Stochastic Processes and their Applications 22 (1986) 315-322
North-Holland

APPROACHING CONSENSUS CAN BE DELICATE WHEN POSITIONS HARDEN

Joel E. COHEN
Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, USA

John HAJNAL
London School of Economics and Political Science, Houghton Street, London WC2A 2AE, England

Charles M. NEWMAN
Department of Mathematics, University of Arizona, Tucson, AZ 85721, USA

Received 24 May 1985
Revised 2 January 1986

A model of consensus leads to examples in which the ergodic behavior of a nonstationary product of random nonnegative matrices depends discontinuously on a continuous parameter. In these examples, a product of random matrices, each of which is a scrambling stochastic matrix, changes from being weakly ergodic (asymptotically of rank 1) with probability 1 to being weakly ergodic with probability 0 as a parameter of the process changes smoothly.

products of random nonnegative matrices * ergodicity * inhomogeneous products * zeta function * strong limit laws * zero-one laws

1. Introduction

Suppose n experts are trying to evaluate some quantity that can be described by a real scalar or real vector. Their initial estimates are respectively F_i^1, $i=1,\ldots,n$. They share and discuss their estimates and form new estimates F_i^2. The process then iterates to yield further estimates F_i^k, $k=1,2,\ldots$, $i=1,\ldots,n$.

Suppose (DeGroot, 1974; Chatterjee and Seneta, 1977) that at each stage $k+1$ of the process, the ith expert forms his or her new estimate as a weighted mean of all prior estimates at stage k:

$$F_i^{k+1}=\sum_{j=1}^{n} a_{ij}^{(k)}F_j^k,\ \sum_{j=1}^{n} a_{ij}^{(k)}=1,\quad k=1,2,\ldots,\quad i=1,\ldots,n.$$

The weighting coefficients $a_{ij}^{(k)}$ may depend on the trial k. If each expert pays no attention to the estimates of the other experts, the weights are given by the identity matrix, $A_k=I$ for all k. The model is open to empirical testing because if any expert's estimate at stage $k+1$ falls outside the convex hull of all estimates at stage k, the model is wrong. Writing F^k for the n-vector with elements F_i^k and $A_k=(a_{ij}^{(k)})$, we have for all $k \geq 2$, $F^k = A_{k-1}A_{k-2} \cdots A_1 F^1$. We shall say that $\{A_k\}_1^\infty$ is *consensual* if for every F^1 the experts will approach consensus, i.e. $|F_i^k - F_j^k| \to 0$ for all $i, j = 1, \ldots, n$ as $k \uparrow \infty$. If $\{A_k\}_1^\infty$ is not consensual, then there exist initial estimates F^1 such that consensus will not occur, i.e. such that, for some i and j, $i \neq j$, $|F_i^k - F_j^k| \not\to 0$.

To allow for the possibility that the evaluation process begins at stage $j > 1$ with some F^j not obtained from an earlier F^{j-1}, define (see Hajnal, 1958) $\{A_k\}_1^\infty$ to be (left) *weakly ergodic* if, for each j, $\{A_k\}_j^\infty$ is consensual. A more detailed definition will be given below in Section 2. Note that $\{A_k\}_1^\infty$ may be consensual but not weakly ergodic when, for example, a single A_k has all its rows equal.

DeGroot (1974) gives necessary and sufficient conditions for $\{A_k\}_1^\infty$ to be consensual when $A_k = A$ for all $k \geq 1$. When $A_k = A$, Berger (1981) observes that, for some initial estimates F^1, the experts will approach consensus even when $\{A_k\}_1^\infty$ is not consensual (for example, if all the experts happen to agree at the outset). Berger gives necessary and sufficient conditions on A and F^1 for the experts to approach consensus. He admits (p. 417) that it is "hard to imagine" that the conditions required of F^1 would be satisfied when A is such that $\{A_k\}$ is *not* consensual.

Chatterjee and Seneta (1977) point out that the experts may approach consensus even if they gradually *harden their positions* by increasing the weight they assign to their own estimates and decreasing the weight they assign to the other estimates.

The purpose of this paper is to show by examples that, when experts harden their positions, a very small change in the process of weighting other experts' estimates can divert the process from moving toward consensus almost surely to remaining in dissension almost surely, or vice versa. More generally, the ergodic behavior of a product of random nonnegative matrices, including e.g. a Markov chain in random environments, can depend discontinuously on a continuous parameter. In the examples to be described, a nonstationary product of random matrices changes from being weakly ergodic with probability 1 (w.p. 1) to being weakly ergodic with probability 0 (w.p. 0) as a parameter of the process changes smoothly.

Other aspects of the dependence on a parameter of the asymptotic behavior of a product of random matrices have been investigated by Kingman (1976), Goldsheid (1980), Cohen (1980), and Kifer (1982). Models of consensus among experts are reviewed by Seneta (1981, Ch. 4) (along with the relevant matrix theory), Wagner and Lehrer (1981), Zidek (1983) and, most comprehensively, Genest and Zidek (1986).

Sections 2 and 3 relate ergodic behavior to zero-one laws for random versions of Riemann's zeta function and give some special examples of discontinuity in ergodic behavior. Section 4 interprets the results of Sections 2 and 3 in terms of the DeGroot-Chatterjee-Seneta model of consensus.

2. Weak ergodicity of stochastic matrix products

All matrices in this paper will be assumed to be $n \times n$, $1 < n < \infty$, and nonnegative, i.e. having every element nonnegative.

If $A_1, A_2, \ldots$ is a sequence of matrices, define $L\{A_k\}$ to be the doubly indexed family of matrices $\{L_{k,m};\ k, m = 1, 2, \ldots\}$ where

$$L_{k,m} = A_{k+m}A_{k+m-1} \cdots A_{k+2}A_{k+1}, \quad k, m = 1, 2, \ldots \tag{1}$$

is the product of m matrices from the sequence $\{A_k\}_1^\infty$ starting from A_{k+1} and multiplying successive factors on the left (L for "left"). We denote the element in row i and column j of $L_{k,m}$ by $(L_{k,m})_{ij}$.

Similarly, define $R\{A_k\} = \{R_{k,m};\ k, m = 1, 2, \ldots\}$ where $R_{k,m} = A_{k+1}A_{k+2} \cdots A_{k+m-1}A_{k+m}$, $k, m = 1, 2, \ldots$, with i, j element $(R_{k,m})_{ij}$.

For any $n \times n$ stochastic matrix $P = (p_{ij})$, $1 < n < \infty$, define

$$\gamma(P) = \tfrac{1}{2} \max_{i,j} \sum_{k=1}^{n} |p_{ik} - p_{jk}|. \tag{2}$$

Then $0 \leq \gamma(P) \leq 1$ and $\gamma(P) = 0$ if and only if all rows of P are identical, i.e. P has rank 1.

A sequence $\{A_k\}_1^\infty$ of stochastic matrices A_k is defined to be left (or right) weakly ergodic if, for all k, $L_{k,m}$ (or $R_{k,m}$) asymptotically has rank 1 as $m \to \infty$; *i.e.* if for all $k \geq 1$, $\lim_{m\to\infty} \gamma(L_{k,m}) = 0$ (or the same with L replaced by R).

Hajnal (1958) discusses only rightward products. Leftward products are introduced and compared to rightward products by Chatterjee and Seneta (1977). For brevity we shall henceforth replace "left and right weakly ergodic" by "ergodic". Chatterjee and Seneta prove that for leftward products of stochastic matrices strong and weak ergodicity are equivalent.

Let $\{B_k\}_{k=1}^\infty$ be a sequence of random stochastic matrices. The ergodicity of $\{B_k\}$ is an asymptotic property which is unaffected by any single B_k, unlike the consensuality of $\{B_k\}_1^\infty$. Let $\{W_k\}_{k=1}^\infty$ be any deterministic or random sequence of permutation matrices. Clearly $\{W_k\}$ is not ergodic.

Let $\{X_k\}_{k=1}^\infty$ be a sequence of real-valued random variables concentrated on $[1, \infty)$. Define the random variable

$$\zeta = \sum_{k=1}^{\infty} k^{-X_k}. \tag{3}$$

Theorem 1. *Suppose there exists positive constants c_1 and c_2 and a positive integer k_0 such that for $k \geq k_0$ and for all i, j,*

$$0 < c_1 k^{-X_k} \leq |(B_k)_{ij} - (W_k)_{ij}| \leq c_2 k^{-X_k} \quad w.p.\ 1. \tag{4}$$

Then

$$P(\{B_k\} \text{ is ergodic}) = P(\zeta = \infty). \tag{5}$$

Proof. Let $d_k = \min_{i,j}|(B_k)_{ij} - (W_k)_{ij}|$, $e_k = \max_{i,j}|(B_k)_{ij} - (W_k)_{ij}|$. Then, by (4), $c_1 k^{-X_k} \leq d_k \leq e_k \leq c_2 k^{-X_k}$. Thus $\zeta < \infty$ implies $\sum e_k < \infty$. By Hajnal's (1958, p. 244) Theorem 6, $\{B_k\}$ then shares the nonergodicity of $\{W_k\}$, i.e. $\zeta < \infty$ implies $\{B_k\}$ is not ergodic.

On the other hand, $\zeta=\infty$ implies $\sum d_k=\infty$, which easily implies $\sum \min_{i,j}(B_k)_{ij}=\infty$, which in turn implies $\{B_k\}$ is ergodic, by the Corollary to Theorem 4 of Chatterjee and Seneta (1977, p. 93). □

This theorem reduces the question of ergodicity for models which satisfy (4) to the question of the divergence of the random zeta function (3), which is the topic of section 3. When the divergence of the series (3) is governed by a zero-one law of probability theory, it comes as no surprise, in the light of (5), that the ergodic behavior of $\{B_k\}$ is discontinuous.

3. Discontinuity in ergodic behavior

We now give conditions under which ζ, defined in (3), converges or diverges almost surely. Define the moment generating function of X_k to be $\phi_k(t)=E(\exp[tX_k])$. If X_k is concentrated on the positive integers only, denote $P[X_k=s]=p_{sk}$, for all positive integers s and k. The next theorem, which concerns independent X_k's, is a direct consequence of Kolmogorov's three-series theorem (Loeve, 1977, I:24a) applied to $Y_k=k^{-X_k}=\exp[-(\ln k)X_k]$. It will be followed by some specific examples.

Theorem 2. *Let $X_1, X_2, \ldots$ be mutually independent. Then* (i) $\zeta<\infty$ *w.p.* 1 *if and only if* $\Phi\equiv\sum_{k=1}^{\infty}\phi(-\ln k)<\infty$. $\zeta=\infty$ *w.p.* 1 *if and only if* $\Phi=\infty$. (ii) *If, for every* k, X_k *is concentrated on the positive integers, then* $\zeta<\infty$ *w.p.* 1 *if and only if* $\rho\equiv\sum_{k=1}^{\infty}p_{1k}k^{-1}<\infty$. $\zeta=\infty$ *w.p.* 1 *if and only if* $\rho=\infty$.

We now turn to some specific examples of $\{X_k\}$. Since the convergence of ζ depends on the distribution of X_k only as $k\to\infty$, we need to specify the distributions only for large k. The criteria given in the examples follow from Theorem 2 and the standard facts that

$$\sum_{k=2}^{\infty}[k(\ln k)^a]^{-1}<\infty \quad \text{if } a>1,$$

$$=\infty \quad \text{if } a\leqslant 1;$$

$$\sum_{k=3}^{\infty}[k(\ln k)(\ln k)^a]^{-1}<\infty \quad \text{if } a>1,$$

$$=\infty \quad \text{if } a\leqslant 1.$$

In Example 1, we use in addition the formula for exponential random variables (e.g. Johnson and Kotz, 1970, p. 210) $\phi_k(-\ln k)=[k(1+\sigma_k\ln k)]^{-1}$.

Example 1. Let $\{X_k\}_1^\infty$ be a sequence of independent exponentially distributed random variables concentrated on $[1, \infty)$ with probability density functions $f_k(x) = 0$, $x < 1$, $f_k(x) = \sigma_k^{-1} \exp[-(x-1)/\sigma_k]$, $x \geq 1$, $\sigma_k > 0$. (i) Then $\zeta = \infty$ w.p. 1 if for some $c > 0$, and for all $k \geqslant 3$, $\sigma_k \leqslant c \ln \ln k$. In particular $\zeta = \infty$ w.p. 1 if $\sigma_k \leqslant c$ with $c > 0$ independent of k or if $\sigma_k \leqslant c(\ln k)^b$ with $b \leqslant 0$. (ii) Also $\zeta < \infty$, w.p. 1 if, for $k \geqslant 3$ and for $c > 0$, $a > 1$, $\sigma_k \geqslant c(\ln \ln k)^a$. In particular, $\zeta < \infty$ w.p. 1 if $\sigma_k \geqslant ck^b$ with $b > 0$ or if $\sigma_k \geqslant c(\ln k)^b$ with $b > 0$.

If X_k has density $f_k(x)$ concentrated on $[1, \infty)$, asymptotically $\phi_k(-\ln k)$ depends on $f_k(x)$ for x near 1 only. For example, if for some $\varepsilon > 0$, $b > 0$, and $0 < c < \infty$, $f_k(x) \leqslant c(x-1)^b$ for $1 \leqslant x \leqslant 1 + \varepsilon$ and all large k, it can be shown that $\Phi < \infty$ so that $\zeta < \infty$ w.p. 1. This follows easily from the estimate

$$\phi_k(-t) = \int_1^\infty \mathrm{e}^{-tx} f_k(x)\,\mathrm{d}x \leqslant \int_1^\infty \mathrm{e}^{-tx} c(x-1)^b\,\mathrm{d}x + \mathrm{e}^{-t(1+\varepsilon)} \int_{1+\varepsilon}^\infty f_k(x)\,\mathrm{d}x$$
$$\leqslant c\Gamma(b+1)\,\mathrm{e}^{-t}/t^{b+1} + \mathrm{e}^{-t(1+\varepsilon)}.$$

Example 2. Let $\{X_k\}_1^\infty$ be a sequence of independent random variables concentrated on the positive integers $1, 2, \ldots$ with $P[X_k = s] = p_{sk}$ as before. (i) Then $\zeta = \infty$ w.p. 1 if, for $k > 3$ and some $c > 0$, $p_{1k} \geqslant c[(\ln k)(\ln \ln k)]^{-1}$, and in particular if $p_{1k} \geqslant c > 0$ with c independent of k; or $p_{1k} \geqslant c(\ln k)^{-a}$ with $a \leqslant 1$; or $p_{1k} \geqslant c[(\ln k)(\ln \ln k)^a]^{-1}$ with $a \leqslant 1$. (ii) Also $\zeta < \infty$ w.p. 1 if, for some $c > 0$, $a > 1$, and all $k \geqslant 3$, $p_{1k} \leqslant c[(\ln k)(\ln \ln k)^a]^{-1}$, and in particular if $p_{1k} \leqslant ck^{-b}$ with $b > 0$ or if $p_{1k} \leqslant c(\ln k)^{-a}$ with $a > 1$.

As a referee points out, Theorem 2 can be illustrated by an example in which $\{X_k\}$ are independently and *identically* distributed on $(1, \infty)$. However, such $\{X_k\}$ have no interpretation in terms of the hardening of positions in an approach to consensus, which is the main application of the theory here, so we omit the example.

The next theorem concerns X_k's which form a positive-integer valued homogeneous Markov chain. If we denote by S_j the "time" of the jth occurrence of the value or state 1 ($S_j = \infty$ if 1 occurs fewer than j times), then it is easy to see that $\zeta < \infty$ if and only if $\sum (S_j)^{-1} < \infty$. The transient and positive recurrent cases of the theorem follow directly from this observation. The null recurrent case requires the additional result that for i.i.d. strictly positive T_i's, $\sum (T_1 + \cdots + T_j)^{-1}$ converges (or diverges) w.p. 1 if and only if $\int_0^1 [1 - E(e^{-tT_1})]^{-1}\,\mathrm{d}t < \infty$ (or $= \infty$). The proof of this fact, which we have not found stated in the literature, is straightforward but lengthy. We follow a referee's request to omit it from the paper; details are available directly from the authors.

Theorem 3. Let $\{X_k\}$ be a homogeneous Markov chain with state space equal to the positive integers. Let T be the (positive) random interval between the first and second occurrences of state 1, i.e. if $X_i = 1$, $X_j = 1$, $i < j$, and $X_k \neq 1$ for $0 < k < i$ and for $i < k < j$, then $T = j - i$. Let $g(t) = E(\mathrm{e}^{-tT})$. (i) If the state 1 is transient, then $\zeta < \infty$ w.p. 1. (ii) If the state 1 is positive recurrent, then $\zeta = \infty$ w.p. 1. (iii) If the state 1 is null recurrent, then $\zeta<\infty$ (or $=\infty$) w.p. 1 if and only if

$$\int_0^1 [1-g(t)]^{-1}\,dt<\infty \quad (\text{or}=\infty).$$

Example 3. Let $X_1=1$ w.p. 1. For positive integers i, j and n, let

$$\begin{aligned} P(X_{n+1}=j\,|\,X_n=i) &= p_i>0 && \text{if } j=i+1,\\ &= q_i=1-p_i && \text{if } j=1,\\ &=0 && \text{otherwise.}\end{aligned}$$

Except for a translation by 1 in the numbering of states, this defines the transition matrix of "the basic example" of Kemeny, Snell and Knapp (1966, p. 83). Let

$$\beta_k=\prod_{j=1}^{k} p_j,\quad k\geq 1;\qquad T=\min\{k\geq 2\colon X_k=1\}-1.$$

Then $P(T\geq k)=P(X_1=1,\ X_2=2,\ldots,X_k=k)=\beta_k$. It is known (Kemeny, Snell and Knapp, 1966, p. 161) that the chain is recurrent if and only if $\lim_{k\to\infty}\beta_k=0$, which is equivalent to $\sum_{i=1}^{\infty} q_i=\infty$; and that, when the chain is recurrent, it is positive recurrent if $E(T)\equiv\sum_{k=1}^{\infty}\beta_k<\infty$ and null recurrent if $\sum_{k=1}^{\infty}\beta_k=\infty$.

It is not difficult to show that the chain $\{X_k\}$ is thus (i) positive recurrent, (ii) null recurrent, or (iii) transient, if for some $a>1$, C in $(0,\infty)$ and integer $K<\infty$, we have, for all $j\geq K$,

(i) $j^{-1}+a/(j\ln j)\leq q_j$;

(ii) $C/(j\ln j)\leq q_j\leq j^{-1}+1/(j\ln j)$;

(iii) $q_j\leq C/[j(\ln j)^a]$.

To analyze the null recurrent case using Theorem 3, one can establish, using elementary but long arguments, that if $1/j\leq q_j\leq 1/j+1/(j\ln j)$, then the chain $\{X_k\}$ is null recurrent with $\zeta=\infty$ w.p. 1, while if, for some $a<0$ and $c>0$, $c/(j\ln j)\leq q_j\leq 1/j+a/(j\ln j)$, then $\{X_k\}$ is null recurrent with $\zeta<\infty$ w.p. 1. Thus if $q_j=1/j+a/(j\ln j)$, $a\leq 0$, the chain is null recurrent; $\zeta=\infty$ w.p. 1 if $a=0$ but $\zeta<\infty$ w.p. 1 if $a<0$.

4. Consensus: Hardening positions

In the model of consensus, suppose, for a very simple example, that on the kth round each expert gives his or her own opinion a weight of $1-k^{-1-\varepsilon}$ and the opinion of every other expert a weight of $k^{-1-\varepsilon}/(n-1)$, where ε is a nonnegative-valued random variable that may depend on k. Let p_k be the probability that $\varepsilon=0$. Then consensus will be approached, in spite of the hardening of positions, if p_k is a positive constant for all k or at least does not decrease too rapidly with increasing k. Our theorems give a precise meaning to the phrase "too rapidly."

More generally, suppose that the weights that an expert attaches to the opinions of other experts are on the kth round uniformly bounded below by $c_1 k^{-X_k}$ and above by $c_2 k^{-X_k}$, with $0 < c_1 < c_2 < \infty$. Here X_k is a random variable characterizing the environment, mood or climate of the experts and of the estimation process. Low values of X_k might reflect amiability among the experts, high values hostility. The behavior of $\{X_k\}$ assumed in the Markov chain of Section 3 might describe an initial "honeymoon," followed by alternating gradual freezes and abruptly renewed thaws. As time k increases, for a given environmental condition X_k, the upper and lower bounds on the weights attached to other experts' estimates gradually decrease, reflecting a hardening of positions. Within these bounds, the actual weight may be complicated functions of the different information available to each expert, of the conflicting interests they serve, of their own prior histories, etc. Theorem 1 considered in conjunction with the examples of Section 3 shows that the line between converging to consensus or not may be remarkably delicate, and the long-run difference may be remarkably sharp.

Acknowledgments

J.E.C. acknowledges fellowships from the John Simon Guggenheim Memorial Foundation, New York, the Center for Advanced Study in the Behavioral Sciences, Stanford, and the John D. and Catherine T. MacArthur Foundation, Chicago; National Science Foundation grants DEB80-11026 and BSR84-07461; and grants to the Center for Advanced Study from the National Institute of Mental Health (5T32MH14581-06) and the Exxon Education Foundation. C.M.N. acknowledges N.S.F. grant MCS80-19384. We are grateful for the help of Robert Axelrod, Morris DeGroot, Mr. and Mrs. William T. Golden and referees.

References

R.L. Berger, A necessary and sufficient condition for reaching a consensus using DeGroot's method, Journal of the Americal Statistical Association 76 (1981) 415-418.

S. Chatterjee and E. Seneta, Towards consensus; some convergence theorems on repeated averaging, Journal of Applied Probability 14 (1977) 89-97.

J.E. Cohen, Convexity properties of products of random nonnegative matrices, Proceedings of the National Academy of Sciences 77 (1980) 3749-3752.

M.H. DeGroot, Reaching a consensus, Journal of the American Statistical Association 69 (1974) 118-121.

C. Genest and J.V. Zidek, Combining probability distributions: A critique and an annotated bibliography, Statistical Science 1 (1986) 114-148.

I.J. Goldsheid, Asymptotic properties of the products of random matrices depending on a parameter, in: R.L. Dobrushin and Ya.G. Sinai, eds., Multicomponent Random Systems (Marcel Dekker, New York, 1980) pp. 239-325.

J. Hajnal, Weak ergodicity in non-homogeneous Markov chains, Proceedings of the Cambridge Philosophical Society 54 (1958) 233-246.

D.L. Isaacson and R.W. Madsen, Markov Chains: Theory and Applications (John Wiley, New York, 1976).

N.L. Johnson and S. Kotz, Continuous Univariate Distributions—1 (Houghton Mifflin, Boston; John Wiley, New York, 1970).

J.G. Kemeny, J.L. Snell and A.W. Knapp, Denumerable Markov Chains (Van Nostrand, Princeton, 1966).

Y. Kifer, Perturbations of random matrix products, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 61 (1982) 83–95.

J.F.C. Kingman, Subadditive processes, in: P.L. Hennequin, ed., Ecole d'Eté de Probabilités de Saint-Flour V - 1976, Lecture Notes in Mathematics 539 (Springer-Verlag, New York, 1976) pp. 168–223.

M. Loeve, Probability Theory vol. 1, 4th ed. (Springer-Verlag, New York, 1977).

E. Seneta, *Non-negative Matrices and Markov Chains*, 2d ed. (Springer-Verlag, New York, 1981).

C. Wagner and K. Lehrer, Rational Consensus in Science and Society: A Philosophical and Mathematical Study (Reidel, Boston, 1981).

J.V. Zidek, Multibayesianity: (i) Consensus of opinion, Lecture notes for the Board of Studies (University of London, May 3, 1983).