Springer Series in Statistics

Springer Series in Statistics

Andersen/Borgan/Gill/Keiding: Statistical Models Based on Counting Processes.
Atkinson/Riani: Robust Diagnostic Regression Analysis.
Atkinson/Riani/Cerioli: Exploring Multivariate Data with the Forward Search.
Berger: Statistical Decision Theory and Bayesian Analysis, 2nd edition.
Borg/Groenen: Modern Multidimensional Scaling: Theory and Applications, 2nd edition.
Brockwell/Davis: Time Series: Theory and Methods, 2nd edition.
Bucklew: Introduction to Rare Event Simulation.
Cappé/Moulines/Rydén: Inference in Hidden Markov Models
Chan/Tong: Chaos: A Statistical Perspective.
Chen/Shao/Ibrahim: Monte Carlo Methods in Bayesian Computation.
Coles: An Introduction to Statistical Modeling of Extreme Values.
David/Edwards: Annotated Readings in the History of Statistics.
Devroye/Lugosi: Combinatorial Methods in Density Estimation.
Efromovich: Nonparametric Curve Estimation: Methods, Theory, and Applications.
Eggermont/LaRiccia: Maximum Penalized Likelihood Estimation, Volume I: Density Estimation.
Fahrmeir/Tutz: Multivariate Statistical Modelling Based on Generalized Linear Models, 2nd edition.
Fan/Yao: Nonlinear Time Series: Nonparametric and Parametric Methods.
Farebrother: Fitting Linear Relationships: A History of the Calculus of Observations 1750-1900.
Federer: Statistical Design and Analysis for Intercropping Experiments, Volume I: Two Crops.
Federer: Statistical Design and Analysis for Intercropping Experiments, Volume II: Three or More Crops.
Ghosh/Ramamoorthi: Bayesian Nonparametrics.
Glaz/Naus/Wallenstein: Scan Statistics.
Good: Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses, 2nd edition.
Good: Permutation Tests: Parametric and Bootstrap Tests of Hypotheses, 3rd edition.
Gouriéroux: ARCH Models and Financial Applications.
Gu: Smoothing Spline ANOVA Models.
Györfi/Kohler/Krzyżak/ Walk: A Distribution-Free Theory of Nonparametric Regression.
Haberman: Advanced Statistics, Volume I: Description of Populations.
Hall: The Bootstrap and Edgeworth Expansion.
Härdle: Smoothing Techniques: With Implementation in S.
Harrell: Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.
Hart: Nonparametric Smoothing and Lack-of-Fit Tests.
Hastie/Tibshirani/Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction.
Hedayat/Sloane/Stufken: Orthogonal Arrays: Theory and Applications.
Heyde: Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation.

(continued after index)

E. Seneta

Non-negative Matrices and Markov Chains

Revised Printing

Springer

E. Seneta
The University of Sydney
School of Mathematics and Statistics
Sydney, N.S.W. 2006
Australia
eseneta@maths.usyd.edu.au

AMS Classification (1980): 15A48, 15A51, 15A52, 60G10, 62E99

Library of Congress Control Number: 2005935207

ISBN-10: 0-387-29765-0 Printed on acid-free paper.
ISBN-13: 978-0387-29765-1

The first edition of this book was published by George Allen & Unwin Ltd. (London,1973)

Printed in the United States of America. (IBT)

9 8 7 6 5 4 3 2 1

springer.com

To my parents

Things are always at their best in the beginning.
Blaise Pascal, *Lettres Provinciales* [1656–1657]

Preface

Since its inception by Perron and Frobenius, the theory of non-negative matrices has developed enormously and is now being used and extended in applied fields of study as diverse as probability theory, numerical analysis, demography, mathematical economics, and dynamic programming, while its development is still proceeding rapidly as a branch of pure mathematics in its own right. While there are books which cover this or that aspect of the theory, it is nevertheless not uncommon for workers in one or another branch of its development to be unaware of what is known in other branches, even though there is often formal overlap. One of the purposes of this book is to relate several aspects of the theory, insofar as this is possible.

The author hopes that the book will be useful to mathematicians; but in particular to the workers in applied fields, so the mathematics has been kept as simple as could be managed. The mathematical requisites for reading it are: some knowledge of real-variable theory, and matrix theory; and a little knowledge of complex-variable; the emphasis is on real-variable methods. (There is only one part of the book, the second part of §5.5, which is of rather specialist interest, and requires deeper knowledge.) Appendices provide brief expositions of those areas of mathematics needed which may be less generally known to the average reader.

The first four chapters are concerned with finite non-negative matrices, while the following three develop, to a small extent, an analogous theory for infinite matrices. It has been recognized that, generally, a research worker will be interested in one particular chapter more deeply than in others; consequently there is a substantial amount of independence between them. Chapter 1 should be read by every reader, since it provides the foundation for the whole book; thereafter Chapters 2–4 have some interdependence as do Chapters 5–7. For the reader interested in the infinite matrix case, Chapter 5 should be read before Chapters 6 and 7. The exercises are intimately connected with the text, and often provide further development of the theory or deeper insight into it, so that the reader is strongly advised to (at least) look over the exercises relevant to his interests, even if not actually wishing to do them. Roughly speaking, apart from Chapter 1, Chapter 2 should be of interest to students of mathematical economics, numerical analysis, combinatorics, spectral theory of matrices, probabilists and statisticians; Chapter 3 to mathematical economists and demographers; and Chapter 4 to probabilists. Chapter 4 is believed to contain one of the first expositions in text-book form of the theory of finite inhomogeneous Markov chains, and contains due regard for Russian-language literature. Chapters 5–7 would at present appear to be of interest primarily to probabilists, although the probability emphasis in them is not great.

This book is a considerably modified version of the author's earlier book *Non-Negative Matrices* (Allen and Unwin, London/Wiley, New York, 1973, hereafter referred to as *NNM*). Since *NNM* used probabilistic techniques throughout, even though only a small part of it explicitly dealt with probabilistic topics, much of its interest appears to have been for people acquainted with the general area of probability and statistics. The title has, accordingly, been changed to reflect more accurately its emphasis and to account for the expansion of its Markov chain content. This has gone hand-in-hand with a modification in approach to this content, and to the treatment of the more general area of inhomogeneous products of non-negative matrices, via "coefficients of ergodicity," a concept not developed in *NNM*.

Specifically in regard to modification, §§2.5–§2.6 are completely new, and §2.1 has been considerably expanded, in Chapter 2. Chapter 3 is completely new, as is much of Chapter 4. Chapter 6 has been modified and expanded and there is an additional chapter (Chapter 7) dealing in the main with the problem of practical computation of stationary distributions of infinite Markov chains from finite truncations (of their transition matrix), an idea also used elsewhere in the book.

It will be seen, consequently, that apart from certain sections of Chapters 2 and 3, the present book as a whole may be regarded as one approaching the theory of Markov chains from a non-negative matrix standpoint.

Since the publication of *NNM*, another English-language book dealing exclusively with non-negative matrices has appeared (A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York, 1979). The points of contact with either *NNM* or its present modification (both of which it complements in that its level, approach, and subject matter are distinct) are few. The interested reader may consult the author's review in *Linear and Multilinear Algebra*, 1980, **9**; and may wish to note the extensive bibliography given by Berman and Plemmons. In the present book we have, accordingly, only added references to those of *NNM* which are cited in new sections of our text.

In addition to the acknowledgements made in the Preface to *NNM*, the author wishes to thank the following: S. E. Fienberg for encouraging him to write §2.6 and Mr. G. T. J. Visick for acquainting him with the non-statistical evolutionary line of this work; N. Pullman, M. Rothschild and R. L. Tweedie for materials supplied on request and used in the book; and Mrs. Elsie Adler for typing the new sections.

Sydney, 1980 E. SENETA

Contents

Glossary of Notation and Symbols

T	usual notation for a non-negative matrix.
A	typical notation for a matrix.
A'	the transpose of the matrix A.
a_{ij}	the (i, j) entry of the matrix A.
$\tilde{T}$	the incidence matrix of the non-negative matrix T.
P	usual notation for a stochastic matrix.
0	zero; the zero matrix.
$\mathbf{x}$	typical notation for a column vector.
$\mathbf{0}$	the column vector with all entries 0.
$\mathbf{1}$	the column vector with all entries 1.
$P_1 \sim P_2$	the matrix P_1 has the same incidence matrix as the matrix P_2.
$\min^+$	the minimum among all strictly positive elements.
R^k, R_k	k-dimensional Euclidean space.
R	set of strictly positive integers; convergence parameter of an irreducible matrix T; a certain submatrix.
I	the identity (unit) matrix; the set of inessential indices.
$i \in R$	i is an element of the set R.
$\mathscr{C} \subset \mathscr{S}$ or $\mathscr{C} \subseteq \mathscr{S}$	$\mathscr{C}$ is a subset of the set $\mathscr{S}$.
$_{(n)}T$	$(n \times n)$ northwest corner truncation of T.
$\Delta_i, \Delta_i(s)$	the principal minor of $(sI - T)$.
$_{(n)}\Delta(\beta)$	$\det [_{(n)}I - \beta, _{(n)}T]$.
$_{(n)}\Delta$	$_{(n)}\Delta(R)$
M.C.	Markov chain.
$\mathscr{E}$	mathematical expectation operator.
G_1	class of $(n \times n)$ regular matrices.
M	class of $(n \times n)$ Markov matrices.
G_2	class of stochastic matrices defined on p. 143.
G_3	class of $(n \times n)$ scrambling matrices.

CHAPTER 1

Fundamental Concepts and Results in the Theory of Non-negative Matrices

We shall deal in this chapter with square non-negative matrices $T = \{t_{ij}\}$, $i, j = 1, \ldots, n$; i.e. $t_{ij} \geq 0$ for all i, j, in which case we write $T \geq 0$. If, in fact, $t_{ij} > 0$ for all i, j we shall put $T > 0$.

This definition and notation extends in an obvious way to row vectors (x') and column vectors (y), and also to expressions such as, e.g.

$$T \geq B \Leftrightarrow T - B \geq 0$$

where T, B and 0 are square non-negative matrices of compatible dimensions.

Finally, we shall use the notation $x' = \{x_i\}$, $y = \{y_i\}$ for both row vectors x' or column vectors y; and $T^k = \{t_{ij}^{(k)}\}$ for kth powers.

Definition 1.1. A square non-negative matrix T is said to be *primitive* if there exists a positive integer k such that $T^k > 0$.

It is clear that if any other matrix $\tilde{T}$ has the same dimensions as T, and has positive entries and zero entries in the same positions as T, then this will also be true of all powers T^k, $\tilde{T}^k$ of the two matrices.

As *incidence matrix* $\tilde{T}$ corresponding to a given T replaces all the positive entries of T by ones. Clearly $\tilde{T}$ is primitive if and only if T is.

1.1 The Perron—Frobenius Theorem for Primitive Matrices[1]

Theorem 1.1. *Suppose T is an $n \times n$ non-negative primitive matrix. Then there exists an eigenvalue r such that:*

(*a*) *r real, >0;*

[1] This theorem is fundamental to the entire book. The proof is necessarily long; the reader may wish to defer detailed consideration of it.

(*b*) *with r can be associated strictly positive left and right eigenvectors*;
(*c*) $r > |\lambda|$ *for any eigenvalue* $\lambda \neq r$;
(*d*) *the eigenvectors associated with r are unique to constant multiples.*
(*e*) *If* $0 \leq B \leq T$ *and* β *is an eigenvalue of B, then* $|\beta| \leq r$. *Moreover,* $|\beta| = r$ *implies* $B = T$.
(*f*) *r is a simple root of the characteristic equation of T.*

PROOF. (*a*) Consider initially a row vector $x' \geq \mathbf{0}'$, $\neq \mathbf{0}'$; and the product $x'T$. Let

$$r(x) = \min_j \frac{\sum_i x_i t_{ij}}{x_j}$$

where the ratio is to be interpreted as '∞' if $x_j = 0$. Clearly, $0 \leq r(x) < \infty$. Now since

$$x_j r(x) \leq \sum_i x_i t_{ij} \text{ for each } j,$$

$$x' r(x) \leq x'T,$$

and so

$$x'\mathbf{1} r(x) \leq x'T\mathbf{1}.$$

Since $T\mathbf{1} \leq K\mathbf{1}$ for $K = \max_i \sum_j t_{ij}$, it follows that

$$r(x) \leq x'\mathbf{1}K/x'\mathbf{1} = K = \max_i \sum_j t_{ij}$$

so $r(x)$ is uniformly bounded above for all such x. We note also that since T, being primitive, can have no column consisting entirely of zeroes, $r(\mathbf{1}) > 0$, whence it follows that

$$r = \sup_{\substack{x \geq \mathbf{0} \\ x \neq \mathbf{0}}} \min_j \frac{\sum_i x_i t_{ij}}{x_j} \tag{1.1}$$

satisfies

$$0 < r(\mathbf{1}) \leq r \leq K < \infty.$$

Moreover, since neither numerator or denominator is altered by the norming of x,

$$r = \sup_{\substack{x \geq \mathbf{0} \\ x'x = 1}} \min_j \frac{\sum_i x_i t_{ij}}{x_j}.$$

Now the region $\{x; x \geq \mathbf{0}, x'x = 1\}$ is compact in the Euclidean n-space R_n, and the function $r(x)$ is an upper-semicontinuous mapping of this region into R_1; hence[1] the supremum, r is actually *attained* for some x, say $\hat{x}$. Thus there exists $\hat{x} \geq \mathbf{0}$, $\neq \mathbf{0}$ such that

$$\min_j \frac{\sum_i \hat{x}_i t_{ij}}{\hat{x}_j} = r,$$

i.e.

$$\sum_i \hat{x}_i t_{ij} \geq r\hat{x}_j; \quad \text{or} \quad \hat{x}'T \geq r\hat{x}' \tag{1.2}$$

[1] see Appendix C.

for each $j = 1, \ldots, n$; with equality for some element of $\hat{x}$.

Now consider

$$z' = \hat{x}'T - r\hat{x}' \geq \mathbf{0}'.$$

Either $z = \mathbf{0}$, or not; if not, we know that for $k \geq k_0$, $T^k > 0$ as a consequence of the primitivity of T, and so

$$z'T^k = (\hat{x}'T^k)T - r(\hat{x}'T^k) > \mathbf{0}',$$

i.e.
$$\frac{\{(\hat{x}'T^k)T\}_j}{\{\hat{x}'T^k\}_j} > r, \text{ each } j,$$

where the subscript j refers to the jth element. This is a contradiction to the definition of r. Hence always

$$z = \mathbf{0},$$

whence
$$\hat{x}'T = r\hat{x}' \tag{1.3}$$

which proves (a).

(b) By iterating (1.3)

$$\hat{x}'T^k = r^k\hat{x}'$$

and taking k sufficiently large $T^k > 0$, and since $\hat{x} \geq \mathbf{0}$, $\neq \mathbf{0}$, in fact $\hat{x}' > \mathbf{0}'$.

(c) Let λ be any eigenvalue of T. Then for some $x \neq \mathbf{0}$ and possibly complex valued

$$\sum_i x_i t_{ij} = \lambda x_j \qquad \left(\text{so that } \sum_i x_i t_{ij}^{(k)} = \lambda^k x_j\right) \tag{1.4}$$

whence
$$|\lambda x_j| = \left|\sum_i x_i t_{ij}\right| \leq \sum_i |x_i| t_{ij},$$

so that
$$|\lambda| \leq \frac{\sum_i |x_i| t_{ij}}{|x_j|}$$

where the right side is to be interpreted as '∞' for any $x_j = 0$. Thus

$$|\lambda| \leq \min_j \frac{\sum_i |x_i| t_{ij}}{|x_j|},$$

and by the definition (1.1) of r

$$|\lambda| \leq r.$$

Now suppose $|\lambda| = r$; then

$$\sum_i |x_i| t_{ij} \geq |\lambda|\,|x_j| = r|x_j|$$

which is a situation identical to that in the proof of part (*a*), (1.2); so that eventually in the same way

$$\sum_i |x_i| t_{ij} = r|x_j|, >0; \qquad j = 1, 2, \ldots, n \tag{1.5}$$

and so $$\sum_i |x_i| t_{ij}^{(k)} = r^k |x_j|, > 0; \qquad j = 1, 2, \ldots, n,$$

i.e. $$\left|\sum_i x_i t_{ij}^{(k)}\right| = |\lambda^k x_j| = \sum_i |x_i t_{ij}^{(k)}| \tag{1.6}$$

where k can be chosen so large that $T^k > 0$, by the *primitivity* assumption on T; but for two numbers $\gamma, \delta \neq 0$, $|\gamma + \delta| = |\gamma| + |\delta|$ if and only if γ, δ have the same direction in the complex plane. Thus writing $x_j = |x_j| \exp i\theta_j$, (1.6) implies $\theta_j = \theta$ is independent of j, and hence cancelling the exponential throughout (1.4) we get

$$\sum_i |x_i| t_{ij} = \lambda |x_j|$$

where, since $|x_i| > 0$ all i, λ is real and positive, and since we are assuming $|\lambda| = r$, $\lambda = r$ (or the fact follows equivalently from (1.5)).

(*d*) Suppose $\boldsymbol{x}' \neq \boldsymbol{0}'$ is a left eigenvector (possibly with complex elements) corresponding to r.

Then, by the argument in (*c*), so is $\boldsymbol{x}'_+ = \{|x_i|\} \neq \boldsymbol{0}'$, which in fact satisfies $\boldsymbol{x}_+ > \boldsymbol{0}$. Clearly

$$\boldsymbol{\eta}' = \hat{\boldsymbol{x}}' - c\boldsymbol{x}'$$

is then also a left eigenvector corresponding to r, for any c such that $\boldsymbol{\eta} \neq \boldsymbol{0}$; and hence the same things can be said about $\boldsymbol{\eta}$ as about $\boldsymbol{x}$; in particular $\boldsymbol{\eta}_+ > \boldsymbol{0}$.

Now either $\boldsymbol{x}$ is a multiple of $\hat{\boldsymbol{x}}$ or not; if not c can be chosen so that $\boldsymbol{\eta} \neq \boldsymbol{0}$, but some element of $\boldsymbol{\eta}$ is; this is impossible as $\boldsymbol{\eta}_+ > \boldsymbol{0}$.

Hence $\boldsymbol{x}'$ is a multiple of $\hat{\boldsymbol{x}}'$.

Right eigenvectors. The arguments (*a*)–(*d*) can be repeated separately for right eigenvectors; (*c*) guarantees that the r produced is the same, since it is purely a statement about eigenvalues.

(*e*) Let $\boldsymbol{y} \neq \boldsymbol{0}$ be a right eigenvector of B corresponding to β. Then taking moduli as before

$$|\beta| \boldsymbol{y}_+ \leq B\boldsymbol{y}_+, \leq T\boldsymbol{y}_+, \tag{1.7}$$

so that using the same $\hat{\boldsymbol{x}}$ as before

$$|\beta| \hat{\boldsymbol{x}}'\boldsymbol{y}_+ \leq \hat{\boldsymbol{x}}' T\boldsymbol{y}_+ = r\hat{\boldsymbol{x}}'\boldsymbol{y}_+$$

and since $\hat{\boldsymbol{x}}'\boldsymbol{y}_+ > 0$,

$$|\beta| \leq r.$$

Suppose now $|\beta| = r$; then from (1.7)

$$ry_+ \leq Ty_+$$

whence, as in the proof of (b), it follows $Ty_+ = ry_+ > \mathbf{0}$; whence from (1.7)

$$ry_+ = By_+ = Ty_+$$

so it must follow, from $B \leq T$, that $B = T$.

(f) The following identities are true for all numbers, real and complex, including eigenvalues of T:

$$\left.\begin{array}{l}(xI - T)\operatorname{Adj}(xI - T) = \det(xI - T)I \\ \operatorname{Adj}(xI - T)(xI - T) = \det(xI - T)I\end{array}\right\} \tag{1.8}$$

where I is the unit matrix and 'det' refers to the determinant. (The relation is clear for x not an eigenvalue, since then $\det(xI - T) \neq 0$; when x is an eigenvalue it follows by continuity.)

Put $x = r$: then any one row of Adj $(rI - T)$ is either (i) a left eigenvector corresponding to r; or (ii) a row of zeroes; and similarly for columns. By assertions (b) and (d) (already proved) of the theorem, Adj $(rI - T)$ is either (i) a matrix with no elements zero; or (ii) a matrix with all elements zero. We shall prove that one element of Adj $(rI - T)$ is positive, which establishes that case (i) holds. The (n, n) element is

$$\det(r_{(n-1)}I - {}_{(n-1)}T)$$

where ${}_{(n-1)}T$ is T with last row and column deleted; and ${}_{(n-1)}I$ is the corresponding unit matrix. Since

$$0 \leq \begin{bmatrix} {}_{(n-1)}T & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} \leq T, \text{ and } \neq T,$$

the last since T is primitive (and so can have no zero column), it follows from (e) of the theorem that no eigenvalue of ${}_{(n-1)}T$ can be as great in modulus as r. Hence

$$\det(r_{(n-1)}I - {}_{(n-1)}T) > 0,$$

as required; *and moreover we deduce that* Adj $(rI - T)$ *has all its elements positive.*

Write $\phi(x) = \det(xI - T)$; then differentiating (1.8) elementwise

$$\operatorname{Adj}(xI - T) + (xI - T)\frac{d}{dx}\{\operatorname{Adj}(xI - T)\} = \phi'(x)I.$$

Substitute $x = r$, and premultiply by $\hat{x}'$;

$$(\mathbf{0}' <)\hat{x}' \operatorname{Adj}(rI - T) = \phi'(r)\hat{x}'$$

since the other term vanishes. Hence $\phi'(r) > 0$ and so r is simple. □

Corollary 1.

$$\min_i \sum_{j=1}^{n} t_{ij} \le r \le \max_i \sum_{j=1}^{n} t_{ij} \tag{1.9}$$

with equality on either side implying equality throughout (i.e. r can only be equal to the maximal or minimal row sum if all row sums are equal).

A similar proposition holds for column sums.

PROOF. Recall from the proof of part (a) of the theorem, that

$$0 < r(\mathbf{1}) = \min_j \sum_i t_{ij} \le r \le K = \max_i \sum_j t_{ij} < \infty. \tag{1.10}$$

Since T' is also primitive and has the same r, we have also

$$\min_j \sum_i t_{ji} \le r \le \max_i \sum_j t_{ji} \tag{1.11}$$

and a combination of (1.10) and (1.11) gives (1.9).

Now assume that one of the equalities in (1.9) holds, but not all row sums are equal. Then by increasing (or, if appropriate, decreasing) the positive elements of T (but keeping them positive), produce a new primitive matrix, with all row sums equal and the same r, in view of (1.9); which is impossible by assertion (e) of the theorem. □

Corollary 2. *Let $\boldsymbol{v}'$ and $\boldsymbol{w}$ be positive left and right eigenvectors corresponding to r, normed so that $\boldsymbol{v}'\boldsymbol{w} = 1$. Then*

$$\text{Adj}\,(rI - T)/\phi'(r) = \boldsymbol{w}\boldsymbol{v}'.$$

To see this, first note that since the columns of Adj $(rI - T)$ are multiples of the same positive right eigenvector corresponding to r (and its rows of the same positive left eigenvector) it follows that we can write it in the form $\boldsymbol{y}\boldsymbol{x}'$ where $\boldsymbol{y}$ is a right and $\boldsymbol{x}'$ a left positive eigenvector. Moreover, again by uniqueness, there exist positive constants c_1, c_2 such that $\boldsymbol{y} = c_1\boldsymbol{w}$, $\boldsymbol{x}' = c_2\boldsymbol{v}'$, whence

$$\text{Adj}\,(rI - T) = c_1 c_2 \boldsymbol{w}\boldsymbol{v}'.$$

Now, as in the proof of the simplicity of r,

$$\boldsymbol{v}'\phi'(r) = \boldsymbol{v}'\,\text{Adj}\,(rI - T) = c_1 c_2 \boldsymbol{v}'\boldsymbol{w}\boldsymbol{v}' = c_1 c_2 \boldsymbol{v}'$$

so that $\boldsymbol{v}'\boldsymbol{w}\phi'(r) = c_1 c_2 \boldsymbol{v}'\boldsymbol{w}$

i.e. $\qquad c_1 c_2 = \phi'(r)$ as required. □

(Note that $c_1 c_2$ = sum of the diagonal elements of the adjoint = sum of the principal $(n-1) \times (n-1)$ minors of $(rI - T)$.)

Theorem 1.1 is the strong version of the Perron–Frobenius Theorem which holds for primitive T; we shall generalize Theorem 1.1 to a wider class of matrices, called *irreducible*, in §1.4 (and shall refer to this generalization as the Perron–Frobenius Theory).

Suppose now the *distinct* eigenvalues of a primitive T are $r, \lambda_2, \ldots, \lambda_t$, $t \le n$ where $r > |\lambda_2| \ge |\lambda_3| \ge \cdots \ge |\lambda_t|$. In the case $|\lambda_2| = |\lambda_3|$ we stipulate that the multiplicity m_2 of λ_2 is at least as great as that of λ_3, *and of any other eigenvalue having the same modulus as* λ_2.

It may happen that a primitive matrix has $\lambda_2 = 0$; an example is a matrix of form

$$T = \begin{pmatrix} a & b & c \\ a & c & b \\ a & c & b \end{pmatrix} > 0 \tag{1.12}$$

for which $r = a + b + c$. This kind of situation gives the following theorem its dual form, the example (1.12) illustrating that in part (b) the bound $(n-1)$ cannot be reduced.

Theorem 1.2. *For a primitive matrix* T:

(a) *if* $\lambda_2 \ne 0$, *then as* $k \to \infty$

$$T^k = r^k w v' + O(k^s |\lambda_2|^k)$$

elementwise, where $s = m_2 - 1$;

(b) *if* $\lambda_2 = 0$, *then for* $k \ge n - 1$

$$T^k = r^k w v'.$$

In both cases w, v' *are any positive right and left eigenvectors corresponding to* r *guaranteed by Theorem* 1.1, *providing only they are normed so that* $v'w = 1$.

PROOF. Let $R(z) = (I - zT)^{-1} = \{r_{ij}(z)\}$, $z \ne \lambda_i^{-1}$, $i = 1, 2, \ldots$ (where $\lambda_1 = r$). Consider a general element of this matrix

$$r_{ij}(z) = \frac{c_{ij}(z)}{(1 - zr)(1 - z\lambda_2)^{m_2} \cdots (1 - z\lambda_t)^{m_t}}$$

where m_i is the multiplicity of λ_i and $c_{ij}(z)$ is a polynomial in z, of degree at most $n - 1$ (see Appendix B).

Here using partial fractions, *in case* (a)

$$r_{ij}(z) = p_{ij}(z) + \frac{a_{ij}}{(1 - zr)} + \sum_{s=0}^{m_2 - 1} \frac{b_{ij}^{(m_2 - s)}}{(1 - z\lambda_2)^{m_2 - s}}$$

+ similar terms for any other non-zero eigenvalues,

where the a_{ij}, $b_{ij}^{(m_2 - s)}$ are constants, and $p_{ij}(z)$ is a polynomial of degree at most $(n - 2)$. Hence for $|z| < 1/r$,

$$r_{ij}(z) = p_{ij}(z) + a_{ij} \sum_{k=0}^{\infty} (zr)^k + \sum_{s=0}^{m_2 - 1} b_{ij}^{(m_2 - s)} \left\{ \sum_{k=0}^{\infty} \binom{-m_2 + s}{k} (-z\lambda_2)^k \right\}$$

+ similar terms for other non-zero eigenvalues.

In matrix form, with obvious notation

$$R(z) = P(z) + A \sum_{k=0}^{\infty} (zr)^k + \sum_{s=0}^{m_2-1} B^{(m_2-s)} \left\{ \sum_{k=0}^{\infty} \binom{-m_2+s}{k} (-z\lambda_2)^k \right\}$$

+ possible like terms.

From Stirling's formula, as $k \to \infty$

$$\binom{-m_2+s}{k} \sim \text{const. } k^{m_2-s-1},$$

so that, identifying coefficients of z^k on both sides (see Appendix B) for large k

$$T^k = Ar^k + 0(k^{m_2-1} |\lambda_2|^k).$$

In case (b), we have merely, with the same symbolism as in case (a)

$$r_{ij}(z) = p_{ij}(z) + \frac{a_{ij}}{(1 - zr)}$$

so that for $k \geq n-1$,

$$T^k = Ar^k.$$

It remains to determine the nature of A. We first note that

$$T^k/r^k \to A \geq 0 \text{ elementwise, as } k \to \infty,$$

and that the series

$$\sum_{k=0}^{\infty} (r^{-1}T)^k z^k$$

has non-negative coefficients, and is convergent for $|z| < 1$, so that by a wellknown result (see e.g. Heathcote, 1971, p. 65).

$$\lim_{x \to 1-} (1 - x) \sum_{k=0}^{\infty} (r^{-1}T)^k x^k = A \text{ elementwise.}$$

Now, for $0 < x < 1$,

$$\sum_{k=0}^{\infty} (r^{-1}T)^k x^k = (I - r^{-1}xT)^{-1} = \frac{\text{Adj } (I - r^{-1}xT)}{\det (I - r^{-1}xT)}$$

$$= \frac{r \text{ Adj } (rx^{-1}I - T)}{x \det (rx^{-1}I - T)}$$

so that $$A = -r \text{ Adj } (rI - T)/c$$

where
$$c = \lim_{x \to 1-} \{-\det(rx^{-1}I - T)/(1-x)\}$$
$$= \frac{d}{dx}[\phi(rx^{-1})]_{x=1}$$
$$= -r\phi'(r)$$

which completes the proof, taking into account Corollary 2 of Theorem 1.1. □

In conclusion to this section we point out that assertion (d) of Theorem 1.1 states that the *geometric multiplicity* of the eigenvalue r is one, whereas (f) states that its *algebraic multiplicity* is one. It is well known in matrix theory that geometric multiplicity one for the eigenvalue of a square arbitrary matrix does not in general imply algebraic multiplicity one. A simple example to this end is the matrix (which is non-negative, but of course not primitive):

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

which has repeated eigenvalue unity (algebraic multiplicity two), but a corresponding left eigenvector can only be a multiple of $\{0, 1\}$ (geometric multiplicity one).

The distinction between geometric and algebraic multiplicity in connection with r in a primitive matrix is slurred over in some treatments of nonnegative matrix theory.

1.2 Structure of a General Non-negative Matrix

In this section we are concerned with a general square matrix $T = \{t_{ij}\}$, i, $j = 1, \ldots, n$, satisfying $t_{ij} \geq 0$, with the aim of showing that the behaviour of its powers T^k reduces, to a substantial extent, to the behaviour of powers of a fundamental type of non-negative square matrix, called *irreducible*. The class of irreducible matrices further subdivides into matrices which are *primitive* (studied in §1.1), and *cyclic* (imprimitive), whose study is taken up in §1.3.

We introduce here a definition, which while frequently used in other expositions of the theory, and so possibly useful to the reader, will be used by us only to a limited extent.

Definition 1.2. A sequence $(i, i_1, i_2, \ldots, i_{t-1}, j)$, for $t \geq 1$ (where $i_0 = i$), from the index set $\{1, 2, \ldots, n\}$ of a non-negative matrix T is said to form a *chain* of length t between the ordered pair (i, j) if

$$t_{ii_1} t_{i_1 i_2} \cdots t_{i_{t-2} i_{t-1}} t_{i_{t-1} j} > 0.$$

Such a chain for which $i = j$ is called a *cycle* of length t between i and itself.

Clearly in this definition, we may without loss of generality impose the restriction that, for fixed (i, j), $i, j \neq i_1 \neq i_2 \neq \cdots \neq i_{t-1}$, to obtain a 'minimal' length chain or cycle, from a given one.

Classification of indices

Let i, j, k be arbitrary indices from the index set $1, 2, \ldots, n$ of the matrix T.

We say that i *leads to* j, and write $i \to j$, if there exists an integer $m \geq 1$ such that $t_{ij}^{(m)} > 0$.[1] If i does not lead to j we write $i \not\to j$. Clearly, if $i \to j$ and $j \to k$ then, from the rule of matrix multiplication, $i \to k$.

We say that i and j *communicate* if $i \to j$ and $j \to i$, and write in this case $i \leftrightarrow j$.

The indices of the matrix T may then be classified and grouped as follows.

(*a*) If $i \to j$ but $j \not\to i$ for some j, then the index i is called *inessential*. An index which leads to no index at all (this arises when T has a row of zeros) is also called inessential.

(*b*) Otherwise the index i is called *essential*. Thus if i is essential, $i \to j$ implies $i \leftrightarrow j$; and there is at least one j such that $i \to j$.

(*c*) It is therefore clear that all essential indices (if any) can be subdivided into *essential classes* in such a way that all the indices belonging to one class communicate, but cannot lead to an index outside the class.

(*d*) Moreover, all inessential indices (if any) which communicate with some index, may be divided into *inessential classes* such that all indices in a class communicate.

Classes of the type described in (*c*) and (*d*) are called *self-communicating* classes.

(*e*) In addition there may be inessential indices which communicate with no index; these are defined as forming an *inessential class* by themselves (which, of course, if not self-communicating). These are of nuisance value only as regards applications, but are included in the description for completeness.

This description appears complex, but should be much clarified by the example which follows, and similar exercises.

Before proceeding, we need to note that the classification of indices (and hence grouping into classes) *depends only on the location of the positive elements, and not on their size*, so any two non-negative matrices with the same incidence matrix will have the same index classification and grouping (and, indeed, *canonical form*, to be discussed shortly).

Further, given a non-negative matrix (or its incidence matrix), classification and grouping of indices is made easy by a *path diagram* which may be described as follows. Start with index 1—this is the zeroth stage;

[1] Or, equivalently, if there is a chain between i and j.

determine all j such that $1 \to j$ and draw arrows to them—these j form the 2nd stage; for each of these j now repeat the procedure to form the 3rd stage; and so on; *but* as soon as an index occurs which has occurred at an earlier stage, ignore further consequents of it. Thus the diagram terminates when every index in it has repeated. (Since there are a finite total number of indices, the process must terminate.) This diagram will represent all possible consequent behaviour for the set of indices which entered into it, which may not, however, be the entire index set. If any are left over, choose one such and draw a similar diagram for it, regarding the indices of the previous diagram also as having occurred 'at an earlier stage'. And so on, till all indices of the index set are accounted for.

EXAMPLE. A non-negative matrix T has incidence matrix

	1	2	3	4	5	6	7	8	9
1	1	1	0	0	0	0	0	0	0
2	1	1	1	0	0	0	1	0	0
3	0	0	0	0	0	0	1	0	0
4	0	0	0	1	0	0	0	0	1
5	0	0	0	0	1	0	0	0	0
6	0	0	1	0	0	1	0	0	0
7	0	0	1	0	0	0	0	0	0
8	0	1	0	0	0	1	0	1	0
9	0	0	0	1	0	0	0	0	1

Thus Diagram 1 tells us {3, 7} is an essential class, while {1, 2} is an inessential (communicating) class.
Diagram 2 tells us {4, 9} is an essential class.

Diagram 1

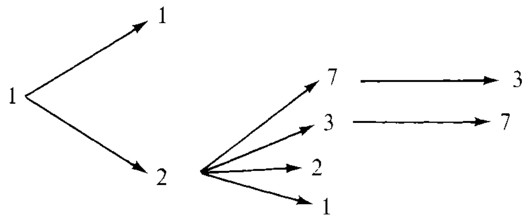


Diagram 2

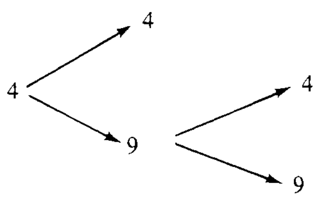


Diagram 3 tells us {5} is an essential class.
Diagram 4 tells us {6} is an inessential (self-communicating) class.
Diagram 5 tells us {8} is an inessential (self-communicating) class.

Diagram 3

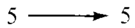


Diagram 4

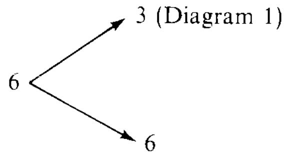
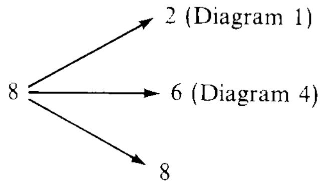


Diagram 5



Canonical Form

Once the classification and grouping has been carried out, the definition 'leads' may be extended to classes in the obvious sense e.g. the statement $\mathscr{C}_1 \rightarrow \mathscr{C}_2 (\mathscr{C}_1 \neq \mathscr{C}_2)$ means that there is an index of $\mathscr{C}_1$ which leads to an index of $\mathscr{C}_2$. Hence all indices of $\mathscr{C}_1$ lead to all indices of $\mathscr{C}_2$, and the statement can only apply to an inessential class $\mathscr{C}_1$.

Moreover, the matrix T may be put into *canonical form* by first relabelling the indices in a specific manner. Before describing the manner, we mention that relabelling the indices using the same index set $\{1, \ldots, n\}$ and rewriting T accordingly merely amounts to performing a *simultaneous permutation* of rows and columns of the matrix. Now such a simultaneous permutation only amounts to a *similarity transformation* of the original matrix, T, so that (*a*) its powers are similarly transformed; (*b*) its spectrum (i.e. set of eigenvalues) is unchanged. Generally any given ordering is as good as any other in a physical context; but the canonical form of T, arrived at by one such ordering, is particularly convenient.

The canonical form is attained by first taking the indices of one essential class (if any) and renumbering them consecutively using the lowest integers, and following by the indices of another essential class, if any, until the essential classes are exhausted. The numbering is then extended to the indices of the inessential classes (if any) which are themselves arranged in an order such that an inessential class occurring earlier (and thus higher in the arrangement) does not *lead* to any inessential class occurring later.

EXAMPLE (continued). For the given matrix T the essential classes are $\{5\}$, $\{4, 9\}$, $\{3, 7\}$; and the inessential classes $\{1, 2\}$, $\{6\}$, $\{8\}$ which from Diagrams 4 and 5 should be ranked in this order. Thus a possible canonical form for T is

$$\begin{array}{c|ccccccccc|}
 & 5 & 4 & 9 & 3 & 7 & 1 & 2 & 6 & 8 \\
\hline
5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
9 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
7 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
2 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
6 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{array}.$$

It is clear that the canonical form consists of square diagonal blocks corresponding to 'transition within' the classes in one 'stage', zeros to the right of these diagonal blocks, but at least one non-zero element to the left of each inessential block unless it corresponds to an index which leads to no other. Thus the general version of the canonical form of T is

$$T = \left[\begin{array}{ccccc|c}
T_1 & 0 & \dots & 0 & \dots & 0 \\
0 & T_2 & & & & \vdots \\
 & 0 & & & & \\
\vdots & & & & & \\
0 & 0 & \dots & & T_z & 0 \\
\hline
 & R & & & & Q
\end{array}\right]$$

where the T_i correspond to the z essential classes, and Q to the inessential indices, with $R \neq 0$ in general, with Q itself having a structure analogous to T, except that there may be non-zero elements to the left of any of its diagonal blocks:

$$Q = \begin{bmatrix} Q_1 & & & \\ & Q_2 & & \\ & & \ddots & 0 \\ S & & & Q_w \end{bmatrix}.$$

Now, in most applications we are interested in the behaviour of the *powers* of T. Let us assume it is in canonical form. Since

$$T^k = \left[\begin{array}{cccc|c} T_1^k & & & & \\ & T_2^k & & & \\ 0 & & \ddots & 0 & 0 \\ & & & T_z^k & \\ \hline & & R_k & & Q^k \end{array}\right], \quad Q^k = \begin{bmatrix} Q_1^k & & & \\ & Q_2^k & & \\ & & \ddots & 0 \\ S_k & & & Q_w^k \end{bmatrix}$$

it follows that a substantial advance in this direction will be made in studying the powers of the diagonal *block submatrices corresponding to self—communicating classes* (the other diagonal block submatrices, if any, are 1×1 zero matrices; the evolution of R_k and S_k is complex, with k). In fact if one is interested in only the essential indices, as is often the case, this is sufficient.

A (sub)matrix corresponding to a single self-communicating class is called *irreducible.*

It remains to show that, *normally*, there is at least one self-communicating (indeed essential) class of indices present for any matrix T; although it is nevertheless *possible* that all indices of a non-negative matrix fall into non self-communicating classes (and are therefore inessential): for example

$$T = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Lemma 1.1. *An $n \times n$ non-negative matrix with at least one positive entry in each row possesses at least one essential class of indices.*

PROOF. Suppose all indices are inessential. The assumption of non-zero rows then implies that for any index i, $i = 1, \ldots, n$, there is at least one j such that $i \to j$, but $j \not\to i$.

Now suppose i_1 is any index. Then we can find a sequence of indices i_2, $i_3, \ldots$ etc. such that

$$i_1 \to i_2 \to i_3 \to \cdots \to i_n \to i_{n+1} \cdots$$

but such that $i_{k+1} \not\to i_k$, and hence $i_{k+1} \not\to i_1, i_2, \ldots,$ or i_{k-1}. However, since the sequence $i_1, i_2, \ldots, i_{n+1}$ is a set of $n + 1$ indices, each chosen from the same n possibilities, $1, 2, \ldots, n$, at least *one index repeats* in the sequence. This is a contradiction to the deduction that no index can lead to an index with a lower subscript. □

We come now to the important concept of the period of an index.

Definition 1.3. If $i \to i$, $d(i)$ is the *period* of the index i if it is the greatest common divisor of those k for which

$$t_{ii}^{(k)} > 0$$

(see Definition A.2 in Appendix A). N.B. If $t_{ii} > 0$, $d(i) = 1$.

We shall now prove that in a communicating class all indices have the same period.

Lemma 1.2. *If* $i \leftrightarrow j$, $d(i) = d(j)$.

PROOF. Let $t_{ij}^{(M)} > 0$, $t_{ji}^{(N)} > 0$. Then for any positive integer s such that $t_{jj}^{(s)} > 0$

$$t_{ii}^{(M+s+N)} \geq t_{ij}^{(M)} t_{jj}^{(s)} t_{ji}^{(N)} > 0,$$

the first inequality following from the rule of matrix multiplication and the non-negativity of the elements of T. Now, for such an s it is also true that $t_{jj}^{(2s)} > 0$ necessarily, so that

$$t_{ii}^{(M+2s+N)} > 0.$$

Therefore $\qquad d(i)$ divides $M + 2s + N - (M + s + N) = s$.

Hence: for every s such that $t_{jj}^{(s)} > 0$, $d(i)$ divides s.

Hence $$d(i) \leq d(j).$$

But since the argument can be repeated with i and j interchanged,

$$d(j) \leq d(i).$$

Hence $d(i) = d(j)$ as required. □

Note that, again, consideration of an incidence matrix is adequate to determine the period.

Definition 1.4. The period of a communicating class is the period of any one of its indices.

EXAMPLE (continued): Determine the periods of all communicating classes for the matrix T with incidence matrix considered earlier.

Essential classes:

$\{5\}$ has period 1, since $t_{55} > 0$.

$\{4, 9\}$ has period 1, since $t_{44} > 0$.

$\{3, 7\}$ has period 2, since $t_{33}^{(k)} > 0$

for every even k, and is zero for every odd k.

Inessential self-communicating classes:

$\{1, 2\}$ has period 1 since $t_{11} > 0$.

$\{6\}$ has period 1 since $t_{66} > 0$.

$\{8\}$ has period 1 since $t_{88} > 0$.

Definition 1.5. An index i such that $i \to i$ is *aperiodic* (acyclic) if $d(i) = 1$. [It is thus contained in an *aperiodic* (*self-communicating*) *class.*]

1.3 Irreducible Matrices

Towards the end of the last section we called a non-negative square matrix, corresponding to a single self-communicating class of indices, irreducible. We now give a general definition, independent of the previous context, which is, nevertheless, easily seen to be equivalent to the one just given. The part of the definition referring to periodicity is justified by Lemma 1.2.

Definition 1.6. An $n \times n$ non-negative matrix T is *irreducible* if for every pair i, j of its index set, there exists a positive integer $m \equiv m(i, j)$ such that $t_{ij}^{(m)} > 0$. An irreducible matrix is said to be cyclic (periodic) with period d, if the period of any one (and so of each one) of its indices satisfies $d > 1$, and is said to be acyclic (aperiodic) if $d = 1$.

The following results all refer to an irreducible matrix with period d. Note that an irreducible matrix T cannot have a zero row or column.

Lemma 1.3. *If $i \to i$, $t_{ii}^{(kd)} > 0$ for all integers $k \geq N_0(=N_0(i))$.*

PROOF.

Suppose $$t_{ii}^{(kd)} > 0,\ t_{ii}^{(sd)} > 0.$$

Then $$t_{ii}^{([k+s]d)} \geq t_{ii}^{(kd)} t_{ii}^{(sd)} > 0.$$

Hence the positive integers $\{kd\}$ such that

$$t_{ii}^{(kd)} > 0,$$

form a closed set under addition, and their greatest common divisor is d. An appeal to Lemma A.3 of Appendix A completes the proof. □

Theorem 1.3. *Let i be any fixed index of the index set $\{1, 2, \ldots, n\}$ of T. Then, for every index j there exists a unique integer r_j in the range $0 \leq r_j < d$ (r_j is called a residue class modulo d) such that*

(*a*) *$t_{ij}^{(s)} > 0$ implies $s \equiv r_j \pmod d$;*[1] *and*
(*b*) *$t_{ij}^{(kd+r_j)} > 0$ for $k \geq N(j)$, where $N(j)$ is some positive integer.*

PROOF. Let $t_{ij}^{(m)} > 0$ and $t_{ij}^{(m')} > 0$.
There exists a p such that $t_{ji}^{(p)} > 0$, whence as before

$$t_{ii}^{(m+p)} > 0 \text{ and } t_{ii}^{(m'+p)} > 0.$$

Hence d divides each of the superscripts, and hence their difference $m - m'$. Thus $m - m' \equiv 0 \pmod d$, so that

$$m \equiv r_j \pmod d.$$

[1] Recall from Appendix A, that this means that if qd is the multiple of d nearest to s from below, then $s = qd + r_j$; it reads 's is equivalent to r_j, modulo d'.

This proves (a).

To prove (b), since $i \to j$ and in view of (a), there exists a positive m such that

$$t_{ij}^{(md+r_j)} > 0.$$

Now, let $N(j) = N_0 + m$, where N_0 is the number guaranteed by Lemma 1.3 for which $t_{ii}^{(sd)} > 0$ for $s \geq N_0$. Hence if $k \geq N(j)$, then

$$kd + r_j = sd + md + r_j, \text{ where } s \geq N_0.$$

Therefore $t_{ij}^{(kd+r_j)} \geq t_{ii}^{(sd)} t_{ij}^{(md+r_j)} > 0$, for all $k \geq N(j)$. □

Definition 1.7. The set of indices j in $\{1, 2, \ldots, n\}$ corresponding to the same residue class (mod d) is called a subclass of the class $\{1, 2, \ldots, n\}$, and is denoted by C_r, $0 \leq r < d$.

It is clear that the d subclasses C_r are disjoint, and their union is $\{1, 2, \ldots, n\}$. It is not yet clear that the composition of the classes does not depend on the choice of initial fixed index i, which we prove in a moment; nor that each subclass contains at least one index.

Lemma 1.4. *The composition of the residue classes does not depend on the initial choice of fixed index i; an initial choice of another index merely subjects the subclasses to a cyclic permutation.*

Proof. Suppose we take a new fixed index i'. Then

$$t_{ij}^{(md+r_{i'}+kd+r'_j)} \geq t_{ii'}^{(kd+r_{i'})} t_{i'j}^{(md+r'_j)}$$

where r'_j denotes the residue class corresponding to j according to classification with respect to fixed index i'. Now, by Theorem 1.3 for large k, m, the right hand side is positive, so that the left hand side is also, whence, in the old classification,

$$md + r_{i'} + kd + r'_j \equiv r_j \pmod d$$

i.e.

$$r_{i'} + r'_j \equiv r_j \pmod d.$$

Hence the composition of the subclasses $\{C_i\}$ is unchanged, and their order is merely subjected to a cyclic permutation σ:

$$\begin{pmatrix} 0 & 1 & \cdots & d-1 \\ \sigma(0) & \sigma(1) & \cdots & \sigma(d-1) \end{pmatrix}.$$ □

For example, suppose we have a situation with $d = 3$, and $r_{i'} = 2$. Then the classes which were C_0, C_1, C_2 in the old classification according to i (according to which $i' \in C_2$) now become, respectively, C'_1, C'_2, C'_0 since we must have $2 + r'_j \equiv r_j$ (mod d) for $r_j = 0, 1, 2$.

Let us now define C_r for all non-negative integers r by putting $C_r = C_{r_j}$ if $r \equiv r_j \pmod d$, using the initial classification with respect to i. Let m be a positive integer, and consider any j for which $t_{ij}^{(m)} > 0$. (There is at least one appropriate index j, otherwise T^m (and hence higher powers) would have ith row consisting entirely of zeros, contrary to irreducibility of T.) Then $m \equiv r_j$ (mod d), i.e. $m = sd + r_j$ and $j \in C_{r_j}$, Now, similarly, let k be any index such that

$$t_{ik}^{(m+1)} > 0.$$

Then, since $m + 1 = sd + r_j + 1$, it follows $k \in C_{r_j+1}$.

Hence it follows that, looking at the ith row, the positive entries occur, for successive powers, in successive subclasses. In particular each of the d cyclic classes is non-empty. If subclassification has initially been made according to the index i', since we have seen the subclasses are merely subjected to a cyclic permutation, the classes still 'follow each other' in order, looking at successive powers, and ith (hence any) row.

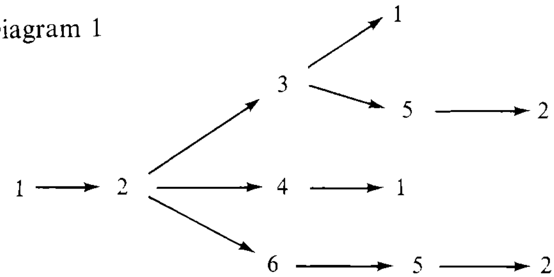
It follows that if $d > 1$ (so there is more than one subclass) a canonical form of T is possible, by relabelling the indices so that the indices of C_0 come first, of C_1 next, and so on. This produces a version of T of the sort

$$T_c = \begin{bmatrix} 0 & Q_{01} & 0 & \cdots & & 0 \\ 0 & 0 & Q_{12} & \cdots & & 0 \\ \cdot & \cdot & 0 & & & \\ \cdot & \cdot & \cdot & & & \\ 0 & \cdot & \cdot & & 0 & Q_{d-2,\,d-1} \\ Q_{d-1,\,0} & 0 & 0 & \cdots & & 0 \end{bmatrix}$$

EXAMPLE: Check that the matrix, whose incidence matrix is given below is irreducible, find its period, and put into a canonical form if periodic.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	0	0	1	1	0	1
3	1	0	0	0	1	0
4	1	0	0	0	0	0
5	0	1	0	0	0	0
6	0	0	0	0	1	0

Diagram 1



Clearly $i \to j$ for any i and j in the index set, so the matrix is certainly irreducible. Let us now carry out the determination of subclasses on the basis of index 1. Therefore index 1 must be in the subset C_0; 2 must be in C_1; 3, 4, 6 in C_2; 1, 5 in C_3; 2 in C_4. Hence C_0 and C_3 are identical; C_1 and C_4; etc., and so $d = 3$. Moreover

$$C_0 = \{1, 5\},\ C_1 = \{2\},\ C_2 = \{3, 4, 6\},$$

so canonical form is

$$\begin{array}{c} \\ 1 \\ 5 \\ 2 \\ 3 \\ 4 \\ 6 \end{array} \begin{array}{cccccc} 1 & 5 & 2 & 3 & 4 & 6 \end{array} \left[\begin{array}{cc|c|ccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array}\right].$$

Theorem 1.4. *An irreducible acyclic matrix T is primitive and conversely. The powers of an irreducible cyclic matrix may be studied in terms of powers of primitive matrices.*

Proof. If T is irreducible, with $d = 1$, there is only one subclass of the index set, consisting of the index set itself, and Theorem 1.3 implies

$$t_{ij}^{(k)} > 0 \quad \text{for} \quad k \geq N(i, j).$$

Hence for $N^* = \max_{i,j} N(i, j)$

$$t_{ij}^{(k)} > 0,\ k \geq N^*, \quad \text{for all } i, j.$$

i.e.
$$T^k > 0 \quad \text{for} \quad k \geq N^*.$$

Conversely, a primitive matrix is trivially irreducible, and has $d = 1$, since for any fixed i, and k great enough $t_{ii}^{(k)} > 0$, $t_{ii}^{(k+1)} > 0$, and the greatest common divisor of k and $k + 1$ is 1.

The truth of the second part of the assertion may be conveniently demonstrated in the case $d = 3$, where the canonical form of T is

$$T_c = \begin{bmatrix} 0 & Q_{01} & 0 \\ 0 & 0 & Q_{12} \\ Q_{20} & 0 & 0 \end{bmatrix},$$

$$\text{and } T_c^2 = \begin{bmatrix} 0 & 0 & Q_{01}Q_{12} \\ Q_{12}Q_{20} & 0 & 0 \\ 0 & Q_{20}Q_{01} & 0 \end{bmatrix},$$

$$T_c^3 = \begin{bmatrix} Q_{01}Q_{12}Q_{20} & 0 & 0 \\ 0 & Q_{12}Q_{20}Q_{01} & 0 \\ 0 & 0 & Q_{20}Q_{01}Q_{12} \end{bmatrix}.$$

Now, the diagonal matrices of T_c^3 (of T_c^d in general) are square and *primitive*, for Lemma 1.3 states that $t_{ii}^{(3k)} > 0$ for all k sufficiently large. Hence

$$T_c^{3k} = (T_c^3)^k,$$

so that powers which are integral multiples of the period may be studied with the aid of the primitive matrix theory of §1.1. One needs to consider also

$$T_c^{3k+1} \quad \text{and} \quad T_c^{3k+2}$$

but these present no additional difficulty since we may write $T_c^{3k+1} = (T_c^{3k})T_c$, $T_c^{3k+2} = (T_c^{3k})T_c^2$ and proceed as before. ■

These remarks substantiate the reason for considering primitive matrices as of prime importance, and for treating them first. It is, nevertheless, convenient to consider a theorem of the type of the fundamental Theorem 1.1 for the broader class of irreducible matrices, which we now expect to be closely related.

1.4 Perron–Frobenius Theory for Irreducible Matrices

Theorem 1.5. *Suppose T is an $n \times n$ irreducible non-negative matrix. Then all of the assertions* (a)–(f) *of Theorem* 1.1 *hold, except that* (c) *is replaced by the weaker statement*: $r \geq |\lambda|$ *for any eigenvalue λ of T. Corollaries* 1 *and* 2 *of Theorem* 1.1 *hold also.*

PROOF. The proof of (a) of Theorem 1.1 holds to the stage where we need to assume

$$z' = \hat{x}'T - r\hat{x}' \geq \mathbf{0}' \quad \text{but} \quad \neq \mathbf{0}'.$$

The matrix $I + T$ is primitive, hence for some k, $(I + T)^k > 0$; hence

$$z'(I + T)^k = \{\hat{x}'(I + T)^k\}T - r\{\hat{x}'(I + T)^k\} > \mathbf{0}$$

which contradicts the definition of r; (b) is then proved as in Theorem 1.6 following; and the rest follows as before, except for the last part in (c). □

We shall henceforth call r the Perron–Frobenius eigenvalue of an irreducible T, and its corresponding positive eigenvectors, the Perron–Frobenius eigenvectors.

The above theorem does not answer in detail questions about eigenvalues λ such that $\lambda \neq r$ but $|\lambda| = r$ in the cyclic case.

The following auxiliary result is more general than we shall require immediately, but is important in future contexts.

Theorem 1.6. (*The Subinvariance Theorem*). *Let T be a non-negative irreducible matrix, s a positive number, and $y \geq \mathbf{0}$, $\neq \mathbf{0}$, a vector satisfying*

$$Ty \leq sy.$$

Then (a) $y > \mathbf{0}$; (b) $s \geq r$, where r is the Perron–Frobenius eigenvalue of T. Moreover, $s = r$ if and only if $Ty = sy$.

Proof. Suppose at least one element, say the ith, of y is zero. Then since $T^k y \leq s^k y$ it follows that

$$\sum_{j=1}^{n} t_{ij}^{(k)} y_j \leq s^k y_i.$$

Now, since T is irreducible, for this i and any j, there exists a k such that $t_{ij}^{(k)} > 0$; and since $y_j > 0$ for some j, it follows that

$$y_i > 0$$

which is a contradiction. Thus $y > \mathbf{0}$. Now, premultiplying the relation $Ty \leq sy$ by $\hat{x}'$, a positive left eigenvector of T corresponding to r,

$$s\hat{x}'y \geq \hat{x}'Ty = r\hat{x}'y$$

i.e.

$$s \geq r.$$

Now suppose $Ty \leq ry$ with strict inequality in at least one place; then the preceding argument, on account of the strict positivity of Ty and ry, yields $r < r$, which is impossible. The implication $s = r$ follows from $Ty = sy$ similarly. ■

In the sequel, any subscripts which occur should be understood as reduced modulo d, to bring them into the range $[0, d-1]$, if they do not already fall in the range.

Theorem 1.7. *For a cyclic matrix T with period $d > 1$, there are present precisely d distinct eigenvalues λ with $|\lambda| = r$, where r is the Perron–Frobenius eigenvalue of T. These eigenvalues are: $r \exp i2\pi k/d$, $k = 0, 1, \ldots, d-1$ (i.e. the d roots of the equation $\lambda^d - r^d = 0$).*

Proof. Consider an arbitrary one, say the ith, of the primitive matrices:

$$Q_{i,i+1} Q_{i+1,i+2} \cdots Q_{i+d-1,i+d}$$

occurring as diagonal blocks in the dth power, T^d, of the canonical form T_c of T (recall that T_c has the same eigenvalues as T), and denote by $r(i)$ its Perron–Frobenius eigenvalue, and by $y(i)$ a corresponding positive right eigenvector, so that

$$Q_{i,i+1} Q_{i+1,i+2} \cdots Q_{i+d-1,i+d}\, y(i) = r(i) y(i).$$

Now premultiply this by $Q_{i-1,i}$:

$$Q_{i-1,i}Q_{i,i+1}Q_{i+1,i+2}\cdots Q_{i+d-2,i+d-1}Q_{i+d-1,i+d}y(i) = r(i)Q_{i-1,i}y(i),$$

and since $Q_{i+d-1,i+d} \equiv Q_{i-1,i}$, we have

$$Q_{i-1,i}Q_{i,i+1}Q_{i+1,i+2}\cdots Q_{i+d-2,i+d-1}(Q_{i-1,i}y(i)) = r(i)(Q_{i-1,i}y(i))$$

whence it follows from Theorem 1.6 that $r(i) \geq r(i-1)$. Thus

$$r(0) \geq r(d-1) \geq r(d-2) \cdots \geq r(0),$$

so that, for all i, $r(i)$ is constant, say $\tilde{r}$, and so there are precisely d dominant eigenvalues of T^d, all the other eigenvalues being strictly smaller in modulus. Hence, since the eigenvalues of T^d are dth powers of the eigenvalues of T, there must be precisely d dominant roots of T, and all must be dth roots of $\tilde{r}$. Now, from Theorem 1.5, the positive dth root is an eigenvalue of T and is r. Thus every root λ of T such that $|\lambda| = r$ must be of the form

$$\lambda = r \exp \mathrm{i}(2\pi k/d),$$

where k is one of $0, 1, \ldots, d-1$, and there are d of them.

It remains to prove that there are no coincident eigenvalues, so that in fact all possibilities $r \exp \mathrm{i}(2\pi k/d)$, $k = 0, 1, \ldots, d-1$ occur.

Suppose that y is a positive $(n \times 1)$ right eigenvector corresponding to the Perron–Frobenius eigenvalue r of T_c (i.e. T written out in canonical form), and let y_j, $j = 0, \ldots, d-1$ be the subvector of components corresponding to subclass C_j.

Thus $$y' = [y'_0, y'_1, \ldots, y'_{d-1}]$$

and $$Q_{j,j+1}y_{j+1} = ry_j.$$

Now, let $\bar{y}_k$, $k = 0, 1, \ldots, d-1$ be the $(n \times 1)$ vector obtained from y by making the transformation

$$y_j \to \exp \mathrm{i}\left(\frac{2\pi jk}{d}\right)y_j$$

of its components as defined above. It is easy to check that $\bar{y}_0 = y$, and indeed that $\bar{y}_k$, $k = 0, 1, \ldots, d-1$ is an eigenvector corresponding to an eigenvector $r \exp \mathrm{i}(2\pi k/d)$, as required. This completes the proof of the theorem. □

We note in conclusion the following corollary on the structure of the eigenvalues, whose validity is now clear from the immediately preceding.

Corollary. *If $\lambda \neq 0$ is any eigenvalue of T, then the numbers $\lambda \exp \mathrm{i}(2\pi k/d)$, $k = 0, 1, \ldots, d-1$ are eigenvalues also. (Thus, rotation of the complex plane about the origin through angles of $2\pi/d$ carries the set of eigenvalues into itself.)*

Bibliography and Discussion

§1.1. and §1.4

The exposition of the chapter centres on the notion of a primitive non-negative matrix as the fundamental notion of non-negative matrix theory. The approach seems to have the advantage of proving the fundamental theorem of nonnegative matrix theory at the outset, and of avoiding the slight awkwardness entailed in the usual definition of irreducibility merely from the permutable structure of T.

The fundamental results (Theorems 1.1, 1.5 and 1.7) are basically due to Perron (1907) and Frobenius (1908, 1909, 1912), Perron's contribution being associated with strictly positive T. Many modern expositions tend to follow the simple and elegant paper of Wielandt (1950) (whose approach was anticipated in part by Lappo-Danilevskii (1934)); see e.g. Cherubino (1957), Gantmacher (1959) and Varga (1962). This is essentially true also of our proof of Theorem 1.1 (= Theorem 1.5) with some slight simplifications, especially in the proof of part (e), under the influence of the well-known paper of Debreu & Herstein (1953), which deviates otherwise from Wielandt's treatment also in the proof of (a). (The proof of Corollary 1 of Theorem 1.1 also follows Debreu & Herstein.)

The proof of Theorem 1.7 is not, however, associated with Wielandt's approach, due to an attempt to bring out, again, the primacy of the primitivity property. The last part of the proof (that all dth roots of r are involved), as well as the corollary, follows Romanovsky (1936). The possibility of evolving §1.4 in the present manner depends heavily on §1.3.

For other approaches to the Perron–Frobenius theory see Bellman (1960, Chapter 16), Brauer (1957b), Fan (1958), Householder (1958), Karlin (1959, §8.2; 1966, Appendix), Pullman (1971), Samelson (1957) and Sevastyanov (1971, Chapter 2). Some of these references do not deal with the most general case of an irreducible matrix, containing restrictions of one sort or another. In their recent treatise on non-negative matrices, Berman and Plemmons (1979) begin with a chapter studying the spectral properties of the set of $n \times n$ matrices which leave a proper cone in R^n invariant, combining the use of the Jordan normal form of a matrix, matrix norms and some assumed knowledge of cones. These results are specialized to non-negative matrices in their Chapter 2; and together with additional direct proofs give the full structure of the Perron–Frobenius theory. We have sought to present this theory in a simpler fashion, at a lower level of mathematical conception and technique.

Finally we mention that Schneider's (1977) survey gives, interalia, a historical survey of the concept of irreducibility.

§1.2 and §1.3

The development of these sections is motivated by probabilistic considerations from the theory of Markov chains, where it occurs in connection with *stochastic* non-negative matrices $P = \{p_{ij}\}$, $i, j = 1, 2, \ldots$, with $p_{ij} \geq 0$ and

$$\sum_j p_{ij} = 1, \qquad i = 1, 2, \ldots.$$

In this setting the classification theory is essentially due to Kolmogorov (1936); an account may be found in the somewhat more general exposition of Chung (1967, Chapter 1, §3), which our exposition tends to follow.

A weak analogue of the Perron–Frobenius Theorem for any square $T \geq 0$ is given as Exercise 1.12. Another approach to Perron–Frobenius-type theory in this general case is given by Rothblum (1975), and taken up in Berman and Plemmons (1979, §2.3).

Just as in the case of stochastic matrices, the corresponding exposition *is not restricted to finite matrices* (this in fact being the reason for the development of this kind of classification in the probabilistic setting), and virtually all of the present exposition goes through for infinite non-negative matrices T, so long as all powers T^k, $k = 1, 2, \ldots$ exist (with an obvious extension of the rule of matrix multiplication of finite matrices). This point is taken up again to a limited extent in Chapters 5 and 6, where infinite T are studied.

The reader acquainted with graph theory will recognize its relationship with the notion of path diagrams used in our exposition. For development along the lines of graph theory see Rosenblatt (1957), the brief account in Varga (1962, Chapters 1 and 2), Paz (1963) and Gordon (1965, Chapter 1). The relevant notions and usage in the setting of non-negative matrices implicitly go back at least to Romanovsky (1936).

Another development, not explicitly graph theoretical, is given in the papers of Pták (1958) and Pták & Sedlaček (1958); and it is utilized to some extent in §2.4 of the next chapter.

Finite stochastic matrices and finite Markov chains will be treated in Chapter 4. The general infinite case will be taken up in Chapter 5.

Exercises

1.1 Find all essential and inessential classes of a non-negative matrix with incidence matrix:

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Find the periods of all self-communicating classes, and write the matrix T in full canonical form, so that the matrices corresponding to all self-communicating classes are also in canonical form.

1.2. Keeping in mind Lemma 1.1, construct a non-negative matrix T whose index set contains no essential class, but has, nevertheless, a self-communicating class.

1.3. Let $T = \{t_{i,j}\}, i,j = 1,2,\ldots,n$ be a non-negative matrix. If, for some fixed i and j, $t_{i,j}^{(k)} > 0$ for some $k \equiv k(i,j)$, show that there exists a sequence $k_1, k_2, \ldots, k_r$ such that

$$t_{i,k_1} t_{k_1,k_2} \cdots t_{k_{r-1},k_r} t_{k_r,j} > 0$$

where $r \leq n-2$ if $i \neq j$, $r \leq n-1$ if $i = j$. Hence show that:
(a) if T is irreducible and $t_{j,j} > 0$ for some j, then $t_{i,j}^{(k)} > 0$ for $k \geq n-1$ and every i; and, hence, if $t_{j,j} > 0$ for every j, then $T^{n-1} > 0$;
(b) T is irreducible if and only if $(I+T)^{n-1} > 0$.

(Wielandt, 1960; Herstein, 1954.)

Further results along the lines of (a) are given as Exercises 2.17–2.19, and again in Lemma 3.9.

1.4. Given $T = \{t_{i,j}\}, i,j = 1,2,\ldots,n$ is a non-negative matrix, suppose that for some power $m \geq 1$, $T^m = \{t_{i,j}^{(m)}\}$ is such that

$$t_{i,i+1}^{(m)} > 0, \; i = 1, 2, \ldots, n-1, \quad \text{and} \quad t_{n,1}^{(m)} > 0.$$

Show that: T is irreducible; and (by example) that it may be periodic.

1.5. By considering the vector $x' = (\alpha, \alpha, 1-2\alpha)$, suitably normed, when: (i)$\alpha = 0$, (ii) $0 < \alpha < \frac{1}{2}$, and the matrix

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 1 & 0 & 2 \end{bmatrix},$$

show that $r(x)$, as defined in the proof of Theorem 1.1 is not continuous in $x \geq \mathbf{0}$, $x'x = 1$.

(Schneider, 1958)

1.6.[1] If r is the Perron–Frobenius eigenvalue of an irreducible matrix $T = \{t_{ij}\}$, show that for any vector $x \in \mathscr{P}$, where $\mathscr{P} = \{x;\, x > \mathbf{0}\}$

$$\min_i \frac{\sum_j t_{ij} x_j}{x_i} \leq r \leq \max_i \frac{\sum_j t_{ij} x_j}{x_i}.$$

(Collatz, 1942)

1.7. Show, in the situation of Exercise 1.6, that equality on either side implies equality on both; and by considering when this can happen show that r is the supremum of the left hand side, and the infimum of the right hand side, over $x \in \mathscr{P}$, and is actually attained as both supremum and infimum for vectors in $\mathscr{P}$.

[1] Exercises 1.6 to 1.8 have a common theme.

1.8. In the framework of Exercise 1.6, show that

$$\max_{x \in \mathscr{P}} \left\{ \min_{y \in \mathscr{P}} \frac{y'Tx}{y'x} \right\} = r = \min_{y \in \mathscr{P}} \left\{ \max_{x \in \mathscr{P}} \frac{y'Tx}{y'x} \right\}.$$

(Birkhoff and Varga, 1958)

1.9.[1] Let B be a matrix with possibly complex elements and denote by B_+ the matrix of moduli of elements of B and β an eigenvalue of B. Let T be irreducible and such that $0 \le B_+ \le T$. Show that $|\beta| \le r$; and moreover that $|\beta| = r$ implies $B_+ = T$, where r is the Perron–Frobenius eigenvalues of T.

(Frobenius, 1909)

1.10. If, in Exercise 1.9, $|\beta| = r$, so that $\beta = re^{i\theta}$, say, it can be shown (Wielandt, 1950) that B has the representation

$$B = e^{i\theta} DTD^{-1}$$

where D is a diagonal matrix whose diagonal elements have modulus one. Show as consequences:

(i) that if $|\beta| = r$, $B_+ = T$;

(ii) that given there are d dominant eigenvalues of modulus r for a given periodic irreducible matrix of period d, they must in fact all be simple, and take on the values $r \exp i(2\pi j/d)$, $j = 0, 1, \ldots, d-1$. (Put $B = T$ in the representation.)

1.11. Let T be an irreducible non-negative matrix and E a non-zero non-negative matrix of the same size. If x is a positive number, show that $A = xE + T$ is irreducible, and that its Perron–Frobenius eigenvalue may be made to equal any positive number exceeding the Perron–Frobenius eigenvalue r of T by suitable choice of x.
(Consider first, for orientation, the situation where at least one diagonal element of E is positive. Make eventual use of the continuity of the eigenvalues of A with x.)

(Birkhoff & Varga, 1958)

1.12. If $T \ge 0$ is any square non-negative matrix, use the canonical form of T to show that the following weak analogue of the Perron–Frobenius Theorem holds: there exists an eigenvalue ρ such that

(a') ρ real, ≥ 0;

(b') with ρ can be associated non-negative left and right eigenvectors;

(c') $\rho \ge |\lambda|$ for any eigenvalue λ of T;

(e') if $0 \le B \le T$ and β is an eigenvalue of B, then $|\beta| \le \rho$.

(In such problems it is often useful to consider a sequence of matrices each $\ge T$ and converging to T elementwise {particularly in relation to (b') here}—Debreu and Herstein (1953).)

1.13. Show in relation to Exercise 1.12, that $\rho > 0$ if and only if T contains a *cycle* of elements.

(Ullman, 1952)

[1] Exercises 1.9 to 1.11 have a common theme.

1.14. Use the relevant part of Theorem 1.4, in conjunction with Theorem 1.2, to show that for an irreducible T with Perron–Frobenius eigenvalue r, as $k \to \infty$

$$s^{-k}T^k \to 0$$

if and only if $s > r$; and if $0 < s < r$, for each pair (i, j)

$$\limsup_{k \to \infty} s^{-k} t_{ij}^{(k)} = \infty.$$

Hence deduce that the power series

$$T_{ij}(z) = \sum_{k=0}^{\infty} t_{ij}^{(k)} z^k$$

have common convergence radius $R = r^{-1}$ for each pair (i, j). (This result is relevant to the development of the theory of countable irreducible T in Chapter 6.)

1.15. Let T be a non-negative matrix. Show that:
(a) $Ty \leq sy$, where $s \neq 0$; $y \geq \mathbf{0}, \neq \mathbf{0} \Rightarrow y > \mathbf{0}$ if and only if T is irreducible;
(b) T has a single non-negative (left or right) eigenvector (to constant multiples) and this eigenvector is positive if and only if T is irreducible.

1.16. If A and B are non-negative matrices such that $0 \leq B \leq A$, $A - B \neq 0$, and $A + B$ is irreducible, show that $\rho(B) < \rho(A)$ where $\rho(\cdot)$ is the eigenvalue alluded to in Exercise 1.12.

1.17. Let T be a non-negative irreducible matrix, s a positive number, and $y \geq \mathbf{0}, \neq \mathbf{0}$ a vector satisfying

$$Ty \geq sy.$$

Show that $r \geq s$, where r is the Perron-Frobenius eigenvector of T, and $s = r$ if and only if $Ty = sy$. [This is a dual result to the (Subinvariance) Theorem 1.6.]

1.18. Suppose T is a non-negative matrix which, by simultaneous permutation of rows and columns may be put in the form

$$\begin{bmatrix} 0 & T_1 & 0 & \cdots & 0 \\ 0 & 0 & T_2 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & T_{d-1} \\ T_d & 0 & 0 & \cdots & 0 \end{bmatrix}$$

where the zero blocks on the diagonal are square. If T has no zero rows or columns, and $T_1 T_2 \cdots T_d$ is irreducible, show using Exercise 1.15(a), that T is irreducible. [*Hint:* Consider $y \geq \mathbf{0}, \neq \mathbf{0}$ partitioned according to $y' = [y_1', y_2', \ldots, y_d']$ where y_i has as many entries as the columns of T_i. Assuming $Ty \leq sy$ for some $s > 0$, show first that $y_1 > \mathbf{0}$, and then that $y_{i+1} > \mathbf{0} \Rightarrow y_i > \mathbf{0}$, $i = 1, \ldots, d$, $y_{d+1} \equiv y_1$.]

(Minc, 1974, Pullman, 1975)

CHAPTER 2

Some Secondary Theory with Emphasis on Irreducible Matrices, and Applications

In this chapter we survey briefly some of the theory which has arisen out of deeper investigation, and generalization, of various aspects of the Perron–Frobenius structure of a non-negative *irreducible* matrix T, with Perron–Frobenius eigenvalue r. Some of the material is of particular relevance in certain applications (e.g. mathematical economics, numerical analysis); these are also briefly discussed.

It is possible to extend several of the results to a reducible matrix T, via canonical form; some such discussion is deferred to the exercises.

2.1 The Equations: $(sI - T)x = c$

In a well-known mathematical–economic setting, to be discussed shortly, it is desired to investigate conditions ensuring positivity ($x > \mathbf{0}$) of solutions to the equation system

$$(sI - T)x = c$$

for any $c \geq \mathbf{0}, \neq \mathbf{0}$. Closely related to this is the question: for what values of s do we have $(sI - T)^{-1} > 0$?

Theorem 2.1. *A necessary and sufficient condition for a solution $x(x \geq \mathbf{0}, \neq \mathbf{0})$ to the equations*

$$(sI - T)x = c \tag{2.1}$$

to exist for any $c \geq \mathbf{0}, \neq \mathbf{0}$ is that $s > r$. In this case there is only one solution x, which is strictly positive and given by

$$x = (sI - T)^{-1}c.$$

PROOF. Suppose first that for some $c \geq \mathbf{0}$, $\neq \mathbf{0}$ a non-negative non-zero solution to (2.1) exists. Then

$$c + Tx = sx$$

i.e.

$$Tx \leq sx$$

with strict inequality for at least one element. This is impossible for $s \leq 0$, and if $s > 0$, Theorem 1.6 implies $s > r$.

Now suppose $s > r$. Then, since $T^k/s^k \to 0$ as $k \to \infty$ (see Exercise 1.14), it follows that

$$(sI - T)^{-1} = s^{-1}(I - s^{-1}T)^{-1} = s^{-1} \sum_{k=0}^{\infty} (s^{-1}T)^k$$

exists, from Lemma B.1 of Appendix B; and moreover, since for any pair i, j, $t_{ij}^{(k)} > 0$ for some $k = k(i, j)$ by irreducibility, it follows that the right hand side is a strictly positive matrix. Hence

$$(sI - T)^{-1} > 0$$

so that

$$(sI - T)^{-1}c > \mathbf{0} \quad \text{for any} \quad c \geq \mathbf{0}, \ \neq \mathbf{0},$$

and clearly

$$x = (sI - T)^{-1}c.$$

□

Corollary 1. *Of those real numbers s for which it exists, $(sI - T)^{-1} > 0$ if and only if $s > r$.*

Corollary 2. *If $s = 1$, then the necessary and sufficient condition stated for the theorem becomes $r < 1$.*

Corollary 3. *If $s = 1$, then $r < 1$ if none of the row (column) sums of T exceed unity, and at least one is less than unity.*

PROOF. Follows directly from Corollary 1 of the Perron–Frobenius Theorem (Theorem 1.1). □

Theorem 2.2. *A condition equivalent to $s > r$ is that*

$$\Delta_i(s) > 0, \qquad i = 1, 2, \ldots, n \tag{2.2}$$

where $\Delta_i(s)$ is the principal minor of $(sI - T)$ which consists of the first i rows and columns of $(sI - T)$.

PROOF.[1] Assume first that $s > r$. Then

$$\Delta_n = \det (sI - T) = \phi(s),$$

[1] We shall write in the proof Δ_i rather than $\Delta_i(s)$ for simplicity, since s is fixed.

in our previous notation, exceeds zero since it is known that $\phi(x) \to \infty$ as $x \to \infty$, and s lies beyond the largest real root r of T. Moreover, since s must exceed the maximal modulus real non-negative eigenvalue of the matrix formed by the first i rows and columns of T, $i < n$ (see Exercise 1.12), it must similarly follow that $\Delta_i > 0$ for $i = 1, 2, \ldots, n - 1$, if $n > 1$.

Assume now that (2.2) holds for some fixed real s. Since each of the Δ_i is a continuous function of the entries of T, it follows that it is possible to replace all the zero entries of T by sufficiently small positive entries to produce a positive matrix $\bar{T}$ with Perron–Frobenius eigenvalue $\bar{r} > r$ by Theorem 1.1(e), for which still $\bar{\Delta}_i > 0$, $i = 1, 2, \ldots, n$. Thus if we can prove that $s > \bar{r}$, this will suffice for what is required.

It follows, then, that it suffices to prove that

$$\Delta_i > 0,\ i = 1, 2, \ldots, n \quad \text{implies} \quad s > r$$

for *positive* matrices T, which is what we now assume about T. We proceed by induction on the dimension n of the matrix T. If $n = 1$, $\Delta_i > 0$ implies $s > t_{11} \equiv T \equiv r$.

Suppose now the proposition (2.2) is true for matrices of dimension n; and for a matrix T of dimension $(n + 1)$ assume

$$\Delta_i > 0,\ i = 1, 2, \ldots, n + 1.$$

If r_n is the Perron–Frobenius eigenvalue of the $(n \times n)$ positive matrix ${}_{(n)}T$ which arises out of crossing out the last row and column of T, we have by induction that $s > r_n$. Let

$$k_{n+1} \equiv \Delta_n/\Delta_{n+1} > 0,$$

and consider the unique solution $\boldsymbol{x} = \boldsymbol{\alpha}_{n+1}$ of the system

$$(sI - T)\boldsymbol{x} = \boldsymbol{f}_{n+1} \tag{2.3}$$

where $\boldsymbol{f}_{n+1} = (0, 0, \ldots, 0, 1)$. In the first instance, since

$$\boldsymbol{\alpha}_{n+1} = (sI - T)^{-1}\boldsymbol{f}_{n+1},$$

it must follow that the $n + 1$ element of $\boldsymbol{\alpha}_{n+1}$ is k_{n+1}, since this is the $(n + 1, n + 1)$ element of $(sI - T)^{-1}$. If we rewrite

$$\boldsymbol{\alpha}_{n+1} = (\boldsymbol{\alpha}_n', k_{n+1})',$$

it follows that $\boldsymbol{x} = \boldsymbol{\alpha}_n$ must satisfy

$$(s_{(n)}I - {}_{(n)}T)\boldsymbol{x} = \boldsymbol{\sigma}$$

where $$\sigma_n = t_{i,n+1}k_{n+1} > 0$$

from the first n equations of (2.3). But, since $s > r_n$, Theorem 2.1 implies that the unique solution, viz. $\boldsymbol{\alpha}_n$, is strictly positive. Hence $\boldsymbol{\alpha}_{n+1} > \mathbf{0}$ also, and the Subinvariance Theorem applied to (2.3) now implies $s > r_{n+1}$, as required. □

Theorem 2.3. *If $s \geq \max_i \sum_{j=1}^n t_{ij}$, and $c_{ij}(s)$ is the cofactor of the ith row and jth column of $sI - T$, then $c_{ii}(s) \geq c_{ij}(s) > 0$, all i, j.*

PROOF.[1] Adj $(sI - T)$ is the transposed matrix of cofactors c_{ij}, and is certainly positive if $s > r$ by Theorem 2.1 above, since

$$0 < (sI - T)^{-1} = \text{Adj } (sI - T)/\phi(s)$$

where $\phi(s) > 0$. Further, if $s = r$ it is also positive, by Corollary 2 to the Perron–Frobenius Theorem. Now, since Corollary 1 of the Perron–Frobenius Theorem asserts that $r \leq \max_i \sum_j t_{ij}$, $\leq s$ by assumption, it follows that all the cofactors are positive.

(i) Consider first the case $s > \max_i \sum_j t_{ij}$.

Now, replace any zero entries of T by a small positive number δ to form a new positive matrix $\bar{T} = \{\bar{t}_{ij}\}$ but such that still $s > \max_i \sum_j \bar{t}_{ij}$. If we can prove $\bar{c}_{ii} \geq \bar{c}_{ij}$ all i, j in this case, this suffices, for, by continuity in δ, letting $\delta \to 0+$, $c_{ij} \leq c_{ii}$.

Thus it suffices to consider a totally positive matrix $T = \{t_{ij}\}$ which we shall now do. Take i, and $j, j \neq i$ fixed (but arbitrary). Replace all elements of the ith row of T: by zeroes, except (i, i)th and (i, j)th where we put $s/2$. Call the new matrix U; it is clearly irreducible, and moreover has all row sums not exceeding s, and all but the ith less than s; thus its Perron–Frobenius eigenvalue $r_U < s$ by Corollary 1 to the Perron–Frobenius Theorem and so

$$0 < \det (sI - U) = -\frac{s}{2}c_{ij} + \frac{s}{2}c_{ii}$$

expanding the determinant by the ith row of $sI - U$ and recalling that the cofactors remain the same as for $sI - T$.

Therefore $$c_{ij} < c_{ii}$$

which is as required.

(ii) $s = \max_i \sum_j t_{ij}$. Take $\delta > 0$ and consider $s + \delta$ in place of s in $(sI - T)$. Then since $c_{ij}(s + \delta) \leq c_{ii}(s + \delta)$ for all $\delta > 0$, from part (i), it follows by continuity that, letting $\delta \to 0$, $c_{ij} \leq c_{ii}$ as required. □

Corollary 1. *If in fact $T > 0$ and $s > \max_i \sum_j t_{ij}$ then the conclusion can be strengthened to $c_{ij}(s) < c_{ii}(s)$ all i, j, $i \neq j$.*

Corollary 2. *If $s \geq \max_j \sum_{i=1}^n t_{ij}$ then $c_{ji}(s) \leq c_{ii}(s)$ for all i, j.*

PROOF. T', the transpose of T, satisfies the conditions of Theorem 2.3. □

[1] We shall write in the proof c_{ij} rather than $c_{ij}(s)$, since s is fixed.

The Open Leontief Model

Consider an economy in which there are n industries, each industry producing exactly one kind of good (commodity). Let x_i be the output of the ith industry, and $t_{ij} \geq 0$ the input of commodity i per unit output of commodity j. It follows then that $t_{ij}x_j$ is the amount of the output of commodity i absorbed in the production of commodity j, and the excess

$$c_i = x_i - \sum_{j=1}^{n} t_{ij}x_j, \; i = 1, \ldots, n$$

is the amount of commodity i available for outside use. Thus the vector $x = \{x_i\}$ may be interpreted as a 'supply' vector, and the vector $c = \{c_i\}$ as a 'demand' vector. A question of importance, therefore, is that of when for a given $c \geq \mathbf{0}, \; \neq \mathbf{0}$, is there a solution $x \geq \mathbf{0}, \; \neq \mathbf{0}$ to the system

$$(I - T)x = c. \tag{2.4}$$

Theorems 2.1 and 2.2 above answer this question in the special case $s = 1$, under the assumption of irreducibility of T. Moreover when the theorems hold, the positivity of $(I - T)^{-1}$, and the fact that $x = (I - T)^{-1}c$, guarantees that an increase in demand of even one good, increases the output of all goods.

If the necessary and sufficient condition $r < 1$ is replaced by the stronger assumption that for all i,

$$\sum_{j=1}^{n} t_{ij} \leq 1, \tag{2.5}$$

with strict inequality for at least one i, then Theorem 2.3 gives the additional result that if only the demand for commodity j increases then the output of commodity j increases by the greatest absolute amount, though all outputs increase.[1] A meaning in economic terms to condition (2.5) may be described as follows. If exactly one unit of each commodity is produced, then $\sum_{j=1}^{n} t_{ij}$ is the total input of commodity i required for this, so (2.5) asserts that not more than one unit of each commodity is required for such production; and at least one commodity is then also available for outside use. The condition dual to (2.5), viz.

$$\sum_{i=1}^{n} t_{ij} \leq 1,$$

with strict inequality for at least one j, may be interpreted as follows. If a unit of each commodity has the same monetary price (a dollar, say), then $\sum_{i=1}^{n} t_{ij}$ is the total cost of producing one unit of commodity j; hence at least one industry is able to pay the factors labour and capital.

[1] See Exercise 2.2.

Example. Consider the matrix

$$T_1 = \begin{bmatrix} \frac{2}{11} & \frac{6}{11} \\ \frac{8}{11} & \frac{4}{11} \end{bmatrix}.$$

T_1 is positive, hence irreducible. The top row sum is less than unity, the bottom one exceeds unity, so that the condition (2.5) is not satisfied. However the column sums are both $\frac{10}{11} < 1$ so that Corollary 2 of Theorem 2.3 does apply for $s = 1$. In fact $r = \frac{10}{11}$ by Corollary 1 of the Perron–Frobenius Theorem, so that Theorem 2.1 holds with $s = 1$. In fact

$$I - T_1 = \begin{bmatrix} \frac{9}{11} & \frac{-6}{11} \\ \frac{-8}{11} & \frac{7}{11} \end{bmatrix}, \quad (I - T_1)^{-1} = \frac{11}{15}\begin{bmatrix} 7 & 6 \\ 8 & 9 \end{bmatrix}$$

which agrees with the assertions of these results.

Note, also, that *if the demand for commodity 1 only increases*, by a single unit, the supply vector increases by $[\frac{77}{15}, \frac{88}{15}]$, *which is a greater increase in supply of commodity 2 than of commodity 1.*

On the other hand if the matrix

$$T_2 = T_1' = \begin{bmatrix} \frac{2}{11} & \frac{8}{11} \\ \frac{6}{11} & \frac{4}{11} \end{bmatrix}$$

is considered, then all of Theorems 2.1 to 2.3 hold and

$$(I - T_2)^{-1} = [(I - T_1)^{-1}]' = \frac{11}{15}\begin{bmatrix} 7 & 8 \\ 6 & 9 \end{bmatrix}$$

so that unit increase in demand of either commodity, the other being held constant, forces a greater increase in supply of that commodity than the other, as expected.

We thus pause to note that (i) it is not possible to increase the demand vector c even in one component while keeping any element of the supply vector x fixed; (ii) it is not necessarily true that an increase in demand for a single commodity forces a greater absolute increase in the supply of that commodity compared to others.

The above discussion has been in terms of absolute changes Δx_i, Δc_i, $i = 1, \ldots, n$ in the supply vector x and demand vector c. Of some interest is also the question of *relative* changes. If $c_i \neq 0$, $x_i \neq 0$ the relative changes are respectively defined by $\Delta x_i/x_i$, $\Delta c_i/c_i$, $i = 1, \ldots, n$, corresponding to some change in c or x. Let Γ_+ consist of those indices $i = 1, \ldots, n$ for which $\Delta c_i > 0$, and Γ_- of those for which $\Delta c_i < 0$. Write $\Delta c = \{\Delta c_i\}$, $\Delta x = \{\Delta x_i\}$.

Theorem 2.4. *Let $s > r$ and $c \geq \mathbf{0}$, $\neq \mathbf{0}$. Then for each $i = 1, \ldots, n$*

$$\min\left\{0, \min_{j \in \Gamma_-} \frac{\Delta x_j}{x_j}\right\} \leq \frac{\Delta x_i}{x_i} \leq \max\left\{0, \max_{j \in \Gamma_+} \frac{\Delta x_j}{x_j}\right\}.$$

PROOF. We shall prove only the right-hand side, recalling that by Theorem 2.1, $x > \mathbf{0}$. It is implicitly assumed also that $\bar{c} = c + \Delta c \geq \mathbf{0}$, $\neq \mathbf{0}$, so that $\bar{x} = x + \Delta x > \mathbf{0}$.

If $\Gamma_+ = \phi$ (the empty set), then $\Delta x = (sI - T)^{-1}\, \Delta c \leq \mathbf{0}$, so the proposition is evidently true, as it is in the case $\Gamma_+ = \{1, 2, \ldots, n\}$. Hence assume that the number of indices in Γ_+ is one of the integers $\{1, 2, \ldots, n-1\}$, and let $\alpha \geq 0$ be the smallest integer such that $t_{i,j}^{(\alpha)} > 0$ for some $i \notin \Gamma_+$ and some $j \in \Gamma_+$; clearly $\alpha \geq 1$ since $t_{i,j}^{(0)} = 0$ for $i \neq j$. Hence for $k = 0, \ldots, \alpha - 1$ all elements of $T^k \Delta c$ with index not in Γ_+ are non-positive. Using the identity for $k \geq 1$

$$(s^k I - T^k)(sI - T)^{-1} = \sum_{h=0}^{k-1} s^{k-1-h} T^h = \sum_{h=1}^{k} s^{k-h} T^{h-1}$$

with $k = \alpha$, and multiplying from the right by Δc

$$(s^\alpha I - T^\alpha)\, \Delta x = \sum_{h=1}^{\alpha} s^{\alpha-h} T^{h-1}\, \Delta c$$

we find that for each $i \notin \Gamma_+$

$$s^\alpha\, \Delta x_i - \sum_{j=1}^{n} t_{i,j}^{(\alpha)}\, \Delta x_j = \sum_{h=1}^{\alpha} s^{\alpha-h} \sum_{j \notin \Gamma_+} t_{i,j}^{(h-1)}\, \Delta c_j \leq 0$$

so that for $i \notin \Gamma_+$

$$s^\alpha\, \Delta x_i - \sum_{j \notin \Gamma_+} t_{i,j}^{(\alpha)}\, \Delta x_j \leq \sum_{j \in \Gamma_+} t_{i,j}^{(\alpha)}\, \Delta x_j.$$

Thus if $\Delta x_j \leq 0$ for all $j \in \Gamma_+$, it follows that

$$(s^\alpha I^* - (T^\alpha)^*)\, \Delta^* x \leq \mathbf{0}$$

where * indicates restriction of the corresponding matrix to indices $i \notin \Gamma_+$. Also, since $T^k/s^k \to 0$ as $k \to \infty$, it follows $\{(T^\alpha)^*/s^\alpha\}^k \to 0$, so by Lemma B.1 of Appendix B, $(s^\alpha I^* - (T^\alpha)^*)^{-1} \geq 0$, whence it follows $\Delta^* x \leq \mathbf{0}$. *Thus if $\Delta x_j \leq 0$ for all $j \in \Gamma_+$, $\Delta x_i \leq 0$ for all $i = 1, \ldots, n$, so the proposition is true again.*

It thus remains to prove it in the case that $\Delta x_j > 0$ for some j in Γ_+, a non-empty and proper subset of $\{1, 2, \ldots, n\}$, in which case it amounts to

$$\Delta x_i / x_i \leq \max_{j \in \Gamma_+} \frac{\Delta x_j}{x_j} \equiv \max_{j = 1, \ldots, n} \frac{\Delta x_j}{x_j}$$

for $i = 1, \ldots, n$. Suppose to the contrary that this last fails, and write

$$\Lambda = \{i : i = 1, \ldots, n;\ \Delta x_i / x_i = \max_j \Delta x_j / x_j\}.$$

There are real numbers $\mu_i, \lambda_i > 0$, $i = 1, \ldots, n$ such that

$$c_i = \mu_i \bar{c}_i, \qquad x_i = \lambda_i \bar{x}_i$$

such that $\mu_i \neq 1$ for $i \in \Gamma_- \cup \Gamma_+$. Since $\Delta x_i / x_i = \lambda_i^{-1} - 1$, $i = 1, \ldots, n$, it follows that for $i \in \Lambda$

$$\lambda_i^{-1} - 1 > \max_{j \in \Gamma_+}(\lambda_j^{-1} - 1) > 0,$$

or equivalently, for $i \in \Lambda$

$$\min_j \lambda_j = \lambda_i < \min_{j \in \Gamma_+} \lambda_j.$$

Now, using the earlier identity again

$$s^k x_i = \sum_{j=1}^n t_{i,j}^{(k)} x_j + \sum_{j=1}^n \sum_{h=1}^k s^{k-h} t_{i,j}^{(h-1)} c_j, \qquad i = 1, \ldots, n$$

so that

$$s^k \bar{x}_i = \sum_{j=1}^n t_{i,j}^{(k)} \frac{\lambda_j \bar{x}_j}{\lambda_i} + \sum_{j=1}^n \sum_{h=1}^k s^{k-h} t_{i,j}^{(h-1)} \frac{\mu_j \bar{c}_j}{\lambda_i}.$$

Since also $(sI - T)\bar{x} = \bar{c}$, also for $i = 1, \ldots, n$ we have analogously from the identity

$$s^k \bar{x}_i = \sum_{j=1}^n t_{i,j}^{(k)} \bar{x}_j + \sum_{j=1}^n \sum_{h=1}^k s^{k-h} t_{i,j}^{(h-1)} \bar{c}_j.$$

From the last two equations

$$0 = \sum_{j=1}^n t_{i,j}^{(k)} \bar{x}_j \left(\frac{\lambda_j}{\lambda_i} - 1 \right) + \sum_{j=1}^n \sum_{h=1}^k s^{k-h} t_{i,j}^{(h-1)} \bar{c}_j \left(\frac{\mu_j}{\lambda_i} - 1 \right).$$

Now, let $\alpha (\geq 1)$ be the smallest positive number such that $t_{i_0, j_0}^{(\alpha)} > 0$ for some $i_0 \in \Lambda$ and some $j_0 \in \{1, 2, \ldots, n\} - \Lambda$. Then

$$0 = \sum_{j=1}^n t_{i_0, j}^{(\alpha)} \bar{x}_j \left(\frac{\lambda_j}{\lambda_{i_0}} - 1 \right) + \sum_{j \in \Lambda} \sum_{h=1}^{\alpha} s^{\alpha-h} t_{i_0, j}^{(h-1)} \bar{c}_j \left(\frac{\mu_j}{\lambda_{i_0}} - 1 \right)$$

since $t_{i_0, j}^{(h-1)} = 0$, $h = 1, \ldots, \alpha$ if $j \in \{1, 2, \ldots, n\} - \Lambda$. Now $\lambda_{i_0} < 1$, and for $j \in \Lambda$, $\mu_j \geq 1$, since $\Lambda \cap \Gamma_+ = \phi$, so $\mu_j / \lambda_{i_0} - 1 > 0$, whence it follows that

$$\sum_{j=1}^n t_{i_0, j}^{(\alpha)} \bar{x}_j \left(\frac{\lambda_j}{\lambda_{i_0}} - 1 \right) \leq 0.$$

Since $\lambda_{i_0} < \lambda_{j_0}$, and $\lambda_{i_0} \leq \lambda_j$ for $j = 1, \ldots, n$, it follows that $t_{i_0, j_0}^{(\alpha)} = 0$ which is a contradiction. □

This theorem, proved under the minimal conditions of Theorem 2.1, shows in particular that if only the demand for commodity j increases, then the output of commodity j increases by the greatest percentage, though all outputs increase.

The economic model just discussed is called the open Leontief model; other economic interpretations may also be given to the same mathematical framework.

We shall mention a *dynamic* model, whose *static* version is formally (i.e. mathematically) identical with the Leontief set-up. Let the elements of x_k denote the output at time stage k of various industries as before; let β, $0 < \beta \leq 1$ be the proportion (the same for each industry) of output at any stage available for internal use, a proportion $1 - \beta$ being needed to meet external demand; let t_{ij} be the amount of output of industry i per unit of input of industry j, at the next time stage, and let $c \geq \mathbf{0}$, $\neq \mathbf{0}$, be an 'external input' vector, the same at all time stages. Then

$$x_{k+1} = \beta T x_k + c.$$

The general solution of this difference equation is

$$x_k = (\beta T)^k x_0 + \left\{ \sum_{i=1}^{k-1} (\beta T)^i \right\} c.$$

If T is assumed irreducible, as usual, then if $(\beta T)^i \to 0$ as $i \to \infty$, x_k converges elementwise to the solution

$$x = (I - \beta T)^{-1} c$$

(see Lemma B.1) of the 'stationary system'. Necessary and sufficient for $(\beta T)^i \to 0$ is $r < \beta^{-1}$ (see Exercise 1.14) which is familiar from Theorems 2.1 and 2.2 with $s = \beta^{-1}$.

Bibliography and Discussion to §2.1

Theorem 2.1, in the form of its Corollary 1, goes back to Frobenius (1912) for positive T.

The form of necessary and sufficient condition embodied in Theorem 2.2 is implicit in the paper of Hawkins & Simon (1949), who deal with positive T; and is explicit in a statement of Georgescu-Roegen (1951). The condition is thus often called the Hawkins–Simon condition in mathematico-economic contexts. Theorem 2.2 in fact holds for a non-negative T which is not necessarily irreducible[1] as was demonstrated by Gantmacher (1959, pp. 85–9; original Russian version: 1954) who himself attributes this result to Kotelyanskii (1952) whose proof he imitates, although in actual fact Kotelyanskii (*a*) considered $T > 0$, (*b*) proved a (non-trivial) variant[2] of

[1] See Exercise 2.4.

[2] See Exercise 2.13, (in §2.2).

Theorem 2.2 where every sign '$>$' in its statement is replaced by '$\geq$'. The Kotelyanskii–Gantmacher assertion was also obtained by Burger (1957) who followed the method of the Hawkins–Simon paper; and by Fan (1958). Further, Morishima (1964, Chapter 1) gives a proof of Theorem 2.2 which, however, makes the implicit assumption that if the last row and column of an irreducible T are deleted, then the matrix remaining is also irreducible. The virtue of the Hawkins–Simon condition, as well as the sufficient condition given in Corollary 3 of Theorem 2.1, is that such conditions may be checked fairly easily from the form of the matrix T, without the necessity of calculating r.

Finally, in relation to Theorem 2.2, it is relevant to mention that less powerful assertions of similar form, but involving the non-negativity of *all* (not just leading) principal minors, begin with Frobenius (1908).

Theorem 2.3 is generally attributed to Metzler (1945, 1951). Debreu & Herstein (1953) give a proof of its Corollary 1 which the proof of the present theorem imitates. (The present statement appears to be marginally more general than those usually encountered.) Theorem 2.4 is due to Sierksma (1979), though the techniques are partly Morishima's (1964; see esp. Chapter 1, Theorem 6).

A simple direct discussion of linear models of production in econometrics, including the Leontief model, is given by Gale (1960, Chapter 9). See also Allen (1965, Chapter 11). The reader interested in an emphasis on non-negative matrix theory in connection with the Leontief model, and alternative interpretations of it, should consult Solow (1952) and Karlin (1959, §§8.1–8.3). We also mention Kemeny & Snell's (1960, §7.7) discussion of this model, in which $\sum_i t_{ij} \leq 1$ all j, but T is not necessarily irreducible, in a Markov chain framework closely related to our development in Chapter 1.

Finally, for a very detailed and extensive mathematical discussion pertaining to the properties of the matrices $sI - T$, and further references, the reader should consult the articles of Wong and of Woodbury in Morgenstern (1954).

Exercises on §2.1

In Exercises 2.1 to 2.3, which have a common theme, T is an ireducible nonnegative matrix with Perron–Frobenius eigenvalue r.

2.1. Let $s > 0$. Show that a necessary and sufficient condition for $r < s$ is the existence of a vector x, $x \geq \mathbf{0}$, $\neq \mathbf{0}$ such that:

$$Tx \leq sx$$

with strict inequality in at least one position.

[The condition $Tx \leq x$, $x \geq \mathbf{0}$, $\neq \mathbf{0}$ is sometimes imposed as a fundamental assumption in simple direct discussions of linear models; in the Leontief model it is tantamount to an assertion that there is at least one demand c, $c \geq \mathbf{0}$, $\neq \mathbf{0}$ which can be met, i.e. the system is 'productive'.]

2.2. Show that if

$$s \geq \sum_{j=1}^{n} t_{ij},$$

with strict inequality for at least one i, then Theorem 2.1 holds, and, moreover, if in the equation system (2.1) $c \geq \mathbf{0}$, $\neq \mathbf{0}$ is increased in the jth component only, then the greatest absolute increase in x is in the jth component also.

2.3. Suppose T is any square non-negative matrix. Show that a *necessary* condition for $(sI - T)^{-1} > 0$ for some s is that T be irreducible, i.e. the situation $(sI - T)^{-1} > 0$ *may only occur if T is irreducible*. (Suppose initially that T has the partitioned form

$$T = \begin{bmatrix} A & 0 \\ C & B \end{bmatrix}$$

where A and B are square and irreducible, 0 is as usual a zero matrix, and $(sI - T)^{-1}$ exists.)

2.4.[1] Suppose T is any square non-negative matrix, with dominant eigenvalue ρ (guaranteed by Exercise 1.12). Show that $\Delta_i > 0$, $i = 1, \ldots, n$ if and only if $s > \rho$.

(Kotelyanskii, 1952)

[Since it is easily shown that $(sI - T)^{-1} \geq 0$ if $s > \rho$—see e.g. Debreu & Herstein (1953)—it follows that the Hawkins–Simon condition ensures *non-negativity* of solution to the Leontief system (2.1), taking $s = 1$, (i.e. (2.4)), even if T is reducible, for any 'demand' vector $c \geq \mathbf{0}$, $\neq \mathbf{0}$.]

2.5. Suppose, in the setting of Exercise 2.4, assuming all cofactors $c_{ij}(1) \geq 0$,

$$1 \geq \max_i \sum_{j=1}^{n} t_{ij}$$

and $\rho < 1$, use induction to show

$$1 > \Delta_1(1) > \Delta_2(1) > \cdots > \Delta_n(1) > 0$$

(Wong, 1954)

2.6. Suppose, in the context of the open Leontief model, that each industry supplies just one industry which differs from itself; and that each of the industries has only one supplier. Show by example that T is not necessarily irreducible. Is it possible to have T primitive in this context? In general, can any of the indices of T be inessential?

2.7. Under the conditions of Theorem 2.4 and using the same notation and approach, show that if $\Gamma_+ \neq \phi$ and $\Delta x_j < 0$ for all $j \in \Gamma_+$, then $\Delta x_j < 0$ for all $i = 1, \ldots, n$.

(Sierksma, 1979)

[1] Exercises 2.4 and 2.5 have a common theme.

2.2 Iterative Methods for Solution of Certain Linear Equation Systems

We consider in this section a system of equations

$$Ax = b \tag{2.6}$$

where $A = \{a_{ij}\}$, $i, j = 1, \ldots, n$ has all its *diagonal elements positive*, and b is some given column vector. The matrix A may be expressed in the form

$$A = D - E - F$$

where $D = \text{diag}\,\{a_{11}, a_{22}, \ldots, a_{nn}\}$ and E and F are respectively strictly lower and upper triangular $n \times n$ matrices, whose entries are the negatives of the entries of A respectively below and above the main diagonal of A. The equation system may, thus, be rewritten,

$$\{I - D^{-1}(E + F)\}x = D^{-1}b.$$

We shall make the assumption that

$$T_1 = D^{-1}(E + F)$$

is non-negative, and write for convenience

$$L = D^{-1}E,\ U = D^{-1}F \quad \text{and} \quad k = D^{-1}b.$$

Thus we are concerned basically with the equation system

$$(I - T)x = k \tag{2.7}$$

where T is non-negative. Thus the equation system considered in §2.1 falls into this framework, as do any systems where A has

(i) positive diagonal and non-positive off-diagonal entries; or
(ii) negative diagonal and non-negative off-diagonal entries (the case (ii) occurs in the theory of continuous parameter Markov processes).

There are two well-known iterative methods for attempting to find a solution x (if one exists) of (2.7) corresponding respectively to the two ways

$$x = (L + U)x + k$$
$$(I - L)x = Ux + k$$

of rewriting the equation system (2.7).

Jacobi iteration: The $(m + 1)$th approximation $x(m + 1)$ is derived from the mth by

$$x(m + 1) = (L + U)x(m) + k.$$

$T_1 = (L + U)$ is *the Jacobi matrix* of the system.

Gauss–Seidel iteration: $x(m+1)$ is related to $x(m)$ by

$$(I - L)x(m+1) = Ux(m) + k$$

i.e.
$$x(m+1) = (I - L)^{-1}Ux(m) + (I - L)^{-1}k.$$

$(I - L)^{-1}$ exists, since L is strictly lower triangular: in fact clearly, because of this last fact,

$$L^k = 0,\ k \geq n, \quad \text{so that}$$

$$(I - L)^{-1} = \sum_{k=0}^{n-1} L^k$$

(see Lemma B.1 of Appendix B).

The matrix $T_2 = (I - L)^{-1}U$ is the Gauss–Seidel matrix of the system.

We shall assume that T_1 is irreducible, with Perron–Frobenius eigenvalue r_1; the matrix $T_2 = (I - L)^{-1}U$ is reducible,[1] so let r_2 denote its dominant (non-negative) eigenvalue in accordance with the content of Exercise 1.12. The significance of the following theorem concerning relative size of r_1 and r_2 will be explained shortly. (Note that irreducibility of T_1 implies $L \neq 0$, $U \neq 0$.)

Theorem 2.5. *One and only one of the following relations is valid:*

$$(1)\ 1 = r_2 = r_1,$$

$$(2)\ 1 < r_1 < r_2,$$

$$(3)\ 0 < r_2 < r_1 < 1.$$

Proof. Let x be a positive eigenvector corresponding to the Perron–Frobenius eigenvalue r_1 of T_1. Then

$$(L + U)x = r_1 x$$

so that
$$(I - r_1^{-1}L)^{-1}(L + U)x = r_1(I - r_1^{-1}L)^{-1}x,$$

i.e.
$$(I - r_1^{-1}L)^{-1}Ux = r_1 x.$$

Thus r_1 is an eigenvalue and x an eigenvector of $(I - r_1^{-1}L)^{-1}U$ which is

$$\sum_{k=0}^{n-1} (r_1^{-1}L)^k U.$$

We deduce immediately (since the presence of the r_1 does not change the position of the positive or zero elements) that r_2, the dominant eigenvalue of

$$(I - L)^{-1}U = \sum_{k=0}^{n-1} L^k U, \textit{ satisfies } r_2 > 0, \quad \text{since} \quad r_1 > 0.$$

[1] See Exercise 2.8.

Further, if $r_1 \geq 1$, $(I - r_1^{-1}L)^{-1}U \leq (I - L)^{-1}U$; it follows that $r_2 \geq r_1$, from the content of Exercise 1.12. In fact, since both incidence matrices are the same, a consideration of Theorem 1.1(*e*) reveals $r_1 > 1 \Rightarrow r_2 > r_1$.

Now, suppose that y, $y \geq \mathbf{0}$, $\neq \mathbf{0}$ is a right eigenvector corresponding to the eigenvalue r_2 of T_2. Thus

$$(I - L)^{-1}Uy = r_2 y$$

so that

$$Uy = r_2(I - L)y,$$

i.e.

$$(r_2 L + U)y = r_2 y.$$

Since $r_2 L + U$ must be irreducible (since $L + U$ is), r_2 must be its Perron–Frobenius eigenvalue, and $y > \mathbf{0}$ (see Theorem 1.6). If $r_2 > 1$, then

$$r_2 L + U \leq r_2(L + U)$$

with strict inequality in at least one position, so that, from the Perron–Frobenius theory, $r_2 < r_2 r_1$, i.e. $1 < r_1$; but if $r_2 = 1$, then $r_1 = 1$. On the other hand, if $r_2 < 1$

$$r_2 L + U \leq L + U$$

with strict inequality in at least one position, so that $r_2 < r_1$.

Let us now summarize our deductions

(i) $r_1 \geq 1 \Rightarrow r_1 \leq r_2$; in fact $r_1 > 1 \Rightarrow r_1 < r_2$.
(ii) $r_2 > 1 \Rightarrow 1 < r_1$.
(iii) $r_2 = 1 \Rightarrow 1 = r_1$.
(iv) $r_2 < 1 \Rightarrow r_2 < r_1$.

The conclusion of the theorem follows. □

Returning now to the Jacobi and Gauss–Seidel methods, we may further investigate both methods by looking at the system, for fixed i:

$$x(m + 1) = T_i x(m) + \boldsymbol{\delta}_i, \qquad i = 1, 2$$

where $\boldsymbol{\delta}_i$ is some fixed vector. If a solution x to the system (2.7) exists then clearly the error vector at time $m + 1$, $\varepsilon(m + 1)$ is given by:

$$\varepsilon(m + 1) \equiv x(m + 1) - x = T_i[x(m) - x]$$

so that

$$\varepsilon(m) = T_i^m[x(0) - x] = T_i^m \varepsilon(0).$$

Thus if $T_i^m \to 0$ as $m \to \infty$ the ith iterative method results in *convergence* to the solution, which is then unique, in fact, since $T_i^m \to 0$ ensures that $(I - T_i)^{-1}$ exists. On the other hand, if $T_i^m \not\to 0$, *even if* a unique solution x is known to occur, the ith iterative method will not in general converge to it.

Before passing on to further discussion, the reader may wish to note that the problem of convergence here in the case $i = 1$ is identical with that of convergence of the dynamic economic model discussed at the conclusion of

§2.1. We have already seen there that since T_1 is irreducible, a necessary and sufficient condition for $T_i^m \to 0$ is $r_1 < 1$.

T_2 on the other hand is reducible, and, while we have not developed appropriate theory to cope with this problem, the Jordan canonical form of a matrix being usually involved, nevertheless the relevant result[1] yields an analogous conclusion viz. that $r_2 < 1$ is necessary and sufficient for $T_2^m \to 0$.

The significance of Theorem 2.5 is now clear: *either both the Jacobi and Gauss–Seidel methods converge, in which case, since $r_2 < r_1$, the latter converges (asymptotically) more quickly;* or neither method converges, in the sense that $T_i^m \not\to 0$, $i = 1, 2$.

In a more general framework, where T_1 is not necessarily assumed irreducible, this statement, together with the analogue of Theorem 2.5, has come to be known as the Stein–Rosenberg Theorem.

Bibliography and Discussion to §2.2

This section barely touches on the discussion of iterative methods of solution of linear equation systems; an exhaustive treatment of this topic with extensive referencing is given by Varga (1962). The two iterative methods discussed and compared here are more accurately known as the *point* Jacobi and *point* Gauss–Seidel iterative methods; both have various other designations. The proof of (our restricted case of) Theorem 2.5 follows the original paper of Stein & Rosenberg (1948); the proof given in Varga (1962) is different. Generalizations of the Stein–Rosenberg Theorem, which enable one to compare the convergence rates of any two iterative methods out of a class of such methods, are available: see e.g. Theorem 3.15 and subsequent discussion in Varga (1962) where appropriate references are given; and Householder (1958).

Exercises on §2.2

2.8. By considering first, for orientation, the example

$$A = \begin{bmatrix} -1 & 2 & 3 \\ 4 & -5 & 6 \\ 7 & 8 & -9 \end{bmatrix},$$

show that (whether T_1 is irreducible or not) T_2 is reducible.

2.9.[2] Suppose we are concerned with the iterative solution of the equation system

$$(I - T)x = b$$

[1] Which goes back to Oldenburger (1940); see e.g. also Debreu & Herstein (1953); Varga (1962), Chapter 1.

[2] Exercises 2.9 and 2.10 have a common theme.

where $T = \{t_{ij}\}$ is irreducible, and its Perron–Frobenius eigenvalue satisfies $r < 1$, but some of the t_{ii}, $i = 1, \ldots, n$ are positive, although all less than unity. One may then be inclined to make use of this 'natural form' of the matrix $A = I - T$ by putting

$$T = \bar{L} + \bar{U}$$

where $\bar{L}$ consists of the corresponding elements of T below the diagonal, and zeros elsewhere, and $\bar{U}$ consists of zeroes below the diagonal, but has its other entries, *on the diagonal* as well as above, the same at T. Show that the discussion of the Jacobi and Gauss–Seidel methods, including the relevant part of Theorem 2.5, goes through with $\bar{L}$, $\bar{U}$ replacing L, U respectively.

(Stein & Rosenberg, 1948)

2.10. In the framework of Exercise 2.9, if r_1 and r_2 have the same meaning as in the development of §2.2, show that $r_1 < r$ (so that the 'natural' Jacobi method described in Exercise 2.9 is less efficient than the standard method; an analogous result is true for the 'natural' Gauss–Seidel method as compared with the standard one).

2.3 Some Extensions of the Perron–Frobenius Structure

The theory of non-negative matrices may be used to determine analogous results for certain other matrices which have related structure, and occur in contexts as diverse as mathematical economics and number theory.

The foremost example occurs in the nature of square real matrices $B = \{b_{ij}\} i, j = 1, \ldots, n$ where $b_{ij} \geq 0$, $i \neq j$. We shall call such matrices ML-matrices, and examine their properties. Closely related properties are clearly possessed by matrices which have the form $-B$, where B is an ML-matrix.

Matrices of the type B or $-B$ are sometimes associated with the names of Metzler and Leontief in mathematical economics; and under an additional condition of the sort

$$\sum_i b_{ij} \leq 0, \text{ all } j; \quad \text{or} \quad \sum_j b_{ij} \leq 0, \text{ all } i,$$

with the names of Minkowski and Tambs-Lyche; and with the notion of a transition intensity matrix in the theory of finite Markov processes.

Matrices of form $A = -B$, where B is an ML-matrix may be written in the form

$$A = sI - T$$

where $s \geq 0$ is sufficiently large to make $T \geq 0$. If in fact this may be done so that also $s \geq \rho(T)$, the spectral radius of T, then A is called an M-matrix. If, further, one can do this so that $s > \rho(T)$ then A is said to be a non-singular M-matrix; since then $A^{-1} = (sI - T)^{-1} \geq 0$ (see Exercise 2.4), it has the inverse-positivity property (the inverse exists and has non-negative entries). Such matrices have already been considered in §2.1 when T is irreducible.

Although we carry forward the discussion in terms of ML-matrices only, it is clear that results such as Theorem 2.6 give information on inverse-positivity also.

Irreducible ML-matrices

An ML-matrix B may always be related to a non-negative matrix $T \equiv T(\mu)$ through the relation

$$T = \mu I + B$$

where $\mu \geq 0$, and is sufficiently large to make T non-negative.

Definition 2.1. An ML-matrix B is said to be irreducible[1] if T is irreducible.

(This definition is merely a convenience; irreducibility can clearly be defined—equivalently—directly in terms of the non-diagonal elements of B, in terms of 'cycles'—see Definition 1.2.)

By taking μ sufficiently large, the corresponding irreducible T can clearly be made *aperiodic* also and thus primitive; e.g. take $\mu > \max_i |b_{ii}|$.

We confine ourselves to irreducible ML-matrices B in this section, in line with the structure of the whole present chapter. The more important facts concerning such matrices are collected in the following theorem; for the reducible case see the exercises to this section.

Theorem 2.6. *Suppose B is an $(n \times n)$ irreducible ML-matrix. Then there exists an eigenvalue τ such that:*

(*a*) *τ is real;*
(*b*) *with τ are associated strictly positive left and right eigenvectors, which are unique to constant multiples;*
(*c*) *$\tau > \text{Re}\,\lambda$ for any eigenvalue λ, $\lambda \neq \tau$, of B (i.e. τ is larger than the real part of any eigenvalue λ of B, $\lambda \neq \tau$);*
(*d*) *τ is a simple root of the characteristic equation of B;*
(*e*) *$\tau \leq 0$ if and only if there exists $y \geq \mathbf{0}$, $\neq \mathbf{0}$ such that $By \leq \mathbf{0}$, in which case $y > \mathbf{0}$; and $\tau < 0$ if and only if there is inequality in at least one position in $By \leq \mathbf{0}$;*
(*f*) *$\tau < 0$ if and only if*

$$\Delta_i > 0,\ i = 1, 2, \ldots, n$$

[1] In the context of numerical analysis, an irreducible ML-matrix, C, is sometimes called *essentially positive*.

where Δ_i is the principal minor of $-B$ formed from the first i rows and columns of $-B$;
(*g*) $\tau < 0$ *if and only if* $-B^{-1} > 0$.

Proof. Writing $B = T - \mu I$ for μ sufficiently large to make T non-negative and irreducible, and noting as a result that if λ_i is an eigenvalue of B, then T has corresponding eigenvalue $\delta_i = \mu + \lambda_i$, and conversely, (*a*), (*b*), (*c*) and (*d*) follow from the Perron–Frobenius theory, τ being identified with $r - \mu$, where r is the Perron–Frobenius eigenvalue of T.

To see the validity of (*e*), let $\lambda_j = x_j + iy_j$ be an eigenvalue of B, $\lambda_j \neq \tau$, and suppose $x_j \geq \tau$. If

$$x_j > \tau, \; \delta_j = \mu + \lambda_j = \mu + x_j + iy_j,$$

where
$$\mu + x_j > \mu + \tau > 0,$$

so that
$$|\delta_j| > r = \mu + \tau$$

which is impossible. On the other hand, if $x_j = \tau$, but $y_j \neq 0$, again

$$|\delta_j| > \mu + \tau;$$

so the only possibility is $x_j = \tau$, $y_j = 0$ i.e. $\lambda_j = \tau$, which is again a contradiction.

The condition $By \leq \mathbf{0}$ may be written as

$$Ty \leq \mu y \qquad (\mu > 0)$$

which is tantamount to $\mu \geq r$ i.e.

$$\tau = r - \mu \leq 0,$$

by the Subinvariance Theorem (Theorem 1.6). The discussion of strict inequality follows similarly (also in Exercise 2.1).

The validity of (*f*) follows from Theorem 2.2; and of (*g*) from Corollary 1 of Theorem 2.1. □

An ML-matrix, B, occurs frequently in connection with the matrix exp (Bt), $t > 0$, in applications where the matrix exp (Bt) is defined (in analogy to the scalar case) as the pointwise limit of the infinite series[1] (which converges absolutely pointwise for each $t > 0$):

$$\sum_{k=0}^{\infty} (Bt)^k/k!$$

Theorem 2.7. *An ML-matrix B is irreducible if and only if* exp $(Bt) > 0$ *for all* $t > 0$. *In this case*

$$\exp(Bt) = \exp(\tau t)\boldsymbol{w}\boldsymbol{v}' + 0(e^{\tau' t})$$

[1] See Lemma B.2 of Appendix B.

elementwise as $t \to \infty$, where w, v' are the positive right and left eigenvectors of B corresponding to the 'dominant' eigenvalue τ of B, normed so that $v'w = 1$; and $\tau' < \tau$.

PROOF. Write $B = T - \mu I$ for sufficiently large $\mu > 0$, so that T is non-negative. Then

$$\exp(Bt) = \exp(-\mu t I)\exp(Tt)$$
$$= \exp(-\mu t)\exp(Tt)$$

and since

$$\exp(Tt) = \sum_{k=0}^{\infty} (tT)^k/k!$$

it follows that $\exp(Bt) > 0$ for any $t > 0$ if and only if T is irreducible, which is tantamount to the required.

Suppose now B is irreducible. Then with judicious choice of μ, $T = \mu I + B$ is primitive, with its Perron–Frobenius eigenvalue $r = \mu + \tau$. Invoking Theorem 1.2, we can write, if $\lambda_2 \neq 0$,

$$T^k = r^k wv' + 0(k^s|\lambda_2|^k), \quad |\lambda_2| < r$$

as $k \to \infty$, where w, v' have the properties specified in the statement of the present theorem, since the Perron–Frobenius eigenvectors of T correspond to those of B associated with τ. If $\lambda_2 = 0$, $T^k = r^k wv'$.

We may therefore write, for some δ, $0 < \delta < r$ that

$$T^k - r^k wv' = Y(k)\,\delta^k$$

where (the elements of) $Y(k) \to 0$ as $k \to \infty$.

Hence

$$\sum_{k=0}^{\infty} \frac{t^k T^k}{k!} - \sum_{k=0}^{\infty} \frac{t^k r^k}{k!} wv' = \sum_{k=0}^{\infty} Y(k)\frac{(t\delta)^k}{k!}$$

i.e. elementwise

$$|\exp(Tt) - \exp(rt)wv'| \leq \exp(t\delta)Y$$

where Y is a matrix of positive elements which bound the corresponding elements of $Y(k)$ in modulus, uniformly for all k. Thus cross multiplying by $\exp(-\mu t)$,

$$|\exp(Bt) - \exp(\tau t)wv'| \leq \exp[t(\delta - \mu)]Y$$

where $\delta - \mu < r - \mu = \tau$. Hence choosing τ' to be any fixed number in the interval $\delta - \mu < \tau' < \tau$, the assertion follows. □

Perron Matrices

Irreducible non-negative matrices T, and matrices B and $-B$ where B is an irreducible ML-matrix are special classes of the set of Perron matrices.

Definition 2.2. An $(n \times n)$ matrix $A = \{a_{ij}\}$ is said to be a *Perron matrix* if $f(A) > 0$ for some polynomial f with real coefficients.

Thus if T is non-negative irreducible, an appropriate polynomial, f, is given by

$$f(x) = \sum_{i=1}^{n} x^i,$$

since the matrix

$$f(T) = \sum_{i=1}^{n} T^i$$

will have a strictly positive contribution for any one of its elements from at least one of the T^i, $i = 1, \ldots, n$ (see Exercise 1.3).

An irreducible ML-matrix B may be written in the form $T - \mu I$, $\mu > 0$, where T is non-negative and primitive, so that $(B + \mu I)^k > 0$ for some positive integer k, and this is a real polynomial in B.

A further important subclass of Perron matrices, not previously mentioned, are the *power-positive* matrices A, i.e. matrices A such that $A^k > 0$ for some positive integer k. In this case, clearly, $f(x) = x^k$.

Perron matrices retain some of the features of the Perron–Frobenius structure of irreducible non-negative matrices discussed in Chapter 1, and we discuss these briefly.

Theorem 2.8. *Suppose A is an $(n \times n)$ Perron matrix. Then there exists an eigenvalue τ such that*

(*a*) *τ is real;*
(*b*) *with τ can be associated strictly positive left and right eigenvectors, which are unique to constant multiples;*
(*c*) *τ is a simple root of the characteristic equation of A.*

PROOF. (*a*) and (*b*): For $x > \mathbf{0}$, let

$$\tau(x) = \min_i \sum_j \frac{a_{ij} x_j}{x_i}.$$

Then, since for each i

$$x_i \tau(x) \le \sum_j a_{ij} x_j$$

$$x\tau(x) \le Ax$$

so that $$\mathbf{1}'x\tau(x) \le \mathbf{1}'Ax$$

i.e. $$\tau(x) \le \mathbf{1}'Ax/\mathbf{1}'x \le K$$

since each element of $\mathbf{1}'A$ is less than or equal to $\max_i \sum_j a_{ij} \equiv K$, so that $\tau(x)$ is bounded above *uniformly* for all $x > 0$. Now, define

$$\tau = \sup_{x > \mathbf{0}} \tau(x).$$

It follows that

$$\tau \geq \tau(\mathbf{1}) = \min_i \sum_{j=1}^{n} a_{ij}.$$

Now, let τ^* be defined by

$$\tau^* = \sup_{z \in \mathscr{C}} \tau(z)$$

where $\mathscr{C}$ is the set of vectors $z = f(A)y$ where $y \geq \mathbf{0}$, $y'y = 1$. Since this set of ys is compact, and the mapping is continuous on the set, $\mathscr{C}$ is compact, and since $z \in \mathscr{C} \Rightarrow z > \mathbf{0}$, clearly $\mathscr{C}$ is a subset of the set $\{x; x > \mathbf{0}\}$. Thus

$$\tau^* \leq \tau \tag{2.8}$$

and since $\tau(x)$ is a continuous mapping from $\{x; x > 0\}$ to R_1, it follows that τ^* is attained for some $z^* \in \mathscr{C}$. Now, since for any $x > 0$

$$Ax - x\tau(x) \geq \mathbf{0}$$

and x may be taken as normed to satisfy $x'x = 1$, (without change in $\tau(x)$) if we multiply this inequality from the left by $f(A)$

$$A(f(A)x) - (f(A)x)\tau(x) \geq \mathbf{0}$$

since $f(A)A = Af(A)$; and since $w = f(A)x \in \mathscr{C}$ it follows that

$$\tau(w) \geq \tau(x)$$

so that certainly

$$\tau^* \geq \tau$$

whence $\quad \tau^* = \tau, \quad$ from (2.8).

Thus we have that for some $z^* > \mathbf{0}$

$$Az^* - \tau z^* \geq \mathbf{0}.$$

Suppose now that the inequality is strict in at least one position. Then, as before

$$A(f(A)z^*) - \tau(f(A)z^*) > \mathbf{0}$$

so that for a $w = f(A)z^* > \mathbf{0}$

$$\tau(w) = \min_i \sum_j \frac{a_{ij} w_j}{w_i} > \tau$$

for each i, and this is a contradiction to the definition of τ. Hence

$$Az^* = \tau z^*,$$

so that, to complete the proof of assertions (a) and (b) of the theorem, it remains first to prove uniqueness to constant multiples of the eigenvector corresponding to τ; and secondly to prove that there is a similar left eigenvector structure.

Since, for arbitrary positive integer k, and *any eigenvector* (possibly with complex elements), $\boldsymbol{b}$

$$A^k\boldsymbol{b} = \tau^k\boldsymbol{b}$$

it follows that

$$f(A)\boldsymbol{b} = f(\tau)\boldsymbol{b}.$$

Now, $f(A)$ is a positive matrix, and since for $\boldsymbol{b}$ we can put $\boldsymbol{z}^*$, $>\boldsymbol{0}$, it follows from the Suninvariance Theorem that $f(\tau)$ is the Perron–Frobenius eigenvalue of $f(A)$ and hence $\boldsymbol{b}$ must be a multiple of $\boldsymbol{z}^*$, by the Perron–Frobenius Theorem.

As regards the left eigenvector structure, it follows that all the above theory will follow through, *mutatis mutandis*, in terms of a real number τ' replacing τ, where

$$\tau' = \sup_{\boldsymbol{x}>\boldsymbol{0}} \left\{ \min_j \frac{\sum_{i=1}^n x_i a_{ij}}{x_j} \right\}$$

so that there exists a unique (to constant multiples) positive left eigenvector $\boldsymbol{c}$ satisfying

$$\boldsymbol{c}'A = \tau'\boldsymbol{c}'$$

and also

$$A\boldsymbol{z}^* = \tau\boldsymbol{z}^*$$

Hence $\quad \boldsymbol{c}'A\boldsymbol{z}^* = \tau'\boldsymbol{c}'\boldsymbol{z}^* = \tau\boldsymbol{c}'\boldsymbol{z}^*, \quad$ and since $\quad \boldsymbol{c}'\boldsymbol{z}^* > 0, \tau' = \tau.$

(c): The fact that there is (to constant multiples) only one real $\boldsymbol{b} \neq \boldsymbol{0}$ such that

$$(\tau I - A)\boldsymbol{b} = \boldsymbol{0}$$

implies that $(\tau I - A)$ is of rank precisely $n - 1$, so that some set of $(n - 1)$ of its columns is linearly independent. Considering the $(n - 1) \times n$ matrix formed from these $(n - 1)$ columns, since it is of full rank $(n - 1)$, it follows[1] that $(n - 1)$ of its rows must be linearly independent. Thus by crossing out a certain row and a certain column of $(\tau I - A)$ it is possible to form an $(n - 1) \times (n - 1)$ non-singular matrix, whose determinant is therefore non-zero. Thus

$$\text{Adj}\ (\tau I - A) \neq 0$$

since at least one of its elements is non-zero. Thus, as in the proof of the Perron–Frobenius Theorem (part (f)), Adj $(\tau I - A)$ has all its elements non-zero, and all positive or all negative, since all columns are real non-zero

[1] By a well-known result of linear algebra.

multiples of the positive right eigenvector, and the rows are real non-zero multiples of the positive left eigenvector, the sign of one element of Adj $(\tau I - A)$ determining the sign of the whole.

Hence $$\phi'(\tau) \neq 0,$$

as in the proof of Theorem 1.1. □

Corollary 1.

$$\min_i \sum_j a_{ij} \leq \tau \leq \max_i \sum_j a_{ij}$$

with a similar result for the columns.

PROOF. We proved in the course of the development, that

$$\min_i \sum_j a_{ij} \leq r \leq \max_j \sum_i a_{ij}$$

and implicitly (through considering τ') that

$$\min_j \sum_i a_{ij} \leq \tau \leq \max_i \sum_j a_{ij}.$$ □

Corollary 2. *Either* Adj $(\tau I - A) > 0$ *or* $-$Adj $(\tau I - A) > 0$. (Proved in the course of (c)).

Corollary 3 (Subinvariance). *For some real* s,

$$Ax \leq sx$$

for some $\boldsymbol{x} \geq \mathbf{0}$, $\neq \mathbf{0}$ *implies* $s \geq \tau$; $s = \tau$ *if and only if* $Ax = sx$.

PROOF. Let $sx - Ax = \boldsymbol{f}; \boldsymbol{f} \geq \mathbf{0}$.

Then $$sc'x - c'Ax = c'f,$$

i.e. $$sc'x - \tau c'x = c'f.$$

Hence, since $s - \tau = c'f/c'x$, the result follows. □

Since the class of Perron matrices includes (irreducible) periodic non-negative matrices, it follows that there may be other eigenvalues of A which have modulus equal to that of τ. This cannot occur for *power-positive* matrices A, which are clearly a generalization of primitive, rather than irreducible, non-negative matrices.

For suppose that λ is any eigenvalue of the power-positive matrix A; thus for some possibly complex valued $\boldsymbol{b} \neq \mathbf{0}$,

$$A\boldsymbol{b} = \lambda \boldsymbol{b}$$

and so $$A^k \boldsymbol{b} = \lambda^k \boldsymbol{b}$$

for each positive integer k, and hence for that k for which $A^k > 0$. It follows as before that for this k, τ^k is the Perron–Frobenius eigenvalue of the primitive matrix A^k, and hence is uniquely dominant in modulus over all other eigenvalues.

Hence, in particular, for this k

$$|\tau^k| > |\lambda^k| \qquad \text{i.e. } |\tau| > |\lambda|$$

if $\lambda^k \neq \tau^k$.

Now suppose $\lambda^k = \tau^k$. Then it follows that any eigenvector $\boldsymbol{b}$ corresponding to λ must be a multiple of a single *positive* vector, since A^k is primitive, corresponding to the Perron–Frobenius eigenvalue τ^k of A^k.

On the other hand this positive vector itself must correspond to the eigenvalue τ of A *itself*: call it z^*.

Thus $$Az^* = \tau z^*, \; Az^* = \lambda z^*$$

Hence $$c'Az^* = \tau c'z^* = \lambda c'z^*$$

Therefore $$\tau = \lambda.$$

In this situation also, τ is clearly a *uniquely dominant eigenvalue in modulus* as with primitive matrices. It may, however, be *negative since for even* k, $A^k > 0$ where $A = -T$, $T > 0$.

Bibliography and Discussion to §2.3

The theory of ML-matrices follows readily from the theory developed earlier for non-negative matrices T and the matrices $sI - T$ derived from them, and no separate discussion is really necessary. Theorem 2.7 is useful, apart from other contexts, in the theory of Markov processes on finite state space, in that it describes the asymptotic behaviour of a probability transition (sub –) matrix exp (Bt) at time t, as t becomes large (B in this context being irreducible), and satisfying, for all i, $\sum_j b_{ij} \leq 0$ (see for example Mandl (1960); Darroch & Seneta (1967)).

Ostrowski (1937; 1956) calls a real $(n \times n)$ matrix $A = \{a_{ij}\}$ an M-matrix if: $a_{ij} \leq 0$, $i \neq j$, and A^{-1} exists, with $A^{-1} \geq 0$, in effect, and deduces certain results for such matrices. We shall not pursue these further in this book, mentioning only the books of Varga (1962) and Berman and Plemmons (1979) for further discussion and references.

The notion of a Perron matrix seems to be due to Dionísio (1963/4, Fasc. 1) who gives the results of Theorem 2.8, his method of proof (as also ours) imitating that of Wielandt (1950) as modified in Gantmacher (1959). It will be noted that this approach differs somewhat from that used in the proof of the Perron–Frobenius Theorem itself in our Chapter 1, that proof following Wielandt almost exclusively. Both proofs have been presented in the text for interest, although the initial stages of Wielandt's proof are not quite appropriate in the general context of a Perron matrix.

Power-positive matrices appear to have been introduced by Brauer (1961), whose paper the interested reader should consult for further details.

There are various other (finite—dimensional) extensions of the Perron–Frobenius structure of non-negative matrices. One such which has received substantial attention is that of an operator H, not necessarily linear, mapping $\{x;\ x \geq \mathbf{0}\}$ into itself, which is *monotone* ($0 \leq x_1 \leq x_2 \Rightarrow Hx_1 \leq Hx_2$), *homogeneous* ($\mathrm{H}(\alpha x) = \alpha \mathrm{H}(x)$ for $0 \leq \alpha < \infty$), and *continuous*, these properties being possessed by any non-negative matrix T. A weak analogue of the irreducibility is

$$\{\mathbf{0} \leq x_1 \leq x_2,\ x_1 \neq x_2\} \Rightarrow \{Hx_1 \leq Hx_2,\ Hx_1 \neq Hx_2\};$$

and for primitivity the right hand side of the implication must be supplemented by the existence of a positive integer m such that $H^m x_1 < H^m x_2$. A theorem much like the Perron–Frobenius Theorem in several respects can then be developed using a method of proof rather similar to that in Chapter 1. The interested reader should consult Brualdi, Parter & Schneider (1966), Morishima (1961, 1964), Solow & Samuelson (1953), and Taylor (1978).

It is also revelant to mention the paper of Mangasarian (1971) who considers the generalized eigenvalue problem $(A - \lambda B)x$ under variants of the assumption that $y'B \geq \mathbf{0}',\ \neq \mathbf{0}' \Rightarrow y'A \geq \mathbf{0}',\ \neq \mathbf{0}'$ to derive a Perron–Frobenius structure; the nature of this generalization becomes clear by putting $B = I$.

Finally we note certain generalizations of Perron–Frobenius structure obtainable by considering matrices which are not necessarily non-negative, but specified *cycles* of whose elements are (the definition of *cycle* is as in §1.2). This kind of theory has been developed in e.g. Maybee (1967); Bassett, Maybee & Quirk (1968); and Maybee & Quirk (1969)—from which further references are obtainable. A central theme of this work is the notion of a *Morishima matrix* (Morishima, 1952), which has properties closely allied to the Perron–Frobenius structure.

An $(n \times n)$ non-negative matrix T, with a view to further generalizations, may itself be regarded as a matrix representation of a linear operator which maps a convex cone of a partially ordered vector space (here the positive orthant of R_n) into itself, in the case when the cone is determined by its set of extreme generators. Berman and Plemmons (1979) give an extensive discussion.

Exercises on §2.3

2.11. Suppose B is an arbitrary $(n \times n)$ ML-matrix. Show that there exists a real eigenvalue ρ^* such that $\rho^* \geq \operatorname{Re} \lambda$ for any eigenvalue λ of B, and that $\rho^* < 0$ if and only if $\Delta_i > 0$, $i = 1, 2, \ldots, n$. (See Exercises to §2.1 of this chapter.)

2.12. In the situation of Exercise 2.11, show that if $\rho^* \leq 0$ then $\Delta_i \geq 0$, $i = 1, 2, \ldots, n$.

2.13. Let $B = \{b_{ij}\}$ be an $(n \times n)$ ML-matrix with $b_{ij} > 0$, $i \neq j$, whose 'dominant' eigenvalue is denoted by ρ^*. Show that (for $n \geq 2$) the condition: $\Delta_n \geq 0$, $\Delta_i > 0$, $i = 1, 2, \ldots, n-1$, ensures that $\rho^* \leq 0$.

Hint: Make use of the identity $(-B)\,\text{Adj}\,(-B) = \Delta_n I$, and follow the proof pattern of the latter part of Theorem 2.2.

[Kotelyanskii (1952) shows that for such B the apparently weaker condition $\Delta_i \geq 0$, $i = 1, \ldots, n$ *implies* $\Delta_i > 0$, $i = 1, \ldots, n-1$. Thus, taking into account also Exercise 2.12 above, $\Delta_i \geq 0$, $i = 1, \ldots, n$ is necessary and sufficient for $\rho^* \leq 0$ for such B.]

2.14. Let $B = \{b_{ij}\}$ be an irreducible $(n \times n)$ ML-matrix with 'dominant' eigenvalue ρ^*. Show that

$$\min_i \sum_{j=1}^{n} b_{ij} \leq \rho^* \leq \max_i \sum_{j=1}^{n} b_{ij}$$

with either equality holding if and only if both hold, making use of the analogous result for a non-negative irreducible matrix T.

If it is assumed, further, that

$$\sum_{j=1}^{n} b_{ij} \leq 0$$

for every i, show that: $\Delta_i \geq 0$, $i = 1, \ldots, n$; and that $\Delta_i > 0$, $i = 1, \ldots, n$ if and only if some row sum of B is *strictly* negative. Hence deduce that $\Delta_n \neq 0$ if and only if some row sum of B is strictly negative. [The variants of this last statement have a long history; see Taussky (1949).]

2.15. If B is an $(n \times n)$ ML-matrix, not necessarily irreducible, but satisfying

$$\sum_{j=1}^{n} b_{ij} \leq 0,$$

show that $\Delta_i \geq 0$, $i = 1, \ldots, n$.

(Ledermann, 1950*a*)

2.16. Show that a non-negative reducible matrix T cannot be a Perron matrix.

2.4 Combinatorial Properties

It has been remarked several times already in Chapter 1 that many properties of a non-negative matrix T depend only on the *positions* of the positive and zero elements within the matrix, and *not on the actual size* of the positive elements. Thus the classification of indices into essential and inessential, values of periods of indices which communicate with themselves, and hence investigation of the properties of irreducibility and primitivity in relation to a given non-negative T, all depend only on the location of the positive entries. This is a consequence of the more general fact that the positions of the positive and zero elements in all powers, T^k, k a positive integer, depend only on the positions in T.

It thus follows that to make a general study of the sequence of the powers T^k, $k = 1, 2, \ldots$ it suffices for example to replace, for each $k \geq 1$, in the matrix T^k the element $t_{ij}^{(k)}$ by unity, whenever $t_{ij}^{(k)} > 0$. The matrices so obtained may still be obtained as appropriate powers of the matrix T so modified (denoted say by $\tilde{T}$), if we accept that the rules of addition and multiplication involving elements 0, 1 pairwise are those of a simple Boolean algebra, i.e. the rules are the usual ones, with the single modification that $1 + 1 = 1$. The basic matrix $\tilde{T}$ has been called the incidence matrix of the matrix T in Chapter 1; such matrices with the Boolean algebra rules for composition as regards elements of their products, are also known in certain contexts as *Boolean relation matrices*.

An alternative formulation for the study of the power structure of a nonnegative matrix is in graph theoretic terms; while a third is in terms of a mapping, F, induced by a matrix T, of its index set $\{1, 2, \ldots, n\}$ into itself. We shall pursue this last approach since it would seem to have the advantage of conceptual simplicity (as compared to the graph-theoretic approach) within the limited framework with which we are concerned in the present chapter. Nevertheless, it should be stressed that all three approaches are quite equivalent, merely being different frameworks for the same topic, viz. the *combinatorial* properties of non-negative matrices.

Denote now by S the set of indices $\{1, 2, \ldots, n\}$ of an irreducible matrix T, and let $L \subset S$. Further, for integer $h \geq 0$, let $F^h(L)$ be the set of indices $j \in S$ such that

$$t_{ij}^{(h)} > 0 \quad \text{for some} \quad i \in L.$$

(If $L = \phi$, the empty set, put $F^h(\phi) = \phi$.)

Thus $F^h(i)$ is the set of $j \in S$ such that $t_{ij}^{(h)} > 0$. Also $F^0(L) = L$ by convention. Further, we notice the following easy consequences of these definitions:

(i) $A \subset B \subset S$, then $F(A) \subset F(B)$.
(ii) If $A \subset S$, $B \subset S$, then $F(A \cup B) = F(A) \cup F(B)$.
(iii) For integer $h \geq 0$, and $L \subset S$,

$$F^{h+1}(L) = F(F^h(L)) = F^h(F(L)).$$

(iv) The mapping F^h may be interpreted as the F-mapping associated with the non-negative matrix T^h, or as the hth iterate (in view of (iii)) of the mapping F associated with the matrix T.

We shall henceforth restrict ourselves to irreducible matrices T; and eventually to primitive matrices T, in connection with a study of the *index of primitivity* of such T.

Definition 2.3. The minimum positive integer $\gamma \equiv \gamma(T)$ for which a primitive matrix T satisfies $T^\gamma > 0$ is called the index of primitivity (or exponent) of T.

Irreducible Matrices

We shall at the outset assume that $n \geq 2$, to avoid trivialities, and note that for an irreducible T, $F(i) \neq \phi$, for each $i \in S$.

Lemma 2.1. $F(S) = S$. *If L is a proper subset of S and $L \neq \phi$ then $F(L)$ contains some index not in L.*

Proof. If $F(S)$ were a proper subset of S, then T would contain a zero column; this is not possible for irreducible T. If for a non-empty proper subset L of S, $F(L) \subset L$, then this also contradicts irreducibility, since then $F^h(L) \subset L$ all positive integer h, and hence for $i \in L$, $j \notin L$, $i \not\rightarrow j$. □

Lemma 2.2. *For $0 \leq h \leq n - 1$, $\{i\} \cup F(i) \cup \cdots \cup F^h(i)$ contains at least $h + 1$ indices.*

Proof. The proposition is evidently true for $h = 0$. Assume it is true for some h, $0 \leq h < n - 1$; then

$$L = \{i\} \cup F(i) \cup \cdots \cup F^h(i)$$

contains at least $h + 1$ indices, and one of two situations occurs:

(*a*) $L = S$, in which case

$$\{i\} \cup F(i) \cup \cdots \cup F^{h+1}(i) = S$$

also, containing $n > h + 1$ elements, so that $n \geq h + 2$, and the hypothesis is verified; or

(*b*) L is a proper non-empty subset of S in which case

$$F(L) = F(i) \cup \cdots \cup F^{h+1}(i)$$

contains at least one index not in L (by Lemma 2.1), and since $i \in L$,

$$\{i\} \cup F(L) = \{i\} \cup F(i) \cup \cdots \cup F^{h+1}(i)$$

contains all the indices of L and at least one not in L, thus containing at least $h + 2$ elements. □

Corollary 1. *If $\{i\} \cup F(i) \cup \cdots \cup F^{h-1}(i)$ is a proper subset of S, then, with the union of $F^h(i)$, at least one new element is added. Thus, if*

$$\{i\} \cup F(i) \cup \cdots \cup F^h(i), \qquad h \leq n - 2,$$

contains precisely $h + 1$ elements, then union with each successive $F^r(i)$, $r = 1, 2, \ldots, n - 2$ adds precisely one new element.

Corollary 2. *For any $i \in S$, $i \cup F(i) \cup \cdots \cup F^{n-1}(i) = S$.* (This was proved directly in Exercise 1.3.)

For purposes of the sequel we introduce a new term.

Definition 2.4. An irreducible ($n \times n$) matrix T is said to be *deterministic* if it is periodic with period $d = n$ (i.e. each cyclic subset of the matrix contains only one index). Equivalently, for such an irreducible matrix, for each $i \in S$ there is only one $j \in S$ such that $t_{ij} > 0$.)

Lemma 2.3. *For $i \in S$, $F^h(i)$ contains at least two indices for some h, $1 \leq h \leq n$ unless the deterministic case obtains.*

PROOF. Since $\{i\} \cup F(i) \cup \cdots \cup F^{n-1}(i) = S$, two cases are possible:

(a) each of the $F^h(i)$, $h = 0, \ldots, n-1$, contains precisely one index, and all differ from each other. Now if $F^n(i)$ contains only one index, either $F^n(i) = \{i\}$ and we are in the deterministic case; or $F^n(i) = F^h(i)$ for some h, $1 \leq h \leq n-1$, which is impossible as irreducibility of T is contradicted. Otherwise $F^n(i)$ contains at least two indices.

(b) some $F^h(i)$, $1 \leq h \leq n-1$ contains at least two indices. □

We now pass on to a study of a general upper bound for $\gamma(T)$ *depending only on the dimension* $n(\geq 2$ by earlier assumption) for primitive T. For subclasses of ($n \times n$) primitive matrices T satisfying additional structural conditions, stronger results are possible.[1]

Theorem 2.9. *For a primitive ($n \times n$) matrix, T, $\gamma \leq n^2 - 2n + 2$.*

PROOF. According to Lemma 2.3, for arbitrary fixed $i \in S$,

$$\{i\} \cup F(i) \cup \cdots \cup F^{n-1}(i) = S,$$

and either (a) $F^h(i)$, $h = 0, \ldots, n-1$ all contain precisely one index, in which case $F^n(i)$ contains at least two[2] indices; or (b) one of the $F^h(i)$, $1 \leq h \leq n-1$ contains at least two.

(a) Since $\{i\} \cup \cdots \cup F^{n-1}(i) = S$, it follows $F(i) \cup \cdots \cup F^n(i) = F(S) = S$, in which case $F^n(i)$ must contain i, and at least one index not i, i.e. one of $F^h(i)$, $h = 1, \ldots, n-1$. Hence for some integer $m \equiv m(i)$, $1 \leq m < n$,

$$F^m(i) \subseteq F^n(i) = F^{m+(n-m)}(i),$$

so that operating repeatedly with F^{n-m}:

$$F^m(i) \subseteq F^{m+(n-m)}(i) \subseteq F^{m+2(n-m)}(i) \subseteq \cdots \subseteq F^{m+(n-1)(n-m)}(i)$$

and by Corollary 2 of Lemma 2.2,

$$F^{m+(n-1)(n-m)}(i) = S.$$

Now

$$\begin{aligned} m + (n-1)(n-m) &= n + (n-2)(n-m) \leq n + (n-2)(n-1) \\ &= n^2 - 2n + 2. \end{aligned}$$

[1] See Exercise 1.3, and Exercises 2.17 and 2.18 of the sequel, and Bibliography and Discussion.

[2] The deterministic case is excluded by assumption of primitivity of T.

(*b*) If one of the $F^h(i)$, $h = 1, 2, \ldots, n-1$ contains at least two indices, we further differentiate between two cases:

(*b*.1) $\{i\} \cup F(i) \cup \cdots \cup F^{n-2}(i) \neq S$. Then by Corollary 1 of Lemma 2.2, each of $F^h(i)$, $h = 0, \ldots, n-2$ adds precisely one new element, and by Corollary 2, $F^{n-1}(i)$ contributes the last element required to make up S. Let $p \equiv p(i)$, $1 \leq p \leq n-1$, be the smallest positive integer such that $F^p(i)$ contains at least two elements. Then there exists an integer m, $0 \leq m < p$ such that $F^m(i) \subseteq F^p(i)$. Proceeding as in (*a*),

$$F^{m+(n-1)(p-m)}(i) = S$$

and

$$\begin{aligned} m + (n-1)(p-m) = p + (n-2)(p-m) &\leq (n-1) + (n-2)(n-1) \\ &= (n-1)^2 < n^2 - 2n + 2. \end{aligned}$$

(*b*.2) $\{i\} \cup F(i) \cup \cdots \cup F^{n-2}(i) = S$. Then

$$S = F(S) = F(i) \cup \cdots \cup F^{n-1}(i),$$

as before, so that for some p, $1 \leq p \leq n-1$, $F^p(i) \supseteq F^0(i) = \{i\}$. Proceeding as before,

$$F^{0+(n-1)p}(i) = S,$$

with

$$(n-1)p \leq (n-1)^2 < n^2 - 2n + 2.$$

Thus combining (*a*) and (*b*), we have that for *each* $i \in S$,

$$F^{n^2-2n+2}(i) = S$$

which proves the theorem. □

Corollary. $\gamma(T) = n^2 - 2n + 2$ *if and only if a simultaneous permutation of rows and columns of T reduces it to an (almost deterministic) form having incidence matrix*

$$\tilde{T} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ & & & & \vdots & \\ 0 & 0 & 0 & & \cdots & 1 \\ 1 & 1 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Proof. From the proof of Theorem 2.9, it may be seen that only in case (*a*) is it possible that there exists a row, say the ith, which becomes entirely positive for the first time only with a power as high as $n^2 - 2n + 2$, and then only with the conditions that $m = 1$ is the unique choice *and*

$$F(i) \subseteq F^n(i) \subseteq \cdots \subseteq F^{1+(n-2)(n-1)}(i) \neq S. \tag{2.9}$$

Ignoring (2.9) for the moment, the other conditions indicate that, taking i as the first index, $F(i)$ as the second, ..., $F^{n-1}(i)$ as the nth, a form with the incidence matrix displayed is the only one possible. It is then easily checked for this form that (2.9) holds, as required, for *each* successive operation on $F(i)$ by F^{n-1} adds *precisely one new* element, up to the $(n-1)$th operation. □

Bibliography and Discussion to §2.4

The approach of the present section is a combination of those of Holladay & Varga (1958) and Pták (1958). Wielandt (1950) stated the result of Theorem 2.9 (without proof) and gave the case discussed in its Corollary as an example that the upper bound is attainable. Theorem 2.9 was subsequently proved by Rosenblatt (1957) using a graph-theoretic approach, and by Holladay & Varga and Pták, in the papers cited. The full force of the Corollary to Theorem 2.9 would seem to be due Dionísio (1963/4, Fasc. 2), using a graph-theoretic approach.

The study of the combinatorial properties of non-negative matrices has expanded rapidly since the early contributions of Rosenblatt, Holladay, Varga, and Pták, usually with the aid of methods which are explicitly graph-theoretic. We shall mention only a limited number of these contributions, from which (or from whose authors) the interested reader should be able to trace others. Perkins (1961), Dulmage & Mendelsohn (1962, 1964) and Heap & Lynn (1964) have further pursued the structure of the powers of primitive T, obtaining sharper bounds than Wielandt's under certain structural conditions. Pullman (1964) has investigated the imprimitive case using a combinatorial approach. Subsequently, Heap & Lynn (1966, I), have considered the 'oscillatory' structure of the sequence of powers of irreducible and reducible T; and (1966, II) the maximum number of positive elements, ρ, which may be contained by a matrix of the sequence T^r, $r = 1, 2, \ldots$ and the positive integer ν for which T^r first contains ρ positive elements. Schwartz (1967) has made a study similar to that of Heap & Lynn (1966, I) for irreducible matrices, using a combinatorial approach.

Exercises on §2.4

All exercises refer to an $(n \times n)$ *primitive* matrix T, unless otherwise stated.

2.17. In Exercise 1.3, of Chapter 1, it was shown that if $t_{ii} > 0$, $i = 1, \ldots, n$, then $\gamma(T) \leq n - 1$. Adapting the reasoning of this exercise show that

(i) If T has exactly one positive diagonal entry, then $\gamma(T) \leq 2(n-1)$. (Rosenblatt, 1957)

(ii) If T has exactly $r \geq 1$ diagonal entries then $\gamma(T) \leq 2n - r - 1$. (Holladay & Varga, 1958)

2.18. If T is an irreducible matrix which has $t_{ij} > 0$ if and only if $t_{ji} > 0$, show that T^2 has *all* its diagonal elements positive. If, in addition, T is primitive, show that $\gamma(T) \le 2(n-1)$.

(Holladay & Varga, 1958)

2.19. Use the fact that $t_{ii}^{(r)} > 0$ for some r, $1 \le r \le n$, and the essence of part (i) of Exercise 2.17 above to obtain the bound $\gamma(T) \le 2n^2 - 2n$ for an *arbitrary* (primitive) T. (This bound, substantially weaker than that of Wielandt in Theorem 2.9, was already known to Frobenius (1912)).

2.20. If $k(T)$ is the *total* number of positive entries in T, show that $k(T^r)$, $r = 1, 2, \ldots$ may not (initially) be non-decreasing, by investigating the (9×9) matrix whose positive entries are $t_{11}, t_{12}, t_{23}, t_{24}, t_{25}, t_{36}, t_{37}, t_{38}, t_{46}, t_{47}, t_{48}, t_{56}, t_{57}, t_{58}, t_{69}, t_{79}, t_{89}, t_{91}$.

(Šidák, 1964)

2.21. In the framework of Exercise 2.20, show that $k(T^r)$ is non-decreasing with r if at least $(n-1)$ diagonal entries of T are positive. (Use Lemma 2.1.)

(Šidák, 1964*b*)

2.22. Let $T = \{t_{ij}\}$, $i, j = 1, \ldots, n$ be a non-negative matrix, not necessarily primitive, and suppose that for some permutation $\{p_1, p_2, \ldots, p_n\}$ of $\{1, 2, \ldots, n\}$, $\prod_{i=1}^{n} t_{ip_i} > 0$. If S is $n \times n$ non-negative, show that $k(T) \le k(TS)$ in the notation of Exercise 2.20; and hence that $k(T^r)$ is nondecreasing with r.

(Vrba, 1973)

2.5 Spectrum Localization

If T is an $n \times n$ irreducible non-negative matrix with Perron-Frobenius eigenvalue r, then in Corollary 1 to Theorem 1.1 of Chapter 1 we obtained the Frobenius-type inclusion

$$s \le r \le S \tag{2.10a}$$

where
$$s = \min_i \sum_{j=1}^{n} t_{ij}, \qquad S = \max_i \sum_{j=1}^{n} t_{ij};$$

with either equality holding if and only if all row sums are equal. A similar proposition holds for column sums. Related inequalities were discussed in Exercises, 1.6 to 1.8 of Chapter 1. It is easy to see from the canonical form of a square non-negative matrix T with spectral radius ρ, using (2.10*a*), that

$$s \le \rho \le S. \tag{2.10b}$$

As regards the whole spectrum, a problem of interest posed by Kolmogorov in 1938, is that of characterizing the region M_n of the complex plane which consists of all points which can be eigenvalues of $(n \times n)$ non-negative matrices T; clearly for the problem to be meaningful it is necessary to impose a condition restricting the size of the spectral radius ρ (assumed positive). If T has a left eigenvector $v' = \{v_i\} > \mathbf{0}'$ corresponding to ρ, or a right eigenvector $w = \{w_i\} > \mathbf{0}$, then the transformation

$$p_{ij} = v_j t_{ji}/v_i \rho \quad \text{or} \quad = t_{ij} w_j / w_i \rho \tag{2.11}$$

$i, j = 1, \ldots, n$, respectively, results in an $(n \times n)$ matrix $P = \{p_{ij}\}$ which has all row sums unity with (since essentially only a similarity transformation is involved) spectrum identical to the spectrum of T except that each eigenvalue of T is divided by ρ. Matrices T for which such a transformation may be performed include those which are irreducible, and those which have identical row sums $a > 0$ (in which case $\rho = a$, by (2.10b), and $\mathbf{1}$ is a corresponding right eigenvector), or which have identical column sums $a > 0$. In the sense of this discussion, it is therefore sufficient to consider the problem for $(n \times n)$ matrices which are stochastic.

Definition 2.5. A non-negative matrix $T = \{t_{ij}\}$, $i, j = 1, \ldots, n$, is called stochastic (or more precisely row stochastic) if

$$\sum_{j=1}^{n} t_{ij} = 1, \qquad \text{each } i.$$

It is called doubly stochastic if also

$$\sum_{i=1}^{n} t_{ij} = 1, \qquad \text{each } j.$$

It is clear that for a stochastic T, $\rho = 1$ and for any eigenvalue λ, $|\lambda| \leq 1$, so M_n for such T is contained in the unit circle.

In the sequel we concern ourselves only with determining a circle in the complex plane, centre origin, which contains all eigenvalues $\lambda \neq 1$ of a particular stochastic T. The procedure is no more difficult for an arbitrary real $(n \times n)$ matrix $A = \{a_{ij}\}$, $i, j = 1, \ldots, n$ with identical row sums a, and we shall carry it through in this setting. Since for a stochastic matrix T, $A = I - T$ has identical row sums (zero), the result (Theorem 2.10) enables us to obtain a circle, centre unity, containing all values $|1 - \lambda|$, where λ is an eigenvalue of T. Thus we concern ourselves (until the Bibliography and Discussion) with spectrum localization by circles.

In the following, for $x = \{x_i\}$, $\|x\|_1 = \sum |x_i|$ (the l_1 norm of x) and f_i denote the vector with unity in the i-th position and zeros elsewhere.

Lemma 2.4. *Suppose $\delta \in R^n$, $n \geq 2$, $\delta'\mathbf{1} = 0$, $\delta \neq \mathbf{0}$. Then for a suitable set $\mathscr{I} = \mathscr{I}(\delta)$ of ordered pairs of indices (i, j), $i, j = 1, \ldots, n$,*

$$\delta = \sum_{(i, j) \in \mathscr{I}} \left(\frac{\eta_{i,j}}{2}\right) \gamma(i, j), \tag{2.12}$$

where $\eta_{i,j} > 0$ and $\sum_{(i,j) \in \mathscr{I}} \eta_{i,j} = \|\delta\|_1$, and $\gamma(i, j) = f_i - f_j$.

PROOF. We proceed by induction on n. The proposition is evidently true for $n = 2$. Suppose it is true for $n = k \geq 2$, and consider $n = k + 1$. Then in $\boldsymbol{\delta} = \{\delta_i\}$, $i = 1, \ldots, k + 1$, choose p so that $|\delta_p| = \max |\delta_i|$, and q so that $\delta_q \neq 0$ and sign $\delta_q = -$sign δ_p. Put $\boldsymbol{\alpha}_1 = \delta_q(f_q - f_p)$, and write $\bar{\boldsymbol{\delta}} = (\boldsymbol{\delta} - \boldsymbol{\alpha}_1)$, which has a zero component in the q-th position, but clearly $\bar{\boldsymbol{\delta}} \neq \mathbf{0}$, $\bar{\boldsymbol{\delta}}'\mathbf{1} = 0$ and $\|\bar{\boldsymbol{\delta}}\|_1 = \|\boldsymbol{\delta}\|_1 - 2|\delta_q|$. We can thus apply the induction hypothesis to $\bar{\boldsymbol{\delta}}$ to obtain the required result for $\boldsymbol{\delta}$, where $\eta_{q,p}/2 = |\delta_q|$. □

Theorem 2.10. *Let $A = \{a_{ij}\}_{i,j=1}^n$ be a matrix with constant row sums a, and suppose λ is an eigenvalue of A other than a. Then*

$$|\lambda| \leq \tfrac{1}{2} \max_{i,j} \sum_{s=1}^{n} |a_{is} - a_{js}|. \tag{2.13}$$

Moreover the right-hand bound may be written in either of the alternative forms

$$a - \min_{i,j} \sum_{s=1}^{n} \min(a_{is}, a_{js}), \qquad \left(\max_{i,j} \sum_{s=1}^{n} \max(a_{is}, a_{js})\right) - a. \tag{2.14}$$

PROOF. Let $z' = (z_1, z_2, \ldots, z_n)$ be an arbitrary row vector of complex numbers. Then for any real $\boldsymbol{\delta} \neq \mathbf{0}$, $\boldsymbol{\delta}'\mathbf{1} = 0$ in view of Lemma 2.4

$$|z'\boldsymbol{\delta}| \leq \sum_{(i,j)\in \mathscr{I}} \left(\frac{\eta_{i,j}}{2}\right)|z_i - z_j| \leq (\tfrac{1}{2}) \max_{i,j} |z_i - z_j| \, \|\boldsymbol{\delta}\|_1.$$

Putting $f(z) = \max_{i,j} |z_i - z_j|$, we have

$$f(Az) = \max_{i,j} \left| \sum_{s=1}^{n} (a_{is} - a_{js})z_s \right| \leq (\tfrac{1}{2}) f(z) \max_{i,j} \sum_{s=1}^{n} |a_{is} - a_{js}|.$$

Thus for any right eigenvector z of A corresponding to an eigenvalue λ of A,

$$|\lambda| f(z) \leq (\tfrac{1}{2}) f(z) \max_{i,j} \sum_{s=1}^{n} |a_{is} - a_{js}|.$$

Now $\lambda \neq a$ implies $f(z) \neq 0$, since $f(z) = 0$ if and only if $z =$ const. $\mathbf{1}$. Hence if $\lambda \neq \alpha$,

$$|\lambda| \leq (\tfrac{1}{2}) \max_{i,j} \sum_{s=1}^{n} |\alpha_{is} - a_{js}|.$$

The alternative forms follow from the identity, valid for any n real pairs (x_i, y_i), $i = 1, \ldots, n$

$$|x_s - y_s| = (y_s + x_s) - 2\min(x_s, y_s) = 2\max(x_s, y_s) - (y_s + x_s). \quad \square$$

Bibliography and Discussion to §2.5

According to Zenger (1972), Theorem 2.10 is due to E. Deutsch, within the framework of Bauer, Deutsch and Stoer (1969). A generally even sharper bound than (2.13) for $\lambda \neq a$ is given by Deutsch and Zenger (1971) viz.

$$|\lambda| \leq \frac{1}{2} \max_{i,j} \left| a_{ii} + a_{jj} - a_{ij} - a_{ji} + \sum_{\substack{k \\ k \neq i,j}} |a_{ik} - a_{jk}| \right|.$$

A bound generally weaker than (2.12), in view of (2.14),

$$|\lambda| \leq \min \left\{ a - \sum_{s=1}^{n} \min_i a_{is}, \left(\sum_{s=1}^{n} \max_i a_{is} \right) - a \right\} \tag{2.15}$$

was obtained by Brauer (1971), and partially by Lynn and Timlake (1969).

Bauer, Deutsch and Stoer (1969) also showed that if T is non-negative and irreducible, then for $\lambda \neq r$, where r is the Perron-Frobenius eigenvalue of T,

$$\tau_B(T) \geq \tau_1(T) \geq |\lambda|/r \tag{2.16}$$

where

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

in terms of the transformed stochastic matrix $P = \{p_{ij}\}$ formed from T via (2.11), and

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

where

$$\begin{aligned} \phi(T) &= \min_{i,j,k,l} \frac{t_{ik} t_{jl}}{t_{jk} t_{il}} \qquad &\text{if } T > 0 \\ &= 0 &\text{if } T \not> 0. \end{aligned}$$

$\tau_B(T)$ and $\tau_1(T)$ are known as "coefficients of ergodicity" (Seneta, 1979) because of certain other properties and uses which will be discussed in Chapter 3 where the inequality $\tau_B(T) \geq \tau_1(T)$ will be obtained (Theorem 3.13) in a slightly more general setting.

The inequality $\tau_B(T) \geq |\lambda|/r$ in terms of the Birkhoff (1957) coefficient of ergodicity $\tau_B(\cdot)$ is due to Hopf (1963).

It is not surprising that some of the above results have been rediscovered in the probabilistic context of ergodicity coefficients. Thus the first of the bounds in (2.15) obtained by Pykh (1973), and in a review of this paper the bound (2.13) stated by Seneta (1974), for stochastic A. A proof of (2.13) was subsequently given, along lines similar to those followed above, by Alpin and Gabassov (1976).

The problem of improving the Frobenius bounds (2.10a) for r in the case in which not all row sums of T are the same, in the guise of determining positive numbers p_1 and p_2 such that

$$s + p_2 \leq r \leq S - p_1,$$

was suggested by Ledermann in 1950. For $T > 0$ such numbers p_1 and p_2, of successively 'sharper' nature were obtained by Ledermann (1950b), Ostrowski (1952) and Brauer (1957a) (see also Medlin (1953)). The case of irreducible T was similarly considered by Ostrowski and Schneider (1960). An account of these and other contributions is given by Marcus and Minc (1964).

Related lines of research for irreducible T, pertain to bounds involving the Perron-Frobenius eigenvector. If $w = \{w_i\} > \mathbf{0}$ is such a right eigenvector, then bounds are available for min w_i/max w_i (perhaps the simplest such is given in Exercise 2.22), min $w_i/\sum_i w_i$ and max $w_i/\sum_i w_i$. See e.g. Lynn and Timlake (1969); de Oliveira (1972); and Berman and Plemmons (1979) for further references.

The problem of determining M_n for stochastic matrices was partially solved by Dmitriev & Dynkin (1945), Dmitriev (1946), and the solution completed by Karpelevich (1951). It turns out that M_n consists of the interior and boundary of a simple curvilinear polygon with vertices on the unit circle, and is symmetric about the real axis. A description of this and related theory is available in the book of de Oliveira (1968, Chapter 2).

A second problem is that of determining, for a given non-negative T, a region which contains all the eigenvalues λ of T. Theorems which determine regions containing all eigenvalues for general (possibly complex valued) matrices in terms of their elements may be used for this purpose. The field here is broad; an account is available in the books of Varga (1962, §1.4 and its discussion) and Marcus & Minc (1964); and the paper of Timan (1972). Theorem 2.10 above is essentially pertinent to *this* problem.[1]

Further, we mention the difficult problem suggested by Suleĭmanova (1949, 1953) of investigating the n-dimensional region $\mathfrak{M}_n$ consisting of n-tuples of (real or complex) numbers, these n-tuples being the set of characteristic roots of some stochastic non-negative matrix T, this being a generalization of Kolmogorov's problem. Associated with this is the problem of conditions on a set of n numbers in order that these form a set of eigenvalues of an $(n \times n)$ stochastic matrix. After Suleĭmanova (1949), this work was carried forward by Perfect (1952, 1953, 1955) for stochastic matrices, and by Perfect & Mirsky (1965, and its references) for doubly stochastic matrices. Further work in this direction has been carried out by the de Oliveira (1968), whose book contains an extensive description and references.

[1] For a useful generalization of the notion of stochasticity and some results relevant to this section, see Haynsworth (1955). See also Barker and Turner (1973) and Fritz, Huppert and Willems (1979).

Vere-Jones (1971) has considered another version of this problem, viz. the investigation of the n-dimensional region consisting of points which are eigenvalue sets of those non-negative T which can be diagonalized by a similarity transformation $Q^{-1}TQ$ by a fixed matrix Q. (See also Buharaev (1968).)

We finally mention the articles of Mirsky (1963, 1964) as sources of further information.

Exercises on §2.5

2.23. Let T be an irreducible non-negative matrix with Perron–Frobenius eigenvalue r and corresponding left and right strictly positive eigenvectors v', w. By using the fact that the matrix defined by (2.11) is stochastic show that

$$S(\min w_i/\max w_i) \leq r \leq s(\max w_i/\min w_i)$$

with s and S having the meaning of (2.10a); and hence that

$$\min w_i/\max w_i \leq (s/S)^{1/2}.$$

Obtain corresponding inequalities in terms of v'.

2.24. Express the vector $\quad \boldsymbol{\delta} = (-9, 4, 4, 1)'$ in the form (2.12).

2.25. For irreducible $T = \{t_{ij}\}$, $i, j = 1, \ldots, n$, show (using (2.16)) that for any eigenvalue $\lambda \neq r$

$$|\lambda|/r \leq (M - m)/(M + m)$$

where $M = \max_{i,j} t_{ij}$, $m = \min_{i,j} t_{ij}$.

(Ostrowski, 1963)

2.26. Calculate τ_B and τ_1 for each of the matrices

$$\begin{bmatrix} 4 & 4 & 48 & 4 \\ 25 & 2 & 8 & 25 \\ 15 & 15 & 15 & 15 \\ 15 & 15 & 15 & 15 \end{bmatrix}, \quad \begin{bmatrix} 9 & 8 & 10 & 9 \\ 12 & 5 & 8 & 11 \\ 9 & 5 & 13 & 9 \\ 9 & 5 & 13 & 9 \end{bmatrix}$$

and verify that $\tau_B \geq \tau_1$.

2.27. Suppose $P = \{p_{ij}\}$, $i, j = 1, \ldots, n$ is any stochastic matrix. Show that $\tau_1(P) < 1$ if and only if no two rows of P are orthogonal (or, alternatively, any two rows intersect). [Stochastic matrices with this property are called "scrambling" and will be considered in Chapter 3 and in detail in Chapter 4.]

2.28. Any real $(n \times n)$ matrix $A = \{a_{ij}\}$ may be adjusted by several devices into an $(n + 1) \times (n + 1)$ matrix with equal row sums a, without changing the spectrum apart from adding on eigenvalue a, so Theorem 2.10 may be used to localize the *whole spectrum* of A.

Show that one such augmentation: consisting in adding a column whose entries are the negatives of the respective row sums of A, and a final row of zeroes (making zero the row sums, a, of the augmented matrix), produces the bound S (see (2.10b)) when $A \geq 0$. Suggest an augmentation which does the same for $A \geq 0$, but where the augmented matrix is still non-negative.

(Seneta, 1979)

2.29. Suppose $T = \{t_{ij}\}$, $i, j = 1, \ldots, n$ is a non-negative matrix with spectral radius $\rho > 0$. Denoting the corresponding right eigenvector by $w = \{w_i\}$ show that

$$t_{rj} \geq t_{kj}, \ j = 1, \ldots, n \Rightarrow w_r \geq w_k.$$

Now supposing T has its r-th row strictly positive, show that $\rho > 0$, and, putting $\theta_{rk} = \max_j (t_{kj}/t_{rj})$, that

$$w_r \theta_{rk} \geq w_k.$$

[*Hint:* Apply the preceding to the matrix $A = \{a_{ij}\}$ formed from T by multiplying its r-th row by θ_{rk} and dividing its r-th column by θ_{rk}, if $\theta_{rk} > 0$.]

(de Oliveira, 1972)

2.30. Denote by $s(T)$ and $S(T)$ the minimal and maximal row sums of a square non-negative matrix T with spectral radius ρ. If A and B are two such matrices show that $s(AB) \geq s(A)s(B)$, $S(AB) \leq S(A)S(B)$ and hence that

$$s(T) \leq \{s(T^2)\}^{1/2} \leq \{s(T^4)\}^{1/4} \leq \cdots \leq \rho \leq \cdots \leq \{S(T^4)\}^{1/4} \leq \{S(T^2)\}^{1/2} \leq S(T).$$

If T is irreducible with Perron-Frobenius eigenvalue r, show that as $p \to \infty$

$$\{s(T^{2^p})\}^{1/2^p} \uparrow r \downarrow \{S(T^{2^p})\}^{1/2^p}.$$

(e.g. Varga, 1962; Yamamoto, 1967)

2.6 Estimating Non-negative Matrices from Marginal Totals

An important practical problem, whose scope and variants are discussed in the notes to this section, may be stated as follows. Given an $(m \times n)$ matrix $A = \{a_{ij}\} \geq 0$ with no row or column zero, when does there exist a non-negative matrix $B = \{b_{ij}\}$ of the same dimensions with specified strictly positive row and column totals (say $B\mathbf{1} = x > \mathbf{0}$, $\mathbf{1}'B = y' > \mathbf{0}'$) and such that $a_{ij} = 0$ implies $b_{ij} = 0$? If such a B exists, how may it be obtained from A?

An intuitively straightforward approach to this problem is to rescale first the rows of a given A so that the vector of row sums is x, then to rescale the columns of the resulting matrix to give y' as the vector of column sums; and then to continue the iterative process by rescaling the row sums, and so on. Under favourable conditions on A, x and y', a matrix B of required form may be expected to result as the limit of this iterative scaling procedure. We shall pursue this approach to a limited extent without making our usual restriction to square matrices (though much work has been done in the setting of A square and irreducible), since in the applications squareness is not a standard requirement.

Taking an iteration to be a row adjustment followed by a column adjustment, we denote the result of $t (t \geq 1)$ iterations by $A^{(2t)} = \{a_{ij}^{(2t)}\}$ with

$$a_{ij}^{(2t)} = r_i^{(t)} a_{ij} c_j^{(t)}, \qquad i = 1, \dots, m; j = 1, \dots, n$$

where

$$c_j^{(t)} = \frac{y_j}{\sum_i r_i^{(t)} a_{ij}} \tag{2.17a}$$

$$r_i^{(t)} = \frac{x_i}{\sum_j a_{ij} c_j^{(t-1)}} \tag{2.17b}$$

where $c_j^{(0)} = 1$. When necessary for clarity we shall write $r_i^{(t)}(A)$ for $r_i^{(t)}$ and $c_j^{(t)}(A)$ for $c_j^{(t)}$; this extended notation makes possible the use of relations such as

$$r_i^{(t+1)}(A) = r_i^{(1)}(A^{(2t)}), \qquad \text{each } i. \tag{2.18}$$

It is also useful to write $A^{(2t-1)} = \{a_{ij}^{(2t-1)}\}$ where

$$a_{ij}^{(2t-1)} = r_i^{(t)} a_{ij} c_j^{(t-1)}$$

for the matrix resulting from t row adjustments and $t - 1$ column adjustments ($t \geq 1$).

Lemma 2.5. *For an A with no zero row or column,*

$$\max_i \left(\frac{r_i^{(t+1)}}{r_i^{(t)}} \right) \leq \max_j \left(\frac{c_j^{(t-1)}}{c_j^{(t)}} \right) \leq \max_i \left(\frac{r_i^{(t)}}{r_i^{(t-1)}} \right)$$

$$\min_i \left(\frac{r_i^{(t+1)}}{r_i^{(t)}} \right) \geq \min_j \left(\frac{c_j^{(t-1)}}{c_j^{(t)}} \right) \geq \min_i \left(\frac{r_i^{(t)}}{r_i^{(t-1)}} \right)$$

for $t \geq 2$; and for $t \geq 1$

$$\min_j \left(\frac{c_j^{(t)}}{c_j^{(t-1)}} \right) \leq \min_i \left(\frac{r_i^{(t)}}{r_i^{(t+1)}} \right) \leq \min_j \left(\frac{c_j^{(t+1)}}{c_j^{(t)}} \right)$$

$$\max_j \left(\frac{c_j^{(t)}}{c_j^{(t-1)}} \right) \geq \max_i \left(\frac{r_i^{(t)}}{r_i^{(t+1)}} \right) \geq \max_j \left(\frac{c_j^{(t+1)}}{c_j^{(t)}} \right).$$

PROOF. By (2.17b) with $t + 1$) in place of t

$$\frac{r_i^{(t)}}{r_i^{(t+1)}} = \sum_j \left(\frac{r_i^{(t)} a_{ij} c_j^{(t-1)}}{x_i} \right) \frac{c_j^{(t)}}{c_j^{(t-1)}}$$

so that, for all i,

$$\left.\begin{aligned} \min_j \left(\frac{c_j^{(t)}}{c_j^{(t-1)}} \right) &\leq \frac{r_i^{(t)}}{r_i^{(t+1)}} \leq \max_j \left(\frac{c_j^{(t)}}{c_j^{(t-1)}} \right) \\ \text{i.e.} \qquad & \\ \max_j \left(\frac{c_j^{(t-1)}}{c_j^{(t)}} \right) &\geq \frac{r_i^{(t+1)}}{r_i^{(t)}} \geq \min_j \left(\frac{c_j^{(t-1)}}{c_j^{(t)}} \right) \end{aligned}\right\} \tag{2.19}$$

since the matrix with (i, j) entry $(r_i^{(t)} a_{ij} c_j^{(t-1)}/x_i)$ is (row) stochastic by (2.17*b*), and so $r_i^{(t)}/r_i^{(t+1)}$ is a convex combination of the elements $c_j^{(t)}/c_j^{(t-1)}$, $j = 1, \ldots, n$.

On the other hand, from (2.17*a*) with $(t + 1)$ in place of t

$$\frac{c_j^{(t)}}{c_j^{(t+1)}} = \sum_i \left(\frac{r_i^{(t+1)}}{r_i^{(t)}}\right) \frac{r_i^{(t)} a_{ij} c_j^{(t)}}{y_i}$$

so

$$\left.\begin{aligned} \min_i \left(\frac{r_i^{(t+1)}}{r_i^{(t)}}\right) \le \frac{c_j^{(t)}}{c_j^{(t+1)}} \le \max_i \left(\frac{r_i^{(t+1)}}{r_i^{(t)}}\right) \\ \text{i.e.} \qquad \\ \max_i \left(\frac{r_i^{(t)}}{r_i^{(t+1)}}\right) \ge \frac{c_j^{(t+1)}}{c_j^{(t)}} \ge \min_i \left(\frac{r_i^{(t)}}{r_i^{(t+1)}}\right) \end{aligned}\right\} \tag{2.20}$$

from which the stated results follow. □

Corollary. *The sequences* $\max_i (r_i^{(t+1)}/r_i^{(t)})$, $\max_j (c_j^{(t)}/c_j^{(t-1)})$ *are non-increasing with* t, *while* $\min_i (r_i^{(t+1)}/r_i^{(t)})$, $\min_j (c_j^{(t)}/c_j^{(t-1)})$ *are non-decreasing.*

In view of the fact that the elements of $A^{(2t)} = \{a_{ij}^{(2t)}\}$ are non-negative and have constant column sum vector y', it follows that $A^{(2t)}$ has at least one limit point (every limit point must be an elementwise finite non-negative matrix with the same column sum vector y', and zeroes in at least the same positions as A). Let A^* denote any limit point.

Lemma 2.6. *If* $A > 0$, *then* $A^* > 0$.

PROOF. From (2.17*a*),

$$c_j^{(t)} = \frac{y_j}{\sum_i r_i^{(t)} a_{ij}} \le \frac{y_j}{(r_i^{(t)} a_j)} \le \frac{y_j}{(a \min_i r_i^{(t)})} \tag{2.21}$$

for all i, where $a = \min a_{ij} > 0$.

From (2.17*b*)

$$\begin{aligned} \sum_j r_i^{(t)} a_{ij} c_j^{(t)} &= \sum_j r_i^{(t)} a_{ij} c_j^{(t-1)} \left(\frac{c_j^{(t)}}{c_j^{(t-1)}}\right) \\ &\ge \min_j \left(\frac{c_j^{(t)}}{c_j^{(t-1)}}\right) \sum_j r_i^{(t)} a_{ij} c_j^{(t-1)} \\ &\ge \min_j \left(\frac{c_j^{(1)}}{c_j^{(0)}}\right) x_i \end{aligned}$$

by the Corollary to Lemma 2.5. Thus

$$r_i^{(t)} \geq \min_j \left(\frac{c_j^{(1)}}{c_j^{(0)}}\right) \frac{x_i}{(\alpha \sum_j c_j^{(t)})}, \qquad (\alpha = \max a_{ij});$$

$$\geq \min_j \left(\frac{c_j^{(1)}}{c_j^{(0)}}\right) \frac{x_i a(\min_i r_i^{(t)})}{(\alpha \sum_j y_j)}, \text{ by (2.21).}$$

Finally

$$c_j^{(t)} = \frac{y_j}{\sum_i r_i^{(t)} a_{ij}} \geq \frac{y_j}{(m\alpha \min_i r_i^{(t)})}.$$

Hence

$$a^{(2t)} = \min_{i,j} a_{ij}^{(2t)} = \min_{i,j} r_i^{(t)} a_{ij} c_j^{(t)}$$

$$> a^2 \min_j \left(\frac{c_j^{(1)}}{c_j^{(0)}}\right) \min_{i,j} \frac{(x_i y_j)}{(m\alpha^2 \sum_j y_j)}$$

$$> 0$$

so $a^{(2t)}$ is bounded away from zero. □

To give the flavour of the general theory, we shall prove subsequent results under a simplifying assumption of *connectedness* of A or A^*. Development of the theory only for the case $A > 0$ ($\Rightarrow A^* > 0$ by Lemma 2.6) obscures the difficulties otherwise present, just as the Perron–Frobenius theory for $T > 0$ obscures the difficulties which enter in the presence of zero elements.

Definition 2.6. An $(m \times n)$ matrix $A = \{a_{ij}\} \geq 0$ is said to be connected if it has no zero row or column, and for any proper non-empty subset I of $\{1, 2, \ldots, m\}$, $F(I) \cap F(I^c) \neq \phi$.

Here, as in §2.4, $F(I) = \{j: a_{ij} > 0, i \in I\}$; we shall write this $F_1(I)$ where confusion is possible. Also, ϕ denotes the empty set and I^c denotes the complement of I in $\{1, 2, \ldots, m\}$. We put $F(\phi) = \phi$. Note that $F(I) \cup F(I^c) = \{1, 2, \ldots, n\}$, since A has no zero row or column. The constraint of connectedness enables us to sharpen the result of Lemma 2.5. We state the result for row adjustments $r_i^{(t)}$ only, though a similar result obtains for column adjustments.

Lemma 2.7. *If $A \geq 0$ is $(m \times n)$ and connected, and for any $t' \geq 2$*

$$\min_i \left(\frac{r_i^{(t')}}{r_i^{(t'-1)}}\right) < \max_i \left(\frac{r_i^{(t')}}{r_i^{(t'-1)}}\right),$$

then

$$\max_i \left(\frac{r_i^{(t)}}{r_i^{(t-1)}} \right) < \max_i \left(\frac{r_i^{(t')}}{r_i^{(t'-1)}} \right)$$

for $t \geq t' + m - 1$.

PROOF. Put $I_0 = \{i;\ r_i^{(t')}/r_i^{(t'-1)} = \min_i\ (r_i^{(t')}/r_i^{(t'-1)})\}$. We have for $t > t'$ by Lemma 2.5

$$\min_j \left(\frac{c_j^{(t'-1)}}{c_j^{(t')}} \right) \geq \min_i \left(\frac{r_i^{(t')}}{r_i^{(t'-1)}} \right).$$

Define

$$J_0 = \left\{ j;\ \min_j \left(\frac{c_j^{(t'-1)}}{c_j^{(t')}} \right) = \min_i \left(\frac{r_i^{(t')}}{r_i^{(t'-1)}} \right) \right\}.$$

If $J_0 \neq \phi$, then since from (2.17*a*)

$$\frac{c_j^{(t'-1)}}{c_j^{(t')}} = \sum_i \left(\frac{r_i^{(t')}}{r_i^{(t'-1)}} \right) \frac{r_i^{(t'-1)} a_{ij} c_j^{(t'-1)}}{y_j}$$

it follows that

$$a_{ij} = 0, \qquad i \in I_0^c, \qquad j \in J_0$$

so $F(I_0^c) \subseteq J_0^c$.

Now again by Lemma 2.5

$$\min_i \left(\frac{r_i^{(t'+1)}}{r_i^{(t')}} \right) \geq \min_j \left(\frac{c_j^{(t'-1)}}{c_j^{(t')}} \right),$$

so define

$$I_1 = \left\{ i;\ \min_i \left(\frac{r_i^{(t'+1)}}{r_i^{(t')}} \right) = \min_j \left(\frac{c_j^{(t'-1)}}{c_j^{(t')}} \right) \right\}.$$

If $I_1 \neq \phi$, then as in the previous step

$$a_{ij} = 0, \qquad i \in I_1, \qquad j \in J_0^c$$

so $F(I_1) \subseteq J_0$. Thus $I_1 \subseteq I_0$. Since A is connected, it is not possible that $I_1 = I_0$, so I_1 is a proper subset of I_0. Suppose $I_1 \neq \phi$. If now

$$\min_i \left(\frac{r_i^{(t'+1)}}{r_i^{(t')}} \right) < \max_i \left(\frac{r_i^{(t'+1)}}{r_i^{(t')}} \right)$$

we can repeat the whole cycle, and continue in this manner. Since I_0 may contain at most $(m-1)$ indices, I_1 may contain at most $(m-2)$, so at worst the process will terminate during the $(m-1)$th repetition of the cycle, whence the conclusion. □

Lemma 2.8. *If $A = \{a_{ij}\} \geq 0$ is $(m \times n)$ and connected, and $B = \{b_{ij}\} \geq 0$ is of the same dimensions and satisfies $B\mathbf{1} = x$ and $\mathbf{1}'B = y'$ for some fixed $x > \mathbf{0}$ and $y > \mathbf{0}$, and is expressible in the form $B = RAC$ for some diagonal matrices $R = \text{diag } r_i$, $C = \text{diag } c_j$, then B is the unique matrix satisfying these conditions and expressible in this manner.*

PROOF. Since $B \geq 0$ has positive row sums, no $r_i, c_j, i = 1, \ldots, m; j = 1, \ldots, n$ may be zero. Suppose $r_i > 0$, $i \in I$ and $r_i < 0$, $i \in I^c$; and $c_j > 0, j \in J$ and $c_j < 0$, $j \in J^c$. Then $a_{ij} = 0$ for $i \in I, j \in J^c$ and $i \in I^c, j \in J$. Since A is connected, a contradiction results unless $I^c = \phi = J^c$ or $I = \phi = J$. Thus we may assume without loss of generality that R and C have positive diagonal.

Suppose $\bar{B} = \bar{R}A\bar{C}$ is another matrix with the properties of B. Then

$$\bar{B} = \bar{R}A\bar{C} = \bar{R}R^{-1}BC^{-1}\bar{C} = PBQ$$

say, where $P = \text{diag } p_i$ and $Q = \text{diag } q_j$ are diagonal matrices with positive diagonal. Since $x = \bar{B}\mathbf{1}$, $y' = \mathbf{1}'\bar{B}$

$$\sum_j p_i b_{ij} q_j = x_i, \qquad i = 1, \ldots, m \tag{2.22a}$$

$$\sum_i p_i b_{ij} q_j = y_j, \qquad j = 1, \ldots, n. \tag{2.22b}$$

Since the matrix $\{b_{ij}/x_i\}$ is row stochastic

$$p_i x_i \min q_j \leq \sum_j p_i b_{ij} q_j \leq x_i p_i \max q_j$$

so by (2.22a)

$$p_i \min q_j \leq 1 \leq p_i \max q_j, \qquad \text{each } i. \tag{2.23a}$$

Similarly, since $\{b_{ij}/y_j\}$ is column stochastic, from (2.22b)

$$q_j \min p_i \leq 1 \leq q_j \max p_i, \qquad \text{each } j. \tag{2.23b}$$

Using (2.23a) and (2.23b)

$$(\min p_i)(\max q_j) = (\max p_i)(\min q_j) = 1. \tag{2.24}$$

Now reorder the rows and columns of $\bar{B}$, B so that

$$p_1 \geq p_2 \geq \cdots \geq p_m; \qquad q_1 \leq q_2 \leq \cdots \leq q_n.$$

Suppose first that for some i_0, j_0, $p_{i_0} q_{j_0} > 1$ and $b_{i_0 j_0} > 0$. If $j_0 = 1$ then $q_{j_0} = \min q_i$ and $1 < p_{i_0} q_{j_0} = p_{i_0} \min q_j \leq \max p_i \min q_j = 1$ by (2.24) which is a contradiction. So suppose $j_0 > 1$. Then there must be a $j_1 < j_0$ such that $p_{i_0} q_{j_1} < 1$ and $b_{i_0 j_1} > 0$, otherwise we have a contradiction to $\sum_j b_{i_0 j} = x_{i_0}$. If $i_0 = 1$ we shall obtain a contradiction to (2.24) again; and if $i_0 > 1$ there is an $i_1 < i_0$ such that $p_{i_1} q_{j_1} > 1$ and $b_{i_1 j_1} > 0$, otherwise we have a contradiction to $\sum_i b_{ij_1} = y_{j_1}$. Continuing in this manner we obtain a sequence $j_0, i_0, j_1, i_1, j_2, i_2, \ldots$ where $j_0 > j_1 > j_2 \ldots$, $i_0 > i_1 > i_2 > \cdots$. Eventually one of the j_k's or one of the i_k's becomes 1. Suppose it is one of the

j_k's; then we have on the next step for some i^*, $1 < p_{i^*} q_1 \le \max_i p_i \min_j q_j$ which is a contradiction to (2.24). Similarly if one of the i_k's becomes 1 first.

A similar argument leading to a contradiction to (2.24) will hold if we suppose that for some $i_0, j_0, p_{i_0} q_{j_0} < 1$ for some $b_{i_0 j_0} > 0$.

Hence for any b_{ij} such that $b_{ij} > 0$, $p_i q_j = 1$. Hence

$$\bar{B} = B. \qquad \square$$

Lemma 2.9. *A limit point $A^* = \{a_{ij}^*\}$ of the sequence $A^{(2t)} = \{a_{ij}^{(2t)}\}$ which satisfies the conditions: $A^*\mathbf{1} = x$, $\mathbf{1}'A^* = y'$, A^* is connected, is the unique limit point with these properties.*

Proof. Since A^* is connected, A is connected. Consider the subsequence of $\{t\}$ through which A^* arises as the limit of $A^{(2t)}$, and let $(r, c) = (\{r_i\}, \{c_j\})$ (either vector of which may have infinite entries) be a limit point of $(r^{(t)}/r_1^{(t)}, c^{(t)}r_1^{(t)})$, where $r^{(t)} = \{r_i^{(t)}\}$, $c^{(t)} = \{c_j^{(t)}\}$, through the same subsequence. Then $r_1 = 1$. Let $I = \{i; r_i = \infty\}$. If I is non-empty, then since for each $i \in I$ there is a j such that $a_{ij} > 0$ (and all a_{ij}^* are finite), it follows that the set $J = \{j; c_j = 0\}$ is nonempty also. Also, I^c is non-empty, and A^* has no zero rows, so J^c is non-empty. It follows that $F_{A^*}(I) \subseteq J$, $F_{A^*}(I^c) \subseteq J^c$, so $F_{A^*}(I) \cap F_{A^*}(I^c) = \phi$, which contradicts the connectedness of A^*. $\square$

An analogous argument beginning with the supposition that the set $\{j; c_j = \infty\}$ is non-empty leads to a contradiction likewise.

Thus all the r_i and c_j are finite, and we may write $A^* = RAC$. Hence by Lemma 2.8 there can be no other limit point with the properties of A^*.

Theorem 2.11. *Suppose $x'\mathbf{1} = y'\mathbf{1}$ and every limit point of the sequence $A^{(2t)} = \{a_{ij}^{(2t)}\}$ is connected. Then $B = \lim_{k\to\infty} A^{(k)}$ exists, satisfies $B\mathbf{1} = x$, $\mathbf{1}'B = y'$, and is expressible in the form $B = RAC$ where R and C are diagonal matrices with positive diagonals [hence B has the same incidence matrix as A].*

Proof. Since $\mathbf{1}'A^{(2t)} = y'$, it follows, as already noted, that $\mathbf{1}'A^* = y'$ for any limit point A^* of $A^{(2t)}$.

Suppose first that for some limit point A^*

$$\min_i \left(\frac{r_i^{(2)}(A^*)}{r_i^{(1)}(A^*)}\right) = \max_i \left(\frac{r_i^{(2)}(A^*)}{r_i^{(1)}(A^*)}\right), \quad = \lambda \text{ say.} \tag{2.25}$$

Then $$r_i^{(2)}(A^*) = \lambda r_i^{(1)}(A^*), \qquad i = 1, \ldots, m.$$

Write $$B = \{b_{ij}\} \quad \text{where}$$

$$b_{ij} = r_i^{(1)}(A^*)a_{ij}^* c_j^{(1)}(A^*) = \lambda^{-1} r_i^{(2)}(A^*)a_{ij}^* c_j^{(1)}(A^*)$$

so that

$$\sum_i b_{ij} = y_j, \; \sum_j b_{ij} = \lambda^{-1} x_i, \quad \text{whence} \sum_j y_j = \lambda^{-1} \sum_i x_i,$$

and since by assumption $\sum_j y_j = \sum_i x_i$, it follows $\lambda = 1$. Now, for each i

$$1 = \frac{r_i^{(2)}(A^*)}{r_i^{(1)}(A^*)} = \lim_{t\to\infty} \left(\frac{r_i^{(2)}(A^{(2t)})}{r_i^{(1)}(A^{(2t)})}\right)$$

through an appropriate subsequence of $\{t\}$, by continuity. Hence for this subsequence

$$1 = \lim_{t\to\infty} \max_i \left(\frac{r_i^{(2)}(A^{(2t)})}{r_i^{(1)}(A^{(2t)})}\right)$$

$$= \lim_{t\to\infty} \max_i \left(\frac{r_i^{(t+2)}(A)}{r_i^{(t+1)}(A)}\right) \quad \text{by (2.18).}$$

Thus, by the Corollary to Lemma 2.5,

$$1 = \lim_{t\to\infty} \max_i \left(\frac{r_i^{(t+1)}(A)}{r_i^{(t)}(A)}\right)$$

through all values t, $t = 1, 2, \ldots$. Similarly, through all t values

$$1 = \lim_{t\to\infty} \min_i \left(\frac{r_i^{(t+1)}(A)}{r_i^{(t)}(A)}\right);$$

whence, for each i, through all t values

$$\left(\frac{r_i^{(t+1)}(A)}{r_i^{(t)}(A)}\right) \to 1. \tag{2.26}$$

Thus for $A^{(2t+1)} = \{a_{ij}^{(2t+1)}\}$, where, recall, $A^{(2t+1)}\mathbf{1} = x$, since

$$a_{ij}^{(2t+1)} = r_i^{(t+1)} a_{ij} c_j^{(t)} = \left(\frac{r_i^{(t+1)}}{r_i^{(t)}}\right) a_{ij}^{(2t)},$$

$$a_{ij}^{(2t+1)} \to a_{ij}^*$$

by (2.26), through an appropriate subsequence of $\{t\}$, for all i, j. Hence $A^*\mathbf{1} = x$; and we already know $\mathbf{1}'A^* = y'$ and that A^* is connected. Lemma 2.9 then shows that A^* is the unique limit point with these properties, and by the assumption of the present theorem that *every* limit point is connected and so has the properties of A^*,

$$A^* = \lim_{t\to\infty} A^{(2t)}.$$

Moreover, by the last statement in the proof of Lemma 2.9, $A^* = RAC$ for some diagonal R and C with positive diagonals. Finally, since B as defined in the course of the proof also has the properties of A^*, and is expressible in the form RAC for some diagonal R and C with positive diagonal, since

$$b_{ij} = r_i^{(1)}(A^*) a_{ij}^* c_j^{(1)}(A^*) = r_i^{(1)}(A^*) r_i a_{ij} c_j c_j^{(1)}(A^*)$$

it follows from Lemma 2.8 that $B = A^*$, which explains the notation of the theorem. Note that B has zeroes in at least the same positions as A.

There remains the important possibility that (2.25) does not hold, i.e. suppose

$$\min_i \left(\frac{r_i^{(2)}(A^*)}{r_i^{(1)}(A^*)}\right) < \max_i \left(\frac{r_i^{(2)}(A^*)}{r_i^{(1)}(A^*)}\right).$$

Then from Lemma 2.7, since A^* is connected,

$$0 < h = \max_i \left(\frac{r_i^{(2)}(A^*)}{r_i^{(1)}(A^*)}\right) - \max_i \left(\frac{r_i^{(m+1)}(A^*)}{r_i^{(m)}(A^*)}\right).$$

Now for t and t' sufficiently large and members of the appropriate subsequence

$$\begin{aligned} 0 < h/2 &< \max_i \left(\frac{r_i^{(2)}(A^{(2t)})}{r_i^{(1)}(A^{(2t)})}\right) - \max_i \left(\frac{r_i^{(m+1)}(A^{(2t')})}{r_i^{(m)}(A^{(2t')})}\right) \\ &= \max_i \left(\frac{r_i^{(2)}(A^{(2t)})}{r_i^{(1)}(A^{(2t)})}\right) - \max_i \left(\frac{r_i^{(2)}(A^{(2(t'+m-1))})}{r_i^{(1)}(A^{(2(t'+m-1))})}\right) \\ &\leq 0 \end{aligned}$$

if $t' + m - 1 \leq t$ by the Corollary to Lemma 2.5, which is a contradiction, since t' and t can certainly be chosen accordingly.

The fact that $A^{(2t-1)} \to B$ as $t \to \infty$ follows from (2.26). □

Corollary. *If $A > 0$ is $(m \times n)$ then [by Lemma 2.6] if $x'\mathbf{1} = y'\mathbf{1}$, then the conclusions of Theorem 2.11 hold.*

Bibliography and Discussion to §2.6

The approach and results of the section are generally those of Bacharach (1965; 1970, Chapter 4). Lemma 2.6 follows Sinkhorn (1964), who also gives a version of Lemma 2.8 in his setting where A is $(n \times n)$, $A > 0$, and B is doubly stochastic ($B\mathbf{1} = \mathbf{1}$, $\mathbf{1}'B = \mathbf{1}'$). The crucial averaging property of a row stochastic matrix $P = \{p_{ij}\}$ on a vector $w = \{w_j\}$ which it multiplies from the left:

$$\min_j w_j \leq \sum_j p_{ij} w_j \leq \max_j w_j$$

and its refinements are used repeatedly in Chapters 3 and 4. Bacharach's results are more extensive (we have given them in abbreviated form, to avoid extensive technicalities but to nevertheless cover the case $A > 0$). For example his main theorem (1965; Theorem 3) states that for an $(m \times n)$ $A \geq 0$ with no zero row or column:

> $A^{(2t)}$ converges to B which satisfies $B\mathbf{1} = x$, $\mathbf{1}'B = y'$ if and only if for any subsets I and J of $\{1, \ldots, m\}$ and $\{1, \ldots, n\}$ respectively
>
> $$a_{ij} = 0 \quad \text{for all} \quad i \in I^c, j \in J \Rightarrow \sum_{i \in I^c} x_i \leq \sum_{j \in J^c} y_j, \; \sum_{i \in I} x_i \geq \sum_{j \in J} y_j. \quad (2.27)$$

(a_{ij} is to be understood as zero if i or $j \in \phi$, as is summation over an empty set.) This, and its Corollary 3 given as our Exercise 2.34, answers the questions posed at the beginning of §2.6. For further discussion it is useful to note (in view of Exercise 2.34) that the theorem just stated may be restated in the following form:

> $A^{(2t)}$ converges to a matrix B which satisfies $B\mathbf{1} = x$, $\mathbf{1}'B = y'$ if and only if there exists a non-negative matrix $B = \{b_{ij}\}$ satisfying these two conditions and the condition $a_{ij} = 0 \Rightarrow b_{ij} = 0$.

We remark at this stage that connectedness in the present context plays much the same role as irreducibility in earlier contexts: it simplifies the development of the general theory.

There have been essentially two separate evolutionary lines for theory of the kind considered in this section. One has motivation and evolution in an economics context, culminating in the book of Bacharach (1970), which gives an extensive bibliography for this line. The other line has had its origins in the general area of mathematical statistics where the motivating papers have been those of Deming and Stephan (1940), Fréchet (work cited in his paper of 1960), and Sinkhorn (1964). Macgill (1977) gives a partial bibliographic survey covering both these main lines, which points out that aspects of the problem studied here may be traced back as far as Kruithof (1937) in the field of telecommunication. The earliest rigorous treatment of aspects of the problem appear to be within an unpublished note of Gorman (1963); see Bacharach (1970; §4.6). We shall give only a brief account of the various settings.

In empirical work in economics, if the elements of a non-negative matrix function of time are known at one time (call the $(m \times n)$ matrix A) but only its row and column sums ($x > \mathbf{0}$, $y' > \mathbf{0}'$ respectively) are known at the next time point, an intuitively simple estimate for the matrix at this time point is $B = \{b_{ij}\}$, where $b_{ij} = r_i a_{ij} c_j$ for some $r = \{r_i\} > \mathbf{0}$, $c = \{c_j\} > \mathbf{0}$ where $B\mathbf{1} = x$, $\mathbf{1}'B = y'$ (if such a B may be found). Bacharach calls such a matrix B *biproportional* to the matrix A, or a biproportional matrix adjustment of A, since its rows are proportional to the rows of A and its columns to the columns (it is often called the RAS adjustment in economics). It is clear that a necessary condition for the existence of a biproportional matrix adjustment B of A, is the existence of a matrix $B \geq 0$, with *zero elements in precisely* *the same positions as* A, and satisfying the further constraints $B\mathbf{1} = x$, $\mathbf{1}'B = y'$. Theorem 2.11 and its Corollary give: (1) sufficient conditions for a biproportional matrix adjustment B for A to exist; (2) provide an iterative procedure, giving $A^{(k)}$, $k = 1, 2, \ldots$ as an algorithm for arriving at B.

If the requirement of biproportionality of B is relaxed in that it is required to satisfy only: $B \geq 0$, $B\mathbf{1} = x$, $\mathbf{1}'B = y'$ and $a_{ij} = 0 \Rightarrow b_{ij} = 0$, then such an "adjustment" will sometimes exist when a biproportional matrix adjustment does not (see Exercise 2.32); and Bacharach's Theorem 3 in the guise of Exercise 2.34 gives necessary and sufficient conditions on A, x and y for this. The alternative statement of the theorem shows that such an adjustment may then be attained by the iterative scaling procedure.

A biproportional adjustment of A may also be thought of as the matrix B minimizing the entropy/information type-expression

$$2 \sum_{i,j} b_{ij} \ln \left(\frac{b_{ij}}{a_{ij}}\right) \tag{2.28}$$

subject to the constraints $B \geq 0$, $a_{ij} = 0 \Leftrightarrow b_{ij} = 0$, $B\mathbf{1} = x$, $\mathbf{1}'B = y'$ (see Macgill (1977) for detail and Bacharach (1970, §6.5) and Wilson (1970) for some interpretations). It is worth noting that if $\sum_{i,j} a_{ij} = \sum_{i,j} b_{ij} = x'\mathbf{1} + y'\mathbf{1}$, and B is close to A, a Taylor series approximation to the expression (2.27) is

$$\sum_{i,j} \frac{(b_{ij} - a_{ij})^2}{b_{ij}} \tag{2.29}$$

which has a formal analogy to the Pearson chi-square (χ^2) statistic for goodness-of-fit in a table with mn classes. Indeed, if a_{ij} is the recorded proportion of outcomes out of r independent trials falling into outcome class (i, j), out of mn mutually exclusive and exhaustive possible outcomes, so $\sum_{i,j} a_{ij} = 1$ and we wish to estimate the probabilities b_{ij}, $i = 1, \ldots, m, j = 1, \ldots, n (\sum_{i,j} b_{ij} = 1)$ subject to the constraints $a_{ij} = 0 \Leftrightarrow b_{ij} = 0$, $\sum_j b_{ij} = b_{i.} = x_i$, $\sum_i b_{ij} = b_{.j} = y_j$ fixed, then if a biproportional adjustment exists (it will in the case $A > 0$), it provides the required estimator matrix. Since in this setting, (2.29) is[1] the χ^2 goodness-of-fit statistic, this estimator will be approximately a "minimum chi-square statistic". In actual fact it will minimize (2.28), and so maximize this same expression when multiplied by $(-1/2)$, which is then known as the relative entropy of the bivariate probability distribution represented by the matrix B to the bivariate sample distribution represented by the matrix A (c.f. Kullback (1959); Akaike (1977)). In a statistical setting the problem was first considered by Deming and Stephan (1949). Statistical literature is cited by Fienberg (1970), who also shows convergence of $A^{(k)}$, $k = 1, 2, \ldots$ in this setting for $A > 0$, which had been

[1] Apart from a factor r.

suggested by Deming and Stephan. Fienberg points out the statistically relevant point regarding the scaling algorithm producing $A^{(k)}$, $k = 1, 2, \ldots$ that interaction structure of the original table A, as defined by the cross product ratios

$$a_{ij}a_{gh}/a_{ih}a_{gj} \qquad (i \neq g, j \neq h)$$

is preserved at each stage of the iterative processes, since

$$a_{ij}^{(2t)} = r_i^{(t)}a_{ij}c_j^{(t)}, \qquad a_{ij}^{(2t-1)} = r_i^{(t)}a_{ij}c_j^{(t-1)}.$$

Amongst the statistical problems studied by Fréchet (see Fréchet (1960), and Thionet (1963) for references) which pertain to our discussion is the following: given two random variables X, Y with respective sample spaces $(\alpha_1, \alpha_2, \ldots, \alpha_m$, all α_i distinct) and $(\beta_1, \ldots, \beta_n$, all β_j distinct), with prescribed marginal distributions $x, y (x > \mathbf{0}, y > \mathbf{0}, x'\mathbf{1} = y'\mathbf{1} = 1)$, is there a bivariate (joint) distribution of (X, Y) consistent with these marginal distributions? In this problem there is no insistence that certain positions of the matrix B, with specified row and column totals x, y' (with $x'\mathbf{1} = y'\mathbf{1} = 1$), sought, be zero. We may use the Corollary to Theorem 2.11 to deduce that a bivariate distribution which has every point (α_i, β_j) $i = 1, \ldots, m$; $j = 1, \ldots, n$ in its sample space (i.e. $B > 0$) will result from applying the iterative scaling algorithm to *any* $(m \times n)$ matrix $A > 0$. The non-uniqueness of solution of this problem is manifest, and demonstrated by Fréchet.

The motivation for the paper of Sinkhorn (1964) is from the estimation theory of Markov chains, and the paper is unconnected with earlier statistical manifestations of the problem of this section. If the transition matrix of a Markov chain is the (stochastic) matrix $P = \{p_{ij}\}$ $i, j = 1, \ldots, n$, then the usual estimate of p_{ij} for fixed i is a_{ij} where a_{ij} is the proportion of transitions to "state" $j, j = 1, \ldots, n$ out of all transitions beginning in state i. Thus $\sum_j a_{ij} = 1$, $i = 1, \ldots, n$. If it is in fact known that P is doubly stochastic (i.e. $P\mathbf{1} = \mathbf{1} = P'\mathbf{1}$), then the iterative scaling procedure $A^{(k)}$, $k = 1, 2, \ldots$ may be expected to produce it. Sinkhorn then proves Lemmas 2.6 and 2.8, and Theorem 2.11 in the setting when $A > 0$ is square and $x = y = \mathbf{1}$. This paper was the first of a sequence of papers by numerous authors. Although Sinkhorn (1967) extended his results to an $(m \times n)$ $A > 0$ and arbitrary $x, y > \mathbf{0}$, most work has focused simply, and using a variety of methods, on *one* mathematical aspect of the problem: For a given $A \geq 0$, $i, j = 1, \ldots, n$, when do there exist diagonal matrices R and C with positive diagonals such that RAC is doubly stochastic, and when are they unique (to scalar factors)? We mention only the papers of Brualdi, Parter and Schneider (1966), Menon (1967, 1968), Sinkhorn & Knopp (1967), and Djoković (1970), where further references may be found. See also Brualdi (1974) and Sinkhorn (1974). Thus in this context attention has been focussed rather on the existence of a biproportional-type matrix adjustment, rather than on the iterative scaling algorithm as a means of producing it.

Exercises on §2.6

2.31. Investigate the behaviour of $A^{(2t)}$, $A^{(2t-1)}$ as t increases for the case $x = y = \mathbf{1}$ where

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Note that the matrix is not connected, and the conclusions of Theorem 2.11 do not hold.

2.32. Show that

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, x = y = \mathbf{1} \Rightarrow A^{(k)} \to A^* = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

as $k \to \infty$ but the limit matrix A^* cannot be expressed in the form RAC. Note that A^* is not connected, and not all conclusions of Theorem 2.11 hold. Show directly that there is no matrix B satisfying $B\mathbf{1} = x$, $\mathbf{1}'B = y'$ and expressible in the form $B = RAC$.

2.33. Suppose $x'\mathbf{1} \neq y'\mathbf{1}$ where $x > \mathbf{0}$, $y > \mathbf{0}$ and for the matrix $A \geq 0$ with no zero row or column

$$\min_i \left(\frac{r_i^{(2)}(A)}{r_i^{(1)}(A)} \right) = \max_i \left(\frac{r_i^{(2)}(A)}{r_i^{(1)}(A)} \right).$$

Show that neither $A^{(2t)}$ nor $A^{(2t-1)}$ changes with t, but $A^{(2t)} \neq A^{(2t-1)}$. Construct an example of an A satisfying these conditions.

2.34. Given an $(m \times n)$ matrix $A = \{a_{ij}\} \geq 0$ with no zero row or column, show that there exists a non-negative matrix $B = \{b_{ij}\}$ of the same dimensions as A, such that $a_{ij} = 0 \Rightarrow b_{ij} = 0$, with $B\mathbf{1} = x > \mathbf{0}$, $\mathbf{1}'B = y' > \mathbf{0}'$ (with x and y fixed) if and only if (2.27) holds. [*Hint*: Use Bacharach's Theorem 3 stated above for the sufficiency.]

(Bacharach, 1965)

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,

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

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 $u_j \geq 0$, and j'' those for which $u_j < 0$, noting that $\sum_j u_j = 0$ (and bearing in mind that the set of j'''s will be empty only if $\boldsymbol{u} = \boldsymbol{0}$). Put

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

Then

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

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 $0 \leq \tau_1(P) \leq 1$), 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 $T \geq 0$ 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 $x, y > 0$, 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 $x = \lambda y$ for some positive $\lambda > 0$, that we may define a quantity $\tau_B(T)$ by

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

which must then satisfy

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

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

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

whence

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

$\tau_B(\cdot)$ 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

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

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 $T = wv' = \{w_i v_j\}$ where $v > 0$; $w \geq 0$, $\neq 0$, then from Exercise 3.1

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

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

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

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 $r \to \infty$

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

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

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

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*:

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

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

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) $\Rightarrow$ (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

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

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:

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

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 ratio $\tau_B(\cdot)$, is essentially more mathematically complex, since the explicit 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 $w^{(p)} = \{w_i^{(p)}\}$, where $w^{(p)} > \mathbf{0}$, $(w^{(p)})'\mathbf{1} = 1$.

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 $v^{(p)} = \{v_i^{(p)}\}$, where $v^{(p)} > \mathbf{0}$, $(v^{(p)})'\mathbf{1} = 1$. 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

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

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

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

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

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

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

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 $v_p' = \{v_j^{(p)}\}$ of (3.17) is independent of $p \geq 0$.*

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

$$x'T_{p,r}/x'T_{p,r}\mathbf{1} \to v_p',$$

whence

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

But $x'H_{p+1} \geq \mathbf{0}'$, $\neq \mathbf{0}'$ (since $x'H_{p+1}\mathbf{1} > 0$ from the row allowability of H_{p+1}) so the limit of the left-hand side is also v_{p+1}'. Hence all v_p have a common value, say v. Moreover $v \geq \mathbf{0}$, $v'\mathbf{1} = 1$, so 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 D) if there exists a probability vector D (i.e. $D \geq \mathbf{0}$, $D'\mathbf{1} = 1$) such that

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

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

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

and call it condition (C). □

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

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

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

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

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

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

and we may write

$$f_i' T_{p,r}/f_i' T_{p,r}\mathbf{1} = v' + E'(r, p, i)$$

where $E(r, p, i) \to \mathbf{0}$ 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 $r \to \infty$

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

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*

$$e_k \xrightarrow{k \to \infty} e \tag{3.19}$$

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

PROOF: Let us assume that the prior conditions and (3.19) hold. Then since by definition of e_k,

$$e_k' = e_k' H_k / e_k' H_k \mathbf{1}$$

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

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

so asymptotic homogeneity, with respect to the (necessarily probability) vector 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

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

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

Corollary. *Under the prior conditions of Theorem* 3.4, *if strong ergodicity with limit vector* $\boldsymbol{v}$ *holds, then* (3.19) *holds with* $e = \boldsymbol{v}$. [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

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

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 $e' > \mathbf{0}'$, 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 $e_k \to e > \mathbf{0}$, 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' ($> \mathbf{0}'$) of H. Hence $e_k \to e$. [This part has followed the proof of Theorem 3.4.] □

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. □

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 $T_{p,r} > 0$, $r \geq t$. □

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

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

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* $T = \{t_{ij}\}$

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

where

$$\phi(T) = \min_{i, j, k, l} \frac{t_{ik} t_{jl}}{t_{jk} t_{il}} \qquad \text{if } T > 0;$$

$$= 0 \qquad \text{if } T \not> 0.$$

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* $\operatorname{osc}(x/y) = \max(x/y) - \min(x/y)$, $x \in R_n$; $y \geq \mathbf{0}$, $\neq \mathbf{0}$ [*this is well-defined since* $\max(x, y) > -\infty$ *and* $\min(x, y) < \infty$], *then* $\infty \geq \operatorname{osc}(x/y) \geq 0$, *and* $\operatorname{osc}(x/y) = 0$ *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*) $\max[(x + cy)/y] = \max(x/y) + c$,
$\min[(x + cy)/y] = \min(x/y) + c$
$\operatorname{osc}[(x + cy)/y] = \operatorname{osc}(x/y)$,
$x \in R_n$; $y \geq \mathbf{0}$, $\neq \mathbf{0}$; $c \in R$.

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

(*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:
$\operatorname{osc}(x'A/y'A) = \max(x'A/y'A) - \min(x'A/y'A)$
$= \{\max([x'A - (\min(x'/y') - \varepsilon) \cdot y'A]/y'A) + \min(x'/y') - \varepsilon\}$
$- \{\min([x'A - (\min(x'/y') - \varepsilon) \cdot y'A]/y'A) + \min(x'/y') - \varepsilon\}$

by Lemma 3.11(*vi*);

$$= ((\operatorname{osc}(x'/y') + 2\varepsilon) \cdot \max([x'A - (\min(x'/y') - \varepsilon) \cdot y'A]/ (\operatorname{osc}(x'/y') + 2\varepsilon) \cdot y'A) - ((\operatorname{osc}(x'/y') + 2\varepsilon) \cdot \min([x'A - (\min(x'/y') - \varepsilon) \cdot y'A]/(\operatorname{osc}(x'/y') + 2\varepsilon)y'A)$$

by Lemma 3.11(*v*);

$$= (\operatorname{osc}(x'/y') + 2\varepsilon)\{\max(z_1'A/z_2'A)/(1 + \max(z_1'A/z_2'A)) - \min(z_1'A/z_2'A)/(1 + \min(z_1'A/z_2'A))\}$$

by Lemma 3.11(*ix*);

$$= (\operatorname{osc}(x'/y') + 2\varepsilon)\{\max(z_1'A/z_2'A)/(1 + \max(z_1'A/z_2'A)) - [1/(1 + \max(z_2'A/z_1'A))]\}$$

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 osc (x'/y') and 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 $\alpha > 0$,

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

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 $k, l = 1, \ldots, n$*

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

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

$$\sup_{x \geq \mathbf{0},\, \neq \mathbf{0}} \left(\frac{x'Tf_k}{x'Tf_l}\right) = \sup_{x \geq \mathbf{0},\, \neq \mathbf{0}} \left(\sum_{i=1}^n x_i f_i'Tf_k \Big/ \sum_{i=1}^n x_i f_i' T f_l\right).$$

Assume without loss of generality that

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

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

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

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

Corollary. *If $T > 0$,*

$$\sup_{x > \mathbf{0}} \left(\frac{x'Tf_k}{x'T\mathbf{f}_l}\right) = \max_i \left(\frac{f_i'Tf_k}{f_i'Tf_l}\right).$$

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

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

$$= \infty \qquad \text{if } T \ngtr 0,\ T \text{ allowable};$$

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

$$\frac{w'Tf_k}{z'Tf_k} \Big/ \frac{w'Tf_l}{z'Tf_l} = \frac{w'Tf_k \cdot z'Tf_l}{w'Tf_l \cdot z'Tf_k}$$

$$\leq \max_i \frac{f_i'Tf_k}{f_i'Tf_l} \cdot \max_j \frac{f_j'Tf_l}{f_j'Tf_k}$$

by the Corollary to Lemma 3.14;

$$= \max_i \left(\frac{t_{ik}}{t_{il}}\right) \cdot \max_j \left(\frac{t_{jl}}{t_{jk}}\right)$$

$$= \max_{i,j} \left(\frac{t_{ik}t_{jl}}{t_{il}t_{jk}}\right)$$

$$\leq \max_{i,j,k,l} \left(\frac{t_{ik}t_{jl}}{t_{il}t_{jk}}\right)$$

$$= \left(\frac{t_{i_0k_0}t_{j_0l_0}}{t_{i_0l_0}t_{j_0k_0}}\right)$$

for some i_0, j_0, k_0, l_0.

Hence

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

On the other hand

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

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

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

[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

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

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

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 osc $(x'/y') = 1$, and we need focus only on osc $(x'T/y'T)$:

$$\begin{aligned} \text{osc } (x'T/y'T) &\geq (x'Tf_k/y'Tf_k) - (x'Tf_l/y'Tf_l) \\ &= (y'_N Tf_k/y'T f_k) - (y'_N Tf_l/y'Tf_l) \end{aligned}$$

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 $N = M \cup \tilde{N}$, $M \cap \tilde{N} = \phi$

$$\begin{aligned} (y'_N Tf_k/y'_N Tf_l) &= [(y'_M Tf_k + y'_{\tilde{N}} Tf_k)/(y'_M Tf_l + y'_{\tilde{N}} Tf_l)] \\ &= [(\mathbf{1}'_M Tf_k + \eta \mathbf{1}'_{\tilde{N}} Tf_k)/(\mathbf{1}'_M Tf_l + \eta \mathbf{1}'_{\tilde{N}} Tf_l)]; \\ &> S - \delta \end{aligned}$$

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

$$\begin{aligned} t/\bar{t} &> (S-\delta)/(I+\delta) \\ &> \theta(T)(1-\varepsilon)^2. \end{aligned}$$

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,

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

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

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

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,*

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

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*);

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

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

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

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

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

since $P\mathbf{1} = \mathbf{1}$, by Lemma 3.11(v) and (vi), noting that for small $\delta > 0$, $x(\delta) = \delta\mathbf{1} + (1 - \delta)\mathbf{1}_S > \mathbf{0}$, $x(\delta) \neq \lambda\mathbf{1}$. 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) $z' \geq \mathbf{0}'$, $\neq \mathbf{0}' \Rightarrow z'A \geq \mathbf{0}'$, $\neq \mathbf{0}'$ 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 $x, y \geq \mathbf{0}$, $\neq \mathbf{0}$ etc.

3.19. Suppose $A = \{a_{ij}\} \geq 0$ 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 $\mathscr{S} = \{s_1, s_2, \cdots\}$ 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 $\mathscr{S}$, and having the following *probability property*: if x_0, $x_1, \ldots, x_{k+1}$ are elements of $\mathscr{S}$, 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

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

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 $\Pi_{i0} = P[X_0 = s_{i_0}]$

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 $\mathbf{\Pi}_0'$.

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

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

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 $\mathbf{\Pi}_k'$ 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

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

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

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

[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 $k > p$

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

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 v', normed so that $v'\mathbf{1} = 1$, be the corresponding positive left eigenvector. Then we have that

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

where v is the column vector of a probability distribution.

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

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

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

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

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

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

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

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. $\mathbf{\Pi}'_0 P^k \to \boldsymbol{v}'$, which is *independent* of $\mathbf{\Pi}_0$, 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

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

$$\leq \sum_{r \in I} p_{ir}^{(k)}.$$

Hence for $k \geq k_0(i)$ 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 $k \geq k_0$, 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

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

$$\leq \theta \sum_{r \in I} p_{ir}^{(mk)}$$

[1] See Exercise 4.11.

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

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

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

Theorem 4.4.

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

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 $m = \{m_i\}$.

Theorem 4.5.

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

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

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 $$m_i = \sum_{j \in I} m_{ij}. \qquad \square$$

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

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

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

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.

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

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

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

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

The general solution is of the form

$$\begin{aligned} x_i &= Az_1^i + Bz_2^i \qquad \text{if } z_1 \neq z_2 \\ &= (A + Bi)z^i \qquad \text{if } z_1 = z_2 = z \end{aligned}$$

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.

$$\begin{aligned} m_1 &= 1 \phantom{{}+qm_1} + pm_2 \\ m_2 &= 1 + qm_1 + pm_3 \\ &\vdots \\ m_{s-1} &= 1 + qm_{s-2}. \end{aligned}$$

Hence we can write in general

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

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 $v = (v'_1, \mathbf{0}')'$ where 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 $\boldsymbol{\Pi}_1 = v_1$. In particular 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 v_1' the stationary distribution corresponding to the primitive submatrix P_1 of P corresponding to the essential states. Let $v' = (v_1', \mathbf{0}')$ be an $1 \times n$ vector. Then as $k \to \infty$*

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

elementwise, where 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 $M = P_1 - \mathbf{1}v_1'$;

then $$M^i = P_1^i - \mathbf{1}v_1'.$$

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

$$R_{k+1} = \sum_{i=0}^{k} Q^{k-i} R \mathbf{1} v_1' + \sum_{i=0}^{k} Q^{k-i} R M^i$$

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 $k \to \infty$

$$\lim_{k \to \infty} R_{k+1} = \sum_{i=0}^{\infty} Q^i R \mathbf{1} v_1' = (I - Q)^{-1} R \mathbf{1} v_1'$$
$$= (I - Q)^{-1}(I - Q)\mathbf{1} v_1'$$
$$= \mathbf{1} v_1'$$

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 $v = \{v_j\}$ 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:

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

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

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

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

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

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

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 $\mathbf{\Pi} = \{\pi_i\}$ 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

$$\begin{aligned} \mathbf{\Pi} &= \frac{\{1, (0.3, 0.6, 0)(I - Q)^{-1}\}}{\{1 + (0.3, 0.6, 0)(I - Q)^{-1}\mathbf{1}\}} \\ &= (\tfrac{8}{35}, \tfrac{15}{35}, \tfrac{12}{35}, 0). \end{aligned}$$

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 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 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

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

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 v its unique stationary distribution vector. Let $R = \lim_{k \to \infty} P^{dk}$. Show that

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

[*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 $0 < t_{j,s}^{(p,r)} \leq 1$. 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 $t_{i,s}^{(p,r)}$ 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 $0 \leq \tau(P) \leq 1$, 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 ($v \geq \mathbf{0}$, $v'\mathbf{1} = 1$): that is, whenever P is stable.

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

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

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 $\{k_r\}$, $r \geq 1$, 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}, $r \geq 1$, 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 $r \geq 1$:

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

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

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

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]

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

[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)

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

so that

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

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

$$\begin{aligned} a(P^{(1)}P^{(2)} \cdots P^{(r)}) &\leq \tau_1(P^{(1)}) a(P^{(2)} \cdots P^{(r)}) \\ &\leq \tau_1(P^{(1)}) \tau_1(P^{(2)}) \cdots \tau_1(P^{(r)}) \tau_1(I) \end{aligned}$$

where I is the unit matrix; i.e.

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

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 $D = \{z'; z \geq \mathbf{0}, z'\mathbf{1} = 1\}$ the quantity, defined for any $(n \times n)$ stochastic matrix P by*

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

satisfies the properties

(*i*) *$\tau(P^{(1)}P^{(2)}) \leq \tau(P^{(1)})\tau(P^{(2)})$, $P^{(1)}$, $P^{(2)}$ 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 $x = f_i$, $y = f_j$, $i \neq j$, 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 $\tau(P) \leq 1$, 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

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

since any real-valued vector $\delta = \{\delta_i\}$ satisfying $\delta \neq \mathbf{0}$ $\delta'\mathbf{1} = 0$ may be written in the form $\delta = \text{const}\ (x - y)$ where x and y are probability vectors, $x \neq y$, 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* $P = \{p_{ij}\}$

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

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

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

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

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

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

suppose
$$\tau_1(P) = \frac{1}{2}\sum_s |p_{i_0 s} - p_{j_0 s}|, \qquad i_0 \neq j_0,$$

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 $\{P_k\}$, $k \geq 1$, of stochastic matrices if*

$$\sum_{k=1}^{\infty} \{1 - c(P_k)\} = \infty.$$

PROOF.

$$\tau_1(T_{p,r}) \leq \prod_{i=1}^{r} \tau_1(P_{p+i})$$

by Lemma 4.3;

$$\leq \prod_{i=1}^{r} c(P_{p+i})$$

by (4.6) and the assertion of the theorem is tantamount to

$$\prod_{i=1}^{\infty} c(P_i) = 0,$$

so

$$\tau_1(T_{p,r}) \to 0 \qquad \text{as } r \to \infty, \quad p \geq 0.$$

The conclusion follows from Lemma 4.1. □

Corollary 1.

$$\alpha(T_{p,r}) = \max_s \max_{i,j} |t_{i,s}^{(p,r)} - t_{j,s}^{(p,r)}| \leq \prod_{i=1}^{r} c(P_{p+i})$$

[This follows from (4.6), since $a(P) \leq \tau_1(P)$].

Corollary 2. *If $c(P_k) \leq c_0 < 1$, $k \geq 1$, (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 $p \geq 0$).*

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 $Q^kP \sim P$ and that P, like Q^kP, will have at least one column positive. □

Lemma 4.5. *If $T_{p,r} \in G_1$, $p \geq 0$, $r \geq 1$, then $T_{p,r} \in M$ for $r \geq t$ 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 $P_k \in G_2$, $k \geq 1$, then $T_{p,r} \in G_1$, $p \geq 0$, $r \geq 1$.

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, $k \geq 1$, 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 $(n_{k+1} \times n_{k+2})$, $k \geq 1$. 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 $\delta_j^* \geq 0$. Show that $\frac{1}{2}\Delta^* = \sum_k \delta_k(\sum_{j'} p_{kj'})$, 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

$$\tfrac{1}{2}\Delta^* \leq \tfrac{1}{2}\Delta \max_{i,h} \sum_{j'} (p_{ij'} - p_{hj'}) \leq \tfrac{1}{2}\Delta\tau_1(P),$$

so that $\Delta^* \leq \Delta c(P)$.

Use this last inequality to prove Theorem 4.9; and the inequali /

$$\Delta^* \leq \Delta\tau_1(P)$$

to prove (c.f. Lemma 4.3)

$$\tau_1(P^{(1)}P^{(2)}) \leq \tau_1(P^{(1)})\tau_1(P^{(2)}).$$

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 $T_{p,r} \in G_1$, $p \geq 0$, $r \geq 1$, and introduce the notation $\gamma_h = \min_{i,j}^{+} p_{ij}(h)$. Show that

$$a(T_{p,r}) \leq \prod_{i=0}^{k-1} c(T_{p+it,t}) \leq \prod_{i=0}^{k-1} \left\{1 - \prod_{j=1}^{t} \gamma_{p+it+j}\right\},$$

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*

$$e_k \to e \qquad (k \to \infty) \tag{4.10}$$

where e_k is the unique stationary distribution vector corresponding to P_k, and e is a limit vector. In the event that either (equivalent) condition holds, $\boldsymbol{D} = e$.

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 v holds, then (4.10) holds with $e = v$.* □

Theorem 4.13. *Assume all P_k, $k \geq 1$, contain a single essential class of indices and satisfy condition (C); and*

$$\tau_1(T_{p,r}) \leq \beta < 1 \qquad (4.11)$$

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 $e_k \xrightarrow{k\to\infty} e$ for a sequence e_k, $k \geq 1$, of stationary distribution vectors of the sequence of stochastic matrices P_k, $k \geq 1$, for some limit vector 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 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 $p \geq 0$

$$\tau_1(T_{p,r}) = \tau_1(T_{p,r-j_0} T_{p+r-j_0,j_0}) \leq \tau_1(T_{p+r-j_0,j_0}) \qquad (4.12)$$

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 e being the unique stationary distribution of P, and also—by Lemma 4.9—the limiting distribution 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 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 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*: $\gamma \mathcal{I}_k \leq P_k \leq \mathbf{1}\mathbf{1}'$ where $\gamma = 1 - c_0$ and $\mathcal{I}_k$ is one of the matrices $\mathbf{1}f'_j$, $j = 1, 2, \ldots, n$.)

(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 $F_0' = (F_1^0, F_2^0, \ldots, F_n^0)$. 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

$$a(U_{p,r}) \leq (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r, \qquad r \geq t,$$

so that for all i, j, s, p, from the definition of $a(\cdot)$

$$|u_{i,s}^{(p,r)} - u_{j,s}^{(p,r)}| \leq (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r, \qquad r \geq t.$$

Proceeding as in the proof of Theorem 4.17, for all $i, s, h = 1, \ldots, n$, $p \geq 0$, $r \geq t$, $k \geq 0$

$$|u_{i,s}^{(p,r+k)} - u_{h,s}^{(p,r)}| \leq (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r, \qquad r \geq t,$$

and letting $k \to \infty$ yields

$$|v_s^{(p)} - u_{h,s}^{(p,r)}| \leq (1-\gamma^t)^{-1}\{(1-\gamma^t)^{1/t}\}^r, \qquad r \geq t,$$

where

$$v_s^{(p)} = \lim_{r\to\infty} u_{i,s}^{(p,r)}, \qquad i = 1, \ldots, n. \qquad \square$$

The following result also gives a condition equivalent to ergodicity for backwards products.

Theorem 4.20. *Backwards products* $U_{p,r}$, $p \geq 0$, $r \geq 1$, *formed from a sequence* P_k, $k \geq 1$, *of stochastic matrices are ergodic if and only if there is only one set of absolute probability vectors* $\boldsymbol{v}_k$, $k \geq 0$, *in which case*

$$U_{p,r} \xrightarrow{r\to\infty} \mathbf{1}\boldsymbol{v}_p', \qquad p \geq 0.$$

Proof. We have for $p \geq 0$, $r \geq 1$, $h \geq 1$

$$U_{p+r,h}U_{p,r} = U_{p,r+h}. \tag{4.15}$$

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 $x \geq 0$, for stochastic matrices V_x, $x \geq 0$. Hence substituting $s_i - p - r$ for h in (4.15) and letting $i \to \infty$ we obtain

$$V_{p+r}U_{p,r} = V_p, \qquad p \geq 0, \quad r \geq 1,$$

so there is always at least one set of absolute probability vectors given by $f_j'V_k$, $k \geq 0$, for any particular fixed $j = 1, \ldots, n$ (i.e. we use the jth row of each V_k, $k \geq 0$).

Now suppose strong ergodicity holds, so that

$$U_{p,r} \xrightarrow{r\to\infty} \mathbf{1}v'_p, \qquad p \geq 0,$$

and suppose $\bar{v}_k$, $k \geq 0$, is any set of absolute probability vectors. Then

$$\bar{v}'_p = \bar{v}'_{p+r} U_{p,r} = \bar{v}'_{p+r}(\mathbf{1}v'_p + E_{p,r}) = v'_p + \bar{v}'_{p+r} E_{p,r}$$

where $E_{p,r} \to 0$ as $r \to \infty$, so, since $\bar{v}'_{p+r}$ is bounded, being a probability vector, $\bar{v}_p = v_p$, $p \geq 0$.

Conversely, suppose there is precisely one set of absolute probability vectors v_k, $k \geq 0$. Then suppose U_q^* is a limit point of $U_{q,r}$ as $r \to \infty$ for fixed but arbitrary $q \geq 0$, so that for some subsequence r_j, $j \geq 1$, of the integers, as $j \to \infty$,

$$U_{q,r_j} \to U_p^*. \tag{4.17}$$

Now use the sequence r_{j+q}, $j \geq 1$, from which to ultimately select the subsequence s_i giving (4.16). Following through the earlier argument, we have $V_x = \mathbf{1}v'_x$, $x \geq 0$, 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),

$$= \mathbf{1}v'_q.$$

Hence for a fixed $q \geq 0$, the limit point is unique, so as $r \to \infty$

$$U_{q,r} \to \mathbf{1}v'_q, \qquad q \geq 0. \qquad \square$$

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, $k \geq 1$, 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

$$\boldsymbol{F}_k = P_k \boldsymbol{F}_{k-1}, \qquad k \geq 1,$$

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.

CHAPTER 5
Countable Stochastic Matrices

We initiate our brief study of non-negative matrices with countable index set by a study of stochastic matrices for two main reasons. Firstly the theory which can be developed with the extra stochasticity assumption provides a foundation whose analytical ideas may readily be generalized to countable matrices which are not necessarily stochastic; and this will be our approach in the next chapter. Secondly the theory of countable stochastic matrices is of interest in its own right in connection with the theory of Markov chains on a countable state space; and indeed it is from this framework that the analytical development of our ideas comes, although we shall avoid probabilistic notions apart from the occasional aside. We shall not deal with inhomogeneous situations at all in the interests of brevity, since the infinite matrix theory is as yet of lesser utility than that for finite matrices.

We adopt the obvious notation: we deal with a square (non-negative) stochastic matrix $P = \{p_{ij}\}$ with a countable (i.e. finite or denumerably infinite index set $\{1, 2, \ldots\}$. We recall from §4.1 that the powers $P^k = \{p_{ij}^{(k)}\}$, $k \geq 1$, ($P^0 = I$) are well defined by the natural extension of the rule of matrix multiplication, and are themselves stochastic. [We note in passing that for an arbitrary non-negative matrix T with denumerably infinite index set, some entries of its powers T^k, $k \geq 2$ may be infinite, so that the powers are not well defined in the same sense.]

5.1 Classification of Indices

Previous Classification Theory

Much of the theory developed in §1.2 for general finite non-negative matrices goes through in the present case.

In particular the notions of one index leading to another, two indices communicating, and the consequent definition of *essential and inessential* indices, and hence classes, remain valid, as also does the notion of *period* of an index, and hence of a self-communicating class containing the index; and so the notion of *irreducibility* of a matrix P and its index set (see §1.3). A *primitive* matrix P may then be *defined* as a matrix corresponding *to an irreducible aperiodic index set*. Moreover, as before, all these notions depend only on the location of the positive elements in P and not on their size.

In actual fact the only notions of §1.2 which do not necessarily hold in the present context are those concerned primarily with pictorial representation, such as the *path diagram*, and the *canonical form*, since we have to deal now with an index set possibly infinite. Nevertheless, things tend to "work" in the same way: and consequently it is worthwhile to keep in mind even these notions as an aid to the classification and power-behaviour theory, even though e.g. it may be impossible as regards the canonical form representation to write two infinite sets of indices (corresponding to two self-communicating classes) as one following the other in sequential order.

Of the *results* proved in §§1.2–1.3, viz. Lemma 1.1, Lemma 1.2, Lemma 1.3, Theorem 1.3, Lemma 1.4 and Theorem 1.4, only Lemma 1.1 does not extend to the present context (even Theorem 1.4 remains partly, though no longer fundamentally, relevant). To see that an infinite stochastic matrix with at least one positive entry in each row does not necessarily possess at least one essential class of indices, it is only necessary to consider the example

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \\ & \vdots & & & \ddots & \end{bmatrix}$$

i.e.

$$\left.\begin{aligned} p_{ij} &= 1 \quad \text{for } j = i + 1 \\ &= 0 \quad \text{otherwise} \end{aligned}\right\} i, j \in \{1, 2, \ldots\}.$$

EXAMPLES. (1)[1] Suppose P has the incidence matrix

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & \\ 1 & 0 & 1 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 1 & 0 & \\ & \vdots & & & \ddots & \end{bmatrix}.$$

Then P is irreducible, since $i \to i + 1$ for each i so that $1 \to i$; also $i \to 1$ for each i. Moreover $p_{11} > 0$, so that index 1 is aperiodic. Hence the entire index set $\{1, 2, \ldots\}$ is aperiodic. Consequently P is primitive.

[1] Relevant to Example (4) of §4.1 if all f_i there are *positive*.

(2) Suppose P has incidence matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & & \\ 1 & 0 & 1 & 0 & 0 & \cdots & \\ 0 & 1 & 0 & 1 & 0 & & \\ 0 & & & & & & \\ \vdots & & \ddots & & \ddots & & \end{bmatrix}.$$

Then (i) index 1 itself forms a single essential aperiodic class. (ii) Each $j \in \{2, 3, \ldots\}$ is inessential, since $j \to j - 1$, so that $j \to 1$; however $\{2, 3, \ldots\}$ is a non-essential self communicating class, since also $j \to j + 1$. Further the subset has period 2, since clearly for each index $j \in \{2, 3, \ldots\}$ "passage" is possible only to each of its adjacent indices, so "return" to any index j can occur at (all) even k in $p_{jj}^{(k)}$, but not at any odd k.

(3) In the counterexample to Lemma 1.1 (in this context) given above, each index is inessential, and each forms a single non-self-communicating inessential class.

New Classification Theory

Although the above, previously developed, classification theory is of considerable value even in connection with possibly infinite P, it is not adequate to cope with problems of possibly extending Perron–Frobenius theory even to some extent to the present context, and so to deal with the problem of asymptotic behaviour as $k \to \infty$ of $P^k = \{p_{ij}^{(k)}\}$, a fundamental problem insofar as this book is concerned.

It is therefore necessary to introduce a rather more sensitive classification of indices which is used in conjunction with the previous one: specifically, we classify each index as recurrent or transient, the recurrent classification itself being subdivided into positive- and null-recurrent. To introduce these notions, we need additional concepts. Write

$$l_{ij}^{(1)} = p_{ij}; \qquad l_{ij}^{(k+1)} = \sum_{\substack{r \\ r \neq i}} l_{ir}^{(k)} p_{rj}, \qquad k \geq 1 \tag{5.1}$$

(where $l_{ij}^{(0)} = 0$, by definition, for all $i, j \in \{1, 2, \ldots\}$). [In the MC framework, $l_{ij}^{(k)}$ is the probability of going from i to j in k steps, without revisiting i in the meantime; it is an example of a "taboo" probability.]

It is readily seen that $l_{ij}^{(k)} \leq p_{ij}^{(k)} (\leq 1)$ by the very definitions of $l_{ij}^{(k)}$, $p_{ij}^{(k)}$.

Thus for $|z| < 1$, the generating functions

$$L_{ij}(z) = \sum_k l_{ij}^{(k)} z^k, \qquad P_{ij}(z) = \sum_k p_{ij}^{(k)} z^k$$

are well defined for all $i, j = 1, 2, \ldots$

Lemma 5.1. *For* $|z| < 1$:

$$P_{ii}(z) - 1 = P_{ii}(z)L_{ii}(z); \qquad P_{ij}(z) = P_{ii}(z)L_{ij}(z), \qquad i \neq j.$$

Further, $|L_{ii}(z)| < 1$, *so that we may write*

$$P_{ii}(z) = [1 - L_{ii}(z)]^{-1}.$$

PROOF. We shall first prove by induction that for $k \geq 1$

$$p_{ij}^{(k)} = \sum_{t=0}^{k} p_{ii}^{(t)} l_{ij}^{(k-t)}. \tag{5.2}$$[1]

The proposition is clearly true for $k = 1$, by virtue of the definitions of $l_{ij}^{(0)}$, $l_{ij}^{(1)}$.

Assume it is true for $k \geq 1$.

$$p_{ij}^{(k+1)} = \sum_r p_{ir}^{(k)} p_{rj}$$

$$= \sum_r \left[\sum_{t=0}^{k} p_{ii}^{(t)} l_{ir}^{(k-t)}\right] p_{rj}$$

by the induction hypothesis:

$$= \sum_{t=0}^{k} \left\{\sum_{r \neq i} l_{ir}^{(k-t)} p_{rj} + l_{ii}^{(k-t)} p_{ij}\right\} p_{ii}^{(t)}$$

$$= \sum_{t=0}^{k-1} \{l_{ij}^{(k-t+1)}\} p_{ii}^{(t)} + \left\{\sum_{t=0}^{k} p_{ii}^{(t)} l_{ii}^{(k-t)}\right\} p_{ij},$$

the first part following from the definition of the $l_{ij}^{(k)}$, in (5.1):

$$= \sum_{t=0}^{k-1} p_{ii}^{(t)} l_{ij}^{(k+1-t)} + p_{ii}^{(k)} p_{ij}$$

by the induction hypothesis;

$$= \sum_{t=0}^{k} p_{ii}^{(t)} l_{ij}^{(k+1-t)}$$

since $p_{ij} = l_{ij}$;

$$= \sum_{t=0}^{k+1} p_{ii}^{(t)} l_{ij}^{(k+1-t)}$$

since $l_{ij}^{(0)} = 0$. This completes the induction.

The first set of relations between the generating functions now follows from the convolution structure of the relation just proved, if one bears in mind that $p_{ij}^{(0)} = \delta_{ij}$, by convention.

[1] This relation is a "last exit" decomposition in relation to passage from i to j in an MC context.

To prove the second part note that since $p_{ii}^{(0)} = 1$, for real z,

$$0 \le z < 1, \qquad 1 \le P_{ii}(z) < \infty.$$

Hence, considering such *real* z, we have

$$1 > 1 - \frac{1}{P_{ii}(z)} = L_{ii}(z).$$

Hence letting $z \to 1-$, $1 \ge L_{ii}(1-) = \sum_{k=0}^{\infty} l_{ii}^{(k)}$.
Thus for complex z satisfying $|z| < 1$,

$$|L_{ii}(z)| \le L_{ii}(|z|) < 1. \qquad \square$$

Let us now define a (possibly infinite) quantity μ_i[1] by the limiting derivative

$$\mu_i = L'_{ii}(1-) \equiv \sum_{k=1}^{\infty} k l_{ii}^{(k)} \le \infty,$$

for each index i.

Definition 5.1. An index i is called *recurrent* if $L_{ii}(1-) = 1$, and *transient* if $L_{ii}(1-) < 1$.

A *recurrent* index i is said to be *positive- or null-recurrent* depending as $\mu_i < \infty$ or $\mu_i = \infty$ respectively. We call such μ_i the mean recurrence measure of a recurrent index i.

The following lemma and its corollary now establish a measure of relation between the old and new terminologies.

Lemma 5.2. *An inessential index is transient.*

PROOF. Let i be an inessential index; if i is not a member of a self-communicating class, i.e. $i \not\to i$ then it follows from the definition of the $l_{ij}^{(k)}$ that $l_{ii}^{(k)} = 0$ all k, so that $L_{ii}(1) = 0 < 1$ as required.

Suppose now i is essential and a member of a self-communicating class, I. Then clearly (since $i \to i$)$l_{ii}^{(k)} > 0$ for some $k \ge 1$. This follows from the definition of the $l_{ij}^{(k)}$'s once more; as does the fact that such an $l_{ii}^{(k)}$ must consist of all non-zero summands of the form

$$p_{ir_1} p_{r_1 r_2} \cdots p_{r_{k-1} i}$$

where $r_j \ne i$, $j = 1, 2, \ldots, k-1$ if $k > 1$, $r_j \to i$; and simply of

$$p_{ii}$$

if $k = 1$. A non-zero summand of this form cannot involve an index u such that $i \to u$, $u \not\to i$.

[1] The "mean recurrence time" of state i in the MC context.

Now in I there must be an index i' such that there exists a $j \notin I$ satisfying $i \to j, j \nrightarrow i$, and $p_{i'j} > 0$, since i is inessential. It follows that there will be an index $q \in I$ such that

$$p_{i'q} > 0.$$

It follows moreover that this element $p_{i'q}$ will be one of the factors, for some k, of one of the non-zero summands mentioned above.

Now consider what happens if the matrix P is replaced by a new matrix $\hat{P} = \{\hat{p}_{rs}\}$ by altering the i'th row (only) of P, by way of putting $\hat{p}_{i'j} = 0$ for the i', j mentioned above, and scaling the other non-zero entries so that they still sum to unity. It follows in particular that $\hat{p}_{i'q} > p_{i'q}$.

It is then easily seen that for some $k \geq 1$

$$\hat{l}_{ii}^{(k)} > l_{ii}^{(k)} > 0$$

and that consequently

$$(1 \geq)\hat{L}_{ii}(1-) > L_{ii}(1-)$$

which yields the required result.[1] □

Corollary. *A recurrent index is essential.*

The following lemma provides an alternative criterion for distinguishing between the recurrence and transience of an index:

Lemma 5.3. *An index j is transient if and only if*

$$\sum_{k=0}^{\infty} p_{jj}^{(k)} < \infty.$$

PROOF. From Lemma 5.1, for real s, $0 \leq s < 1$

$$P_{jj}(s) = [1 - L_{jj}(s)]^{-1}$$

Letting $s \to 1-$ yields $P_{jj}(1-) \equiv \sum_{k=0}^{\infty} p_{jj}^{(k)} < \infty$ if and only if $L_{jj}(1-) < \infty$. □

Corollary. *An index j is recurrent if and only if*

$$\sum_{k=0}^{\infty} p_{jj}^{(k)} = \infty.$$

In fact it is also true that if j is transient, then

$$\sum_{k=0}^{\infty} p_{ji}^{(k)} < \infty$$

[1] Which is, incidentally, trivial to prove by probabilistic, rather than the present analytical, reasoning.

for every index i such that $i \to j$, so that $p_{ji}^{(k)} \to 0$ as $k \to \infty$; but before we can prove this, we need to establish further relations between the generating functions $P_{ij}(z)$, and between the generating functions $L_{ij}(z)$.

Lemma 5.4. *For* $|z| < 1$, *and all* $i, j = 1, 2, \ldots$

$$P_{ij}(z) = z \sum_r P_{ir}(z)p_{rj} + \delta_{ij}$$

$$L_{ij}(z) = z \sum_r L_{ir}(z)p_{rj} + zp_{ij}(1 - L_{ii}(z)).$$

PROOF. The first relation follows directly from an elementwise consideration of the relation $P^{k+1} = P^kP$; and the second from substitution for the $P_{ij}(z)$ in terms of the $L_{ij}(z)$ via Lemma 5.1. □

Corollary 1. *For each pair of indices* i, j, *such that* $i \to j$, $L_{ji}(1-) < \infty$.

PROOF. For s such that $0 < s < 1$, and all i, j,

$$s \sum_r L_{ir}(s)p_{rj} \leq L_{ij}(s)$$

from the second of the relations in Lemma 5.4 (noting $L_{ii}(s) < 1$). Iterating the above inequality k times,

$$s^k \sum_r L_{ir}(s)p_{rj}^{(k)} \leq L_{ij}(s), \qquad \text{all } i, j, k \geq 1;$$

in particular

$$s^k L_{ir}(s)p_{rj}^{(k)} \leq L_{ij}(s), \qquad \text{all } i, j, r, k \geq 1.$$

Thus, with the appropriate substitutions

$$s^k L_{ji}(s)p_{ij}^{(k)} \leq L_{jj}(s). \tag{5.3}$$

Now if $i \to j$, but $j \not\to i$, the assertion of the Corollary 1 is trivial, so suppose that $i \to j$ and $j \to i$. Then there is a k such that $p_{ij}^{(k)} > 0$; and also from its definition, $L_{ji}(s) > 0$. Letting $s \to 1-$ in (5.3) yields $L_{ji}(1-) < \infty$ since $L_{jj}(1-) \leq 1$. Thus $0 < L_{ji}(1-) < \infty$. Thus in this case the Corollary is also valid. □

Corollary 2. *If* j *is transient, and* $i \to j$,

$$\sum_{k=0}^{\infty} p_{ji}^{(k)} < \infty$$

(*and so* $p_{ji}^{(k)} \to 0$ *as* $k \to \infty$). *If* j *is recurrent, and* $i \leftrightarrow j$,

$$\sum_{k=0}^{\infty} p_{ji}^{(k)} = \infty.$$

Proof. For $i = j$, this is covered by Lemma 5.3 and its Corollary. For $0 < s < 1$, and $i \neq j$, from Lemma 5.1

$$P_{ji}(s) = P_{ji}(s)L_{ji}(s).$$

Let $s \to 1-$.

Consider j transient first: then $P_{jj}(1-) < \infty$ by Lemma 5.3, and $L_{ji}(1-) < \infty$ by Corollary 1 above, so the result follows.

If j is recurrent, $P_{jj}(1-) = \infty$, and since $i \leftrightarrow j$, $0 < L_{ji}(1-)$, $< \infty$; so that $P_{ji}(1-) = \infty$ as required. □

The problem of limiting behaviour for the $p_{ji}^{(k)}$ as $k \to \infty$ for recurrent indices is a deeper one, and we treat it in a separate section.

5.2 Limiting Behaviour for Recurrent Indices

Theorem 5.1. *Let i be a recurrent aperiodic*[1] *index, with mean recurrence measure $\mu_i \leq \infty$. Then as $k \to \infty$*

$$p_{ii}^{(k)} \to \mu_i^{-1}.$$

We shall actually prove a rather more general analytical form of the theorem.

Theorem 5.1′. *Let $f_0 = 0, f_j \geq 0, j \geq 1$, with*

$$\sum_{j=0}^{\infty} f_j = 1.$$

Assume that the g.c.d. *of those j for which $f_j > 0$, is unity*[2]*, and let the sequence $\{u_k\}$, $k \geq 0$, be defined by*

$$u_0 = 1; \qquad u_k = \sum_{j=0}^{k} f_j u_{k-j}, \; k \geq 1.$$

It follows (e.g. by induction) that $0 \leq u_k \leq 1$, $k \geq 0$. Then as $k \to \infty$

$$u_k \to \frac{1}{\sum_{j=1}^{\infty} jf_j}.$$

[Theorem 5.1 now follows by putting $l_{ii}^{(j)} = f_j$, $p_{ii}^{(k)} = u_k$ for $j, k \geq 0$, in view of the basic relation

$$p_{ii}^{(k)} = \sum_{j=0}^{k} p_{ii}^{(j)} l_{ii}^{(k-j)} \equiv \sum_{j=0}^{k} l_{ii}^{(j)} p_{ii}^{(k-j)}, \; k \geq 1$$

[1] See Exercises 5.1 and 5.2 for an indication of the case of a periodic index.

[2] See Appendix A.

deduced in the proof of Lemma 5.1, in conjunction with our convention $p_{ii}^{(0)} = 1$.]

Proof. (*i*) We first note from Appendix A, that of those j for which $f_j > 0$, there is a finite subset $\{j_1, j_2, \ldots, j_r\}$ such that 1 is its g.c.d., and by the Corollary to Lemma A.3 any integer $q \geq N_0$, for some N_0 is expressible in the form

$$q = \sum_{t=1}^{r} p_t j_t,$$

with the p_t non-negative integers, depending on the value of q.

(*ii*) Put $r_t = \sum_{j=t}^{\infty} f_{j+1}$, noting $r_0 = 1$; then for $k \geq 1$,

$$r_0 u_k = \sum_{j=0}^{k} f_j u_{k-j} = \sum_{j=0}^{k-1} (r_j - r_{j+1}) u_{k-j-1},$$

$$r_0 u_k + \sum_{j=0}^{k-1} r_{j+1} u_{k-(j+1)} \equiv \sum_{j=0}^{k} r_j u_{k-j} = \sum_{j=0}^{k-1} r_j u_{(k-1)-j}.$$

Thus for all $k \geq 1$,

$$\sum_{j=0}^{k} r_j u_{k-j} = r_0 u_0 = 1.$$

(*iii*) Since $0 \leq u_k \leq 1$, there exists a subsequence $\{k_v\}$ of the positive integers such that

$$\lim_{v \to \infty} u_{k_v - j} = \alpha \equiv \limsup_{k \to \infty} u_k \leq 1.$$

Let j be such that $f_j > 0$, and suppose

$$\limsup_{v \to \infty} u_{k_v - j} = \beta, \qquad (0 \leq \beta \leq \alpha).$$

Suppose $\alpha > \beta$. Then

$$\begin{aligned} u_{k_v} &= \sum_{r=0}^{k_v} f_r u_{k_v - r} \\ &\leq \sum_{\substack{r=0 \\ r \neq j}}^{M} f_r u_{k_v - r} + f_j u_{k_v - j} + \varepsilon \end{aligned}$$

where $0 < \varepsilon < (\alpha - \beta) f_j$ and $M \geq j$ is chosen so that

$$r_M = \sum_{r=M+1}^{\infty} f_r < \varepsilon,$$

since $0 \leq u_k \leq 1$. Letting $v \to \infty$

$$\alpha \leq \alpha \sum_{\substack{r=0 \\ r \neq j}}^{M} f_r + f_j \beta + \varepsilon$$

i.e. letting $M \to \infty$,

$$\alpha \le \alpha \sum_{\substack{r=0 \\ r \ne j}}^{\infty} f_r + f_j \beta + \varepsilon$$

$$= \alpha \sum_{r=0}^{\infty} f_r - (\alpha - \beta) f_j + \varepsilon$$

i.e.

$$\alpha \le \alpha - (\alpha - \beta) f_j + \varepsilon.$$

Taking into account the choice of ε, this is a contradiction. Hence $\beta = \alpha$; thus we may take $\{k_v\}$ to be a sequence such that

$$\lim_{v \to \infty} u_{k_v - j} = \alpha$$

This is true for any j such that $f_j > 0$, the sequence $\{k_v\}$ depending on j. If we consider the set of j, $\{j_1, \ldots, j_r\}$ specified in part (i) of the proof, then by a suitable number of repetitions of the argument just given, for any fixed nonnegative integers $p_1, p_2, \ldots, p_r$ there exists a (refined) sequence $\{k_v\}$ such that

$$\lim_{v \to \infty} u_{k_v - (p_1 j_1 + p_2 j_2 + \cdots + p_r j_r)} = \alpha$$

i.e. for any fixed integer $q \ge N_0$ (from part (i) of the proof)

$$\lim_{v \to \infty} u_{k_v - q} = \alpha \tag{5.4}$$

where subsequence $\{k_v\}$ depends on q.

Denote the sequence $\{k_v\}$ corresponding to $q = N_0$ by $\{k_v^{(0)}\}$. Repeat again the argument of (iii) in starting from *this* sequence to obtain a subsequence of it $\{k_v^{(1)}\}$ for which (5.4) holds, for $q = 1$, etc. It follows that, from the Cantor diagonal selection principle, there is a subsequence of the integers (we shall still call it $\{k_v\}$) such that

$$\lim_{v \to \infty} u_{k_v - q} = \alpha$$

for *every* $q \ge N_0$.

Now let

$$\gamma = \liminf_{k \to \infty} u_k (0 \le \gamma \le \alpha \le 1).$$

By an argument analogous but "dual" to the one above,[1] there exists a subsequence $\{n_v\}$ of the integers such that

$$\lim_{v \to \infty} u_{n_v - q} = \gamma$$

for all $q \ge N_0$.

[1] But effectively using Fatou's Lemma rather than the epsilon argument.

(*iv*) Now introduce new subsequences of the positive integers, $\{s_v\}$, $\{t_v\}$ defined for sufficiently large v by

$$s_v = k_v - N_0, \qquad t_v = n_v - N_0.$$

Then
$$\lim_{v\to\infty} u_{s_v - p} = \lim_{v\to\infty} u_{k_v-(N_0+p)} = \alpha$$

for $p \geq 0$, and similarly

$$\lim_{v\to\infty} u_{t_v-p} = \gamma.$$

Now
$$1 = \sum_{p=0}^{s_v} r_p u_{s_v-p},$$

$$1 = \sum_{p=0}^{t_v} r_p u_{t_v-p} \leq \sum_{p=0}^{M} r_p u_{t_v-p} + \varepsilon$$

where $\varepsilon > 0$ is arbitrary, and M such that $r_M < \varepsilon$. From these relations, as $v \to \infty$, taking lim inf and using Fatou's lemma in the first; and lim sup in the second, and subsequently $M \to \infty$; and $\varepsilon \to 0$:

$$1 \geq \alpha \sum_{p=0}^{\infty} r_p, \qquad 1 \leq \gamma \sum_{p=0}^{\infty} r_p$$

where $0 \leq \gamma \leq \alpha$.

Further

$$\sum_{p=0}^{\infty} r_p = \sum_{p=0}^{\infty} \sum_{j=p}^{\infty} f_{j+1} = \sum_{j=0}^{\infty} \sum_{p=0}^{j} f_{j+1} = \sum_{j=0}^{\infty} (j+1) f_{j+1}, = \mu, \text{ say.}$$

Thus if $\mu \equiv \sum_{p=0}^{\infty} r_p = \infty$, $\Rightarrow \alpha = 0$; thus $\alpha = \gamma = \mu^{-1}$, as required. If

$$\mu < \infty, \qquad \mu^{-1} \geq \alpha \geq \gamma \geq \mu^{-1};$$

thus $\alpha = \gamma = \mu^{-1}$, as required.

This completes the proof. □

Corollary 1. *If i is a recurrent aperiodic index and j is any index such that $j \to i$ then as $k \to \infty$*

$$p_{ij}^{(k)} \to \mu_i^{-1} L_{ij}(1-).$$

PROOF.

$$p_{ij}^{(k)} = \sum_{t=0}^{k} p_{ii}^{(t)} l_{ij}^{(k-t)} = \sum_{t=0}^{k} p_{ii}^{(k-t)} l_{ij}^{(t)}.$$

Now since
$$L_{ij}(1-) = \sum_{k=0}^{\infty} l_{ij}^{(k)} < \infty$$

(by Corollary 1 to Lemma 5.4), we may use the Dominated Convergence Theorem to conclude that as $k \to \infty$ (by Theorem 5.1)

$$p_{ij}^{(k)} \to \mu_i^{-1} \sum_{t=0}^{\infty} l_{ij}^{(t)}$$

which is the required conclusion. □

Corollary 2. *If i is in fact positive recurrent,*

$$\sum_j L_{ij}(1-) \le \mu_i < \infty.$$

Proof. By stochasticity and since a recurrent index is essential

$$1 = \sum_j p_{ij}^{(k)} = \sum_{\substack{j \\ j \to i}} p_{ij}^{(k)} + \sum_{\substack{j \\ j \not\to i}} p_{ij}^{(k)} = \sum_{\substack{j \\ j \to i}} p_{ij}^{(k)}$$

so that by Fatou's Lemma

$$1 \ge \mu_i^{-1} \sum_j L_{ij}(1-) \text{ as required.}$$ □

[Note that by similar argument, Corollary 1 is trivially true if $j \not\to i$].

It is possible to develop (as in the case of a finite matrix) a rather more unified theory in the important case of a single essential class of indices, i.e. in the case of an irreducible P. We now pass to this case.

5.3 Irreducible Stochastic Matrices

Since Lemma 1.2 of Chapter 1 continues to hold in the present context, we know that every index of an irreducible countable P has the *same* period, d.

Properties, like this, possessed in common by all indices of a single essential class of indices, are often called *solidarity* properties. We shall go on to show that transience, null-recurrence, and positive-recurrence are all solidarity properties, and to describe the limiting behaviour as $k \to \infty$ of P^k as a whole.

Moreover an important role in the present theory is played by *subinvariant row vectors* (measures), and we treat this topic in some depth also. The results which are given for these are the (one-sided) analogues, for countable irreducible *stochastic* P, of the crucial Subinvariance Theorem of Chapter 1 for finite irreducible non-negative T.

Theorem 5.2. *All indices corresponding to an irreducible stochastic P are transient, or all are null-recurrent, or all are positive-recurrent.*

Proof. We first note the elementary inequalities for any two indices r and j

$$p_{jj}^{(k+N+M)} \geq p_{jr}^{(N)} p_{rr}^{(k)} p_{rj}^{(M)}$$
$$p_{rr}^{(k+N+M)} \geq p_{rj}^{(M)} p_{jj}^{(k)} p_{jr}^{(N)}$$

which follow directly from $P^{k+N+M} = P^N P^k P^M = P^M P^k P^N$. Now since $r \leftrightarrow j$, it follows that M and N can be chosen so that

$$p_{rj}^{(M)} > 0, \qquad p_{jr}^{(N)} > 0.$$

Now the index r must be transient, or null-recurrent, or positive-recurrent. We treat the cases separately.

(i) If r is transient

$$\sum_{k=0}^{\infty} p_{rr}^{(k)} < \infty$$

from Lemma 5.3; this implies by the second of the elementary inequalities that

$$\sum_{k=0}^{\infty} p_{jj}^{(k)} < \infty$$

so (any other index) j is also transient.

(ii) If r is null-recurrent, from Theorem 5.1

$$p_{rr}^{(k)} \to 0 \qquad \text{as} \quad k \to \infty;$$

thus by the second of the elementary inequalities

$$p_{jj}^{(k)} \to 0 \quad \text{as} \quad k \to \infty.$$

On the other hand, since r is recurrent, from the Corollary to Lemma 5.3 it follows that

$$\sum_{k=0}^{\infty} p_{rr}^{(k)} = \infty;$$

hence the first of the elementary inequalities implies

$$\sum_{k=0}^{\infty} p_{jj}^{(k)} = \infty$$

so that by Lemma 5.3, j is recurrent. Now if j is assumed positive recurrent, we know from Theorem 5.1 (and its periodic version) that as $k \to \infty$

$$p_{jj}^{(k)} \nrightarrow 0;$$

which is a contradiction. Hence j is also null-recurrent.

(iii) If r is positive recurrent, j must be also; otherwise a contradiction to the positive recurrence of r would arise from (i) or (ii). □

This theorem justifies the following definition.

Definition 5.2. An irreducible P is said to be transient, or null-recurrent, or positive-recurrent depending on whether any one of its indices is transient, or null-recurrent, or positive-recurrent, respectively.

Definition 5.3. For a stochastic P, a row vector x', $x' \geq \mathbf{0}'$, $x' \neq \mathbf{0}'$ satisfying

$$x'P \leq x'$$

is called a *subinvariant* measure. If in fact $x'P = x'$ the x' is called an *invariant* measure. [Note that a positive multiple of such a measure is still such a measure.]

Lemma 5.5. *For a stochastic irreducible P, a subinvariant measure always exists. One such is given by the vector $\{L_{ij}(1-)\}$, $j = 1, 2$, [for arbitrary fixed i], whose elements are positive and finite.*

Proof. From Lemma 5.4, for i fixed but arbitrary, and $j = 1, 2, \ldots$, and $0 < s < 1$

$$L_{ij}(s) = s \sum L_{ir}(s)p_{rj} + sp_{ij}(1 - L_{ii}(s)). \tag{5.5}$$

Letting $s \to 1-$ and using Fatou's Lemma

$$L_{ij}(1-) \geq \sum_r L_{ir}(1-)p_{rj} + p_{ij}(1 - L_{ii}(1-))$$

whence, since $L_{ii}(1-) \leq 1$, the row vector $\{L_{ij}(1-)\}$, $j = 1, 2, \ldots$ is a subinvariant measure. We know that since $i \leftrightarrow j$

$$0 < L_{ij}(1-) < \infty$$

from Corollary 1 of Lemma 5.4. □

Corollary. *For fixed i, $j = 1, 2, 3, \ldots$*

$$L_{ij}(1-) = \sum_r L_{ir}(1-)p_{rj} + p_{ij}(1 - L_{ii}(1-)).$$

Proof. From the proof above, since

$$s \sum_r L_{ir}(s)p_{rj} \leq \sum_r L_{ir}(1-)p_{rj} \leq L_{ij}(1-) < \infty$$

we may use the Dominated Convergence Theorem in (5.5) in letting $s \to 1-$, to obtain the "finer" result stated. □

Lemma 5.6. *Any subinvariant measure for irreducible stochastic P has all its entries positive.*

Proof. For such a measure x'

$$x'P \leq x'$$

so that

$$x'P^k \leq x'$$

and in particular

$$\sum_i x_i p_{ij}^{(k)} \leq x_j$$

i.e. for any i, j

$$x_i p_{ij}^{(k)} \leq x_j.$$

Select fixed i so that $x_i > 0$; for any fixed j, since $i \to j$, there is a k such that $p_{ij}^{(k)} > 0$. Hence

$$x_j > 0 \quad \text{all} \quad j = 1, 2, \ldots$$ □

Lemma 5.6 makes sensible the formulation of the following result.

Theorem 5.3. *If $x' = \{x_i\}$ is any subinvariant measure corresponding to irreducible stochastic P, then for fixed but arbitrary i, and all $j = 1, 2, \ldots$*

$$x_j/x_i \geq \bar{x}_{ij}$$

where[1]

$$\bar{x}_{ij} = (1 - \delta_{ij})L_{ij}(1-) + \delta_{ij}$$

and $\bar{x}_{ij}$, $j = 1, 2, \ldots$ is also a subinvariant measure (with ith element unity).

[This theorem therefore says that out of all subinvariant measures normed to have a fixed element unity, there is one which is minimal.]

PROOF. That $\{\bar{x}_{ij}\}$, $j = 1, 2, \ldots$ is subinvariant is readily checked from the statement of Corollary 1 of Lemma 5.5, since $L_{ii}(1-) \leq 1$.

We prove the rest by induction. Let $\{y_j\}$ be any subinvariant measure with $y_i = 1$. It is required to show that for every j

$$y_j \geq (1 - \delta_{ij}) \sum_{k=0}^{\infty} l_{ij}^{(k)} + \delta_{ij}$$

or equivalently that for every j, and every $m \geq 1$

$$y_j \geq (1 - \delta_{ij}) \sum_{k=0}^{m} l_{ij}^{(k)} + \delta_{ij}.$$

Now we have for all j,

$$y_j \geq \sum_r y_r p_{rj} \geq y_i p_{ij} = l_{ij}^{(1)}$$

and by assumption

$$y_i = 1,$$

[1] δ_{ij} is the Kronecker delta, as before.

so the proposition is true for $m = 1$ and all j. Assume it is true for $m \geq 1$: then for each j

$$\begin{aligned} y_j &\geq \sum_r y_r p_{rj} \\ &= \sum_{r \neq i} y_r p_{rj} + p_{ij} \\ &\geq \sum_{r \neq i} \sum_{k=0}^{m} l_{ir}^{(k)} p_{rj} + p_{ij} \\ &= \sum_{k=1}^{m} \sum_{r \neq i} l_{ir}^{(k)} p_{rj} + p_{ij} \end{aligned}$$

and from the definition of the $l_{ij}^{(k)}$

$$\begin{aligned} &= \sum_{k=1}^{m} l_{ij}^{(k+1)} + l_{ij}^{(1)} \\ &= \sum_{k=0}^{m} l_{ij}^{(k+1)} = \sum_{k=0}^{m+1} l_{ij}^{(k)}. \end{aligned}$$

Hence

$$y_j \geq \sum_{k=0}^{m+1} l_{ij}^{(k)}, \quad \text{all} \quad j,$$

and moreover

$$y_i = 1$$

by given, so induction is true for $m + 1$; This completes the proof. □

Theorem 5.4. *For a recurrent*[1] *matrix P an invariant measure exists, and is unique (to constant multiples).*

A subinvariant measure which is not invariant exists if and only if P is a transient matrix; one such subinvariant measure is then given by $\{x_{ij}\}, j = 1, 2, \ldots$

PROOF. If P is recurrent, $L_{ii}(1-) = 1$ for all i, so existence of an invariant measure $\{x_{ij}\}, j = 1, 2, \ldots$ follows from the Corollary of Lemma 5.5. Now let $y = \{y_i\}$ be any subinvariant measure scaled so that $y_i = 1$ for fixed i. Then

$$z = \{y_j - x_{ij}\}, \qquad j = 1, 2, \ldots$$

satisfies $z \geq \mathbf{0}$, (from Theorem 5.3) and $z_i = 0$; and if $z \neq \mathbf{0}$, z is clearly a subinvariant measure, since y is subinvariant and $\{x_{ij}\}$ is invariant: but $z_i = 0$ and this is not possible by Lemma 5.6.

Hence $z = \mathbf{0}$, which completes the proof for the recurrent case.

If P is transient, $L_{ii}(1-) < 1$ for each i, hence a subinvariant measure which is not invariant may be taken as $\{x_{ij}\}, j = 1, 2, \ldots$ for fixed i from the

[1] i.e. positive-recurrent or null-recurrent

Corollary of Lemma 5.5; there is strict inequality in the subinvariance equations only at the ith position. Suppose now a subinvariant measure $y = \{y_j\}$, normed so that $y_i = 1$, exists but is not invariant, in relation to a stochastic P. Then

$$y_j \geq \bar{x}_{ij}, \qquad j = 1, 2, \ldots$$

If P is recurrent, then from the first part y must be invariant, since $y_j = \bar{x}_{ij}$, which is a contradiction. Hence P must be transient. □

Corollary. *If x' is a subinvariant measure for recurrent P, it is, in fact, an invariant measure, and is a positive multiple of $\{\bar{x}_{ij}\}$, $j = 1, 2, \ldots$*

Theorem 5.5. (*General Ergodic Theorem*[1]). *Let P be a primitive*[2] *stochastic matrix. If P is transient or null-recurrent, then for each pair of indices i, j, $p_{ij}^{(k)} \to 0$ as $k \to \infty$.*

If P is positive recurrent, then for each pair i, j

$$\lim_{k \to \infty} p_{ij}^{(k)} = \mu_j^{-1}$$

and the vector $v = \{\mu_j^{-1}\}$ is the unique invariant measure of P satisfying $v'\mathbf{1} = 1$. [v' is thus of course the unique stationary distribution.]

Proof. If P is transient, the result follows from Corollary 2 of Lemma 5.4 (in fact if P is merely irreducible); if P is null-recurrent, from Corollary 1 of Theorem 5.1.

If P is positive recurrent on the other hand, for any pair i, j

$$p_{ij}^{(k)} \to \mu_i^{-1} L_{ij}(1-) = \mu_i^{-1} \bar{x}_{ij}$$

(since in the recurrent case $L_{ii}(1-) = 1$).

From the Corollary to Theorem 5.4, for fixed i

$$\sum_r \bar{x}_{ir} p_{rj} = \bar{x}_{ij}, \qquad j = 1, 2, \ldots$$

where (from Corollary 2 of Theorem 5.1)

$$\sum_r \bar{x}_{ir} \leq \mu_i < \infty.$$

Hence $u' = \{u_j\}$ where

$$u_j = \bar{x}_{ij} / \sum_r \bar{x}_{ir}$$

is the unique invariant measure of P satisfying $u'\mathbf{1} = 1$ (and hence is independent of the initial choice of i).

[1] cf. Theorem 4.2.

[2] For the irreducible periodic case see Exercises 5.1 and 5.2.

Now since $u'P = u'$ it follows $u'P^k = u'$ so that

$$u_j = \sum_r u_r p_{rj}^{(k)}, \qquad \text{all } j.$$

Since $\sum_r u_r < \infty$, we have, by dominated convergence, letting $k \to \infty$

$$u_j = \sum_r u_r \mu_r^{-1} \bar{x}_{rj}. \tag{5.6}$$

Now suppose that in fact

$$\sum_j \bar{x}_{rj} < \mu_r \qquad \text{for some } r.$$

Then summing over j in (5.6) (using Fubini's Theorem)

$$\sum_j u_j < \sum_r u_r$$

which is impossible. Hence

$$\sum_j \bar{x}_{rj} = \mu_r \qquad \text{for all } r = 1, 2, \ldots$$

and so
$$u_j = \mu_i^{-1} \bar{x}_{ij} \qquad \text{for all } i;$$

and by putting $i = j$, we see

$$u_j = v_j = \mu_j^{-1}$$

as required, since $\bar{x}_{ii} = 1$. □

Corollary. *If P is an irreducible transient or null-recurrent matrix, there exists no invariant measure $\boldsymbol{v}'$ satisfying $\boldsymbol{v}'\mathbf{1} < \infty$.*

PROOF. In either case[1] we know $p_{ij}^{(k)} \to 0$ as $k \to \infty$. Suppose a measure of the required sort exists: then

$$v_j = \sum_i v_i p_{ij} = \sum_i v_i p_{ij}^{(k)}.$$

Since $\sum_i v_i < \infty$, by dominated convergence

$$v_j = \sum_i v_i \left(\lim_{k \to \infty} p_{ij}^{(k)} \right) = 0,$$

for each $j = 1, 2, \ldots$, which is a contradiction to $\boldsymbol{v} \geq \mathbf{0}, \neq \mathbf{0}$. □

5.4 The "Dual" Approach; Subinvariant Vectors

In §§5.1–5.3 of this chapter we developed the theory of classification of indices, subinvariant measures, and asymptotic behaviour of powers P^k, for

[1] See Exercise 5.2 for the periodic situation.

a countable stochastic matrix P, in terms of the quantities $l_{ij}^{(k)}$ and consequently the $L_{ij}(1-)$.

There is a more classical approach, dual to the one we have presented, for developing the index classification and behaviour of powers theory, which is in many respects (as will be seen) more natural than the one just given, since it accords better with the fact that the row sums of P are unity. However (although it too leads to Theorem 5.5), it is not suitable for dealing with subinvariant measures, which multiply P from the left, and are particularly important in the stability theory of Markov chains: but rather with *subinvariant vectors*.

Definition 5.4. For a stochastic P, a column vector $\boldsymbol{u}$, $\boldsymbol{u} \geq \mathbf{0}$, $\boldsymbol{u} \neq \mathbf{0}$ satisfying

$$P\boldsymbol{u} \leq \boldsymbol{u}$$

is called a *subinvariant vector*. If in fact $P\boldsymbol{u} = \boldsymbol{u}$, the vector $\boldsymbol{u}$ is called an *invariant vector*. [A positive multiple of such is still such a vector.]

The stochasticity of P ensures than an invariant vector always exists, viz. $\mathbf{1}$ is such, for $P\mathbf{1} = \mathbf{1}$. While subinvariant vectors are of little importance in the stability theory of MC's, they are of substantial interest in the description of long term behaviour of transient chains, i.e. of their Martin exit boundary theory, to which we shall pass shortly.

It is thus appropriate to sketch here the alternative development, although all proofs will be left to the interested reader.[1]

The development is in terms of the quantities $f_{ij}^{(k)}$, $k \geq 0$, $i, j = 1, 2, \ldots$ defined by

$$f_{ij}^{(1)} = p_{ij}; \qquad f_{ij}^{(k+1)} = \sum_{\substack{r \\ r \neq j}} p_{ir} f_{rj}^{(k)}, \qquad k \geq 1,$$

where $f_{ij}^{(0)} = 0$ by definition. [In the MC framework, $f_{ij}^{(k)}$ is the probability of going from i to j in k steps, without visiting j in the meantime; it is also an example of a "taboo" probability, and the set $\{f_{ij}^{(k)}\}$, $k = 0, 1, 2, \ldots$ is known as the first passage time distribution from i to j.]

The generating functions, well defined $|z| < 1$,

$$F_{ij}(z) = \sum_k f_{ij}^{(k)} z^k, \qquad P_{ij}(z) = \sum_k p_{ij}^{(k)} z^k$$

are related by (the analogue of Lemma 5.1)

$$P_{ii}(z) = [1 - F_{ii}(z)]^{-1}; \qquad P_{ij}(z) = F_{ij}(z) P_{jj}(z), \qquad i \neq j$$

from which it is seen that $F_{ii}(z) = L_{ii}(z)$, $|z| < 1$, so that $f_{ii}^{(k)} = l_{ii}^{(k)}$. From this it is seen that the classification theory of indices is the same, and Theorem 5.1 is equally relevant to the present situation.

[1] See Exercise 5.4.

On the other hand, instead of Lemma 5.4 we have from considering now $P^{k+1} = PP^k$:

Lemma 5.4.$^{\text{D}}$. *For $|z| < 1$, and all $i, j = 1, 2, \ldots$*

$$P_{ij}(z) = z \sum_r p_{ir} P_{rj}(z) + \delta_{ij}$$

$$F_{ij}(z) = z \sum_r p_{ir} F_{rj}(z) + z p_{ij}(1 - F_{jj}(z)).$$

Corollary 1. *For each pair of indices i, j, $F_{ij}(1-) \le 1$.*

Corollary 2. *If j is transient,*

$$\sum_{k=0}^{\infty} p_{ij}^{(k)} < \infty$$

for all i (and so $p_{ij}^{(k)} \to 0$ as $k \to \infty$). If i is recurrent and $i \leftrightarrow j$,

$$\sum_{k=0}^{\infty} p_{ij}^{(k)} = \infty.$$

Instead of Corollary 1 to Theorem 5.1, we have that if j is a *recurrent aperiodic index*, and i is any other index,

$$p_{ij}^{(k)} \to \mu_j^{-1} F_{ij}(1-).$$

The results are adequate to prove the analogue of Theorem 5.2. The analogue of Lemma 5.5 is trivial in view of the vector $\boldsymbol{u} = \mathbf{1}$; and the analogue of Lemma 5.6 holds from similar argument.

The minimal subinvariant vector is given, for fixed but arbitrary j and $i = 1, 2, \ldots$ by

$$\bar{u}_{ij} = (1 - \delta_{ij}) F_{ij}(1-) + \delta_{ij};$$

and the analogue of Theorem 5.4 reads:

Theorem 5.4.$^{\text{D}}$. *For a recurrent matrix P, an invariant vector exists and is a positive multiple of the vector* $\mathbf{1}$, *and $\bar{u}_{ij} = 1$, all i, j.*

A subinvariant vector which is not invariant exists if and only if P is a transient matrix; one such subinvariant vector is given by $\{\bar{u}_{ij}\}$, $i = 1, 2, \ldots$.

Theorem 5.5 can be proved also, using slightly different emphasis.

Some of the notes made in the present section will be required in the next.

5.5 Potential and Boundary Theory for Transient Indices[1]

In this section only we shall relax the assumption that $P = \{p_{ij}\}$, $i, j \geq 1$ is stochastic in a minor way, assuming only that

$$p_{ij} \geq 0, \qquad 1 \geq \sum_j p_{ij} > 0 \qquad \text{each } i.$$

In the strictly substochastic case, i.e. where for some i, $1 > \sum_j p_{ij}$, we may think of P as a part of an enlarged stochastic matrix $\bar{P} = \{\bar{p}_{ij}\}$, on the index set $\{0, 1, 2, \ldots\}$, where $\bar{p}_{ij} = p_{ij}$, $i, j \geq 1$; $\bar{p}_{i0} = 1 - \sum_j p_{ij}$, $i \geq 1$, $\bar{p}_{00} = 1$: in this case

$$\bar{P} = \begin{bmatrix} 1 & \mathbf{0}' \\ \boldsymbol{p}_1 & P \end{bmatrix}, \qquad \boldsymbol{p}_1 \geq \mathbf{0}, \qquad \neq \mathbf{0}$$

so that

$$\bar{P}^k = \begin{bmatrix} 1 & \mathbf{0}' \\ \boldsymbol{p}_k & P^k \end{bmatrix};$$

so in either situation it is meaningful to study the (well-defined) powers P^k of the matrix $P = \{p_{ij}\}$, $i, j = 1, 2, \ldots$

It is readily seen that Lemma 5.4^D and its Corollaries continue to hold in the present more general situation; and in the strictly substochastic case of P, the $P_{ij}(z)$ and $F_{ij}(z)$ for $i, j \geq 1$ coincide with the corresponding quantities $\bar{P}_{ij}(z)$, $\bar{F}_{ij}(z)$ of the expanded matrix $\bar{P}$ for $i, j \geq 1 (\bar{f}_{0j}^{(k)} = 0$ for all $k \geq 0$ if $j \geq 1)$.

We shall now extend slightly the notion of a subinvariant vector, and introduce the more usual terminology which generally occurs in this context.

Definition 5.5. A column vector[2] $\boldsymbol{u} \geq \mathbf{0}$ satisfying

$$P\boldsymbol{u} \leq \boldsymbol{u}$$

is said to be a *superregular vector* for P. If in fact $P\boldsymbol{u} = \boldsymbol{u}$, $\boldsymbol{u}$ is said to be *regular*.

Thus the vector $\mathbf{0}$ is regular for P; and the vector $\mathbf{1}$ is superregular (but is not regular if P is strictly substochastic).

If we define the vector $\{\bar{u}_{ij}\}$, $i = 1, 2, \ldots$ for fixed but arbitrary $j \geq 1$ as in the previous section, it will similarly be superregular for P: in fact

$$\bar{u}_{ij} = \sum_k p_{ik}\bar{u}_{kj} + \delta_{ij}(1 - F_{jj}(1-)), \qquad i = 1, 2, \ldots \tag{5.7}$$

[1] This section is of more specialist interest; and in the part dealing with the Poisson–Martin representation requires deeper mathematics than the rest of the book.

[2] Assumed elementwise finite as usual.

[so that strict inequality may occur only in position $i = j$ as regards the superregularity equations]; and the minimality property in this case (obtained with an almost identical induction proof) will take the form that for any superregular $\boldsymbol{u}$,

$$u_i \geq \bar{u}_{ij} u_j, \qquad i, j = 1, 2, \ldots$$

(since we no longer have the assurance that $\boldsymbol{u} > \mathbf{0}$).

Since $\{\bar{u}_{ir}\}$, $i = 1, 2, \ldots$ is itself superregular, we have, by putting $u_i = \bar{u}_{ir}$, $i = 1, 2, \ldots$:

The Fundamental Inequality: $\bar{u}_{ir} \geq \bar{u}_{ij}\bar{u}_{jr}$, all $i, j, r \geq 1$.

We shall for the rest of the section make the

Basic Assumption 1.

(*i*) The index set of P is in fact *all the positive integers*, R. (The case of finite P is excluded here for the first time.)
(*ii*) All indices are *transient*: i.e. $F_{ii}(1-) < 1$ for all $i \geq 1$.

[Note that we do *not* necessarily have irreducibility of P.]

By Corollary 2 of Lemma 5.4^{D} of the previous section

$$g_{ij} = \sum_{k=0}^{\infty} p_{ij}^{(k)} < \infty \qquad \text{for all } i, j \in R$$

so that the non-negative matrix $G = \{g_{ij}\}$ on $i, j \in R$ (i.e. $R \times R$) is elementwise finite. This matrix will be used to define potentials on R; it is sometimes called the *Green's function*, or *kernel* defined on $(R \times R)$ corresponding to P.

Potential Theory

Since $\boldsymbol{u} \geq P\boldsymbol{u}$ for a vector $\boldsymbol{u}$ superregular on R, it follows

$$\boldsymbol{u} \geq P\boldsymbol{u} \geq P^2\boldsymbol{u} \geq \cdots \geq P^k\boldsymbol{u} \geq \mathbf{0}.$$

Denote the limit vector of the monotone decreasing non-negative vector sequence $\{P^k\boldsymbol{u}\}$ by $P^{\infty}\boldsymbol{u}$.

Lemma 5.7. *For a superregular vector $\boldsymbol{u}$, $P^{\infty}\boldsymbol{u}$ is regular on R, and moreover*

$$\boldsymbol{u} = P^{\infty}\boldsymbol{u} + G(\boldsymbol{u} - P\boldsymbol{u}).$$

PROOF.

$$P^{\infty}\boldsymbol{u} = \lim_{k \to \infty} P^{r+k}\boldsymbol{u} = \lim_{k \to \infty} P^r(P^k\boldsymbol{u})$$

(by Fubini's Theorem), and since $P^k\boldsymbol{u} \leq \boldsymbol{u}$, and $P^k\boldsymbol{u} \downarrow P^{\infty}\boldsymbol{u}$,

$$= P^r\left(\lim_{k\to\infty} P^k\boldsymbol{u}\right)$$

by dominated convergence,

$$= P^r(P^{\infty}\boldsymbol{u}).$$

Taking $r = 1$ shows that $P^{\infty}\boldsymbol{u}$ is regular.
Now, we may write $\boldsymbol{u}$ as

$$\boldsymbol{u} = P^{r+1}\boldsymbol{u} + \sum_{k=0}^{r} (P^k\boldsymbol{u} - P^{k+1}\boldsymbol{u}), \qquad \text{for } r \geq 1$$

$$= P^{r+1}\boldsymbol{u} + \sum_{k=0}^{r} P^k[\boldsymbol{u} - P\boldsymbol{u}].$$

Letting $r \to \infty$ yields the result, since $G = \sum_{k=0}^{\infty} P^k$. □

Definition 5.6. If $\boldsymbol{v} \geq \mathbf{0}$ (elementwise finite as usual) is a column vector, then $G\boldsymbol{v}$ is called its potential [$G\boldsymbol{v}$ may have some entries infinite.]

Lemma 5.8. *If a potential $G\boldsymbol{v}$ is everywhere finite, it (a) determines $\boldsymbol{v}$; and, (b) is superregular.*

PROOF. It is easily checked that

$$G = PG + I$$

from the definition of G; applying this to (an elementwise finite) vector $\boldsymbol{v}$ we have

$$G\boldsymbol{v} = P(G\boldsymbol{v}) + \boldsymbol{v}.$$

Thus if $G\boldsymbol{v}$ is everywhere finite, $P(G\boldsymbol{v})$ is finite, and

$$\boldsymbol{v} = G\boldsymbol{v} - P(G\boldsymbol{v}) = (I - P)(G\boldsymbol{v})$$

which proves the first part; further in this situation clearly

$$G\boldsymbol{v} \geq P(G\boldsymbol{v})$$

which is equivalent to the second part. □

Lemma 5.9. *A necessary and sufficient condition for a superregular vector $\boldsymbol{u}$ to be a (finite) potential is $P^{\infty}\boldsymbol{u} = \mathbf{0}$.*

PROOF. Suppose a non-negative vector $\boldsymbol{u}$ is an elementwise finite potential; then for some (elementwise finite) $\boldsymbol{v} \geq 0$

$$\boldsymbol{u} = G\boldsymbol{v},$$

$$= P\boldsymbol{u} + \boldsymbol{v}$$

from the proof of Lemma 5.8, i.e.

$$P^k u - P^{k+1} u = P^k v$$

i.e.
$$\sum_{k=0}^{r} (P^k u - P^{k+1} u) = \sum_{k=0}^{r} P^k v$$

$$u - P^{r+1} u = \left(\sum_{k=0}^{r} P^k \right) v.$$

Let $r \to \infty$;

$$u - P^{\infty} u = Gv, = u \qquad \text{by definition of } u \text{ and } v.$$

Therefore
$$P^{\infty} u = \mathbf{0}.$$

Now let u be a superregular vector such that $P^{\infty} u = \mathbf{0}$. Define v, non-negative and elementwise finite by

$$v = u - Pu.$$

Hence as before

$$\sum_{k=0}^{r} P^k v = \sum_{k=0}^{r} (P^k u - P^{k+1} u)$$

and letting $r \to \infty$

$$Gv = u - P^{\infty} u$$

$$= u$$

by assumption. Hence u is a potential for v. □

Theorem 5.6. *A superregular vector u may be decomposed into a sum*

$$u = r + g$$

where g is a potential and r is regular. The decomposition is unique; in fact

$$r = P^{\infty} u, \qquad g = G(u - Pu).$$

PROOF. Lemma 5.7 already asserts the possibility of decomposition with the specific forms for r and g stated. To prove uniqueness let

$$u = r + g$$

where r is regular and g is a potential; g is necessarily elementwise finite, since u is, by its very definition.

Then

$$P^k u = P^k r + P^k g$$

i.e.
$$P^k u = r + P^k g$$

since r is regular; let $k \to \infty$, taking into account g is also superregular, in view of Lemma 5.8

$$P^{\infty}u = r + P^{\infty}g$$
$$= r$$

since $P^{\infty}g = \mathbf{0}$ by Lemma 5.9.

Hence

$$u = P^{\infty}u + g$$

and the fact that $g = G(u - Pu)$ follows from Lemma 5.7. □

Theorem 5.7. *Let u be a superregular vector and h an (elementwise finite) potential. Then (a) the elementwise minimum, $c = \min(u, h)$ is also a potential; and (b) there exists a non-decreasing sequence of (finite) potentials $\{h_n\}$ converging (elementwise) to u.*

PROOF. (*a*) Since h is a finite potential, $P^{\infty}h = \mathbf{0}$ (Lemma 5.9).

Thus since $\quad h \geq c \geq \mathbf{0}, \quad$ then $P^k h \geq P^k c \geq \mathbf{0}$.

It follows, letting $k \to \infty$, that $P^{\infty}c = \mathbf{0}$.

We shall now prove c is superregular (it is clearly elementwise finite).

$$Pc \leq Pg \leq h$$

since h, being a potential, is superregular; and from definition of c. Similarly

$$Pc \leq Pu \leq u$$

from definition of c and u. Hence

$$Pc \leq \min(u, h) = c.$$

Hence by Lemma 5.9, c must be a potential.

(*b*) Let g and h be any two finite potentials; then clearly $g + h$ is superregular (since both g and h are) and

$$P^{\infty}(g + h) = P^{\infty}g + P^{\infty}h = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

by Lemma 5.9. Hence again by Lemma 5.9, $g + h$ is a finite potential; thus a sum of finite potentials is itself a finite potential.

Now let f_j be the column vector with unity in the jth position and zeros elsewhere: since $g_{jj} > 0$ always, Gf_j is a potential having its jth position nonzero, at least.

Let

$$g_j = x(j)Gf_j$$

where $x(j)$ is positive integer chosen so that $x(j)g_{jj} > u(j)$ where $u = \{u(j)\}$ is the given superregular vector. Then g_j is a finite potential (being a finite sum of finite potentials), and d_n defined by

$$d_n = \sum_{j=1}^{n} g_j$$

is also. It follows that $d_n(j) > u(j)$, $j = 1, 2, \ldots, n$.

Consider now the sequence $\{h_n\}$ defined by

$$h_n = \min(u, d_n),$$

h_n being a potential from the first part (*a*) of the theorem. Moreover $\{h_n\}$ is a non-decreasing sequence of finite potentials converging elementwise to u.

□

The reader acquainted with various versions of potential theory in physics will recognize that the above results are formally analogous. Theorem 5.6 is, for example, the analogue of the Riesz Decomposition Theorem; some further discussion will be given at the end of the chapter. We shall have occasion to make use of some of the previously developed results shortly.

The Martin Exit Boundary; the Poisson–Martin Integral Representation for a Superregular Vector

The basic purpose of this subsection, in keeping with the general aims of the book, is to develop an important representation of a superregular vector which is similar to, but rather more sophisticated than, the Riesz decomposition given by Lemma 5.7 and Theorem 5.6. In this connection we need to develop first a small amount of boundary theory, the Martin exit boundary being of additional importance in the study of long-term behaviour of the trajectories of infinite Markov chains which consist of transient states only. This probabilistic framework, which we shall not pursue here; nevertheless motivates the making of an additional basic assumption, which can in effect be reasoned (from the probabilistic framework) to be made without essential loss of generality.

Basic Assumption 2. The index 1 leads to every index, i.e. $1 \to j$, $j = 1, 2, \ldots$ [This implies that in $G = \{g_{ij}\}$, $g_{1j} > 0$ for all $j \in R$.]

Define for all $i, j \in R$

$$K(i, j) = \frac{\bar{u}_{ij}}{\bar{u}_{1j}} \qquad \text{(where } \bar{u}_{1j} > 0 \text{ for all } j \in R \text{ since } 1 \to j)$$

the $\bar{u}_{ij} = (1 - \delta_{ij})F_{ij}(1-) + \delta_{ij}$ having been defined in §5.4. [It may be useful to note also, on account of the relation between the $F_{ij}(z)$ and $P_{ij}(z)$, that in fact

$$K(i, j) = \frac{g_{ij}}{g_{1j}}, \qquad \text{all } i, j \in R.]$$

From the Fundamental Inequality deduced earlier in this section

$$\bar{u}_{1j} \geq \bar{u}_{1i}\bar{u}_{ij}$$

so that

$$K(i, j) \leq 1/\bar{u}_{1i} \qquad \text{for all } i, j;$$

and further, since for fixed $j \in R$, $\{\bar{u}_{ij}\}$ is superregular, then so is $K(i, j)$.

Let us now define on $R \times R$ the function

$$d(v_1, v_2) = \sum_{i \in R} |K(i, v_1) - K(i, v_2)| \bar{u}_{1i} w_i$$

where $\{w_i\}$ is any sequence of strictly positive numbers such that

$$\sum_{i \in R} w_i < \infty$$

(e.g. we might take $w_i = 2^{-i}$).

We shall now show that $d(.,.)$ is a *metric* on R i.e.

$$0 \leq d(v_1, v_2) < \infty \qquad \text{for } v_1, v_2 \in R$$

and:

(*i*) $d(v_1, v_2) = 0$ if and only if $v_1 = v_2$;
(*ii*) $d(v_1, v_2) = d(v_2, v_1)$;
(*iii*) $d(v_1, v_3) \leq d(v_1, v_2) + d(v_2, v_3)$.

The non-negativity is trivial; and finiteness of the bivariate function $d(.,.)$ follows from the fact that

$$d(v_1, v_2) \leq 2 \sum_{i \in R} w_i < \infty$$

on account of the triangle inequality on the real numbers, and the bound on $K(i, j)$ established above. Similarly (*ii*) and (*iii*) are trivial; and in fact the only non-obvious proposition which needs to be demonstrated is that

$$d(v_1, v_2) = 0 \Rightarrow v_1 = v_2.$$

Suppose not; suppose that for some v_1, v_2, $d(v_1, v_2) = 0$ and $v_1 \neq v_2$. Now $d(v_1, v_2) = 0$ implies

$$K(i, v_1) = K(i, v_2), \qquad \textit{all } i \in R$$

since $\bar{u}_{1i} w_i > 0$ for all i. Thus

$$\sum_{i \in R} p_{v_2 i} K(i, v_1) = \sum_{i \in R} p_{v_2 i} K(i, v_2).$$

But since $v_1 \neq v_2$ by assumption, this becomes

$$K(v_2, v_1) = K(v_2, v_2) - \frac{(1 - F_{v_2 v_2}(1-))}{\bar{u}_{1 v_2}}$$

from the subinvariance equations (5.7) for $\{\bar{u}_{ij}\}$, $i = 1, 2, \ldots$ for fixed j, where inequality occurs only in the $i = j$ position. Thus

$$K(v_2, v_1) < K(v_2, v_2)$$

which is a contradiction to $K(i, v_1) = K(i, v_2)$ for all $i \in R$.

Thus R is metrized by the metric d, and so we may henceforth speak of the *metric space* (R, d).

Now this metric space is *not necessarily complete*; i.e. we cannot say that for every Cauchy sequence $\{j_n\}$ in (R, d) (so that $d(j_n, j_m) < \varepsilon$, $n, m \geq N_0$) there exists a limit point $z \in (R, d)$ such that $d(j_n, z) \to 0$ as $n \to \infty$. In other words (R, d) does not necessarily contain all its *limit points*. The following lemma will help us understand the process of making (R, d) complete, by transferring the problem to the real line.

Lemma 5.10. *A sequence $\{j_n\}$ is Cauchy in the metric space (R, d) if and only if the sequence of real numbers $K(i, j_n)$ is Cauchy (in respect to the usual metric on the real line) for each $i \in R$*

PROOF. From the definition of $d(\ldots)$ if $\{j_n\}$ is Cauchy in (R, d), since $\bar{u}_{1i} w_i > 0$, $\{K(i, j_n)\}$ is Cauchy on the real line for each i.

Conversely, if $\{K(i, j_n)\}$ is Cauchy for each i, let $\varepsilon > 0$ and choose a finite subset of indices $E \subset R$ such that

$$\sum_{i \in R - E} w_i < \frac{\varepsilon}{4}.$$

Choose M sufficiently large so that

$$\bar{u}_{1i} | K(i, j_m) - K(i, j_n) | < \frac{\varepsilon}{2 \sum_{i \in R} w_i}$$

for each $i \in E$ and $n, m \geq M$.

Then writing symbolically

$$d(j_m, j_n) = \sum_{i \in E} + \sum_{i \in R - E}$$

in the definition of $d(\ldots)$ we have

$$\begin{aligned} d(j_m, j_n) &\leq \frac{\varepsilon \sum_{i \in E} w_i}{2 \sum_{i \in R} w_i} + \frac{\varepsilon}{2} \\ &\leq \varepsilon \end{aligned}$$

for $n, m \geq M$. □

Suppose now that $\{j_n\}$ is any Cauchy sequence in (R, d). Then, by the result just proved, $\{K(i, j_n)\}$ is Cauchy for each fixed $i \in R$. Thus for each $i \in R$

$$\lim_{n \to \infty} K(i, j_n) = K(i, x)$$

where $K(i, x)$ is notation we adopt for the limit, which is of course some real number; of course $x \equiv x(\{j_n\})$ is not an entity we can "picture" in general. but we only need to work with it insofar as we need only to know $K(i, x)$, $i \in R$. Define now for any $j \in R$

$$d(j, x) = \sum_{i \in R} |K(i, j_n) - K(i, x)| \bar{u}_{1i} w_i.$$

Then if $\{j_n\}$ is the Cauchy sequence corresponding to x

$$d(j_n, x) = \sum_{i \in R} |K(i, j_n) - K(i, x)| \bar{u}_{1i} w_i$$
$$\to 0$$

as $n \to \infty$, by dominated convergence. Thus if we *extend* the metric $d(\ldots)$ from R to operate *also* on any *new* points x added in this way, in the obvious manner, and put $x = y$, if and only if $K(i, x) = K(i, y)$ for all $i \in R$, then we shall have an *extended*, and now *complete*, metric space (R^*, d). The set R is thus *dense* in (R^*, d), which is therefore *separable*. In fact more is now true:

Lemma 5.11. *The metric space (R^*, d) is compact.*

PROOF. [Compactness in metric space is equivalent to the Weierstrass property: that every infinite sequence has a limit point in the metric space.]

Let $\{k_n\}$ be any sequence in (R^*, d). We know that for each $i \in R$, and for any $k \in R$ and so (by obvious extension) for any $k \in R^*$,

$$0 \leq K(i, k) \leq 1/\bar{u}_{1i}.$$

Thus for any fixed $i \in R$, $K(i, k_n)$ is a bounded sequence of real numbers; it then follows from the Bolzano–Weierstrass Theorem and the Cantor diagonal refinement procedure that there exists a subsequence $\{k_{n_j}\}, j = 1, 2, \ldots$ of $\{k_n\}$ such that

$$\lim_{j \to \infty} K(i, k_{n_j}) \text{ exists for } every\ i \in R.$$

Thus the sequence $K(i, k_{n_j})$ is Cauchy on the real line for every $i \in R$; and repeating now the argument in the second part of the proof of Lemma 5.10 ($k_{n_j} \in R^*$ now, and not necessarily to R) it follows that $\{k_{n_j}\}$ is Cauchy in (R^*, d) and, this last being complete, there exists a limit point ξ in (R^*, d) as required. □

Definition 5.7. The set $R^* - R$ is called the Martin exit boundary of R, induced by the matrix P, *relative to index* 1.[1]

1. Now let $\mathfrak{R}$ be the σ-field of Borel sets of $(R^* d)$, i.e. the minimum σ-field containing the open sets of (R^*, d). [It is obvious that each $i \in R$ is itself a closed set. Thus R itself is closed, and so $R^* - R$ is a Borel set.]

[1] By suitable rearrangement of rows and columns of P, the development of the theory entails no essential loss in generality in working in relation to index 1.

2. Further, consider $K(i, x)$ for fixed $i \in R$, as a function of $x \in (R^*, d)$. Then for $x_1, x_2 \in (R^*, d)$
$$d(x_1, x_2) < \delta \Rightarrow |K(i, x_1) - K(i, x_2)| < \varepsilon$$
from the definition of the metric, where $\varepsilon > 0$ is arbitrary, and δ appropriately, chosen. Thus $K(i, x)$ is continuous on (R^*, d); and since (R^*, d) is *compact*, $K(i, x)$ is *uniformly continuous* on (R^*, d) for i fixed.
3. Let $\{P_n(.)\}$ be any sequence of probability measures[1] on the Borel sets $\mathfrak{R}$ (i.e. measures such that $P_n(R^*) = 1$). Then since (R^*, d) is compact, there exists a subsequence $\{n_i\}$, $i \geq 1$, of the positive integers such that the subsequence $\{P_{n_i}(.)\}$ converges weakly to a limit probability measure $\{P(.)\}$. [This is the generalized analogue of the Helly selection principle for probability distributions on a closed line segment $[a, b]$. We shall use generalizations of another Helly theorem to compact metric space below. For proofs in the metric space setting, see Parthasarathy (1967), pp. 39–52; the theorem just used is on p. 45.]

Theorem 5.8. *Any superregular vector* $\boldsymbol{u} = \{u(i)\}$ *has representation*
$$u(i) = \int_{R^*} K(i, x)\mu(dx), \qquad i \in R$$
where $\mu(.)$ *is some finite measure on* $\mathfrak{R}$ *independent of* i.

PROOF. The case $\boldsymbol{u} = \boldsymbol{0}$ is trivial; assume $\boldsymbol{u} \neq \boldsymbol{0}$.

We shall first prove the proposition for $\boldsymbol{u} = \boldsymbol{h}$ where $\boldsymbol{h}$ is a finite non-zero potential. There exists a non-negative vector $\boldsymbol{k} = \{k(i)\}$ such that
$$\boldsymbol{h} = G\boldsymbol{k},$$
i.e.
$$h(i) = \sum_{j \in R} g_{ij}k(j)$$
$$= \sum_{j \in R} \frac{g_{ij}}{g_{1j}} g_{1j}k(j) = \sum_{j \in R} K(i, j)g_{1j}k(j).$$
Hence if we define a measure on $\mathfrak{R}$ by
$$\mu(\{j\}) = g_{1j}k_j, j \in R, \qquad \mu(R^* - R) = 0,$$
then we have the required representation for $\boldsymbol{h}$, since
$$\mu(R^*) = \sum_{j \in R} g_{1j}k(j) = h(1);$$
$h(1) > 0$ since for all $j \in R$, $g_{1j} > 0$ by the Basic Assumption 2; and $k(j) > 0$ for at least one $j \in R$, as $\boldsymbol{h} \neq \boldsymbol{0}$ by assumption.

Now for a superregular vector $\boldsymbol{u} \neq \boldsymbol{0}$, let $\{\boldsymbol{h}_n\}$ be a non-decreasing sequence of potentials converging to $\boldsymbol{u}$ elementwise (by Theorem 5.7) and let the

[1] The assumption of probability measures is not strictly necessary to the sequel, but is convenient; especially for the reader acquainted with some probabilistic measure theory.

corresponding measures as deduced above be $\{\mu_n(\,.\,)\}$, where

$$\mu_n(R^*) = h_n(1) > 0.$$

We know $h_n(1) \uparrow u(1)$ as $n \to \infty$, so that $u(1) > 0$. Thus

$$h_n(i) = \int_{R^*} K(i, x)\mu_n(dx)$$

where the sequence $\{\mu_n(\,.\,)\}$ does not depend on i.

Therefore $\quad u(i) = \lim_{n\to\infty} h_n(i) = \lim_{n\to\infty} \int_{R^*} K(i, x)\mu_n(dx)$

$$= \lim_{n\to\infty} h_n(1) \int_{R^*} K(i, x)\frac{\mu_n(dx)}{h_n(1)}$$

where $P_n(\,.\,) = \mu_n(\,.\,)/h_n(1)$ is a probability measure on $\mathfrak{R}$. Hence by the generalized Helly selection principle mentioned in 3 above, and by the generalized Helly–Bray Lemma, since $K(i, x)$ is bounded and continuous in $x \in R^*$, for fixed $i \in R$

$$u(i) = u(1) \int_{R^*} K(i, x)P(dx)$$

where $P(\,.\,)$ is the limit probability measure on $\mathfrak{R}$ of an appropriate subsequence $\{P_{n_i}(\,.\,)\}$.

The theorem now follows by putting

$$\mu(\,.\,) = u(1)P(\,.\,). \qquad \square$$

5.6 Example[1]

We consider the infinite stochastic matrix P, defined on the index set $R = \{1, 2, \ldots\}$ by

$$p_{i,\,i+1} = p_i(>0)$$
$$p_{i,\,1} = 1 - p_i = q_i(>0), \quad i \in R.$$

Thus we may represent P in the form

$$P = \begin{bmatrix} q_1 & p_1 & 0 & 0 & \cdots \\ q_2 & 0 & p_2 & 0 & \\ q_3 & 0 & 0 & p_3 & \\ \vdots & & & & \ddots \end{bmatrix}.$$

The index set R clearly forms a single essential aperiodic class, and hence P is primitive.

[1] See Example (4) of §4.1 for another view of this example.

For $k \geq 1$

$$l_{11}^{(k)} = f_{11}^{(k)} = p_0 p_1 \cdots p_{k-1} q_k, \qquad (p_0 = 1)$$
$$= (p_0 p_1 \cdots p_{k-1}) - (p_0 p_1 \cdots p_k)$$

since $q_k = 1 - p_k$. Hence

$$L_{11}(1-) \equiv F_{11}(1-) \equiv \sum_{k=1}^{\infty} f_{11}^{(k)} = p_0 - \lim_{k\to\infty} (p_0 p_1 \cdots p_k)$$
$$= p_0 - \lim_{k\to\infty} \alpha_k$$

where $\alpha_k = p_0 p_1 \cdots p_k$ has a limit $\alpha_\infty \geq 0$ as $k \to \infty$ since it is positive and decreasing with k. Hence

$$L_{11}(1-) = F_{11}(1-) = 1 - \alpha_\infty.$$

Thus index 1 (and thus every index) is transient if and only if $\alpha_\infty > 0$ (and is recurrent if $\alpha_\infty = 0$).

In the case $\alpha_\infty = 0$:

$$\mu = \sum_{k=1}^{\infty} k f_{11}^{(k)} = \sum_{k=1}^{\infty} k(\alpha_{k-1} - \alpha_k)$$
$$= \sum_{k=1}^{\infty} \sum_{j=1}^{k} (\alpha_{k-1} - \alpha_k)$$
$$= \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} (\alpha_{k-1} - \alpha_k) = \sum_{j=1}^{\infty} \alpha_{j-1}.$$

Hence

$$\mu = \sum_{j=0}^{\infty} \alpha_j,$$

and we obtain positive recurrence of the matrix P if $\sum_{j=0}^{\infty} \alpha_j < \infty$ and null-recurrence otherwise (when $\alpha_\infty = 0$).

We shall next evaluate in terms of the α_is the quantities $F_{ij}(1-)$, $i \neq j$; this is really of interest only when $\alpha_\infty > 0$, i.e. in the case of transience of the matrix P, for otherwise[1] $F_{ij}(1-) = 1$, all $i, j \in R$, but we shall not assume that $\alpha_\infty > 0$ necessarily. [The case $F_{ii}(1-)$, $i \in R$ is—at least in the transient case—more difficult to evaluate, but, as we shall see, we shall not need if for consideration of the Martin exit boundary.]

First we note that

$$F_{ij}(1-) = 1 \qquad \text{if } 1 \leq i < j$$

for this quantity is readily seen to be the "absorption probability" from index i of the set $1 \leq i < j$ [to which there corresponds a strictly substoch-

[1] See Exercise 5.4 for the general proposition.

astic primitive matrix ${}_{(j-1)}P$ (the $(j-1)\times(j-1)$ northwest corner truncation of P)] into the single "absorbing index" of remaining indices $\{j, j+1, \ldots\}$. (See §4.2).

Secondly

$$F_{ij}(1-) = F_{i1}(1-)F_{1j}(1-), \qquad \text{if } i > j \geq 1$$

which follows from the definition of $F_{ij}(1-)$ in this special situation, i.e. in view of the previous result

$$F_{ij}(1-) = F_{i1}(1-), \qquad \text{if } i > j \geq 1.$$

Thus we need only to find $F_{i1}(1-)$ for $i > 1$. Now

$$\begin{aligned} f_{i1}^{(k)} &= q_i,\ k = 1 \\ &= p_i p_{i+1} \cdots p_{i+k-2} q_{i+k-1},\ k \geq 2 \end{aligned}$$

from the special structure of P.

Hence

$$\begin{aligned} F_{i1}(1-) &= q_i + \sum_{k=2}^{\infty} [p_i p_{i+1} \cdots p_{i+k-2} - p_i p_{i+1} \cdots p_{i+k-1}] \\ &= q_i + p_i - \lim_{k\to\infty} p_i p_{i+1} \cdots p_{i+k} \\ &= q_i + p_i - \lim_{k\to\infty} \frac{p_0 p_1 \cdots p_{i+k}}{p_0 p_1 \cdots p_{i-1}} \\ &= 1 - \frac{\alpha_\infty}{\alpha_{i-1}}. \end{aligned}$$

Thus, finally

$$F_{ij}(1-) = \begin{cases} 1 & \text{if } i < j, \\ 1 - \dfrac{\alpha_\infty}{\alpha_{i-1}} & \text{if } i > j \geq 1. \end{cases}$$

Let us now pass on to the Martin boundary theory, assuming transience henceforth i.e. $\alpha_\infty > 0$.

Then

$$\boxed{\begin{aligned} \bar{u}_{ij} &= 1 & &\text{if } 1 \leq i \leq j \\ &= 1 - \frac{\alpha_\infty}{\alpha_{i-1}} & &\text{if } i > j, \end{aligned}}$$

so that

$$K(i, j) = \frac{\bar{u}_{ij}}{\bar{u}_{1j}} = \bar{u}_{ij}.$$

Hence (relative to index 1 as usual) we have the metric on $(R \times R)$

$$\begin{aligned} d(v_1, v_2) &= \sum_{i \in R} w_i \bar{u}_{1i} |K(i, v_1) - K(i, v_2)| \\ &= \sum_{i \in R} w_i |K(i, v_1) - K(i, v_2)|. \end{aligned}$$

Assume, for further working, that we are dealing with $v_1 \neq v_2$; so without loss of generality, let us assume $v_1 < v_2$; then

$$\begin{aligned} d(v_1, v_2) = \sum_{i=1}^{v_1} w_i(1 - 1) + \sum_{i=v_1+1}^{v_2} w_i \left| 1 - \frac{\alpha_\infty}{\alpha_{i-1}} - 1 \right| \\ + \sum_{i=v_2+1}^{\infty} w_i \left| 1 - \frac{\alpha_\infty}{\alpha_{i-1}} - \left(1 - \frac{\alpha_\infty}{\alpha_{i-1}}\right) \right|, \end{aligned}$$

i.e. for $v_1 < v_2$

$$\boxed{d(v_1, v_2) = \sum_{v_1 < i \leq v_2} w_i \frac{\alpha_\infty}{\alpha_{i-1}}}$$

where, of course, $0 < \alpha_\infty / \alpha_{i-1} < 1$.

Thus the metric space (R, d) is *isometric* to the metric space which is that subset of the real line (with the ordinary modulus-difference metric) consisting of points

$$v_1, v_1 + v_2, v_1 + v_2 + v_3, \ldots$$

where $$v_i = w_i \alpha_\infty / \alpha_{i-1}, \qquad i \in R.$$

Now every *non-terminating* subsequence of this new metric space converges to the same point, viz.

$$\sum_{i=1}^{\infty} v_i \equiv \sum_{i=1}^{\infty} w_i \frac{\alpha_\infty}{\alpha_{i-1}} < \infty.$$

Hence the Cauchy completion of (R, d), viz. the Martin exit boundary of (R, d) relative to index 1 induced by the matrix P, consists of a *single point*, which we may call, therefore, "infinity".

Bibliography and Discussion

Our discussion of countable stochastic matrices is rather selective; excellent treatments now exist in textbook and monograph form, and the reader desiring a sustained and more extensive development of most of the material presented in this chapter, as well as extensive bibliography, should consult the books of Feller (1968), Chung (1967) and Kemeny, Snell & Knapp (1966), expecially Chung's.

As mentioned earlier in the book, study of countable stochastic matrices was initiated by Kolmogorov (1936), by whom the notions of essential and inessential indices, which we use throughout the book, were introduced; and followed closely by many contributions of Doeblin, although we have not touched to any significant extent on this later author's work in this context.

The standard approach to the basic theory is in terms of the quantities $\{f_{ij}^{(k)}\}$. We have worked in terms of the quantities $\{l_{ij}^{(k)}\}$ partly so as not to duplicate this approach yet again, although for stochastic P it is in several ways not quite so natural as the other. The fundamental reason, however, for using the $\{l_{ij}^{(k)}\}$ is that even in the stochastic context it is the natural tool as regards the development of the probabilistically important theory of sub-invariant and invariant measures, and makes a reasonably unified treatment possible; whereas, e.g. in Feller (1968), a change of approach is necessary when passing on to the topic of invariant measures. Finally, it is necessary to mention that in §§5.1, 5.3 and 5.4, the present author was substantially influenced in approach by the papers of Vere-Jones (1962, 1967), which are not restricted to the stochastic framework, and play a substantial role in the development of the next chapter also.

Theorem 5.1′ is due to Erdös, Feller & Pollard (1949). The part of Theorem 5.4 referring to recurrent P is due to Derman (1954). A necessary and sufficient condition for the existence of an invariant measure for a transient P is due to Harris (1957) and Veech (1963). Theorem 5.5 is due to Kolmogorov (1936) and Feller (1968; 1st ed. 1950). For a probabilistic rather than analytical proof of Theorem 5.5 the reader should consult Orey (1962).

A consideration of Perron–Frobenius-type spectral properties of denumerable stochastic irreducible P as operators on various sequence spaces (i.e. from a functional-analytic point of view) has been given by Nelson (1958), Kendall (1959), Sidák (1962, 1963, 1964*a*) (see also Holmes (1966)), and Moy (1965), and these results provide an interesting comparison with those obtained in the present chapter. The reader interested in this topic should also consult the references in the notes at the conclusion of the next chapter, which refer to a more general (specifically, non-stochastic) situation.

Much of §5.5 is adapted from the papers of Doob (1959) and Moy (1967*a*). Extensive discussions of the potential theory in the same framework as in §5.5 may be found in the book of Kemeny, Snell & Knapp (1966) and the earlier exposition of Neveu (1964).

Exercises

5.1. Let i be an index with period $d > 1$ corresponding to the stochastic matrix $P = \{p_{ij}\}$. Show that in relation to the matrix

$$P^d = \{p_{ij}^{(d)}\},$$

i is aperiodic, and show also that for $0 \le s < 1$,

$$L_{ii}(s) = L_{ii}^{(d)}(s^{1/d})$$

by first showing

$$P_{ii}(s) = P_{ii}^{(d)}(s^{1/d})$$

where the superscript (d) pertains to the matrix P^d.

Use this to show that if i is positive or null recurrent, or transient in relation to P, it is also this, respectively, in relation to P^d; and that in the recurrent case, as $k \to \infty$

$$p_{ii}^{(kd)} \to d/\mu_i.$$

5.2. Use the final result of Exercise 5.1 to conclude that for an irreducible stochastic matrix P (not necessarily aperiodic) for all $i, j = 1, 2, \dots$

$$p_{ij}^{(k)} \to 0$$

if the matrix P is transient, or null-recurrent; and that this occurs for *no* pair (i, j) if P is positive recurrent.

5.3. Show, if i is an index corresponding to a *finite* stochastic matrix P, that i is transient if and only if i is inessential; and is positive recurrent otherwise.

5.4. Develop in full the "dual" approach to the theory as indicated in §5.4. Prove in particular the analogues of Theorems 5.3 and 5.4. Prove also Theorem 5.5 from this new viewpoint, noting in the course of this that for a recurrent stochastic P, $F_{ij}(1-) = 1$ for all $i, j = 1, 2, \dots$

5.5 Show that a stochastic irreducible matrix P is positive-recurrent if and only if there exists a *subinvariant* measure x' for P such that $x'\mathbf{1} < \infty$. Repeat for an invariant measure.

5.6. An infinite stochastic irreducible $P = \{p_{ij}\}$ satisfies

$$\sum_{j=1}^{N} p_{ij} \ge \delta > 0$$

uniformly for all i in the index set of P, where N is some fixed positive integer. Show that P is positive-recurrent.

Hint: Consider $\sum_{j=1}^{N} p_{ij}^{(k)}$.

5.7. For the Example to which §5.6 is devoted, attempt to solve the invariant equations $x'P = x'$, and hence deduce a necessary and sufficient condition for positive-recurrence of this particular P, using the result of Exercise 5.5. Does this concur with the results of §5.6 on classification?

Find μ_i for each $i = 1, 2, \dots$ in the positive-recurrent case.

5.8. Let $P = \{p_{ij}\}$ and $P_n = \{p_{ij}(n)\}$, $n = 1, 2, \dots$ be transition probability matrices of irreducible recurrent Markov chains each defined on the states $\{1, 2, \dots\}$, and let $\{v_i\}$ and $\{v_i^{(n)}\}$, $n = 1, 2, \dots$ be the corresponding invariant measures, normalized so that $v_1 = v_1^{(n)} = 1$ for all n. Assume that $p_{ij}(n) \to p_{ij}$ for all i, j as $n \to \infty$.

Show that $\{v_i^*\}$, defined by $v_i^* = \liminf_{n\to\infty} v_i^{(n)}$, $i = 1, 2, \ldots$, is a subinvariant measure corresponding to P; and hence show that

$$v_i^* = \lim_{n\to\infty} v_i^{(n)}, \qquad i = 1, 2, \ldots$$

Finally, deduce that $v_i = v_i^*$, $i = 1, 2, \ldots$

5.9. If P is an infinite recurrent doubly stochastic matrix, show that it must be null-recurrent.

5.10. Show that the structure of the Martin exit boundary $B = R^* - R$ does not depend on the choice of weights $\{w_i\}$ in the definition of the metric $d(v_1, v_2)$ on $(R \times R)$, so long as $w_i > 0$ all i, $\sum_{i\in R} w_i < \infty$. (Use Lemma 5.10.)

5.11. Let $P = \{p_{ij}\}$ be an irreducible substochastic matrix on $R \times R$ with one index transient. Show that all $i \in R$ are also transient, and that $x' = \{x_i\}$ defined by $x_i = g_{1i}/g_{11}$ is a strictly positive superregular *row* vector for P. Further show that the matrix $\hat{P} = \{\hat{p}_{ij}\}$, where

$$\hat{p}_{ij} = x_j p_{ji}/x_i, \qquad i, j \in R$$

is irreducible, substochastic, and all its indices are transient.

Hint: Show $$\hat{g}_{ij} = x_j g_{ji}/x_i.$$

The Martin *entrance* boundary (relative to index 1) for matrix P is the Martin exit boundary for the matrix $\hat{P}$.

Show that[1] for $i, j \in R$

$$\hat{K}(i, j) = K(j, i)/K(j, 1).$$

5.12. Using the results of Exercise 5.11 in relation to P as given in §5.6, with $\alpha_\infty > 0$, show that

$$\hat{K}(i, j) = \begin{cases} 1, & j = 1, \quad i \geq 1 \\ \left[1 - \dfrac{\alpha_\infty}{\alpha_{j-1}}\right]^{-1}, & 1 < j \leq i \\ 1, & j > i \geq 1 \end{cases}$$

and hence that the space $(\hat{R}, \hat{d})$ is *itself complete*, so that the Martin *entrance* boundary of P, corresponding to index 1, is empty.

5.13. The *substochastic* irreducible matrix $P = \{p_{ij}\}$ on $R = \{1, 2, \ldots\}$ described by

$$p_{j,j+1} = p_j > 0, \qquad p_{j,j-1} = q_j > 0, \qquad p_{j,j} = 1 - p_j - q_j \geq 0$$

for $j \geq 1$, describes the "transitions" between the inessential indices (those of R) of a (space inhomogeneous) random walk on the non-negative integers with "absorbing barrier" at the "origin", 0. For this matrix, it can be shown[2] that

[1] Capped symbols refer to the situation in relation to matrix $\hat{P}$.

[2] Seneta (1967*b*).

$G = \{g_{ij}\}$ is given by

$$g_{ij} = \begin{cases} \bar{q}_i \dfrac{\sum_{s=0}^{j-1} \rho_s}{q_j \rho_{j-1}}, & 1 \le j \le i \\[2ex] \bar{q}_j \dfrac{\sum_{s=0}^{i-1} \rho_s}{q_j \rho_{j-1}}, & 1 \le i \le j \end{cases}$$

where

$$\bar{q}_i = \frac{\sum_{s=i}^{\infty} \rho_s}{\sum_{s=0}^{\infty} \rho_s}, \qquad \rho_0 = 1, \qquad \rho_i = \frac{q_1 q_2 \cdots q_i}{p_1 p_2 \cdots p_i}$$

and $\bar{q}_i$ (the "absorption probability" into the origin from $i \ge 1$) is to be understood as unity if $\sum \rho_s$ diverges.

Calculate $K(i, j)$, $i, j \in R$, and hence deduce the structure of the Martin exit boundary.

Repeat with the Martin entrance boundary, using the comments of Exercise 5.11, and $\hat{K}(i, j)$.

CHAPTER 6
Countable Non-negative Matrices

Countable stochastic matrices P, studied in the previous chapter, have two (related) essential properties not generally possessed by countable non-negative matrices $T = \{t_{ij}\}$, $i, j = 1, 2, \ldots$. In the first instance the powers P^k, $k \geq 1$, are all well defined (using the obvious extension of matrix multiplication); secondly the matrix P (and its powers) have row sums unity.

It is inconvenient in the general case to deal with (as usual elementwise finite) matrices T whose powers may have infinite entries. For the sequel we therefore make the

Basic Assumption 1: $T^k = \{t_{ij}^{(k)}\}$, $k \geq 1$, are all elementwise finite.

The second role played by the "row-sums unity" assumption involved in the stochasticity of P, is that this last may be expressed in the form $P\mathbf{1} = \mathbf{1}$, whence we may expect that even in the case where P is actually infinite, our study of Perron–Frobenius-type structure of P may still be centred about the "eigenvalue" unity, and the "right eigenvector" playing a role similar to the finite case will be $\mathbf{1}$. The reader examining the details of the last chapter will see that this is in fact the approach which was adopted.

In the case of general non-negative T the asymmetric stochasticity assumption is absent, and we may, in the first instance, expect to need to resolve the problem (restricting ourselves to the irreducible case) of the natural analogue of the Perron–Frobenius eigenvalue. Indeed the resolution of this problem, and associated problems concerning "eigenvectors", makes this theory more fundamental even for transient stochastic P (as will be seen), than that of the last chapter.

For convenience we shall confine ourselves to those T which satisfy the

Basic Assumption 2: T is irreducible.

Recall that in the countable context irreducibility is defined precisely as in the finite case, viz. for each $i, j = 1, 2, \ldots$ there exists $k \equiv k(i, j)$ such that $t_{ij}^{(k)} > 0$. It also follows as in Chapter 1, that all indices of T have a common finite period $d \geq 1$ (for further detail of this kind see §5.1).

6.1 The Convergence Parameter R, and the R-Classification of T

Theorem 6.1. *The power series*

$$T_{ij}(z) = \sum_{k=0}^{\infty} t_{ij}^{(k)} z^k, \qquad i, j = 1, 2, \ldots$$

all have common convergence radius R, $0 \leq R < \infty$, for each pair i, j.

Proof. Denote the convergence radius of $\sum_{k=0}^{\infty} t_{ij}^{(k)} z^k$ by R_{ij}. On account of the non-negativity of the coefficients in the power series,

$$R_{ij} = \sup_{s \geq 0} \left(s: \sum_k t_{ij}^{(k)} s^k < \infty \right).$$

We now consider the obvious inequalities, frequently used hitherto,

$$t_{ij}^{(v+k)} \geq t_{ij}^{(v)} t_{jj}^{(k)}$$

$$t_{jj}^{(v+k)} \geq t_{ji}^{(v)} t_{ij}^{(k)}$$

$$t_{ii}^{(v+k)} \geq t_{ij}^{(k)} t_{ji}^{(v)}$$

$$t_{ij}^{(v+k)} \geq t_{ii}^{(k)} t_{ij}^{(v)}.$$

The first inequality clearly implies that

$$R_{ij} \leq R_{jj}$$

(consider v constant and such that $t_{ij}^{(v)} > 0$, using irreducibility; and form the relevant power series), while the second analogously implies that

$$R_{jj} \leq R_{ij}.$$

Thus we have that $R_{ij} = R_{jj}$ for all $i, j = 1, 2, \ldots$. The third and fourth inequalities imply, respectively, that

$$R_{ii} \leq R_{ij}, \qquad R_{ij} \leq R_{ii}$$

so that $R_{ij} = R_{ii}$ for all $i, j = 1, 2, \ldots$.

Thus we may write R for the common value of the R_{ij}.

It remains to show only that $R < \infty$. This can be seen via Lemma A.4 of Appendix A, which implies, in view of

$$t_{ii}^{(k+r)d} \geq t_{ii}^{(kd)} t_{ii}^{(rd)}, \qquad k, r \geq 0$$

where d is the period of i, that

$$\lim_{n\to\infty} \{t_{ii}^{(nd)}\}^{1/n}$$

exists and is positive and so

$$\lim_{n\to\infty} \{t_{ii}^{(nd)}\}^{1/nd} \equiv \limsup_{k\to\infty} \{t_{ii}^{(k)}\}^{1/k}$$

exists and is positive, and by a well-known theorem[1] the limit is $1/R_{ii} \equiv 1/R$.

□

From Lemma A.4 of Appendix A we obtain the additional information:

Corollary. $t_{ii}^{(nd)} \leq R^{-nd}$

i.e.

$$R^k t_{ii}^{(k)} \leq 1$$

for all $i = 1, 2, \ldots, k \geq 0$; *and*

$$\lim_{n\to\infty} \{t_{ii}^{(nd)}\}^{1/nd} = R^{-1}.$$

N.B. It is possible to avoid the use of Lemma A.4 altogether here,[2] through an earlier introduction of the quantities $F_{ij}(z)$, $L_{ij}(z)$ which we carry out below; its use has been included for unity of exposition, and for the natural role which it plays in this theory.

Definition 6.1. The common convergence radius, R, $0 \leq R < \infty$, of the power series

$$\sum_{k=0}^{\infty} t_{ij}^{(k)} z^k$$

is called the *convergence parameter* of the matrix T.

Basic Assumption 3: We deal only with matrices T for which $R > 0$.

Note that for a countable substochastic (including stochastic) matrix $T = P$, $R \geq 1$, clearly. Further, it is clear if T is finite, that $1/R = r$, the Perron–Frobenius eigenvalue of T (in the aperiodic case this is obvious from Theorem 1.2 of Chapter 1, and may be seen to be so in the period d case by considering a primitive class of T^d). For this reason the quantity $1/R$ is sometimes used in the general theory, and bears the name of *convergence norm* of T.

To proceed further, we define quantities $f_{ij}^{(k)}$, $l_{ij}^{(k)}$ as in the last chapter; write $f_{ij}^{(0)} = l_{ij}^{(0)} = 0$, $f_{ij}^{(1)} = l_{ij}^{(1)} = t_{ij}$ and thereafter write inductively

$$f_{ij}^{(k+1)} = \sum_{\substack{r \\ r \neq j}} t_{ir} f_{rj}^{(k)}, \qquad l_{ij}^{(k+1)} = \sum_{\substack{r \\ r \neq i}} l_{ir}^{(k)} t_{rj} \qquad i, j = 1, 2, \ldots$$

[1] e.g. Titchmarsh (1939, §7.1).

[2] See Exercise 6.1.

Clearly $f_{ij}^{(k)}, l_{ij}^{(k)} \le t_{ij}^{(k)}$, and so

$$F_{ij}(z) = \sum_{k=0}^{\infty} f_{ij}^{(k)} z^k, \qquad L_{ij}(z) = \sum_{k=0}^{\infty} l_{ij}^{(k)} z^k$$

are convergent for $|z| < R$, and as in the previous chapter

$$T_{jj}(z)F_{ij}(z) = T_{ii}(z)L_{ij}(z) = T_{ij}(z), \qquad i \ne j;$$
$$L_{ii}(z) = F_{ii}(z); \qquad T_{ii}(z) = (1 - L_{ii}(z))^{-1}$$

for $|z| < R$; the very last equation follows from

$$T_{ii}(z) - 1 = T_{ii}(z)L_{ii}(z)$$

for $|z| < R$, through the fact that for $|z| < R$

$$|L_{ii}(z)| \le L_{ii}(|z|)$$

and for $0 \le s < R$

$$L_{ii}(s) = 1 - [T_{ii}(s)]^{-1} < 1$$

as before. We thus have, letting $s \to R-$, that

$$L_{ii}(R-) \equiv F_{ii}(R-) \le 1.$$

Finally let us put

$$\mu_i(R) = RL'_{ii}(R-) = \sum_k k l_{ii}^{(k)} R^k.$$

Definition 6.2. An index i is called *R-recurrent* if $L_{ii}(R-) = 1$ and *R-transient* if $L_{ii}(R-) < 1$.

An *R-recurrent* index i is said to be *R-positive* or *R-null* depending as $\mu_i(R) < \infty$ or $\mu_i(R) = \infty$ respectively.

Clearly, an index i is R-transient if and only if $T_{ii}(R-) < \infty$; in this case $t_{ii}^{(k)} R^k \to 0$ as $k \to \infty$. To go further, we can adapt the results of Chapter 5 for our convenience in the following way: proceeding as in Lemma 5.4, we obtain for $|z| < R$ that

$$L_{ij}(z) = z \sum_r L_{ir}(z) t_{rj} + z t_{ij}(1 - L_{ii}(z))$$

which yields eventually that $L_{ij}(R-) < \infty$ for all $i, j (i \leftrightarrow j$ because of irreducibility) and as in Lemma 5.5 we obtain that there is always a row vector $\boldsymbol{x}', \boldsymbol{x}' \ge \boldsymbol{0}', \ne \boldsymbol{0}'$ satisfying

$$R\boldsymbol{x}'T \le \boldsymbol{x}' \tag{6.1}$$

one such being given by the vector $\{L_{ij}(R-)\}$, $j = 1, 2, \ldots$ [for arbitrary fixed i].

Definition 6.3. Any $x' \geq \mathbf{0}'$, $\neq \mathbf{0}'$ satisfying

$$Rx'T \leq x'$$

is called an *R-subinvariant* measure. If in fact $Rx'T = x'$, x' is called an *R-invariant* measure.

As in Lemma 5.6 any R-subinvariant measure has all its entries positive.

Now for a given R-subinvariant measure $x' = \{x_i\}$ define an, in general, substochastic matrix $P = \{p_{ij}\}$ by

$$p_{ij} = Rx_j t_{ji}/x_i, \qquad i, j = 1, 2, \ldots$$

so that $P^k = \{p_{ij}^{(k)}\}$ has its elements given by

$$p_{ij}^{(k)} = R^k x_j t_{ji}^{(k)}/x_i.$$

It is readily checked P is irreducible. Let us then consider two cases.

(1) P is strictly substochastic, i.e. at least one row sum is less than unity. Then as at the outset of §5.5, each of the indices $i = 1, 2, \ldots$ may be considered inessential in an enlarged matrix $\bar{P}$, hence transient from Lemma 5.2, whence for each $i = 1, 2, \ldots$

$$\sum_k p_{ii}^{(k)} < \infty. \tag{6.2}$$

We shall in the rest of this section include this case with the case where P is *stochastic* and *transient*, in which case (6.2) holds also.

(2) P is stochastic, in which case it may be transient, null-recurrent or positive-recurrent.

It can be seen that the transformation will enable us to fall back, to a large extent, on the results presented in Chapter 5 by considering the matrix P, and then translating the results for the matrix T.

Thus for example an index i of P is transient if and only if

$$\sum_{k=0}^{\infty} p_{ii}^{(k)} < \infty; \qquad \Leftrightarrow \sum_{k=0}^{\infty} t_{ii}^{(k)} R^k < \infty,$$

i.e. if and only if i in T is R-transient. Moreover since P (like T) is irreducible, all its indices are either transient or recurrent; *thus all indices of T must be R-transient, or all must be R-recurrent.* The last case can only occur if P is stochastic.

Further note that for $|z| < 1$

$$[1 - L_{ii}^{(P)}(z)]^{-1} = P_{ii}(z) = \sum_{k=0}^{\infty} p_{ii}^{(k)} z^k = \sum_{k=0}^{\infty} t_{ii}^{(k)} R^k z^k$$

$$= [1 - L_{ii}^{(T)}(Rz)]^{-1}$$

where the superscripts P and T indicate which matrix is being referred to. Hence for $0 \le s < 1$

$$\frac{1 - L_{ii}^{(P)}(s)}{1 - s} = \frac{1 - L_{ii}^{(T)}(Rs)}{1 - s}$$

and letting $s \to 1-$ in the R-recurrent case of T ($=$ the recurrent case of P)

$$L_{ii}^{(P)\prime}(1-) = RL_{ii}^{(T)\prime}(R-).$$

Thus i in P is positive recurrent if and only if i in T is R-positive, etc., and since all i in recurrent P are positive recurrent, or all are null-recurrent, for *R-recurrent T all indices are R-positive or R-null.* Thus it is now sensible to make the:

Definition 6.4. T is called R-transient, R-positive or R-null if one of its indices is R-transient, R-positive or R-null respectively.

It is clear, further, that if T is R-transient or R-null, for any i, j, $R^k t_{ij}^{(k)} \to 0$ as $k \to \infty$.

If T is R-positive,

$$R^k t_{ij}^{(k)}$$

is clearly still bounded as $k \to \infty$ for any fixed pair of indices, (i, j) since

$$x_i R^k t_{ij}^{(k)}/x_j = p_{ji}^{(k)} \le 1.$$

On the other hand, no matter what the R-classification of T,

$$\beta^k t_{ij}^{(k)}$$

is unbounded as $k \to \infty$ for any pair (i, j), if $\beta > R$. For suppose not.

Then
$$\beta^k t_{ij}^{(k)} \le K_{ij} = \text{const.}$$

for all k. Let z satisfy $|z| < \beta$. Then clearly

$$T_{ij}(z) = \sum_{k=0}^{\infty} z^k t_{ij}^{(k)}$$

converges for such z and hence converges for z satisfying $\beta > |z| > R$, which are outside the radius of convergence of $T_{ij}(z)$; which is a contradiction.

In conclusion to this section we note that the *common convergence radius* of all $T_{ij}(z)$, and the *common R-classification of all indices* in an irreducible T are further examples of so-called *solidarity* properties, enjoyed by all indices of an irreducible non-negative matrix (another example is the common period d of all indices).

6.2 R-Subinvariance and Invariance; R-Positivity

As in Theorem 5.3 if $x' = \{x_i\}$ is any R-subinvariant measure corresponding to our (irreducible, $R > 0$) T, then for fixed but arbitrary i, and all $j = 1, 2, \ldots$

$$x_j/x_i \geq \bar{x}_{ij}$$

where
$$\bar{x}_{ij} = (1 - \delta_{ij})L_{ij}(R-) + \delta_{ij}$$

and $\{\bar{x}_{ij}\}$, $j = 1, 2, \ldots$ is also an R-subinvariant measure [with ith element unity]. We then have in the same manner as before for Theorem 5.4:

Theorem 6.2. *For an R-recurrent matrix T an R-invariant measure always exists, and is a constant multiple of* $\{\bar{x}_{ij}\}$, $j = 1, 2, \ldots$.

An R-subinvariant measure which is not invariant exists if and only if T is R-transient; one such is then given by $\{\bar{x}_{ij}\}$, $j = 1, 2, \ldots$.

N.B. The analogous discussion with the quantities $F_{ij}(R-)$ would not be, in the present context, essentially different from the above, since there is no "asymmetric" assumption such as stochasticity present, which tends to endow T and T' with somewhat divergent properties. (In any case it is easily shown that the common convergence radius R is the same for T and T', and the subinvariant *vector* properties of T may be evolved from the subinvariant *measure* properties of T'.)

We now trivially extend the scope of our investigations of subinvariance, by saying:

Definition 6.5. $x' = \{x_i\}$, $x' \geq \mathbf{0}'$, $\neq \mathbf{0}'$ is a *β-subinvariant* measure for $\beta > 0$, if

$$\beta x'T \leq x'$$

elementwise. The definition of β-invariance is analogous.

Theorem 6.3.

(*a*) *If* $x' = \{x_i\}$ *is a β-subinvariant measure, then* $x > \mathbf{0}$ *and*

$$x_j/x_i \geq \delta_{ij} + (1 - \delta_{ij})L_{ij}(\beta);$$

(*b*) $L_{ii}(\beta) \leq 1$ *for all i and* $\beta \leq R$;

(*c*) *for* $\beta \leq R$, $L_{ij}(\beta) < \infty$ *for all i, j and for fixed i* $\{L_{ij}(\beta)\}$ *constitutes a left β-subinvariant measure, remaining subinvariant if* $L_{ii}(\beta)$ *is replaced by unity;*

(*d*) *no β-subinvariant measure can exist for* $\beta > R$.

Proof. In view of the theory of this chapter and Chapter 5, the only proposition which is not clear from already established methods is (*d*), and we shall prove only this.

Suppose $x' \geq 0', \neq 0'$ is a β-subinvariant measure for $\beta > R$. Then since

$$\beta^k x' T^k \leq x'$$

we have for a $\bar{\beta}$ such that $R < \bar{\beta} < \beta$

$$x'\left(\sum_{k=0}^{\infty} (\bar{\beta}T)^k\right) \leq (1 - \bar{\beta}\,|\,\beta)^{-1}x'.$$

But, since $\bar{\beta} > R$ the convergence radius of $\sum_{k=0}^{\infty} t_{ij}^{(k)}z^k$, the left hand side cannot be elementwise finite, which is a contradiction. □

We now pass on to a deeper study of the interaction of β-subinvariance and invariance properties as they relate to the important case of R-positivity of T.

Theorem 6.4. *Suppose $x' = \{x_i\}$ is a β-invariant measure and $y = \{y_i\}$ a β-invariant vector of T. Then T is R-positive if*

$$y'x = \sum_i y_i x_i < \infty,$$

in which case $\beta = R$, x' is (a multiple of) the unique R-invariant measure of T and y is (a multiple of) the unique R-invariant vector of T.

Conversely, if T is R-positive, and x', y are respectively an invariant measure and vector, then $y'x < \infty$.

PROOF. We have

$$\beta x'T = x', \qquad \beta T y = y$$

where $0 < \beta$, and $\beta \leq R$, the last following from Theorem 6.3(*d*), [and $x > 0$, $y > 0$ from Theorem 6.3(*a*)]. Form the stochastic matrix $P = \{p_{ij}\}$ where

$$p_{ij} = \beta x_j t_{ji}/x_i.$$

Suppose first $\sum_i x_i y_i < \infty$, and norm so that $\sum_i x_i y_i = 1$; and put $v = \{v_i\}$, where $v_i = x_i y_i$. Then

$$\sum_i v_i p_{ij} = x_j \beta \sum_i t_{ji} y_i = x_j y_j = v_j$$

for each i. Thus

$$v'P = v'$$

and $v'\mathbf{1} = 1$. Thus irreducible stochastic P has a stationary distribution, and by the Corollary to Theorem 5.5 this is possible only if P is positive-recurrent.

Hence

$$p_{ij}^{(k)} = \beta^k t_{ji}^{(k)} x_j / x_i \not\to 0$$

as $k \to \infty$; since for $\beta < R$, $\beta^k t_{ji}^{(k)} \to 0$, from the definition of R, and as we know $\beta \leq R$, it must follow $\beta = R$, and that T is in fact R-positive, for otherwise $R^k t_{ji}^{(k)} \to 0$ as $k \to \infty$. The rest follows from Theorem 6.2.

Conversely, assume T is R-positive, and x' and y are respectively an R-invariant measure and vector. Form the stochastic matrix $P = \{p_{ij}\}$ where

$$p_{ij} = Rx_j t_{ji}/x_i.$$

Since T is R-positive, P is positive-recurrent, and so has a unique invariant measure (to constant multiples) v, which moreover satisfies $v'\mathbf{1} < \infty$. (Theorem 5.5 and its periodic version.)

Now consider the row vector $\{x_i y_i\}$; this is in fact an invariant measure for P, since

$$\sum_i x_i y_i p_{ij} = Rx_j \sum_i t_{ji} y_i = x_j y_j$$

for each $j = 1, 2, \ldots$. Hence

$$\sum_j x_j y_j < \infty. \qquad \square$$

Theorem 6.5. *If T is an aperiodic R-positive matrix, then as $k \to \infty$*

$$R^k t_{ij}^{(k)} \to x_j y_i \Big/ \sum_j x_j y_j, \qquad > 0,$$

where x', y are an invariant measure and vector of T respectively.

PROOF. Form the usual positive-recurrent stochastic matrix $P = \{p_{ij}\}$ (which is now aperiodic)

$$p_{ij} = Rx_j t_{ji}/x_i,$$

so that $p_{ij}^{(k)} = R^k x_j t_{ji}^{(k)}/x_i$.

From the body of the proof of the second part of the last theorem

$$\frac{R^k x_j t_{ji}^{(k)}}{x_i} = p_{ij}^{(k)} \to v_j = \frac{x_j y_j}{\sum_j x_j y_j}$$

since $v = \{v_j\}$ constitutes the unique invariant measure of P normed to sum to unity (Theorem 5.5). The assertion follows immediately. $\square$

6.3 Consequences for Finite and Stochastic Infinite Matrices

(*i*) Finite Irreducible T

It has already been mentioned that in this case $1/R = r$, where r is the Perron–Frobenius eigenvalue of T. Clearly, since $T^k/r^k \not\to 0$, such T are R-positive; the unique R-invariant measure and vector are of course the left and right Perron–Frobenius eigenvectors corresponding to the eigenvalue r. A consequence of the new theoretical development is that for each i,

$$L_{ii}(r^{-1}) = 1.$$

Theorem 6.5 is of course a weak version of Theorem 1.2 of Chapter 1. [The question of when the rate of convergence is (uniformly) geometric in the case of infinite T does not have a simple answer—see the discussion following Theorem 6.6 in the next section, and the Bibliography and Discussion.]

(*ii*) Infinite Irreducible Stochastic Matrices

For $T = P$, where P is an infinite irreducible stochastic matrix, it has already been noted that $R \geq 1$. Viewing the development of those results of §5.3 which relate to the problem of subinvariant measures of such a P, without specialization to recurrent P, such as Theorem 5.3, it is seen without difficulty from Theorem 6.3 that in essence the discussion there pertains in general to β-subinvariance, where $\beta \equiv 1 \leq R$, this choice of β having been to some extent dictated by analogy with finite stochastic P, where $1 = 1/r = R$ (although of course invariance itself is of physical significance in Markov chain theory[1]).

However it is now clear from the earlier section of this chapter that in fact β-subinvariance, with $\beta = 1$, may in general be of less profound significance than R-subinvariance if it may happen that $R > 1$. On the other hand since finite irreducible stochastic P are always R-positive, with $R = 1/r = 1$, we do not expect to be able to improve on the results of Chapter 5 for infinite R-positive T (and perhaps also R-null T) where we would expect $R = 1$.

Theorem 6.6. *If an irreducible stochastic P is positive-recurrent or null-recurrent, it is R-positive or R-null (respectively) with $R = 1$. Conversely an R-positive or R-null stochastic P with $R = 1$ implies (respectively) positive-recurrence and null-recurrence.*

There exist stochastic P which are transient with $R > 1$; such P may still be R-positive.

PROOF. If P is recurrent $L_{ij}(1-) = 1$ for any index i, by definition of recurrence. On the other hand $L_{ii}(R-) \leq 1$ and for $0 < \beta < R$, clearly

$$L_{ii}(\beta-) < L_{ii}(R-).$$

Hence $R = 1$; and consequently P is recurrent. The rest of the first assertion is trivial, as is its converse.

We demonstrate the final part of the theorem[2] by the example discussed in §5.6, with

$$p_i = \{\tfrac{1}{2} + (\tfrac{1}{2})^{i+1}\}/\{\tfrac{1}{2} + (\tfrac{1}{2})^i\}, \qquad i \geq 1.$$

[1] See Derman (1955).

[2] See Exercise 6.5 for an example of a transient P which has $R > 1$ and is R-transient.

Then

$$l_{11}^{(k)} = f_{11}^{(k)} = p_0 p_1 \cdots p_{k-1}(1 - p_k), \qquad (p_0 = 1)$$
$$= (\tfrac{1}{2})^{k+1}.$$

Therefore $$F_{11}(z) = \sum_{k=1}^{\infty} (\tfrac{1}{2})^{k+1} z^k = \left(\frac{z}{4}\right)\left(1 - \frac{z}{2}\right)^{-1}.$$

Hence $F_{11}(1) = \frac{1}{2}$ so that P is transient.

Moreover $$P_{11}(z) = [1 - F_{11}(z)]^{-1} = 2(2 - z)(4 - 3z)^{-1}$$

at least for $|z| < 1$. Hence by analytic continuation, the power series $P_{11}(z)$ has convergence radius $\frac{4}{3}$ so that

$$R = \tfrac{4}{3}\ (> 1).$$

Also for $0 < s < \frac{4}{3}$

$$F'_{11}(s) = (2 - s)^{-2}$$

so $$F'_{11}(\tfrac{4}{3}-) = \tfrac{9}{4} < \infty$$

so that P is R-positive. □

Transient stochastic P with $R > 1$ are sometimes called *geometrically transient* in virtue of the following result.

Theorem 6.7. *A stochastic irreducible P satisfies*

$$p_{ij}^{(k)} = 0(\delta^k), \tag{6.3}$$

as $k \to \infty$, for some δ, $0 < \delta < 1$, and for some fixed pair of indices i and j, if and only if P has convergence parameter $R > 1$. In this case (6.3) *holds uniformly in δ for each pair i, j, for any fixed δ satisfying $R^{-1} \le \delta < 1$; and for no $\delta < R^{-1}$ for any pair (i, j).*

Proof. Consider the power series

$$P_{ij}(z) = \sum_{k=0}^{\infty} p_{ij}^{(k)} z^k.$$

Since $p_{ij}^{(k)} \le \text{const } \delta^k$, it follows that the power series converges for $|\delta z| < 1$, i.e. $|z| < \delta^{-1}$, and since $\delta^{-1} > 1$ it follows that the common convergence radius R of all such series satisfies $R \ge \delta^{-1} > 1$ [and so, by Theorem 6.6, P is transient]. Conversely if P has convergence parameter $R > 1$, then since for any fixed pair i, j

$$p_{ij}^{(k)} R^k$$

is bounded as $k \to \infty$, it follows

$$p_{ij}^{(k)} = 0(R^{-k})$$

as $k \to \infty$. Moreover this is true for any pair of indices i, j. If $\delta < R^{-1}$, $\delta^{-k} p_{ij}^{(k)}$ is unbounded as $k \to \infty$ {since $\beta^k t_{ij}^{(k)}$ is for any $\beta > R$}. □

Note that in general, however, in this situation of geometric transience one can only assert that

$$p_{ij}^{(k)} \leq C_{ij} R^{-k}$$

where C_{ij} is a constant which *cannot* be replaced by a uniform constant for all pairs of indices (i, j).

Corollary. *If P is recurrent it is not possible to find a pair of indices* (i, j) and a δ satisfying $0 < \delta < 1$ such that

$$p_{ij}^{(k)} = 0(\delta^k), \qquad \text{as } k \to \infty.$$

{This corollary is of course of chief significance for null-recurrent P.}

The analogous question to these considerations for positive recurrent matrices P (R-positive, with $R = 1$, stochastic P) in case P is also aperiodic, is that of whether and when the convergence:

$$|p_{ij}^{(k)} - v_j| \to 0 \tag{6.4}$$

(see Theorems 5.5 and 6.5) where $\boldsymbol{v}' = \{v_j\}$ is the unique stationary measure of P satisfying $\boldsymbol{v}'\mathbf{1} = 1$, takes place at a geometric rate for every pair of indices (i, j). This situation is called *geometric ergodicity*; it is known that if a geometric convergence rate to zero obtains for *one pair of indices* (i, j), the same geometric convergence rate is applicable to all indices, so that geometric ergodicity is a (uniform) *solidarity result* for a positive recurrent aperiodic stochastic set of indices. However it *does not always* hold in such a situation; i.e. positive recurrent aperiodic stochastic P exist for which the rate in (6.4) is geometric for *no pair* of indices (i, j).

6.4 Finite Approximations to Infinite Irreducible T

Since the theory evolved in this chapter to this point for non-negative T satisfying our basic assumptions coincides with Perron–Frobenius results if T is actually finite, and is thus a natural extension of it to the countable case, a question of some theoretical as well as computational interest pertains to whether, in the case of an infinite T, its Perron–Frobenius-type structure is reflected to some extent in its $(n \times n)$ northwest corner truncations ${}_{(n)}T$, and increasingly so as $n \to \infty$.

To make this investigation easier, it seems natural to make a further assumption for this section, that at least the irreducibility structure of T should be to some extent reflected in the structure of the truncation ${}_{(n)}T$:

Assumption 4: All but at most a finite number of truncations ${}_{(n)}T$, $n \geq 1$, are irreducible.[1]

We shall adopt the convention that a preceding subscript (n) refers to the truncation ${}_{(n)}T$; thus e.g. (for sufficiently large n) ${}_{(n)}R$ is the convergence parameter of ${}_{(n)}T$, i.e. the Perron–Frobenius eigenvalue of ${}_{(n)}T$ is ${}_{(n)}R^{-1}$.

Theorem 6.8. ${}_{(n+1)}R < {}_{(n)}R$ *for all irreducible* ${}_{(n)}T$. ${}_{(n)}R \downarrow R$ *as* $n \to \infty$.

PROOF. Let ${}_{(n)}y = \{{}_{(n)}y_i\}$ be the positive right Perron–Frobenius eigenvector of ${}_{(n)}T$; then the matrix ${}_{(n)}P = \{{}_{(n)}p_{ij}\}$ defined by

$${}_{(n)}p_{ij} = {}_{(n)}R\ {}_{(n)}t_{ij}\ {}_{(n)}y_j / {}_{(n)}y_i$$

is stochastic and irreducible, for all n for which ${}_{(n)}T$ is irreducible.

Now the $(n \times n)$ northwest corner truncation of ${}_{(n+1)}P$ must be strictly substochastic, or ${}_{(n+1)}P$ could not be irreducible. On the other hand since this truncation has entries

$${}_{(n+1)}R\ {}_{(n+1)}t_{ij}\ {}_{(n+1)}y_j / {}_{(n+1)}y_i,$$

$i, j = 1, \ldots, n$, and for these indices ${}_{(n+1)}t_{ij} = {}_{(n)}t_{ij}$, it follows that it has positive entries in the same position as ${}_{(n)}P$ and so is irreducible; consequently *its* Perron–Frobenius eigenvalue is *strictly less* than unity. On the other hand its convergence radius is clearly ${}_{(n)}R/{}_{(n+1)}R$, since ${}_{(n+1)}t_{ij} = {}_{(n)}t_{ij}$, $i, j = 1, 2, \ldots, n$.

Therefore $\qquad {}_{(n)}R/{}_{(n+1)}R > 1, \qquad$ as required.

Thus ${}_{(n)}R \downarrow {}_{(\infty)}R$ for some ${}_{(\infty)}R$ satisfying $R \leq {}_{(\infty)}R < {}_{(n)}R$, since ${}_{(n)}t_{11}^{(k)} \leq t_{11}^{(k)}$ implies that $R \leq {}_{(n)}R$.

Now, we have

$${}_{(n)}R \sum_{j=1}^{n} {}_{(n)}t_{ij}\ {}_{(n)}y_j = {}_{(n)}y_i$$

where we assume that for every n (sufficiently large) $\{{}_{(n)}y_i\}$ has been scaled so that ${}_{(n)}y_1 = 1$. Put

$$y_j^* = \liminf_{n \to \infty} {}_{(n)}y_j$$

for each $j = 1, 2, \ldots$, so that $y_1^* = 1$, $\infty \geq y_j^* \geq 0$. By Fatou's Lemma

$${}_{(\infty)}R \sum_{j=1}^{\infty} t_{ij} y_j^* \leq y_i^*, \qquad i = 1, 2, \ldots.$$

Iterating this, we find

$${}_{(\infty)}R^k \sum_{j=1}^{\infty} t_{ij}^{(k)} y_j^* \leq y_i^*$$

[1] This assumption does not in fact result in essential loss of generality—see the Bibliography and Discussion to this chapter.

and taking $i = 1$, using the fact that $y_1^* = 1$ and the irreducibility of T, implies $y_j^* < \infty$ for all j. Hence $y^* = \{y_i^*\}$ is an ${}_{(\infty)}R$-subinvariant vector; hence $y^* > \mathbf{0}$, and ${}_{(\infty)}R \le R$, the last by analogy with part (*d*) of Theorem 6.3.

Hence finally

$${}_{(\infty)}R = R$$

which is as required. □

Theorem 6.9. *Let* ${}_{(n)}x' = \{{}_{(n)}x_i\}$ *and* ${}_{(n)}y = \{{}_{(n)}y_i\}$ *be the left and right Perron–Frobenius eigenvectors, normed so that* ${}_{(n)}x_1 = 1 = {}_{(n)}y_1$, *of* ${}_{(n)}T$, *where the infinite matrix T is R-recurrent. Then*

$$\lim_{n\to\infty} {}_{(n)}y_i = y_i, \qquad \lim_{n\to\infty} {}_{(n)}x_i = x_i$$

exist, for each i, and $x' = \{x_i\}$ *and* $y = \{y_i\}$ *are the unique R-invariant measure and vector, respectively, of T, normed so that* $x_1 = 1 = y_1$.

PROOF. We give this for the y only, as is adequate. From the proof of Theorem 6.8

$$y^* = \{y_i^*\}, \qquad y_i^* = \liminf_{n\to\infty} {}_{(n)}y_i$$

is always an R-subinvariant vector of T, and when T is R-recurrent (as at present) it must be the unique R-invariant vector of T with first element unity, y. (Theorem 6.2.)

Let i^* be the first index for which

$$y_{i^*}^* = \liminf_{n\to\infty} {}_{(n)}y_{i^*} < \limsup_{n\to\infty} {}_{(n)}y_{i^*} \le \infty.$$

Then there exists a subsequence $\{n_r\}$ of the integers such that

$$\bar{y}_{i^*} \equiv \lim_{r\to\infty} {}_{(n_r)}y_{i^*} = \limsup_{n\to\infty} {}_{(n)}y_{i^*} \le \infty.$$

If we repeat the relevant part of the argument in the proof of Theorem 6.8, as in the proof of this theorem, but using the subsequence $\{n_r\}$, $r = 1, 2, \ldots$ rather than that of all integers, we shall construct an R-invariant vector of T, $\bar{y} = \{\bar{y}_i\}$ with first element unity, but with

$$\bar{y}_{i^*} > y_{i^*}^* \equiv y_{i^*}$$

which is impossible, by uniqueness of the R-invariant vector y with first element unity. □

In the following theory we concentrate on how the determinantal and cofactor properties of the matrices $[{}_{(n)}I - R_{(n)}T]$ relate to those of the infinite matrix $[I - RT]$. Since ${}_{(n)}R > R$, the reader will recognize here that there is a relationship between this theory and that of §2.1 of Chapter 2. In actual fact we shall not pursue the relation to any great extent, persisting rather in the investigation of the approximative properties of the finite truncations to the Perron–Frobenius-type structure of T. Some investigation of the determinantal properties of $[{}_{(n)}I - \beta_{(n)}T]$ as $n \to \infty$, for $\beta \le R$ however occurs in the Exercises.[1]

Write,[2]

$${}_{(n)}\Delta(\beta) = \det\,[{}_{(n)}I - \beta \cdot {}_{(n)}T]; \qquad {}_{(n)}\Delta \equiv {}_{(n)}\Delta(R).$$

${}_{(n)}c_{ij}(\beta) =$ cofactor of the (i, j) entry of $[{}_{(n)}I - \beta \cdot {}_{(n)}T]$; $\qquad {}_{(n)}c_{ij} \equiv {}_{(n)}c_{ij}(R)$.

We note that for $0 \le \beta \le R$,

$${}_{(n)}\Delta(\beta) > 0$$

since $\beta < {}_{(n)}R$ [this follows from a fact already used in the proof of Theorem 1.1, that for a real square matrix H, the characteristic function $\det[\lambda I - H]$ is positive for λ exceeding the largest real root, if one such exists].

We shall now consider only the quantities ${}_{(n)}\Delta$, $\{{}_{(n)}c_{ij}\}$.

Theorem 6.10. (*a*) *As* $n \to \infty$,

$${}_{(n)}c_{ji}/{}_{(n)}c_{ii} \uparrow \bar{x}_{ij} > 0 \qquad (i, j = 1, 2, \ldots),$$

where, for fixed i, $\{\bar{x}_{ij}\}$, $j = 1, 2, \ldots$ *is the minimal left* R*-subinvariant measure*[3] *of* T.

(*b*) $\lim_{n\to\infty} {}_{(n)}\Delta \ge 0$ *exists; and if* $\lim {}_{(n)}\Delta > 0$, T is R-transient, $\lim_{n\to\infty} {}_{(n)}c_{ij} \ge 0$ *exists for every pair* (i, j), *and all these limits are positive or zero together. For an* R*-transient matrix*, ${}_{(n)}\Delta$ *and* ${}_{(n)}c_{ij}$ *have positive or zero limits together.*

Proof. Since ${}_{(n)}R > R$, and ${}_{(n)}c_{ji}/{}_{(n)}\Delta$ is the (i, j) entry of $[{}_{(n)}I - R_{(n)}T]^{-1}$, it follows from Lemma B.1 of Appendix B that

$${}_{(n)}c_{ji} = {}_{(n)}\Delta \cdot {}_{(n)}T_{ij}(R), \qquad \text{so that } {}_{(n)}c_{ji} > 0 \text{ for } i, j = 1, 2, \ldots$$

Further, when $i \ne j$, since for $|z| < {}_{(n)}R$, ${}_{(n)}T_{ij}(z) = {}_{(n)}L_{ij}(z)(1 - {}_{(n)}L_{ii}(z))^{-1}$ it follows that

$${}_{(n)}L_{ij}(R) \cdot {}_{(n)}\Delta = {}_{(n)}c_{ji}[1 - {}_{(n)}L_{ii}(R)], \qquad i \ne j$$

$${}_{(n)}\Delta = {}_{(n)}c_{ii}[1 - {}_{(n)}L_{ii}(R)].$$

Hence for $i \ne j$,

$${}_{(n)}c_{ji}/{}_{(n)}c_{ii} = {}_{(n)}L_{ij}(R)$$

and since it is clear that ${}_{(n)}l_{ij}^{(k)} \uparrow l_{ij}^{(k)}$ as $n \to \infty$,

$${}_{(n)}c_{ji}/{}_{(n)}c_{ii} \uparrow L_{ij}(R)(< \infty)$$

[1] See Exercise 6.8.

[2] Noting that the definition of these symbols differs here from that in Chapter 2, slightly.

[3] As defined at the outset of §6.2. A similar proposition will, of course, hold for R-subinvariant vectors.

as $n \to \infty$, which proves (a). [The inequality and the rest of the assertion follows from Theorems 6.2 and 6.3.]

We now pass to the proof of (b). By cofactor expansion along the $(n+1)$th column of: $[{}_{(n+1)}I - R \cdot {}_{(n+1)}T]$ *with its jth row and column deleted,*

$$ {}_{(n+1)}c_{jj} = (1 - R \cdot t_{n+1,n+1})_{(n+1)}c_{jj/n+1,n+1} - R \sum_{\substack{k=1\\k\neq j}}^{n} t_{k,n+1} \cdot {}_{(n+1)}c_{jj/k,n+1} $$

$$ = {}_{(n+1)}c_{jj/n+1,n+1} - R \sum_{\substack{k=1\\k\neq j}}^{n+1} t_{k,n+1} \cdot {}_{(n+1)}c_{jj/k,n+1}; $$

and
$$ {}_{(n+1)}c_{jj/n+1,n+1} = {}_{(n)}c_{jj}; $$

where ${}_{(n+1)}c_{jj/k,n+1}$ is the $(k, n+1)$ cofactor of this matrix. The corresponding matrix formed from ${}_{(n+1)}T$ by crossing out its jth row and column, may not be irreducible, but none of its eigenvalues can exceed $1/{}_{(n+1)}R$ in modulus.[1] Hence

$$ {}_{(n+1)}c_{jj/k,n+1} \geq 0 $$

in the same manner as in the first step of the proof of this theorem. Thus

$$ 0 < {}_{(n+1)}c_{jj} \leq {}_{(n)}c_{jj}, $$

so that

$$ \lim_{n\to\infty} {}_{(n)}c_{jj} \geq 0 $$

exists; let us put c_{jj} for the value of this limit. Thus as $n \to \infty$

$$ {}_{(n)}\Delta \downarrow \lim_{n\to\infty} {}_{(n)}c_{ii}[1 - {}_{(n)}L_{ii}(R)] $$
$$ = c_{ii}[1 - L_{ii}(R)] \geq 0 $$

exists; and if the limit is positive $L_{ii}(R) < 1$, so T is R-transient. Further since

$$ {}_{(n)}c_{ji}/{}_{(n)}c_{ii} \to \bar{x}_{ij} > 0 $$

it follows that if for some i, $c_{ii} > 0$ then the limit

$$ c_{ji} = \lim_{n\to\infty} {}_{(n)}c_{ji} $$

exists and $c_{ji} > 0$, for all $j = 1, 2, \ldots$ and if $c_{ii} = 0$,

$$ 0 = c_{ji} = \lim_{n\to\infty} {}_{(n)}c_{ji} $$

exists. Thus we have that in the matrix whose (i, j) entry is c_{ji} each row is either strictly positive or zero.

[1] See e.g. Exercise 1.12, (e').

By making the same considerations as hitherto for the *transpose* of T, T', we shall obtain the same conclusion about the columns of this same matrix, whose (i, j) entry is c_{ji}; so that finally *all* c_{ji} are positive or zero as asserted.

Finally, if T is R-transient, $L_{ii}(R) < 1$, and since

$$\lim {}_{(n)}\Delta = c_{ii}[1 - L_{ii}(R)]$$

the conclusion that $\lim {}_{(n)}\Delta$ and the c_{ij} are positive or zero together, follows. □

Corollary. *For any pair* (i, j)

$$\frac{{}_{(n)}c_{ji}}{{}_{(n)}c_{ii}} \le \bar{x}_{ij} \le \{\bar{x}_{ji}\}^{-1} \le \left|\frac{{}_{(n)}c_{ij}}{{}_{(n)}c_{jj}}\right|^{-1};$$

with both sides converging to $\bar{x}_{ij} = \{\bar{x}_{ji}\}^{-1}$ *as* $n \to \infty$, *if* T is R-recurrent.

Proof. The inequalities

$$\frac{{}_{(n)}c_{ji}}{{}_{(n)}c_{ii}} \le \bar{x}_{ij}; \qquad \{\bar{x}_{ji}\}^{-1} \le \left|\frac{{}_{(n)}c_{ij}}{{}_{(n)}c_{jj}}\right|^{-1},$$

are given by the theorem, as is the convergence in each. Let $\{x(i)\}$ be any R-subinvariant measure of T. Then by the minimality property of $\{\bar{x}_{ij}\}$

$$\bar{x}_{ij} \le x(j)/x(i) = \{x(i)/x(j)\}^{-1} \le \{\bar{x}_{ji}\}^{-1} \tag{6.5}$$

which completes the proof of the inequality.

If T is R-recurrent, $\{\bar{x}_{ij}\}$, $j = 1, 2, \ldots$ is the only R-invariant measure of T with ith element unity. Hence all inequalities in (6.5) become equalities, which completes the assertion. □

Thus we can compute precise bounds, from ${}_{(n)}T$ *for the unique R-invariant measure with ith element unity of an R-recurrent T.*

6.5 An Example

We illustrate a number of the results and ideas of this chapter by an example which has the feature that the countable non-negative matrix $T = \{t_{ij}\}$ considered is more conveniently treated in doubly infinite form rather than having index set $\{1, 2, \ldots\}$ as could be achieved by relabelling. It will be taken up again in Chapter 7.

Suppose $T = \{t_{ij}\}_{i,j=-\infty}^{\infty}$ is defined by

$$t_{ij} = \begin{cases} bs_{i-1}, & j = i - 1 \\ (1 - 2b)s_i, & j = i \\ bs_{i+1}, & j = i + 1 \\ 0 & \text{otherwise} \end{cases}$$

where $b > 0$, $s_0 = 1$, $s_i > 0$, $0 \le |i| < h + 1$, where $h \le \infty$ and $s_i = 0$ otherwise, but $t_{h+1,h} = t_{-h-1,-h} = 0$; and

$$s \overset{\text{def}}{=} \sup_{i \neq 0} s_i < (1 - 2b), \tag{6.6}$$

the right-hand side being independent of h.

Let $\boldsymbol{u}^{(k)} = \{u_i^{(k)}\}$, $|i| = 0, 1, 2, \ldots$, $k \ge 0$, and suppose

$$\boldsymbol{u}^{(k+1)} = T\boldsymbol{u}^{(k)}, \qquad k \ge 0,$$

where $\boldsymbol{u}^{(0)} \ge \mathbf{0}$, $\mathbf{1}'\boldsymbol{u}^{(0)} = 1$ and $u_0^{(0)} > 0$, so $\boldsymbol{u}^{(k)} \ge \mathbf{0}$, $\neq \mathbf{0}$ for all $k \ge 0$ (in particular $u_0^{(k)} > 0$ for all $k \ge 0$). Further, since $\mathbf{1}'T$ when restricted to its "central" $(2h + 1)$ elements is $\{(1 - b)s_{-h}, s_{-h+1}, \ldots, s_{-1}, s_0, s_1, \ldots, s_{h-1}, (1 - b)s_h\}$, it follows that $\infty > \mathbf{1}'\boldsymbol{u}^{(k)}$ for all $k \ge 0$, so that

$$\boldsymbol{a}^{(k)} = \boldsymbol{u}^{(k)}/\mathbf{1}'\boldsymbol{u}^{(k)}, \qquad k \ge 0,$$

satisfies $\mathbf{1}'\boldsymbol{a}^{(k)} = 1$, $\boldsymbol{a}^{(k)} \ge \mathbf{0}$, $a_0^{(k)} > 0$.

We shall determine the asymptotic behaviour as $k \to \infty$ of $t_{ij}^{(k)}$. The related vectors $\boldsymbol{u}^{(k)}$, $\boldsymbol{a}^{(k)}$ shall for the present be used as an aid in the proof only, though there are evident connections with the notions of ergodicity of not necessarily stochastic matrices, in the manner discussed in Chapter 4.

First notice that

$$\begin{aligned} \boldsymbol{a}^{(k+1)} = \boldsymbol{u}^{(k+1)}/\mathbf{1}'\boldsymbol{u}^{(k+1)} &= T\boldsymbol{u}^{(k)}/\mathbf{1}'T\boldsymbol{u}^{(k)} \\ &\ge T\boldsymbol{u}^{(k)}/\boldsymbol{s}'\boldsymbol{u}^{(k)} = T\boldsymbol{a}^{(k)}/\boldsymbol{s}'\boldsymbol{a}^{(k)} \end{aligned} \tag{6.7}$$

where $\boldsymbol{s} = \{s_j\}$. Thus

$$\begin{aligned} a_0^{(k+1)} &\ge \{bs_{-1}a_{-1}^{(k)} + (1 - 2b)s_0 a_0^{(k)} + bs_1 a_1^{(k)}\}/\boldsymbol{s}'\boldsymbol{a}^{(k)} \\ &\ge (1 - 2b)a_0^{(k)}/\boldsymbol{s}'\boldsymbol{a}^{(k)} \end{aligned}$$

since $s_0 = 1$,

$$\ge (1 - 2b)a_0^{(k)}/(s + (1 - s)a_0^{(k)})$$

since

$$\boldsymbol{s}'\boldsymbol{a}^{(k)} \le a_0^{(k)} + s\sum_{i \neq 0} a_i^{(k)} = s + (1 - s)a_0^{(k)}.$$

Thus

$$(a_0^{(k+1)})^{-1} \le (1 - s)/(1 - 2b) + (s/(1 - 2b))(a_0^{(k)})^{-1}$$

whence, iterating

$$(a_0^{(k+1)})^{-1} \le \left\{\frac{(1 - s)}{(1 - 2b)}\right\}\left\{\sum_{r=0}^{k}\left(\frac{s}{(1 - 2b)}\right)^r + \left(\frac{s}{(1 - 2b)}\right)^{k+1}(a_0^{(0)})^{-1}\right\}.$$

In view of (6.6), $a_0^{(k)}$, $k \geq 0$, is therefore uniformly bounded from 0, say

$$\left.\begin{aligned} a_0^{(k)} &\geq \delta > 0, \qquad k \geq 0 \\ \text{and} \qquad \liminf_{k \to \infty} a_0^{(k)} &\geq 1 - 2b(1-s)^{-1} > 0 \end{aligned}\right\} \tag{6.8}$$

where δ is independent of h.

In the case of *finite* h, the finite matrix $\{t_{ij}\}_{i,j=-h}^{h}$ is primitive and the asymptotic behaviour of $\boldsymbol{u}^{(k)}$ is evident from Theorem 1.2. In particular, if $(\boldsymbol{x}^{(h)})'$, $\boldsymbol{y}^{(h)}$ are the Perron–Frobenius eigenvectors of this matrix normed so that $(\boldsymbol{x}^{(h)})'\boldsymbol{y}^{(h)} = 1$, $\mathbf{1}'\mathbf{y}^{(h)} = 1$ and $r(h)$ the Perron–Frobenius eigenvalue, then as $k \to \infty$

$$[r(h)]^{-k}T^k \to \boldsymbol{y}^{(h)}(\boldsymbol{x}^{(h)})'$$

if the vectors $\boldsymbol{y}^{(h)}$, $(\boldsymbol{x}^{(h)})'$ are made infinite in both directions by augmenting them by zeros, and (cf. Theorem 4.6)

$$\boldsymbol{a}^{(k)} \to \boldsymbol{y}^{(h)}.$$

Let us now turn to the case of *infinite* h. Consider T modified so that all entries are replaced by zero except t_{ij}, $i, j = -n, \ldots, -1, 0, 1, \ldots, n$ which are as for the original infinite T (thus we are effectively considering a $(2n+1) \times (2n+1)$ truncation of this infinite T). For the modified matrix taking $\boldsymbol{u}^{(0)} = \boldsymbol{y}^{(n)}$, where $\boldsymbol{y}^{(n)}$ is the infinite extended probability-normed right Perron–Frobenius eigenvector of the truncation, as considered above, it follows from (6.7) that for the modified matrix $\boldsymbol{a}^{(k)} = \boldsymbol{y}^{(n)}$ for all k, and from (6.8) that $y_0^{(n)} \geq \delta > 0$ (where $\boldsymbol{y}^{(n)} = \{y_i^{(n)}\}$), for all n. Using Cantor's diagonal selection argument, we may ensure that elementwise as $k \to \infty$

$$\boldsymbol{y}^{(n_k)} \to \boldsymbol{y}$$

through some subsequence $\{n_k\}$ of the integers, where $0 < \mathbf{1}'\boldsymbol{y} \leq 1$ (since $\mathbf{1}'\boldsymbol{y}^{(n)} = 1$, $y_0^{(n)} \geq \delta$). Since from (6.7), for $|i| \leq n_k$

$$y_i^{(n_k)} = \frac{\{bs_{i-1}y_{i-1}^{(n_k)} + (1-2b)s_i y_i^{(n_k)} + bs_{i+1}y_{i+1}^{(n_k)}\}}{\boldsymbol{s}'\boldsymbol{y}^{(n_k)} - bs_{-n_k}y_{-n_k}^{(n_k)} - bs_{n_k}y_{n_k}^{(n_k)}}.$$

It follows from the case $i = 0$ (since $y_0^{(n_k)} \geq \delta > 0$) that the limit as $k \to \infty$ of the denominator exists, and if we denote it by β^{-1}, $0 < \beta < \infty$ and

$$\boldsymbol{y} = \beta T \boldsymbol{y}.$$

Now put $\boldsymbol{x}' = \{s_i y_i\}$. It is readily checked that $\boldsymbol{x}'T = \beta^{-1}\boldsymbol{x}'$ and, clearly $\boldsymbol{x}'\boldsymbol{y} = \sum_i y_i^2 s_i < \infty$. Thus by Theorem 6.4, $\beta = R$, and T is R-positive (it is evidently irreducible), with $\boldsymbol{y}$ being a multiple of the unique R-invariant vector and $\boldsymbol{x}'$ of the R-invariant measure. Since T is evidently also aperiodic, by Theorem 6.5 for each (i, j), as $k \to \infty$

$$R^k t_{ij}^{(k)} \to y_i x_j \Big/ \sum_r y_r x_r = y_i s_j y_j \Big/ \sum_r s_r y_r^2 > 0.$$

Notice that apart from Theorem 6.4 and 6.5, we have used ideas akin to Theorem 6.9. We shall develop this theme further in Chapter 7 and extend the conclusions of the example there.

Had we developed the theory of §6.2 more extensively, we would have been able to show also that for each (i, j), as $k \to \infty$

$$t_{ij}^{(k)} \Big/ \sum_h t_{hj}^{(k)} \to y_i / \mathbf{1}'y.$$

Bibliography and Discussion

The strikingly elegant extension of Perron–Frobenius theory to infinite irreducible non-negative matrices T is as presented in §6.1–§6.3 basically due to D. Vere-Jones (1962, 1967) where a more detailed treatment and further results may be found. The reader is referred also to Albert (1963), Pruitt (1964), Kendall (1966), and Moy (1967*a*) in this connection; the last reference gives a particularly detailed treatment of the periodic case of T (see also Šidák (1964*a*)) and a generalization to irreducible T of the Martin boundary and potential theory of §5.5. The latter part of the proof of Theorem 6.1 is due to Kingman (1963, §8), and Theorem 6.4 is due to Kendall (1966).

As regards the Basic Assumptions 1–3 on T, consequences of relaxing 2 (i.e. of not necessarily assuming irreducibility) are examined briefly by Kingman (1963, §6) and Tweedie (1971, §2). The role of 1 and 2 and the consequences of their relaxation are examined by Kendall (1966) and Mandl & Seneta (1969; §2).

For infinite irreducible stochastic P the study of problems of geometric rate of convergence to their limit was initiated by Kendall (1959) and it is to him that the term *geometric ergodicity*, and the proof of its being a solidarity property, are due. The study was continued by Kendall (1960); Vere-Jones (1962) proved the uniform solidarity results for geometric convergence rate for transient and positive-recurrent P discussed in Theorem 6.7 and its following remarks: and further contributions were made by Vere-Jones (1963), Kingman (1963) and Vere-Jones (1966).

For examples of the usage of the Vere-Jones R-theory see also Seneta & Vere-Jones (1966), Moy (1967*b*), Daley (1969), and Kuich (1970*a*). Study of countable non-negative T as operators on sequences spaces has been undertaken by Putnam (1958, 1961) and, comprehensively, by Vere-Jones (1968).

The contents of §6.4 are taken from Seneta (1967*a*), a paper largely motivated by some earlier results and ideas of Sarymsakov (1953*b*; 1954, §§22–24) for infinite primitive stochastic P. Thus the result ${}_{(n)}R \downarrow {}_{(\infty)}R \geq 1$ is due to him, as is the idea of the proof of Theorem 6.10; although in the absence at the time of Vere-Jones' extension of the Perron–Frobenius theory to infinite irreducible T, his results were necessarily rather weaker. The present proof of Theorem 6.8 is adapted from Mandl & Seneta (1969) and differs from the proof in Seneta (1967*a*). The Corollary to Theorem 6.10 is essentially due to Kendall (see Seneta, 1967*a*).

A discussion of the fact that Assumption 4 in §6.4 does not result in essential loss of generality is given in Seneta (1968*a*) as a result of correspondence between Kendall and Seneta; and to some extent in Sarymsakov (1953*b*, pp. 11–12). Some computational aspects of the two alogorithms of the theory are briefly discussed in this 1968 paper also, and analysed, extended and implemented in Allen, Anderssen & Seneta (1977).

The example of §6.5 is adapted from Moran (1976).

Exercises

6.1. Prove the last part of Theorem 6.1 (avoiding the use of Lemma A.4) by assuming to the contrary that $R = \infty$ and using the relation $T_{ii}(z) = (1 - L_{ii}(z))^{-1}$.
Hint: For some k, $l_{ii}^{(k)} > 0$.

6.2. Show that if the Basic Assumption 3 is not made, then the situation where the convergence parameter R is 0 must be defined as R-transient to accord with the present theory for $R > 0$.

(Kendall, 1966)

6.3. Show that if T is periodic, with period d, then there exist "eigenvectors" (with possibly complex elements) corresponding to each of the "eigenvalues" $R^{-1} e^{2\pi i h/d}$, $h = 0, 1, 2, \ldots, d - 1$ if T possesses an R-invariant vector.

6.4. Suppose A and B are finite or infinite irreducible non-negative matrices, and R_A, $R_B \geq 0$ their convergence parameters. Suppose X is a non-negative, non-zero matrix such that AX, XB are elementwise finite, and that

$$AX = XB.$$

Show that, if each row of X has only a finite number of non-zero elements, $R_B \leq R_A$; and if each column of X has only a finite number of non-zero elements $R_A \leq R_B$. (Thus if X is *both row-finite* and *column-finite* in the above sense, $R_A = R_B$.)
Hint: Show first $A^k X = X B^k$.

(Generalized from Kuich, 1970*b*)

6.5. A semi-unrestricted random walk with reflecting barrier at the origin, as described by the stochastic matrix P of Example (3) of Chapter 4, §4.1, with $a = q$ can be shown to have

$$F_{00}(z) = qz + \tfrac{1}{2}(1 - \sqrt{[1 - 4pqz^2]}).$$

(Recall that in this instance the index set is $\{0, 1, 2, \ldots\}$.)
Show that P is transient if and only if $p > q$; in which case $R = (4pq)^{-1/2}$ and P is R-transient.
Hint: $1 - 4pq = (p + q)^2 - 4pq$.

6.6. Let $A = \{a_{ij}\}$ be an infinite matrix defined by $a_{i,i} = 0$, $a_{i,2i} = c_1$, $a_{i,2i+1} = c_2$, $a_{2i,i} = d$, $a_{2i+1,i} = d$ $(i \geq 1)$, $a_{i,j} = 0$ otherwise; c_1, c_2, $d > 0$. Show A is irreducible.

Construct a non-negative matrix X which is row and column finite such that $AX = XB$ where $B = \{b_{ij}\}$, is given by

$$b_{i,i+1} = c_1 + c_2, \qquad i \geq 1; \qquad b_{i,i-1} = d, \qquad i \geq 2,$$

all other $b_{ij} = 0$.

For the case $c_1 + c_2 = p$, $0 < p < 1$, it can be shown that for the matrix B,

$$F_{11}(z) = \tfrac{1}{2}(1 - \sqrt{[1 - 4pqz^2]}).$$

Use this fact and the result of Exercise 6.5 to deduce that

$$R_A = \{4d(c_1 + c_2)(d + c_1 + c_2)\}^{-1/2}.$$

Hint: Take $X = \{x_{ij}\}$ to be such that its jth column contains only zeros, apart from a "block" of elements with values 1, of length 2^{j-1}.

(Kuich, 1970*b*)

6.7. Show that if T has period d, and all the assumptions of §6.4 are satisfied, all $_{(n)}T$ for sufficiently large n have period d also.

(Seneta, 1967*a*)

6.8. Carry through as far as possible the arguments of Theorem 6.10 in the case of the more general matrix $[I - \beta T]$, $0 < \beta \leq R$. Show in particular that

$$d(\beta) \overset{\text{def}}{=} \lim_{n \to \infty} \det [_{(n)}I - \beta \cdot {}_{(n)}T] \geq 0$$

exists for $0 \leq \beta \leq R$ [the approach to the limit being monotone decreasing as n increases]. Thus $d(\beta)$ may in a sense be regarded as a (modified) characteristic polynomial of T; note that $d(R) = 0$ certainly for R-recurrent T, as one might expect.

CHAPTER 7

Truncations of Infinite Stochastic Matrices

For irreducible infinite stochastic matrices $T = P$ the approximative behaviour of finite truncations, touched on in §6.4, acquires probabilistic significance inasmuch as such matrices may be regarded in the role of transition matrices of countable Markov chains.

It is, first, of some interest to glance at the consequences of evolving the theory of §6.4 with $\beta = 1$ rather than $\beta = R$ (in the notation of Chapter 6), although we know now, in view of Theorem 6.6, that it is only for transient P that there may be differing results. On the other hand, even with recurrent P, the stochasticity of P will imply a more specialized structure of results than for general R-recurrent T.

Second, an important problem in applied work involving countable Markov chains is that, even though every entry of the transition matrix $P = \{p_{ij}\}$, $i, j = 1, 2, \ldots$ may be precisely specified, and even though the matrix is irreducible and positive-recurrent, the unique stationary distribution, which is of central importance, may not be analytically determinable. Recall[1] that for a matrix P with the specified structure, the stationary distribution is the unique invariant measure x' of P satisfying $x'\mathbf{1} = 1$, to which, in this case, we shall also refer as the *probability-normed* invariant measure. Evidently finite approximative techniques, based on successive finite truncations of P, for the probability-normed invariant measure are necessary for computer implementation. Results such as Theorem 6.9 and Theorem 6.10(a) provide limiting results which enable us to calculate from finite truncations, approximations to the *unit-normed* invariant measure (i.e. with some specific element unity). One might suppose in practice that, once an adequate approximation to the unit-normed measure has been obtained,

[1] See Theorem 5.5 and Exercise 5.5.

one for the stationary distribution can be obtained merely by dividing each of the elements in the approximation by their sum. Nevertheless algorithms should be developed for the direct computation of the probability-normed measure, and this is a different problem which we shall therefore examine. Allied problems, on which we shall also touch, are those of numerical stability and conditioning with increasing truncation size, and rate of convergence.

7.1 Determinantal and Cofactor Properties

In the sequel, ${}_{(n)}P$ is the $(n \times n)$ northwest corner truncation of P and the other quantities follow notation preceding, and in the body of, Theorem 6.10, if in that notation we replace ${}_{(n)}T$ by ${}_{(n)}P$.

Theorem 7.1. *If for stochastic irreducible P all but at most a finite number of truncations ${}_{(n)}P$, $n \geq 1$, are irreducible,*[1] *as $n \to \infty$:*

(*a*) ${}_{(n)}c_{ji}(1)/{}_{(n)}c_{ii}(1) \uparrow x_{ij} > 0$, $(i, j = 1, 2, \ldots)$ *where, for fixed i, $\{x_{ij}\}$, $j = 1, 2, \ldots$ constitutes the minimal subinvariant measure*[2] *of P.*

(*b*) $\lim {}_{(n)}\Delta(1) \geq 0$ *exists; if* $\lim {}_{(n)}\Delta(1) > 0$, *$P$ is transient.* $c_{ij}(1) = \lim {}_{(n)}c_{ij}(1)$ *exists for every pair (i, j) and all these limits are positive or zero together. For a transient matrix ${}_{(n)}\Delta(1)$ and ${}_{(n)}c_{ij}(1)$ have positive or zero limits together.*

(*c*) *If P is recurrent, $c_{ji}(1)$ is independent of i.*

(*d*) *There exist recurrent P where all $c_{ij}(1)$ are positive; and transient P where* $\lim {}_{(n)}\Delta(1) > 0$.

Proof. Proof of (*a*) and (*b*) follows that of Theorem 6.10, taking into account Theorem 5.3. [See Exercise 6.8 for a direction of further generalization.]

We need to prove (*c*) only in case the $c_{ij}(1)$ are positive and P is recurrent. In the manner of the proof of Theorem 6.10, approached with the quantities ${}_{(n)}F_{ij}$ rather than ${}_{(n)}L_{ij}$, we arrive at

$$c_{ji}(1) = c_{jj}(1)\{(1 - \delta_{ij})F_{ij}(1-) + \delta_{ij}\}$$

and since $F_{ij}(1-) = 1$ for all (i, j) since P is recurrent (Theorem 5.4^{D}) the result follows.

To demonstrate the validity of (*d*), we return to the stochastic P investigated in §5.6, which we note has all its truncations irreducible. Moreover

$${}_{(n)}\Delta(1) = \det [{}_{(n)}I - {}_{(n)}P] = (1 - q_n)p_1 p_2 \cdots p_{n-1}, = \alpha_n$$

[1] As remarked in §6.4, this assumption does not result in essential loss of generality.

[2] In the sense of Theorem 5.3.

in the notation of §5.6, the determinant being evaluated by first adding each column in turn to the first; and then expanding by the nth row. On the other hand for $1 \leq j \leq n$

$$_{(n)}c_{jj} = p_0 p_1 \cdots p_{j-1}, \qquad (p_0 = 1)$$

by deleting the jth row and column of $[_{(n)}I - {}_{(n)}P]$, adding all columns in turn to the first in the resulting matrix; and finally expanding the determinant by its first column.

Thus for $j = 1, 2, \ldots$

$$c_{jj} = \alpha_{j-1}$$

irrespective of whether P is recurrent or transient. On the other hand P is transient if and only if $\alpha_\infty > 0$ and so (in this case) P is transient if and only if

$$\alpha_\infty = \lim_{n\to\infty} {}_{(n)}\Delta(1) > 0. \qquad \square$$

Corollary. *Supposing P is as in Theorem 7.1 and recurrent and $\{x(i)\}$ is any invariant measure then as $n \to \infty$*

$$_{(n)}c_{ji}(1)/_{(n)}c_{ii}(1) \uparrow x(j)/x(i) \downarrow {}_{(n)}c_{jj}(1)/_{(n)}c_{ij}(1). \tag{7.1}$$

[This is merely a restricted version of the Corollary to Theorem 6.10.]

There may be probabilistic grounds for calling those transient P where $\lim {}_{(n)}\Delta(1) > 0$, *strongly transient* since in the example considered in the above proof and an appropriate generalization,[1] the positivity of this limit indicates that there is a *positive probability* of a corresponding Markov chain "going to infinity" by the "shortest possible" route (in these cases well defined) no matter what the initial state index, the limit itself being this probability if the chain starts at state-index 1.

The use of cofactor/determinant methods in the style of Theorems 6.10 and 7.1 has a number of further useful features which will become evident after the following preliminary lemma.

Lemma 7.1. *Let $\{y_{ij}, i, j = 1, \ldots, n\}$ be n^2 arbitrary numbers and define the matrix $X = \{x_{ij}\}, i, j = 1, \ldots, n$ by $x_{ij} = -y_{ij}, i, j = 1, \ldots, n, i \neq j$, and $x_{ii} = \sum_{k=1}^{n} y_{ik}, i = 1, \ldots, n$. Then*

$$\det X = \sum_J y_{1j_1} y_{2j_2} \cdots y_{nj_n}$$

where the summation is over a subset J of the set J^* of all ordered n-tuples $(j_1, \ldots, j_n)$, each entry chosen from $\{1, 2, \ldots, n\}$.

PROOF. Evidently

$$\det X = \sum_{J^*} \delta(j_1, j_2, \ldots, j_n) y_{1j_1} y_{2j_2} \cdots y_{nj_n}$$

[1] See Exercise 7.1.

where the $\delta(j_1, \ldots, j_n)$ are integers. Indeed, putting $y_{1j_1} = y_{2j_2} = \cdots = y_{nj_n} = 1$ and all other $y_{ij} = 0$, we see that $\delta(j_1, j_2, \ldots, j_n)$ is the determinant of the resulting X, which we may call $X(j_1, j_2, \ldots, j_n)$. It will therefore suffice to prove that det $X(j_1, j_2, \ldots, j_n) = 0$ or 1. If $j_i \neq i$ for any $i = 1, \ldots, n$, then each row sum of $X(j_1, j_2, \ldots, j_n)$ is zero, and hence this matrix is singular. Suppose for some i, say $i = i^*$, $j_i = i$. Then the i^*th row of $X(j_1, j_2, \ldots, j_n)$ has unity in the i^*th position and zeros elsewhere, and deleting the i^*th row and column again gives a matrix of the form of $X(j_1, j_2, \ldots, j_n)$ but of dimension $(n-1) \times (n-1)$. Hence the result follows by expanding det $X(j_1, j_2, \ldots, j_n)$ by the i^*th row and induction on n. □

Corollary. *If $X(i, i)$ is the cofactor of the (i, i) entry of X, then*

$$X(i, i) = \sum y_{1j_1} y_{2j_2} \cdots y_{i-1, j_{i-1}} y_{i+1, j_{i+1}} \cdots y_{nj_n}$$

the summation being over some subset J_i of ordered $(n-1)$ tuples $(j_1, j_2, \ldots, j_{i-1}, j_{i+1}, \ldots, j_n)$ with each element selected from the numbers $\{1, 2, \ldots, i-1, i+1, \ldots, n\}$.

PROOF.

$$X(i, i) = \sum \delta_i(j_1, j_2, \ldots, j_{i-1}, j_{i+1}, \ldots, j_n)$$

$$\times\, y_{1j_1} y_{2j_2} \cdots y_{i-1, j_{i-1}} y_{i+1, j_{i+1}} \cdots y_{nj_n}$$

with summation over all ordered $(n-1)$ tuples where the $\delta_i(\cdots)$ are integers. Again

$$\delta_i(j_1, j_2, \ldots, j_{i-1}, j_{i+1}, \ldots, j_n)$$

is $X(i, i)$ if we put all $y_{ij} = 0$ except

$$y_{1j_1} = y_{2j_2} = \cdots = y_{i-1, j_{i-1}} = y_{i+1, j_{i+1}} = \cdots = y_{nj_n} = 1,$$

and the form of the $(n-1) \times (n-1)$ matrix of which this is the determinant is that of the body of the lemma. □

The crux of the above results is that det X and $X(i, i)$ are each expressed as a sum of products of the y_{ij}'s with a *plus* sign for every summand.

It is clear that if $Q = \{q_{ij}\}$, $i, j = 1, \ldots, n$ is an irreducible strictly substochastic matrix of the nature of ${}_{(n)}P$ in Theorem 7.1, then $I - Q$ has the form of the matrix X in Lemma 7.1 if we put $y_{ij} = q_{ij}$, $i \neq j$, and $y_{ii} = 1 - \sum_{k=1}^{n} q_{ik}$. Returning specifically to the setting of Theorem 7.1 we have

$$y_{ii} = 1 - \sum_{k=1}^{n} p_{ik} = \sum_{k=n+1}^{\infty} p_{ik}$$

so that, by Lemma 7.1 and its Corollary,

$$ {}_{(n)}\Delta(1) = \sum p_{1k_1} p_{2k_2} \cdots p_{nk_n} \tag{7.2a}$$

with summation over a subset of ordered n tuples $(k_1, k_2, \ldots, k_n)$ where each element is selected from $\{1, 2, \ldots\}$; and

$$_{(n)}c_{ii}(1) = \sum p_{1k_1} p_{2k_2} \cdots p_{i-1, k_{i-1}} p_{i+1, k_{i+1}} \cdots p_{nk_n} \tag{7.2b}$$

with summation over a subset of ordered $(n-1)$ tuples $(k_1, k_2, \ldots, k_{i-1}, k_{i+1}, \ldots, k_n)$ with each element selected from $\{1, 2, \ldots, i-1, i+1, \ldots\}$.

Theorem 7.2. *Suppose the matrix $P = \{p_{ij}\}$ satisfies the conditions of Theorem 7.1,*[1] *and a stochastic matrix $\tilde{P} = \{\tilde{p}_{ij}\}$ satisfies for each (i, j)*

$$(1 + \varepsilon_i)^{-1} p_{ij} \leq \tilde{p}_{ij} \leq p_{ij}(1 + \varepsilon_i), \tag{7.3}$$

where $0 \leq \varepsilon_i$ and

$$1 + \varepsilon = \prod_{i=1}^{\infty} (1 + \varepsilon_i) < \infty.$$

Then $\tilde{P}$ is transient if and only if P is.

Proof. Let $_{(n)}\tilde{\Delta}(1)$ and $_{(n)}\tilde{c}_{ii}(1)$ have the obvious meaning for $\tilde{P}$, which, in view of (7.3) has the same incidence matrix as P and has coincident truncation structure.

Using the elementary inequalities

$$\prod_{\substack{k=1 \\ k \neq i}}^{n} (1 + \varepsilon_k) \leq \prod_{k=1}^{n} (1 + \varepsilon_k) \leq \prod_{k=1}^{\infty} (1 + \varepsilon_k) = 1 + \varepsilon$$

we obtain from (7.2)–(7.3) that

$$\begin{aligned}(1 + \varepsilon)^{-2} {}_{(n)}c_{ii}(1)/_{(n)}\Delta(1) &\leq {}_{(n)}\tilde{c}_{ii}(1)/_{(n)}\tilde{\Delta}(1) \\ &\leq (1 + \varepsilon)^2 {}_{(n)}c_{ii}(1)/_{(n)}\Delta(1). \end{aligned} \tag{7.4}$$

If P is transient, $\sum_{r=0}^{\infty} p_{ii}^{(r)} < \infty$ by Lemma 5.3, and, as in the proof of Theorem 6.10, by dominated convergence

$$\frac{_{(n)}c_{ii}(1)}{_{(n)}\Delta(1)} \uparrow \sum_r p_{ii}^{(r)}.$$

Hence by (7.4)

$$\lim_{n \to \infty} \frac{_{(n)}\tilde{c}_{ii}(1)}{_{(n)}\tilde{\Delta}(1)} < \infty$$

and the limit must be $\sum_r \tilde{p}_{ii}^{(r)}$, so $\tilde{P}$ is transient. That $\tilde{P}$ transient implies P transient is proved similarly. □

[1] Again, the restriction of ireducibility of truncations is made for convenience rather than necessity.

Corollary. *In the event that P and $\tilde{P}$ are recurrent and $\boldsymbol{x}'$ is the unique (to constant multiples) invariant measure of P, then $\tilde{P}$ has invariant measure $\tilde{\boldsymbol{x}}'$ satisfying*

$$(1+\varepsilon)^{-2}\boldsymbol{x}' \le \tilde{\boldsymbol{x}}' \le (1+\varepsilon)^{2}\boldsymbol{x}'.$$

PROOF. In view of (7.3) and (7.2b)

$$\frac{(1+\varepsilon)^{-2}{}_{(n)}c_{jj}(1)}{{}_{(n)}c_{ii}(1)} \le \frac{{}_{(n)}\tilde{c}_{jj}(1)}{{}_{(n)}\tilde{c}_{ii}(1)} \le \frac{(1+\varepsilon)^{2}{}_{(n)}c_{jj}(1)}{{}_{(n)}c_{ii}(1)}.$$

Now

$$\frac{{}_{(n)}c_{jj}(1)}{{}_{(n)}c_{ii}(1)} = \left(\frac{{}_{(n)}c_{jj}(1)}{{}_{(n)}c_{ji}(1)}\right)\Bigg/\left(\frac{{}_{(n)}c_{ii}(1)}{{}_{(n)}c_{ji}(1)}\right)$$

and as $n \to \infty$ from Theorem 7.1 and its Corollary putting $\boldsymbol{x}' = \{x(i)\}$,

$$\frac{{}_{(n)}c_{ii}(1)}{{}_{(n)}c_{ji}(1)} \downarrow \frac{x(i)}{x(j)}, \qquad \frac{{}_{(n)}c_{jj}(1)}{{}_{(n)}c_{ji}(1)} \downarrow \frac{1}{F_{ij}(1-)} = 1$$

so

$${}_{(n)}c_{jj}(1)/{}_{(n)}c_{ii}(1) \to x(j)/x(i)$$

and similarly for the matrix $\tilde{P}$. Thus

$$(1+\varepsilon)^{-2}[x(j)/x(i)] \le \tilde{x}(j)/\tilde{x}(i) \le (1+\varepsilon)^{2}[x(j)/x(i)].$$

Thus taking i fixed and $j = 1, 2, \ldots$ yields the result. □

In the case that P is positive-recurrent and satisfies the conditions of Theorem 7.1, let us write $\boldsymbol{\pi}' = \{\pi_i\}$ for the unique probability-normed invariant measure (unique stationary distribution). Let us take i fixed in (7.1) and then rewrite (7.1) as

$${}_{(n)}\alpha(j) \uparrow \pi_j/\pi_i \downarrow {}_{(n)}\beta(j).$$

Then

$$\pi_j \le \pi_j \Bigg/ \sum_{k=1}^{n} \pi_k \le {}_{(n)}\beta(j) \Bigg/ \sum_{k=1}^{n} {}_{(n)}\alpha(k),$$

and by dominated convergence as $n \to \infty$,

$$\sum_{k=1}^{n} {}_{(n)}\alpha(k) \uparrow \sum_{k=1}^{\infty} \frac{\pi_k}{\pi_i} = \frac{1}{\pi_i}$$

so

$${}_{(n)}\beta(j) \Bigg/ \sum_{k=1}^{n} {}_{(n)}\alpha(k) \downarrow \pi_j, \qquad j = 1, 2, \ldots. \tag{7.5}$$

Thus the left-hand side provides a direct pointwise estimator of π' converging elementwise from above, providing one solution to the problem of numerically approximating the stationary distribution as mentioned in the preliminary remarks to this chapter.

Attention thus focuses, for a given finite truncation $_{(n)}P$, on the question of *computation* of the cofactors which enter into relation (7.5); or (7.1) if a unit-normed version of the invariant measure for recurrent P is required. Since we may fix i, we require for specific n under consideration, $_{(n)}c_{ji}(1)$, $_{(n)}c_{jj}(1)$, $_{(n)}c_{ij}(1)$, $j = 1, \ldots, n$, that is: the ith row and column and the diagonal of the matrix Adj $({}_{(n)}I - {}_{(n)}P)$. {In the case of a unit-normed measure, if both upper *and* lower approximative bounds for $x(j)/x(i)$ are not required, (7.1) shows that the ith row of the adjoint matrix will suffice.} Putting

$$_{(n)}h_{ij} = {}_{(n)}c_{ji}(1)/{}_{(n)}\Delta(1),$$

and

$$_{(n)}\boldsymbol{h}_i' = \{{}_{(n)}h_{ij}\},\ j = 1, \ldots, n,\quad {}_{(n)}\boldsymbol{g}_i = \{{}_{(n)}h_{ji}\},\ j = 1, \ldots, n,$$

from the fact that

$$\begin{aligned}\text{Adj } ({}_{(n)}I - {}_{(n)}P)({}_{(n)}I - {}_{(n)}P) &= ({}_{(n)}I - {}_{(n)}P) \text{ Adj } ({}_{(n)}I - {}_{(n)}P)\\ &= {}_{(n)}\Delta(1)_{(n)}I\end{aligned}$$

it follows that

$$_{(n)}\boldsymbol{h}_i'({}_{(n)}I - {}_{(n)}P) = {}_{(n)}\boldsymbol{f}_i', \qquad ({}_{(n)}I - {}_{(n)}P)_{(n)}\boldsymbol{g}_i = {}_{(n)}\boldsymbol{f}_i \tag{7.6}$$

where $_{(n)}\boldsymbol{f}_i$ is our customary notation for a vector with unity in the ith position and zeros elsewhere. The required vectors are therefore obtained as solutions to linear equation systems involving the matrix $({}_{(n)}I - {}_{(n)}P)$ and taking suitable ratios of elements allows, *in theory*, implementation of (7.1) or (7.5).

It is clear in regard to *numerical properties* of the approximations that the structure of the matrix $({}_{(n)}I - {}_{(n)}P)$ as $n \to \infty$ must be considered in respect of numerical stability and conditioning; and that such considerations are important for any *practical* (computer) implementation of the algorithms (7.1) or (7.5).

For each n sufficiently large, the matrix $_{(n)}P$ is of the form $Q = \{q_{ij}\}$ where Q is irreducible, strictly substochastic, so that $Q^k \to 0$ as $k \to \infty$. Hence $(I - Q)^{-1} = \sum_{k=0}^{\infty} Q^k > 0$ (Lemma B.1 of Appendix B), and moreover the matrix $A = \{a_{ij}\} = I - Q$ with whose inversion we are in practice concerned, is also diagonally dominant viz.

$$|a_{ii}| \geq \sum_{\substack{j \\ j \neq i}} |a_{ij}|$$

with strict inequality for at least one i (on account of strict substochasticity). This last property is a favourable one in regard to numerical aspects of the systems (7.6) as $n \to \infty$. We shall not dwell on this here;[1] and shall confine ourselves to a discussion of the sensitivity of $(I - Q')^{-1}$ to perturbations in the matrix $I - Q'$. We shall henceforth, therefore, be confining our remarks to the first of the systems (7.6) when written in transposed form.

The sensitivity of B^{-1} to perturbations in a nonsingular matrix $B = \{b_{ij}\}$, $i, j = 1, \ldots, n$ may be measured by the *condition number*

$$\kappa(B) = \|B^{-1}\| \, \|B\|$$

where $\|\cdot\|$ is a vector norm in R_n. In dealing with a matrix of the form $B = (I - Q')$ where $Q = \{q_{ij}\}$ is an irreducible strictly substochastic matrix, it is natural to use the l_1 norm yielding

$$\|B\| = \max_j \left\{\sum_i |b_{ij}|\right\}.$$

Thus

$$\|Q'\| \leq 1, \qquad 1 - q_{11} \leq \|I - Q'\| \leq \max_j 2(1 - q_{jj}) \leq 2$$

noting that $q_{jj} < 1$ for each j, and since if r is the Perron-Frobenius eigenvalue of Q', $(1 - r)^{-1}$ is the Perron-Frobenius eigenvalue of $(I - Q')^{-1}$, it follows that

$$(1 - r)^{-1} \leq \|(I - Q')^{-1}\| \leq \sum_{k=0}^{\infty} \|Q'\|^k = (1 - \|Q'\|)^{-1} \leq \infty.$$

Thus

$$(1 - q_{11})/(1 - r) \leq \kappa(I - Q') \leq 2(1 - \|Q'\|)^{-1}. \tag{7.7}$$

Clearly with $Q = {}_{(n)}P$, and ${}_{(n)}r$ denoting its Perron–Frobenius eigenvalue, since we are considering infinite recurrent P, it follows from Theorems 6.6 and 6.8, since $R = 1$, and ${}_{(n)}r = 1/{}_{(n)}R$, that ${}_{(n)}r \uparrow 1$. Hence from (7.7), as $n \to \infty$

$$\kappa({}_{(n)}I - {}_{(n)}P') \to \infty,$$

so the system

$$({}_{(n)}I - {}_{(n)}P')_{(n)}\boldsymbol{h}_i = {}_{(n)}f_i \tag{7.8}$$

suffers increasing ill-conditioning with increasing n. (By symmetry, using the l_∞ norm, similar remarks would apply to the other equation system of the pair (7.6).) While this does not necessarily mean that direct solution of (7.6) for ${}_{(n)}\boldsymbol{h}_i'$ and/or ${}_{(n)}\boldsymbol{g}_i$ followed by suitable adjustment to implement (7.1) or (7.5) will produce increasingly numerically inaccurate approximations with increasing n, several reasons exist,[2] in the case of *positive-recurrent* P, for

[1] See Bibliography and Discussion to this chapter for references.

[2] See Bibliography and Discussion to this chapter.

considering rather than (7.5) as the finite approximation procedure for the stationary distribution $\pi' = \{\pi_j\}$, the possibility of the procedure (for fixed i)

$$_{(n)}\boldsymbol{h}_i'/_{(n)}\boldsymbol{h}_i'\mathbf{1} \to \boldsymbol{\pi}' \tag{7.9}$$

in case it can be shown to hold. An obvious probabilistically aesthetic reason is that the left-hand side of (7.9), unlike that of (7.5), is itself a probability vector and the relation (7.9) is a manifestation of "convergence in distribution". The reader should note also that the approximations can be calculated from the single system (7.8), again unlike the situation (7.5).

We shall establish in the next section that (7.9) holds at least for certain classes of irreducible positive-recurrent P. We shall refer to it as the probability algorithm.

7.2 The Probability Algorithm

In this section we deal (except where otherwise stated) with positive-recurrent stochastic P. A rather obvious question which may be asked if it is sought to approximate $\boldsymbol{\pi}'$ by a sequence of finite probability vectors, is whether this may be done by the finite stationary distribution vectors $_{(n)}\boldsymbol{\pi}'$ corresponding to the (irreducible) finite stochastic matrices $_{(n)}\bar{P}$ formed from $_{(n)}P$ by augmenting one specific column of $_{(n)}P$. We first show that the left-hand side of (7.9) may be regarded as such a $_{(n)}\boldsymbol{\pi}'$, which result thus establishes an *equivalence between two algorithms*: that based on (7.8) for the unadjusted truncations $_{(n)}P$, and that based on an $_{(n)}\bar{P}$. This differing viewpoint of the finite approximate vectors $_{(n)}\boldsymbol{\pi}'$, further, enables us to consider *different methods for their calculation* from the numerically relevant standpoint of condition (number) as $n \to \infty$, which in the case of calculation directly from (7.8) appeared unsatisfactory, but may be moreso from a different computational approach.

The equivalence mentioned above is based on the following result,[1] where $Q = \{q_{ij}\}$ is an $(n \times n)$ substochastic matrix satisfying $Q^k \to 0$ as $k \to \infty$. Thus the matrix $(I - Q)^{-1} = \sum_{k=0}^{\infty} Q^k \geq 0$ exists by Lemma B.1 of Appendix B.

Lemma 7.2. *Let* $\boldsymbol{x}(\geq \mathbf{0}, \neq \mathbf{0})$ *be the unique solution of*

$$(I - Q')\boldsymbol{x} = \boldsymbol{b} \qquad (\boldsymbol{b} \geq \mathbf{0}, \neq \mathbf{0}).$$

Then the matrix

$$P = Q + (I - Q)\mathbf{1}\boldsymbol{b}'/\boldsymbol{b}'\mathbf{1}$$

is stochastic, and if $\boldsymbol{\pi}$ *is any stationary distribution vector corresponding to it,*

$$\boldsymbol{\pi}' = \boldsymbol{x}'/\boldsymbol{x}'\mathbf{1}.$$

[1] Which indicates how the assumption of irreducibility of $_{(n)}P$ may be dispensed with.

PROOF. It is obvious $P \geq 0$, $P\mathbf{1} = \mathbf{1}$. A finite stochastic matrix always has at least one stationary distribution vector[1] π' (thus $\pi' \geq \mathbf{0}'$, $\pi'\mathbf{1} = 1$, $\pi'P = \pi'$). Thus

$$\pi'Q + \pi'(I - Q)\mathbf{1}\boldsymbol{b}'/\boldsymbol{b}'\mathbf{1} = \pi'$$

so

$$\pi'(I - Q) = (\pi'(I - Q)\mathbf{1}/\boldsymbol{b}'\mathbf{1})\boldsymbol{b}'$$

so π' is a multiple of x' whence the conclusion follows. □

Corollary. *Under the assumptions of the lemma, P has a unique stationary distribution vector π', and this is the unique solution to the linear equation $n \times (n + 1)$ system*

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

and of the corresponding $(n \times n)$ system if any column of $I - P$ (and the corresponding zero entry of the right-hand side) is omitted.

PROOF. The uniqueness of π' follows from the uniqueness of x'. Then as in the discussion in §4.2 (preceding Definition 4.3), it follows that the index set of P contains a single essential class of indices, and the other conclusions are as in that section. □

It follows that if we take $Q = {}_{(n)}P$ and $\boldsymbol{b} = f_i$, and if ${}_{(n)}P^k \to 0$ as $k \to \infty$ (certainly so if ${}_{(n)}P$ is assumed irreducible), then $P = {}_{(n)}\bar{P}$ [that is, ${}_{(n)}P$ augmented in the ith column to make it stochastic], and from (7.8)

$$ {}_{(n)}\pi' = {}_{(n)}\boldsymbol{h}_i'/{}_{(n)}\boldsymbol{h}_i'\mathbf{1} \tag{7.10}$$

as required.

The next step is to show that ${}_{(n)}\pi' \to \pi'$ elementwise as $n \to \infty$ at least for some classes of positive-recurrent matrices P.

Lemma 7.3. *Suppose that we can find a subsequence $\{{}_{(n_s)}P\}$, $s \geq 1$, of truncations, each member of which satisfies ${}_{(n_s)}P^k \to 0$ as $k \to \infty$,[2] and such that, by changing one column only (not necessarily the same for each s), ${}_{(n_s)}P$ can be made stochastic in such a way that those $n_s - 1$ stationary equations, of the whole n_s equations*

$$ {}_{(n_s)}\boldsymbol{v}'{}_{(n_s)}\bar{P} = {}_{(n_s)}\boldsymbol{v}',$$

which do not involve the relevant column, coincide with $(n_s - 1)$ of the first n_s stationary equations, of the infinite system

$$\boldsymbol{v}'P = \boldsymbol{v}'.$$

[1] See Lemma 1.1, and remarks in §4.2.

[2] For irreducible P this involves no loss of generality—see Exercise 7.3.

Then (using the previous notation for stationary distribution vectors)

$$_{(n_s)}\pi_j = \pi_j \Big/ \sum_{r=1}^{n_s} \pi_r, \qquad j = 1, \ldots, n_s.$$

In particular, as $s \to \infty$, $_{(n_s)}\pi_j \downarrow \pi_j, j \geq 1$.

PROOF. From the Corollary to Lemma 7.2, $(n_s - 1)$ of the equations

$$_{(n_s)}\boldsymbol{v}' \, _{(n_s)}\bar{P} = {}_{(n_s)}\boldsymbol{v}'$$

together with $_{(n_s)}\boldsymbol{\pi}'\mathbf{1} = 1$ determine the stationary distribution vector $_{(n_s)}\boldsymbol{\pi}'$ uniquely. Under the assumptions of the lemma, therefore

$$_{(n_s)}\pi_j = \pi_j \Big/ \sum_{k=1}^{n_s} \pi_k, \qquad j = 1, \ldots, n_s. \qquad \square$$

Note that in this situation we have monotone convergence from above, comparable to (7.5).

Two kinds of infinite stochastic matrices $P = \{p_{ij}\}$, assumed *a priori* positive-recurrent which satisfy the conditions of Lemma 7.3 are the following.

(*a*) *Generalized Renewal Matrices.*
These satisfy

$$p_{ij} = 0, \qquad i > j > 1$$

and are a generalization of the infinite matrix considered in §5.6 [See also the Bibliography and Discussion to §§3.1–3.2.] If we consider $_{(n)}\bar{P}$ formed from $_{(n)}P$ by replacing p_{i1}, $i = 1, \ldots, n$ by

$$\bar{p}_{i1} = p_{i1} + \sum_{j=n+1}^{\infty} p_{ij}$$

so only the first column of $_{(n)}P$ is affected, then clearly the *last* $(n - 1)$ stationary equations of $_{(n)}\bar{P}$ are also stationary equations of P. (7.10) holds with $i = 1$.

(*b*) *Upper-Hessenberg Matrices.*
A matrix of this type satisfies

$$p_{ij} = 0 \qquad i > j + 1$$

so all entries below the subdiagonal are zero. Clearly the *first* $(n - 1)$ stationary equations of $_{(n)}\bar{P}$ formed by augmenting the *last* column of $_{(n)}P$ are also stationary equations of P. Thus (7.10) holds with $i = n$ in this case, and hence with *i not fixed irrespective of n*. By Lemma 7.3 therefore

$$_{(n)}\boldsymbol{h}'_n / {}_{(n)}\boldsymbol{h}'_n \mathbf{1} \downarrow \boldsymbol{\pi}'$$

in the sense of elementwise convergence.

Definition 7.1. An infinite $P = \{p_{ij}\}$, $i, j = 1, 2, \ldots$ is said to be a Markov matrix if all the elements of at least one column are uniformly bounded away from zero.[1]

We shall show that for a Markov matrix, *which may not be irreducible*, a unique stationary distribution vector always exists, and may be approximated by ${}_{(n)}\pi'$ calculated from appropriately augmented truncations ${}_{(n)}P$.

Theorem 7.3. *Let $P = \{p_{ij}\}$, $i, j = 1, 2, \ldots$ be a Markov matrix, and assume without loss of generality that $p_{i1} \geq \delta > 0$, $i \geq 1$. Let stochastic ${}_{(n)}\bar{P}$ be formed from ${}_{(n)}P$ by augmentation of the first column (if necessary).[2] Then ${}_{(n)}\bar{P}$ and P have unique stationary distribution vectors ${}_{(n)}\pi'$, π' respectively, and*

$$ {}_{(n)}\pi' \to \pi' $$

elementwise.

PROOF. ${}_{(n)}\bar{P}$ is evidently a finite Markov matrix (its 1st column is positive), and hence is, clearly, regular i.e. contains a single essential class of indices which is aperiodic. Hence by the discussion in §4.2 for stochastic matrices with a single essential class, there is a unique stationary distribution vector ${}_{(n)}\pi'$. This clearly satisfies

$$ {}_{(n)}\pi'[{}_{(n)}I - ({}_{(n)}\bar{P} - \delta \mathbf{1} f'_1)] = \delta f'_1 \tag{7.11} $$

where

$$ Q \overset{\text{def}}{\equiv} {}_{(n)}\bar{P} - \delta \mathbf{1} f'_1 $$

is clearly substochastic with all row sums $(1 - \delta)$. Hence (e.g. by Theorem 4.3), $Q^k \to 0$ as $k \to \infty$, so $(I - Q)^{-1}$ exists (Lemma B.1 of Appendix B) and is non-negative, whence the system (7.11) has a unique solution given by

$$ {}_{(n)}\pi' = \delta f'_1 \sum_{s=0}^{\infty} ({}_{(n)}\bar{P} - \delta \mathbf{1} f'_1)^s, \qquad \geq \delta f'_1. \tag{7.12} $$

Let $\pi^* = \{\pi_j^*\}$, $j \geq 1$, be given by

$$ \pi_j^* = \liminf_{n \to \infty} {}_{(n)}\pi_j, \qquad j \geq 1, $$

where, from (7.12), $\pi_1^* \geq \delta$. Hence by Fatou's Lemma, since ${}_{(n)}\pi'\mathbf{1} = 1$

$$ 0 < \delta \leq (\pi^*)'\mathbf{1} \leq 1. \tag{7.13} $$

Rearranging (7.11) as

$$ {}_{(n)}\pi'({}_{(n)}\bar{P} - \delta \mathbf{1} f'_1) + \delta f'_1 = {}_{(n)}\pi' $$

[1] This is consistent with Definition 4.7.

[2] See Exercise 7.7.

and again applying Fatou's Lemma

$$(\pi^*)'(P - \delta \mathbf{1} f'_1) + \delta f'_1 \leq (\pi^*)' \tag{7.14}$$

whence from (7.13)

$$(\pi^*)'P \leq (\pi^*)'.$$

Now, in fact equality must hold at all entries of this equation, otherwise by stochasticity of P and (7.13)

$$(\pi^*)'\mathbf{1} < (\pi^*)'\mathbf{1}.$$

Hence there is elementwise equality in (7.14), and hence $(\pi^*)'\mathbf{1} = 1$. Hence P has a stationary distribution vector, viz. π^*.

Now, it is readily seen that P contains a single essential class of indices (those which communicate with index 1) which is also aperiodic $(p_{11} \geq \delta > 0)$, any other indices being inessential and hence transient (Lemma 5.2). Hence, as $k \to \infty$, P^k has an elementwise limit matrix, which is the zero-matrix unless the index 1 is positive-recurrent (see Chapter 5). The case of a zero limit matrix, in view of

$$(\pi^*)'P^k = (\pi^*)' \qquad k \geq 1$$

leads as $k \to \infty$, by dominated convergence to $\mathbf{0}' = (\pi^*)'$ which is a contradiction. Hence the index 1 is positive-recurrent, and there is a unique stationary distribution vector corresponding to the single essential class. An argument, analogous to that given in §4.2 for a finite matrix containing a single essential class, now reveals that there is a unique stationary distribution vector π', being this vector for the positive-recurrent class augmented by zero entries for any inessential indices in P.

Thus $\pi^* = \pi$. Suppose now that for some j,

$$\limsup_{n\to\infty} {}_{(n)}\pi_j > \liminf_{n\to\infty} {}_{(n)}\pi_j = \pi_j^*.$$

Select a subsequence $\{n_k\}$ so that, for that j,

$$\lim_{k\to\infty} {}_{(n_k)}\pi_j = \limsup_{n\to\infty} {}_{(n)}\pi_j.$$

Repeating the previous argument through the subsequence $\{n_k\}$ we arrive at a contradiction to the uniqueness of π. Hence

$$\lim_{n\to\infty} {}_{(n)}\pi_j = \pi_j, \qquad j = 1, 2, \ldots.$$ □

Corollary. *If P is, additionally, assumed irreducible, then*

$${}_{(n)}\pi' = {}_{(n)}\boldsymbol{h}'_1 / {}_{(n)}\boldsymbol{h}'_1\mathbf{1} \to \pi'.$$

Proof. Exercise 7.3 shows that ${}_{(n)}P$ satisfies the role of Q in Lemma 7.2 so (7.10) is applicable. □

The above reasoning in part enables us to deduce (a generally unverifiable) necessary and sufficient condition for

$$_{(n)}\pi' \to \pi'$$

in the case of a P which is positive-recurrent. We have in general

$$_{(n)}\pi' {}_{(n)}\bar{P} = {}_{(n)}\pi'$$

where $_{(n)}\bar{P}$ is $_{(n)}P$ (where Exercise 7.3 shows that $_{(n)}P$ satisfies the role of Q in Lemma 7.2), augmented in some column (not necessarily the same for each n) to make it stochastic. Define $\pi^* = \{\pi_j^*\}$ by

$$(0 \le)\pi_j^* = \liminf_{n\to\infty} {}_{(n)}\pi_j, \qquad j \ge 1.$$

By Fatou's Lemma $0 \le (\pi^*)'\mathbf{1} \le 1$, and

$$(\pi^*)'P \le (\pi^*)'.$$

Now suppose the set of probability vectors $\{_{(n)}\pi\}$, $n \ge 1$, is *tight*[1] (viz. any infinite subsequence itself possesses an infinite subsequence converging elementwise to a probability vector). Suppose for some j, say $j = j_0$,

$$\limsup_{n\to\infty} {}_{(n)}\pi_j > \liminf_{n\to\infty} {}_{(n)}\pi_j = \pi_j^*.$$

Suppose $\{n_i\}$, $\{n_k\}$ are subsequences such that

$$\lim_{i\to\infty} {}_{(n_i)}\pi_{j_0} = \limsup_{n\to\infty} {}_{(n)}\pi_{j_0}, \qquad \lim_{k\to\infty} {}_{(n_k)}\pi_{j_0} = \liminf_{n\to\infty} {}_{(n)}\pi_{j_0}.$$

Using tightness and taking subsequences if necessary

$$\lim_{i\to\infty} {}_{(n_i)}\pi = \pi_1, \qquad \lim_{k\to\infty} {}_{(n_k)}\pi = \pi_2$$

where π_1 and π_2 are probability vectors *which differ in at least the* j_0th *element*. But, by Fatou's Lemma, applied to the corresponding finite systems, it follows that

$$\pi_1' P \le \pi_1', \qquad \pi_2' P \le \pi_2'$$

whence summing both sides clearly $\pi_1' P = \pi_1'$, $\pi_2' P = \pi_2'$, and since P is positive-recurrent, $\pi_1 = \pi_2$ which is a contradiction. Hence

$$\lim_{n\to\infty} {}_{(n)}\pi_j = \pi_j^*, \qquad j \ge 1,$$

where $\pi^* = \{\pi_j^*\}$ is the unique stationary distribution vector π of P (π^* must be a probability vector by tightness of $\{_{(n)}\pi\}$).

Conversely, if $_{(n)}\pi \to \pi$ where π is the unique stationary distribution vector of P, then the sequence $\{_{(n)}\pi\}$ is obviously tight. Thus tightness of $\{_{(n)}\pi\}$ is necessary and sufficient for $_{(n)}\pi \to \pi$.

[1] This notion is well-known in the theory of weak convergence of probability measures, though the equivalence of our definition and the usual one is known as Prokhorov's Theorem.

Returning to the case of a Markov matrix P, it is interesting to envisage the calculation of ${}_{(n)}\boldsymbol{\pi}$, for fixed n, *directly from the linear equation system* (7.11) transposed, rather than (as for a P which is also irreducible) from (7.8) and (7.10) with $i = 1$ which incurs problems of numerical condition as $n \to \infty$.[1] We have already noted that with

$$Q = {}_{(n)}\bar{P} - \delta \mathbf{1} f'_1,$$

$$Q\mathbf{1} = (1 - \delta)\mathbf{1}$$

so if we assume for convenience of application of (7.7) that, as there, Q is irreducible, obviously $r = 1 - \delta$, $\|Q'\| = 1 - \delta$, whence

$$(1 - q_{11})/\delta \le \kappa(I - Q') \le 2/\delta$$

so "good conditioning" with increasing n persists for this computational procedure.

Example. The following is a structurally simple irreducible infinite matrix P, with irreducible truncations, which is both a generalized renewal matrix and a Markov matrix. $P = \{p_{ij}\}$ is defined by

$$p_{ij} = \begin{cases} p_j, & j \ge i \ge 1, \\ \sum_{r=1}^{i-1} p_r, & j = 1, i \ge 2, \\ 0 & \text{otherwise,} \end{cases}$$

where $\boldsymbol{p} = \{p_j\}, j \ge 1$, is a probability distribution with each element positive. It is easily shown from the stationary equations that the unique stationary distribution vector $\boldsymbol{\pi} = \{\pi_j\}$ satisfies

$$\pi_j = \pi_1 p_j \Big/ \prod_{r=2}^{j} (1 - p_r), \qquad j \ge 2. \tag{7.15}$$

Thus for $j \ge 2$,

$$\begin{aligned} {}_{(n)}\pi_j &= \pi_j \Big/ \sum_{k=1}^{n} \pi_k \\ &= \left\{ p_j \Big/ \prod_{r=2}^{j} (1 - p_r) \right\} \Big/ \left\{ 1 + \sum_{k=2}^{n} \left(p_k \Big/ \prod_{r=2}^{k} (1 - p_r) \right) \right\} \end{aligned}$$

and

$${}_{(n)}\pi_1 = \pi_1 \Big/ \sum_{k=1}^{n} \pi_k = 1 \Big/ \left\{ 1 + \sum_{k=2}^{n} \left(p_k \Big/ \prod_{r=2}^{k} (1 - p_r) \right) \right\}.$$

[1] See also Exercise 7.9.

For a situation such as obtains here, and in general where

$$_{(n)}\pi_j = \pi_j \Big/ \sum_{k=1}^{n} \pi_k, \qquad j \geq 1,$$

we see that

$$_{(n)}\pi_j - \pi_j = \pi_j \left(\sum_{k=n+1}^{\infty} \pi_k \Big/ \sum_{k=1}^{n} \pi_k \right) \sim \pi_j \sum_{k=n+1}^{\infty} \pi_k$$

(recall that if P is positive-recurrent, $\pi > \mathbf{0}$, from Lemma 5.6) as $n \to \infty$, so we are able to make deductions about *convergence rate*. Specifically, using the present example, we see directly from (7.15) that

$$\sum_{k=n+1}^{\infty} \pi_k \sim \text{const.} \sum_{k=n+1}^{\infty} p_k$$

so the convergence rate of $_{(n)}\pi_j$ to π_j with n for fixed j may be arbitrarily slow or arbitrarily fast, in dependence on corresponding choice of $\boldsymbol{p} = \{p_j\}$.

7.3 Quasi-stationary Distributions

The purpose of this section is to set into a probabilistic framework results on the pointwise convergence of the Perron–Frobenius eigenvectors of truncations (supposed irreducible): that is, to consider in a probabilistically relevant framework truncation results of the kind of Theorem 6.9 (and Theorem 6.8), whereas so far in this chapter we have followed through, rather, the consequences of Theorem 6.10, which in essence avoids the Perron–Frobenius structure of truncations.

The motivating result here is Theorem 4.6 which shows if Q is a finite primitive strictly substochastic matrix, then its *probability-normed left Perron–Frobenius eigenvector* has a probabilistic interpretation akin to the stationary-limiting distribution corresponding to primitive stochastic P.

We shall thus call the probability-normed left Perron–Frobenius eigenvector of a finite irreducible substochastic matrix Q (which may be stochastic) a quasi-stationary distribution. This notion thus generalizes that of a stationary distribution (which is unique for irreducible stochastic P). The next generalization needed is that of a positive-recurrent stochastic matrix P which we may apply to the countable case. The following shows how to do this.

Lemma 7.4. *Let Q be an irreducible substochastic (perhaps not strictly) matrix with convergence parameter R, and suppose Q is R-positive. Then either $R > 1$, or $R = 1$ and the matrix is stochastic and positive-recurrent.*

PROOF. Since the elements $q_{ij}^{(k)}$ of Q^k may be interpreted as probabilities it is clear that the convergence radius R of

$$P_{ij}(z) = \sum_{k=0}^{\infty} q_{ij}^{(k)} z^k,$$

satisfies $R \geq 1$ (see Theorem 6.1) as we have already noted in Chapter 6. If $R = 1$, since $Q\mathbf{1} \leq \mathbf{1}$ from substochasticity of Q, $\mathbf{1}$ is an R-subinvariant-vector, and by the assumed R-positivity and (a dual of) Theorem 6.2, $Q\mathbf{1} = \mathbf{1}$, so Q is stochastic; and in view of Theorem 6.6 is positive-recurrent. □

Definition 7.2. An irreducible substochastic R-positive matrix Q is said to have a quasi-stationary distribution vector $\boldsymbol{v}$, if its unique (to constant multiples) R-invariant measure x' may be probability-normed, in which case $\boldsymbol{v} = x/x'\mathbf{1}$.

An analogue of Theorem 4.6 can now be proved {had we developed the theory of §6.2 more extensively for Q of the kind described in Definition 7.2} which lends analogous meaning to $\boldsymbol{v}$ as in the finite case.[1] We shall not do this, but turn to the truncation setting.

If we assume Q infinite irreducible substochastic and R-positive, and all but a finite number of its northwest corner truncations $_{(n)}Q$, $n \geq 1$, irreducible (for convenience), of interest is the asymptotic behaviour as $n \to \infty$ of the quasi-stationary distributions, $_{(n)}\boldsymbol{v}$, of $_{(n)}Q$. In particular, if Q has a quasi-stationary distribution, $\boldsymbol{v}$, the question arises: under what conditions is it true that $_{(n)}\boldsymbol{v} \to \boldsymbol{v}$ elementwise, in analogy with our investigations of §7.2? Theorem 6.9 provides the information that the *unit-normed* Perron–Frobenius left-eigenvectors converge to the *unit-normed* R-invariant measure of Q, but we have no information on the consequences of probability norming.

We may, however, apply the condition of *tightness* of the set of quasi-stationary distributions $\{_{(n)}\boldsymbol{v}\}$ in a manner analogous to that in §7.2.

Theorem 7.4. *For infinite irreducible substochastic and R-positive Q with all but a finite number of* $_{(n)}Q$, $n \geq 1$, *irreducible, if the set of probability distributions* $\{_{(n)}\boldsymbol{v}\}$, $n \geq 1$, *is tight (i.e. any infinite subsequence possesses an infinite subsequence converging elementwise to a probability vector) then Q possesses a quasistationary distribution vector* $\boldsymbol{v}$, *and* $_{(n)}\boldsymbol{v} \to \boldsymbol{v}$. *Conversely, if for some probability distribution vector* $\boldsymbol{v}$, $_{(n)}\boldsymbol{v} \to \boldsymbol{v}$ *then the set* $\{_{(n)}\boldsymbol{v}\}$ *is tight, and* $\boldsymbol{v}$ *is the (unique) quasistationary distribution vector of Q.*

PROOF. If $_{(n)}R$ denotes the convergence parameter of $_{(n)}Q$ (i.e. the reciprocal of its Perron–Frobenius eigenvalue), we have from Theorem 6.8 that

[1] See Seneta and Vere-Jones (1966).

${}_{(n)}R \downarrow R$. For the rest we need only imitate arguments in §7.2, and shall do so for completeness.

Suppose $\{{}_{(n)}\boldsymbol{v}\}$ is tight and suppose for some j, say $j = j_0$, and appropriate subsequences $\{n_i\}$, $\{n_k\}$

$$\lim_{i\to\infty} {}_{(n_i)}v_{j_0} = \limsup_{n\to\infty} {}_{(n)}v_{j_0} > \lim_{k\to\infty} {}_{(n_k)}v_{j_0}$$

$$= \liminf_{n\to\infty} {}_{(n)}v_{j_0}. \tag{7.16}$$

By tightness, taking subsequences if necessary

$$\lim_{i\to\infty} {}_{(n_i)}\boldsymbol{v} = \boldsymbol{v}_1, \qquad \lim_{k\to\infty} {}_{(n_k)}\boldsymbol{v} = \boldsymbol{v}_2$$

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are probability vectors which differ in at least the j_0th element. By Fatou's Lemma applied to the systems

$${}_{(n_i)}R\, {}_{(n_i)}\boldsymbol{v}'\, {}_{(n_i)}Q = {}_{(n_i)}\boldsymbol{v}', \qquad {}_{(n_k)}R\, {}_{(n_k)}\boldsymbol{v}'\, {}_{(n_k)}Q = {}_{(n_k)}\boldsymbol{v}'$$

we find

$$R\boldsymbol{v}_1' Q \le \boldsymbol{v}_1', \qquad R\boldsymbol{v}_2' Q \le \boldsymbol{v}_2'.$$

By R-positivity of Q and Theorem 6.2,

$$R\boldsymbol{v}_1' Q = \boldsymbol{v}_1', \qquad R\boldsymbol{v}_2' Q = \boldsymbol{v}_2'.$$

and $\boldsymbol{v}_1 = \boldsymbol{v}_2$ which is a contradiction to (7.16). Hence

$$\lim_{n\to\infty} {}_{(n)}v_j = v_j, \qquad j \ge 1,$$

exists, by tightness $\boldsymbol{v} = \{v_j\}$ is a probability vector, and by a repetition of the foregoing argument, is a quasi-stationary vector of Q, and the unique one.

Conversely, if ${}_{(n)}\boldsymbol{v} \to \boldsymbol{v}$ for some probability vector $\boldsymbol{v}$, then $\{{}_{(n)}\boldsymbol{v}\}$ is tight. A repetition of the foregoing argument establishes that $\boldsymbol{v}$ must be a quasi-stationary distribution of Q, and hence the unique one. □

Corollary 1. *If Q has no quasi-stationary distribution, $\{{}_{(n)}\boldsymbol{v}\}$ cannot converge elementwise to a probability vector.*

Corollary 2. *If $Q = P$ is assumed stochastic, but the assumption of R-positivity is dropped entirely, and ${}_{(n)}\boldsymbol{v} \to \boldsymbol{v}$ for some probability vector $\boldsymbol{v}$, then P is positive-recurrent, and $\boldsymbol{v}$ is its unique stationary distribution.*

PROOF. As before we find

$$R\boldsymbol{v}'P \le \boldsymbol{v}'.$$

Since $\boldsymbol{v}'\mathbf{1} = 1$, and $R \ge 1$, a contradiction results unless $R = 1$ and

$$\boldsymbol{v}'P = \boldsymbol{v}'. \qquad \square$$

EXAMPLE. We return to the example of §6.5 in the light of the present section.

The (doubly infinite) matrix T considered there when $h = \infty$ has the property that

$$Q = T'$$

is strictly substochastic since

$$Q\mathbf{1} = T'\mathbf{1} = s = \{s_j\}$$

and $s_0 = 1$ while the other s_j (see (6.6)) are positive but uniformly bounded from unity. The finite matrix $\{t_{ij}\}$, $i, j = -n, \ldots, -1, 0, 1, \ldots, n$ of dimensions $(2n + 1) \times (2n + 1)$ when transposed may be denoted by ${}_{(n)}Q$ and is the $(2n + 1) \times (2n + 1)$ "centre" truncation of Q. It has been noted that Q and ${}_{(n)}Q$ are irreducible. The vector $y^{(n)}$ is the quasi-stationary distribution vector of ${}_{(n)}Q$, and, in relation to the matrix ${}_{(h)}Q$ the conclusion

$$a^{(k)} \xrightarrow{k \to \infty} y^{(h)}$$

is a manifestation of Theorem 4.6. It is, further, shown that

$$y_0^{(n)} \geq \delta > 0,$$

and hence through some subsequence $\{n_k\}$ of the integers

$$y^{(n_k)} \to y, \qquad \text{with } 0 < \mathbf{1}'y \leq 1$$

and Q is R-positive, with y' a multiple of the unique R-invariant measure of Q.

Hence Q has a quasi-stationary distribution: viz. $y/y'\mathbf{1}$, but it is not clear that $\mathbf{1}'y = 1$ (i.e. that y itself forms the quasi-stationary distribution vector); nor that $y^{(n)} \to y$ as $n \to \infty$. Tightness of the sequence $\{y^{(n)}\}$ would suffice, from Theorem 7.4, to establish these facts, and we shall now show that tightness obtains, *under certain additional assumptions*.

We again consider (doubly infinite) Q modified so that all entries are replaced by zero except the (i, j) entries where $i, j = -n, \ldots, -1, 0, 1, \ldots, n$; and consider $y^{(n)}$ to be the infinite extended left probability-normed Perron–Frobenius eigenvector of the truncation ${}_{(n)}Q$.

The additional constraints which we now impose are

$$s_0 = 1 > s_1 = s_{-1} \geq s_2 = s_{-2} \geq s_3 = s_{-3} \geq \cdots.$$

The consequent symmetry about the index 0 in the matrix ${}_{(n)}Q$ clearly induces a symmetry about this index in the vector $y^{(n)} = \{y_i^{(n)}\}$

$$y_i^{(n)} = y_{-i}^{(n)}, \qquad i = 1, \ldots, n. \tag{7.17}$$

Now,

$$y_i^{(n)} = \{bs_{i-1}y_{i-1}^{(n)} + (1 - 2b)s_i y_i^{(n)} + bs_{i+1}y_{i+1}^{(n)}\}/\gamma(n)$$

$|i| \leq n$, where $\gamma(n) = s'y^{(n)} - 2bs_n y_n^{(n)}$. Thus

$$\gamma(n) \sum_{i=-n}^{n} i^2 y_i^{(n)} = (1-2b) \sum_{i=-n}^{n} i^2 s_i y_i^{(n)}$$

$$+ b\left(\sum_{i=-n}^{n} i^2 s_{i-1} y_{i-1}^{(n)} + \sum_{i=-n}^{n} i^2 s_{i+1} y_{i+1}^{(n)} \right)$$

$$= (1-2b) \sum_{i=-n}^{n} i^2 s_i y_i^{(n)}$$

$$+ b\left(\sum_{j=-n-1}^{n-1} (j+1)^2 s_j y_j^{(n)} + \sum_{j=-n+1}^{n+1} (j-1)^2 s_j y_j^{(n)} \right).$$

On expanding the squared terms, and using the assumed symmetry of the s_j's about 0:

$$= (1-2b) \sum_{i=-n}^{n} i^2 s_i y_i^{(n)}$$

$$+ 2b\left(\sum_{j=-n}^{n-1} j^2 s_j y_j^{(n)} + \sum_{j=-n}^{n-1} s_j y_j^{(n)} + \sum_{j=-n}^{n-1} j s_j y_j^{(n)} - \sum_{j=-n+1}^{n} j s_j y_j^{(n)} \right)$$

$$\leq (1-2b) \sum_{i=-n}^{n} i^2 s_i y_i^{(n)} + 2b\left(\sum_{j=-n}^{n-1} j^2 s_j y_j^{(n)} + \sum_{j=-n}^{n-1} s_j y_j^{(n)} \right)$$

$$\leq \sum_{i=-n}^{n} i^2 s_i y_i^{(n)} + 2b \sum_{j=-n}^{n} s_j y_j^{(n)}.$$

Then, if we put $M_2(n) = \sum_{i=-n}^{n} i^2 y_i^{(n)}$ (the "second moment" of the probability distribution $y^{(n)}$), we have

$$M_2(n) \leq \frac{\sum_i i^2 s_i y_i^{(n)} + 2b \sum_j s_j y_j^{(n)}}{\sum_j s_j y_j^{(n)} - 2bs_n y_n^{(n)}}$$

$$= \left\{ 2b + \frac{\sum_i i^2 s_i y_i^{(n)}}{\sum_j s_j y_j^{(n)}} \right\} \left\{ 1 - \frac{2bs_n y_n^{(n)}}{\sum_j s_j y_j^{(n)}} \right\}^{-1}. \tag{7.18}$$

Let us now focus on the two terms on the right of (7.18) in turn. Since, putting $v_j = s_j$, $j \neq 0$, $v_0 = s_1 = s$

$$\sum_i i^2 v_i y_i^{(n)} - \left(\sum_i i^2 y_i^{(n)} \right) \left(\sum_k v_k y_k^{(n)} \right)$$

$$= \frac{1}{2} \sum_{i,k} (i^2 - k^2)(v_i - v_k) y_i^{(n)} y_k^{(n)} \leq 0$$

by the assumed symmetry and monotonicity of the s_i with increasing $|i|$

$$\begin{aligned} 2b + \frac{\sum_i i^2 s_i y_i^{(n)}}{\sum_j s_j y_j^{(n)}} &\le 2b + \frac{M_2(n) \sum_k v_k y_k^{(n)}}{y_0^{(n)}(1-s) + \sum_j v_j y_j^{(n)}} \\ &= 2b + \left\{M_2(n) \Big/ \left(\frac{1 + y_0^{(n)}(1-s)}{\sum_k v_k y_k^{(n)}}\right)\right\} \\ &\le 2b + \left\{\frac{sM_2(n)}{(s + y_0^{(n)}(1-s))}\right\} \end{aligned}$$

since $\sum_j v_j y_j^{(n)} \le s \sum_j y_j^{(n)} = s$. From (7.18) where we shall put $\omega(n) = 2bs_n y_n^{(n)}/\sum_j s_j y_j^{(n)}$, and the above,

$$M_2(n) \le \{2b + \{sM_2(n)/(s + y_0^{(n)}(1-s))\}\}\{1 - \omega(n)\}^{-1},$$

so

$$M_2(n) \le 2b\left\{1 - \omega(n) - \frac{s}{s + y_0^{(n)}(1-s)}\right\}^{-1}$$

providing the expression $\{\cdots\}$ is positive. Now

$$\begin{aligned} \omega(n) = \frac{2bs_n y_n^{(n)}}{\sum_j s_j y_j^{(n)}} &\le \min\left(\frac{2bs_n y_n^{(n)}}{s_0 y_0^{(n)}}, \frac{2bs_n y_n^{(n)}}{2s_n y_n^{(n)}}\right) \\ &\le \min\left(\frac{2bs_n}{y_0^{(n)}}, b\right) \end{aligned}$$

and recalling that

$$\liminf_{n\to\infty} y_0^{(n)} \ge 1 - 2b(1-s)^{-1}$$

a number of possible additional conditions will ensure that $\{\cdots\}$ is ultimately positive and

$$\limsup_{n\to\infty} M_2(n) \equiv M_2 < \infty$$

e.g. $s_n \to 0$ as $n \to \infty$ (which ensures $\omega(n) \to 0$); or merely $s < (1-2b)(1-b)$.

It remains to show that the boundedness of $M_2(n)$ ensures the tightness of the set of probability distribution vectors $\{y^{(n)}\}$ in the present circumstances. That tightness obtains now follows from well-known arguments in probability theory, but for completeness we shall give a direct proof. For any $c > 0$

$$\sum_{|i|>c} y_i^{(n)} \le c^{-2} \sum_{|i|>c} i^2 y_i^{(n)} \le c^{-2}(M_2 + \delta)$$

for some positive $\delta > 0$ independent of n; so for sufficiently large c,

$$\sum_{|i|>c} y_i^{(n)} < \varepsilon \quad \text{or, equivalently} \quad \sum_{|i|\le c} y_i^{(n)} > 1 - \varepsilon$$

for all n. We have already seen that there is a subsequence $\{n_k\}$ of the positive integers such that $y^{(n_k)} \to y$ as $k \to \infty$ where $0 < \mathbf{1}'y \le 1$. We could similarly, *starting* from any infinite *subsequence* of the integers, obtain a further subsequence for which a similar proposition was true, so let $\{n_k\}$ denote such a further subsequence, and $y = \{y_i\}$ the limit vector corresponding to it. Then for arbitrary ε, and correspondingly chosen c, applying the above inequalities with $n = n_k$, and letting $k \to \infty$, we obtain

$$\sum_{|i| \leq c} y_i \geq 1 - \varepsilon;$$

and letting $c \to \infty$ yields $\mathbf{1}'y \geq 1 - \varepsilon$ so $\mathbf{1}'y = 1$, which establishes tightness by establishing that y is a probability vector.

Bibliography and Discussion

Theorem 7.1 and its Corollary are based on the approach of Seneta (1967a, 1968a). Lemma 7.1 is taken from Takahashi (1973), following Bott and Mayberry (1954). The application of results such as this to give conclusions such as Theorem 7.2 and its Corollary via a truncation argument is due to Tweedie (1980), whose statement and proof of such results is somewhat different. The argument of the remainder of §7.1, and of §7.2 is condensed and synthesized from Golub and Seneta (1974), Allen, Anderssen and Seneta (1977), and Seneta (1980). Related papers are Golub and Seneta (1973) and Paige, Styan and Wachter (1975). The reader interested in practical situations where such approximative procedures from finite truncations are necessary should consult Beckmann, McGuire and Winsten (1956) and Bithell (1971).

The method of "stochasticizing" truncations of an infinite stochastic matrix was suggested by Sarymsakov (1945a) and used by him for other purposes. As regards approximative utility in regard to the infinite stationary distribution vector of a positive-recurrent matrix P, the *theoretical* results presently available are those in our §7.2, taken from Golub and Seneta (1974) and Seneta (1980). The papers of Allen, Anderssen and Seneta (1977) and Seneta (1980) concern themselves with numerical and computational aspects.

Lemma 7.4 is from Seneta and Vere-Jones (1966) where an extensive discussion with a number of examples of the notion of a quasi-stationary distribution in Markov chains with a denumerable infinity of states is given. Theorem 7.4 is suggested by §4 of Seneta (1980). The tightness argument for the example is adapted from Moran (1976).

Exercises

7.1. "Slowly spreading" infinite stochastic matrices $P = \{p_{ij}\}$, $i, j \geq 1$ are defined by the condition

$$p_{i,\, i+r} = 0, \qquad i \geq 1, \qquad r \geq 2$$

(i.e. all elements above the superdiagonal are zero). Irreducibility, which we assume, clearly implies $p_{i,\,i+1} > 0, i \geq 1$: this situation clearly subsumes both the Example of §5.6, used in the proof of Theorem 6.11, and that of Exercise 6.5 above.

Show that

$$\Delta_n(1) = \det\,[{}_{(n)}I - {}_{(n)}P] = \prod_{i=1}^{n} p_{i,\,i+1}$$

so that

$$\lim_{n\to\infty} \Delta_n(1) > 0 \Leftrightarrow \sum_{i=1}^{\infty} (1 - p_{i,\,i+1}) < \infty.$$

Note that this condition is far from satisfied in the situation of Exercise 6.5, even if $p > q$ i.e. $p > \frac{1}{2}$ (after adjusting the index set in that question approximately).

(Adapted from Kemeny, 1966)

7.2. For the matrix $P = \{p_{ij}\}$, $i, j \geq 1$, defined by

$$p_{i,\,i+1} = p_i, \qquad i \geq 1$$
$$p_{i,\,i-1} = q_i = 1 - p_i, \qquad i \geq 2$$
$$p_{1,\,1} = q_1$$

where $0 < p_i < 1$ for all $i \geq 1$, show that an invariant measure $x' = \{x(i)\}$ always exists and is unique to constant multiples by solving the difference equation

$$p_{i-1}x(i-1) + q_{i+1}x(i+1) = x(i), \qquad i \geq 1$$

under the auxiliary condition $x(2) = (p_1/q_2)x(1)$, where $x(1)$ is taken as positive but otherwise arbitrary.

Hence use the result of Exercise 7.1 to show that there exist stochastic P for which an invariant measure exists, even though

$$\lim_{n\to\infty} {}_{(n)}\Delta(1) > 0$$

(again, contrary to expectations, but again only in the transient case).

7.3. Show that if P is an infinite irreducible stochastic matrix and Q is any square northwest corner truncation of P, then $Q^k \to 0$ as $k \to \infty$. (*Hint:* Assuming Q is $(n \times n)$, so its index set is $I = \{1, 2, \ldots, n\}$, show that in P, for each $i \in I$, $i \to \{n+1, n+2, \ldots\}$, and invoke Theorem 4.3.)

7.4. Assume P is an infinite stochastic matrix with a single essential class of indices, which, further, contains an infinite number of indices. Show that if Q is any square northwest corner truncation of P, then $Q^k \to 0$ as $k \to \infty$. (This result generalizes that of Exercise 7.3.)

7.5. Assume P is an infinite stochastic matrix, which contains one and only one positive-recurrent class of indices, which consists of an infinite number of indices. (All other indices are thus null-recurrent or transient). Show that the conclusion of Exercise 7.4 continues to hold.

Construct further results along the lines of Exercises 7.3–7.5.

7.6. Suppose the infinite stochastic matrix P is defined by $P = \mathbf{1}\boldsymbol{p}'$ where $\boldsymbol{p}$ is a probability vector whose elements are all positive. Show that P is positive recurrent. If ${}_{(n)}\boldsymbol{p}'$ comprises the first n elements of $\boldsymbol{p}'$, whose sequence of elements is now assumed non-increasing, show that

$$\kappa({}_{(n)}I - {}_{(n)}P') \sim 2/(1 - \mathbf{1}'\boldsymbol{p}_n).$$

(This simple example shows that the condition number of the system (7.8) may approach infinity arbitrarily slowly or fast.)

(Seneta, 1976)

7.7. Show by example that an infinite stochastic P which is a Markov matrix may have finite truncations ${}_{(n)}P$ which are eventually all stochastic. Relate this result to that of Exercise 7.4, to generalize the Corollary to Theorem 7.3.

7.8. Noting the arbitrariness in rate in Exercise 7.6 where P is a structurally simple example of a Markov matrix, and the arbitrariness in the example concluding §7.2, obtain explicit forms for ${}_{(n)}\pi'$ and π' for the Markov matrix of Exercise 7.6 and investigate the pointwise convergence rate. (*Hint:* For a probability vector $\boldsymbol{p}$, $\boldsymbol{p}'(\mathbf{1}\boldsymbol{p}') = \boldsymbol{p}'$).

7.9. Show that the linear system (7.11) may be written as

$$(({}_{(n)}I - f_1\mathbf{1}')({}_{(n)}I - {}_{(n)}P') + \delta f_1\mathbf{1}')_{(n)}\pi = \delta f_1$$

in terms of the (unaugmented) $(n \times n)$ northwest corner truncation ${}_{(n)}P$.

APPENDICES

APPENDIX A
Some Elementary Number Theory

We consider the set of all integers, both non-negative and negative.

Lemma A.1. *Any subset, S, of the integers containing at least one non-zero element and closed under addition and subtraction contains a least positive element and consists of all multiples of this integer.*

PROOF. Let $a \in S$, $a \neq 0$. Then S contains the difference $a - a = 0$, and also $0 - a = -a$. Consequently there is at least one positive element, $|a|$, in S, and hence there is a smallest positive element, b, in S. Now, S must contain all integral multiples of b, for if it contains nb, $n = 1, 2, \ldots$ etc., then it must contain $(n + 1)b = nb + b$, and we know it contains b. Moreover, $(-n)b = 0 - (nb)$ is the difference of two elements in S, for $n = 1, 2, \ldots$ and so S contains all negative multiples of b also.

We now show that S can contain nothing but integral multiples of b. For if c is any element of S, there exist integers q and r such that $c = bq + r$, $0 \leq r < b$ (qb is the multiple of b closest to c from below; we say $c \equiv r(\text{mod } b)$). Thus $r = c - bq$ must also be in S, since it is a difference of numbers in S. Since $r \in S$, $r \geq 0$, $r < b$, and b is the least positive integer in S, it follows $r = 0$, so that $c = qb$. □

Definition A.1. Every positive integer which divides all the integers $a_1, a_2, \ldots, a_k$ is said to be a *common divisor* of them. The largest of these common divisors, is said to be the *greatest common divisor* (g.c.d.). This number is a well defined positive number if not all $a_1, a_2, \ldots, a_k$ are zero, which we assume henceforth.

Lemma A.2. *The greatest common divisor of $a_1, a_2, \ldots, a_k$, say d, can be expressed as a "linear combination", with integral coefficients, of $a_1, a_2, \ldots,$*

a_k, *i.e.*

$$d = \sum_{i=1}^{k} b_i a_i, \; b_i \text{ integers.}$$

PROOF. Consider the set S of all numbers of the form $\sum_{i=1}^{k} b_i a_i$. For any two such

$$\sum_{i=1}^{k} b_i^{(1)} a_i \pm \sum_{i=1}^{k} b_i^{(2)} a_i = \sum_{i=1}^{k} (b_i^{(1)} \pm b_i^{(2)}) a_i,$$

and hence the set S of such numbers is closed under addition and subtraction, hence by Lemma A.1 consists of all multiples of some minimum positive number

$$v = \sum_{i=1}^{k} b_i a_i.$$

Thus d, the greatest common divisor of $a_1, \ldots, a_k$, must divide v, so that $0 < d \leq v$. Now each a_i is itself a member of S (choose the bs so that $b_i = 1$ and all other bs zero), so that each a_i is a multiple of v, by Lemma A.1. Thus a contradiction arises unless $d = v$, since d is supposed to be the greatest common divisor of the as. □

Definition A.2. Let a_i, $i = 1, 2, \ldots$ be an infinite set of positive integers. If d_k is the greatest common divisor of $a_1, \ldots, a_k$, then the greatest common divisor of a_i, $i = 1, 2, \ldots$, is defined by

$$d = \lim_{k \to \infty} d_k.$$

The limit $d \geq 1$ clearly exists (since the sequence $\{d_k\}$ is non-increasing).

Moreover d is an integer, and must be attained after a finite number of k, since all the d_ks are integers.

Lemma A.3. *An infinite set of positive integers, $V = \{a_i, i \geq 1\}$, which is closed under addition (i.e. if two numbers are in the set, so is their sum), contains all but a finite number of positive multiples of its greatest common divisor.*

PROOF. We can first divide all elements in the infinite set V by the greatest common divisor d, and thus reduce the problem to the case $d = 1$, which we consider henceforth.

By the fact that $d = 1$ must be the greatest common divisor of some finite subset, $a_1, a_2, \ldots, a_k$, of V, it follows from Lemma A.2 that there is a linear combination of these as such that

$$\sum_{i=1}^{k} a_i n_i = 1,$$

where the n_is are integers. Let us rewrite this as

$$m - n = 1 \tag{A.1}$$

where m is the sum of the positive terms, and $-n$ the sum of the negative terms. Clearly both n and m are positive integer linear combinations of the a_is and so belong to the set V. Now, let q be any integer satisfying $q \geq n \times (n - 1)$; and write

$$q = an + b, \qquad 0 \leq b < n$$

where a is a positive integer, $a \geq (n - 1)$. Then using (A.1)

$$q = an + b(m - n)$$

$$q = (a - b)n + bm$$

so that q is also in the set V. Hence all sufficiently large integers are in the set V, as required. □

Corollary. *If $a_1, a_2, \ldots, a_k$ are positive integers with g.c.d. unity, then any sufficiently large positive integer q may be expressed as*

$$q = \sum_{i=1}^{k} a_i p_i$$

where the p_i are non-negative integers.

We conclude that this subsection by an application which is of relevance in Chapter 6, and whose proof involves, in a sense, a tightening of the kind of argument given in Lemma A.3.

Lemma A.4.[1] *Let u_i $(i = 0, 1, 2, \ldots)$ be non-negative numbers such that, for all $i, j \geq 0$*

$$u_{i+j} \geq u_i u_j.$$

Suppose the set V of those integers $i \geq 1$ for which $u_i > 0$ is non-empty and has g.c.d., say d, which satisfies $d = 1$. Then

$$u = \lim_{n \to \infty} u_n^{1/n}$$

exists and satisfies $0 < u \leq \infty$; further, for all $i \geq 0$, $u_i \leq u^i$.

Proof. The set V is closed under addition, in virtue of $u_{i+j} \geq u_i u_j$, and since $d = 1$, by Lemma A.3, V contains all sufficiently large integers. Hence for any $r \in V$ there exists an $s \in V$, $s > r$, such that the g.c.d. of r and s is unity. Thus by Lemma A.2 we have

$$1 = b_1 r + b_2 s$$

[1] Due to Kingman (1963). This is an analogue of the usual theorems on supermultiplicative or subadditive functions (e.g. Hille & Phillips 1957, Theorem 7.6.1; Khintchine, 1969, §7).

for some integers b_1, b_2. Not both of b_1 and b_2 can be strictly positive, since $r, s \geq 1$. Assume for the moment

$$b_1 \equiv a > 0, \qquad -b_2 \equiv b \geq 0.$$

Let n by any integers such that $n \geq rs$, and let $k = k(n)$ be the smallest positive integer such that

$$b/r \leq k/n \leq a/s;$$

such an integer certainly exists since

$$n(a/s - b/r) = n(ar - bs)/rs = n/rs \geq 1.$$

Then

$$n = Ar + Bs$$

where $A = na - k(n)s$, $B = k(n)r - nb$ are non-negative integers. Now, from the assumption of the Lemma,

$$u_n = u_{Ar+Bs} \geq u_{Ar}u_{Bs} \geq \cdots \geq u_r^A u_s^B, \qquad (n \geq rs).$$

Thus

$$u_n^{1/n} \geq u_r^{a-(ks/n)} u_s^{(kr/n)-b}.$$

Letting $n \to \infty$, and noting $k(n)/n \to b/r$, we see that

$$\liminf_{n\to\infty} u_n^{1/n} \geq u_r^{a-bs/r} u_s^{b-b} = u_r^{1/r}.$$

This holds for any $r \in V$ and so for any sufficiently large integer r. Hence

$$\liminf_{n\to\infty} u_n^{1/n} \geq \limsup_{r\to\infty} u_r^{1/r}$$

which shows

$$u = \lim_{n\to\infty} u_n^{1/n}$$

exists. Further, again from the inequality $u_{i+j} \geq u_i u_j$,

$$u_{ki} \geq u_i^k, \qquad i, k \geq 0$$

so that

$$(u_{ki}^{1/ki})^i \geq u_i$$

and letting $k \to \infty$,

$$u^i \geq u_i, \qquad i \geq 1$$

and since $u_i > 0$ for some $i \geq 1$ the proof is complete, apart from the case $i = 0$, which asserts

$$1 \geq u_0.$$

The truth of this follows trivially by putting $i = j = 0$ in the equality, to obtain $u_0 \geq u_0^2$.

We need now to return to the possibility that $b_1 \leq 0$, $b_2 > 0$. If we write $b_2 = a$, $-b_1 = b$, so that $1 = as - br$, then the roles of s and r need to be changed in the subsequent argument and, apart from this $k = k(n)$ needs to be chosen as the *largest* integer such that

$$b/s \leq k/n \leq a/r.$$

This leads eventually once more to

$$\liminf_{n \to \infty} u_n^{1/n} \geq u_r^{1/r}$$

and the rest is as before. □

Corollary.

$$u = \sup_n u_n^{1/n}.$$

APPENDIX B
Some General Matrix Lemmas

Lemma B.1. *If A is a finite $n \times n$ matrix with real or complex elements such that $A^k \to 0$ elementwise as $k \to \infty$, then $(I - A)^{-1}$ exists and*

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$$

convergence being elementwise. ($A^0 = I$ by definition.)

PROOF. First note that

$$(I - A)(I + A + \cdots + A^{k-1}) = I - A^k.$$

Now, for k sufficiently large, A^k is uniformly close to the zero matrix, and so $I - A^k$ is to I, and is therefore non-singular. (More specifically, by the continuity of the eigenvalues of a matrix if its elements are perturbed, the eigenvalues of $I - A^k$ must be close to those of I for large k, the latter being all 1; hence $I - A^k$ has no zero eigenvalues, and is therefore non-singular.) Taking determinants

$$\det(I - A)\det(I + A + \cdots + A^{k-1}) = \det(I - A^k) \neq 0,$$

therefore $$\det(I - A) \neq 0.$$

Therefore

$$(I - A)^{-1} \text{ exists, and } I + A + \cdots + A^{k-1} = (I - A)^{-1}(I - A^k).$$

Letting $k \to \infty$ completes the proof of the assertion. □

Corollary. *Given an $n \times n$ matrix A, for all complex z sufficiently close to* 0, $(I - zA)^{-1}$ *exists, and*

$$(I - zA)^{-1} = \sum_{k=0}^{\infty} z^k A^k$$

in the sense of elementwise convergence.

PROOF. Define δ by

$$\delta = \max_{i,j} |a_{ij}|.$$

Then putting $A^k = \{a_{ij}^{(k)}\}$, it follows from matrix multiplication that $|a_{ij}^{(2)}| \leq n\,\delta^2$, and, in general

$$|a_{ij}^{(k)}| \leq n^{k-1}\,\delta^k, \qquad \text{all } i, j = 1, 2, \ldots, n.$$

Hence $A^k z^k \to 0$ if $|z| < (n\,\delta)^{-1}$, and the result follows from Lemma B.1. □

Note: We shall call the quantity $R(z) \equiv (I - zA)^{-1}$ the *resolvent* of A, for those z for which it exists, although this name is generally given to $(zI - A)^{-1}$.

The above Corollary also provides an analytical method for finding A^k for arbitrary k for real matrices of small dimension. It is necessary only to find the resolvent, and pick out the coefficient of z^k. Note that

$$(I - zA)^{-1} = \text{Adj}\,(I - zA)/\det\,(I - zA),$$

and the roots of $\det\,(I - zA) = z^n \det\,(z^{-1}I - A)$ for $z \neq 0$, are given by $z_i = \lambda_i^{-1}$ where the λ_i are the non-zero eigenvalues of A. Hence

$$\det\,(I - zA) = \prod_{j=1}^{n} (1 - z\lambda_j)$$

where the λ_j are the eigenvalues of A, and the singularities of the *resolvent* are all poles, possibly non-simple, since the elements of the *adjoint* matrix are polynomials of degree at most $n - 1$.

EXAMPLE. Find P^k for the (stochastic) matrix:

$$P = \begin{bmatrix} 1 & 0 & 0 \\ p_1 & p_2 & p_3 \\ 0 & 0 & 1 \end{bmatrix}, \quad p_i > 0, \quad i = 1, 2, 3, \quad \sum_{i=1}^{3} p_i = 1.$$

$$(I - zP) = \begin{bmatrix} 1 - z & 0 & 0 \\ -zp_1 & 1 - zp_2 & -zp_3 \\ 0 & 0 & 1 - z \end{bmatrix}$$

$$\det\,(I - zP) = (1 - z)^2(1 - zp_2).$$

Therefore

$$(I - zP)^{-1} = (1 - z)^{-2}(1 - zp_2)^{-1}$$
$$\times \begin{bmatrix} (1-z)(1-zp_2) & 0 & 0 \\ zp_1(1-z) & (1-z)^2 & zp_3(1-z) \\ 0 & 0 & (1-z)(1-zp_2) \end{bmatrix}.$$

By the use of partial fractions and the separating of components, we get eventually

$$(I - zP)^{-1} = (1 - z)^{-1} \begin{bmatrix} 1 & 0 & 0 \\ p_1(1-p_2)^{-1} & 0 & p_3(1-p_2)^{-1} \\ 0 & 0 & 1 \end{bmatrix}$$
$$+ (1 - p_2 z)^{-1} \begin{bmatrix} 0 & 0 & 0 \\ -p_1(1-p_2)^{-1} & 1 & -p_3(1-p_2)^{-1} \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus equating coefficients of z^k,

$$P^k = \begin{bmatrix} 1 & 0 & 0 \\ p_1(1-p_2)^{-1} & 0 & p_3(1-p_2)^{-1} \\ 0 & 0 & 1 \end{bmatrix}$$
$$+ p_2^k \begin{bmatrix} 0 & 0 & 0 \\ -p_1(1-p_2)^{-1} & 1 & -p_3(1-p_2)^{-1} \\ 0 & 0 & 0 \end{bmatrix}.$$

Lemma B.2. *If $A = \{a_{ij}\}$ is a finite $n \times n$ matrix with real or complex elements, the matrix series*

$$\sum_{k=0}^{\infty} t^k \frac{A^k}{k!}$$

converges elementwise; for each $t > 0$ the limit matrix is denoted by $\exp(tA)$, *in analogy to the scalar case. If B is also an $(n \times n)$ matrix, then if $AB = BA$*

$$\exp(A + B) = \exp(A) \cdot \exp(B).$$

Proof. Let $\delta = \max_{i,j} |a_{ij}|$. As in the Corollary to Lemma B.1

$$|a_{ij}^{(k)}| \le n^{k-1} \delta^k, \qquad \text{all } i, j = 1, 2, \ldots, n$$

and elementwise (absolute) convergence follows from the fact that

$$\sum_{k=1}^{\infty} t^k \frac{n^{k-1} \delta^k}{k!}$$

converges for any $t, \delta, n > 0$.

The second part of the assertion follows from this simply by multiplying out the series for $\exp(A)$, $\exp(B)$, and collecting appropriate terms, as is permissible under the circumstances. □

APPENDIX C
Upper Semi-continuous Functions

Let $f(x)$ be a mapping of a subset $\mathscr{A}$ of the Euclidean n-space R_n, into R_1 extended by the values $+\infty$ and $-\infty$.

Definition C.1. The function f is said to be upper semicontinuous on $\mathscr{A}$, if

$$\limsup_{k \to \infty} f(x_k) \leq f(x_0)$$

for any $x_0 \in \mathscr{A}$, where $\{x_k\}$ is any sequence contained in $\mathscr{A}$ such that $x_k \to x_0$ as $k \to \infty$.

Lemma C.1. (*a*) *For any (finite, denumerable, or even non-countable) set of functions* Λ *which are upper semi-continuous on* $\mathscr{A}$,

$$h(x) = \inf_{f \in \Lambda} f(x)$$

is also upper semi-continuous on $\mathscr{A}$.

(*b*) *An upper semi-continuous function defined on a compact* $\mathscr{A}$ *attains its supremum for some* $x_0 \in \mathscr{A}$.

PROOF. (*a*) By definition for every $f(x) \in \Lambda$

$$h(x) \leq f(x), \qquad \forall x \in \mathscr{A}.$$

Thus for any $x_0 \in \mathscr{A}$ and $x_k \to x_0$, with $\{x_k\}$ in $\mathscr{A}$

$$h(x_k) \leq f(x_k), \qquad f \in \Lambda$$

so that
$$\limsup_{k \to \infty} h(x_k) \leq \limsup_{k \to \infty} f(x_k) \leq f(x_0)$$

for any $f \in \Lambda$.

Thus $$\limsup_{k\to\infty} h(x_k) \le \inf_{f\in\Lambda} f(x_0) = h(x_0)$$

(*b*) Let $a = \sup_{x\in\mathscr{A}} f(x)$ for an f upper-semicontinuous on $\mathscr{A}$. We can find a sequence $\{x_k\} \subset \mathscr{A}$ such that

$$f(x_k) \to a.$$

Since $\mathscr{A}$ is a compact, there exists a convergent subsequence $\{x_{k_i}\}$, $i = 1, 2, \ldots$ converging to a value $x_0 \in \mathscr{A}$. By upper semicontinuity of f

$$a = \limsup_{i\to\infty} f(x_{k_i}) \le f(x_0) \le a.$$

Thus $$f(x_0) = a.$$ □

Bibliography

The following list includes all titles referred to in the text, and a very few others.

Akaike, H., (1977). On entropy maximization principle. In *Applications of Statistics*, ed. P. R. Krishnaiah. North Holland, Amsterdam, pp. 27–41.

Albert, E., (1963). Markov chains and λ-invariant measures. *J. Math. Anal. Applicns.*, **6**, 404–18.

Allen, B., Anderssen, R. S., and Seneta, E., (1977). Computation of stationary measures for infinite Markov chains. *TIMS Studies in the Management Sciences* **7** (*Algorithmic Methods in Probability*, ed. M. F. Neuts), pp. 13–23. North Holland, Amsterdam. [More extensively as: Numerical investigation of infinite stochastic matrices. Technical Report No. 48, Computer Centre, Australian National University, 1975.]

Allen, R. G. D., (1965). *Mathematical Economics* (2nd edition). Macmillan, London.

Anthonisse, J. M. and Tijms, H., (1977). Exponential convergence of products of stochastic matrices. *J. Math. Anal. Applicns.*, **59**, 360–4.

Bacharach, M., (1965). Estimating non-negative matrices from marginal data. *International Economic Review*, **6**, 294–310.

— (1970). *Biproportional Matrices and Input-Output Change*. Cambridge University Press, Cambridge.

Barker, G. P. and Turner, R. E. L., (1973). Some observations on the spectra of cone preserving maps. *Linear Algebra and Applicns.*, **6**, 149–53.

Bassett, L., Maybee, J., and Quirk, J., (1968). Qualitative economics and the scope of the correspondence principle. *Econometrica*, **36**, 544–63.

Bauer, F. L., (1965). An elementary proof of the Hopf inequality. Numenische Math., **7**, 331–7.

Bauer, F. L., Deutsch, E., and Stoer, J., (1969). Abschätzungen für die Eigenwerte positiver linearer Operatoren. *Linear Algebra and Applicns.*, **2**, 275–301.

Beckmann, M., McGuire, C. B., and Winsten, C., (1956). *Studies in the Economics of Transportation*. Cowles Commission for Research in Economics, Yale University Press, p. 42.

Bellman, R., (1960). *Introduction to Matrix Analysis*. McGraw Hill, New York.

Berman, A. and Plemmons, R. J., (1979). *Nonnegative Matrices in the Mathematical Sciences.* Academic Press, New York.

Bernardelli, H., (1941). Population waves. *Journal of the Burma Research Society*, **31**, 1–18.

Bernstein, S. N., (1946). *Teoriya Veroiatnostei* (4th edition). Gostehizdat, Moscow–Leningrad. (See esp. pp. 203–13, 465–84, which are reprinted in Bernstein (1964).)

— (1964). Classification of Markov chains and their matrices [in Russian]. *Sobranie Sochineniy: Tom IV, Teoriya Veroiatnostei i Matematicheskaia Statistika [1911–1946]*, pp. 455–83. Izd. Nauka.

Birkhoff, G., (1946). Tres observaciones sobre el algebra lineal. *Univ. Nac. Tucumán Rev.*, Ser A **5**, 147–50.

— (1957). Extensions of Jentzsch's theorem. *Trans. Amer. Math. Soc.*, **85**, 219–27.

— (1967). *Lattice Theory* (3rd edition). Amer. Math. Soc. Colloq. Publicns., Vol. XXV, Providence, R.I.

Birkhoff, G. and Varga, R. S., (1958). Reactor criticality and non-negative matrices. *J. Soc. Industr. Appl. Math.*, **6**, 354–77.

Bithell, J. F., (1971). Some generalized Markov chain occupancy processes and their application to hospital admission systems. *Review of the International Statistical Institute*, **39**, 170–84.

Blackwell, D., (1945). Finite non-homogeneous chains. *Ann. Math.*, **46**, 594–99.

Blagoveshchenskii, Yu. N., (1960). On ergodicity for schemes of series of Markov chains with finite number of states and discrete time [in Russian]. *Izvestia A.N. Uzbek. S.S.R.* Ser. fiz. mat. nauk **4**, 7–15.

Bott, R. and Mayberry, J. P., (1954). Matrices and trees. In Morgenstern (1954).

Brauer, A., (1952). Limits for the characteristic roots of a matrix IV: Applications to stochastic matrices. *Duke Math. J.*, **19**, 75–91.

— (1957a). The theorems of Ledermann and Ostrowski on positive matrices. *Duke Math. J.*, **24**, 265–74.

— (1957b). A new proof of theorems of Perron and Frobenius on non-negative matrices, I. Positive matrices. *Duke Math. J.*, **24**, 367–78.

— (1961). On the characteristic roots of power-positive matrices. *Duke Math. J.*, **28**, 439–45.

Brualdi, R., (1974). The DAD theorem for arbitrary row sums. *Proc. Amer. Math. Soc.*, **45**, 189–94.

Brualdi, R. A., Parter, S. V., and Schneider, H., (1966). The diagonal equivalence of a non-negative matrix to a stochastic matrix. *J. Math. Anal. Applicns.*, **16**, 31–50.

Buchanan, M. L. and Parlett, B. N., (1966). The uniform convergence of matrix powers. *Numerische Math.*, **9**, 51–4.

Buharaev, R. G., (1968). Sets of processions of eigenvalues of stochastic matrices. *Soviet Math. Dokl.*, **9**, 64–7.

Burger, E., (1957). Eine Bemerkung über nicht-negative Matrizen. *Z. Angew. Math. Mech.*, **37**, 227.

Bushell, P. J., (1973). On the projective contraction ratio for positive linear mappings. *J. London Math. Soc.*, **6**, 256–58.

Chatterjee, S. and Seneta, E., (1977). Towards consensus: some convergence theorems on repeated averaging. *J. Appl. Prob.*, **14**, 89–97.

Cherubino, S., (1957). *Calcolo Delle Matrici.* Edizione Cremonese, Roma.

Chung, K. L., (1967). *Markov Chains with Stationary Transition Probabilities* (2nd edition). Springer, Berlin.

Cohen, J. E., (1979a). Contractive inhomogeneous products of non-negative matrices. *Math. Proc. Cambridge Phil. Soc.*, **86**, 351–64.

— (1979b). Ergodic theorems in demography. *Bull. Amer. Math. Soc.*, **1**, 275–95.

Cohn, H., (1974). A ratio limit theorem for the finite non-homogeneous Markov

chains. *Israel J. Math.*, **19**, 329–34.
— (1976). Finite non-homogeneous Markov chains: asymptotic behaviour. *Adv. Appl. Prob.*, **8**, 502–16.
Collatz, L., (1942). Einschliessungenssatz für die characteristischen Zahlen von Matrizen. *Math. Zeit.*, **48**, 221–6.

Daley, D. J., (1969). Quasi-stationary behaviour of a left-continuous random walk. *Ann. Math. Statist.*, **40**, 532–9.
Dalkey, N. C., (1969). The Delphi method: an experimental study of group opinion. The Rand Corporation, RM-5888.
Darroch, J. N. and Seneta, E., (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Prob.*, **4**, 192–6.
Debreu, G. and Herstein, I. N., (1953). Nonnegative square matrices. *Econometrica*, **21**, 597–607.
Deming, W. E. and Stephan, F. F., (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.*, **11**, 427–44.
Denardo, E. V., (1977). Periods of connected networks and powers of non-negative matrices. *Mathematics of Operations Research*, **2**, 20–4.
Derman, C., (1954). A solution to a set of fundamental equations in Markov chains. *Proc. Amer. Math. Soc.*, **5**, 332–4.
— (1955). Some contributions to the theory of denumerable Markov chains. *Trans. Amer. Math. Soc.*, **79**, 541–55.
Deutsch, E. and Zenger, C., (1971). Inclusion domains for the eigenvalues of stochastic matrices. *Numerische Math.*, **18**, 182–92.
Dionísio, J. J., (1963/4, Fasc. 1). Matrizes não negativas. *Rev. Fac. Ciências Lisboa*, Ser A **10**, 5–35.
— (1963/4, Fasc. 2). Sobre gráficos fortemente conexos e matrices não negativas. *Rev. Fac. Ciências Lisboa*, Ser A **10**, 139–50.
Djoković, D. Ž., (1970). Note on non-negative matrices. *Proc. Amer. Math. Soc.*, **25**, 80–2.
Dmitriev, N. A., (1946). Characteristic roots of stochastic matrices [in Russian]. *Izvestia A.N. S.S.S.R.* Ser. mat., **10**, 167–84.
Dmitriev, N. A. and Dynkin, E. B., (1945). On the characteristic numbers of stochastic matrices [in Russian]. *Doklady A.N. S.S.S.R.*, **49**, 159–62.
Dobrushin, R. L., (1956). Central limit theorem for non-stationary Markov chains, I, II. *Theory Prob. Appl.*, **1**, 65–80, 329–83. (English translation.)
Doeblin, W., (1937). Le cas discontinu des probabilités en chaîne. *Publ. Fac. Sci. Univ. Masaryk* (Brno), no. **236**.
— (1938). Exposé de la théorie des chaînes simples constantes de Markoff à un nombre fini d'états. *Revue Mathématique* (*Union Interbalkanique*), **2**, 77–105.
Doob, J. L., (1959). Discrete potential theory and boundaries. *J. Math. Mech.*, **8**, 433–58.
Dulmage, A. L. and Mendelsohn, N. S., (1962). The exponent of a primitive matrix. *Canad. Math. Bull.*, **5**, 241–4.
— (1964). Gaps in the exponent set of primitive matrices. *Illinois J. Math.*, **8**, 642–6.

Erdös, P., Feller, W., and Pollard, H., (1949). A theorem on power series. *Bull. Amer. Math. Soc.*, **55**, 201–4.

Fan, K., (1958). Topological proofs for certain theorems on matrices with non-negative elements. *Monatsh. Math.* **62**, 219–37.
Feldstein, M. S. and Rothschild, M., (1974). Towards an economic theory of replacement investment. *Econometrica*, **42**, 393–423.

Feller, W., (1968). *An Introduction to Probability Theory and its Applications*, vol. 1 (3rd edition). Wiley, New York.
Fiedler, M. and Pták, V., (1962). On matrices with non-positive off-diagonal elements and positive principal minors. *Czechoslovak Math. J.*, **12**, 382–400.
Fienberg, S. E., (1970). An iterative procedure for estimation in contingency tables. *Ann. Math. Statist.*, **41**, 907–17.
Fortet, R., (1938). Sur l'itération des substitutions algébriques linéaires à une infinité de variables et ses applications à la théorie des probabilites en chaîne. *Revista de Ciências (Lima)*, **40**, 185–261, 337–447, 481–528.
Fréchet, M., (1938). *Recherches théoriques modernes sur le calcul des probabilités. Second livre. Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles.* (Reprinted in 1952 with a new supplement and a note of Paul Lévy.) Gauthier-Villars, Paris.
— (1960). Sur les tableaux dont les marges et des bornes sont données. *Review of the International Statistical Institute*, **28**, 10–32.
Fritz, F.-J., Huppert, B., and Willems, W., (1979). *Stochastische Matrizen*. Springer, Berlin.
Frobenius, G., (1908). Über Matrizen aus positiven Elementen. *S.-B. Preuss. Akad. Wiss.* (Berlin), 471–6; (1909), 514–8.
— (1912). Über Matrizen aus nicht negativen Elementen. *S.-B. Preuss. Akad. Wiss.* (Berlin), 456–77.
Furstenberg, H. and Kesten, H., (1960). Products of random matrices. *Ann. Math. Statist.*, **31**, 457–69.

Gale, D., (1960). *The Theory of Linear Economic Models*. McGraw-Hill, New York.
Gantmacher, F. R., (1959). *Applications of the Theory of Matrices* (translated by J. L. Brenner). Interscience, New York.
Georgescu-Roegen, N., (1951). Some properties of a generalized Leontief model. Chapter 10 in *Activity Analysis of Production and Allocation*, ed. T. C. Koopmans. Wiley, New York.
Golub, G. and Seneta, E., (1973). Computation of the stationary distribution of an infinite Markov matrix. *Bull. Austral. Math. Soc.*, **8**, 333–41.
— (1974). Computation of the stationary distribution of an infinite stochastic matrix of special form. *Bull. Austral. Math. Soc.*, **10**, 255–61.
Golubitsky, M., Keeler, E. B., and Rothschild, M., (1975). Convergence of the age structure: applications of the projective metric. *Theoretical Population Biology*, **7**, 84–93; Addendum: *ibid.*, **10**, (1976), 413.
Gordon, P., (1965). *Théorie des Chaînes de Markov Finies et ses Applications*. Dunod, Paris.
Gorman, W. M., (1963). Estimating trends in Leontief matrices: a note on Mr Bacharach's paper. Duplicated note, Nuffield College, Oxford.
de Groot, M. H., (1974). Reaching a consensus. *J. Amer. Statist. Assoc.*, **69**, 118–21.

Hadamard, J., (1928). Sur le battage des cartes et ses relations avec la Mécanique Statistique. *C. R. Congrès Int. des Mathématiciens de Bologne*, **5**, 133–9.
Hajnal, J., (1956). The ergodic properties of non-homogeneous finite Markov chains. *Proc. Cambridge Phil. Soc.*, **52**, 67–77.
— (1958). Weak ergodicity in non-homogeneous Markov chains. *Proc. Cambridge Phil. Soc.*, **54**, 233–46.
— (1976). On products of non-negative matrices. *Math. Proc. Cambridge Phil. Soc.*, **79**, 521–30.
Harary, F., (1970). A matrix approach to nonstationary chains. *Operations Res.*, **18**, 1168–81.

Harris, T. E., (1957). Transient Markov chains with stationary measures. *Proc. Amer. Math. Soc.*, **8**, 937–42.
Hawkins, D. and Simon, H. A., (1949). Note: Some conditions of macroeconomic stability. *Econometrica*, **17**, 245–8.
Haynsworth, E. V., (1955). Quasi-stochastic matrices. *Duke Math. J.*, **22**, 15–24.
Heap, B. R. and Lynn, M. S., (1964). The index of primitivity of a non-negative matrix. *Numerische Math.*, **6**, 120–41.
— (1966). The structure of powers of non-negative matrices. *J. SIAM Appl. Math.*, **14**, I: The index of convergence, 610–39; II: The index of maximum density, 762–77.
Heathcote, C. R., (1971). *Probability: Elements of the Mathematical Theory*. George Allen and Unwin, London.
Herstein, I. N., (1954). A note on primitive matrices. *Amer. Math. Monthly*, **61**, 18–20.
Hille, E. and Phillips, R. S., (1957). *Functional Analysis and Semi-Groups* (revised edition). American Math. Soc. Colloquium Publications, **31**, Providence, R. I.
Holladay, J. C. and Varga, R. S., (1958). On powers of nonnegative matrices. *Proc. Amer. Math. Soc.*, **9**, 631–4.
Holmes, P. T., (1966). On some spectral properties of a transition matrix. *Sankhyā*, Ser A **28**, 205–14; Corrigendum **29**, (1967), 106.
Hopf, E., (1963). An inequality for positive integral linear operators. *J. Math. Mech.*, **12**, 683–92.
Hostinsky, B., (1931). *Méthodes générales du Calcul des Probabilités*. Mém. Sc. Math. no. 51. Gauthier-Villars, Paris.
— (1950). Novější práce o Markovových řetězech a příbuzných úlohách. *Čas. Pěst. Mat. Fys.*, **74**, 48–62.
Householder, A. S., (1958). The approximate solution of matrix problems. *J. Assoc. Comput. Mach.*, **5**, 204–43.
Howard, R. A., (1960). *Dynamic Programming and Markov Processes.* M.I.T. Press, Cambridge, Mass.

Isaacson, D. and Madsen, R., (1974). Positive columns for stochastic matrices. *J. Appl. Prob.*, **11**, 829–35.
— (1976). *Markov Chains.* Wiley, New York.
Iosifescu, M., (1972). On two recent papers on ergodicity in non-homogeneous Markov chains. *Ann. Math. Statist.*, **43**, 1732–6.
— (1977). *Lanţuri Markov Finite şi Aplicaţii.* ed. Tehnică, Bucharest.

Joffe, A. and Spitzer, F., (1966). On multitype branching processes with $\rho \leq 1$. *J. Math. Anal. Applicns.*, **19**, 409–30.

Karlin, S., (1959). *Mathematical Methods and Theory in Games, Programming, and Economics.* Addison-Wesley, Reading, Mass.
— (1966). *A First Course in Stochastic Processes.* Academic Press, New York.
Karpelevich, F. I., (1949). On the characteristic roots of matrices with non-negative coefficients [in Russian]. *Uspehi Mat. Nauk*, **4**, 177–8.
— (1951). On the characteristic roots of matrices with non-negative elements [in Russian]. *Izvestia A.N. S.S.S.R.* Ser. mat., **15**, 361–83.
Kaucky, J., (1930). Quelques remarques sur les chaînes de Markoff [in Czech, French summary]. *Publ. Fac. Sci. Univ. Masaryk* (Brno), no. **131**.
Keilson, J. H. and Styan, G. P. H., (1973). Markov chains and M-matrices: inequalities and equalities. *J. Math. Anal. Applicns.*, **41**, 439–59.
Kemeny, J. G., (1966). Representation theory for denumerable Markov chains. *Trans. Amer. Math. Soc.*, **125**, 47–62.

Kemeny, J. G. and Snell, L. J., (1960). *Finite Markov Chains*. Van Nostrand, Princeton, N.J.

Kemeny, J. G., Snell, L. J. and Knapp, A. W., (1966). *Denumerable Markov Chains*. Van Nostrand, Princeton, N.J.

Kendall, D. G., (1959). Unitary dilations of Markov transition operators, and the corresponding integral representations for transition-probability matrices. In *Probability and Statistics* (Cramér Memorial Volume), ed. U. Grenander. Almqvist and Wiksell, Stockholm and New York, pp. 139–61.

— (1960). Geometric ergodicity and the theory of queues. In *Mathematical Methods in the Social Sciences*, ed. K. J. Arrow, S. Karlin, P. Suppes. Stanford.

— (1966). Contribution to discussion in "Quasi-stationary distributions and time-reversion in genetics" by E. Seneta [with discussion]. *J. Roy. Statist. Soc.*, Ser B **28**, 253–77.

Kennedy, D. P., (1978). On sets of countable non-negative matrices and Markov decision processes. *Adv. Appl. Prob.*, **10**, 633–46.

Khintchine, A. Y., (1969). *Mathematical Methods in the Theory of Queueing* (2nd edition) (translation D. M. Andrews and M. H. Quenouille, with additional notes by E. Wolman). Griffin, London.

Kingman, J. F. C., (1961). A convexity property of positive matrices. *Quart. J. Math., Oxford* (2), **12**, 283–4.

— (1963). The exponential decay of Markov transition probabilities. *Proc. London Math. Soc.*, **13**, 337–58.

— (1975). Geometrical aspects of the theory of non-homogeneous Markov chains. *Proc. Cambridge Phil. Soc.*, **77**, 171–83.

Klimko, L. A. and Sucheston, L., (1968). On probabilistic limit theorems for a class of positive matrices. *J. Math. Anal. Applicns.*, **24**, 191–201.

Kolmogorov, A. N., (1931). Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, **104**, 415–58.

— (1936a). Anfangsgründe der Theorie der Markoffschen Ketten mit unendlichen vielen möglichen Zuständen. *Mat. Sbornik* (N.S.), **1**, 607–10; *Bull. Univ. Moscou* [in Russian], **1** (1937), no. 3, 1–16.

— (1936b). Zur Theorie der Markoffschen Ketten. *Math. Ann.*, **112**, 155–60.

Konečný, M., (1931). Sur la théorie des chaînes de Markoff. *Publ. Fac. Sci. Univ. Masaryk* (Brno), no. **147**.

Kotelyanskii, D. M., (1952). Some properties of matrices with positive elements [in Russian]. *Mat. Sbornik*, **31**, 497–506.

Kozniewska, I., (1962). Ergodicité et stationnarité des chaînes de Markoff variables à un nombre fini d'états possibles. *Colloq. Math.*, **9**, 333–46.

Kruithof, J., (1937). Calculation of telephone traffic. (Translation 2663, Post Office Research Department Library, London.)

Kuich, W., (1970a). On the entropy of context-free languages. *Information and Control*, **16**, 173–200.

— (1970b). On the convergence parameter of infinite non-negative matrices. *Monatsh. Math.*, **74**, 138–44.

Kullback, S., (1959). *Information Theory and Statistics*. Wiley, New York.

Lappo-Danilevskii, J. A., (1934). Mémoires sur la théorie des systèmes des équations différentielles linéaires. *Trav. Inst. Phys.—Math. Stekloff. Acad. Sci. U.R.S.S.*, **1**. [All 3 volumes of Lappo-Danilevskii's work on the same topic, of which this is the first, were reprinted by Chelsea, N.Y.]

Larisse, J., (1966). Sur une convergence presque sûre d'un produit infini de matrices stochastiques ayant le même nombre de classes ergodiques. *C.R. Acad. Sci. Paris*, **262**, 913–5.

Larisse, J. and Schützenberger, M. P., (1964). Sur certaînes chaînes de Markov non

homogènes. *Publ. Inst. Statist. Univ. Paris*, **8**, 57–66.
Ledermann, W., (1950a). On the asymptotic probability distribution for certain Markoff processes. *Proc. Cambridge Phil. Soc.*, **46**, 581–94.
— (1950b). Bounds for the greatest latent root of a positive matrix. *J. London Math. Soc.*, **25**, 265–8.
Leslie, P. H., (1945). On the use of matrices in certain population mathematics. *Biometrika*, **33**, 183–212.
Lewis, E. G., (1942). On the generation and growth of a population. *Sankhyā*, **6**, 93–6.
Lopez, A., (1961). *Problems in Stable Population Theory*. Office of Population Research, Princeton University, Princeton.
Lynn, M. S. and Timlake, W. P., (1969). Bounds for Perron eigenvectors and subdominant eigenvalues of positive matrices. *Linear Algebra and Applicns.*, **2**, 143–52.

McFarland, D. D., (1969). On the theory of stable populations: a new and elementary proof of the theorems under weaker assumptions. *Demography*, **6**, 301–22.
Macgill, S. M., (1977). Theoretical properties of biproportional matrix adjustments. *Environment and Planning* A, **9**, 687–701.
Malécot, G., (1944). Sur un problème de probabilités en chaîne que pose la génétique. *C.R. Acad. Sci. Paris*, **219**, 379–81.
Mandl, P., (1960). On the asymptotic behaviour of probabilities within groups of states of a homogeneous Markov process [in Russian]. *Čas. Pěst. Mat.*, **85**, 448–56.
— (1967). An iterative method for maximizing the characteristic root of positive matrices. *Rev. Roum. Math. Pures et Appl.*, **12**, 1317–22.
Mandl, P. and Seneta, E., (1969). The theory of non-negative matrices in a dynamic programming problem. *Austral. J. Statist.*, **11**, 85–96.
Mangasarian, O. L., (1971). Perron–Frobenius properties of $Ax - \lambda Bx$. *J. Math. Anal. Applicns.*, **36**, 86–102.
Marcus, M. and Minc, H. (1964). *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Rockleigh, N.J.
— (1967). On a conjecture of B. L. van der Waerden. *Proc. Cambridge Phil. Soc.*, **63**, 305–9.
Markov, A. A., (1906). Extension of the law of large numbers to dependent quantities [in Russian]. *Izv. Fiz.-Matem. Obsch. Kazan Univ.*, (2nd ser.) **15**, 135–56. (Also in Markov (1951), pp. 339–61.)
— (1907). Investigation of an important case of dependent trials [in Russian]. *Izvestia Akad. Nauk S.P.-B.* (*6*), **1**, 61–80.
— (1924). *Ischislenie Veroiatnostei* (4th edition) (esp. pp. 552–81). Gosizdat, Moscow.
— (1951). *Izbrannie Trudy*. A.N. S.S.S.R., Leningrad.
Maybee, J. S. (1967). Matrices of class J_2. *J. Res. Nat. Bur. Standards*, **71B**, 215–24.
Maybee, J. S. and Quirk, J., (1969). Qualitative problems in matrix theory. *SIAM Review* **11**, 30–51.
Medlin, G. W., (1953). On bounds for the greatest characteristics root of a matrix with positive elements. *Proc. Amer. Math. Soc.*, **4**, 769–71.
Menon, M. V., (1967). Reduction of a matrix with positive elements to a doubly stochastic matrix. *Proc. Amer. Math. Soc.*, **18**, 244–7.
— (1968). Matrix links, an extremization problem, and the reduction of a non-negative matrix to one with prescribed row and column sums. *Canad. J. Math.*, **20**, 225–32.
Metzler, L. A., (1945). A multiple-county theory of income transfers. *J. Political Economy*, **59**, 14–29.

— (1951). Taxes and subsidies in Leontief's input-output model. *Quart. J. Economics*, **65**, 433–8.
Meyer, C. D., (1975). The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Review*, **17**, 443–64.
— (1978). An alternative expression for the mean first passage matrix. *Linear Algebra and Applicns.*, **22**, 41–7.
Miller, H. D., (1961). A convexity property in the theory of random variables defined on a finite Markov chain. *Ann. Math. Statist.*, **32**, 1260–70.
Minc, H., (1970). On the maximal eigenvector of a positive matrix. *SIAM J. Numer. Anal.*, **7**, 424–7.
— (1974). The structure of irreducible matrices. *Linear and Multilinear Algebra*, **2**, 85–90.
Minkowski, H., (1900). Zur Theorie der Einheiten in den algebraischen Zahlkorpern. *Gesammelte Abhandlungen*, vol. 1 (1911), 316–9.
Mirsky, L., (1963). Results and problems in the theory of doubly-stochastic matrices. *Z. Wahrscheinlichkeitstheorie*, **1**, 319–34.
— (1964). Inequalities and existence theorems in the theory of matrices. *J. Math. Anal. Applicns.*, **9**, 99–118.
von Mises, R., (1931). *Wahrscheinlichkeitsrechnung*. Fr. Deuticke, Leipzig and Vienna.
Moran, P. A. P., (1962). Polysomic inheritance and the theory of shuffling. *Sankhyā*, Ser A, **24**, 63–72.
— (1976). Global stability of genetic systems governed by mutation and selection. *Math. Proc. Cambridge Phil. Soc.*, **80**, 331–6.
Morgenstern, O. (ed.), (1954). *Economic Activity Analysis*. Wiley, New York.
Morishima, M., (1952). On the laws of change of the price system in an economy which contains complementary commodities. *Osaka Economic Papers*, **1**, 101–13.
— (1961). Generalizations of the Frobenius-Wielandt theorems for non-negative square matrices. *J. London Math. Soc.*, **36**, 211–20.
— (1964). *Equilibrium, Stability, and Growth*. Clarendon Press, Oxford.
Mott, J. L., (1957). Conditions for the ergodicity of non-homogeneous finite Markov chains. *Proc. Roy. Soc. Edinburgh*, section A **64**, 369–80.
Mott, J. L. and Schneider, H., (1957). Matrix norms applied to weakly ergodic Markov chains. *Archiv der Mathematik*, **8**, 331–3.
Moy, S.-T. C., (1965). λ-continuous Markov chains. *Trans. Amer. Math. Soc.*, **117**, 68–91; Part II *ibid.* **120**, 83–107.
— (1967a). Ergodic properties of expectation matrices of a branching process with countably many types. *J. Math. Mech.*, **16**, 1207–26.
— (1967b). Extension of a limit theorem of Everett, Ulam and Harris on multitype branching processes to a branching process with countably many types. *Ann. Math. Statist.*, **38**, 992–9.
Mukherjea, A., (1979). Limit theorems: stochastic matrices, ergodic Markov chains, and measures on semigroups. In *Probabilistic Analysis and Related Topics*, Vol. 2, Academic Press, New York. pp. 143–203.
Mukherjea, A. and Nakassis, A., (1977). On the limit of the convolution iterates of a probability measure on $(n \times n)$ stochastic matrices. *J. Math. Anal. Applicns.*, **60**, 392–7.

Nelson, E., (1958). The adjoint Markoff process. *Duke Math. J.*, **25**, 671–90.
Neveu, J., (1964). Chaînes de Markov et théorie du potentiel. *Ann. Fac. Sci. Univ. Clermont—Ferrand.*, **24**, 3 (Math.), 37–89.

Oldenburger, R., (1940). Infinite powers of matrices and characteristic roots. *Duke Math. J.*, **6**, 357–61.

de Oliveira, G. N., (1968). *Sobre Matrizes Estócasticas e Duplamente Estócasticas.* Coimbra. (Also as: *Rev. Fac. Ci. Univ. Coimbra*, **41**, 15–221.)
— (1972). On the characteristic vectors of a matrix. *Linear Algebra and Applicns.*, **5**, 189–96.
Orey, S., (1962). An ergodic theorem for Markov chains. *Z. Wahrscheinlichkeitstheorie*, **1**, 174–6.
Ostenc, É., (1934). Sur le principe ergodique dans les chaînes de Markov à elements variables. *C.R. Acad. Sci. Paris*, **199**, 175–6.
Ostrowski, A. M., (1937). Über die Determinanten mit überwiegender Hauptdiagonale. *Comment. Math. Helv.*, **10**, 69–96.
— (1952). Bounds for the greatest latent root of a positive matrix. *J. London Math. Soc.*, **27**, 253–6.
— (1956). Determinanten mit überwiegender Hauptdiagonale und die absolut Konvergenz von linearen Iterationsprozessen. *Comment. Math. Helv.*, **30**, 175–210.
— (1963). On positive matrices. *Math. Ann.*, **150**, 276–84.
— (1964). Positive matrices and functional analysis. In: *Recent Advances in Matrix Theory*, ed. H. Schneider. University of Wisconsin Press, pp. 81–101.
Ostrowski, A. M. and Schneider, H. (1960). Bounds for the maximal characteristic root of a non-negative irreducible matrix. *Duke Math. J.*, **27**, 547–53.

Paige, C. C., Styan, G. P. H., and Wachter, P. G., (1975). Computation of the stationary distribution of a Markov chain. *J. Statist. Comput. Simul.*, **4**, 173–86.
Parthasarathy, K. R., (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.
Paz, A., (1963). Graph-theoretic and algebraic characterizations of some Markov processes. *Israel J. Math.*, **1**, 169–80.
— (1965). Definite and quasidefinite sets of stochastic matrices. *Proc. Amer. Math. Soc.*, **16**, 634–41.
— (1971). *Introduction to Probabilistic Automata.* Academic Press, New York.
Paz, A. and Reichaw, M., (1967). Ergodic theorems for sequences of infinite stochastic matrices. *Proc. Cambridge Phil. Soc.*, **63**, 777–84.
Perfect, H., (1952). On positive stochastic matrices with real characteristic roots. *Proc. Cambridge Phil. Soc.*, **48**, 271–6.
— (1953). Methods of constructing certain stochastic matrices. *Duke Math. J.*, **20**, 395–404; Part II, *ibid* **22** (1955), 305–11.
Perfect, H. and Mirsky, L., (1965). Spectral properties of doubly-stochastic matrices. *Monatsh. Math.* **69**, 35–57.
Perkins, P., (1961). A theorem on regular matrices. *Pacific J. Math.*, **11**, 1529–33.
Perron, O., (1907). Zur Theorie der Matrices. *Math. Ann.*, **64**, 248–63.
Pollard, J. H., (1973). *Mathematical Models for the Growth of Human Populations.* University Press, Cambridge.
Pruitt, W. E., (1964). Eigenvalues of non-negative matrices. *Ann. Math. Statist.*, **35**, 1797–800.
Pták, V., (1958). On a combinatorial theorem and its application to non-negative matrices [in Russian and English]. *Czechoslovak Math. J.*, **8**, 487–95.
Pták, V. and Sedláček, J., (1958). On the index of imprimitivity of non-negative matrices [in Russian and English]. *Czechoslovak Math. J.*, **8**, 496–501.
Pullman, N. J., (1964). On the number of positive entries in the powers of a non-negative matrix. *Canadian Math. Bull.*, **7**, 525–37.
— (1966). Infinite products of substochastic matrices. *Pacific J. Math.*, **16**, 537–44.
— (1971). A geometric approach to the theory of nonnegative matrices. *Linear Algebra and Applicns.*, **4**, 297–312.
— (1975). A note on a theorem of Minc on irreducible non-negative matrices. *Linear and Multilinear Algebra*, **2**, 335–6.

Pullman, N. and Styan, G. P. H., (1973). The convergence of Markov chains with nonstationary transition probabilities and constant causative matrix. *Stoch. Proc. Appl.*, **1**, 279–85.
Putnam, C. R., (1958). On bounded matrices with non-negative elements. *Canad. J. Math.*, **10**, 587–91.
— (1961). A note on non-negative matrices. *Canad. J. Math.*, **13**, 59–62. (A correction to the 1958 paper.)
Pykh, Yu. A., (1973). The characteristic numbers of stochastic matrices [in Russian]. *Doklady A.N. S.S.S.R.* **211**, 1299–1301.

Quine, M. P., (1972). The multitype Galton-Watson process with ρ near 1. *Adv. Appl. Prob.*, **4**, 429–52.

Rheinboldt, W. C. and Vandergraft, J. S., (1973). A simple approach to the Perron–Frobenius theory for positive operators on general partially-ordered finite-dimensional linear spaces. *Mathematics of Computation*, **27**, 139–45.
Robert, F., (1973). *Matrices Non-negatives et Normes Vectorielles* (cours de DEA) (mimeographed lecture notes).
Romanovsky, V. I., (1936). Recherches sur les chaînes de Markoff. *Acta Math.*, **66**, 147–251.
— (1949). *Diskretnii Tsepi Markova.* G.I.T.-T.L. Moscow-Leningrad. (Translated in 1970 as *Discrete Markov Chains*, by E. Seneta. Wolters-Noordhoff, Groningen.)
Rosenblatt, D., (1957). On the graphs and asymptotic forms of finite Boolean relation matrices and stochastic matrices. *Naval Res. Logist. Quart.*, **4**, 151–67.
Rothblum, U. G., (1975). Algebraic eigenspaces of nonnegative matrices. *Linear Algebra and Applicns.*, **12**, 281–92.

Samelson, H., (1957). On the Perron–Frobenius theorem. *Michigan Math. J.*, **4**, 57–9.
Sarymsakov, T. A., (1945a). Sur les chaînes de Markoff à une infinité dénombrable d'états possibles. *C.R. (Doklady) Acad. Sci. U.R.S.S.*, **47**, 617–9.
— (1945b). Sur une synthèse des deux méthodes d'exposer la théorie des chaînes discrètes de Markoff. *C.R. (Doklady) Acad. Sci. U.R.S.S.*, **48**, 159–61.
— (1953a). On the ergodic principle for inhomogeneous Markov chains [in Russian]. *Doklady A.N. S.S.S.R.*, **90**, 25–8.
— (1953b). On the matrix method of investigation of Markov chains with a countable set of possible states [in Russian]. *Sredneaziatskii Gos. Univ.: Trudy*, Novaia Seria, vyp. **46**, Fiz.-mat. nauki, kniga 9, pp. 1–46. Izd. S.-A. G.U., Tashkent.
— (1954). *Osnovi Teorii Processov Markova.* G.I.T.-T.L., Moscow.
— (1956). On the theory of inhomogeneous Markov chains [in Russian]. *Doklady A.N. Uzbek. S.S.R.*, **8**, 3–7.
— (1958). On inhomogeneous Markov chains [in Russian]. *Doklady A.N. S.S.S.R.*, **120**, 465–7.
— (1961). Inhomogeneous Markov chains [in Russian]. *Teor. Veroiat. Primenen.*, **6**, 194–201.
Sarymsakov, T. A. and Mustafin, H. A., (1957). On an ergodic theorem for inhomogeneous Markov chains [in Russian]. *Sredneaziatskii Gos. Univ. im V.I. Lenina: Trudy*, Novaia Seria, vyp. **74**, Fiz-mat. nauki, kniga 15, pp. 1–38. Izd. A. N. Uzbekskoi S.S.R., Tashkent.
Schneider, H. (1958). Note on the fundamental theorem of irreducible non-negative matrices. *Proc. Edinburgh Math. Soc.*, **2**, 127–30.
— (1977). The concepts of irreducibility and full indecomposability of a matrix in the works of Frobenius, König and Markov. *Linear Algebra and Applicns.*, **18**, 139–62.
Schwartz, Š., (1967). Some estimates in the theory of non-negative matrices. *Czechoslovak Math. J.*, **17**, 399–407.

Seneta, E., (1967a). Finite approximations to infinite non-negative matrices. *Proc. Cambridge Phil. Soc.*, **63**, 983–992; Part II: refinements and applications. *ibid.* **64**, (1968a) 465–470.
— (1967b). On the maxima of absorbing Markov chains. *Austral. J. Statist.*, **9**, 93–102.
— (1968b). The principle of truncations in applied probability. *Comm. Math. Univ. Carolinae*, **9**, 533–9.
— (1973a). On strong ergodicity of inhomogeneous products of finite stochastic matrices. *Studia Math.*, **46**, 241–7.
— (1973b). On the historical development of the theory of finite inhomogeneous Markov chains. *Proc. Cambridge Phil. Soc.*, **74**, 507–13.
— (1973c). *Non-Negative Matrices.* Allen and Unwin, London.
— (1974). Review #11167. *Math. Rev.*, **48**, No. 6, p. 1296.
— (1976). The condition of truncations of a stochastic matrix. *Communications in Statistics, Simulation and Computation.* **B5**, 144–5.
— (1979). Coefficients of ergodicity: structure and applications. *Adv. Appl. Prob.*, **11**, 576–90.
— (1980). Computing the stationary distribution for infinite Markov chains. *Linear Algebra and Applicns.*, **34**, 259–67.
Seneta, E. and Sheridan, S., (1981). Strong ergodicity of non-negative matrix products. *Letters in Linear Algebra.* In *Linear Algebra and Applicns.*, (to appear).
Seneta, E. and Vere-Jones, D., (1966). On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *J. Appl. Prob.*, **3**, 403–34.
Sevastyanov, B. A., (1971). *Vetviaschiesia Processi.* Nauka, Moscow.
Sheridan, S., (1979). *Inhomogeneous Products of Non-Negative Matrices.* Honours Year Project, Department of Statistics, School of General Studies, Australian National University, Canberra.
Šidák, Z., (1962). Eigenvalues of operators in denumerable Markov chains. *Trans. 3rd Prague Conf. on Inf. Theory Stat. Decision Fns., Random Processes.* (Published in 1964 by Publishing House of the Czechoslovak Acad. Sciences, Prague, pp. 641–56.)
— (1963). Němkteré věty a příklady z teorie operátorů ve spočetných Markovových řetezcích. *Čas. Pěst. Mat.*, **88**, 457–78.
— (1964a). Eigenvalues of operators in l_p-spaces in denumerable Markov chains. *Czechoslovak Math. J.*, **14**, 438–43.
— (1964b). O počtu kladných prvků v mocninách nezáporné matice. *Čas Pěst. Mat.*, **89**, 28–30.
Sierksma, G., (1979). Non-negative matrices; the open Leontief model. *Linear Algebra and Applicns.*, **26**, 175–201. (There is a sequel, arising out of an error: T. T. Read and G. Sierksma: Some remarks on the paper "Non-negative matrices; the open Leontief model", Report WS-61/7901, Econometric Institute, University of Groningen.)
Sinkhorn, R., (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, **35**, 876–9.
— (1967). Diagonal equivalence to matrices with prescribed row and column sums. *Amer. Math. Monthly*, **74**, 402–5; II *Proc. Amer. Math. Soc.*, **45**, 195–8.
Sinkhorn, R. and Knopp, P., (1967). Concerning non-negative matrices and doubly stochastic matrices. *Pacific J. Math.*, **21**, 343–8.
Sirazhdinov, S. H., (1950). The ergodic principle for non-homogeneous Markov chains [in Russian]. *Doklady A.N. S.S.S.R.*, **71**, 829–30.
Solow, R., (1952). On the structure of linear models. *Econometrica*, **20**, 29–46.
Solow, R. and Samuelson, P. A., (1953). Balanced growth under constant returns to scale. *Econometrica*, **21**, 412–24.

Stein, P. and Rosenberg, R. L., (1948). On the solution of linear simultaneous equations by iteration. *J. London Math. Soc.*, **23**, 111–8.
Suleĭmanova, H. R., (1949). Stochastic matrices with real characteristic numbers [in Russian]. *Doklady A.N. S.S.S.R.*, **66**, 343–5.
— (1953). On the characteristic numbers of stochastic matrices [in Russian]. *Uchenie Zap. Mosk. Ped. Inst.*, **71**, 167–77.

Takahashi, Y., (1973). On the effects of small deviations in the transition matrix of a finite Markov chain. *J. Operations Research Soc. of Japan*, **16**, 104–29.
Tambs-Lyche, R., (1928). Un théorème sur les déterminants. *Det. Kong. Vid. Selskab.*, Forh. **I**, nr. 41, 119–20.
Taussky, O., (1949). A recurring theorem on determinants. *Amer. Math. Monthly*, **56**, 672–6.
Taylor, G. C., (1978). Stability in recurrence systems whose recurrence relations possess a certain positivity property. *SIAM J. Appl. Math.*, **35**, 48–67.
Thionet, P., (1963). Sur certaines variantes des projections du tableau d'échanges interindustriels. *Bulletin of the International Statistical Institute* (Proceedings of the 34th Session, Ottawa) **40**, Book 1, 119–31.
Thomasian, A. J., (1963). A finite criterion for indecomposable channels. *Ann. Math. Statist.*, **34**, 337–8.
Timan, O. Z., (1972). On the boundaries of Brauer for characteristic numbers of indecomposable stochastic matrices [in Russian]. *Zh. Vychisl. Mat. i Mat. Fiz.*, **12**, 192–8.
Titchmarsh, E. C., (1939). *The Theory of Functions* (2nd edition). Oxford University Press, London.
Tweedie, R. L., (1971). Truncation procedures for non-negative matrices. *J. Appl. Prob.*, **8**, 311–20.
Tweedie, R. L., (1980). Perturbations of countable Markov chains and processes. *Ann. Inst. Statist. Math.*, **32**, 283–90.

Ullman, J. L., (1952). On a theorem of Frobenius. *Michigan Math. J.*, **1**, 189–93.

Varga, R. S., (1962). *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs, N.J.
Veech, W., (1963). The necessity of Harris' condition for the existence of a stationary measure. *Proc. Amer. Math. Soc.*, **14**, 856–60.
Vere-Jones, D., (1962). Geometric ergodicity in denumerable Markov chains. *Quart. J. Math. Oxford* (2), **13**, 7–28.
— (1963). On the spectra of some linear operators associated with queueing systems. *Z. Wahrscheinlichkeitstheorie*, **2**, 12–21.
— (1966). Note on a theorem of Kingman and a theorem Chung. *Ann. Math. Statist.*, **37**, 1844–6.
— (1967). Ergodic properties of non-negative matrices—I. *Pacific J. Math.*, **22**, 361–86; —II *ibid.* **26** (1968), 601–20.
— (1971). Finite bivariate distributions and semigroups of non-negative matrices. *Quart. J. Math. Oxford* (2), **22**, 247–70.
Vrba, A., (1973). On powers of non-negative matrices. *Čas. Pěst. Mat.*, **98**, 298–9.

van der Waerden, B. L., (1926). Aufgabe 45. *Jber. Deutsch. Math. Verein.*, **35**, 117.
Wielandt, H., (1950). Unzerlegbare, nicht negative Matrizen. *Math. Zeit.*, **52**, 642–8.
Wilson, A. G., (1970). *Entropy in Urban and Regional Modelling*. Pion, London.
Winkler, R. L., (1968). The consensus of subjective probability distributions. *Management Science*, **15**, B61–B75.
Wolfowitz, J., (1963). Products of indecomposable, aperiodic, stochastic matrices. *Proc. Amer. Math. Soc.*, **14**, 733–7.

Wong, Y. K., (1954). Inequalities for Minkowski-Leontief matrices. In: *Economic Activity Analysis* (ed. O. Morgenstern), pp. 201–81. Wiley, New York.

Yamamoto, T., (1967). On the extreme values of the roots of matrices. *J. Math. Soc. Japan*, **19**, 173–8.

Zenger, C., (1972). A comparison of some bounds for the non-trivial eigenvalues of stochastic matrices. *Numerische Math.*, **19**, 209–11.

Author Index

Subject Index

Additional Bibliography. Coefficients of Ergodicity.

The appearance of this book in 1981 helped generate considerable subsequent work on coefficients of ergodicity, the focus of Chapters 3 and 4 in particular. The following lists related literature which has come to the author's attention since then. It is based on material prepared to accompany a keynote presentation at the 14-th International Workshop on Matrices and Statistics. (Auckland, New Zealand, March 29 – April 1, 2005.)

Artzrouni, M. and Li, X. (1995) A note on the coefficient of ergodicity of a column- allowable non-negative matrix. *Linear Algebra and its Applications*, **214**, 93-101.

Choi, M.-D., Ruskai, M.B. and Seneta, E. (1994) Equivalence of certain entropy contraction coefficients. *Linear Algebra and its Applications*, **208/209**, 29-36.

Cohen, J.E., Iwasa, Y., Rautu, Gh., Ruskai, M.B., Seneta, E. and Zbaganu, C. (1992) Relative entropy under mappings by stochastic matrices. *Linear Algebra and its Applications*, **179**, 211-235.

Cohen, J.E. and Sellers, P.H. (1982) Sets of nonnegative matrices with positive inhomogeneous products. *Linear Algebra and its Applications*, **47**, 185-192.

Cohn, H. (1989) Products of stochastic matrices and applications. *Internat.J. Math. and Math. Sci.*, **12**, No. 2, 209-233.

Cohn, H. and Nerman, O. (1990) On products of nonnegative matrices. *Ann. Prob.*, **18**, No.4, 1806-1815.

Federgruen, A. (1981) The rate of convergence for backwards products of a convergent sequence of finite Markov matrices. *Stochastic Processes and their Applications*,**11** , 187-192.

Fleischer, I. and Joffe, A. (1995a) Ratio ergodicity for nonhomogeneous Markov chains in general state spaces. *Journal of Theoretical Probability*, **8**, 31-37.

---------------------------- (1995b) Preservation of ergodicity under perturbation. *Math. Zeitschr.*, **218**, 219-222.

Funderlic, R.E. and Meyer,C.D. (1986) Sensitivity of the stationary distribution vector for an ergodic Markov chain. *Linear Algebra and its Applications*,**76** , 1-17.

Hajnal, J. (1993) Shuffling with two matrices. *Contemporary Mathematics*, **149**, 271-187.

Hartfiel, D.J.(1981) A general theory of measures for non-negative matrices. *Linear Algebra and its Applications*,**35** , 21-35.

-------------- (1984) An ergodic result for the limit of a sequence of substochastic products $xA_1 \ldots A_k$. *Linear and Multilinear Algebra*, **15**, 193-205.

--------------- (1996) Similitudes and the l_1 norm. *Linear Algebra and its Applications*,**245** , 305-319.

--------------- (1998). *Markov Set-Chains*. Lecture Notes in Mathematics No.1695. Springer, Berlin.

--------------- (1999) Coefficients of ergodicity for imprimitive matrices. *Commun.Statist.-Stochastic Models*. **15**(1), 81-88.

-------------- (2000) Difference equations in a general setting. *Linear Algebra and its Applications*,**305** , 173-185.

-------------- (2002) *Nonhomogeneous Matrix Products*. World Scientific,Singapore.

Hartfiel, D.J. and Seneta,E. (1990) A note on semigroups of regular stochastic matrices. *Linear Algebra and its Applications,* **141**, 47-51.

------------------------------ (1994) On the theory of Markov set-chains. *Adv Appl. Prob.,* **26**, 947-964.

Hunter,J.J. (1986) Stationary distributions of perturbed Markov chains. *Linear Algebra and its Applications*,**82** , 201-214.

Ipsen, I. and Meyer,C.D. (1994) Uniform stability for Markov chains. *SIAM J. Matrix Anal. Appl.*, **15**, 1061-1074.

Iosifescu, M. (1980) *Finite Markov Processes and Their Applications.* Wiley, Chichester.

Kijima, M. (1997) *Markov Processes for Stochastic Modeling.* Chapman and Hall, London.

Kirkland, S. J.,and Neumann, M.(2000) Regular Markov chains for which the transition matrix has large exponent. *Linear Algebra and its Applications*,**316** , 45-65.

Kirkland, S. J., Neumann, M. and Shader, B.L. (1998a) Applications of Paz's inequality to perturbation bounds for Markov chains. *Linear Algebra and its Applications*,**268** , 183-196.

-- (1998b) On a bound on algebraic connectivity: the case of equality. *Czechoslovak Math.* J., **48(123),** 65-76.

Le Gland, F. and Mevel, L. (2000) Basic properties of the projective product with application to products of column-allowable nonnegative matrices. *Math. Control Systems,* **13**, 41-62.

Lešanovský, A.(1990) Coefficients of ergodicity generated by nonsymmetrical vector norms. *Czechoslovak Math.*J., **40(115),** 284-294.

Meyer, C.D.(1980) The condition of a finite Markov chain and perturbation bounds for the limiting probabilities. *SIAM J. Algebraic Discrete Methods*, **1**, 273-283.

-------------- (1994) Sensitivity of the stationary distribution of a Markov chain. *SIAM J. Matrix Anal. Appl.*, **15**, 715-728.

Paŭn, U. (2001) Ergodic theorems for finite Markov chains. *Math. Reports (Bucharest)*, **3 (53)**, No. 4, 383-390.

---------- (2002) A class of ergodicity coefficients and applications. *Math. Reports (Bucharest)*, **4 (54)**, No. 2, 225-232.

Rhodius, A. (1993) On explicit forms for ergodicity coefficients. *Linear Algebra and its Applications,***194** , 71-83.

-------------- (1997) On the maximum of ergodicity coefficients, the Dobrushin ergodicity coefficient, and products of stochastic matrices. *Linear Algebra and its Applications,***253** , 141-154.

Rothblum, U.G. and Tan, C.P. (1985) Upper bounds on the maximum modulus of subdominant eigenvalues of non-negative matrices. *Linear Algebra and its Applications,***66** , 45-86.

Schweitzer, P.J. (1986) Posterior bounds on the equilibrium distribution of a finite Markov chain. *Commun. Statist.-Stochastic Models*, **2**, 323-338.

Seneta, E. (1982) Entropy and martingales in Markov chain models. *J.Appl. Prob.*,**19A** (Moran Festschrift), 367-381.

----------- (1984a) On the limiting set of non-negative matrix products. *Statistics and Probability Letters*, **2**, 159-163.

------------ (1984b) Spectrum localization by ergodicity coefficients for stochastic matrices. *Linear and Multilinear Algebra,* **14,** 343-347.

------------ (1984c) Explicit forms for ergodicity coefficients and spectrum localization. *Linear Algebra and its Applications,***60** , 187-197.

------------- (1988a) Sensitivity to perturbation of the stationary distribution: some refinements. *Linear Algebra and its Applications,***108** , 121-126.

------------- (1988b) Perturbation of the stationary distribution measured by ergodicity coefficients. *Adv. Appl. Prob.*, **20**, 228-230.

-------------- (1991) Sensitivity analysis, ergodicity coefficients and rank-one updates for finite Markov chains. In: W.J. Stewart (ed.) *Numerical Solution of Markov Chains*. Marcel Dekker, New York, pp.121-129.

-------------- (1993a) Explicit forms for ergodicity coefficients of stochastic matrices. *Linear Algebra and its Applications,***191** , 245-252.

-------------- (1993b) Applications of ergodicity coefficients of homogeneous Markov chains. *Contemporary Mathematics*, **149**, 189-199.

-------------- (1993c) Sensitivity of finite Markov chains under perturbation. *Statistics and Probability Letters*, **17**, 163-168.

-------------- (1993d) Ergodicity for products of infinite stochastic matrices. *Journal of Theoretical Probability*, **6,** 345-352.

-------------- (1996) Markov and the birth of chain dependence theory. *International Statistical Review*, **64**, 255-263.

-------------- (1998) Complementation in stochastic matrices and the GTH algorithm. *SIAM J. Matrix Anal Appl.*, **19,** 556-563.

Seneta, E. and Sheridan, S. (1981) Strong ergodicity of non-negative matrix products. *Linear Algebra and its Applications*, **37**, 277-292.

Seneta, E.and Tan, C.P. (1984) The Euclidean and Frobenius ergodicity coefficients and spectrum localization. *Bull. Malaysia Math. Soc.*, **(7) 1**, 1-7.

Srozhiddinov, Kh. M. (1990) Strict estimates of the spectral radius of nonnegative indecomposable matrix.(Russian) *Izv. Akad. Nauk Tadzh. SSR, Otd. Fiz.-Mat.Khim. Geol. Nauk*, No.**4(118)**, 57-61.

Tan, C.P. (1982) A functional form for a particular coefficient of ergodicity. *J.Appl. Prob.*, **19,** 858-863.

----------- (1983) Coefficients of ergodicity with respect to vector norms. *J.Appl. Prob.*, **20,** 277-287.

----------- (1984a) Spectrum localization of an ergodic stochastic matrix. *Bulletin of the Institute of Mathematics Academia Sinica (Taiwan)*, **12,** No. 2, 147-151.

----------- (1984b) A bound problem in modelling of computer systems and queueing networks. *Mathematical Computer Performance and Reliability*. G. Iazeolla, P.J. Courtois and A. Hordijk (Eds.) Elsevier, North Holland, 303-311.

------------ (1986) Spectrum localization using Hölder norms. *Houston Journal of Mathematics*, **12**, No.3, 441-449.

------------ (1987) Lower bounds for tau coefficients and operator norms, using composite matrix norms. *Bull.Austral.Math. Soc.*, **35**, 49-57.

Veselý.P. (1992) Two contributions to the theory of coefficients of ergodicity. *Czechoslovak Math*.J., **42(117),** 73-88.

---------- (1994) Majorants of matrix norms and spectrum localization. *Czechoslovak Math*.J., **44(119),** 141-161.

Corrigenda.

p.61. In exercise **2.20.** 1964 should read 1964b. In exercise **2.22** $k(T) \le k(TS)$ should read $k(S) \le k(ST)$ (or $k(S) \le k(TS)$).

pp. 69-71. Several typos related to occurrence of *min* and omission of $\sum$.

p.77, mid-page. (2.27) should read (2.28).

p.82, line 7↓ . $= 1$ should be replaced by $= 0$.

p.91. Assertion in exercise **3.6** is not valid, and exercise **3.7** is affected.

p.110, line1↓ . "Theorem 3.9" should read "Theorem 3.11".

p.141. In statement of Corollary 1, α should be replaced by a.

p.166. At end of Proof of Lemma 5.3, replace ∞ by 1.

p.185, mid-page. Replace ***Pg*** by ***Ph***

p. 189, line 3↑ . Replace " itself is closed" by "itself is Borel"

p.220. In the expression just above *Hint:* replace $(d+c_1+c_2)$ by $(d+c_1+c_2)^2$

p.224. At the end of the statement of the **Corollary,** replace $\{1, 2, \dots, i-1, i+1, \dots n\}$ by $\{1, 2, \dots, i-1, i, i+1, \dots n\}$

Springer Series in Statistics

(continued from p. ii)

Huet/Bouvier/Poursat/Jolivet: Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples, 2nd edition.
Ibrahim/Chen/Sinha: Bayesian Survival Analysis.
Jolliffe: Principal Component Analysis, 2nd edition.
Knottnerus: Sample Survey Theory: Some Pythagorean Perspectives.
Kolen/Brennan: Test Equating: Methods and Practices.
Kotz/Johnson (Eds.): Breakthroughs in Statistics Volume I.
Kotz/Johnson (Eds.): Breakthroughs in Statistics Volume II.
Kotz/Johnson (Eds.): Breakthroughs in Statistics Volume III.
Küchler/Sørensen: Exponential Families of Stochastic Processes.
Kutoyants: Statistical Influence for Ergodic Diffusion Processes.
Lahiri: Resampling Methods for Dependent Data.
Le Cam: Asymptotic Methods in Statistical Decision Theory.
Le Cam/Yang: Asymptotics in Statistics: Some Basic Concepts, 2nd edition.
Liu: Monte Carlo Strategies in Scientific Computing.
Longford: Models for Uncertainty in Educational Testing.
Manski: Partial Identification of Probability Distributions.
Mielke/Berry: Permutation Methods: A Distance Function Approach.
Molenberghs/Verbeke: Models for Discrete Longitudinal Data.
Pan/Fang: Growth Curve Models and Statistical Diagnostics.
Parzen/Tanabe/Kitagawa: Selected Papers of Hirotugu Akaike.
Politis/Romano/Wolf: Subsampling.
Ramsay/Silverman: Applied Functional Data Analysis: Methods and Case Studies.
Ramsay/Silverman: Functional Data Analysis, 2nd edition.
Rao/Toutenburg: Linear Models: Least Squares and Alternatives.
Reinsel: Elements of Multivariate Time Series Analysis, 2nd edition.
Rosenbaum: Observational Studies, 2nd edition.
Rosenblatt: Gaussian and Non-Gaussian Linear Time Series and Random Fields.
Särndal/Swensson/Wretman: Model Assisted Survey Sampling.
Santner/Williams/Notz: The Design and Analysis of Computer Experiments.
Schervish: Theory of Statistics.
Seneta: Non-negative Matrices and Markov Chains, revised printing.
Shao/Tu: The Jackknife and Bootstrap.
Simonoff: Smoothing Methods in Statistics.
Singpurwalla and Wilson: Statistical Methods in Software Engineering: Reliability and Risk.
Small: The Statistical Theory of Shape.
Sprott: Statistical Inference in Science.
Stein: Interpolation of Spatial Data: Some Theory for Kriging.
Taniguchi/Kakizawa: Asymptotic Theory of Statistical Inference for Time Series.
Tanner: Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, 3rd edition.
van der Laan: Unified Methods for Censored Longitudinal Data and Causality.
van der Vaart/Wellner: Weak Convergence and Empirical Processes: With Applications to Statistics.
Verbeke/Molenberghs: Linear Mixed Models for Longitudinal Data.
Weerahandi: Exact Statistical Methods for Data Analysis.
West/Harrison: Bayesian Forecasting and Dynamic Models, 2nd edition.