Z. Wahrscheinlichkeitstheorie verw. Gebiete
44, 155–158 (1978)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

A Decision Theoretical Characterization of Weak Ergodicity

Bo Lindqvist

Dept. of Mathematics, University of Oslo, Blindern, N-Oslo 3, Norway

Summary. The relation between the ergodic coefficient and deficiency relative to the least informative experiment is investigated. The result is applied to nonhomogeneous Markov chains (NMC's). Our main result can be described as follows: Given an NMC, define the experiments $\mathscr{E}_n^{(j)}$ for $n \geqq 1$ 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, $\mathscr{E}_n^{(j)}$ converges as $n \to \infty$ (with respect to deficiencies) to the least informative experiment. It is finally shown that in the homogeneous case, the rate of convergence is always exponential.

1. Introduction

In [4] we studied the experiment $\mathscr{E}_n$ 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 $\mathscr{E}_n$ 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 [3]. 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. This leads to a new characterization of weak ergodicity of NMC's.

The basic tool in our study is the ergodic coefficient introduced by Dobrušin [1] and used in the study of weak ergodicity of NMC's by e.g. Paz [6], Madsen [5] and Iosifescu [2]. The relation between the ergodic coefficient and deficiencies is derived using ideas from Torgersen [7]. A survey of the general theory of deficiencies is given by Torgersen [8].

2. The Ergodic Coefficient and Deficiencies

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

0044-3719/78/0044/0155/$01.00

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

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

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

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

$$\|\rho\| = \sup_{x \in \chi} \|\rho(x, \cdot)\|.$$

If $\rho(x, \cdot)$ for any $x \in \chi$ is a probability measure on $\mathscr{B}$, then ρ will be called *a Markov kernel from* $(\chi, \mathscr{A})$ *to* $(\mathscr{Y}, \mathscr{B})$. If in addition ρ does not depend upon x, then ρ is called a *constant* Markov kernel.

Let P be a Markov kernel from $(\chi, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$. The *ergodic coefficient of* $P, \alpha(P)$, is given by

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

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

Let P be a Markov kernel from $(\chi, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$. Then by Iosifescu [2] there exists a constant Markov kernel E such that $P = E + R$ and

$$\|R\| \leqq 2\varepsilon(P). \tag{2}$$

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

$$P(x, C) = \int P''(y, C) P'(x, dy); \qquad x \in \chi, \ C \in \mathscr{C}. \tag{3}$$

As is proved by Dobrušin [1]

$$\varepsilon(P) \leqq \varepsilon(P') \varepsilon(P''). \tag{4}$$

The rest of this section is devoted to relating $\varepsilon(P)$ to the concept of deficiencies, as defined by LeCam [3].

The deficiency $\delta(\mathscr{E}, \mathscr{F})$ of an experiment $\mathscr{E}$ relative to an experiment $\mathscr{F}$ measures the loss, under the least favorable conditions, by basing ourselves on $\mathscr{E}$ rather than on $\mathscr{F}$. If $\delta(\mathscr{E}, \mathscr{F}) = 0$ then we say that $\mathscr{E}$ is more informative than $\mathscr{F}$ and write this $\mathscr{E} \geqq \mathscr{F}$.

Let $(\Theta, \mathscr{T})$ and $(\chi, \mathscr{A})$ be measurable spaces. Interpreting $(\chi, \mathscr{A})$ as the sample space and Θ as the parameter set, we shall let the experiment $\mathscr{E}_P$ be defined by $\mathscr{E}_P = (\chi, \mathscr{A}, P(\theta, \cdot);\ \theta \in \Theta)$, where P is a Markov kernel from $(\Theta, \mathscr{T})$ to $(\chi, \mathscr{A})$. Let now Q be a Markov kernel from $(\Theta, \mathscr{T})$ to $(\mathscr{Y}, \mathscr{B})$, and assume that $\mathscr{E}_P$ is a dominated experiment. Then, by Theorem 3 in LeCam [3], we have

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

where infimum is taken over all almost Markov kernels M from $(\chi, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$, (see [8]).

The Markov kernel PM is defined by (3), noting that the integration is valid even if $P''(y,\cdot)$ is not a measure on $(\mathscr{X},\mathscr{C})$.

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

$$l(Q)\underset{\mathrm{def}}{=}\delta(\mathscr{L},\mathscr{E}_Q)=\inf_P\|P-Q\| \tag{5}$$

where infimum is taken over all constant Markov kernels P.

Let P be a Markov kernel from $(\Theta,\mathscr{T})$. The fundamental relation between the ergodic coefficient and deficiencies is given by

$$\varepsilon(P)\leqq l(P)\leqq 2\varepsilon(P). \tag{6}$$

The right hand inequality of (6) follows from the considerations leading to (2). The left hand inequality is proved as follows:

Let $\eta>0$. Then by (5) there is a constant Markov kernel Q such that $\|P-Q\|\leqq l(P)+\eta$. Let Q_0 denote the probability measure $Q(\theta,\cdot)$. Then

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

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

$$\|P(\theta',\cdot)-P(\theta'',\cdot)\|\leqq\|P(\theta',\cdot)-Q_0\|+\|P(\theta'',\cdot)-Q_0\|\leqq 2l(P)+2\eta.$$

From (1) follows, since η was arbitrarily chosen, that $\varepsilon(P)\leqq l(P)$.

3. Application to Non-Homogeneous Markov Chains

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

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

where composition of Markov kernels is defined by (3).

Following Iosifescu [2] we shall say that an NMC is *weakly ergodic* if $\lim_{n\to\infty}\varepsilon({}^jP^n)=0$ for all $j\geqq 0$.

Given an NMC we may for any $j\geqq 0$ define a sequence of experiments $\{\mathscr{E}_n^{(j)}\}_{n\geqq 1}$ where $\mathscr{E}_n^{(j)}$ is the experiment of observing the chain at time $n+j$, the state at time j being the unknown parameter. Hence $\mathscr{E}_n^{(j)}$ is an experiment with parameter space χ_j defined by the Markov kernel ${}^jP^n$ from $(\chi_j,\mathscr{A}_j)$ to $(\chi_{n+j},\mathscr{A}_{n+j})$.

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

Proof. $\mathscr{E}_n^{(j)}$ is defined by ${}^jP^{n+j}$. Hence $\delta(\mathscr{L}, \mathscr{E}_n^{(j)}) = l({}^jP^{n+j})$, which by (6) tends to 0 as $n \to \infty$ if and only if $\varepsilon({}^jP^{n+j}) \to 0$. But this means by definition that the NMC is weakly ergodic. □

We shall finally study homogeneous Markov chains (HMC). A HMC is completely determined by a measurable state space $(\chi, \mathscr{A})$ and a Markov kernel P from $(\chi, \mathscr{A})$ to $(\chi, \mathscr{A})$. Now the n-step transition probabilities ${}^jP^n$ equal $P^n = PP \dots P$, so the sequences $\{\mathscr{E}_n^{(j)}\}_{n \geqq 1}$ are identical for $j = 0, 1, \dots$. Hence it is enough to consider the experiments $\mathscr{E}_n$, $n \geqq 1$, with parameter space χ, and which are given by the Markov kernel P^n. Theorem 1 immediately implies

Corollary 1. *An HMC is weakly ergodic if and only if the sequence $\{\mathscr{E}_n\}$ converges to the minimal informative experiment.*

From [4] it follows that this result holds if χ is a finite set. Furthermore, [4] proves that $\delta(\mathscr{L}, \mathscr{E}_n)$ converges to 0 with exponential speed. The next theorem extends this result to the case of general state space.

Theorem 2. *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 (6) 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 $\varepsilon(P^{n_0}) = \eta < 1$ for some $n_0 \geqq 1$. Then given n, choose i and $0 \leqq j < n_0$ such that $n = in_0 + j$. By (4)

$$\varepsilon(P^n) = \varepsilon(P^{in_0+j}) \leqq [\varepsilon(P^{n_0})]^i \, \varepsilon(P^j) \leqq c\eta^{n/n_0}$$

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

References

1. Dobrušin, R.L.: Central limit theorem for nonstationary Markov chains, I, II. Theor. Probability Appl. **1**, 65–80 and 329–383 (1956)
2. Iosifescu, M.: On two recent papers on ergodicity in nonhomogeneous Markov chains. Ann. Math. Statist. **43**, 1732–1736 (1972)
3. LeCam, L.: Sufficiency and approximate sufficiency. Ann. Math. Statist, **35**, 1419–1455 (1964)
4. Lindqvist, B.: How fast does a Markov chain forget the initial state? A decision theoretical approach. Scand. J. Statist, **4**, 145–152 (1977)
5. Madsen, R.W.: A note on some ergodic theorems of A. Paz. Ann. Math. Statist, **42**, 405–408 (1971)
6. Paz, A.: Ergodic theorems for infinite probabilistic tables. Ann. Math. Statist, **41**, 539–550 (1970)
7. Torgersen, E.N.: Deviations from total information and from total ignorance as measures of information. Statist. research report, Univ. of Oslo, (1976)
8. Torgersen, E.N.: Comparison of statistical experiments. Scand. J. Statist, **3**, 186–208 (1976)

Received September 12, 1977