AD-755 469

ON THE HISTORICAL DEVELOPMENT OF THE THEORY OF FINITE INHOMOGENEOUS MARKOV CHAINS

E. Seneta

Princeton University

Prepared for:

Office of Naval Research

January 1973

ON THE HISTORICAL DEVELOPMENT OF THE THEORY OF FINITE INHOMOGENEOUS MARKOV CHAINS

by

E. Seneta

Te 'inical Report 30, Series 2
Department of Statistics
PRINCETON UNIVERSITY
January 1973

16

1a

ON THE HISTORICAL DEVELOPMENT OF THE THEORY OF FINITE INHOMOGENEOUS MARKOV CHAINS

by

E. Seneta+

Resume

The main purpose of the note is to compare necessary and sufficient conditions for weak ergodicity of finite inhomogeneous Markov chains given by Doeblin (1937), and Hajnal (1958), the former paper being little known; and more generally to expand on the nature and consequences of Doeblin's approach as compared to Hajnal's in some detail. A consequence is some insight into the relation between various "coefficients of ergodicity".

Key Words and Phrases: finite inhomogeneous Markov chains; history of probability; coefficients of ergodicity; weak ergodicity; Markov matrices; scrambling matrices.

AMS (1970) subject classification. Primary: 60J05.

+ Permanent address: Australian National University, Canberra.
Partial support by the Office of Naval Research on contract N00014-67-A-0151-0017 awarded to the Department of Statistics, Princeton University, is gratefully acknowledged.

1. Introduction. In this note all matrices are of fixed size $n \times n$. Let $\{P_k\}$, $k \geq 1$ be a sequence of stochastic matrices (i.e. matrices with non-negative entries and unit row sums); and let $T_{r,k} = \{t_{i,j}^{(r,k)}\}$ be the stochastic matrix defined by

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

for $r \geq 0$, $k \geq 1$.

The sequence $\{P_k\}$ is said to be weakly ergodic (in the sense of Kolmogorov) if for all $i,j,s = 1, \ldots n$ and $r \geq 0$

$$(1.1) \qquad (t_{i,s}^{(r,k)} - t_{j,s}^{(r,k)}) \to 0$$

as $k \to \infty$.

The earliest sufficient condition (since it is in large measure due to Markov himself) for weak ergodicity, as presented in the textbook of Bernstein (1946), states that weak ergodicity obtains if

$$(1.2) \qquad \sum_{i=1}^{\infty} \lambda(P_i) = \infty$$

where, for a stochastic $P = \{p_{i,j}\}$,

$$(1.3) \qquad \lambda(P) = \max_j (\min_i p_{ij}) .$$

In the Russian literature this is known as Markov's theorem; the final assertion is a consequence of the inequality for all

$i,j = 1,\ldots,n$; $r \geq 0$ and $k \geq 1$

$$\sum_s |t_{i,s}^{(r,k)} - t_{j,s}^{(r,k)}| \leq 2 \prod_{s=r+1}^{r+k} (1 - \lambda(P_s))$$

i.e.

$$a(T_{r,k}) \leq \prod_{s=r+1}^{r+k} (1 - \lambda(P_s)) \tag{1.4}$$

where

$$a(P) = \frac{1}{2} \max_{i,j} \sum_s |p_{i,s} - p_{j,s}| \quad .$$

The reasoning leading to (1.4) has been substantially refined in more recent times to yield,

$$a(T_{r,k}) \leq \prod_{s=r+1}^{r+k} a(P_s) \quad , \tag{1.5}$$

(Dobrušin, 1956; Paz and Reichaw, 1967). This last inequality sharpens the well-known one of Hajnal (1958);

$$b(T_{r,k}) \leq \prod_{s=r+1}^{r+k} \{1 - \beta(P_s)\} \tag{1.6}$$

where

$$b(P) = \max_s \max_{i,j} |p_{i,s} - p_{j,s}| \quad , \quad \beta(P) = \min_{i,j} \sum_s \min(p_{i,s}, p_{j,s}) \tag{1.7}$$

since Paz (1970) and Iosifescu (1972) show that

(1.8) $$\beta(P) = 1 - a(P)$$

while (1.6) itself implies

$$b(P) \leq 1 - \beta(P) \ , = a(P) \ .$$

The first necessary and sufficient condition for weak ergodicity is often ascribed to Hajnal (1958, Theorem 3), and there is little doubt that he gave the first proof involving such a condition, although it is necessary to mention an analogous and simultaneous announcement of Sarymsakov (1958), given in a broader context. However, in a little-known summary paper, Doeblin (1937) announces a condition of a different kind which he asserts is necessary and sufficient; and promises publication of this, and other material announced in the paper in various periodicals. So far as the present author can determine, a further paper containing a proof of this particular result never appeared, possibly due to Doeblin's premature death in World War II. In actual fact, the truth of his assertion follows immediately from e.g. that of Hajnal (1958, p. 239), as we shall note in the sequel. It is nevertheless interesting to speculate on the manner in which Doeblin may have arrived at his result in relation to the knowledge available at the time and this is the main purpose of the present note. Such investigation provides some insight into the relation between various "coefficients of ergodicity" which are used in the study of such non-homogeneous situations. We confine ourselves to the case of finite state-space, since it appears to the present author from the more recent papers cited above, that there is some but no substantial, loss in so doing, as compared to either the countable or general state space situation, at least at the present time.

A secondary purpose of this note is to demonstrate that the development of the theory of inhomogeneous products of finite stochastic matrices as a whole, as put forward by Hajnal, can be achieved perhaps more simply by basing one's ideas on the approach of Doeblin. In particular we shall refer to another characterizati of weak ergodicity (following the necessary and sufficient condition given above) in Doeblin's paper, which coincides with Hajnal's Theorem 4; and compare the roles played by "scrambling matrices" and "Markov matrices" in the two theoretical approaches.

The reader interested in the more recent developments in the subject should consult the references cited; we mention that Doeblin's condition itself was motivated by the announcement of a sufficient condition (which it subsumes) of Ostenc (1934).

---

2. Coefficients of Ergodicity. We shall denote by the term coefficient of ergodicity any function $\mu(\cdot)$ continuous on the set of $(n \times n)$ stochastic matrices P when P is regarded as a point in Euclidean n^2-dimensional space, and satisfying $0 \leq \mu(P) \leq 1$. A coefficient of ergodicity shall be called proper if

$$(2.1) \qquad \mu(P) = 1 \text{ if and only if } P = \underset{\sim}{1}\,\underset{\sim}{v}'$$

for some probability vector $\underset{\sim}{v}$ (i.e. all rows of P are identical).

We shall be concerned with the situation where $1 - m(\cdot)$ is a proper coefficient of ergodicity, and $\mu(\cdot)$ a coefficient of ergodicity (not necessarily proper) such that

$$(2.2) \qquad m(P^{(1)}P^{(2)}\ldots.P^{(k)}) \leq C \prod_{i=1}^{k} (1 - \mu(P^{(i)}))$$

for every finite set of stochastic matrices $P^{(i)}$, $i = 1,\ldots.,k$ and every k, where C is a constant which may depend on $\mu(\cdot)$ and $m(\cdot)$ (but not on the nature of the finite set of P's chosen). We see from Section 1 that $\beta(\cdot)$ and $1 - b(\cdot)$ are both proper coefficients of ergodicity; and (1.4)-(1.6) are all manifestations of (2.2).

The following proposition is a consequence of these definitions. The proof is totally analogous to the short demonstration of Hajnal's Theorem 3, although Hajnal deals with specific coefficients, and is omitted. (The ideas of the proof occur elsewhere in the present note in any case.)

<u>Theorem</u> 1. <u>Suppose that we are given</u> m <u>and</u> μ <u>such that (2.2) is satisfied (for both parts of this theorem). A given sequence</u> $\{P_i\}$ <u>of stochastic matrices is weakly ergodic if there exists a strictly increasing subsequence</u> $\{i_j\}$, $j = 1,2,\ldots.$ <u>of the positive integers such that</u>

$$(2.3) \qquad \sum_{j=1}^{\infty} \mu(T_{i_j},\, i_{j+1} - i_j) = \infty \quad .$$

Conversely, if $\{P_i\}$ *is a weakly ergodic sequence, and* $\mu(\cdot)$ *of (2.2) is also proper, then (2.3) is satisfied for some strictly increasing subsequence* $\{i_j\}$ *of the positive integers*.

Corollary. *If both the* μ *and* $1-m$ *of (2.2) are proper, then (2.3) is both necessary and sufficient for weak ergodicity of a specific sequence* $\{P_i\}$ *of stochastic matrices*.

Thus a necessary and sufficient condition can be formulated in terms of any two *specific proper* coefficients of ergodicity for which (2.2) can be shown to hold. The difficulty occurs in demonstrating this last; the more difficult part of e.g. Hajnal's paper lies in demonstrating that (2.2) holds, which as can be seen from (1.6) is attained with

$$(2.4) \qquad C = 1 \;, \quad \mu(P) = \beta(P) \;, \quad 1-m(P) = b(P) \;.$$

The $\lambda(\cdot)$ defined by (1.3) is not a proper coefficient of ergodicity, and, while the sufficiency part of Theorem 1 gives Markov's theorem, λ cannot be used directly in formulating a necessary and sufficient condition.

Now, as Hajnal points out in a slightly different context, clearly, for every P

$$(2.5) \qquad \beta(P) \geq \alpha(P) \; (\geq \lambda(P))$$

where

$$(2.6) \qquad \alpha(P) = \sum_s (\min_i \; p_{i,s}) \;.$$

It is readily checked that $\alpha(P)$ is a proper coefficient of ergodicity, and in view of (2.5) and (1.6) may be used with the $m(P)$ of (2.4) in a specific instance of Theorem 1 to give

$$(2.7) \qquad \sum_{j=1}^{\infty} \alpha(T_{i_j, i_{j+1}-i_j}) = \infty$$

as a necessary and sufficient condition for weak ergodicity of a specific sequence $\{P_i\}$. This is Doeblin's assertion.

It is, however, possible to arrive at the assertion that (2.7) is sufficient for weak ergodicity directly from an application of Markov's theorem. (The necessity of the condition (2.3) in Theorem 1, as also for this particular case, hinges only on the fact that if weak ergodicity obtains, $\mu(T_{r,k}) \to 1$ for each $r \geq 0$ as $k \to \infty$). It appears not unlikely that this is the manner in which Doeblin proceeded. We formulate the "comparison" principle involved in general terms first.

Lemma 1. *Suppose that (2.2) is satisfied for some* m *and* μ (μ *not necessarily proper); and let* $\nu(\cdot)$ *be any coefficient of ergodicity (not necessarily proper). If for any sequence* $\{P^{(i)}\}$ *of stochastic matrices for which the left-hand side diverges*

$$(2.8) \qquad \sum_{i=1}^{\infty} \nu(P^{(i)}) = \infty \quad \Rightarrow \quad \sum_{i=1}^{\infty} \mu(P^{(i)}) = \infty ,$$

then for a particular sequence $\{P_i\}$ *the existence of a strictly increasing subsequence of the positive integers such* that

$$(2.9) \qquad \sum_{j=1}^{\infty} \nu(T_{i_j, i_{j+1} - i_j}) = \infty$$

is sufficient for the weak ergodicity of $\{P_i\}$.

Proof: Take $r \geq 0$ fixed but arbitrary, and consider k large in $T_{r,k}$. Let j_1 be such that i_{j_1} is the minimal number of the sequence $\{i_j\}$ to satisfy $i_j \geq r + 1$; and $i_{j(k)}$ the maximal number to satisfy $i_j < k + r$.

Then since

$$T_{r,k} = T_{r,i_{j_1}-r} \; T_{i_{j_1}, i_{j(k)}-i_{j_1}} \; T_{i_{j(k)}, k+r-i_{j(k)}}$$

$$= T_{r,i_{j_1}-r} \prod_{j=j_1}^{j=j(k)-1} T_{i_j, i_{j+1}-i_j} \; T_{i_{j(k)}, k+r-i_{j(k)}}$$

it follows from (2.2) that

$$m(T_{r,k}) \leq C(1-\mu(T_{r,i_{j_1}-r}))\{\prod_{j=j_1}^{j=j(k)-1} (1-\mu(T_{i_j, i_{j+1}-i_j}))\} \times$$

$$\times ((1-\mu(T_{i_{j(k)}, k+r-i_{j(k)}}))$$

$$\leq C \prod_{j=j_1}^{j=j(k)-1} (1-\mu(T_{i_j, i_{j+1}-i_j}))$$

and the right hand side diverges to zero as $k \to \infty$, in view of (2.8) and (2.9) .

Corollary. The coefficient of ergodicity $\alpha(\cdot)$ defined by (2.6) satisfies condition (2.8), with $\mu = \lambda$, and $m = a$.

Proof:

$$\sum_{i=1}^{\infty} \alpha(P^{(i)}) = \sum_{i=1}^{\infty} \sum_{s=1}^{n} (\min_r p_{r,s}^{(i)})$$

$$= \sum_{s=1}^{n} \sum_{i=1}^{\infty} (\min_r p_{r,s}^{(i)})$$

so that divergence of the left hand side implies

$$\sum_{i=1}^{\infty} (\min_r p_{r,s}^{(i)}) = \infty \quad \text{for some } s \,;$$

which in turn implies

$$\sum_{i=1}^{\infty} \max_s (\min_r p_{r,s}^{(i)}) = \sum_{i=1}^{\infty} \lambda(P^{(i)}) = \infty \,.$$

This corollary is merely a manifestation in part of the obviously close relation between $\lambda(P)$ and $\alpha(P)$; clearly $\lambda(P) > 0$ if and only if $\alpha(P) > 0$ (so a Markov matrix is equivalently defined by either requirement, as will be seen from its definition in §3.2).

---

3. Comparison Between the Two Approaches . In this section we shall focus attention on a brief, direct comparison of the proper coefficients of ergodicity $\alpha(\cdot)$ and $\beta(\cdot)$ with a view to demonstrating that, insofar as the theoretical matters pertaining to weak ergodicity touched on in Hajnal's paper

are concerned, either may be used with equal convenience.

3.1. Coincidence Probabilities.

If we consider each of two systems independently undergoing trials governed by an inhomogeneous Markov chain governed by the sequence $\{P_i\}$, then Doeblin asserts that no matter at which state, (corresponding to one of the integers $1,\ldots,n$) each of the systems begins, they will be in the same state at the same time on an infinite number of occasions with probability 1 , if and only if the sequence $\{P_i\}$ is weakly ergodic. The same proposition is stated and proved in Theorem 4 of Hajnal's paper.

The necessity of weak ergodicity in this proof is not related to coefficients of ergodicity; the proof of sufficiency, however leans heavily on the inequality

$$\sum_{s=1}^{n} p_{1,s}\, p_{2,s} \geq \beta^2(P)/n \tag{3.1}$$

for any P . Now

$$n \sum_{s=1}^{n} p_{1,s}\, p_{2,s} \geq n \sum_{s=1}^{n} (\min_i p_{i,s})^2, \geq \left(\sum_{s=1}^{n} (\min_i p_{i,s})\right)^2$$

from the Cauchy-Schwartz inequality, so that

$$\sum_{s=1}^{n} p_{1,s}\, p_{2,s} \geq \alpha^2(P)/n$$

and the remainder of Hajnal's proof of sufficiency holds in terms of $\alpha(\cdot)$.

3.2. Scrambling Matrices and Markov Matrices.

A stochastic matrix P is called scrambling if and only if $\beta(P) > 0$ where $\beta(\cdot)$ is defined in (2.4). A stochastic matrix P is called Markov if and only if $\lambda(P) > 0$ where $\lambda(P)$ is defined in (1.3). In his Lemma 1, Hajnal shows that the scrambling property is monotone and preserved in a product, whatever other stochastic matrices may follow a scrambling matrix, by showing that for any stochastic $P = \{p_{i,j}\}$, $Q = \{q_{i,j}\}$

$$\beta(P) \leq \beta(PQ) \quad .$$

The same is true of $\alpha(\cdot)$, for

$$\begin{aligned}\alpha(PQ) &= \sum_{j=1}^{n} \min_{i} \{\sum_{k=1}^{n} p_{i,k}\, q_{k,j}\} \\ &\geq \sum_{j=1}^{n} \sum_{k=1}^{n} (\min_{i} p_{i,k}) q_{k,j} \\ &= \sum_{k=1}^{n} (\min_{i} p_{i,k}) = \alpha(P)\end{aligned}$$

so that

$$\alpha(P) \leq \alpha(PQ) \; .$$

It is also true, more fundamentally, that analogously to Hajnal's Lemma 2, if either P or Q is a Markov matrix, then so is PQ (a Markov stochastic matrix, recall, is merely one with an entirely positive column).

There remains only one result, Hajnal's Theorem 1, which we have not touched on implicitly or explicitly, in his development of weak ergodicity theory. This theorem characterizes scrambling matrices in terms of regular matrices (a regular stochastic matrix is one having a single eigenvalue of modulus unity, counting repeated eigenvalues as distinct), so it is not possible to find an analogue in this framework for Markov matrices.

We mention, however, one more result, important in applications, where Markov matrices are just as convenient as scrambling matrices. Let G_1 be the class of $(n \times n)$ regular stochastic matrices, and let M be the class of $(n \times n)$ Markov matrices. Let t be the number of distinct types (with regard to location of positive elements, but not their actual values) of matrices in G_1. Finally, let $\{P_i\}$ be a sequence of stochastic matrices.

Theorem 2. If for each $r \geq 0$, $T_{r,k} \in G_1$ for all $k \geq 1$, then $T_{r,k} \in M$ for $k \geq t+1$.

This result is due to Sarymsakov and Mustafin (1957); although the reader may prefer the simpler approach of Wolfowitz (1963, Lemmas 3 and 4 where the word "scrambling" may be replaced by "Markov" without altering the proofs.)

The remarks of this section may serve to indicate -- and the theme is further expanded in the book of the present author (Seneta, 1973, Chapter 4) -- that, in spite of the fact that the notion of a scrambling stochastic matrix may regarded as the more fundamental, since a matrix may be scrambling but not Markov -- frequently the simpler, and much earlier notion of a Markov matrix will suffice.

A historical note on the concept of "scrambling matrix" itself (apart from the marginal reference to Dobrušin already cited): it appears to have been exploited by Sarymsakov (1956, 1958) as well as Hajnal (1958).

REFERENCES

Bernstein, S.N. (1946). Teoriya Veroiatnostei, Moscow-Leningrad.

Dobrušin, R.L. (1956). Central limit theorem for non-stationary Markov chains, I, II. Theor. Probability Appl., **1**, 65-80 and 329-383 (English translation).

Doeblin, W. (1937). Le cas discontinu des probabilités en chaîne. Publ. Fac. Sci. Univ. Masaryk (Brno). No. **236**.

Hajnal, J. (1958). Weak ergodicity in nonhomogeneous Markov chains. Proc. Cambridge Phil. Soc., **54**, 233-246.

Iosifescu, M. (1972). On two recent papers on ergodicity in nonhomogeneous Markov chains. Ann. Math. Statist., **43**, 1732-1736.

Ostenc, É. (1934). Sur le principe ergodique dans les chaînes de Markov à elements variables. C.R. Acad. Sci Paris, **199**, 175-176.

Paz, A. (1970). Ergodic theorems for infinite probabilistic tables. Ann. Math. Statist., **41**, 539-550.

Paz, A. and Reichaw, M. (1967). Ergodic theorems for sequences of infinite stochastic matrices. Proc. Cambridge Phil. Soc., **63**, 777-784.

Sarymsakov, T. A. (1956). On the theory of inhomogeneous Markov chains. [in Russian] Dokl. A.N. Uzbek. S.S.R. **8**, 3-7.

Sarymsakov, T.A. (1958). On inhomogeneous Markov chains. Dokl. A.N. S.S.S.R., 120, 465-467.

Sarymsakov, T.A. and Mustafin, H.A. (1957). On an ergodic theorem for inhomogeneous Markov chains. [in Russian] Sredneaziatskii Gos. Universitet 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, 1957.

Seneta, E. (1973). Non-Negative Matrices. Allen and Unwin, London, (in press).

Wolfowitz, J. (1963). Products of indecomposable, aperiodic, stochastic matrices. Proc. Amer. Math. Soc., 14, 733-737.