STATISTICAL RESEARCH REPORT No 6
Institute of Mathematics July 1977
University of Oslo

A CHARACTERIZATION OF WEAK ERGODICITY IN TERMS OF CONVERGENCE OF EXPERIMENTS

by

Bo Lindqvist

ABSTRACT

The relation between the ergodic coefficient and deficiency relative to the least informative and the most informative experiment is investigated. The results are applied to non-homogeneous Markov chains (NMC's). Our main result can be described as follows: Given an NMC, define the experiments $\mathcal{E}_n^{(j)}$ for $n \geq 1$ consisting in observing the (n+j)-th state of the chain, the j-th state being the unknown parameter. Then the chain is weakly ergodic if and only if for any j, $\mathcal{E}_n^{(j)}$ converges as $n \to \infty$ to the least informative experiment. It is finally shown that in the homogeneous case, the rate of convergence is always exponential.

Key words: ergodic coefficient, least informative experiment, totally informative experiment, deficiency, weak ergodicity, convergence of experiments.

1. Introduction and summary.

Lindqvist (1977) studies the experiment $\mathcal{E}_n$ obtained by observing the n-th state of a finite Markov chain in order to obtain information about the initial state. As a particular result it follows that $\mathcal{E}_n$ converges to the least informative experiment if and only if the Markov chain is ergodic. Here convergence means convergence with respect to the deficiency introduced by LeCam (1964). It is furthermore proved that the rate of convergence is exponential.

The present paper extends these results to the case of non-homogeneous Markov chains (NMC's) with arbitrary state spaces. Given an NMC, we shall consider the experiments $\mathcal{E}_n(j)$ for $n \geq 1$, $j \geq 0$ consisting in observing the (n+j)-th state of the chain, the j-th state being the unknown parameter. We prove that $\mathcal{E}_n(j)$ converges, for any j, to the least informative experiment if and only if the NMC is weakly ergodic. This corresponds well to the common interpretation of weak ergodicity as "loss of memory". It is finally proved that the rate of convergence is exponential for homogeneous chains.

Weak ergodicity of NMC's is studied by e.g. Paz (1970), Madsen (1971) and Iosifescu (1972), who make use of the ergodic coefficient introduced by Dobrusin (1956). In section 2 of this paper we study, using ideas and results from Torgersen (1976a), the relation between the ergodic coefficient and deficiencies. In section 3 these results are combined with the theory of weak ergodicity in order to derive our main results.

2. The ergodic coefficient and deficiencies.

Let μ be a signed measure on some measurable space $(\mathcal{X},\mathcal{A})$. By $\|\mu\|$ we shall mean the usual total variation norm, i.e.

$$\|\mu\| = \mu(A) - \mu(B)$$

where A and B are the positive and negative part, respectively, of the Hahn decompostion for μ.

Let $(\mathcal{X},\mathcal{A})$ and $(\mathcal{Y},\mathcal{B})$ be measurable spaces and let ρ be an $\mathcal{A}$-measurable measure on $\mathcal{B}$, i.e. ρ is a real function on $\mathcal{X}\times\mathcal{B}$ such that

(i) $\rho(x,\cdot)$ is a signed measure on $\mathcal{B}$ for any $x\in\mathcal{X}$

(ii) $\rho(\cdot,B)$ is an $\mathcal{A}$-measurable function for any $B\in\mathcal{B}$.

We shall let the norm of an $\mathcal{A}$-measurable measure on $\mathcal{B}$ be given by

$$\|\rho\| = \sup_{x\in\mathcal{X}} \|\rho(x,\cdot)\|$$

If $\rho(x,\cdot)$ for any $x\in\mathcal{X}$ is a probability measure on $\mathcal{B}$, then ρ will be called <u>a Markov kernel from $(\mathcal{X},\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$</u>.

The ergodic coefficient was introduced by Dobrusin (1956). We shall state the definition and some basic properties (see also Iosifescu, 1972).

Let P be a Markov kernel from $(\mathcal{X},\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$. <u>The ergodic coefficient of P</u>, $\alpha(P)$, is defined by

(2.1) $\quad \alpha(P) = 1 - \sup|P(x',B) - P(x'',B)|$

where supremum is taken over all $x', x'' \in \chi$ and all $B \in \mathcal{B}$.

Clearly $0 \leq \alpha(P) \leq 1$. It is seen that $\alpha(P) = 1$ if and only if P does not depend upon x. In this case P is called a constant Markov kernel.

As is noted by Dobrusin, $\alpha(P)$ can be expressed in terms of the total variation norm $\|\cdot\|$ as follows:

$$(2.2) \qquad \alpha(P) = 1 - \tfrac{1}{2} \sup_{x', x'' \in \chi} \|P(x', \cdot) - P(x'', \cdot)\|$$

(This is easily seen from (2.1) considering the Hahn decomposition of the measure $P(x', \cdot) - P(x'', \cdot)$)

For convenience, we shall introduce the functional $\epsilon(P) \underset{\text{def}}{=} 1 - \alpha(P)$.

If Ω is a set, then let $r_\Omega = 0$ if Ω is infinite and let $r_\Omega = k^{-1}$ if Ω is finite with k points.

Lemma 2.1.

Let P be a Markov kernel from $(\chi, \mathfrak{A})$ to $(\mathcal{Y}, \mathcal{B})$. Then there exists a constant Markov kernel E such that $P = E + R$ and

$$\|R\| \leq 2(1 - r_\chi)\, \epsilon(P)$$

Remark: Iosifescu (1972) states the same result, except that the right hand side is $2\epsilon(P)$. Hence our result is somewhat stronger in the case of finite χ.

Proof: If χ is infinite, let x_0 be some fixed member of χ and let E be the constant Markov kernel defined by $E(x, B) = P(x_0, B)$; $x \in \chi, B \in \mathcal{B}$. Then if $R = P - E$,

$$\|R\| = \sup_{x \in \chi} \|P(x,\cdot) - P(x_o,\cdot)\|$$

$$\leq \sup_{x',x'' \in \chi} \|P(x',\cdot) - P(x'',\cdot)\| = 2\,\epsilon\,(P)$$

by (2.2).

If χ is finite with k points, then define E by

$$E(x,B) = k^{-1} \sum_{x' \in \chi} P(x',B) \;;\; x \in \chi, B \in \mathcal{B}$$

and let $R = P-E$. For fixed $x_o \in \chi$ we get

$$\|R(x_o,\cdot)\| = \|k^{-1} \sum_{x \neq x_o} [P(x_o,\cdot) - P(x,\cdot)]\|$$

$$\leq k^{-1}(k-1)\,\epsilon\,(P) \qquad \text{by (2.2)}$$

Let P' and P'' be Markov kernels, respectively from $(\chi,\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$ and from $(\mathcal{Y},\mathcal{B})$ to $(\mathcal{Z},\mathcal{C})$. Then the composition $P = P'P''$ is defined to be the Markov kernel from $(\chi,\mathcal{A})$ to $(\mathcal{Z},\mathcal{C})$ defined by

$$\text{(2.3)} \qquad P(x,C) = \int P''(y,C)\,P'(x,dy) \;;\; x \in \chi, C \in \mathcal{C} .$$

As is proved by Dobrusin (1956)

$$\text{(2.4)} \qquad \epsilon(P) \leq \epsilon(P')\,\epsilon\,(P'')$$

Example 1. Let χ and $\mathcal{Y}$ be (at most) countable sets. In this case we shall always assume that the σ-field $\mathcal{A}$ ($\mathcal{B}$) consists of all subsets of χ($\mathcal{Y}$) . Any Markov kernel P from χ to $\mathcal{Y}$ may now be represented by a Markov matrix $(p_{ij})_{i \in \chi,\; j \in \mathcal{Y}}$, where

$$P(i,A) = \sum_{j \in A} p_{ij} \;;\; i \in \chi,\ A \subseteq \mathcal{Y}.$$

As is shown by Dobrusin (1956) (see also Isaacson and Madsen, 1976), we can write

$$\alpha(P) = \inf_{i,j \in \chi} \sum_{k \in \mathcal{Y}} (p_{ik} \wedge p_{jk})$$

and

$$\varepsilon(P) = \tfrac{1}{2} \sup_{i,j \in \chi} \sum_{k \in \mathcal{Y}} |p_{ik} - p_{jk}|$$

We note that the composition $P = P'P''$ of two Markov kernels P' and P'' is now given by the usual matrix product of the corresponding Markov matrices.

The rest of this section is devoted to relating $\alpha(P)$ and $\varepsilon(P)$ to the concept of deficiencies, as defined by Le Cam (1964). A survey of the theory of deficiencies is given by Torgersen (1976 b).

The deficiency $\delta(\mathcal{E},\mathcal{F})$ of an experiment $\mathcal{E}$ relative to an experiment $\mathcal{F}$ measures the loss, under the least favorable conditions, by basing ourselves on $\mathcal{E}$ rather than on $\mathcal{F}$. We have $0 \leq \delta(\mathcal{E},\mathcal{F}) \leq 2(1-r_{\Theta})$, where Θ denotes the parameter set. If $\delta(\mathcal{E},\mathcal{F}) = 0$ then we say that $\mathcal{E}$ is more informative than $\mathcal{F}$ and write this $\mathcal{E} \geq \mathcal{F}$.

Let $(\Theta,\mathcal{T})$ and $(\chi,\mathcal{A})$ be measurable spaces. Interpreting $(\chi,\mathcal{A})$ as the sample space and Θ as the parameter set, we shall let the experiment $\mathcal{E}_P$ be defined by $\mathcal{E}_P = (\chi,\mathcal{A},P(\theta,\cdot)\;;\;\theta \in \Theta)$, where P is a Markov kernel from $(\Theta,\mathcal{T})$ to $(\chi,\mathcal{A})$. Let now Q be a Markov kernel from

$(\Theta,\mathcal{T})$ to $(\mathcal{Y},\mathcal{B})$, and assume that $\mathcal{E}_P$ is a dominated experiment. Then, by theorem 3 in LeCam (1964), we have

$$\delta(\mathcal{E}_P,\mathcal{E}_Q) = \inf_M \|PM - Q\|$$

where infimum is taken over all almost Markov kernels M from $(\chi,\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$, i.e. all real valued functions M from $\chi\times\mathcal{B}$ satisfying

(i) $M(\cdot,B)$ is $\mathcal{A}$-measurable for any fixed $B\in\mathcal{B}$.

(ii) $0 = M(\cdot,\emptyset)\leq M(\cdot,B)\leq M(\cdot,\mathcal{Y}) = 1$ a.e. $P(\theta,\cdot)$; $\theta\in\Theta$ for each $B\in\mathcal{B}$.

(iii) $M(\cdot, \bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} M(\cdot,B_i)$ a.e. $P(\theta,\cdot)$; $\theta\in\Theta$ when $B_1,B_2,\ldots$ are disjoint sets in $\mathcal{B}$.

The composition PM is defined by (2.3), noting that the integration is valid even if $P''(y,\cdot)$ is not a measure on $(\mathcal{Z},\mathcal{C})$. However, conditions (i) - (iii) above imply that PM is always a Markov kernel.

Let $\mathcal{L}$ denote the least informative experiment, i.e. an experiment satisfying $\mathcal{E}\geq\mathcal{L}$ for any experiment $\mathcal{E}$. $\mathcal{L}$ may be represented by any experiment $\mathcal{E}_P$ for which P is a constant Markov kernel. $\mathcal{L}$ is obviously a dominated experiment. Let Q be a Markov kernel from $(\Theta,\mathcal{T})$ to $(\mathcal{Y},\mathcal{B})$. Since P constant implies that PM is a constant Markov kernel for any almost Markov kernel M, it is seen that we have

$$(2.5)\qquad l(Q) \underset{\text{def}}{=} \delta(\mathcal{L},\mathcal{E}_Q) = \inf_P \|P-Q\|$$

where infimum is taken over all constant Markov kernels P.

Let $\mathcal{M}$ denote a totally informative experiment, i.e. an experiment such that $\mathcal{M} \geq \mathcal{E}$ for any $\mathcal{E}$. $\mathcal{M}$ may be represented by any experiment $\mathcal{E}_Q$ for which $Q(\theta',\cdot)$ and $Q(\theta'',\cdot)$ are mutually singular measures whenever $\theta' \neq \theta''$. If P is a Markov kernel from $(\Theta,\mathcal{T})$, then we shall define $m(P) = \delta(\mathcal{E}_P,\mathcal{M})$. It can be shown that $m(P) = 2$ whenever $\mathcal{E}_P$ is a dominated experiment not equivalent to $\mathcal{M}$ and Θ is not countable. Hence we shall be concerned with the functional $m(P)$ only in case Θ is at most countable. In this case $\mathcal{M}$ may be represented by the identity matrix I and we have $m(P) = \inf_M \|PM - I\|$ where infimum is taken over all Markov matrices M of suitable dimension.

Theorem 2.2

Let P be a Markov kernel from $(\Theta,\mathcal{T})$. Then

$$\varepsilon(P) \leq l(P) \leq 2(1-r_\Theta)\,\varepsilon(P)$$

Proof: Let $\eta > 0$. Then by (2.5) there is a constant Markov kernel Q such that $\|P-Q\| \leq l(P) + \eta$. Let Q_0 denote the probability measure $Q(\theta,\cdot)$. Then

$$\|P(\theta,\cdot) - Q_0\| \leq l(P) + \eta \quad \text{for all } \theta \in \Theta.$$

Hence, if $\theta' \neq \theta''$

$$\begin{aligned}\|P(\theta',\cdot) - P(\theta'',\cdot)\| &\leq \|P(\theta',\cdot) - Q_0\| + \|P(\theta'',\cdot) - Q_0\| \\ &\leq 2l(P) + 2\eta .\end{aligned}$$

From (2.2) follows since η was arbitrainly chosen, that $\epsilon(P) \leq l(P)$.

The right hand inequality of the theorem follows from lemma 2.1 and (2.5).

Corollary 2.3.

If Θ has two points, then $l(P) = \epsilon(P)$.

Corollary 5.5 of Torgersen (1976 a) states that if Θ has k points, then

(2.6) $\quad m(P)/(k(k-1)) \leq 2(1-r_\Theta) - l(P) \leq m(P)$

The right hand inequality follows directly from the triangle inequality for deficiencies, since

$$2(1-r_\Theta) = \delta(\mathcal{L},\mathcal{M}) \leq \delta(\mathcal{L},\mathcal{E}_P) + \delta(\mathcal{E}_P,\mathcal{M}) .$$

From (2.6) and theorem 2.2. we get

Theorem 2.4.

(i) $m(P) \geq 2(1-r_\Theta)\,\alpha(P)$

(ii) If Θ has two points, then $\frac{1}{2}m(P) \leq \alpha(P) \leq m(P)$

Remark: If Θ has more than two points, then there exists no constant $c > 0$ such that $c\, m(P) \leq \alpha(P)$ for all P .
In fact we may have $\alpha(P) = 0$ and $m(P) \neq 0$. This happens if at least two, but not all, measures $P(\theta,\cdot)$ are mutually singular.

As will be seen from example 2, the inequalities of theorem 2.4 can not be sharpened. In the example we prove in fact statement (ii) of theorem 2.4 in the case where P is given by a 2×2-matrix, without using formula (2.6).

Example 2: Let P be a Markov kernel from $\{1,2\}$ into $\{1,2\}$ defined by the 2×2-matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

By corollary 2.3, $l(P) = \epsilon(P)$, which by the formula of example 1 equals $|1-\alpha-\beta|$. Hence $\alpha(P) = 1 - |1-\alpha-\beta|$. It remains to compute $m(P)$. By definition, $m(P) = \inf_M \|PM - I\|$, where infimum is taken over all 2×2-Markov matrices M. A geometrical approach, which will not be given here, leads us to

$$m(P) = 2 \cdot \frac{[\alpha\vee\beta]\wedge[(1-\alpha)\vee(1-\beta)]}{1+|\alpha-\beta|}$$

If the cases $\alpha+\beta \lesseqgtr 1$ and $\alpha \lesseqgtr \beta$ are treated separetely, then it is now easy to show that $m(P)\geq\alpha(P)$ and $m(P)\leq 2\,\alpha(P)$. We finally show that these inequalities can not be sharpened. Put $0<\alpha=\beta<\frac{1}{2}$. Then $m(P)=\alpha(P)=2\alpha$. If we put $\beta=0$, then $m(P)=2\alpha/(1+\alpha)$ and $\alpha(P)=\alpha$. Hence $m(P)/\alpha(P)\uparrow 2$ as $\alpha\downarrow 0$. In fact it is not difficult to prove that $m(P)<2\alpha(P)$ holds whenever $\alpha(P)\neq 0$.

3. Application to non-homogeneous Markov chains.

By Dobrusin (1956) (see also Iosifescu, 1972), a non-homogeneous Markov chain (NMC) can be considered as a sequence of measurable state spaces $(X_j,\mathfrak{A}_j)$ and Markov kernels ${}^jP$ from $(X_j,\mathfrak{A}_j)$ to $(X_{j+1},\mathfrak{A}_{j+1})$; $j = 0,1,2,\ldots$. Then ${}^jP(x_j,A_{j+1})$ is the probability of being in $A_{j+1}\in X_{j+1}$ at time $j+1$, conditional on being in $x_j \in X_j$ at time j . The n-step transition probability ${}^jP^n$ is a Markov kernel from $(X_j, \mathcal{A}_j)$ to $(X_{j+n}, \mathcal{A}_{j+n})$; $j \geq 0$, $n \geq 1$, defined by

$$ {}^jP^n = ({}^jP)({}^{j+1}P) \ldots ({}^{n+j-1}P) $$

where composition of Markov kernels is defined in section 2.

Hence by (2.4)

$$ \varepsilon({}^jP^n) \leq \prod_{i=j}^{n+j-1} \varepsilon({}^iP) \; ; \; j \geq 0 \; , \; n \geq 1 $$

Definition 3.1.

An NMC is said to be weakly ergodic if $\lim_{n \to \infty} \varepsilon({}^jP^n) = 0$ for all $j \geq 0$.

The following result is taken from Iosifescu (1972):

Theorem 3.2.

An NMC is weakly ergodic if and only if either one of the following conditions is fulfilled.

(i) For any $j \geq 0$ there is a sequence of constant Markov kernels $\{{}^jE_n\}_{n \geq 1}$ such that

$$ \lim_{n \to \infty} \| {}^jP^n - {}^jE_n \| = 0 $$

(ii) There exists a strictly increasing sequence $(j_k)_{k \geq 1}$ of natural numbers such that

$$ \sum_{k \geq 1} \alpha({}^{j_k}P^{j_{k+1}-j_k}) \text{ diverges} $$

(iii) For an arbitrarily fixed $0 < \varepsilon < 1$ there is a function f mapping the set of natural numbers into itself such that

$$\liminf_{j\to\infty} \alpha({}^{j}P^{f(j)}) \geq \epsilon$$

Given an NMC we may for any $j \geq 0$ define a sequence of experiments $\{\mathcal{E}_n^{(j)}\}_{n\geq 1}$ where $\mathcal{E}_n^{(j)}$ is the experiment of observing the chain at time $n+j$, the state at time j being the unknown parameter. Hence $\mathcal{E}_n^{(j)}$ is an experiment with parameter space χ_j defined by the Markov kernel ${}^{j}P^{n}$ from $(\chi_j, \mathcal{O}\!\mathcal{L}_j)$ to $(\chi_{n+j}, \mathcal{O}\!\mathcal{L}_{n+j})$. The following results are direct consequences of theorem 3.2.

Corollary 3.3.

An NMC is weakly ergodic if and only if for any $j \geq 0$, the sequence $\{\mathcal{E}_n^{(j)}\}_{n\geq 1}$ converges to the minimal informative experiment (with respect to deficiencies).

Proof: $\mathcal{E}_n^{(j)}$ is defined by ${}^{j}P^{n+j}$. Hence $\delta(\mathcal{L}, \mathcal{E}_n^{(j)}) = 1({}^{j}P^{n+j})$, which by theorem 2.2 tends to 0 as $n \to \infty$ if and only if $\epsilon({}^{j}P^{n+j}) \to 0$. But this means by definition 3.1 that the NMC is weakly ergodic.

Remark: The above corollary is also an immediate consequence of (2.5) and theorem 3.2 (i).

Corollary 3.4.

Given a weakly ergodic NMC such that the spaces χ_j ; $j \geq 0$ are at most countable, then

(i) There exists a strictly increasing sequence $(j_k)_{k\geq 1}$ of natural numbers such that

$$\sum_{k \geq 1} \delta(\mathcal{E}^{(j_k)}_{j_{k+1}-j_k}, \mathcal{M}) \quad \text{diverges}$$

(ii) for any $0 < \varepsilon < 1$ there is a function f mapping the set of natural numbers into itself such that

$$\liminf_{j \to \infty} \delta(\mathcal{E}^{(j)}_{f(j)}, \mathcal{M}) \geq \varepsilon$$

If, for all $j \geq 0, X_j$ has exactly two points, then each of conditions (i) and (ii) imply that the given NMC is weakly ergodic.

Proof: These results follow from theorem 2.4 and theorem 3.2.

We shall finally study homogeneous Markov chains (HMC). A HMC is completely determined by a measurable state space $(X, \mathcal{A})$ and a Markov kernel P from $(X, \mathcal{A})$ to $(X, \mathcal{A})$. Now the n-step transition probabilities ${}^jP^n$ equal $P^n = P \cdot P \cdots P$, so the sequences $\{\mathcal{E}_n^{(j)}\}_{n \geq 1}$ are identical for $j = 0, 1, \ldots$. Hence it is enough to consider the experiments $\mathcal{E}_n, n \geq 1$, with parameter space X, and which are given by the Markov kernel P^n. From corollary 3.4 we get

Corollary 3.5.

An HMC is weakly ergodic if and only if the sequence $\{\mathcal{E}_n\}$ converges to the minimal informative experiment.

From Lindqvist (1977) follows that this result holds if X is a finite set. Furthermore, Lindqvist (1977) proves that $\delta(\mathcal{L}, \mathcal{E}_n)$ converges to 0 with exponential speed. The next theorem extends this result to the case of general state space.

Theorem 3.6.

Assume that the sequence $\{\mathscr{E}_n\}$ is constructed from a HMC. If $\delta(\mathscr{L},\mathscr{E}_n) \to 0$ as $n \to \infty$, then the rate of convergence is exponential.

Proof: It is by theorem 2.2 enough to prove that $\varepsilon(P^n) \to 0$ with exponential speed whenever P defines a weakly ergodic HMC. Since by assumption $\varepsilon(P^n) \to 0$ we must have $\varepsilon(P^{n_o}) = \eta < 1$ for some $n_o \geq 1$. Then given n, choose i and $0 \leq j < n_o$ such that $n = in_o + j$. By (2.4)

$$\varepsilon(P^n) = \varepsilon(P^{in_o+j}) \leq [\varepsilon(P^{n_o})]^i \, \varepsilon(P^j) \leq c\eta^{n/n_o}$$

for some $c > 0$ independent of n. The result follows.

Remark: That the convergence of $\delta(\mathscr{L},\mathscr{E}_n)$ is not exponential in general for NMC's, is seen from the following example.

Example 3. Let an NMC with state spaces $X_j = \{1,2\}$ for all $j \geq 0$ be given by the transition matrices

$$^jP = \begin{pmatrix} 1 - \frac{1}{j+2} & \frac{1}{j+2} \\ 0 & 1 \end{pmatrix} \quad ; \quad j = 0,1,2,\ldots$$

Here $\varepsilon(^jP) = 1 - (j+2)^{-1}$ for $j = 0,1,2,\ldots$. It is easily verified that if P and Q are 2×2-Markov matrices, then $\varepsilon(PQ) = \varepsilon(P)\,\varepsilon(Q)$. Hence

$$\varepsilon(^jP^n) = \varepsilon(^jP)\,\varepsilon(^{j+1}P)\ldots\varepsilon(^{j+n-1}P) = (j+1)/(n+j+1)$$

so $\varepsilon(^jP^n) \to 0$ as $n \to \infty$ for any $j \geq 0$. However, the convergence is not exponential.

Remark: From theorem V.4.3. of Isaacson and Madsen (1976) follows that the chain considered in example 3 is in fact strongly ergodic (for definition, see e.g. Iosifescu, 1972). A criterion for exponential convergence of $\delta(\mathcal{L}, \mathcal{E}_n^{(j)})$ in the case of strongly ergodic NMC's follows from the result of Cheng-Chi Huang et al. (1976).

R E F E R E N C E S

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

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

Isaacson, D.L. and Madsen, R.W. (1976). Markov chains: Theory and applications. Wiley & Sons, New York.

Le Cam, L. (1964). Sufficiency and approximate sufficiency. Ann. Math. Statist, 35, 1419-1455.

Lindqvist, B. (1977). How fast does a Markov chain forget the initial state. A decision theoretical approach. To appear in Scand. J. Statist.

Madsen, R.W. (1971). A note on some ergodic theorems of A. Paz. Ann. Math. Statist, 42, 405-408.

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

Torgersen, E.N. (1976 a). Deviations from total information and from total ignorance as measures of information. Stat. research rep. no. 3, Univ of Oslo.

Torgersen, E.N. (1976 b). Comparison of Statistical Experiments. Scand. J. Statist, 3, 186-208.

Cheng-Chi Huang, Isaacson, D. and Vinograde, B. (1976). The rate of convergence of certain nonhomogeneous Markov chains. Z. Wahrscheinlichkeitstheorie verw. Gebiete. 35, 141-146.