Ergodicity of Age Structure in Populations with Markovian Vital Rates, I: Countable States

JOEL E. COHEN*

This paper establishes new ergodic theorems for population age structure. Let $A_1, A_2, \ldots$ be denumerable Leslie matrices, and for $\ell = 1, 2$, $\mathbf{m}^{(\ell)}(0)$ be age structures (vectors with elements $m_i^{(\ell)}(0)$), satisfying the assumptions of the Coale-Lopez theorem. Let $\{A(t, \omega_\ell)\}^\infty_{t=1}$ be sample paths of a discrete-time Markov chain with sample space $\{A_k\}^\infty_{k=1}$, and $\mathbf{m}^{(\ell)}(t) = A(t, \omega_\ell)A(t-1, \omega_\ell) \cdots A(1, \omega_\ell)\mathbf{m}^{(\ell)}(0)$. Then for $b = 1, 2, \ldots$ (*weak stochastic ergodicity*) $\lim_{t\to\infty} (E(m_j^{(1)}(t)/m_i^{(1)}(t))^b - E(m_j^{(2)}(t)/m_i^{(2)}(t))^b) = 0$ if the chain is finite and weakly ergodic (see [4]) or denumerable and weakly ergodic (see [13]). The limit holds and (*strong stochastic ergodicity*) $\{\mathbf{m}^{(\ell)}(t)\}^\infty_{t=1}$ converge in distribution, if the chain is homogeneous, aperiodic, positive recurrent, and uniformly geometrically ergodic.

1. INTRODUCTION

The ergodic theorems of demography establish that the present age structure of a unisexual, closed population is independent of the population's age structure in the sufficiently remote past, but depends entirely on the recent history of vital (birth and death) rates.

The new ergodic theorems in this paper are based on the following model (from which, for the moment, the technical details are omitted). Suppose that at each instant in discrete time, the array of age-specific birth and death rates to which a unisexual closed population is subject is drawn from a set S of such arrays. Given two initial populations, and two initial arrays of vital rates from S, suppose that the vital rates in the next instant of time are chosen from S according to a Markov chain independently for each of the two populations. It will be proved that the difference between corresponding moments of the age structures of the two populations vanishes.[1]

This theorem is stronger than existing ergodic theorems for the age structure of populations in that the histories of vital rates of the two populations are independent sample paths of a stochastic process which does not ignore its own recent history.

In a paper neglected by demographers, Furstenberg and Kesten [3] study a similar model, although their results are restricted to stationary processes. Their bounded positive matrices may be identified with block-products of Leslie matrices (arrays of vital rates which are defined later).

Smith and Wilkinson [16] and Athreya and Karlin [1] analyze branching processes in which the probability generating function of the offspring distribution is a random variable subject to influence by its prior history. In multitype branching processes, the mean value matrix is equivalent to the Leslie matrix [14]. However, only in [1] are multitype processes mentioned, and in that case only extinction appears to be studied.

In work which has appeared since this article was originally submitted, Le Bras [10], continuing earlier heuristic analysis [9], identifies states of a finite irreducible stationary Markov chain with arrays of vital rates. His conclusions seem related to those of both [3] and this article.

This paper proves the mutual convergence of moments of age structure when the arrays of vital rates are chosen from a denumerable (not necessarily finite) set S of such arrays, according to a Markov chain which is not necessarily stationary or homogeneous in time, regardless of the initial age structure or initial array of vital rates in the population.

Section 2 states some necessary known Theorems 1 and 2, and the new Theorems 3 and 4, and proposes extensions of the new theorems. Section 3 proves the new theorems.[2]

Because of limited space, the demographic and biological motivation for these theorems cannot be presented in detail here. The empirical stimulus for the present model is a dissatisfaction with assuming fixed vital rates, deterministically changing vital rates (periodic or otherwise), or stochastic vital rates which, at every time t, are independent of the vital rates at all times before t. *A priori* a Markovian model of age-specific vital rates seems a step toward reality.

The plausibility and utility of the Markovian assumption can be tested, because the model suggests a scheme for incorporating historical human data into a new

* Joel E. Cohen is professor, The Rockefeller University, New York, NY 10021. This research was supported in part by Harvard University, The Population Council, The National Science Foundation, and King's College Research Centre, Cambridge, Eng. The author wishes to thank Ansley J. Coale, Lloyd Demetrius, John Hajnal, Louis Henry, Henry Horn, Norman Kaplan, Niels Keiding, Harry Kesten, Nathan Keyfitz, Hervé Le Bras, Ronald Lee, Herman Leonard, Robert M. May, David D. McFarland, Don R. McNeil, Robert K. Merton, Jr., John H. Pollard, Michael Rothschild, Paul A. Samuelson, Roger Schofield, Zenas M. Sykes and Geoffrey S. Watson for helpful comments.

[1] The age structures also converge in distribution or in law [8] if the Markov chain is homogeneous. A proof of this, and an explicit means of finding the limiting distribution, will appear subsequently.

[2] A future paper will extend these results to uncountably many arrays of vital rates and provide a numerical illustration.

© Journal of the American Statistical Association
June 1976, Volume 71, Number 354
Theory and Methods Section

method of population projection. Arrange all the age-specific effective fertility and survival coefficients customarily used in age-specific population projections into a vector. Fit a linear first-order autoregressive scheme to a historically observed sequence of such vectors. (Such a scheme, which does require an uncountable state space of vectors of vital rates, is merely a simple particular specification of the Markovian model, not the only one possible.) Use the estimated parameters and an initial array of vital rates to project a distribution of arrays of future vital rates. Given an initial age structure, this distribution of future vital rates implies a distribution of projected subsequent age structures and population sizes. The empirical merit of this scheme, or of other possible parametric specifications of the Markovian model, remains to be determined.

Other applications in view include the study of the use of different habitats (with corresponding different arrays of vital rates) by mobile animal populations; and the comparison of age structures of biological populations on islands which are subject to similar patterns of weather but not necessarily identical sequences of weather conditions (with, again, corresponding sequences of vital rates).

2. THEOREMS, OLD AND NEW

Closed, unisexual populations are observed at discrete times $t = 0, 1, \ldots$. Individuals are identified by their age at last birthday, starting from age 0 and stopping at age $K \geq 1$. The age structure of a population is given by a $(K+1)$-vector of nonnegative real numbers in which the jth element is the (not necessarily integral) number of individuals of age j at last birthday, $j = 0, \ldots, K$.

Define C to be the (uncountable) set of all $(K+1) \times (K+1)$ nonnegative real matrices of the (Leslie) form

$$A = \begin{bmatrix} f_0 & f_1 & \ldots & f_{K-1} & f_K \\ s_0 & 0 & \ldots & 0 & 0 \\ 0 & s_1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & s_{K-1} & 0 \end{bmatrix} \quad (2.1)$$

where all elements other than the first row and the subdiagonal are zero, subject to the following conditions.

i. There is some number $\epsilon > 0$, and there are at least two distinct integers i_0, j_0, where $0 \leq i_0, j_0 \leq K$, such that the greatest common divisor of $i_0 + 1$ and $j_0 + 1$ is one and for every matrix A in C, $f_{i_0} > \epsilon$ and $f_{i_0} > \epsilon$.
ii. For all A in C, $\epsilon < s_j \leq 1$ for all $j = 0, \ldots, K-1$.
iii. There is some $M' > \epsilon > 0$ such that for all A in C and for all j, $0 \leq j \leq K$, $f_j < M' < \infty$.

In (2.1), f_j is interpreted as the "effective" fertility of individuals in the jth age class, that is, the number born to individuals aged j who survive to the start of the next time interval. s_j is the proportion of individuals aged j at the start of a time interval who survive to become aged $j+1$ at the start of the next time interval.

Lopez [11] proves the following theorem of weak ergodicity.

Theorem 1: Let $\mathbf{m}(0)$ and $\mathbf{n}(0)$ be two nonnegative $(K+1)$-vectors with finite elements $m_j(0)$ and $n_j(0)$, $0 \leq j \leq K$, such that for at least some j' and some j'', $0 \leq j', j'' \leq \max(i_0, j_0)$, $m_{j'}(0) > 0$ and $n_{j''}(0) > 0$. (These vectors $\mathbf{m}(0)$ and $\mathbf{n}(0)$ are the two initial population vectors.) If, for $t \geq 1$,

$$\begin{aligned} \mathbf{m}(t) &= A(t)A(t-1) \cdots A(1)\mathbf{m}(0) , \\ \mathbf{n}(t) &= A(t)A(t-1) \cdots A(1)\mathbf{n}(0) , \end{aligned} \quad (2.2)$$

where $A(\theta)$, $\theta = 1, \ldots, t$ are any elements of C, then for all i, j, $0 \leq i, j \leq K$,

$$\lim_{t\to\infty} ([m_j(t)/n_j(t)] - [m_i(t)/n_i(t)]) = 0 . \quad (2.3)$$

Corollary 1: Let $\mathbf{e}_j$ be a $(K+1)$-vector with elements δ_{jk}, $k = 0, 1, \ldots, K$, where $\delta_{jk} = 1$ if $j = k$ $\delta_{jk} = 0$ if $j \neq k$. (Thus, $\mathbf{e}_j^T \cdot \mathbf{m}(t) = m_j(t)$.) Let $\mathbf{m}(0)$ and $\mathbf{n}(0)$ satisfy the assumptions of Theorem 1. Let t be any time greater than a positive constant n (calculated from K, i_0, j_0 in [11]). Let $A(\theta)$, $B(\theta)$, $\theta = 1, \ldots, t$, be any elements from C. Then for any $\delta > 0$, however small, there exists a time $s > t$ such that for any A_{j_τ}, $\tau = t+1, \ldots, s$, from C,

$$\left[\frac{\mathbf{e}_j^T A_{j_s} A_{j_{s-1}} \cdots A_{j_{t+1}} A(t) \cdots A(1)\mathbf{m}(0)}{\mathbf{e}_i^T A_{j_s} A_{j_{s-1}} \cdots A_{j_{t+1}} A(t) \cdots A(1)\mathbf{m}(0)} - \frac{\mathbf{e}_j^T A_{j_s} \cdots A_{j_{t+1}} B(t) \cdots B(1)\mathbf{n}(0)}{\mathbf{e}_i^T A_{j_s} \cdots A_{j_{t+1}} B(t) \cdots B(1)\mathbf{n}(0)}\right] < \delta . \quad (2.4)$$

Proof of Corollary 1: From the algebraic identity

$$\left[\frac{m_j(s)}{n_j(s)} - \frac{m_i(s)}{n_i(s)}\right] = \left(\frac{m_j(s)}{n_j(s)}\right)\left(\frac{n_j(s)}{n_i(s)}\right)\left[\frac{n_i(s)}{n_j(s)} - \frac{m_i(s)}{m_j(s)}\right], \quad (2.5)$$

a necessary and sufficient condition for the vanishing of either quantity in square brackets to imply the vanishing of the other, as $s \to \infty$, is that for any fixed $t < s$ there exist constants α_t, β_t, such that for s large enough,

$$0 < \alpha_t \leq (m_j(s)/n_j(s))(n_j(s)/n_i(s)) \leq \beta_t < \infty . \quad (2.6)$$

From Lopez's proof of Theorem 1 [11, p. 55], if $\mathbf{m}(0)$ and $\mathbf{n}(0)$ satisfy the boundedness and positivity assumptions of the theorem, so also will any $\mathbf{m}(t) = A(t) \cdots A(1)\mathbf{m}(0)$ and any $\mathbf{n}(t) = B(t) \cdots B(1)\mathbf{n}(0)$. After time $n (= N_0$ in Pollard's notation [15]), all components of $\mathbf{m}(t)$ and $\mathbf{n}(t)$ are positive, so the ratio $m_j(t)/n_j(t)$ is defined for all j, $0 \leq j \leq K$. Then for any $t > n$, $m_j(t)/n_j(t)$ can never become either zero or infinite.

So at any time $t > n$, for all j, there exist numbers $r_0(t)$, $R_0(t)$ such that $0 < r_0(t) \leq m_j(t)/n_j(t) \leq R_0(t) < \infty$. From time $t+1$ onward, the same sequence of matrices is applied to both $\mathbf{m}(t)$ and $\mathbf{n}(t)$. Lopez's proof of Theorem 1 implies, then, that for all $s > t$,

$$0 < r_0(t) \leq m_j(s)/n_j(s) \leq R_0(t) < \infty , \quad j = 0, \ldots, K . \quad (2.7)$$

From [11, p. 56], at any time $s > n$,

$$0 < R^{-1} \leq n_j(t)/n_i(t) \leq R < \infty \ , \quad i, j = 0, \ldots, K \ . \tag{2.8}$$

Taking $\alpha_t = r_0(t)/R$ and $\beta_t = R_0(t)R$ gives (2.6). Theorem 1 guarantees that the left side of (2.5) vanishes as $s - t \rightarrow \infty$; hence, the corollary.

Theorem 2 (from [19]): Let S be a denumerable state space with at least two distinct elements and $I = \{1, 2, \ldots\}$ the set of indices of elements of S. Suppose an irreducible, aperiodic, positive recurrent geometrically ergodic Markov chain on S has time-homogeneous one-step transition probabilities $\mathbf{P} = \{P_{ij}\}_{i,j \in I}$, where P_{ij} is the conditional probability that the state of the chain at time $t + 1$ will be the state indexed by j, given that the state of the chain at time t is the state indexed by i. Then there exists a unique sequence $\{\pi_j, j \in I\}$ satisfying

$$\begin{aligned} \pi_j > 0 \ , \quad j \in I \ , \\ \sum_{j \in I} \pi_j = 1 \ , \\ \pi_j = \sum_{i \in I} \pi_i P_{ij} \ , \quad \text{all} \quad j \in I \ , \end{aligned} \tag{2.9}$$

$$|P_{ij}^{(t)} - \pi_j| \leqslant M_{ij}\rho^t \ , \quad i, j \in I, M_{ij} < \infty, \rho < 1 \ ,$$

where $P_{ij}^{(t)}$ is the t-step transition probability from i to j, $P_{ij}^{(1)} = P_{ij}$, and ρ is independent of i, j.

Such a chain will be called uniformly geometrically ergodic if $M_{ij} \leq M < \infty$, $i, j \in I$. Every time-homogeneous finite irreducible aperiodic Markov chain is uniformly geometrically ergodic.

Theorem 3 (strong stochastic ergodicity): Let S be a denumerable subset, containing at least two distinct elements, of the set C of Leslie matrices defined before Theorem 1. Let I be the set of indices of the elements $\{A_1, A_2, \ldots\}$ in S. Let $\mathbf{m}(0)$ and $\mathbf{n}(0)$ satisfy the hypotheses of Theorem 1. Suppose

$$\begin{aligned} \mathbf{m}(t) = A(t)A(t-1) \cdots A(2)A_1\mathbf{m}(0) \ , \\ \mathbf{n}(t) = B(t)B(t-1) \cdots B(2)A_2\mathbf{n}(0) \ , \end{aligned} \tag{2.10}$$

where the sequence $A(2), A(3), \ldots$ of matrices from S is determined by a time-homogeneous irreducible aperiodic positive recurrent uniformly geometrically ergodic Markov chain on the state space S starting from state A_1, and the sequence $B(2), B(3), \ldots$ of matrices from S is independently determined by the same Markov chain starting from the state A_2. More precisely, for all k, ℓ in I, $t = 1, 2, \ldots$,

$$\begin{aligned} P[A(t+1) = A_\ell \,|A(t) = A_k] = P_{k\ell} \ , \\ A(1) = A_1 \text{ with probability } 1 \ ; \\ P[B(t+1) = A_\ell \,|B(t) = A_k] = P_{k\ell} \ , \\ B(1) = A_2 \text{ with probability } 1 \end{aligned} \tag{2.11}$$

Then for every integer $b \geq 1$, $0 \leq i, j < K$,

$$\lim_{t \to \infty} (E(m_j(t)/m_i(t))^b - E(n_j(t)/n_i(t))^b) = 0 \tag{2.12}$$

and $m_j(t)/m_i(t)$ and $n_j(t)/n_i(t)$ converge in distribution as $t \rightarrow \infty$.

Define a transition matrix P of a denumerable-state Markov chain as regular if it is aperiodic and irreducible. Suppose a Markov chain has the successive regular transition matrices $P(1)$, $P(2)$, ... Let $H(t) = P(1), P(2) \cdots P(t)$. Let I be the index set of states. The chain is weakly ergodic in the sense of [4] if, for every i, j, ℓ in I, and for every $\epsilon > 0$, there exists t such that

$$|h_{ij}(t) - h_{\ell j}(t)| < \epsilon \tag{2.13}$$

where $h_{ij}(t)$ is the i, jth element of $H(t)$. For such a chain $\lim_{t \to \infty} h_{ij}(t)$ may not exist. The chain is weakly ergodic in the sense of [13] if, for every choice for the origin of time and for every $\epsilon > 0$, there exists t such that for every i, j, ℓ in I, (2.13) holds.

Theorem 4 (weak stochastic ergodicity): Let S be a subset of the set C of Leslie matrices defined before Theorem 1.

Let $\mathbf{m}(t)$, $\mathbf{n}(t)$ be defined as in (2.10). Here, the sequence $A(2), A(3), \ldots$ of matrices from S is determined by a weakly ergodic inhomogeneous Markov chain with regular transition matrices $P(1), P(2), \ldots$ and state space S, starting from state A_1. The sequence $B(2)$, $B(3), \ldots$ of matrices from S is independently determined by the same Markov chain starting from the state A_2. More precisely, for all k, ℓ in the index set I, $t = 1, 2, \ldots$,

$$\begin{aligned} P[A(t+1) = A_\ell \,|A(t) = A_k] = P_{k\ell}(t) \ , \\ A(1) = A_1 \text{ with probability } 1 \ ; \\ P[B(t+1) = A_\ell \,|B(t) = A_k] = P_{k\ell}(t) \ , \\ B(1) = A_2 \text{ with probability } 1 \end{aligned} \tag{2.14}$$

i. If the chain is weakly ergodic in the sense of Hajnal, assume the number $N \geqslant 2$ of states in S is finite.

ii. If the chain is weakly ergodic in the sense of Paz, assume the the number of states in S is denumerable.

Under either Assumption i or ii, for every integer $b \geq 1$, (2.12) holds.

Neither Theorem 3 nor Theorem 4 asserts that the convergence (2.12) is uniform over the order b of the moment. Only (2.12) will be proved here. After seeing these results and a draft of the paper cited in Footnotes 1 and 2, Norman Kaplan (via personal communication, July 1975) proved concisely the convergence in distribution asserted in Theorem 3.

Generalizations of Theorems 3 and 4 may affect either the operator, which is here narrowly specified as a Leslie matrix, or the stochastic process governing the choice of operator, here, particular kinds of Markov chains.

Theorems 3 and 4 obviously apply, by [3, Lemma 3.3], if Leslie matrices are replaced by strictly positive matrices bounded above and bounded away from zero. The theorems probably apply if Leslie matrices are replaced by certain classes of power-positive matrices, positive linear operators satisfying certain contraction conditions (see [6, 7]), nonlinear operators of Solow-Samuelson type [17], or certain Banach space operators [2]. Theorems equivalent to Theorems 3 and 4 probably apply (conditional on nonextinction) if the Leslie or strictly positive matrices are interpreted as the expected value matrices of multitype branching processes in random environments (see [1, 16]), or if the vital rates of Leslie matrices are viewed as parameters of Bernoulli trials (see [12, 14]).

Theorems 3 and 4 obviously apply if the order of dependency in the Markov chain is increased from first to any fixed order.

The theorems probably apply if the Markov chain is replaced by an ergodic Markov chain in which the transition matrix is itself an iid random variable (see [18]), or indeed by any strictly stationary metrically transitive discrete-time stochastic process.

It may also be possible to relax the assumption that the stochastic process generating the sequence of operators applied to one population is identical to the stochastic process for the other population. Hajnal [5] shows how close the transition matrices of two Markov chains have to be for them to have the same type of long-run behavior. Such conditions may suffice to guarantee, for a homogeneous process, convergence in distribution of the age structures.

3. PROOFS OF THEOREM 3 AND THEOREM 4

Proof of Theorem 3: Let s, t, $s > t > 1$, be the two time periods, and b be any positive integer. Then, under the assumptions of Theorem 3, the following moments exist by (2.8) and are equal to

$$E(m_j(s)/m_i(s))^b = \sum_{j_2\in I}\sum_{j_3\in I}\cdots\sum_{j_t\in I}\cdots\sum_{j_s\in I} P_{1j_2}P_{j_2j_3}\cdots P_{j_{t-1}j_t}\cdots P_{j_{s-1}j_s}\left[\frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_t}\cdots A_{j_2}A_1\mathbf{m}(0)}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_t}\cdots A_{j_2}A_1\mathbf{m}(0)}\right]^b \quad (3.1)$$

and $E[(n_j(s)/n_i(s))^b]$ is identical to (3.1), save that $\mathbf{n}(0)$ replaces $\mathbf{m}(0)$, A_2 replaces A_1, and P_{2j_2} replaces P_{1j_2}. Let $\delta > 0$ be an arbitrary small positive number, and let $\mathbf{m}^*(t) = A_1{}^t\mathbf{m}(0)$, taking t as fixed for the moment. Now choose $s - t$ to be large enough to satisfy the following two conditions. (The possibility of finding such a value of $s - t$, given any δ, is guaranteed by Corollary 1.) First, for every $\mathbf{m}(t) = A_{j_t}\cdots A_{j_2}\mathbf{m}(0)$, there exists a number $\delta_1(j_t, \ldots, j_2)$ such that $|\delta_1(j_t, \ldots, j_2)| < \delta$ and

$$\frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}(t)}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}(t)} = \frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)} + \delta_1(j_t, \ldots, j_2) . \quad (3.2)$$

Second, for every $\mathbf{n}(t) = A_{j_t}\cdots A_{j_2}A_2\mathbf{n}(0)$, there exists a number $\delta_2(j_t, \ldots, j_2)$ such that $|\delta_2(j_t, \ldots, j_2)| < \delta$ and

$$\frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{n}(t)}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{n}(t)} = \frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)} + \delta_2(j_t, \ldots, j_2) . \quad (3.3)$$

Now by the binomial theorem, for any finite integer b, $(x+\delta)^b = x^b + \delta C$, where, if x and δ are uniformly bounded above, so is C. Hence, (3.1) may be rewritten, and (3.2) and the binomial theorem may be used to obtain

$$\begin{aligned}
E(m_j(s)/m_i(s))^b &= \sum_{j_2}\cdots\sum_{j_t} P_{1j_2}\cdots P_{j_{t-1}j_t}\cdot\left(\sum_{j_{t+1}}\cdots\sum_{j_s} P_{j_tj_{t+1}}\cdots P_{j_{s-1}j_s}\cdot\left[\frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_{t+1}}(A_{j_t}\cdots A_{j_2}A_1\mathbf{m}(0))}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_{t+1}}(A_{j_t}\cdots A_{j_2}A_1\mathbf{m}(0))}\right]^b\right)\\
&= \sum_{j_2}\cdots\sum_{j_t} P_{1j_2}\cdots P_{j_{t-1}j_t}\cdot\left(\sum_{j_{t+1}}\cdots\sum_{j_s} P_{j_tj_{t+1}}\cdots P_{j_{s-1}j_s}\cdot\left[\frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)} + \delta_1(j_t, \ldots, j_2)\right]^b\right)\\
&= \sum_{j_2}\cdots\sum_{j_t} P_{1j_2}\cdots P_{j_{t-1}j_t}\cdot\left(\sum_{j_{t+1}}\cdots\sum_{j_s} P_{j_tj_{t+1}}\cdots P_{j_{s-1}j_s}\cdot\left[\frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)}\right]^b\right)\\
&\quad + \sum_{j_2}\cdots\sum_{j_s} P_{1j_2}\cdots P_{j_{s-1}j_s}\delta_1(j_t, \ldots, j_2)\cdot B_1(j_s, \ldots, j_2)\\
&= \sum_{j_t} P_{1j_t}{}^{(t-1)}\left(\sum_{j_{t+1}}\cdots\sum_{j_s} P_{j_tj_{t+1}}\cdots P_{j_{s-1}j_s}\cdot\left[\frac{\mathbf{e}_j{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^TA_{j_s}\cdots A_{j_{t+1}}\mathbf{m}^*(t)}\right]^b\right)\\
&\quad + \sum_{j_2}\cdots\sum_{j_s} P_{1j_2}\cdots P_{j_{s-1}j_s}\delta_1(j_t, \ldots, j_2)\cdot B_1(j_s, \ldots, j_2) , \quad (3.4)
\end{aligned}$$

where $B_1(j_s, \ldots, j_2)$ is given explicitly in (3.8).

The last step of (3.4) follows because $P_{1j_t}{}^{(t-1)}$, the $(t-1)$-step transition probability from state 1 to state j_t, is simply the sum over all $(t-2)$-tuples $(j_2, \ldots, j_{t-1})$ of the products of one-step transition probabilities.

Now choose t large enough so that, for all j_t in I,

$$P_{\ell,j_t}{}^{(t-1)} = \pi_{j_t} + \gamma_\ell(j_t) , \quad |\gamma_\ell(j_t)| < \delta/2 , \quad \ell = 1, 2 . \quad (3.5)$$

The possibility of finding a value of t such that (3.5) holds uniformly in all j_t is assured by uniform geometric ergodicity. Using (3.5) in (3.4) gives

$$E(m_j(s)/m_i(s))^b$$

$$= \sum_{j_t} \pi_{j_t}\left(\sum_{j_{t+1}} \cdots \sum_{j_s} P_{j_t j_{t+1}} \cdots P_{j_{s-1}j_s} \cdot \left[\frac{\mathbf{e}_j{}^T A_{j_s} \cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^T A_{j_s} \cdots A_{j_{t+1}}\mathbf{m}^*(t)}\right]^b\right)$$

$$+ \gamma_1(j_t)\left(\sum_{j_{t+1}} \cdots \sum_{j_s} P_{j_t j_{t+1}} \cdots P_{j_{s-1}j_s} \cdot \left[\frac{\mathbf{e}_j{}^T A_{j_s} \cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^T A_{j_s} \cdots A_{j_{t+1}}\mathbf{m}^*(t)}\right]^b\right)$$

$$+ \sum_{j_2} \cdots \sum_{j_s} P_{1j_2} \cdots P_{j_{s-1}j_s}\delta_1(j_t, \ldots, j_2) \cdot B_1(j_s, \ldots, j_2) \ . \quad (3.6)$$

An identical argument gives an expression for $E[(n_j(s)/n_i(s))^b]$ which is the right side of (3.6) with γ_1, P_{1j_2}, δ_1 and B_1 replaced, respectively, by γ_2, P_{2j_2}, δ_2, and B_2.

Then the absolute difference is given by

$$|E(m_j(s)/m_i(s))^b - E(n_j(s)/n_i(s))^b|$$

$$= \Bigg|(\gamma_1(j_t) - \gamma_2(j_t))\left(\sum_{j_{t+1}} \cdots \sum_{j_s} P_{j_t j_{t+1}} \cdots P_{j_{s-1}j_s} \cdot \left[\frac{\mathbf{e}_j{}^T A_{j_s} \cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^T A_{j_s} \cdots A_{j_{t+1}}\mathbf{m}^*(t)}\right]^b\right) + \sum_{j_2} \cdots \sum_{j_s} [P_{1j_2} \cdots P_{j_{s-1}j_s}\delta_1(j_t, \ldots, j_2)B_1(j_s, \ldots, j_2) - P_{2j_2} \cdots P_{j_{s-1}j_s}\delta_2(j_t, \ldots, j_2)B_2(j_s, \ldots, j_2)]\Bigg|$$

$$\le \delta R^b + 2\delta B = \delta(R^b + 2B), \quad (3.7)$$

where R is the bound in (2.8) and the bound B satisfying

$$B \ge B_\ell(j_s, \ldots, j_2) = \sum_{h=1}^{b} \binom{b}{h}\left[\frac{\mathbf{e}_j{}^T A_{j_s} \cdots A_{j_{t+1}}\mathbf{m}^*(t)}{\mathbf{e}_i{}^T A_{j_s} \cdots A_{j_{t+1}}\mathbf{m}^*(t)}\right]^{b-h} \cdot [\delta_\ell(j_t, \ldots, j_2)]^{h-1} \ , \quad \ell = 1, 2 \ , \quad (3.8)$$

depends on R and b and does not increase as δ decreases. Since the last member of (3.7) can be made arbitrarily small by choice of δ, the difference between the moments in the first member of (3.7) vanishes. Thus, (2.12) is proved.

Proof of Theorem 4: The proof is a repetition of the proof of Theorem 3, except for two changes. First, for all i, j_{t+1} in I and all $t = 1, 2, \ldots, P_{ij_{t+1}}(t)$ replaces $P_{ij_{t+1}}$. Second, instead of (3.5), choose t large enough so that, for all j_t in I

$$\begin{aligned} P_{2j_t}{}^{(t-1)} &\equiv h_{2j_t}(t) \\ &= P_{1j_t}{}^{(t-1)} + \gamma(j_t) \ , \quad |\gamma(j_t)| < \delta \ ; \quad (3.9) \\ &\equiv h_{1j_t}(t) + \delta(j_t) \ . \end{aligned}$$

The possibility of finding a value of t such that (3.9) holds uniformly in all j_t is now guaranteed by (2.13) and, in Case i, the finiteness of S. From (3.6) onward, the proof is adjusted for the replacement of π_{j_t} by $h_{1j_t}(t)$ or $h_{2j_t}(t)$.

[*Received March 1974. Revised May 1975.*]

REFERENCES

[1] Athreya, K.B. and Karlin, S., "On Branching Processes with Random Environments: I. Extinction Probabilities; II. Limit Theorems," *Annals of Mathematical Statistics*, 42, Nos. 5 and 6 (1971), 1499–520; 1843–58.

[2] Demetrius, Lloyd, "Multiplicative Processes—II," *Mathematical Biosciences*, 20 (1974), 345–57.

[3] Furstenberg, H. and Kesten, H., "Products of Random Matrices," *Annals of Mathematical Statistics*, 31, No. 2 (1960), 457–69.

[4] Hajnal, John, "The Ergodic Properties of Nonhomogeneous Finite Markov Chains," *Proceedings of the Cambridge Philosophical Society*, 52, Part 1 (1956), 67–77.

[5] ———, "Weak Ergodicity in Nonhomogeneous Markov Chains," *Proceedings of the Cambridge Philosophical Society*, 54, Part 2 (1958), 233–46.

[6] Hopf, Eberhard, "An Inequality for Positive Linear Integral Operators," *Journal of Mathematics and Mechanics*, 12, No. 5 (1963), 683–92.

[7] ———, "Remarks on My Paper, 'An Inequality for Positive Linear Integral Operators,'" *Journal of Mathematics and Mechanics*, 12, No. 6 (1963), 889–92.

[8] Kendall, M.G. and Stuart, A., *Advanced Theory of Statistics, Vol. 1*, 3rd ed., London: Charles W. Griffin & Co., Ltd., 1969.

[9] Le Bras, Hervé, "Eléments pour une Théorie des Populations Instables," *Population*, 26 (May–June 1971), 525–72.

[10] ———, "Populations Stables Aléatoires," *Population*, 29 (May–June 1974), 435–64.

[11] Lopez, A., *Problems in Stable Population Theory*, Princeton, N.J.: Office of Population Research, 1961.

[12] Namkoong, G., "Persistence of Variances for Stochastic, Discrete-Time Population Growth Models," *Theoretical Population Biology*, 3, No. 4 (1972), 507–18.

[13] Paz, A., *Introduction to Probabilistic Automata*, New York: Academic Press, Inc., 1971.

[14] Pollard, John H., "On the Use of the Direct Matrix Product in Analyzing Certain Stochastic Population Models," *Biometrika*, 53, Nos. 3 and 4 (1966), 397–415.

[15] ———, *Mathematical Models for the Growth of Human Populations*, London: Cambridge University Press, 1973.

[16] Smith, Walter L. and Wilkinson, William E., "Branching Processes in Markovian Environments," *Duke Mathematical Journal*, 38 (1971), 749–63.

[17] Solow, Robert M. and Samuelson, Paul A., "Balanced Growth Under Constant Returns to Scale," *Econometrica*, 21 (July 1953), 412–24.

[18] Takahashi, Y., "Markov Chains with Random Transition Matrices," *Kodai Mathematical Seminar Reports*, 21 (1969), 426–47.

[19] Vere-Jones, D., "Geometric Ergodicity in Denumerable Markov Chains," *Quarterly Journal of Mathematics, Ser. 2*, 13 (1962), 7–28.