma B.1 of Appendix B (and $= \sum_{k=0}^{\infty} Q^k$ elementwise).

Now let $x_{i\rho}^{(k)}$ be the probability of absorption by time k into E_ρ from $i \in I$. Then the elementary theorems of probability, plus the Markov property enable us to write

$$x_{i\rho}^{(k)} = x_{i\rho}^{(1)} + \sum_{r \in I} p_{ir} x_{r\rho}^{(k-1)} \qquad \text{(Backward Equation)},$$

$$x_{i\rho}^{(k)} = x_{i\rho}^{(k-1)} + \sum_{r \in I} p_{ir}^{(k-1)} x_{r\rho}^{(1)} \qquad \text{(Forward Equation)}.$$

The Forward Equation tells us that

$$(1 \geq) x_{i\rho}^{(k)} \geq x_{i\rho}^{(k-1)}$$

so that $\lim_{k\to\infty} x_{i\rho}^{(k)}$ exists, and it is plausible to interpret this (and it can be rigorously justified) as $x_{i\rho}$.

If we now take limits in the Backward Equation as $k \to \infty$

$$x_{i\rho} = x_{i\rho}^{(1)} + \sum_{r \in I} p_{ir} x_{r\rho},$$

an equation whose validity is intuitively plausible. Rewriting this in matrix terms,

$$\boldsymbol{x}_\rho = \boldsymbol{x}_\rho^{(1)} + Q\boldsymbol{x}_\rho, \qquad (I - Q)\boldsymbol{x}_\rho = \boldsymbol{x}_\rho^{(1)}$$

from which the statement of the theorem follows. □

The matrix $[I - Q]^{-1}$ plays a vital role in the theory of finite absorbing chains (as does its counterpart in the theory of transient infinite chains to be considered in the next chapter) and it is sometimes called the *fundamental matrix* of absorbing chains. We give one more instance of its use.

Let Z_{ij} be the number of visits to state $j \in I$ starting from $i \in I$. ($Z_{ii} \geq 1$). Then

$$Z_i = \sum_{j \in I} Z_{ij}, \qquad i \in I$$

is the time to absorption of the chain starting from $i \in I$. Let $m_{ij} = \mathscr{E}(Z_{ij})$ and $m_i = \mathscr{E}(Z_i)$ be the expected values of Z_{ij} and Z_i respectively, and $M = \{m_{ij}\}_{i,\, j \in I}$, and $\boldsymbol{m} = \{m_i\}$.

Theorem 4.5.

$$M = (I - Q)^{-1}$$

$$\boldsymbol{m} = M\mathbf{1} = (I - Q)^{-1}\mathbf{1}.$$

PROOF. Recall that $Q^0 = I$ by definition. Let $Y_{ij}^{(k)} = 1$ if $X_k = j$, $Y_{ij}^{(k)} = 0$ if $X_k \neq j$, the process $\{X_k\}$ having started at $i \in I$. Then

$$\mathscr{E}(Y_{ij}^{(k)}) = p_{ij}^{(k)} \cdot 1 + (1 - p_{ij}^{(k)}) \cdot 0$$
$$= p_{ij}^{(k)}, \qquad k \geq 0.$$

Moreover

$$Z_{ij} = \sum_{k=0}^{\infty} Y_{ij}^{(k)}$$

the sum on the right being effectively finite for any realization of the process, since absorption occurs in finite time. By positivity

$$m_{ij} = \mathscr{E}(Z_{ij}) = \sum_{k=0}^{\infty} \mathscr{E}(Y_{ij}^{(k)}) = \sum_{k=0}^{\infty} p_{ij}^{(k)},$$

$i, j \in I$. Thus

$$M = \sum_{k=0}^{\infty} Q^k \qquad \text{elementwise}$$
$$= (I - Q)^{-1} \qquad \text{(Lemma B.1 of Appendix B)}$$

and since

$$Z_i = \sum_{j \in I} Z_{ij},$$

it follows that

$$m_i = \sum_{j \in I} m_{ij}.$$ □

Finally, in connection with the fundamental matrix, the reader may wish to note that, in spite of the elegant matrix forms of Theorems 4.4 and 4.5, it may still be easier to solve the corresponding linear equations for the desired quantities in actual problems. These are

$$x_{i\rho} = x_{i\rho}^{(1)} + \sum_{r \in I} p_{ir} x_{r\rho}, \qquad i \in I. \qquad \text{(Theorem 4.4)}$$

$$m_i = 1 + \sum_{r \in I} p_{ir} m_r, \qquad i \in I. \qquad \text{(Theorem 4.5)}$$

and we shall do so in the following example.

EXAMPLE: (Random walk between two absorbing barriers). (See §4.1). Here there are two essential classes E_0, E_s consisting of one state each (the absorbing barriers). The inessential states are $I = \{1, 2, \ldots, s-1\}$ where we assume $s > 1$, and the matrix Q is given by

$$Q = \begin{bmatrix} 0 & p & 0 & \cdots & 0 & 0 \\ q & 0 & p & \cdots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & q & 0 \end{bmatrix}$$

with $x_{i0}^{(1)} = \delta_{i1} q$, $x_{is}^{(1)} = \delta_{i,\, s-1} p$, $i \in I$, δ_{ij} being the Kronecker delta.

(a) *Probability of eventual absorption into E_s*. We have that

$$x_{is} = x_{is}^{(1)} + \sum_{r \in I} p_{ir} x_{rs}, \qquad i = 1, 2, \ldots, s-1,$$

i.e.

$$\begin{aligned} x_{1s} &= \qquad\qquad p x_{2s} \\ x_{2s} &= q x_{1s} + p x_{3s} \\ &\ \vdots \\ x_{s-2,s} &= q x_{s-3,s} + p x_{s-1,s} \\ x_{s-1,s} &= p + q x_{s-2,s}. \end{aligned}$$

Write for convenience $x_i \equiv x_{is}$. Then if we define $x_0 = 0$, $x_s = 1$, the above equations can be written in unified form as

$$\begin{cases} x_i = q x_{i-1} + p x_{i+1}, & i = 1, 2, \ldots, s-1 \\ x_0 = 0, \quad x_s = 1 \end{cases}$$

and it is now a matter of solving this difference equation under the stated boundary assumptions.

The general solution is of the form

$$\begin{aligned} x_i &= A z_1^i + B z_2^i \qquad \text{if } z_1 \neq z_2 \\ &= (A + Bi) z^i \qquad \text{if } z_1 = z_2 = z \end{aligned}$$

where z_1 and z_2 are the solutions of the characteristic equation

$$pz^2 - z + q = 0$$

viz., $z_1 = 1$, $z_2 = q/p$ bearing in mind that $1 - 4pq = (p - q)^2$.

Hence:

(*i*) if $q \neq p$, we get, using boundary conditions to fix A and B:

$$x_i = \{1 - (q/p)^i\}/\{1 - (q/p)^s\}, \qquad i \in I.$$

(*ii*) if $q = p = \frac{1}{2}$

$$x_i = i/s, \qquad i \in I.$$

(*b*) *Mean time to absorption*. We have that

$$m_i = 1 + \sum_{r \in I} p_{ir} m_r, \qquad i \in I$$

i.e.

$$\begin{aligned} m_1 &= 1 \qquad\quad + p m_2 \\ m_2 &= 1 + q m_1 + p m_3 \\ &\ \vdots \\ m_{s-1} &= 1 + q m_{s-2}. \end{aligned}$$

Hence we can write in general

$$\begin{cases} m_i = 1 + q m_{i-1} + p m_{i+1}, & i = 1, 2, \ldots, s-1 \\ m_0 = 0, \quad m_s = 0. \end{cases}$$

We have here to deal with an inhomogeneous difference equation, the homogeneous part of which is as before, so that the general solution to it is as before plus a particular solution to the inhomogeneous equation. It can be checked that

(*i*) $q \neq p$; $i/(q - p)$ is a particular solution, and that taking into account boundary conditions

$$m_i = i/(q-p) - \{s/(q-p)\}\{1 - (q/p)^i\}/\{1 - (q/p)^s\}, \qquad i \in I.$$

(*ii*) $q = p = \frac{1}{2}$; $-i^2$ is a particular solution and hence

$$m_i = i(s - i), \qquad i \in I.$$

The following theorem is analogous to Theorem 4.2 in that when attention is focused on behavior within the set I of *inessential states*, under similar structural conditions on the set I, then a totally analogous result obtains.

Theorem 4.6. *Let Q, the submatrix of P corresponding to transitions between the inessential states of the MC corresponding to P, be primitive, and let there be a positive probability of $\{X_k\}$ beginning in some $i \in I$. Then for $j \in I$ as $k \to \infty$*

$$P[X_k = j \mid X_k \in I] \to v_j^{(2)} \Big/ \sum_{j \in I} v_j^{(2)}$$

where $v^{(2)} = \{v_j^{(2)}\}$ is a positive vector independent of the initial distribution, and is, indeed, the left Perron–Frobenius eigenvector of Q.

Proof. Let us note that if $\mathbf{\Pi}_0$ is that part of the initial probability vector restricted to the initial states then

$$P[X_k \in I] = \mathbf{\Pi}_0' Q^k \mathbf{1}, > 0$$

since by primitivity $Q^k > 0$ for k large enough, and $\mathbf{\Pi}_0 \neq \mathbf{0}$, Moreover the vector of the quantities

$$P[X_k = j \mid X_k \in I], \qquad j \in I$$

is given by

$$\mathbf{\Pi}_0' Q^k / \mathbf{\Pi}_0' Q^k \mathbf{1}.$$

The limiting behaviour follows on letting $k \to \infty$ from Theorem 1.2 of Chapter 1, the contribution of the right Perron–Frobenius eigenvector dropping out between numerator and denominator. □

Chains Whose Index Set Contains a Single Essential Class

On account of the extra stochasticity property inherent in non-negative matrices P which act as transition matrices of Markov chains, the properties of irreducible non-negative matrices (to an extent) hold for stochastic matrices which, apart from a single essential class (which gives rise to a stochastic irreducible submatrix, P_1), may also contain some inessential indices. The reason, in elementary terms, is that if P is written in canonical form (as in the preceding discussion):

$$P = \begin{bmatrix} P_1 & 0 \\ R & Q \end{bmatrix}$$

then $Q^k \to 0$ elementwise (in accordance with Theorem 4.3). P_1^k exhibits the behaviour of an irreducible matrix in accordance with Chapter 1, with the simplifications due to the stochasticity of P_1. The effect of P_1 thus dominates that of Q; the concrete manifestation of this will become evident in subsequent discussion.

Firstly (compare Theorem 4.1) a corresponding MC has a unique stationary distribution, which is essentially the stationary distribution corresponding to P_1. For, an $n \times 1$ vector $\boldsymbol{v} = (\boldsymbol{v}_1', \mathbf{0}')'$ where $\boldsymbol{v}_1$ is the unique stationary distribution corresponding to P_1, the chain being assumed in canonical form, is clearly a stationary distribution of P; and suppose *any* vector $\boldsymbol{\Pi}$ satisfying $\boldsymbol{\Pi}'P = \boldsymbol{\Pi}'$, $\boldsymbol{\Pi}'\mathbf{1} = 1$ is correspondingly partitioned, so that $\boldsymbol{\Pi}' = \{\boldsymbol{\Pi}_1', \boldsymbol{\Pi}_2'\}$. Then

$$\boldsymbol{\Pi}_1' P_1 + \boldsymbol{\Pi}_2' R = \boldsymbol{\Pi}_1'$$
$$\boldsymbol{\Pi}_2' Q = \boldsymbol{\Pi}_2'$$

From the second of these it follows that $\boldsymbol{\Pi}_2 Q^k = \boldsymbol{\Pi}_2$, so, by Theorem 4.3, $\boldsymbol{\Pi}_2 = \mathbf{0}$; so, from the first equation $\boldsymbol{\Pi}_1' P_1 = \boldsymbol{\Pi}_1'$; and $\boldsymbol{\Pi}_1'\mathbf{1} = 1$, whence $\boldsymbol{\Pi}_1 = \boldsymbol{v}_1$. In particular $\boldsymbol{v}$ is the unique stationary distribution.

It is evident[1] that an MC which contains at least two essential classes will not have a single stationary distribution, and hence chains with a single such class may be characterized as having a single stationary distribution. This vector is the unique solution $\boldsymbol{\Pi}$ to the linear equation system

$$\boldsymbol{\Pi}'\{\mathbf{1}, I - P\} = \{1, \mathbf{0}'\}$$

where the matrix $\{\mathbf{1}, I - P\}$ is $n \times (n+1)$. This uniqueness implies this matrix is of rank n, and hence contains n linearly independent columns. The last n columns are linearly dependent, since $(I - P)\mathbf{1} = \mathbf{0}$; but the vector $\mathbf{1}$ combined with any $(n-1)$ columns of $I - P$ clearly gives a linearly independent set.

It follows from the above discussion that for an MC containing a single essential class of indices and transition matrix P, any $(n-1)$ of the equations $\boldsymbol{\Pi}'P = \boldsymbol{\Pi}'$ are sufficient to determine the stationary distribution vector to a constant multiple and the additional condition $\boldsymbol{\Pi}'\mathbf{1} = 1$ then specifies it completely. The resulting $(n \times n)$ linear equation system may be used for the practical calculation of the stationary distribution, which will have zero entries corresponding to any inessential states.

[1] See Exercise 4.12.

To develop further the theory of such Markov chains, we define at this stage a regular[1] stochastic matrix, and hence a regular Markov chain as one with a regular transition matrix. This notion will play a major role in the remainder of this chapter.

Definition 4.3. An $n \times n$ stochastic matrix is said to be *regular* if its essential indices form a single essential class, which is aperiodic.

Theorem 4.7. *Let P be the transition matrix of a regular MC, in canonical form, and $\boldsymbol{v}_1'$ the stationary distribution corresponding to the primitive submatrix P_1 of P corresponding to the essential states. Let $\boldsymbol{v}' = (\boldsymbol{v}_1', \mathbf{0}')$ be an $1 \times n$ vector. Then as $k \to \infty$*

$$P^k \to \mathbf{1}\boldsymbol{v}'$$

elementwise, where $\boldsymbol{v}'$ is the unique stationary distribution corresponding to the matrix P, the approach to the limit being geometrically fast.

PROOF. Apart from the limiting behaviour of $p_{ij}^{(k)}$, $i \in I, j \in E$, this theorem is a trivial consequence of foregoing theory, in this and the preceding section.

If we write

$$P^k = \begin{bmatrix} P_1^k & 0 \\ R_k & Q^k \end{bmatrix}$$

it is easily checked (by induction, say) that putting $R_1 = R$

$$R_{k+1} = \sum_{i=0}^{k} Q^i R P_1^{k-i} = \sum_{i=0}^{k} Q^{k-i} R P_1^i$$

so that we need to examine this matrix as $k \to \infty$.

Put $M = P_1 - \mathbf{1}\boldsymbol{v}_1'$;

then
$$M^i = P_1^i - \mathbf{1}\boldsymbol{v}_1'.$$

Now from Theorem 4.2 we know that each element of M^i is dominated by $K_1 \rho_1^i$, for some $K_1 > 0$, $0 < \rho_1 < 1$, independent of i, for every i.

Moreover

$$R_{k+1} = \sum_{i=0}^{k} Q^{k-i} R \mathbf{1}\boldsymbol{v}_1' + \sum_{i=0}^{k} Q^{k-i} R M^i$$

and we also know from Theorem 4.3 that each element of Q^i is dominated by $K_2 \rho_2^i$ for some $K_2 > 0$, $0 < \rho_2 < 1$ independent of i, for every i. Hence each component of the right hand sum matrix is dominated by

$$K_3 \sum_{i=0}^{k} \rho_2^{k-i} \rho_1^i$$

for some $K_3 > 0$, and hence $\to 0$ as $k \to \infty$.

[1] Our usage of "regular" differs from that of several other sources, especially of Kemeny and Snell (1960).

Hence, as $k \to \infty$

$$\lim_{k\to\infty} R_{k+1} = \sum_{i=0}^{\infty} Q^i R \mathbf{1} \boldsymbol{v}_1' = (I - Q)^{-1} R \mathbf{1} \boldsymbol{v}_1'$$
$$= (I - Q)^{-1}(I - Q)\mathbf{1}\boldsymbol{v}_1'$$
$$= \mathbf{1}\boldsymbol{v}_1'$$

as required. □

Both Theorems 4.2 and 4.7 express conditions under which the probability distribution: $P[X_k = j]$, $j = 1, 2, \ldots, n$ approaches a limit distribution $\boldsymbol{v} = \{v_j\}$ as $k \to \infty$, independent of the initial distribution $\boldsymbol{\Pi}_0$ of $\{X_k\}$. This tendency to a limiting distribution *independent of the initial distribution* expresses a tendency to equilibrium regardless of initial state; and is called the *ergodic property* or *ergodicity*.[1] Theorem 4.6 shows that when attention is focused on behaviour within the set I of *inessential states*, then under a similar structural condition on the set I, an analogous result obtains.

Absorbing-chain Techniques

The discussion given earlier focussing on the behaviour within the set I of inessential states if such exist, for a reducible chain, has wider applicability in the context of MC's containing a single essential class of states in general, and irreducible MC's in particular, though this initially seems paradoxical. We shall give only an informal discussion of some aspects of this topic.

If P is $(n \times n)$ stochastic and irreducible, write $A = I - P$ and ${}_{(n-1)}A = {}_{(n-1)}I - {}_{(n-1)}P$ the $(n-1) \times (n-1)$ northwest corner truncation[1] of A. We may write, with obvious notation:

$$A = \begin{bmatrix} {}_{(n-1)}A & -\boldsymbol{c} \\ -\boldsymbol{d}' & a \end{bmatrix} \qquad \text{where } \boldsymbol{c}, \boldsymbol{d} \geq \mathbf{0}.$$

Now since $i \to n$, $i = 1, \ldots, n-1$ (since P is irreducible), it follows that the modified MC with stochastic transition matrix

$$\begin{bmatrix} {}_{(n-1)}P & \boldsymbol{c} \\ \mathbf{0}' & 1 \end{bmatrix}$$

has the states $\{1, \ldots, n-1\}$ inessential, with "absorbing" state n, so ${}_{(n-1)}P$ plays the role of Q in Theorem 4.3, and in particular ${}_{(n-1)}A^{-1}$ exists and is non-negative. Indeed, by Theorem 4.5, its entries give expected numbers of

[1] See Exercises 4.9 and 4.10.

visits starting from any state $i \in \{1, \ldots, n-1\}$ to any state on this set before "absorption" into state n in the modified chain, and sums of these for a fixed initial state in $\{1, \ldots, n-1\}$ gives the mean time to absorption in n. In regard to the *original* chain, described by P, these mean times have the interpretation of *mean first passage time from* $i \in \{1, \ldots, n-1\}$ to n. We shall take up extension of this important point shortly. For the moment, the reader may wish to check that the unique stationary distribution $\boldsymbol{\Pi}'$, for the chain described by P, which is determined by the equations

$$\boldsymbol{\Pi}'A = \mathbf{0}', \qquad \boldsymbol{\Pi}'\mathbf{1} = 1$$

is given explicitly by the expression

$$\boldsymbol{\Pi}' = \{\boldsymbol{d}'{}_{(n-1)}A^{-1}, 1\}/(1 + \boldsymbol{d}'{}_{(n-1)}A^{-1}\mathbf{1})$$

since $\boldsymbol{c} = ({}_{(n-1)}I - {}_{(n-1)}P)\mathbf{1} = {}_{(n-1)}A\mathbf{1}$.

Clearly the state n holds no special significance in the above argument, which shows that such absorbing chain considerations may be used to obtain expressions for all first passage times from any state to any *other* state in the MC governed by P, and that such considerations may be used to provide expressions for the stationary probability vector.

In an MC with a single essential class of indices, every other state leads *to a specified state in the essential class*. If we regard any such specified state as playing the role of the state n in the above discussion, it is clear that the entire discussion for irreducible P given above applies in this situation also.

For an MC started from state i, the first passage time from i to j in general is the number of transitions ("time") until the process first enters j, if $j \neq i$; or (as shown) the number of transitions until it next enters j, if $j = i$. We have discussed above a method for obtaining the expected first passage time, μ_{ij}, (in specified situations) from i to j, $j \neq i$. The question concerning the expected first passage time from a state to itself has been left open. Denote this quantity for an essential state i by μ_{ii}, or just μ_i: we may clearly treat the question within the framework of an irreducible chain, and do so henceforth. This quantity is more commonly called the mean-recurrence time of state i, and in the purely analytical treatment of Chapter 5 (see Definition 5.1) is called a mean recurrence measure. A simple conditional expectation argument, conditioning on the first step, shows that

$$\mu_{jj} = p_{jj}1 + \sum_{\substack{i \\ i \neq j}} p_{ji}(1 + \mu_{ij})$$
$$= 1 + \sum_{\substack{i \\ i \neq j}} p_{ji}\mu_{ij}$$

so once μ_{ij}, $i \neq j$, $i = 1, \ldots, n$ are all known, μ_{jj} may be calculated. Indeed, put $j = n$, to accord with our previous discussion. Then it follows, in terms of the notation introduced above that

$$\mu_{nn} = 1 + \boldsymbol{d}'{}_{(n-1)}A^{-1}\mathbf{1}$$

whence we obtain, as a bonus from the preceding discussion, the important result that

$$\mu_{nn} = 1/\pi_n$$

where $\Pi = \{\pi_i\}$ is the unique stationary distribution for the MC governed by P.

More generally, for an irreducible MC, it follows

$$\mu_{ii} = 1/\pi_i, \qquad i = 1, \ldots, n,$$

which is a fundamental and intuitively pleasing result of MC theory, and makes clear the intimate connection between mean first-passage times and the stationary distribution.

EXAMPLE (A simple dynamic stochastic inventory model). A toy shop stocks a certain toy. Initially there are 3 items on hand. Demand for the toy during any week is a random variable independent of demand in any other week, and if $p_k = Pr\{k$ toys demanded during a week$\}$, then $p_0 = 0.6$, $p_1 = 0.3$, $p_2 = 0.1$. Orders received when supply is exhausted are not recorded. The shopkeeper may only replenish stock at weekends, according to the policy: do not replenish if there is any stock on hand, but if there is no stock on hand obtain two more items.

Calculate the expected number of weeks to first replenishment, and the limiting-stationary distribution.

Denote by X_n the number of items of stock at the end of week n (just before the weekend). The state space is $\{0, 1, 2, 3\}$, $X_0 = 3$, and $\{X_n\}$ is a Markov chain with transition matrix:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.1 & 0.3 & 0.6 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0.1 & 0.3 & 0.6 & 0 \\ 0 & 0.1 & 0.3 & 0.6 \end{bmatrix} \end{array}.$$

Thus state 3 is inessential, and states 0, 1, 2 form a single essential class. We require μ_{30}. This is the third element of the vector $(I - Q)^{-1}\mathbf{1}$ where

$$Q = \begin{bmatrix} 0.6 & 0 & 0 \\ 0.3 & 0.6 & 0 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}, \quad \text{so } (I - Q)^{-1} = \begin{bmatrix} \frac{5}{2} & 0 & 0 \\ \frac{15}{8} & \frac{5}{2} & 0 \\ \frac{130}{64} & \frac{15}{8} & \frac{5}{2} \end{bmatrix},$$

and since $(I - Q)^{-1}\mathbf{1} = (\frac{5}{2}, \frac{35}{8}, \frac{205}{32})'$, $\mu_{30} = \frac{205}{32}$. The unique stationary distribution is consequently given by the vector

$$\Pi = \frac{\{1, (0.3, 0.6, 0)(I - Q)^{-1}\}}{\{1 + (0.3, 0.6, 0)(I - Q)^{-1}\mathbf{1}\}}$$

$$= (\tfrac{8}{35}, \tfrac{15}{35}, \tfrac{12}{35}, 0).$$

Bibliography and Discussion to §§4.1–4.2

There exists an enormous literature on finite homogeneous Markov chain theory; the concept of a Markov chain is generally attributed to A. A. Markov (1907), although some recognition is also accorded to H. Poincaré in this connection. We list here only the *books* which have been associated with the significant development of this subject, and which may thus be regarded as milestones in its development, referring the reader to these for further earlier references: Markov (1924), Hostinsky (1931), von Mises (1931), Fréchet (1938), Bernstein (1946), Romanovsky (1949), Kemeny & Snell (1960). [The reader should notice that these references are not quite chronological, as several of the books cited appeared in more than one edition, the latest edition being generally mentioned here.] An informative sketch of the early history of the subject has been given by W. Doeblin (1938), and we adapt it freely here for the reader's benefit, in the next two paragraphs.

After the first world war the topic of homogeneous Markov chains was taken up by Urban, Lévy, Hadamard, Hostinsky, Romanovsky, von Mises, Fréchet and Kolmogorov. Markov himself had considered the case where the entries of the finite transition matrix $P = \{p_{ij}\}$ were all positive, and showed that in this case all the $p_{ij}^{(k)}$ tend to a positive limit independent of the initial state, s_i, a result rediscovered by Lévy, Hadamard, and Hostinsky. In the general case $(p_{ij} \geq 0)$ Romanovsky (under certain restrictive hypotheses) and Fréchet, in noting the problem of the calculation of the $p_{ij}^{(k)}$ was essentially an algebraic one, showed that the $p_{ij}^{(k)}$ are asymptotically periodic, Fréchet then distinguishing three situations: the positively regular case, where $p_{ij}^{(k)} \to p_j > 0$, all i, j; the regular case, where $p_{ij}^{(k)} \to p_j \geq 0$ all i, j; the non-oscillating case where $p_{ij}^{(k)} \to p_j^i$, for all i, j; and also the general singular case. Fréchet linked the discussion of these cases to the roots of the characteristic equation of the matrix P. Hostinsky, von Mises, and Fréchet found necessary and sufficient conditions for positive regularity.[1] Finally, Hadamard (1928) gave, in the special case pertaining to card shuffling, the reason for the asymptotic periodicity which enters in the singular case, by using non-algebraic reasoning.

On the other hand the matrix $P = \{p_{ij}\}$ is a matrix of non-negative elements; and these matrices were studied extensively before the first world war by Perron and, particularly, by Frobenius. The remarkable results of Frobenius which enable one to analyze immediately the singular case, were not utilized until somewhat later in chain theory. The first person to do so was probably von Mises (1931), who, in his treatise, deduced a number of important theorems for the singular case from the results of Frobenius. The schools of Fréchet and of Hostinsky remained unaware of this component of von Mises' works, and ignorant also of the third memoir of Frobenius on

[1] See Exercise 4.10 as regards the regular case; Doeblin omits these references.

non-negative matrices. In 1936 Romanovsky, certainly equally unaware of the same work of von Mises, deduced, also from the theorems of Frobenius, by a quite laborious method, theorems more precise than those of von Mises. Finally, Kolmogorov gave in 1936 a complete study of Markov chains with a countable number of states, which is applicable therefore to finite chains.

The present development of the theory of finite homogeneous Markov chains is no more than an introduction to the subject, as the reader will now realise; it deals, further, only with ergodicity problems, whereas there are many problems more probabilistic in nature, such as the Central Limit Theorem, which have not been touched on, because of the nature of the present book. Our approach is, of course, basically from the point of view (really a consequence) of the Perron–Frobenius theory, into which elements of the Kolmogorov approach have been blended. The reader interested in a somewhat similar, early, development, would do well to consult Doeblin's (1938) paper; and a sequel by Sarymsakov (1945).

The subsection on "absorbing chain techniques" has sought to give an elementary flavour of the approach to finite MC theory proposed by Meyer (1975, 1978) and espoused by Berman and Plemmons (1979). A matrix approach to the theory of finite MC's grounded in the elements of linear algebra, with heavy emphasis on spectral structure, has recently been given by Fritz, Huppert and Willems (1979).

Some further discussion pertaining *specifically* to the case of countable, rather than finite state space (or, correspondingly, index set) will be found in the next chapter.

Exercises on §4.2

(All these exercises refer to homogeneous Markov chains.)

4.1. Let P be an irreducible stochastic matrix, with period $d = 3$. Consider the asymptotic behaviour, as $k \to \infty$, of P^{3k}, P^{3k+1}, P^{3k+2} respectively, in relation to the unique stationary distribution corresponding to P. Extend to arbitrary period d.

Hint: Adapt Theorem 1.4 of Chapter 1.

4.2. Find the unique stationary distribution vector $\boldsymbol{v}$ for a random walk between two reflecting barriers, assuming s is odd. (See Example (2) of §4.1.) Apply the results of Exercise 4.1, to write down $\lim_{k\to\infty} P^{2k}$ and $\lim_{k\to\infty} P^{2k+1}$ in terms of the elements of $\boldsymbol{v}$.

4.3. Use either the technique of Appendix B, or induction, to find P^k for arbitrary k, where (stochastic) P is given by

$$(i)\ \begin{pmatrix} p_1 & q_1 \\ p_2 & q_2 \end{pmatrix}; \qquad (ii)\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}; \qquad (iii)\ \begin{pmatrix} 0 & 1 & 0 \\ q & 0 & p \\ 0 & 1 & 0 \end{pmatrix}.$$

4.4. Consider two urns A and B, each of which contains m balls, such that the total number of balls, $2m$, consists of equal numbers of black and white members. A ball is removed simultaneously from each urn and put into the other at times $k = 1, 2, \ldots$.

Explain why the number of white balls in urn A before each transfer forms a Markov chain, and find its transition probabilities.

Give intuitive reasons why it might be expected that the limiting/stationary distribution $\{v_i\}$ of the number of white balls in urn A is given by the hypergeometric probabilities

$$v_i = \frac{\binom{m}{m-i}\binom{m}{i}}{\binom{2m}{m}}, \qquad i = 0, 1, 2, \ldots, m$$

and check that this is so.

4.5. Let P be a finite stochastic matrix (i.e. a stochastic matrix all of whose column sums are also unity).

(*i*) Show that the states of the Markov chain corresponding to P are all essential.

(*ii*) If P is irreducible and aperiodic, find $\lim P^m$ as $m \to \infty$.

4.6. A Markov chain $\{X_k\}$, $k = 0, 1, 2, \ldots$ is defined on the states $0, 1, 2, \ldots, 2N$, its transition probabilities being given by

$$p_{j,i} = \binom{2N}{i}\left(\frac{j}{2N}\right)^i\left(1 - \frac{j}{2N}\right)^{2N-i},$$

$j, i = 0, 1, \ldots, 2N$. Investigate the nature of the states.

Show that for $m \geq 0$, $j = 0, 1, 2, \ldots, 2N$,

$$\mathscr{E}[X_{m+1} \mid X_m = j] = j,$$

and that consequently

$$\mathscr{E}[X_{m+1} \mid X_0 = j] = j.$$

Hence, or otherwise, deduce the probabilities of eventual absorption into the state 0 from the other states.

(Malécot, 1944)

4.7. A Markov chain is defined on the integers $0, 1, 2, \ldots, a$, its transition probabilities being specified by

$$p_{i,i+1} = \frac{1}{2}\left(\frac{a-i-1}{a-i}\right),$$

$$p_{i,i-1} = \frac{1}{2}\left(\frac{a-i+1}{a-i}\right),$$

$i = 1, 2, \ldots, a-1$, with states 0 and a being absorbing.

Find the mean time to absorption, m_i, starting from $i = 1, 2, \ldots, a-1$.

Hints: (1) The state $a - 1$ is reflecting. (2) Use the substitution $z_i = (a-i)m_i$.

4.8. Let L be an $(n \times n)$ matrix with zero elements on the diagonal and above, and U an $(n \times n)$ matrix with zero elements below the diagonal. Suppose that $P_1 = L + U$ is stochastic.
(*i*) Show that $L^n = 0$, and hence (with the help of probabilistic reasoning, or otherwise), that the matrix $P_2 = (I - L)^{-1}U$ is stochastic.
(*ii*) Show by example that even if P_1 is irreducible and aperiodic, P_2 may be reducible.

4.9. Theorem 4.7 may be regarded as asserting that a sufficient condition for ergodicity is the regularity of the transition matrix P. Show that regularity is in fact a necessary condition also.

4.10. Use the definition of regularity and the result of the preceding exercise to show that a necessary and sufficient condition on the matrix P for ergodicity is that there is only one eigenvalue of modulus unity (counting any repeated eigenvalues as distinct).

(Kaucky, 1930; Konečný, 1931)

4.11. Show that any inessential state leads to an essential state.
Hint: Use a contradiction argument, as in the proof of Lemma 1.1 of Chapter 1.

4.12. Show that if an n-state MC contains at least two essential classes of states, then any weighted linear combination of the stationary distribution vectors corresponding to each such class, each appropriately augmented by zeros to give an $(n \times 1)$ vector, is a stationary distribution of the chain.

4.13. Denote by M_j the class of $(n \times n)$ stochastic matrices P such that for some power k, and hence for all higher powers, P^k has its jth column positive. Denote by G_1 the class of regular $(n \times n)$ stochastic matrices. Show that $G_1 = \bigcup_{j=1}^n M_j$, while $\bigcap_{j=1}^n M_j$ is the set of primitive $(n \times n)$ stochastic matrices. [See Exercise 3.12.]

4.14. Suppose P is irreducible and stochastic, with period d, and $\mathbf{v}$ its unique stationary distribution vector. Let $R = \lim_{k\to\infty} P^{dk}$. Show that

$$R \sum_{k=0}^{d-1} \frac{P^k}{d} = \mathbf{1}\mathbf{v}'.$$

[*Hint:* Consider P in canonical form, and use the methods of §1.4.]

4.3 Finite Inhomogeneous Markov Chains and Coefficients of Ergodicity

In this section, as already foreshadowed in §4.1 of this chapter, we shall adopt the notation of Chapter 3 except that we shall use $P_k = \{p_{ij}(k)\}$ instead of $H_k = \{h_{ij}(k)\}$, $i, j = 1, \ldots, n$ to emphasize the stochasticity of P_k, and we

shall be concerned with the asymptotic behaviour of the forward product[1]

$$T_{0,r} \equiv P_1 P_2 \cdots P_r \equiv \prod_{k=1}^{r} P_k$$

as $k \to \infty$.

Naturally, both Theorems 3.3 and 3.7, for example, are applicable here, and it is natural to begin by examining their implications in the present context, where, we note, each $T_{p,r}$ is stochastic also.

Under the conditions of the first of these, as $r \to \infty$, for all i, j, p, s

$$\frac{t_{i,s}^{(p,r)}}{t_{j,s}^{(p,r)}} \to W_{i,j}^{(p)} > 0$$

where the limit is independent of s. Put for sufficiently large r,

$$\frac{t_{i,s}^{(p,r)}}{t_{j,s}^{(p,r)}} = W_{i,j}^{(p)} + \varepsilon(i, j, s, p, r).$$

Thus, using the stochasticity of $T_{p,r}$

$$1 = \sum_{s=1}^{n} t_{i,s}^{(p,r)} = W_{i,j}^{(p)} + \sum_{s=1}^{n} t_{j,s}^{(p,r)}\varepsilon(i, j, s, p, r)$$

where also $0 < t_{j,s}^{(p,r)} \le 1$. Letting $r \to \infty$, it follows that the additional assumption has led to:

$$1 = W_{i,j}^{(p)}, \qquad \text{all } i, j, p,$$

so that the rows tend not only to proportionality, but indeed equality, although their nature still depends on r in general. Indeed we may write

$$\frac{t_{i,s}^{(p,r)}}{t_{j,s}^{(p,r)}} - 1 = \frac{t_{i,s}^{(p,r)} - t_{j,s}^{(p,r)}}{t_{j,s}^{(p,r)}} \to 0$$

as $r \to \infty$, which on account of the present boundedness of $t_{j,s}^{(p,r)}$ implies

$$t_{i,s}^{(p,r)} - t_{j,s}^{(p,r)} \to 0 \tag{4.2}$$

as $r \to \infty$, for each i, j, p, s. This conclusion is a weaker one than that preceding it, and so we may expect to obtain it under weaker assumptions (although in the present stochastic context) than given in Theorem 3.3. In fact, since this kind of assertion does not involve a ratio, conditions imposed in the former context, to ensure positivity of denominator, *inter alia*, may be expected to be subject to weakening.

Under the conditions of Theorem 3.7, in addition to the present stochasticity assumption, we have, simply, that

$$t_{i,j}^{(p,r)} \to v_j, \qquad j = 1, 2, \ldots, n \tag{4.3}$$

where $\mathbf{v}' = \{v_j\}$ is the unique invariant distribution of the limit primitive matrix P.

[1] Backwards products of stochastic matrices are of interest also: see §4.6.

(4.2) and (4.3) are manifestations of weak and strong ergodicity respectively in the MC sense.

Definition 4.4. We shall say that *weak ergodicity* obtains for the MC (i.e. sequence of stochastic matrices P_i) if

$$t_{i,s}^{(p,r)} - t_{j,s}^{(p,r)} \to 0$$

as $r \to \infty$ for each i, j, s, p. (Note that it is sufficient to consider $i \neq j$.)

This definition does not imply that the $t_{i,s}^{(p,r)}$ themselves tend to a limit as $r \to \infty$, merely that the rows tend to equality (= "independence of initial distribution") but are still in general dependent on r.

Definition 4.5. If weak ergodicity obtains, and the $t_{i,s}^{(pr)}$ themselves tend to a limit for all i, s, p as $r \to \infty$, then we say *strong ergodicity* obtains.

Hence strong ergodicity requires the elementwise existence of the limit of $T_{p,r}$ as $r \to \infty$ for each p, in addition to weak ergodicity. It is clear from Definition 3.4 and Lemma 3.5 that the definition of strong ergodicity here is completely consistent with that given in the more general setting.

A stochastic matrix with identical rows is sometimes called *stable*. Note that if P is stable, $P^2 = P$, and so $P^r = P$.

Thus we may, in a consistent way, speak of weak ergodicity as tendency to stability.

As in Chapter 3, a convenient approach to the study of both weak and strong ergodicity is by means of an appropriate contraction coefficient; such coefficients in this stochastic setting are more frequently called coefficients of ergodicity. In contrast to Chapter 3, we shall first introduce some general notions for this concept, then specialize to those we shall use in the sequel. It will be seen that the notions of stochastic matrices with a positive column, and the quantity $\tau_1(P)$ already encountered, *interalia*, within the context of Theorems 3.1 and 2.10, are central to this discussion.

Definition 4.6. We call any scalar function $\tau(\cdot)$ continuous on the set of $(n \times n)$ stochastic matrices (treated as points in R_{n^2}) and satisfying $0 \le \tau(P) \le 1$, a *coefficient of ergodicity*. It is then said to be *proper* if

$$\tau(P) = 0 \qquad \text{if and only if } P = \mathbf{1}v'$$

where v is any probability vector ($v \ge \mathbf{0}$, $v'\mathbf{1} = 1$): that is, whenever P is stable.

Lemma 4.1. *Weak ergodicity of forward products is equivalent to*

$$\tau(T_{p,r}) \to 0, \qquad r \to \infty, \qquad p \ge 0$$

where $\tau(\cdot)$ is a proper coefficient of ergodicity.

PROOF. Take p fixed but arbitrary, and suppose $\tau(T_{p,r}) \to 0$, $r \to \infty$. Suppose then $t_{i,s}^{(p,r)} - t_{j,s}^{(p,r)} \to 0$ for all i, j, s as $r \to \infty$ is false. Then there is a subsequence $\{k_r\}$, $r \ge 1$, of the positive integers and an $\varepsilon > 0$ such that the Euclidean distance between T_{p,k_r} and *every* stable matrix is at least ε. Since T_{p,k_r}, $r \ge 1$, is stochastic and the set of stochastic matrices is compact (bounded and closed) in R_{n^2}, we may, by selecting a subsequence of $\{k_r\}$ if necessary, assume $T_{p,k_r} \to P^*$ where P^* is stochastic, and by the assumptions on τ, $\tau(T_{p,k_r}) \to 0 = \tau(P^*)$, whence P^* is stable, and hence a contradiction results. The converse follows easily by continuity of $\tau(\cdot)$. □

Theorem 4.8. *Suppose $m(\cdot)$ and $\tau(\cdot)$ are proper coefficients of ergodicity and for any r stochastic matrices $P^{(i)}$, $i = 1, \ldots, r$ with each $r \ge 1$:*

$$m(P^{(1)}P^{(2)} \cdots P^{(r)}) \le \prod_{i=1}^{r} \tau(P^{(i)}). \tag{4.4}$$

Then weak ergodicity of forward products[1] *$T_{p,r}$ formed from a given sequence $\{P_k\}$, $k \ge 1$, obtains if and only if there is a strictly increasing sequence of positive integers $\{k_s\}$, $s = 0, 1, 2, \ldots$ such that*

$$\sum_{s=0}^{\infty} \{1 - \tau(T_{k_s, k_{s+1}-k_s})\} = \infty. \tag{4.5}$$

PROOF. (Similar to Theorem 3.2; Exercise 4.15). □

Examples of *proper* coefficients of ergodicity are (in terms of $P = \{p_{ij}\}$; see Theorem 3.1) evidently:

$$\tau_1(P) = \frac{1}{2} \max_{i,j} \sum_{s=1}^{n} |p_{is} - p_{js}| \equiv 1 - \min_{i,j} \sum_{s=1}^{n} \min(p_{is}, p_{js});$$

$$a(P) = \max_{s} \max_{i,j} |p_{is} - p_{js}|;$$

$$b(P) = 1 - \sum_{s=1}^{n} \left(\min_i p_{is}\right).$$

An example of an *improper* coefficient of ergodicity is

$$c(P) = 1 - \max_{s} \left(\min_i p_{is}\right),$$

where[2]

$$a(P) \le \tau_1(P) \le b(P) \le c(P) \tag{4.6}$$

[1] For the analogous result relating to backwards products see Theorem 4.18.

[2] See Exercise 4.16.

with $c(P) < 1$ if and only if P has a positive column. Theorem 3.1 enables us to deduce a concrete manifestation of (4.4), for if we substitute in it $w = \{w_j\}$ with $w_j = t_{j,s}^{(p,r)}$, $P = P_{p+1}$, we have from (3.4)

$$\max_{h,h'} |t_{h,s}^{(p,r)} - t_{h',s}^{(p,r)}| \le \tau_1(P_{p+1}) \max_{j,j'} |t_{j,s}^{(p+1,r-1)} - t_{j',s}^{(p+1,r-1)}|$$

so that

$$a(T_{p,r}) \le \tau_1(P_{p+1}) a(T_{p+1,r-1}).$$

More generally for *any* sequence $\{P^{(i)}\}$, $i \ge 1$, of stochastic matrices, and each $r \ge 1$,

$$a(P^{(1)}P^{(2)} \cdots P^{(r)}) \le \tau_1(P^{(1)}) a(P^{(2)} \cdots P^{(r)})$$
$$\le \tau_1(P^{(1)})\tau_1(P^{(2)}) \cdots \tau_1(P^{(r)})\tau_1(I)$$

where I is the unit matrix; i.e.

$$a(P^{(1)}P^{(2)} \cdots P^{(r)}) \le \prod_{i=1}^{r} \tau_1(P^{(i)}) \tag{4.7}$$

since $\tau_1(I) = 1$. By (4.6) it follows that (4.4) also holds with $m = a$, and $\tau = b$ (or $\tau = c$, taking into account the Corollary to Theorem 4.8). A "homogeneous" inequality of form (4.4), in that both $m(\cdot)$ and $\tau(\cdot)$ are the same, may be obtained analogously to (3.7) by considering a metric $d(x', y')$ on the sets of probability row vectors.

Lemma 4.2. *For a metric d on the set $D = \{z'; z \ge \mathbf{0}, z'\mathbf{1} = 1\}$ the quantity, defined for any $(n \times n)$ stochastic matrix P by*

$$\tau(P) = \sup_{\substack{x', y' \in D \\ x \ne y}} \frac{d(x'P, y'P)}{d(x', y')}$$

satisfies the properties

(i) $\tau(P^{(1)}P^{(2)}) \le \tau(P^{(1)})\tau(P^{(2)})$, $P^{(1)}$, $P^{(2)}$ *stochastic;*
(ii) $\tau(P) = 0$ *for stochastic P if and only if P is stable.*

Proof. The only non-obvious part of this assertion is $\tau(P) = 0 \Rightarrow P = \mathbf{1}v'$ where $v' \in D$. Now $\tau(P) = 0 \Rightarrow (x - y)'P = \mathbf{0}'$ for any two probability vectors x, y, and $(x - y)\mathbf{1} = 0$. Taking $x = f_i$, $y = f_j$, $i \ne j$, where f_k is, as usual, the vector with zeroes everywhere except unity in the kth position, it follows that the ith and jth rows of P are the same, for arbitrary i, j. □

This lemma provides a means of generating coefficients of ergodicity, providing the additional constraints inherent in their definition, of continuity and that $\tau(P) \le 1$, are satisfied for any P. There are a number of well-known metrics defined on sets of probability distributions[1] such as D, and

[1] See Exercise 4.17.

other candidates for investigation are metrics corresponding to any vector norm $\|\cdot\|$ on R_n, or C_n (the set of n-length vectors with complex valued entries), i.e.

$$d(x', y') = \|x' - y'\|. \tag{4.8}$$

Obvious choices for investigation here are the l_p norms

$$\|x'\|_p = \left\{\sum_{i=1}^{n} |x_i|^p\right\}^{1/p} \qquad \left(\|x'\|_\infty = \max_i |x_i|\right)$$

where $x' = \{x_i\}$.

For any metric of the form (4.8) the definition of $\tau(P)$ according to Lemma 4.2 is

$$\tau(P) = \sup_{\substack{x', y' \in D \\ x \ne y}} \frac{\|(x - y)'P\|}{\|(x - y)'\|} = \sup_{\|\delta\| = 1} \|\delta' P\|$$

since any real-valued vector $\delta = \{\delta_i\}$ satisfying $\delta \ne \mathbf{0}$ $\delta'\mathbf{1} = 0$ may be written in the form $\delta = \text{const}\,(x - y)$ where x and y are probability vectors, $x \ne y$, and $\text{const} = \frac{1}{2}\sum_i |\delta_i| = \sum_i \delta_i^+ = -\sum_i \delta_i^-$ where $a^+ = \max(a, 0)$, $a^- = \min(a, 0)$.

The following result provides another concrete manifestation of (4.4), and establishes $\tau_1(\cdot)$ as an analogue, in the present stochastic setting, of $\tau_B(\cdot)$ of Chapter 3.

Lemma 4.3. *For stochastic $P = \{p_{ij}\}$*

$$\sup_{\substack{\|\delta'\| = 1 \\ \delta' = 0}} \|\delta' P\|_1 = \tau_1(P) = \frac{1}{2} \max_{i,j} \sum_{s=1}^{n} |p_{is} - p_{js}|$$

so that $\tau_1(\cdot)$ is a proper coefficient of ergodicity satisfying

$$\tau_1(P^{(1)}P^{(2)}) \le \tau_1(P^{(1)})\tau_1(P^{(2)}),$$

for any stochastic $P^{(1)}$, $P^{(2)}$.

Proof. By Lemma 2.4, any real $\delta = \{\delta_i\}$ satisfying $\|\delta'\| = 1$, $\delta'\mathbf{1} = 0$ may be written

$$\delta = \sum_{(i,j) \in \mathscr{I}} \left(\frac{\eta_{i,j}}{2}\right) \gamma(i, j)$$

where

$$\eta_{i,j} > 0, \qquad \sum_{(i,j) \in \mathscr{I}} \eta_{i,j} = 1, \qquad \gamma(i, j) = f_i - f_j$$

a suitable set $\mathscr{I} = \mathscr{I}(\delta)$ of ordered pairs of indices (i, j), $i, j = 1, \ldots, n$.

Hence

$$\|\delta' P\|_1 = \sum_s \left| \sum_r \delta_r p_{rs} \right| \le \sum_s \sum_{(i,j)\in\mathscr{I}} \left(\frac{\eta_{ij}}{2}\right) |p_{is} - p_{js}|$$
$$\le \frac{1}{2} \max_{i,j} \sum |p_{is} - p_{js}|$$
$$= \tau_1(P).$$

We may construct a δ such that $\|\delta'\|_1 = 1$, $\delta' \mathbf{1} = 0$, $\|\delta' P\|_1 = \tau_1(P)$ as follows:

suppose
$$\tau_1(P) = \frac{1}{2} \sum_s |p_{i_0 s} - p_{j_0 s}|, \qquad i_0 \ne j_0,$$

and take $\delta = \frac{1}{2}(f_{i_0} - f_{j_0})$. The final part of the assertion follows from Lemma 4.2. □

Corollary. *Weak ergodicity of forward products of a sequence of $(n \times n)$ stochastic matrices is equivalent to*

$$\tau_1(T_{p,r}) \to 0, \qquad r \to \infty, \qquad p \ge 0.$$

4.4 Sufficient Conditions for Weak Ergodicity

In this section we apply the general notions of §4.3 and earlier chapters to obtain conditions for weak ergodicity.

Definition 4.7. An $(n \times n)$ stochastic matrix P is called a Markov matrix if $c(P) < 1$, i.e. at least one column of P is entirely positive. We shall also need repeatedly the notion of a regular stochastic matrix (Definition 4.3), and shall denote the set of such $(n \times n)$ matrices by G_1 (as in Exercise 4.13). The class of $(n \times n)$ Markov matrices is denoted by M; obviously $M \subset G_1$. We shall introduce further classes of stochastic matrices, G_2 and G_3 in the course of this section.

The following theorem is the oldest, and in a sense the most fundamental (as we shall see from the sequel) result on weak ergodicity, which we shall treat with minimal recourse to $\tau_1(\cdot)$, which however will play a substantial role, analogous to that of $\tau_B(\cdot)$ in Chapter 3, in the discussion of strong ergodicity. It was initially proved by direct contractivity reasoning.[1]

[1] See Exercise 4.18.

Theorem 4.9. *Weak ergodicity obtains for forward products formed from a sequence $\{P_k\}$, $k \ge 1$, of stochastic matrices if*

$$\sum_{k=1}^{\infty} \{1 - c(P_k)\} = \infty.$$

PROOF.

$$\tau_1(T_{p,r}) \le \prod_{i=1}^{r} \tau_1(P_{p+i})$$

by Lemma 4.3;

$$\le \prod_{i=1}^{r} c(P_{p+i})$$

by (4.6) and the assertion of the theorem is tantamount to

$$\prod_{i=1}^{\infty} c(P_i) = 0,$$

so

$$\tau_1(T_{p,r}) \to 0 \qquad \text{as } r \to \infty, \quad p \ge 0.$$

The conclusion follows from Lemma 4.1. □

Corollary 1.

$$\alpha(T_{p,r}) = \max_s \max_{i,j} |t_{i,s}^{(p,r)} - t_{j,s}^{(p,r)}| \le \prod_{i=1}^{r} c(P_{p+i})$$

[This follows from (4.6), since $a(P) \le \tau_1(P)$].

Corollary 2. *If $c(P_k) \le c_0 < 1$, $k \ge 1$, (i.e. all P_k are "uniformly Markov") then weak ergodicity obtains at a rate which is at least geometric with parameter c_0 (for every $p \ge 0$).*

The following sequence of arguments including Theorem 4.10 parallels that leading to Theorem 3.7.

Lemma 4.4. *If P and Q are stochastic, $Q \in G_1$ and PQ or QP has the same incidence matrix as P (i.e. $PQ \sim P$ or $QP \sim P$), then $P \in M$.*

PROOF. Since $Q \in G_1$, $Q^k \in M$ for some k. Assuming first $PQ \sim P$, it follows $PQ^k \sim P$, so P, like PQ^k, has at least those columns positive that are positive in Q^k. If we assume $QP \sim P$, it follows that $Q^k P \sim P$ and that P, like $Q^k P$, will have at least one column positive. □

Lemma 4.5. *If $T_{p,r} \in G_1$, $p \ge 0$, $r \ge 1$, then $T_{p,r} \in M$ for $r \ge t$ where t is the number of distinct incidence matrices corresponding to G_1.*

PROOF. For a fixed p, there are some numbers a, b satisfying $1 \le a < b \le t + 1$ such that

$$P_{p+1}P_{p+2} \cdots P_{p+a}P_{p+a+1} \cdots P_{p+b} \sim P_{p+1}P_{p+2} \cdots P_{p+a}$$

since the number of distinct incidence matrices is t. Hence

$$T_{p,a}T_{p+a,b-a} \sim T_{p,a}.$$

Since $T_{p+a,b-a} \in G_1$, by Lemma 4.4, $T_{p,a} \in M$. Thus $T_{p,r}$, $r \ge a$, has a strictly positive column (not necessarily the same one for each r). □

The following result is analogous to Theorem 3.3 of Chapter 3.

Theorem 4.10. *If* $T_{p,r} \in G_1$, $p \ge 0$, $r \ge 1$, *and*[1]

$$\min_{i,j}{}^{+} p_{ij}(k) \ge \gamma > 0 \tag{4.9}$$

uniformly for all $k \ge 1$, *then weak ergodicity obtains, at a uniform geometric rate for all* $p \ge 0$. [*In particular,* (4.9) *holds if the sequence* $\{P_k\}$ *has each of its elements selected from a numerically finite set of stochastic matrices.*]

PROOF. Consider p fixed but arbitrary and r "large": then

$$T_{p,r} = T_{p,t}T_{p+t,t}T_{p+2t,t} \cdots T_{p+(k-1)t,t}\overline{T}_{(p,r)} = T_{p,kt}\overline{T}_{(p,r)}$$

where k is the largest positive integer such that $kt \le r$, t has the meaning of Lemma 4.5, and $\overline{T}_{(p,r)}$ is some stochastic (possibly the unit) matrix. Since by Lemma 4.5, $T_{p+it,t}$ is Markov, from (4.9)

$$c(T_{p+it,t}) \le 1 - \gamma^t.$$

By Corollary 1 of Theorem 4.9 and the last equation

$$a(T_{p,r}) \le \prod_{i=0}^{k-1} c(T_{p+it,t}) \le (1 - \gamma^t)^k \le (1 - \gamma^t)^{(r/t)-1}$$

i.e.

$$a(T_{p,r}) \le (1 - \gamma^t)^{-1}\{(1 - \gamma^t)^{1/t}\}^r,$$

and letting $r \to \infty$ completes the result, since $\gamma < 1$. □

The assumption that $T_{p,r} \in G_1$, $p \ge 0$, $r \ge 1$ in Theorem 4.10 is a restrictive one on the basic sequence P_k, $k \ge 1$, and from the point of view of utility, conditions on the individual matrices P_k are preferable. To this end, we introduce the classes G_2 and G_3 of $(n \times n)$ stochastic matrices.

[1] Recall that $\min^+$ is the minimum over the positive elements. We may call (4.9) condition (C) in accordance with (3.18) of Chapter 3.

Definition 4.8. (*i*) $P \in G_2$ if (*a*) $P \in G_1$; (*b*) $QP \in G_1$ for any $Q \in G_1$; (*ii*) $P \in G_3$ if $\tau_1(P) < 1$, i.e. if given two rows α and β, there is at least one column γ such that $p_{\alpha\gamma} > 0$ and $p_{\beta\gamma} > 0$.[1]

If $P \in G_3$, P is called a *scrambling* matrix; the present definitions of such matrices is entirely consistent with the more general Definition 3.2. It is also clear from the definition of the class G_2 that if $P_k \in G_2$, $k \ge 1$, then $T_{p,r} \in G_1$, $p \ge 0$, $r \ge 1$.

Theorem 4.11. $M \subset G_3 \subset G_2 \subset G_1$.

PROOF. The implication $M \subset G_3$ (any Markov matrix is scrambling) is obvious, and that $G_2 \subset G_1$ follows from the definition of G_2.

To prove $G_3 \subset G_2$, consider a scrambling P in canonical form, so its essential classes of indices are also in canonical form if periodic. (Clearly the scrambling property is invariant under simultaneous permutation of rows and columns.)

It is now easily seen that if there is more than one essential class the scrambling property fails by judicious selection of rows α and β in different essential classes; and if an essential class is periodic, the scrambling property fails by choice of α and β in different cyclic subclasses. Thus $P \in G_1$.

Now consider a scrambling $P = \{p_{ij}\}$, not necessarily in canonical form. Then for any stochastic $Q = \{q_{ij}\}$, QP is scrambling, for take any two rows α, β and consider the corresponding entries

$$\sum_{k=1}^{n} q_{\alpha k}p_{kj}, \qquad \sum_{r=1}^{n} q_{\beta r}p_{rj}$$

in the jth column of QP. Then there exist k, r such that $q_{\alpha k} > 0$, $q_{\beta r} > 0$ by stochasticity of Q. By the scrambling property of P, there exists j such that $p_{kj} > 0$ and $p_{rj} > 0$. Hence QP is scrambling, and so $QP \in G_1$ by the first part of the theorem.

Hence, putting both parts together, $P \in G_2$. □

Corollary. *For any stochastic* Q, QP *is scrambling, for fixed scrambling* P.

This corollary motivates the following result.

Lemma 4.6. *If* P *is scrambling, then so are* PQ *and* QP *for any stochastic* Q. *The word "scrambling" may be replaced with "Markov".*

PROOF. In view of what has gone before in this section, we need prove only that $p \in G_3 \Rightarrow PQ \in G_3$ for any stochastic Q. Since $P = \{p_{ij}\}$ is scrambling, for any pair of indices (α, β) there is a $j = j(\alpha, \beta)$ such that $p_{\alpha j} > 0$, $p_{\beta j} > 0$. There is a k such that $q_{jk} > 0$ by stochasticity of Q. Hence the kth column of PQ has positive entries in the α, β rows. □

[1] See Exercise 2.27.

Matrices in G_3 thus have two special properties:

(*i*) it is easy to verify whether or not a matrix $\in G_3$;
(*ii*) if all P_k are scrambling, then $T_{p,k}$ is scrambling, $p \geq 0$, $k \geq 1$, and in particular $T_{p,k} \in G_1$.

Lemma 4.7. $P, Q \in G_2 \Rightarrow PQ, QP \in G_2$.

Proof. See Exercise 4.20. □

Thus G_2 is closed under multiplication, but G_1 is not.[1]

Lemma 4.8. *If P has incidence matrix of form $\tilde{P} = I + C$ where C is an incidence matrix (such P are said to be "normed"), and $Q \in G_1$, then QP and $PQ \in G_1$. [In particular, if also $P \in G_1$, $P \in G_2$.]*

Proof. See Exercise 4.20. □

This result permits us to demonstrate that G_3 is a proper subset of G_2. A stochastic matrix P whose incidence matrix is

$$\tilde{P} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

evidently satisfies the condition of Lemma 4.8 and is clearly a member of G_1, so $P \in G_2$. However, the 3rd and 4th rows do not intersect, so P is not scrambling.

To conclude this section we remark that another generalization of scrambling matrices useful in a certain context in regard to verification of the condition

$$T_{p,r} \in G_1, \qquad p \geq 0, \qquad r \geq 1$$

is given in the following Bibliography and Discussion.

Bibliography and Discussion to §§4.3–4.4

The definition of weak ergodicity is due to Kolmogorov (1931), who proves weak ergodicity for a sequence P_k, $k \geq 1$, of finite or infinite stochastic matrices under a restrictive condition related to the Birkhoff coefficient of ergodicity $\tau_B(\cdot)$ (see Seneta, 1979), rather than coefficients of the kind discussed here. (See also Sarymsakov (1954), §4, for a repetition of Kolmogorov's

[1] See Exercise 4.21.

reasoning. The statement of Theorem 4.8 is in essence due to Doeblin (1937); our proof follows Hajnal (1958). The coefficient of ergodicity $b(\cdot)$ is likewise due to Doeblin (1937); while $a(\cdot)$ and $\tau_1(\cdot)$, via Theorem 3.1, are already implicit in Markov (1906)—see Bibliography and discussion to §§3.1–3.2. Lemmas 4.2 and 4.3 are largely due to Dobrushin (1956) (see also Hajnal (1958) and Paz and Reichaw (1967)). For a historical survey see Seneta (1973b).

Theorem 4.9 and its Corollaries are due to Bernstein (1946) (see Bernstein (1964) for a reprinting of the material). It is known in the Russian literature as "Markov's Theorem"; Bernstein's reasoning is by way of the contractive result presented as our Exercise 4.18 which, the reader will perceive, is obtained by the same argument as (Markov's) Theorem 3.1.

Lemmas 4.4, 4.5, 4.7, and 4.8, and Theorem 4.10 are all due to Sarymsakov (1953a; summary) and Sarymsakov and Mustafin (1957), although the simple proof of Lemma 4.5 as presented here is essentially that of Wolfowitz (1963). Lemma 4.6 (also announced by Sarymsakov, 1956), and a number of other properties of scrambling matrices are proved by Hajnal (1958). The introduction of the class G_2 is again due to Sarymsakov and Mustafin. The elegant notion of a scrambling matrix was exploited by Sarymsakov (1956) and Hajnal (1958).

Weaker versions of Theorem 4.10, pertaining to the situation where each P_k is chosen from a fixed finite set K, were obtained by Wolfowitz (1963) and Paz (1963, 1965). The substantial insight into the problem due to Sarymsakov and Mustafin consists in noting that the theory can be developed in terms of the (at first apparently restrictive) notion of a Markov matrix, and made to depend, in the end, on (Bernstein's) Theorem 4.9.

In coding (information, probabilistic automation) theory, where the term SIA is generally used in place of *regular*, the situation is somewhat different to that considered above (see Paz, 1971, Chapter 2, §§3–4) in that, instead of considering all products of the form $T_{p,r}$ formed from a *given sequence* P_k, $k \geq 1$, of stochastic matrices, one considers *all possible sequences* which can be formed from a *fixed finite set* ("alphabet") K, and the associated forward products ("words") $T_{0,r}$, $r \geq 1$, ($T_{0,r}$ is said to be a word of length r) for each. The analogous *basic assumption* (to $T_{p,r} \in G_1, p \geq 0, r \geq 1$) is then that all words in the member matrices of the alphabet K be regular (in which case all forward products formed from any particular sequence are uniformly weakly ergodic by Theorem 4.10.) By Lemma 4.5, a necessary condition for all words to be regular is the existence of an integer r_0 such that all words of length $r \geq r_0$ are scrambling; this condition is also sufficient (Exercise 4.28); and indeed, from Lemma 4.5, there will then be such an r_0 satisfying $r_0 \leq t$. Thus in the present context the validity (or not) may certainly be verified in a finite number of operations, and substantial effort has been dedicated to obtaining bounds on the amount of labour involved. Bounds of this kind were obtained by Thomasian (1963) and Wolfowitz (1963), but a best possible result is due to Paz (1965) who shows that if an r_0 exists, then the smallest value it can take satisfies

$$\min r_0 \le \tfrac{1}{2}(3^n - 2^{n+1} + 1)$$

and the bound is sharp. Thus one needs only check all words of lengths progressively increasing to this upper bound to see if for some specific length all words are scrambling (by Lemma 4.6 all words of greater length will be scrambling). It is clear that a large number of words may still need to be checked.

The question arises whether by restricting the set K to particular types of matrices, the basic assumption is more easily verifiable. This is the case if $K \subset G_3$ (Lemma 4.6, Theorem 4.11); and more generally if $K \subset G_2$ (Lemma 4.7, Theorem 4.11). A substantial, but not wholly conclusive, study of the class G_2 occurs in Sarymsakov and Mustafin (1957), as do studies of related classes of matrices. Of interest therefore are several possible conditions on the elements of the set K such that any word of length $(n-1)$ is scrambling (then all words are regular, by Exercise 4.28).

One condition of the kind mentioned is announced by Sarymsakov (1958), and proved, rather lengthily, in Sarymsakov (1961). Suppose P is a stochastic matrix with the property that if A and $\tilde{A}$ are any two disjoint non-empty sets of its indices than *either* $F(A) \cap F(\tilde{A}) \neq \phi$; *or* $F(A) \cap F(\tilde{A}) = \phi$ and $\#(F(A) \cup F(\tilde{A})) > \#(A \cup \tilde{A})$. Here for any set of indices B, $F(B)$ is the set of "one-step" consequent indices (in line with the definition and notation of §2.4); $\#$ denotes the "number of indices in". Clearly, scrambling matrices satisfy this condition, since these are precisely the stochastic matrices for which $F(A) \cap F(\tilde{A}) \neq \phi$ for all A, $\tilde{A}$. To see very simply that a product of $(n-1)$ matrices *of this kind* is scrambling, notice that if any two rows intersect in P, then they intersect in PQ, for any stochastic Q (proof as in Lemma 4.6). Now suppose there are two rows which do not intersect in a product of $(n-1)$ matrices each satisfying Sarymsakov's property: call these A, $\tilde{A}$; and write (as earlier) $F^k(B)$ for the set of kth-stage consequents of any non-empty set B (i.e. after multiplying the first k matrices together). Then our supposition is $\phi = F^{n-1}(A) \cap F^{n-1}(\tilde{A})$, so

$$\#(F^{n-1}(A) \cup F^{n-1}(\tilde{A})) > \#(F^{n-2}(A) \cup F^{n-2}(\tilde{A})) > \cdots > \#(A \cup \tilde{A}) = 2;$$

i.e. $\#(F^{n-1}(A) \cup F^{n-1}(\tilde{A})) > n$, a contradiction to the supposition. Thus if K consists of a (finite numer of) matrices of Sarymsakov's class, the basic assumption will be satisfied. However, to verify for a *particular* P that it does not belong to this class, the number of pairs of sets A, $\tilde{A}$ which need to be checked is, by the partition argument of Paz (1971, p. 90), again $\tfrac{1}{2}(3^n - 2^{n+1} + 1)$.

Another condition on K ensuring that any word of length $(n-1)$ is not only scrambling, but is, indeed, a Markov matrix, is given by Anthonisse and Tijms (1977).

Recent contributions to the theory of inhomogeneous Markov chains relevant to §§4.3–4.4 have also been made by Paz (1971), Iosifescu (1972, 1977), Kingman (1975), Cohn (1976), Isaacson and Madsen (1976), and Seneta (1979). Treatments within monographs are given by Bernstein (1946, 1964), Paz (1971), Seneta (1973), Isaacson and Madsen (1976) and Iosifescu (1977); these provide references additional to those listed in this book.

Although all our theoretical development for weak ergodicity has been for a sequence P_k, $k \ge 1$, of stochastic matrices which are all $(n \times n)$, we have mentioned earlier (in connection with the Pólya Urn scheme) that some inhomogeneous Markov chains do not have a constant state space; and in general one should examine the situation where if P_k is $(n_k \times n_{k+1})$, then P_{k+1} is $(n_{k+1} \times n_{k+2})$, $k \ge 1$. It is then still possible to carry some of the theory through (the argument of Exercise 4.18 still holds, for example). Writings on finite inhomogeneous chains of the Russian school (e.g. Bernstein, 1946; Sarymsakov, 1958, 1961) have tended to adhere to this framework.

We have not considered the case of products of infinite stochastic matrices here; the reader is referred to the monographs cited.

Exercises on §§4.3–4.4

4.15. Prove Theorem 4.8.

4.16. Prove (4.6) [See Exercise 3.8 for a partial proof.]

4.17. Familiar metrics on the set of probability distributions are the P. Lévy distance and the supremum distance. Show that if these are considered within the set D (of length n probability vectors), then the coefficient of ergodicity $\tau(P)$ (defined in Lemma 4.2) generated is identical to that generated by the l_∞ norm. (Detail on this coefficient may be found in Seneta, 1979.)

4.18. Suppose $\boldsymbol{\delta} = \{\delta_i\}$ is a real vector satisfying $\boldsymbol{\delta} \neq \mathbf{0}$, $\boldsymbol{\delta}'\mathbf{1} = 0$, and $\boldsymbol{\delta}^* = (\delta_i^*)$ is defined by $(\boldsymbol{\delta}^*)' = \boldsymbol{\delta}'P$ for stochastic P. Let $\Delta = \sum |\delta_i| = \|\boldsymbol{\delta}'\|_1$, $\Delta^* = \sum |\delta_i^*| = \|(\boldsymbol{\delta}^*)'\|_1$, and j' denote a typical index for which $\delta_j^* \ge 0$. Show that $\tfrac{1}{2}\Delta^* = \sum_k \delta_k(\sum_{j'} p_{kj'})$, and proceed as in the proof of Theorem 3.1, with δ_k playing the role of u_k and $\sum_{j'} p_{kj'}$ playing the role of w_k to show

$$\tfrac{1}{2}\Delta^* \le \tfrac{1}{2}\Delta \max_{i,h} \sum_{j'} (p_{ij'} - p_{hj'}) \le \tfrac{1}{2}\Delta\tau_1(P),$$

so that $\Delta^* \le \Delta c(P)$.

Use this last inequality to prove Theorem 4.9; and the inequali' /

$$\Delta^* \le \Delta\tau_1(P)$$

to prove (c.f. Lemma 4.3)

$$\tau_1(P^{(1)}P^{(2)}) \le \tau_1(P^{(1)})\tau_1(P^{(2)}).$$

4.19. In the notation of the statement and proof of Theorem 4.10, suppose (4.9) does not necessarily hold, but continue to assume $T_{p,r} \in G_1$, $p \ge 0$, $r \ge 1$, and introduce the notation $\gamma_h = \min_{i,j}^{+} p_{ij}(h)$. Show that

$$a(T_{p,r}) \le \prod_{i=0}^{k-1} c(T_{p+it,t}) \le \prod_{i=0}^{k-1} \left\{1 - \prod_{j=1}^{t} \gamma_{p+it+j}\right\},$$

so that if

$$\sum_{i=0}^{\infty} \prod_{j=1}^{t} \gamma_{p+it+j} = \infty, \qquad p \geq 0,$$

weak ergodicity obtains.

4.20. Prove Lemmas 4.7 and 4.8.

4.21. Show by example that the set G_1 is not closed under multiplication. Show, however, that if $P \in G_1$, $Q \in G_1$ then either both or neither of PQ, $QP \in G_1$.

4.22. Show by example that it is possible that $A, B, C \in G_1$, such that AB, BC, $AC \in G_1$, but $ABC \notin G_1$, although $BAC \in G_1$. (Contrast with the result of Exercise 4.15.)

(Sarymsakov, 1953*a*)

4.23. Suppose that weak ergodicity obtains for a sequence $\{P_i\}$ of stochastic matrices, not necessarily members of G_1. Show that for each fixed $p \geq 0$, there exists a strictly increasing sequence of integers $\{m_i\}$, $i \geq 1$, such that

$$T_{m_i, m_{i+1}} \in M, \qquad i \geq 0$$

where $m_0 = p$.

(Sarymsakov (1953*a*); Sarymsakov & Mustafin (1957))

Hint: A row of an $n \times n$ stochastic matrix has at least one entry $\geq n^{-1}$.

4.24. Discuss the relation between M, G_2 and G_3 when the dimensions of the stochastic matrices are $n \times n$, where $n = 2, 3$. Discuss the relation between G_2 and G_3 when $n \geq 5$.

4.25. Show by examples that a Markov matrix is not necessarily a "normed" matrix of G_1 (i.e. a matrix of G_1 with positive diagonal); and vice versa. Thus neither of these classes contains the other.

4.26. Show that, for $n = 4$, a scrambling matrix $P = \{p_{ij}\}$ is "nearly Markov", in that there is a column j such that $p_{i_1, j} > 0$, $p_{i_2, j} > 0$, $p_{i_3, j} > 0$ for distinct i_1, i_2, i_3. Extend to $n > 4$.

Hint. For an $n \times n$ scrambling matrix there are $n(n-1)/2$ distinct pairs of (row) indices, but only n actual (column) indices.

4.27. Show that if P^k is scrambling for some positive integer k, then $P \in G_1$.

(Paz, 1963)

4.28. Let $P_1, \ldots, P_k$ be a finite set of stochastic matrices of the same order.

Show that if there is an r_0 such that all words in the Ps of length at least r_0 are scrambling, then each word in the $Ps \in G_1$.

Hint: Use Exercise 4.27.

(Paz, 1965)

4.29. Let $\{P_i\}$ be a weakly ergodic sequence of stochastic matrices, and let $\det\{P_i\}$ denote, as usual, the determinant of P_i. Show that

$$\sum_{i=1}^{\infty} (1 - |\det\{P_i\}|) = \infty.$$

(Sirazhdinov, 1950)

4.30. Let us call a stochastic matrix $P = \{p_{ij}\}$ *quasi-Markov* if for a proper subset A of index set $\mathscr{S} = \{1, 2, \ldots, n\}$

$$\sum_{j \in A} p_{ij} > 0, \qquad \text{for each } i \in \mathscr{S}.$$

[A Markov matrix is thus one where A consists of a single index].

Show that a scrambling matrix is quasi-Markov, but (by examples) that a quasi-Markov matrix $\notin G_1$ (i.e. is not regular) necessarily.

Hint: Use the approach of Exercise 4.26.

4.5 Strong Ergodicity for Forward Products

We have already noted that the definition of strong ergodicity for forward products $T_{p,r} = \{t_{i,j}^{(p,r)}\}$, $p \geq 0$, $r \geq 1$, of stochastic matrices formed from a sequence P_k, $k \geq 1$, is subsumed by that of the more general context of row allowable matrices considered in §3.3. Thus here the forward products are said to be strongly ergodic if for all i, j, p

$$t_{i,j}^{(p,r)} \xrightarrow{r \to \infty} v_j^{(p)}$$

independently of i. The limit vector $\boldsymbol{v}_p = \{v_j^{(p)}\}$ is again evidently a probability vector, and, as in §3.3, is easily shown to be independent of p.

Indeed virtually all the theory of §3.3 goes through without changes of proof for sometimes slightly more general structure of quantities being considered, to compensate for the stochasticity of the underlying sequence P_k, $k \geq 1$. We shall generally not need to give anew formally either definitions or proofs as a consequence. Firstly asymptotic homogeneity here reduces to the existence of a probability vector $\boldsymbol{D}$ such that $\boldsymbol{D}'P_k \to \boldsymbol{D}'$ as $k \to \infty$, and condition (C), as already noted (4.9) to $0 < \gamma \leq \min^+ p_{ij}(k)$.

Lemma 4.9. *Strong ergodicity of $T_{p,r}$, $p \geq 0, r \geq 1$ (with limit vector $\boldsymbol{v}$) implies asymptotic homogeneity (with respect to $\boldsymbol{v}$) of the sequence P_k, $k \geq 1$.*

PROOF. As for Lemma 3.6 (condition (C) is not needed). □

Theorem 4.12. *If all P_k, $k \geq 1$, contain a single essential class of indices, and condition (C) is satisfied, then asymptotic homogeneity of the P_k (with respect to a probability vector $\boldsymbol{D}$) is equivalent to*

$$\boldsymbol{e}_k \to \boldsymbol{e} \qquad (k \to \infty) \tag{4.10}$$

where $\boldsymbol{e}_k$ is the unique stationary distribution vector corresponding to P_k, and $\boldsymbol{e}$ is a limit vector. In the event that either (equivalent) condition holds, $\boldsymbol{D} = \boldsymbol{e}$.

PROOF. As for Theorem 3.4, with the change that we take $\mathscr{I}_k$ to be members of the finite set of all incidence matrices $\mathscr{I}(j)$, $j = 1, \ldots, t$ *containing a single essential class of indices* (and reference to irreducible matrices is generally replaced by reference to matrices of this kind). □

Corollary. *Under the prior conditions of Theorem 3.4, if strong ergodicity with limit vector $\boldsymbol{v}$ holds, then (4.10) holds with $\boldsymbol{e} = \boldsymbol{v}$.* □

Theorem 4.13. *Assume all P_k, $k \geq 1$, contain a single essential class of indices and satisfy condition (C); and*

$$\tau_1(T_{p,r}) \leq \beta < 1 \tag{4.11}$$

for all $r \geq t$ (for some $t \geq 1$), uniformly in $p \geq 0$. Then asymptotic homogeneity is necessary and sufficient for strong ergodicity.

Proof. As for Theorem 3.5, *mutatis mutandis*. In particular we use $\tau_1(\cdot)$ in place of $\tau_B(\cdot)$ and the corresponding distance generated by $\|\cdot\|_1$ in place of the projective distance, and do not need the strict positivity of vectors and matrices inherent in the use of the projective distance. □

Corollary. *If (4.11) holds, and $\boldsymbol{e}_k \xrightarrow{k\to\infty} \boldsymbol{e}$ for a sequence $\boldsymbol{e}_k$, $k \geq 1$, of stationary distribution vectors of the sequence of stochastic matrices P_k, $k \geq 1$, for some limit vector $\boldsymbol{e}$, then strong ergodicity holds.* □

Theorem 4.14. *If $P_k \to P$ (elementwise) as $k \to \infty$, where $P \in G_1$ (i.e. P is regular), then strong ergodicity obtains, and the limit vector $\boldsymbol{v}$ is the unique stationary distribution vector of P.*

Proof. As in the proof of Theorem 3.6; again we use $\tau_1(\cdot)$ in place of $\tau_B(\cdot)$, positivity of matrices is not needed, and the proof is somewhat simpler. Since P is regular, there is a $j_0 \geq 1$ such that P^{j_0} is Markov. Now for $p \geq 0$

$$\tau_1(T_{p,r}) = \tau_1(T_{p,r-j_0} T_{p+r-j_0,j_0}) \leq \tau_1(T_{p+r-j_0,j_0}) \tag{4.12}$$

for $r \geq j_0$. As $\tau_1(T_{k,j_0}) \to \tau_1(P^{j_0})$ as $k \to \infty$, by the continuity of τ_1, where $\tau_1(P^{j_0}) < 1$ since P^{j_0} is scrambling, so for $k \geq \alpha_0$, say, $\tau_1(T_{k,j_0}) \leq \beta < 1$. Hence for $r \geq j_0 + \alpha_0$, $= t$, say, from (4.12)

$$\tau_1(T_{p,r}) \leq \beta < 1$$

for all $p \geq 0$. This is condition (4.11) of Theorem 4.13. As in the proof of Theorem 3.6, it is easy to prove that the conditions of the Corollary to Theorem 4.13 are otherwise satisfied (with $\boldsymbol{e}$ being the unique stationary distribution of P, and also—by Lemma 4.9—the limiting distribution $\boldsymbol{v}$ in the strong ergodicity). □

Theorem 4.15. *If $T_{p,r} \in G_1$, $p \geq 0$, $r \geq 1$, and condition (C) is satisfied, asymptotic homogeneity is necessary and sufficient for strong ergodicity.*[1]

Proof. From Theorem 4.13, we need only verify that (4.11) holds, since $T_{p,r} \in G_1$, $p \geq 0$, $r \geq 1 \Rightarrow P_k \in G_1$, $k \geq 1$. From Lemma 4.5, $T_{p,r} \in M$ for $r \geq t$, and for such r by (4.6) and condition (C)

$$\tau_1(T_{p,r}) \leq \tau_1(T_{p,t}) \leq c(T_{p,t}) \leq 1 - \gamma^t < 1.$$ □

[1] This result is a strong ergodicity version of Theorem 4.10, and the analogue of Theorem 3.7.

We conclude this section with a uniformity result analogous to Theorem 3.8.

Theorem 4.16. *Suppose $\mathscr{A}$ is any set of stochastic matrices such that $\mathscr{A} \subset G_1$ and each matrix satisfies condition (C). For $H \in \mathscr{A}$ let $e(H)$ be the unique stationary distribution vector, and suppose $\boldsymbol{x}$ is any probability vector. Then for $r \geq t$, where t is the number of distinct incidence matrices corresponding to G_1,*

$$\|x'H^r - e'(H)\|_1 \leq K\beta^{r/t}$$

where $K > 0$, $0 \leq \beta < 1$, both independent of H and $\boldsymbol{x}$.

Proof. Proceeding as in the proof of Theorem 4.10 and by (4.6), for $r \geq t$

$$\tau_1(T_{p,r}) \leq \prod_{i=0}^{k-1} \tau_1(T_{p+it,t})$$
$$\leq \prod_{i=0}^{k-1} c(T_{p+it,t}) \leq (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r$$

where t is the number of distinct incidence matrices corresponding to G_1, where k is the largest positive integer such that $kt \leq r$. Hence for any $H \in \mathscr{A}$

$$\tau_1(H^r) \leq (1-\gamma^t)^{-1}\beta^{r/t}$$

where $\beta = (1-\gamma^t)$. Hence

$$\|x'H^r - e'(H)\|_1 = \|x'H^r - e'(H)H^r\|_1 \leq \tau_1(H^r)\|x' - e'(H)\|_1$$

by Lemma 4.3;

$$< 2\tau_1(H^r).$$

Hence the result follows by taking $K = 2(1-\gamma^t)^{-1}$. □

Bibliography and Discussion to §4.5

This section has been written to closely parallel §3.3; the topics of both sections are treated in unified manner (as is manifestly possible) in Seneta and Sheridan (1981). Theorem 4.14 is originally due to Mott (1957) and Theorem 4.15 to Seneta (1973a). It is clear that, in Theorem 4.16, t can be taken as any upper bound over $(n \times n)$ $P \in G_1$ for the least integer r for which P^r has a positive column; according to Isaacson and Madsen (1974), we may take $t = (n-1)(n-2) + 1 = n^2 - 3n + 3$ (cf. Theorem 3.8).

One of the earliest theorems on strong ergodicity is due to Fortet (1938, p. 524) who shows that if P is regular and

$$\sum_k \|P_k - P\|_\infty < \infty$$

where

$$\|A\|_\infty = \max_i \left\{\sum_j |a_{ij}|\right\}$$

with $A = \{a_{ij}\}$, $i, j = 1, \ldots, n$, then as $r \to \infty$, lim $(r \to \infty)T_{p,r}$ exists, $p \geq 0$; this result is subsumed by Theorem 4.14. For a sequence of uniform Markov matrices P_k, $k \geq 1$, Theorem 4.15, which is the strong ergodicity extension of the Sarymsakov–Mustafin Theorem (Theorem 4.10), was obtained by Bernstein (1946) (see also Mott (1957) and Exercise 4.32). The notion of asymptotic homogeneity is due in this context to Bernstein (1946, 1964). Important early work on strong ergodicity was also carried out by Kozniewska (1962); a presentation of strong ergodicity theory along the lines of Bernstein and Kozniewska without use of the explicit notion of coefficient of ergodicity may be found in Seneta (1973c, §4.3), and in part in the exercises to the present section. The interested reader should also consult the papers of Hajnal (1956) and Mott and Schneider (1957); and the books of Isaacson and Madsen (1976) and Iosifescu (1977) for further material and references.

There is also a very large literature pertaining to probabilistic aspects of non-homogeneous finite Markov chains other than weak and strong ergodicity, e.g. the Central Limit Theorem and Law of the Iterated Logarithm. This work has been carried on largely by the Russian school; a comprehensive reference list to it may be found in Sarymsakov (1961), and an earlier one in Doeblin (1937). Much of the early work is due to Bernstein (see Bernstein (1964) and other papers in the same collection).

Exercises on §4.5

4.31. We say the sequence of stochastic matrices P_k, $k \geq 1$, is *asymptotically stationary* if there exists a probability vector $\boldsymbol{D}$ such that

$$\lim_{r\to\infty} \boldsymbol{D}'T_{p,r} = \boldsymbol{D}', \qquad p \geq 0.$$

Show that (*i*) asymptotic stationarity implies asymptotic homogeneity; and, more generally: (*ii*) asymptotic stationarity implies

$$\lim_{p\to\infty} \boldsymbol{D}'T_{p,r} = \boldsymbol{D}', \qquad r \geq 1.$$

(Kozniewska, 1962; Seneta, 1973a)

4.32. If the sequence of stochastic matrices P_k, $k \geq 1$, is uniformly Markov (i.e. $c(P_k) \leq c_0 < 1$, $k \geq 1$), show that asymptotic homogeneity is necessary and sufficient for strong ergodicity. (*Hint*: $\gamma \mathscr{I}_k \leq P_k \leq \mathbf{1}\mathbf{1}'$ where $\gamma = 1 - c_0$ and $\mathscr{I}_k$ is one of the matrices $\mathbf{1}\boldsymbol{f}'_j$, $j = 1, 2, \ldots, n$.)

(Bernstein, 1946)

4.33. Use the results of Exercises 4.31 and 4.32 to show that for a sequence of uniformly Markov matrices, asymptotic homogeneity and asymptotic stationarity are equivalent.

(Bernstein, 1946)

4.34. Show that if weak ergodicity obtains for the forward products $T_{p,r}$, $p \geq 0$, $r \geq 1$, formed from a sequence P_k, $k \geq 1$, of stochastic matrices, then asymptotic stationarity is equivalent to strong ergodicity.

(Kozniewska, 1962)

4.35. If $P_k = P$, $k \geq 1$, (i.e. all P_k's have common value P), show that weak and strong ergodicity are equivalent.

(Kozniewska, 1962)

4.6 Backwards Products

As in §3.1 we may consider general (rather than just forward) products $H_{p,r} = \{h_{i,j}^{(p,r)}\}$, $p \geq 0$, $r \geq 1$, formed from a given sequence P_k, $k \geq 1$, of stochastic matrices: $H_{p,r}$ is a product formed in any order from $P_{p+1}, P_{p+2}, \ldots, P_{p+r}$. From Lemma 4.3, it follows that

$$\tau_1(H_{p,r}) \leq \prod_{i=1}^{r} \tau_1(P_{p+i})$$

and it is clear from §§4.3–4.4 that a theory of weak ergodicity for such arbitrary products may be developed to some extent.[1]

Of particular interest from the point of view of applications is the behaviour as $r \to \infty$ of the backwards products

$$U_{p,r} = \{u_{i,j}^{(p,r)}\} = P_{p+r} \cdots P_{p+2}P_{p+1}, \qquad p \geq 0, \quad r \geq 1$$

for reasons which we now indicate.

A group of individuals, each of whom has an estimate of an unknown quantity engage in an information-exchanging operation. This unknown quantity may be the value of an unknown parameter, or a probability. When the individuals are made aware of each others' estimates, they modify their own estimate by taking into account the opinion of others; each individual weights the several estimates according to his opinion of their reliabilities. To obtain a quantitative formulation of the model, suppose there are n individuals, and let their initial estimates be given by the entries of the vector $\boldsymbol{F}_0 = (F_1^0, F_2^0, \ldots, F_n^0)$. Let $p_{ij}(1)$ be the initial weight which the ith individual attaches to the opinion of the jth individual. After the first interchange of information, the ith individual's estimate becomes

$$F_i^1 = \sum_j p_{ij}(1)F_j^0,$$

where the $p_{ij}(1)$'s can be taken to be normalized so that

$$\sum_j p_{ij}(1) = 1, \qquad i = 1, \ldots, n.$$

[1] See Exercise 4.36.

Clearly, the F_i's may be elements of any convex set in an appropriate linear space, rather than just real numbers; in particular, they may be probability distributions. Now write $P_k = \{p_{ij}(k)\}$, $k \geq 1$, $i, j = 1, \ldots, n$, where $p_{ij}(k)$ is the weight attached by the ith individual to the estimate of the jth individual after k interchanges of information, properly normalized. If $\boldsymbol{F}_k$ is the estimate vector resulting, then

$$\boldsymbol{F}_k = P_k \boldsymbol{F}_{k-1} = P_k P_{k-1} \cdots P_1 \boldsymbol{F}_0 = U_{0,k} \boldsymbol{F}_0$$

where the P_k, $k \geq 1$, are each stochastic matrices. The interest is clearly in the behaviour of $U_{0,k}$ as $k \to \infty$, with respect to: (1) whether *consensus* tends to be obtained (i.e. whether the elements of $\boldsymbol{F}_k$ tend to become the same), clearly a limited interpretation of what has been called weak ergodicity; and (*ii*) whether the opinions tend to stabilize at the same fixed opinion, a limited interpretation of what has been called strong ergodicity, for backwards products.

One may also think of the set P_k, $k \geq 1$, in this context as one-step transition matrices corresponding to an inhomogeneous Markov chain starting in the infinitely remote past and ending at time 0, P_k being the transition matrix at time $-k$, and $U_{p,r} = P_{p+r} \cdots P_{p+2} P_{p+1}$ as the r-step transition matrix between time $-(p+r)$ and time $-p$. In this setting it makes particular sense to consider the existence of a set of probability vectors $\boldsymbol{v}_k$, $k \geq 0$, such that

$$\boldsymbol{v}'_{r+p} U_{p,r} = \boldsymbol{v}'_p, \qquad p \geq 0, \quad r \geq 1, \tag{4.13}$$

the set $\boldsymbol{v}_k$, $k \geq 0$, then having the interpretation of *absolute probability vectors* at the various "times" $-k$, $k \geq 0$. We shall given them this name in general.

Finally we shall use the definitions of weak and strong ergodicity analogous to those for forward products (Definitions 4.4 and 4.5) by saying *weak ergodicity* obtains if

$$u_{i,s}^{(p,r)} - u_{j,s}^{(p,r)} \to 0 \tag{4.14}$$

as $r \to \infty$ for each i, j, s, p and *strong ergodicity* obtains if weak ergodicity obtains and the $u_{i,s}^{(p,r)}$ themselves tend to a limit for all i, s, p as $r \to \infty$ (in which case the limit of $u_{i,s}^{(p,r)}$ is independent of i; *but not necessarily*, as with forward products, of p[1]). The reason for the informality of definitions is the following:

Theorem 4.17. *For backwards products $U_{p,r}$, $p \geq 0$, $r \geq 1$, weak and strong ergodicity are equivalent.*

Proof. We need prove only that weak ergodicity implies strong ergodicity. Fix $p \geq 0$ and $\varepsilon > 0$; then by weak ergodicity

$$-\varepsilon \leq u_{i,s}^{(p,r)} - u_{j,s}^{(p,r)} \leq \varepsilon$$

[1] See Exercise 4.39.

for $r \geq r_0(p)$, uniformly for all $i, j, s = 1, \ldots, n$. Since

$$U_{p,r+1} = P_{p+r+1} U_{p,r},$$

$$\sum_{j=1}^{n} p_{hj}(p+r+1)(u_{i,s}^{(p,r)} - \varepsilon) \leq \sum_{j=1}^{n} p_{hj}(p+r+1) u_{j,s}^{(p,r)}$$

$$\leq \sum_{j=1}^{n} p_{hj}(p+r+1)(u_{i,s}^{(p,r)} + \varepsilon)$$

i.e.

$$u_{i,s}^{(p,r)} - \varepsilon \leq u_{h,s}^{(p,r+1)} \leq u_{i,s}^{(p,r)} + \varepsilon.$$

By induction

$$u_{i,s}^{(p,r)} - \varepsilon \leq u_{h,s}^{(p,r+k)} \leq u_{i,s}^{(p,r)} + \varepsilon$$

for all $i, s, h = 1, \ldots, n$, $p \geq 0$, $r \geq r_0(p)$, $k \geq 0$. Putting $i = h$, it is evident that $u_{i,s}^{(p,r)}$ is a Cauchy sequence, so $\lim\ (r \to \infty) u_{i,s}^{(p,r)}$ exists. □

Hence we need only speak of ergodicity of the $U_{p,r}$, and it is sufficient to prove "weak ergodicity" (4.14). We may, on the other hand, handle weak ergodicity easily through use of coefficients of ergodicity as in §§4.3–4.4, since scalar relations in terms of these, such as (4.4), (4.7), and that given by Lemma 4.3, are "direction-free".

Theorem 4.18. *Suppose $m(\cdot)$ and $\tau(\cdot)$ are proper coefficients of ergodicity satisfying* (4.4). *Ergodicity of backwards products $U_{p,r}$ formed from a given sequence P_k, $k \geq 1$, obtains if and only if there is a strictly increasing sequence of positive integers $\{k_s\}$, $s = 0, 1, 2, \ldots$ such that*

$$\sum_{s=0}^{\infty} \{1 - \tau(U_{k_s,\, k_{s+1}-k_s})\} = \infty.$$

Proof. As indicated for Theorem 4.8. □

Results for backwards products analogous to Theorems 4.9 and 4.14 for weak and strong ergodicity of forward products respectively, are set as Exercises 4.36 and 4.38. Lemma 4.5 for forward products has its analogue in Exercise 4.37, which can be used to prove the analogue of Theorem 4.10 (weak ergodicity for forward products) and Theorem 4.15 (strong ergodicity).

Theorem 4.19. *If for each $p \geq 0$, $r \geq 1$, $U_{p,r} \in G_1$ and*

$$\min_{i,j}{}^{+}\, p_{ij}(k) \geq \gamma > 0$$

uniformly for all $k \geq 1$, then ergodicity obtains at a uniform geometric rate for all $p \geq 0$. [This is true in particular if $U_{p,r} \in G_1$, $p \geq 0$, $r \geq 1$, and the sequence

$\{P_k\}$ has its elements selected from a numerically finite set of stochastic matrices.]

PROOF. Proceeding analogously to the proof of Theorem 4.10, with the same meaning for t, we obtain

$$a(U_{p,r}) \le (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r, \qquad r \ge t,$$

so that for all i, j, s, p, from the definition of $a(\cdot)$

$$|u_{i,s}^{(p,r)} - u_{j,s}^{(p,r)}| \le (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r, \qquad r \ge t.$$

Proceeding as in the proof of Theorem 4.17, for all $i, s, h = 1, \dots, n$, $p \ge 0$, $r \ge t$, $k \ge 0$

$$|u_{i,s}^{(p,r+k)} - u_{h,s}^{(p,r)}| \le (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r, \qquad r \ge t,$$

and letting $k \to \infty$ yields

$$|v_s^{(p)} - u_{h,s}^{(p,r)}| \le (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r, \qquad r \ge t,$$

where

$$v_s^{(p)} = \lim_{r\to\infty} u_{i,s}^{(p,r)}, \qquad i = 1, \dots, n. \qquad \square$$

The following result also gives a condition equivalent to ergodicity for backwards products.

Theorem 4.20. *Backwards products $U_{p,r}$, $p \ge 0$, $r \ge 1$, formed from a sequence P_k, $k \ge 1$, of stochastic matrices are ergodic if and only if there is only one set of absolute probability vectors $\boldsymbol{v}_k$, $k \ge 0$, in which case*

$$U_{p,r} \xrightarrow{r\to\infty} \mathbf{1}\boldsymbol{v}'_p, \qquad p \ge 0.$$

PROOF. We have for $p \ge 0$, $r \ge 1$, $h \ge 1$

$$U_{p+r,h} U_{p,r} = U_{p,r+h}. \tag{4.15}$$

Now, since the set of stochastic matrices is compact in R_{n^2}, we may use the Cantor diagonal argument to select a subsequence of the positive integers s_i such that as $i \to \infty$

$$U_{x, s_i - x} \to V_x \tag{4.16}$$

for each $x \ge 0$, for stochastic matrices V_x, $x \ge 0$. Hence substituting $s_i - p - r$ for h in (4.15) and letting $i \to \infty$ we obtain

$$V_{p+r} U_{p,r} = V_p, \qquad p \ge 0, \quad r \ge 1,$$

so there is always at least one set of absolute probability vectors given by $f'_j V_k$, $k \ge 0$, for any particular fixed $j = 1, \dots, n$ (i.e. we use the jth row of each V_k, $k \ge 0$).

Now suppose strong ergodicity holds, so that

$$U_{p,r} \xrightarrow{r\to\infty} \mathbf{1}\boldsymbol{v}'_p, \qquad p \ge 0,$$

and suppose $\bar{\boldsymbol{v}}_k$, $k \ge 0$, is any set of absolute probability vectors. Then

$$\bar{\boldsymbol{v}}'_p = \bar{\boldsymbol{v}}'_{p+r} U_{p,r} = \bar{\boldsymbol{v}}'_{p+r}(\mathbf{1}\boldsymbol{v}'_p + E_{p,r}) = \boldsymbol{v}'_p + \bar{\boldsymbol{v}}'_{p+r} E_{p,r}$$

where $E_{p,r} \to 0$ as $r \to \infty$, so, since $\bar{\boldsymbol{v}}'_{p+r}$ is bounded, being a probability vector, $\bar{\boldsymbol{v}}_p = \boldsymbol{v}_p$, $p \ge 0$.

Conversely, suppose there is precisely one set of absolute probability vectors $\boldsymbol{v}_k$, $k \ge 0$. Then suppose U_q^* is a limit point of $U_{q,r}$ as $r \to \infty$ for fixed but arbitrary $q \ge 0$, so that for some subsequence r_j, $j \ge 1$, of the integers, as $j \to \infty$,

$$U_{q,r_j} \to U_p^*. \tag{4.17}$$

Now use the sequence r_{j+q}, $j \ge 1$, from which to ultimately select the subsequence s_i giving (4.16). Following through the earlier argument, we have $V_x = \mathbf{1}\boldsymbol{v}'_x$, $x \ge 0$, by assumed uniqueness of the set of absolute probability distributions. But, from (4.17)

$$U_q^* = \lim_{i\to\infty} U_{q, s_i - q} = V_q$$

from (4.16),

$$= \mathbf{1}\boldsymbol{v}'_q.$$

Hence for a fixed $q \ge 0$, the limit point is unique, so as $r \to \infty$

$$U_{q,r} \to \mathbf{1}\boldsymbol{v}'_q, \qquad q \ge 0. \qquad \square$$

Bibliography and Discussion to §4.6

The development of this section follows Chatterjee and Seneta (1977) to whom Theorems 4.17–4.19 are due. Theorem 4.20 is due to Kolmogorov (1936b); for a succinct reworking see Blackwell (1945). In the special case where the P_k, $k \ge 1$, are drawn from a finite alphabet K, Theorem 4.19 was also obtained by Anthonisse and Tijms (1977).

Our motivating model for the estimate-modification process which has been used in this section is due to de Groot (1974), whose own motivation is the problem of attaining agreement about subjective probability distributions. He gives a range of references; a survey is given by Winkler (1968). Another situation of applicability arises in forecasting, where several individuals interact with each other while engaged in making the forecast (Delphi method; see Dalkey (1969)). The scheme

$$\boldsymbol{F}_k = P_k \boldsymbol{F}_{k-1}, \qquad k \ge 1,$$

represents an inhomogeneous version of a procedure described by Feller (1968) as "repeated averaging"; Feller, like de Groot, considers only the case where all P_k are the same, i.e. $P_k = P$, $k \geq 1$. In this special case, Theorem 4.19 and Exercise 4.38 are essentially due to de Groot.

For the rather complex behaviour of $U_{p,r}$ as $r \to \infty$ *in general* the reader should consult Blackwell (1945, Theorem 3), Pullman (1966, Theorem 1); and Cohn (1974) for an explanation in terms of the tail σ-field of a reverse Markov chain.

A further relevant reference is Mukherjea (1979, Section 3).

Exercises on §4.6

4.36. Defining weak ergodicity for arbitrary products $H_{p,r}$, $p \geq 0$, $r \geq 1$, by

$$h_{i,s}^{(p,r)} - h_{j,s}^{(p,r)} \to 0$$

as $r \to \infty$ for each i, j, s, p, show that weak ergodicity is equivalent to $\tau_1(H_{p,r}) \to 0$ as $r \to \infty$, $p \geq 0$. Show that sufficient for such weak ergodicity is $\sum_{k=1}^{\infty} \{1 - \tau_1(P_k)\} = \infty$.

4.37. Show that if $U_{p,r} \in G_1$, $p \geq 0$, $r \geq 1$, then $U_{p,r} \in M$ for $r \geq t$, where t is the number of distinct incidence matrices corresponding to G_1. (*Hint:* Lemmas 4.4 and 4.5.)

4.38. Show that if $P_k \to P$ (elementwise) as $k \to \infty$, where $P \in G_1$, then ergodicity holds for the backward products $U_{p,r}$, $p \geq 0$, $r \geq 1$. (*Hint:* Show that there exists a p_0 and a t such that $c(U_{p,t}) \leq c_0 < 1$ uniformly for $p \geq p_0$, and use the approach of the proof of Theorem 4.10.)

(Chatterjee and Seneta, 1977)

4.39. Suppose the backward products $U_{p,r}$, $p \geq 0$, $r \geq 1$, formed from a sequence P_k, $k \geq 1$, of stochastic matrices, are ergodic, so that as $r \to \infty$

$$U_{p,r} \to \mathbf{1}v'_p, \qquad p \geq 0,$$

where the limit vectors may or may not depend on p. By using the fact that $U_{p,r}\mathbf{1}v' = \mathbf{1}v'$ for any probability vector v', construct another sequence for which the limit vectors are not all the same (i.e. depend on p).

4.40. An appropriate analogy for backwards products to the class G_2 given by Definition 4.8 in the G'_2 of stochastic matrices, defined by $P \in G'_2$ if (*a*) $P \in G_1$; (*b*) $PQ \in G_1$ for any $Q \in G_1$. Discuss why even more useful might be the class $\bar{G}_2 = \{P;\, P \in G_1;\, PQ, QP \in G_1 \text{ for any } Q \in G_1\} = G_2 \cap G'_2$. Show that $\bar{G}_2$ is a strictly larger class than G_3, the class of $(n \times n)$ scrambling matrices.

PART II

COUNTABLE NON-NEGATIVE MATRICES