Adv. Appl. Prob. **9**, 18–37 (1977)
Printed in Israel

ERGODICITY OF AGE STRUCTURE IN POPULATIONS WITH MARKOVIAN VITAL RATES. II. GENERAL STATES

JOEL E. COHEN, *The Rockefeller University, New York*

Abstract

The age structure of a large, unisexual, closed population is described here by a vector of the proportions in each age class. Non-negative matrices of age-specific birth and death rates, called Leslie matrices, map the age structure at one point in discrete time into the age structure at the next. If the sequence of Leslie matrices applied to a population is a sample path of an ergodic Markov chain, then: (i) the joint process consisting of the age structure vector and the Leslie matrix which produced that age structure is a Markov chain with explicit transition function; (ii) the joint distribution of age structure and Leslie matrix becomes independent of initial age structure and of the initial distribution of the Leslie matrix after a long time; (iii) when the Markov chain governing the Leslie matrix is homogeneous, the joint distribution in (ii) approaches a limit which may be easily calculated as the solution of a renewal equation. A numerical example will be given in Cohen (1977).

AGE STRUCTURE; POPULATION DYNAMICS; ERGODIC THEOREMS OF DEMOGRAPHY; PRODUCTS OF RANDOM MATRICES; MULTITYPE PROCESSES IN RANDOM ENVIRONMENTS

1. Results and example

1.1. *Summary.* In human and other biological populations, the numbers of births and deaths sometimes fluctuate more than would be predicted by binomial sampling from processes with fixed underlying vital rates. A central problem of demography and of general population biology is to find intuitively reasonable, mathematically tractable and empirically successful models for observed and future variations in underlying vital rates.

The model studied in Cohen ((1976); hereafter referred to as Part I) and in this Part II is a three-tiered structure.

At the lowest level is a sequence $\{y_n\}$ of age structures of a closed, unisexual population observed at discrete instants of time $n = 0, 1, \cdots$. Since attention focuses here on age *structure*, these vectors y_n are normalized so that at each n their elements sum to 1.

Received in revised form 7 June 1976.

At the middle level is a sequence of operators $\{x_n\}$ representing the action of age-specific vital rates on age structure; x_n maps y_{n-1} into y_n, $n = 1, 2, \cdots$.

At the highest level is a model for passing from x_{n-1} to x_n. In the classical theory of stable populations, the model at the highest level is an identity operator: $x_n = x_{n-1}$. Under that model, for certain choices of y_0 and x_1, y_n approaches the unique limiting stable age distribution determined by x_1 but not by y_0. In the weak ergodic theorem of Lopez (see Golubitsky, Keeler and Rothschild (1975)), the model at the highest level is an arbitrary determinate sequence $\{x_n\}$ in which the elements are chosen from a set X of possible operators. Under that model, for certain choices of y_0 and X, y_n approaches a possibly time-varying sequence which depends entirely on the sequence $\{x_n\}$ and not on y_0.

In both Parts I and II, the model for passing from x_{n-1} to x_n is a Markov chain, with certain ergodic and other properties, on a state space (or sample space) X of possible operators.

In Part II, the behavior of the age structures $\{y_n\}$ is described by four theorems. They are stated formally in Section 1.2. The corollaries deal with cases of practical interest in demography and Markov chains.

Theorem 1 observes that, assuming smoothly behaved transition functions in the Markov chain on X and smoothly behaved operators x in X, the joint process (x_n, y_n) of operators and points (vital rates and age structures) is a Markov chain (although $\{y_n\}$ by itself is not in general a Markov chain). The transition function of this bivariate chain is written out explicitly in terms of the transition function governing $\{x_n\}$.

Theorem 2 says that under additional smoothness and ergodic conditions on the chain on X, and assuming the abstract equivalent of the weak ergodic theorem of demography, the bivariate chain $\{(x_n, y_n)\}$ also has ergodic features. The idea of the proof is simply to divide a long period from time n to time $n + m$ into two long periods. The first, from n to $n + r$, is long enough for the chain on X to forget its past. The second, from $n + r$ to $n + m$, is long enough for contractions on Y to obliterate the effects of the values of y at time $n + r$ and, *a fortiori*, at time n.

Theorem 3 supposes that the chain on X in Theorem 2 is homogeneous. Then the bivariate chain converges in distribution to an invariant long-run distribution which may be calculated explicitly by solving a linear integral equation.

Theorem 4 says that under even stronger conditions on the smoothness of action of the operators from X, the bivariate chain $\{(x_n, y_n)\}$ satisfies a stronger ergodic condition; when the chain $\{X_n\}$ of operators is homogeneous, the rate of convergence in distribution of the bivariate chain $\{(x_n, y_n)\}$ is exponential.

Corollary 1 suggests a possible model for estimation of the transition function of the chain on X from historical data on vital rates.

Corollary 2 applies Theorems 1 to 3 to finite Markov chains with random transition matrices. Each transition matrix is required to fall in a uniformly bounded class of scrambling matrices.

Corollary 3 rests on the concept, due to Hajnal (1976), of an ergodic set of matrices. The corollary shows that the same contractive property of the operators and of the stochastic process determining the choice of successive operators assures the results of Theorems 2 and 3. This corollary permits the operators to belong to a finite class of non-negative matrices that is so general that applications of the theorem in genetics and economics become obvious.

1.2. *Setting, definitions and results.* Let N be the set of natural numbers $\{1,2,\cdots\}$, R the set of all finite real numbers $(-\infty, +\infty)$, and P the set of non-negative finite reals $[0,\infty)$. If S and S' are any sets, S^c is the complement of S, 2^S is the family of all subsets of S, and $S\triangle S' = (S \cup S') \cap (S \cap S')^c$ is the symmetric difference of S and S'. Used with sets, $+$ means disjoint union; thus $S + S'$ means $S \cup S'$, and moreover $S \cap S' = \emptyset$. $\mathrm{Lim}_n(\cdot)$ means the limit of $(\cdot)$ as $n \to \infty$, n in N unless otherwise indicated. If $\mathcal{S}$ is a family of sets, $\sigma(\mathcal{S})$ is the minimal σ-field generated by $\mathcal{S}$, $\mathcal{R}(\mathcal{S})$ the ring generated by $\mathcal{S}$. If x is a $k \times k$ real matrix, k in N, $x = (x(i,j))$, define $\| x \|_X = \Sigma_{i,j=1}^k | x(i,j)|$. If y is a k-vector (a column), $y = (y(i))$, define $\| y \| = \Sigma_{i=1}^k | y(i)|$.

The elements of the set $Y(X)$ will be denoted by $y(x)$ with or without affixes, e.g. y', y_1, y_n (x', x_1, x_n), and similarly for sets A (B) belonging to the field $\mathcal{A}(\mathcal{B})$. Elements of Z will sometimes be denoted z and sometimes (x,y) with corresponding affixes if any. Thus z' is the same as (x', y') without further comment.

The transpose of a vector is indicated by a suffix $^{\mathrm{Tr}}$.

Let $(\Omega, \mathcal{F}, \boldsymbol{P})$ be a probability space, and $\{X_n\}$ a sequence for n in N of measurable functions from Ω into a measurable space $(X, \mathcal{A})$ where X is a set and $\mathcal{A}$ is a σ-field of subsets of X. For every A in $\mathcal{A}$ and x in X, let $Q_n(A) = \boldsymbol{P}\{\omega \in \Omega : X_n(\omega) \in A\} \equiv \boldsymbol{P}[X_n \in A]$ and let $P_n^m(x, A) = \boldsymbol{P}[X_{n+m} \in A \mid X_n = x]$. Assume $\{X_n\}$ form a Markov chain. If $P_n^1 = P_1^1$ for all n in N, the chain is homogeneous. We sometimes abbreviate $P_n^1 = P_n$.

Definition 1. The chain $\{X_n\}$ is weakly ergodic if and only if, for every n, for every $\varepsilon > 0$, for every A in $\mathcal{A}$ and for every x, x' in X, there exists m_0 such that for all $m \geqq m_0$, $| P_n^m(x, A) - P_n^m(x', A)| < \varepsilon$.

Definition 2. The chain $\{X_n\}$ is uniformly weakly ergodic if and only if, for every n, for every $\varepsilon > 0$, and for A in $\mathcal{A}$, there exists m_0 such that for all $m \geqq m_0$, $\sup_{x,x' \in X} | P_n^m(x, A) - P_n^m(x', A)| < \varepsilon$.

Definition 3. The chain $\{X_n\}$ is S-ergodic if and only if, for every n and for every $\varepsilon > 0$, there exists m_0 such that for all $m \geqq m_0$,

$$\Delta_{nm}(\{X_n\}) \equiv \sup_{x, x' \in X} \sup_{A \in \mathscr{A}} |P_n^m(x, A) - P_n^m(x', A)| < \varepsilon.$$

This condition is identical to Griffeath's (1975) 'S-uniform ergodicity'.

These definitions apply to both homogeneous and inhomogeneous chains.

Any chain with a finite state space which is weakly ergodic is S-ergodic; and conversely.

We now list the assumptions needed to ensure the validity of Theorem 1 below. Proofs of the theorems will be found in Section 2. Suppose (Y, d) is a pseudometric space (of age structures), $\mathscr{B}$ the Borel σ-field of subsets of Y generated by open spheres, and $(X, \mathscr{A}, \mu_X)$ is a σ-finite non-negative measure space (of vital rates operators) such that $\mathscr{A}$ is the σ-field generated by a topology on X. Let $Z = X \times Y$ and let $\mathscr{C} = \sigma(\mathscr{A} \times \mathscr{B})$. (If, as in most applications, X and Y are separable metric spaces, then $\mathscr{C} = \mathscr{A} \times \mathscr{B}$ (Billingsley (1968), p. 225).)

Suppose that the application of an operator $x \in X$ to an element $y \in Y$ yields another element of Y denoted by xy. Assume that the image xy is a jointly continuous function of both x and y. Let B/y denote the set of all operators x in X which, when applied to an element y in Y, produce an image xy falling in $B \in \mathscr{B}$, i.e., $B/y = \{x \in X : xy \in B\}$. (Obviously$(B/y)y \subset B$.) Then suppose that B/y is a uniformly continuous function of y, i.e. for any $\varepsilon > 0$, there exists $\delta > 0$ such that, for any y, y' in Y, if $d(y, y') < \delta$, then $\mu_X((B/y) \triangle (B/y')) < \varepsilon$.

Now let $\{X_n\}$ form a Markov chain, i.e. for every n let P_n be a regular (Loève (1963), p. 137) Markov transition function. Let both the initial distribution $Q_1(\cdot)$ and $P_n^m(x, \cdot)$, $x \in X$, be μ_X-continuous. We define $\mathfrak{A}$ to be the Borel field (whose points are the sets A in $\mathscr{A}$) which is generated by the family of open spheres $S(A, r) = \{A' \in \mathscr{A} : \mu_X(A \triangle A') < r\}$, for all $A \in \mathscr{A}$ and $r > 0$. Then we assume further that $(x, A) \to P_n^m(x, A)$ is jointly measurable, for every n and m; that is, if $p \in [0, 1]$, then $\{(x, A) : x \in X,\ A \in \mathscr{A} \text{ and } P_n^m(x, A) \leqq p\} \in \sigma(A \times \mathfrak{A})$.

For a given y_0 in Y, define $\{Y_n\}$ inductively by $Y_0(\Omega) = y_0$, $Y_n(\omega) = X_n(\omega) Y_{n-1}(\omega)$. Define $\{Z_n\}$, with sample probability space $(Z, \mathscr{C})$, by $Z_n(\omega) = (X_n(\omega), Y_n(\omega))$. Finally, define $G_n : X \times Y \times (\mathscr{A} \times \mathscr{B}) \to [0, 1]$ by $G_n(x, y, A \times B) = P_n(x, A \cap (B/y))$.

Theorem 1. With the definitions and assumptions listed above,

(i) G_n is a regular conditional probability.

(ii) There is a unique extension of G_n to a regular conditional probability which maps $X \times Y \times \mathscr{C}$ to $[0, 1]$. (Again, this is immediate if X and Y are separable metric spaces.)

(iii) Z_n is a Markov chain with one-step transition probability function given by $\boldsymbol{P}[Z_{n+1} \in C \mid Z_n = z] = G_n(x, y, C)$, $z \in Z$, $C \in \mathscr{C}$, $n \in N$, and with initial probability distribution F_1 determined by the unique extension to $\mathscr{C}$ of the function $F_1 : \mathscr{A} \times \mathscr{B} \to [0,1]$ defined by $F_1(A \times B) = Q_1(A \cap (B/y_0))$ where Q_1 is the distribution of X_1.

Theorem 2. Under the assumptions of Theorem 1, the one-step transition probability functions $P_n(x, \cdot)$ of $\{X_n\}$ are expressible as integrals of density functions. Suppose that these densities are uniformly bounded above, and that $\{X_n\}$ is S-ergodic (Definition 3).

Suppose, analogously to the weak ergodic theorem of demography, that for every $\delta > 0$ there exists m_0 such that for all $m \geqq m_0$, for all initial elements y, $y' \in Y$, and all subsequent sequences of operators $x_1, \cdots, x_m$ from X, $d(x_m \cdots x_1 y, x_m \cdots x_1 y') < \delta$.

Then $\{Z_n\}$ is a uniformly weakly ergodic Markov chain (Definition 2).

Theorem 3. Under the assumptions of Theorem 2, let $F_n : \mathscr{C} \to [0,1]$ be the distribution of Z_n, n in N. Suppose, after some time n_0 in N which, without loss of generality, we shall take to be $n_0 = 1$, the one-step transition functions of the chain X_n are homogeneous in time, that is, $P_n = P_1$ for n in N.

Then (i) there is a probability $F(\cdot) : \mathscr{C} \to [0,1]$ such that, for every n in N, C in $\mathscr{C}$, $\lim_m \sup_{(x,y)\in Z} | G_n^m(x, y, C) - F(C)| = 0$.

(ii) F satisfies the renewal equation $F(C) = \int_X \int_Y F(dx \times dy) G_n(x, y, C)$, for all C in $\mathscr{C}$, and any n in N.

(iii) F is an invariant distribution of the chain Z_n, that is, if $F_1 = F$, then $F_n = F$ for all n in N.

(iv) $\text{Lim}_n \sum_{k=1}^n g(Z_k)/n = \int_Z g(z)F(dz)$ almost surely, for any Borel function g for which the integral exists.

(v) Let Z be a metric space, $\mathscr{C}$ the σ-field of Borel subsets of Z. Let $g : Z \to R$ be any bounded, real, measurable function that is continuous almost everywhere with respect to F. Then for all n in N, x' in X, y' in Y, $\lim_m | \int_Z g(z) G_n^m(x', y', dz) - \int_Z g(z)F(dz)| = 0$.

Theorem 4. Under the conditions of Theorem 1, replace the assumption that B/y is a uniformly continuous function of y by the stronger assumption of uniform equicontinuity, namely, for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all B in $\mathscr{B}$ and for all y, y' in Y, if $d(y, y') < \delta$ then $\mu_X(B/y \triangle B/y') < \varepsilon$.

Then:

(i) Under the additional assumptions of Theorem 2, $\{Z_n\}$ is an S-ergodic Markov chain (Definition 3).

(ii) If $\{X_n\}$ is also homogeneous as in Theorem 3, then the chain $\{Z_n\}$ is exponentially convergent (Loève (1963)), that is, there exist $a > 0$ and $b > 0$ such that, for every n, m in N, $\sup_{z\in Z} \sup_{C\in\mathscr{C}} | G_n^m(x, y, C) - F(C)| \leqq ae^{-bm}$.

We now indicate, as corollaries, two important applications of the above results.

Corollary 1 (age-structured populations).

(i) Let k in N be greater than 1, and k', k'' in N satisfy $k' < k'' \leqq k$, g.c.d. $(k', k'') = 1$. Let L, U in P satisfy $0 < L < 1$, $L < U$. Let X be the set of $k \times k$ real matrices of the form

$$x = \begin{pmatrix} f_1 & \cdots & f_{k-1} & f_k \\ s_1 & \ddots & 0 & 0 \\ 0 & \ddots & s_{k-1} & 0 \end{pmatrix}$$

satisfying $L \leqq s_j \leqq 1$, $L \leqq f_{k'}, f_{k''} \leqq U$, and $0 \leqq f_j \leqq U$ for $j \neq k', k''$, $j = 1, 2, \cdots, k$. Non-zero elements may occur only in the first row and subdiagonal. s_j occurs in row $j+1$ and column j, $j = 1, \cdots, k-1$.

For each matrix x in X, let $p(x)$ be the vector $p(x)^{\text{Tr}} = (f_1 \cdots f_k, s_1 \cdots s_{k-1}) \equiv (p_1(x), \cdots, p_{2k-1}(x))$ in P^{2k-1}. Let I_j be the open interval (excluding both end points) of possible values of $p_j(x)$, $x \in X$. For example, $I_j = (L, 1)$ for $j = k+1, \cdots, 2k-1$. Define $p(X) = I_1 \times I_2 \times \cdots \times I_{2k-1}$; $p(X)$ is an open rectangular parallelepiped in P^{2k-1} which excludes all the $(2k-2)$-dimensional faces. Let $\text{Int}(X) = \{x \in X : p(x) \in p(X)\}$; $\text{Int}(X)$ is the interior of X. If the topology on X is induced by the metric $d_X(x, x') = \| x - x' \|_X$, then $\mathcal{A}$ is the Borel field. Take μ_X to be Lebesgue measure on P^{2k-1}.

(ii) Let $Y = \{y \in P^k : y(j) > 0$ for some $j \leqq k''$ and $\| y \| = 1\}$. Y is of dimension $k-1$. Algebraically, each y in Y is a column k-vector with jth element $y(j)$. For y, y' in Y, define $d(y, y') = \| y - y' \|$. $\mathcal{B}$ is the Borel field in Y with typical element B. Let y_0, y_0' be in Y.

(iii) Let $Z = X \times Y$, with elements $z = (x, y)$; and $\mathcal{C} = \sigma(\mathcal{A} \times \mathcal{B})$ with elements C. (Since Z is separable, $\mathcal{C} = \mathcal{A} \times \mathcal{B}$.) The product metric ρ_Z on $Z \times Z$ is $\rho_Z(z, z') = \max(d_X(x, x'), d(y, y'))$.

(iv) Define $xy = x \cdot y / \| x \cdot y \|$ where $\cdot$ means the usual multiplication of a matrix (on the left) by a column vector (on the right). (Thus $B / y = \{x \in X : \| x \cdot y \| b = x \cdot y$, for some $b \in B\}$.)

(v) Let $\{X_n\}$, n in N, be a Markov chain with sample probability space $(X, \mathcal{A})$. For each n in N, let $p_n(x, x')$ be the conditional density of X_{n+1} at x' given $X_n = x$. Let $p_n(x, x')$ be jointly continuous in x and x'; let $\sup_{x \in X} \sup_{x' \in X} p_n(x, x') < M < \infty$. Assume 'uniform positivity': there exists δ, $0 < \delta < 1$, such that for every n in N, there exists $A_{n+1} \in \mathcal{A}$ satisfying $\mu_X A_{n+1} \geqq \delta$ and $\inf_{x \in X} \inf_{x' \in A_{n+1}} p_n(x, x') \geqq \delta$.

(vi) Let $x_1 \in \text{Int}(X)$. Define $Y_n(\omega) = X_n(\omega) \cdot Y_{n-1}(\omega) / \| X_n(\omega) \cdot Y_{n-1}(\omega) \|$, and $Z_n = (X_n, Y_n)$.

Then the conclusions of Theorems 1, 2 and 3 apply. In particular, $\{Z_n\}$ is uniformly weakly ergodic; the age structure $\{Y_n\}$ converges in distribution; and the factor by which total population size changes in the interval from time n to $n+1$, averaged over all sample paths beginning from $Z_1 = z_1$, converges as

$n \to \infty$ to $\lambda = \int_{z \in Z} \int_{x' \in X} (\| x' \cdot y \| / \| y \|) \cdot P_1(x, dx') F(dz)$, where $z \equiv (x, y)$ and F is given by Theorem 3. Moreover, the time-average rate of growth for any particular sample path converges almost surely to λ, i.e.

$$\lim_n n^{-1} \sum_{j=0}^{n-1} \| X_{j+1}(\omega) \cdot Y_j(\omega) \| / \| Y_j(\omega) \| = \lambda \quad \text{almost surely } (\omega).$$

Remarks. We now specify a concrete Markov chain $\{X_n\}$ in order to show that these apparently abstract and stringent conditions are met by a process which is demographically plausible and statistically natural for the analysis of historical data.

Let $T_1(a, \cdot), \cdots, T_{2k-1}(a, \cdot)$, where the parameter a is in R, be $2k-1$ continuous functions with continuous derivatives from $p(X)$ into R defined by

$$T_j(a, p(x)) = \ln\left(-\ln\left(\sum_{i=1}^{j} p_i(x)/(kU)\right) - a\right)$$

$$= \ln\left(-\ln\left(\sum_{i=1}^{j} f_i/(kU)\right) - a\right)$$

for $j = 1, \cdots, k$;

$$T_j(a, p(x)) = -\ln\left(\left[\prod_{i=k+1}^{j} p_i(x) - L^{j-k}\right] \Big/ \left[1 - \prod_{i=k+1}^{j} p_i(x)\right]\right)$$

$$= -\ln\left(\left[\prod_{i=1}^{j-k} s_i - L^{j-k}\right] \Big/ \left[1 - \prod_{i=1}^{j-k} s_i\right]\right)$$

for all $j = k+1, \cdots, 2k-1$.

Since $\partial T_j / \partial p_i(x) = 0$ whenever $i > j$ and $\partial T_i / \partial p_i(x) < 0$ everywhere on $p(X)$, the Jacobian of $T = (T_1, \cdots, T_{2k-1})^{\mathrm{Tr}}$ never vanishes on $p(X)$. Hence $T^{-1}: R^{2k-1} \to p(X)$ exists, is continuous and has continuous derivatives.

Define the map $S: R^{2k-1} \to R^{2k-1}$ via

$$S_j(q) = \alpha_1 + \beta_1 q_j, \quad j = 1, \cdots, k$$

$$= \alpha_2 + \beta_2 q_j, \quad j = k+1, \cdots, 2k-1,$$

where $q^{\mathrm{Tr}} = (q_1, \cdots, q_{2k-1})$ and $S^{\mathrm{Tr}} = (S_1, \cdots, S_{2k-1})$.

Define $H: \mathrm{Int}(X) \to \mathrm{Int}(X)$ as acting identically to $T^{-1}(b, S \circ T(a, \cdot))$ on $p(X)$.

Then let $p_n(\cdot, \cdot) = p(\cdot, \cdot)$ where, for x, $x' \in \mathrm{Int}(X)$,

$$\text{(A)} \qquad p(x, x') = \frac{\exp\{-\frac{1}{2} p(x' - H(x))^{\mathrm{Tr}} \cdot V^{-1} \cdot p(x' - H(x))\}}{\int_{x'' \in \mathrm{Int}(X)} \exp\{-\frac{1}{2} p(x'' - H(x))^{\mathrm{Tr}} \cdot V^{-1} \cdot p(x'' - H(x))\} dx''}.$$

In other words, if $X_n(\omega) = x$, let $X_{n+1}(\omega) = H(x) + \varepsilon$, where $p(\varepsilon)$, the vector containing the possibly non-zero elements of the matrix ε, has a truncated multivariate normal distribution. The truncation excludes all values of $H(x)+\boldsymbol{\varepsilon}$ falling outside of $\text{Int}(X)$; the untruncated normal distribution of $p(\boldsymbol{\varepsilon})$ has means $\mathbf{0}$ and a non-singular matrix $\boldsymbol{V}$ of finite variances and covariances (e.g., but not necessarily, $\boldsymbol{V}=\sigma^2\boldsymbol{I}$ where $0<\sigma^2<\infty$).

The transformation H may be interpreted as follows: $\Sigma_{i=1}^{j} p_i(x)$ is (except for mortality adjustments required by the discrete time interval) the cumulated fertility through the jth age category, and $\Sigma_{i=1}^{j} p_i(x)/(kU)$ is cumulated fertility as a fraction of the maximum possible gross rate of reproduction. If the age categories correspond to age intervals of equal width and if cumulated fertility is described by the Gompertz distribution (Brass (1974), p. 552), then, for appropriate choice of constant a, $T_j(a,p(x))$ is linear in j, $j=1,\cdots,k$. $T_j(a,p(x))$, $j=k+1,\cdots,2k-1$ are logit transformations of the probabilities $\Pi_{i=1}^{j-k} s_i$ of surviving from birth to age category $j-k$.

Empirical studies reviewed by Brass (1974) suggest that short-term variation in age-specific fertility can be represented as a single linear transformation of the Gompertz-transformed rates, and similarly for the logit-transformed age-specific survival rates. The choice of a, b, α_i, β_i, $i=1,2$, is determined by examination of particular data. When $a=b$, $\alpha_1=\alpha_2=0$, $\beta_1=\beta_2=1$, H is the identity map.

The transformation H simply rearranges the elements of any x into a vector, takes the Gompertz transform of the cumulated fertility and the logit transform of the cumulated survival described by x, applies a linear transform to each, and inversely transforms the result to an element $H(x)$ of $\text{Int}(X)$.

Corollary 2 (Markov chains with random transition matrices). (i) Let $k>1$, k in N, $0<L\leqq 1$. Let $X\subset P^{k\times k}$ be the set of all $k\times k$ row stochastic matrices $x=(x(i,j))$, $\Sigma_{j=1}^{k} x(i,j)=1$, such that x is irreducible and aperiodic and that for any rows i_1, i_2 there is a column j such that $x(i_1,j)\geqq L$, $x(i_2,j)\geqq L$. (The column j is not required to be uniform over x in X.) Let $d_X(x,x')=\|x-x'\|_X$. Let θ be the metric topology of $P^{k\times k}$, $\mathcal{A}$ the Borel field, μ_X Lebesgue measure in Euclidean space of dimension $k(k-1)$.

(ii) Let $Y=\{y\in P^k:\|y\|=1\}$, $d(y,y')=\|y-y'\|$ for any y, y' in Y. Interpret each y algebraically as a *row* k-vector. y_0 in Y is arbitrary; $\mathcal{B}$ is the Borel field in Y with typical element B; μ is Lebesgue measure in Euclidean $(k-1)$-space.

(iii) $Z=X\times Y$ with elements $z=(x,y)$; and $\mathcal{C}=\mathcal{A}\times\mathcal{B}$ with elements C.

(iv) Define $xy=y\cdot x$ where $\cdot$ means multiplication of a row vector (on the left) by a matrix (on the right).

(v) Let $\{X_n\}$, n in N, be a Markov chain with sample probability space $(X,\mathcal{A})$. For each n in N, let $p_n(x,x')$ be the conditional density of X_{n+1} at x' given $X_n=x$. Let $p_n(x,x')$ be jointly continuous in x and x'; let $\sup_{x\in X}\sup_{x'\in X} p_n(x,x')<M<\infty$. Assume uniform positivity (defined in Corollary 1).

(vi) Choose any $x_1\in X$. Let $Y_0(\Omega)=y_0$ and $X_1(\Omega)=x_1$. For $n>0$, let $Y_n(\omega)=Y_{n-1}(\omega)\cdot X_n(\omega)$, $Z_n=(X_n,Y_n)$.

Then, interpreting the notation as defined above, Theorems 1, 2 and 3 apply.

Remarks. Except for the uniform lower bound L, the set X in Corollary 2 is the class of 'scrambling matrices' defined by Hajnal ((1958), p. 235). The Markov matrices independently studied by Takahashi ((1969), p. 438, his Lemma 7) are special scrambling matrices. X is a special kind of ergodic set (Definition 4 below, due to Hajnal (1976)).

Hajnal ((1956), pp. 76–77) suggests the possibility of studying non-homogeneous Markov chains whose transition matrices are determined by a stochastic process. Corollary 2 may be viewed as one interpretation of that suggestion.

Takahashi (1969), in another possible interpretation which is apparently independent of Hajnal (1956), (1958), assumes that X_n, X_m $(m \neq n)$ are independently (though, in his Theorem 8, p. 441, not necessarily identically) distributed in the set of *all* stochastic $k \times k$ matrices, and finds conditions on the distributions which imply almost sure uniform contraction: for all $\varepsilon > 0$, all $\delta > 0$, and any two probability row k-vectors y_0, y_0' in Y as defined in Corollary 2, there exists n_0 in N such that for $n \geqq n_0$,

$$P[\| y_0 X_1(\omega) \cdots X_n(\omega) - y_0' X_1(\omega) \cdots X_n(\omega) \| < \varepsilon] > 1 - \delta.$$

The *same* sequence of stochastic matrices is applied to y_0' as to y_0; hence this result establishes an almost sure version of the sure condition, which Hajnal (1958) calls ergodicity in the weak sense, assumed in our Theorem 2. Hybrids of our Corollary 2 and Takahashi's Theorem 8 can be imagined.

Definition 4 (Hajnal (1976)). An ergodic set $H(s, g, r)$ is a set of $s \times s$ non-negative square matrices with at least one positive element in each row and in each column such that any product of g factors which are members of $H(s, g, r)$ is positive (i.e., every element of the product is positive and finite) and such that for each h in $H(s, g, r)$, $\min^+(h)/\max^+(h) > r > 0$. Here $\min^+(h)$ and $\max^+(h)$ are the smallest and largest of the positive elements of h; s and g are in N, $r > 0$ is in P.

The sets X defined in the preceding corollaries are ergodic sets if $k'' = k$ in Corollary 1. Hajnal (1976) describes many more examples.

Corollary 3 (finite ergodic sets of operators). Let $X = H_1(s_1, g_1, r_1)$ be an ergodic set containing s_2 distinct members (s_2 finite) labelled $x_1, \cdots, x_{s_2}$, and let $S = H_2(s_2, g_2, r_2)$ be an ergodic set each of whose members is stochastic. (This means that if the elements of t in S are $t(i, j)$, then $\Sigma_{j=1}^{s_2} t(i, j) = 1$, $i = 1, \cdots, s_2$.) Let $\{t_n\}_{n=1}^{\infty}$ be an infinite sequence, with repetitions possible, of members of S, and let $\{X_n\}_{n=1}^{\infty}$ be a Markov chain with state space X such that $P[X_{n+1} = x_j \mid X_n = x_i] = t_n(i, j)$. Let Y be the set of all positive column s_1-vectors with elements which sum to 1.

Define $\mathscr{B}$ to be the family of all Borel sets B in Y. Define $B/y = \{x \in X : x \cdot y/\| x \cdot y \| \in B\}$. Let $Z = \{(x, y): x \in X,\ y \in Y\}$ and let $\mathscr{C}$ be the set of all sets $C = A \times B$ where $A \subset X$, $B \in \mathscr{B}$. Let $y_0 \in Y$. Define $\{Y_n\}_{n=0}^{\infty}$ to be the family of random variables with sample space Y such that $Y_0 = y_0$ with probability 1, and for $n > 0$, $Y_n = X_n \cdot Y_{n-1}/\| X_n \cdot Y_{n-1} \|$. Let $\{Z_n\}_{n=1}^{\infty} = \{(X_n, Y_n)\}_{n=1}^{\infty}$. For $n \in N$, $x \in X$, $y \in Y$, $A \subset X$, $B \in \mathscr{B}$, define $G_n(x, y, A \times B) = P_n(x, A \cap (B/y))$ where $P_n(x, A) = \Sigma_{x_j \in A}\, t_n(x, x_j)$.

Then:

(i) G_n is a regular conditional probability which maps $X \times Y \times \mathscr{C}$ to $[0, 1]$.

(ii) Z_n is a Markov chain with one-step transition probability function given by $\boldsymbol{P}[Z_{n+1} \in C \mid Z_n = (x, y)] = G_n(x, y, C)$, $(x, y) \in Z$, $C \in \mathscr{C}$, $n \in N$.

(iii) Z_n is uniformly weakly ergodic. In particular, for any two members y_0 and y_0' of Y, if $\cdot$ means ordinary matrix or matrix-vector multiplication,

$$\limsup_n \sup_{y_0, y_0' > 0} \sup_{x_1(\omega_1), x_1(\omega_2) \in X} \big| \boldsymbol{P}[X_n(\omega_1) \cdots X_1(\omega_1) \cdot y_0/\| X_n(\omega_1) \cdots X_1(\omega_1) \cdot y_0 \| \in B] - \boldsymbol{P}[X_n(\omega_2) \cdots X_1(\omega_2) \cdot y_0'/\| X_n(\omega_2) \cdots X_1(\omega_2) \cdot y_0' \| \in B] \big| = 0.$$

(iv) When $t_n = t$ for all n, then the five conclusions of Theorem 3 follow. In particular, for every positive y_0, suppressing ω

$$\lim_n \boldsymbol{P}[X_n = x_j \text{ and } X_n \cdots X_1 \cdot y_0/\| X_n \cdots X_1 \cdot y_0 \| \in B] = F(x_j, B), j = 1, \cdots, s_2,$$

where $F : X \times \mathscr{B} \to [0, 1]$ is the limiting joint probability distribution. F is the solution of

$$F(x_j, B) = \int_{y \in Y} \sum_{i=1}^{s_2} F(x_i, dy) t(i, j) I_B(x_j \cdot y/\| x_j \cdot y \|)$$

where $I_B(y) = 1$ if $y \in B$, $I_B(y) = 0$ if $y \notin B$.

2. Proofs

2.1. *Proof of Theorem* 1.

Lemma 1. For A, $A' \in \mathscr{A}$, if $\rho(A, A') = \mu_X(A \triangle A')$, then $(\mathscr{A}, \rho)$ is a pseudometric space which is homeomorphic to a pseudometric space of diameter at most one. An additive vector- or scalar-valued function on $\mathscr{A}$ which is μ_X-continuous is continuous on $(\mathscr{A}, \rho)$; $A \cup A'$, $A \cap A'$, $A \triangle A'$, and A^c are continuous functions of A and A'.

Proof. Kelley ((1955), p. 121) and Dunford and Schwartz ((1958), p. 158).

Lemma 2. Y_n is $\mathscr{B}$-measurable.

Proof. Y_0 is measurable since $\Omega \in \mathcal{F}$. Since $y_n = x_n y_{n-1}$ is continuous, the compound map $Y_n(\omega) = X_n(\omega) Y_{n-1}(\omega)$ is measurable. Use induction.

Proof of (i). Regularity means (Loève (1963), p. 137) (a) for every $A \times B$ in $\mathcal{A} \times \mathcal{B}$, $G_n(\cdot, \cdot, A \times B)$ is $\mathcal{C}$-measurable; and (b) for every z in Z, $G_n(x, y, \cdot)$ is a probability on $\mathcal{A} \times \mathcal{B}$.

(a) The map from Z to $X \times \mathcal{A}$ given by $z = (x, y) \to (x, B/y)$ is jointly continuous by assumption. The map $(x, B/y) \to (x, A \cap B/y)$ is jointly continuous by Lemma 1. The map $(x, A \cap B/y) \to P_n(x, A \cap B/y) = G_n(x, y, A \times B)$ is jointly measurable by assumption. Hence the composed map is $\mathcal{C}$-measurable.

(b) Given (x, y) in Z, we show that $G_n(x, y, A \times B) \geqq 0$, $G_n(x, y, X \times Y) = 1$, and G_n is σ-additive. First, $G_n(x, y, A \times B) \geqq 0$ since $P_n(x, A \cap (B/y)) \geqq 0$. Second, for any y in Y, $X \subset Y/y$ because if $x \in X$ then $xy \in Y$. But also $X \supset Y/y$ by definition. Hence $X = Y/y$. So $G_n(x, y, X \times Y) = P_n(x, X \cap (Y/y)) = P_n(x, X) = 1$. Third, we show initially that G_n is additive on disjoint elements of $\mathcal{A} \times \mathcal{B}$. Let $A \times B = A_1 \times B_1 + A_2 \times B_2$. Then $A_1 \cap A_2 = \varnothing$ or $B_1 \cap B_2 = \varnothing$. Now for any y in Y and any B', B'' in $\mathcal{B}$, $(B' \cap B'')/y = (B'/y) \cap (B''/y)$. In the present situation, letting $B_1 = B'$, $B_2 = B''$ gives

$$A_1 \cap (B_1/y) \cap A_2 \cap (B_2/y) = (A_1 \cap A_2) \cap ((B_1 \cap B_2)/y) = \varnothing$$

so that

$$(A_1 \cap (B_1/y)) \cup (A_2 \cap (B_2/y)) = (A_1 \cap (B_1/y)) + (A_2 \cap (B_2/y)).$$

Then

$$G_n(x, y, A \times B) = P_n(x, (A_1 \cap (B_1/y)) \cup (A_2 \cap (B_2/y))) = P_n(x, A_1 \cap (B_1/y)) \\ + P_n(x, A_2 \cap (B_2/y)) = G_n(x, y, A_1 \times B_1) + G_n(x, y, A_2 \times B_2)$$

by additivity of P_n on disjoint elements of $\mathcal{A}$. Finally, to show that G_n is σ-additive, it remains only to show that G_n is continuous from above at $\varnothing$ (Kingman and Taylor (1966), p. 56, Theorem 3.2 (iii)). Let $A_j \times B_j$, $j \in N$ be a decreasing sequence of sets in $\mathcal{A} \times \mathcal{B}$ with limit $\varnothing$; write $A_j \times B_j \downarrow \varnothing$. Then $A_j \downarrow \varnothing$ or $B_j \downarrow \varnothing$ or both. If $A_j \downarrow \varnothing$ then $\lim_j P_n(x, A_j) = 0$ for each x in X, n in N, and $P_n \geqq G_n$. If $B_j \downarrow \varnothing$, then for any y in Y, $B_j/y \downarrow \varnothing$, so $\lim_j P_n(x, A_j \cap (B_j/y)) \leqq \lim_j P_n(x, B_j/y) = 0$, for any x in X, y in Y, and $n \in N$. So G_n is σ-additive.

Proof of (ii). We first show (b) there is a unique extension of $G_n(x, y, \cdot)$ from the domain $\mathcal{A} \times \mathcal{B}$ to the domain $\mathcal{C} = \sigma(\mathcal{A} \times \mathcal{B})$ and that this extension is a probability on $\mathcal{C}$, for every x in X, y in Y; then we show (a) $G_n(\cdot, \cdot, C)$ is $\mathcal{C}$-measurable, for every C in $\mathcal{C}$.

(b) $\mathcal{A} \times \mathcal{B}$ is a semi-ring (Kingman and Taylor (1966), pp. 15, 134). So

(Theorem 3.5 of Kingman and Taylor (1966), p. 66) there is a unique additive extension of $G_n(x, y, \cdot)$ to a measure (which is also non-negative) on the ring $\mathcal{R}(\mathcal{A} \times \mathcal{B})$ generated by $\mathcal{A} \times \mathcal{B}$. This ring is actually a field, so $G_n(x, y, \cdot)$ is a probability on $\mathcal{R}(\mathcal{A} \times \mathcal{B})$. Since the σ-ring generated by $\mathcal{R}(\mathcal{A} \times \mathcal{B})$ is actually a σ-field $\mathcal{C}$ and $G_n(x, y, \cdot)$ is bounded on $\mathcal{R}(\mathcal{A} \times \mathcal{B})$, there is a unique extension of $G_n(x, y, \cdot)$ to a (non-negative) measure on $\mathcal{C}$ (Theorem 4.2 of Kingman and Taylor (1966), p. 77). The extension is a probability since $G_n(x, y, X \times Y) = 1$.

(a) Let $\mathcal{C}_0 = \{C \in \mathcal{C}$: for every z in Z, the map $G_n(\cdot, \cdot, C): z \to G_n(x, y, C)$ is measurable$\}$. By (i) $\mathcal{C}_0 \supset \mathcal{A} \times \mathcal{B}$. If $C \in \mathcal{R}(\mathcal{A} \times \mathcal{B})$, then $C = \Sigma_1^s C_i$, $C_i \in \mathcal{A} \times \mathcal{B}$ for some finite s in N (Theorem 1.4 of Kingman and Taylor (1966), p. 17). For such C, by (b) above, $G_n(\cdot, \cdot, C) = \Sigma_1^s G_n(\cdot, \cdot, C_i)$ which is a continuous function of measurable functions and therefore measurable. So $\mathcal{C}_0 \supset \mathcal{R}(\mathcal{A} \times \mathcal{B})$. Finally, if $\{C_j\}_{j=1}^{\infty}$ is a monotone sequence of sets in $\mathcal{R}(\mathcal{A} \times \mathcal{B})$, then for each fixed n, $\{G_n(\cdot, \cdot, C_j)\}_{j=1}^{\infty}$ is a monotone sequence of measurable functions. So (Theorem 5.4 (iii) of Kingman and Taylor (1966), p. 106) $\lim_j G_n(\cdot, \cdot, C_j)$ is measurable. By continuity of $G_n(x, y, \cdot)$ for every $z \in Z$, $\lim_j G_n(\cdot, \cdot, C_j) = G_n(\cdot, \cdot, \lim_j C_j)$. Hence $\lim_j C_j \in \mathcal{C}_0$ and $\mathcal{C}_0$ is a monotone class containing $\mathcal{R}(\mathcal{A} \times \mathcal{B})$. Then $\mathcal{C}_0 \supset \mathcal{C}$ by the corollary of the monotone class theorem (Kingman and Taylor (1966), p. 18), or by the $\pi-\lambda$ theorem (Blumenthal and Getoor (1968), p. 5, Theorem 2.2).

Proof of (iii). Y_n is $\mathcal{B}$-measurable by Lemma 2. Given $Y_n(\omega) = y_n$, $Y_{n+1}(\omega) = X_{n+1}(\omega) y_n$ by construction. But X_{n+1} given $X_n(\omega) = x_n$ is conditionally independent of $X_1(\omega), \cdots, X_{n-1}(\omega)$ because $\{X_n\}$ is a Markov chain. Thus for $C \in \mathcal{C}$, $\boldsymbol{P}[(X_{n+1}(\omega), \ Y_{n+1}(\omega)) \in C \mid Z_j(\omega) = z_j, \ j = 1, \cdots, n] = \boldsymbol{P}[Z_{n+1}(\omega) \in C \mid Z_n(\omega) = z_n]$. So $\{Z_n\}$ is a Markov chain.

To show that G_n is the one-step transition probability function of Z_n, by (ii), it suffices to establish that $\boldsymbol{P}[Z_{n+1} \in C \mid Z_n = z] = G_n(x, y, C)$ for $C = A \times B \in \mathcal{A} \times \mathcal{B}$. Now $Y_{n+1} = X_{n+1} y$ if $Y_n = y$. Then $X_{n+1} \in A$ and $Y_{n+1} \in B$ if and only if $X_{n+1} \in A$ and $X_{n+1} y \in B$ if and only if $X_{n+1} \in A$ and $X_{n+1} \in B/y$ if and only if $X_{n+1} \in A \cap (B/y)$. Hence $\boldsymbol{P}[Z_{n+1} \in C \mid Z_n = z] = P_n(x, A \cap (B/y)) = G_n(x, y, C)$.

For $C = A \times B \in \mathcal{A} \times \mathcal{B}$, $F_1(C) = \boldsymbol{P}[Z_1 \in C] = Q_1(A \cap (B/y_0))$ by the same argument. The extension of F_1 to C in $\mathcal{C}$ repeats the argument of (ii).

2.2. *Proof of Theorem* 2.

Lemma 3. Under the assumptions of Theorem 1, for every n in N and x in X, there is a density function $p_n(x, \cdot)$ determined up to μ_X-null sets such that for every A in $\mathcal{A}$, $P_n(x, A) = \int_A p_n(x, x') \mu_X(dx')$.

Proof. In view of the μ_X-continuity assumed in Theorem 1, the Radon–Nikodym theorem applies.

Lemma 4. Under the assumptions of Theorem 1, if there is a finite constant $M > 0$ such that, for all n in N and x in X, $p_n(x, \cdot) \leqq M$, then $P_n(x, \cdot)$ is uniformly continuous in A, uniformly in $n \in N$ and $x \in X$; that is, for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $A, A' \in \mathcal{A}$, if $\rho(A, A') < \delta$ then for every n in N and x in X, $|P_n(x, A) - P_n(x, A')| < \varepsilon$.

Proof. Choose $\delta = \varepsilon / M$. Then $|P_n(x, A) - P_n(x, A')| \leqq \int_{A \triangle A'} M\mu_X(dx') = M\mu_X(A \triangle A') < \varepsilon$.

Lemma 5. Under the assumptions of Theorem 2, for each B in $\mathcal{B}$, the family of maps from Y to $[0, 1]$ given by $\{G_n(x, \cdot, A \times B) : n \in N,\ x \in X,\ A \in \mathcal{A}\}$ is uniformly equicontinuous; that is, for every $\varepsilon > 0$ and every B in $\mathcal{B}$, there exists $\delta > 0$ such that for any y, y' in Y, if $d(y, y') < \delta$, then for every n in N, x in X and A in $\mathcal{A}$, $|G_n(x, y, A \times B) - G_n(x, y', A \times B)| < \varepsilon$.

Proof. Choose B in $\mathcal{B}$ and $\varepsilon > 0$. By Lemma 4, there exists $\delta' > 0$ such that for all A', A'' in $\mathcal{A}$, if $\rho(A', A'') < \delta'$ then for every n in N and x in X, $|P_n(x, A') - P_n(x, A'')| < \varepsilon$. Now for all $A, A_1, A_2 \in \mathcal{A}$, $(A \cap A_1) \triangle (A \cap A_2) = A \cap (A_1 \triangle A_2) \subset A_1 \triangle A_2$, so $\rho(A \cap A_1, A \cap A_2) \leqq \rho(A_1, A_2)$. Theorem 1 assumes that, given B, there exists $\delta > 0$ such that for all y, y' in Y, if $d(y, y') < \delta$ then $\rho(B/y, B/y') < \delta'$. Letting $A_1 = B/y$, $A_2 = B/y'$, $A' = A \cap A_1 = A \cap (B/y)$, $A'' = A \cap A_2 = A \cap (B/y')$ gives $|G_n(x, y, A \times B) - G_n(x, y', A \times B)| = |P_n(x, A \cap (B/y)) - P_n(x, A \cap (B/y'))| < \varepsilon$ whenever $d(y, y') < \delta$.

Lemma 6. Under the assumptions of Theorem 2, for every $\varepsilon > 0$ and every C in $\mathcal{C}$, there exists l_0 in N such that for all n in N, all $l \geqq l_0$ in N, x in X, y, y' in Y and $x_1, \cdots, x_l$ in X, $|G_n(x, x_l \cdots x_1 y, C) - G_n(x, x_l \cdots x_1 y', C)| < \varepsilon$.

Proof. Let $\mathcal{C}_1 = \{C \in \mathcal{C}$: for every $\varepsilon > 0$ there exists l_0 in N such that for all n in N, all $l \geqq l_0$ in N, all $x, x_1, \cdots, x_l$ in X, all y, y' in Y, $|G_n(x, x_l \cdots x_1 y, C) - G_n(x, x_l \cdots x_1 y', C)| < \varepsilon\}$. First, $\mathcal{C}_1 \supset \mathcal{A} \times \mathcal{B}$. For let $\varepsilon > 0$ and $C \in \mathcal{A} \times \mathcal{B}$, $C = A \times B$. By Lemma 5, for this B there exists $\delta > 0$ such that for any y^*, y^{**} in Y, if $d(y^*, y^{**}) < \delta$, then for every n in N, x in X and A' in $\mathcal{A}$ $|G_n(x, y^*, A' \times B) - G_n(x, y^{**}, A' \times B)| < \varepsilon$. By assumption of Theorem 2, there exists l_0 in N such that, for all y, y' in Y, any $l \geqq l_0$, l in N, and all $x_1, \cdots, x_l$ in X, if $y^* = x_l \cdots x_1 y$ and $y^{**} = x_l \cdots x_1 y'$, then $d(y^*, y^{**}) < \delta$. This is the desired l_0.

Secondly, $\mathcal{C}_1 \supset \mathcal{R}(\mathcal{A} \times \mathcal{B})$. For let $\varepsilon > 0$ and $C \in \mathcal{R}(\mathcal{A} \times \mathcal{B})$. Again (Theorem 1.4 of Kingman and Taylor (1966), p. 17), $C = \Sigma_{j=1}^{s} C_j$, $C_j \in \mathcal{A} \times \mathcal{B}$. By Theorem 1 (ii), $G_n(\cdot, \cdot, C) = \Sigma_{j=1}^{s} G_n(\cdot, \cdot, C_j)$. Then there exist $l(j)$ such that whenever $l \geqq l(j)$, l in N, then for all n in N, all $x, x_1, \cdots, x_l$ in X, all y, y' in Y, $|G_n(x, x_l \cdots x_1 y, C_j) - G_n(x, x_l \cdots x_1 y', C_j)| < \varepsilon / s$, for $j = 1, \cdots, s$. Choose $l_0 = \max_j \{l(j)\}$. This is the desired l_0.

Finally, let $\{C_j\}_{j=1}^{\infty}$ be a monotone sequence, $C_j \in \mathcal{R}(\mathcal{A} \times \mathcal{B})$, $C = \lim_j C_j$. For C in $\mathcal{C}$, let $\pi C = \{x \in X : \text{for some } y \text{ in } Y, (x, y) \in C\}$. Then πC is in $\mathcal{A}$, $\mu_X \pi C$ is defined, and for every n in N, x in X, y in Y, $G_n(x, y, C) \leqq P_n(x, \pi C)$. Choose $\varepsilon > 0$. Since $\mu_X \pi C = \lim_j \mu_X \pi C_j$ and for all n in N, x in X, y in Y, $G_n(x, y, \cdot)$ is μ_X-continuous, there exists j_0 in N such that for all $j \geqq j_0$, $G_n(x, y, C - C_j) < \varepsilon/4$, uniformly in n, x and y, by Lemma 4. Moreover since $\mathcal{C}_1 \supset \mathcal{R}(\mathcal{A} \times \mathcal{B})$, there exists l_0 such that for all n in N, all $l \geqq l_0$ in N, all $x, x_1, \cdots, x_l$ in X, and all y, y' in Y, $|G_n(x, x_l \cdots x_1 y, C_{j_0}) - G_n(x, x_l \cdots x_1 y', C_{j_0})| < \varepsilon/2$. Then for this l_0, whenever $l \geqq l_0$, l in N, $|G_n(x, x_l \cdots x_1 y, C) - G_n(x, x_l \cdots x_1 y', C)| < \varepsilon$.

Thus $\mathcal{C}_1$ contains the monotone class generated by $\mathcal{R}(\mathcal{A} \times \mathcal{B})$, so $\mathcal{C}_1 = \mathcal{C}$ by the monotone class theorem (Kingman and Taylor (1966), p. 18).

Lemma 7. If $\{X_n\}$ is any Markov chain on the measurable space $(X, \mathcal{A})$, and if $P_n^m(x, A)$ is the regular transition probability function from $x \in X$ at time n into $A \in \mathcal{A}$ at time $n + m$, then

$$\Delta_{nm}(\{X_n\}) = \sup_{x, x' \in X} \sup_{A \in \mathcal{A}} \{P_n^m(x, A) - P_n^m(x', A)\}$$

$$= \sup_{x, x' \in X} \tfrac{1}{2} \int_X |P_n^m(x, dx'') - P_n^m(x', dx'')|.$$

Proof. Loève ((1963), p. 367).

Proof of Theorem 2. Define $\delta(y, y') = 1$ if $y = y'$, $\delta(y, y') = 0$ if $y \neq y'$. Then $\delta(y, y')\delta(y', y'') = \delta(y, y'')$. Define $r = [m/2] =$ the integral part of $m/2$. We shall always assume $m \geqq 2$. Then $\lim_m r = \lim_m (m - r) = \infty$. Define $G_n^m(x, y, C) = \boldsymbol{P}[Z_{n+m} \in C \mid Z_n = z]$ for all n, m in N.

For n, m in N, y_n in Y, define $y_{n,m}(y_n) = x_{n+m} \cdots x_{n+1} y_n$ as an explicit function of y_n and an implicit function of $x_{n+1}, \cdots, x_{n+m}$ in X. Then $y_{n,m}(y_n) = y_{n+r,m-r}(y_{n,r}(y_n))$. Let y be an arbitrary fixed element of Y. Whenever

$$y_{n+r,m-r-1}(y_{n,r}(y_n)) = x_{n+m-1} \cdots x_{n+r+1} y_{n,r}(y_n)$$

and

$$y_{n+r,m-r-1}(y) = x_{n+m-1} \cdots x_{n+r+1} y$$

occur in the same equation, we interpret the expression $x_{n+m-1} \cdots x_{n+r+1}$ to be the same both times it occurs, so that $y_{n+r,m-r-1}(y_{n,r}(y_n))$ and $y_{n+r,m-r-1}(y)$ have a common factor consisting of the leftmost $m - r - 1$ elements from X.

Now choose n in N, $\varepsilon > 0$ and C in $\mathcal{C}$. By Lemma 6, there exists m_0' large enough that for $m \geqq m_0'$, m in N, $|\eta_{n,m}| < \varepsilon/4$ and $|\eta'_{n,m}| < \varepsilon/4$ uniformly in y_n, y_n' in Y and uniformly in $x_{n+1}, \cdots, x_{n+m-1}$ in X, where

$$\eta_{n,m} = G_{n+m-1}(x_{n+m-1}, y_{n+r,m-r-1}(y_{n,r}(y_n)), C)$$

$$- G_{n+m-1}(x_{n+m-1}, y_{n+r,m-r-1}(y), C)$$

and

$$\eta'_{n,m} = G_{n+m-1}(x_{n+m-1}, y_{n+r,m-r-1}(y_{n,r}(y'_n)), C) - G_{n+m-1}(x_{n+m-1}, y_{n+r,m-r-1}(y), C).$$

The second term in these definitions does not depend on y_n, y'_n or $x_{n+1}, \cdots, x_{n+r}$. $\eta_{n,m}$ and $\eta'_{n,m}$ are implicitly functions of y, of $x_{n+1}, \cdots, x_{n+m-1}$ and, respectively, of y_n, y'_n but not of x_n or x'_n.

Then $G_n(x, y, A \times B) = P_n(x, A \cap (B/y))$ implies that $G_n(x, y, dx' \times dy') = P_n(x, dx' \cap (dy'/y)) = P_n(x, dx')\delta(y', x'y)$. The iterated regular conditional probabilities theorem (Loève (1963), p. 137) applied to the Markov chain Z_n gives

$$\begin{aligned}
G_n^m(x_n, y_n, C) &= \int_{z_{n+1}\in Z} P_n(x_n, dx_{n+1})\delta(y_{n+1}, x_{n+1}y_n) \\
&\quad \cdots \int_{z_{n+m-1}\in Z} P_{n+m-2}(x_{n+m-2}, dx_{n+m-1}) \\
&\quad \cdot \delta(y_{n+m-1}, x_{n+m-1}y_{n+m-2}) G_{n+m-1}(x_{n+m-1}, y_{n+m-1}, C) \\
&= \int_X \cdots \int_X P_n(x_n, dx_{n+1}) \cdots G_{n+m-1}(x_{n+m-1}, x_{n+m-1} \cdots x_{n+1}y_n, C) \\
&= \int_{x_{n+1}\in X} \cdots \int_{x_{n+r}\in X} P_n(x_n, dx_{n+1}) \cdots P_{n+r-1}(x_{n+r-1}, dx_{n+r}) \\
&\quad \cdot \int_{x_{n+r+1}\in X} \cdots \int_{x_{n+m-1}\in X} P_{n+r}(x_{n+r}, dx_{n+r+1}) \\
&\quad \cdots G_{n+m-1}(x_{n+m-1}, y_{n+r,m-r-1}(y_{n,r}(y_n)), C) \\
&= \int_X \cdots \int_X P_n(x_n, dx_{n+1}) \cdots P_{n+r-1}(x_{n+r-1}, dx_{n+r}) \\
&\quad \cdot \int_X P_{n+r}(x_{n+r}, dx_{n+r+1}) \cdots [G_{n+m-1}(x_{n+m-1}, y_{n+r,m-r-1}(y), C) + \eta_{n,m}] \\
&= \int_{x_{n+1}\in X} \cdots \int_{x_{n+m-1}\in X} \eta_{n,m} P_n(x_n, dx_{n+1}) \cdots P_{n+m-2}(x_{n+m-2}, dx_{n+m-1}) \\
&\quad + \int_{x_{n+r}\in X} P_n^r(x_n, dx_{n+r}) \int_{x_{n+r+1}\in X} \cdots \int_{x_{n+m-1}\in X} P_{n+r}(x_{n+r}, dx_{n+r+1}) \\
&\quad \cdots G_{n+m-1}(x_{n+m-1}, y_{n+r,m-r-1}(y), C) \\
&= T_1 + T_2
\end{aligned}$$

where T_i is the ith term on the right. Replacing x_n by x'_n and $\eta_{n,m}$ by $\eta'_{n,m}$ gives expressions of identical form for $G_n^m(x'_n, y'_n, C) = T'_1 + T'_2$. Then

$|G_n^m(x_n, y_n, C) - G_n^m(x'_n, y'_n, C)| = |T_1 + T_2 - T'_1 - T'_2| \leqq |T_1| + |T'_1| + |T_2 - T'_2| \leqq \varepsilon/4 + \varepsilon/4 + \int_{x_{n+r} \in X} |P_n^r(x_n, dx_{n+r}) - P_n^r(x'_n, dx_{n+r})|$. Since $\{X_n\}$ is S-ergodic, there exists m_0 in N at least as big as m'_0 chosen earlier so that whenever $m \geqq m_0$, the integral on the right is less than $\varepsilon/2$. Then, whenever $m \geqq m_0$, $|G_n^m(x_n, y_n, C) - G_n^m(x'_n, y'_n, C)| < \varepsilon$, uniformly in z_n and z'_n in Z.

2.3. *Proof of Theorem* 3. Since $P_n = P_1$, we abbreviate the one-step transition functions G_n^1 to G and the m-step transition functions G_n^m to G^m.

Then (Loève (1963), p. 366) for all n, m in N,

$$\text{(B)} \qquad F_{n+m}(C) = \int_Z F_n(dz) G^m(x, y, C)$$

and

$$\text{(C)} \qquad G^{n+m}(x, y, C) = \int_Z G^n(x, y, dz') G^m(x', y', C).$$

For the remainder of this proof, the domain of integration is Z.

(i) From (C), $G^{n+m}(x, y, C) - G^m(x, y, C) = \int G^n(x, y, dz') G^m(x', y', C) - \int G^n(x, y, dz') G^m(x, y, C) = \int G^n(x, y, dz')[G^m(x', y', C) - G^m(x, y, C)]$. Choose $\varepsilon > 0$. By Theorem 2, for this C there exists m_0 in N such that if $m \geqq m_0$, $|G^{n+m}(x, y, C) - G^m(x, y, C)| < \int G^n(x, y, dz') \cdot \varepsilon = \varepsilon$, for all n in N. By Cauchy, $\lim_m G^m(x, y, C)$ exists and by Theorem 2, this limit is a function of C only; call it $F(C)$. $F(C)$ is countably additive by the Nikodym corollary of the Vitali–Hahn–Saks theorem (Dunford and Schwartz (1958), p. 160), and therefore a probability. To show uniformity in (x, y) for a given n and C, choose $\varepsilon > 0$, and choose a particular (x, y) in Z, say (x_0, y_0). By the result just proved, there exists m_0 in N such that if $m \geqq m_0$, $|G_n^m(x_0, y_0, C) - F(C)| < \varepsilon/2$. By Theorem 2, there exists m_1 in N such that if $m \geqq m_1$ then $\sup_{z \in Z} |G_n^m(x_0, y_0, C) - G_n^m(x, y, C)| < \varepsilon/2$. Let $m_2 = \max(m_0, m_1)$. Then for $m \geqq m_2$, and any (x, y) in Z,

$$|G_n^m(x, y, C) - F(C)| \leqq |G_n^m(x, y, C) - G_n^m(x_0, y_0, C)| + |G_n^m(x_0, y_0, C) - F(C)| < \varepsilon.$$

(ii) Setting $m = 1$ and applying $\lim_n$ to Equation (C) gives $F(C) = \lim_n \int G^n(x, y, dz') G(x', y', C)$. For any characteristic function I_C of a set $C \in \mathscr{C}$, $\lim_n \int G^n(x, y, dz') I_C(z') = \lim_n G^n(x, y, C) = F(C) = \int F(dz') I_C(z')$. Therefore for any simple functions $f_j : Z \to [0, 1]$, $\lim_n \int G^n(x, y, dz') f_j(z') = \int F(dz') f_j(z')$. Since $G(\cdot, \cdot, C)$ is non-negative, uniformly bounded by 1, and measurable, choose f_j, j in N, to be a sequence of simple functions increasing uniformly to $G(\cdot, \cdot, C)$ (Kingman and Taylor (1966), p. 104). Choose $\varepsilon > 0$. Then there exists j_0 such that, uniformly in z, $|f_j(z) - G(x, y, C)| < \varepsilon/3$ whenever $j \geqq j_0$ and there exists n' such that $|\int G^n(x, y, dz') f_j(z') - \int F(dz') f_j(z')| < \varepsilon/3$ whenever $n \geqq n'$.

Thus for all $n \geqq n'$,

$$\left|\int G^n(x,y,dz')G(x',y',C)-\int F(dz')G(x',y',C)\right|$$
$$\leqq \left|\int G^n(x,y,dz')G(x',y',C)-\int G^n(x,y,dz')f_j(z')\right|$$
$$+\left|\int G^n(x,y,dz')f_j(z')-\int F(dz')f_j(z')\right|$$
$$+\left|\int F(dz')f_j(z')-\int F(dz')G(x',y',C)\right|<\varepsilon.$$

Thus

$$F(C)=\lim_n\int G^n(x,y,dz')G(x',y',C)$$
$$=\int \lim_n G^n(x,y,dz')G(x',y',C)=\int F(dz')G(x',y',C).$$

(iii) In Equation (B), set $m = 1$, $F_n = F$. Then by (ii), $F_{n+1} = F$. Use induction.

(iv) By (iii), if $F_1 = F$, then the sequence $g(Z_n)$ is stationary. By Theorem 2, it is indecomposable, so that the σ-field of invariant events is $\{\emptyset, Z\}$. The claim then follows from the stationarity theorem (Loève (1963), p. 421).

(v) Result (i) implies weak convergence of $G_n^m(x,y,\cdot)$ to F (as $m \to \infty$), which implies (v) (Billingsley and Topsøe (1967), p. 1).

2.4. *Proof of Theorem* 4.

Lemma 8. Under the assumptions of Theorem 4 (i) the family of maps from Y to $[0,1]$ given by $\{G_n(x,\cdot,C): n \in N,\ x \in X,\ C \in \mathscr{C}\}$ is uniformly equicontinuous; that is, for every $\varepsilon > 0$, there exists $\delta > 0$ such that for any y, y' in Y, if $d(y,y') < \delta$, then for every n in N, x in X, C in $\mathscr{C}$,

$$|G_n(x,y,C)-G_n(x,y',C)|<\varepsilon.$$

Proof. Choose $\varepsilon > 0$. Repeat the argument of Lemma 5 using the uniform equicontinuity with respect to B to establish uniform equicontinuity for all x in X and all $A \times B$ in $\mathscr{A} \times \mathscr{B}$. Extend G_n to uniform equicontinuity for all x in X and all C in $\mathscr{C}$ by repeating the argument of Theorem 1 (ii).

Proof of (i). Choose n in N and $\varepsilon > 0$. Drawing on Lemma 8, repeat the argument of Theorem 2 without initially conditioning on C in $\mathscr{C}$. The argument then concludes uniformly in z_n, z'_n in Z and uniformly in C in $\mathscr{C}$.

Proof of (ii). Every homogeneous S-ergodic chain is exponentially convergent.

2.5. *Proof of Corollary* 1. Uniform positivity. We must confirm that the particular $p_n(x, x')$ defined in (A) satisfies the general conditions assumed in the first paragraph of (v) of Corollary 1. First, $p_n(x, x')$ in (A) is jointly continuous in x and x', and $p_n(x, x') < M < \infty$. Second, let $\delta' = \mu_X(\text{Int}(X)) = (U-L)^2 U^{k-2}(1-L)^{k-1}$. Then $\delta' > 0$. Since $\sup_{x,x' \in \text{Int}(X)} d_X(x, x') < \infty$, $\delta'' \equiv \inf_{x,x' \in \text{Int}(X)} p_n(x, x') > 0$. So if $\delta \equiv \min(\delta', \delta'')$, uniform positivity is satisfied.

Proof that Theorem 1 *applies.* The assumptions of Theorem 1 which remain to be verified in the context of Corollary 1 are that (1) B/y is a uniformly continuous function of y, and (2) the map $(x, A) \rightarrow P_n^m(x, A) = \int_A p(x, x')dx'$ is jointly measurable.

(1) Since $\| x \cdot y \| \neq 0$, the set $B/y = \{x \in X : x \cdot y / \| x \cdot y \| \in B\}$ is a continuous function from Y to $\mathcal{A}$ by inspection. (Y, d) is a compact metric space. If $\rho(A, A') = \mu_X(A \triangle A')$, for A, A' in $\mathcal{A}$, then $(\mathcal{A}, \rho)$ is a pseudometric space. The proof of Theorem 2.4 of Kingman and Taylor ((1966), p. 37) extends to a range space which is a pseudometric space so B/y is uniformly continuous. Thus given B in $\mathcal{B}$, for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all y, y' in Y, if $\| y - y' \| < \delta$ then $\mu_X((B/y) \triangle (B/y')) < \varepsilon$.

(2) The map $(x, A) \rightarrow P_n^m(x, A)$ is jointly continuous, hence jointly measurable.

Proof that Theorem 2 *applies.* The assumptions of Theorem 2 which remain to be verified in the context of Corollary 1 are that (1) $\{X_n\}$ is S-ergodic, and (2) for all $\delta > 0$ there exists m_0 such that for all $m \geqq m_0$ and for all $x_1, \cdots, x_m \in X$, $d(x_m \cdots x_1 y_0, x_m \cdots x_1 y_0') < \delta$.

(1) Uniform positivity is an obvious analog of the generalized Markov condition (Loève (1963), p. 369) for homogeneous chains on general state spaces. A calculation exactly parallel to Loève's shows that $\Delta_{nm} \leqq (1-\delta^2)^m$. Thus assumption (v) of Corollary 1 guarantees that $\{X_n\}$ is not merely S-ergodic, but is exponentially convergent (Loève (1963), p. 367), even when $\{X_n\}$ is not homogeneous. Here $\delta \equiv \min(\delta', \delta'')$ as in 2.5 above.

(2) The weak ergodic theorem of demography is proved with elegance by Golubitsky, Keeler and Rothschild ((1975), p. 89). This theorem implies (2).

Proof that Theorem 3 *applies.* The transition probability density function in (A) is homogeneous. Theorem 3 (v) applies since $Z = X \times Y$ with the metric ρ_Z is a separable metric space and $\mathcal{C}$ is the Borel σ-field.

For ω in Ω, total population size changes from time n to time $n+1$ by the factor $\lambda_n(\omega) = \| X_{n+1}(\omega) \cdot Y_n(\omega) \| / \| Y_n(\omega) \|$. ($\cdot$ means matrix–vector multiplication.) By construction of $Y_n(\omega)$, $\| Y_n(\omega) \| \neq 0$ surely so $\lambda_n(\omega)$ is defined, and $\lambda_n(\omega)$ is bounded surely by construction. Then $\lambda(z_n) \equiv E_\omega(\lambda_n(\omega) \mid Z_n = (x_n, y_n)) = \int_X (\| x \cdot y_n \| / \| y_n \|) P_n(x_n, dx)$ is a bounded, positive, continuous function of z_n which gives the expected factor of change in population size from n to $n+1$ conditional on Z_n. If only the initial conditions Z_1 are known, then

$$E_\omega(\lambda_n(\omega) \mid Z_1 = z_1) = \int_Z \int_X (\|x \cdot y_n\| / \|y_n\|) P_n(x_n, dx) G^{n-1}(x_1, y_1, dz_n)$$

is the expected factor of change from n to $n+1$. In the homogeneous case, $\lim_n E_\omega(\lambda_n(\omega) \mid Z_1 = z_1) = \int_Z \lambda(z) F(dz) = \lambda$.

If $F_1 = F$, then since $\{Z_n\}$ is a stationary ergodic sequence of random vectors, the sequence of random variables $\{\lambda_n\}$ is also stationary and ergodic (Breiman (1968), pp. 105, 119), hence $\lim_n n^{-1} \Sigma_{j=0}^{n-1} \lambda_j(\omega) = \lambda$ almost surely (ω).

2.6. *Proof of Corollary* 2. The analog of the weak ergodic theorem of demography required by Theorem 2 is immediate from Lemma 3 of Hajnal ((1958), p. 237).

2.7. *Proof of Corollary* 3.

(i) Using counting measure on X and Lebesgue $(s_1 - 1)$-measure on Y, the measurability of $G_n(\cdot, \cdot, A \times B)$ is immediate, and $G_n(x, y, \cdot)$ is obviously a probability. The extension to $\mathscr{C}$ is immediate since X and Y are separable metric spaces.

(ii) By proof of Theorem 1 (iii).

(iii) $P_n(x, \cdot)$ is defined in terms of a transition density $t_n(x, \cdot)$ which is uniformly bounded by 1. Theorem 3 of Hajnal (1976) implies both that $\{X_n\}$ is S-ergodic (in fact, exponentially convergent) and that the analog of the weak ergodic theorem of demography assumed in our Theorem 2 also holds. Because X is finite, our Lemmas 4 and 5 are trivial, and do not require the continuity assumptions of our Theorem 1. The result then follows by the arguments for Lemmas 6 and 7 and Theorem 2.

(iv) By proof of Theorem 3.

Acknowledgements

I thank J. Hajnal, D. G. Kendall and B. D. Ripley for extremely helpful suggestions and guidance. B. D. Ripley's sustained contributions to several formulations and proofs are appreciated. I enjoyed the hospitality of King's College Research Centre and the Statistical Laboratory of the University of Cambridge during 1974–75. I thank Harvard University, the U.S. National Science Foundation, and King's College Cambridge for support, and the University of Cambridge Computing Service for computing facilities.

References

BILLINGSLEY, P. (1968) *Convergence of Probability Measures.* Wiley, New York.

BILLINGSLEY, P. AND TOPSØE, F. (1967) Uniformity in weak convergence. *Z. Wahrscheinlichkeitsth.* 7, 1–16.

BLUMENTHAL, R. M. AND GETOOR, R. K. (1968) *Markov Processes and Potential Theory.* Academic Press, New York.

BRASS, W. (1974) Perspectives in population prediction: illustrated by the statistics of England and Wales. *J. R. Statist. Soc.* A **137**, 532–583.

COHEN, J. E. (1976) Ergodicity of age structure in populations with Markovian vital rates. I. Countable states. *J. Amer. Statist. Assoc.* **71**, 335–339.

COHEN, J. E. (1977) Ergodicity of age structure in populations with Markovian vital rates. III. Mean and approximate variance. *Adv. Appl. Prob.* **9** (3).

DOBRUSHIN, R. L. (1956) Central limit theorem for nonstationary Markov chains. I, II. *Theor. Prob. Appl.* **1**, 65–79, 329–383.

DUNFORD, N. AND SCHWARTZ, J. T. (1958) *Linear Operators. I: General Theory.* Interscience, New York.

GOLUBITSKY, M., KEELER, E. B. AND ROTHSCHILD, M. (1975) Convergence of the age-structure: applications of the projective method. *Theoret. Pop. Biol.* **7**, 84–93.

GRIFFEATH, D. (1975) A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitsth.* **31**, 95–106.

HAJNAL, J. (1956) The ergodic properties of nonhomogeneous finite Markov chains. *Proc. Camb. Phil. Soc.* **52**, 67–77.

HAJNAL, J. (1958) Weak ergodicity in nonhomogeneous Markov chains. *Proc. Camb. Phil. Soc.* **54**, 233–246.

HAJNAL, J. (1976) On products of non-negative matrices. *Math. Proc. Camb. Phil. Soc.* **79**, 521–530.

KELLEY, J. L. (1955) *General Topology.* Van Nostrand, Princeton, N.J.

KINGMAN, J. F. C. AND TAYLOR, S. J. (1966) *Introduction to Measure and Probability.* Cambridge University Press.

LOÈVE, M. (1963) *Probability Theory*, 3rd edn. Van Nostrand, Princeton, N.J.

PITMAN, J. W. (1974) Uniform rates of convergence for Markov chain transition probabilities. *Z. Wahrscheinlichkeitsth.* **29**, 193–227.

TAKAHASHI, Y. (1969) Markov chains with random transition matrices. *Kodai Math. Seminar Rep.* **21**, 426–447.