SIAM J. APPL. MATH.
Vol. 36, No. 1, February, 1979

0036-1399/79/3601-0016 $01.00/0

THE CUMULATIVE DISTANCE FROM AN OBSERVED TO A STABLE AGE STRUCTURE*

JOEL E. COHEN†

Abstract. In a closed, unisexual, age-structured population with age-specific birth and death rates which are constant in time, the vector describing a census by age categories will, as time increases, approach proportionality to the stable age structure implied by the vital rates. This stable age structure is the dominant eigenvector of a demographic projection matrix which carries out the action on a census vector of the age-specific vital rates. We show that the elementwise convergence to zero of the discounted deviations from the stable age structure is complete and exponential. The sum, over all time, of the signed discounted deviations may be easily calculated from a fundamental matrix based on the projection matrix. These results are proved for any primitive nonnegative square matrix. In the demographic context, these results suggest alternatives to an index which has been used to measure the distance from an observed to a stable age structure.

1. Introduction and results. The theory of stable populations considers closed, unisexual, age-structured populations. Individuals are subject to birth and death rates which depend on age and are constant in time (Keyfitz [8], Coale [2]). Suppose time and age are measured in the same discrete unit. Let K be a census vector, that is, a nonnegative, nonzero column n-vector giving the number of individuals in each of n age classes at the present time, taken as $t=0$. Let M be an $n\times n$ projection matrix. The i,jth element of M measures the contribution to age class i at time $t+1$ of one individual in age class j at time t. We assume only that M is nonnegative and primitive (i.e. that every element of M^t is positive, for some t). These assumptions hold for at least every observed human population if age classes beyond the last age of reproduction are excluded. It is a well-known consequence of the Perron–Frobenius theorem that M has a simple real root (eigenvalue) λ with modulus larger than the modulus of every other root and that

$$\lim_{t\to\infty}\lambda^{-t}M^t=B \tag{1}$$

where B is a finite positive $n\times n$ matrix. Consequently if

$$K(t)=M^tK,\qquad t=0,1,2,\cdots, \tag{2}$$

is the result of projecting the census vector K forward t time units via M, then, as the strong ergodic theorem of demography observes,

$$\lim_{t\to\infty}\lambda^{-t}[K(t)-\lambda^tBK]=0. \tag{3}$$

The zero on the right of (3) is an n-vector with each element zero.

Let the sum of the first t terms of the sequence on the left of (3) be

$$S_t(M,K)=\sum_{m=0}^{t-1}\lambda^{-m}M^mK-tBK,\qquad t=1,2,\cdots. \tag{4}$$

* Received by the editors February 15, 1978, and in revised form June 30, 1978.
† Rockefeller University, New York, New York 10021. This work was supported in part by the National Science Foundation under Grant BMS74-13276.

We shall show that the terms in (3) approach 0 so rapidly that the sequence of sums in (4) approaches a finite limit as $t \to \infty$. This limit may be calculated directly:

$$\lim_{t\to\infty} S_t(M, K) = (Z - B)K \tag{5}$$

where

$$Z = (I + B - M/\lambda)^{-1}. \tag{6}$$

We shall prove that Z exists and satisfies (5). Moreover, the total number of individuals in age class i at time i resulting from an initial population consisting of only one individual in age class j and no others at time 0 is $M_{ij}^{(t)}$, the i,jth element of M^t, $i, j = 1, \cdots, n$, and

$$\lim_{t\to\infty} \sum_{i=1}^{n} \sum_{m=0}^{t-1} \lambda^{-m} M_{ii}^{(t)} - t = \textstyle\sum_{i=1}^{n} z_{ii} - 1. \tag{7}$$

Here z_{ii} is the ith diagonal element of Z.

In the sequence of n-vectors in the brackets on the left of (3), the ith element may be sometimes positive, sometimes negative. Hence in the sums (4), the ith element may be the sum of positive and negative terms. The finiteness of the limiting sum does not depend on any cancellation of positive and negative terms. Specifically, let

$$R_t(M, K) = \sum_{m=0}^{t-1} \lambda^{-m} |M^m K - \lambda^m BK|, \qquad t = 1, 2, \cdots, \tag{8}$$

where $|M^m K - BK|$ is an n-vector in which the ith element is the absolute value of the ith element of the n-vector $M^m K - BK$. We shall show that

$$\lim_{t\to\infty} R_t(M, K) < \infty. \tag{9}$$

The limit in (9), and hence in (5), is approached exponentially fast. We know no simple exact expression for this limit, unlike the limit in (5).

If

$$k = K \Big/ \sum_{i=1}^{n} K_i \tag{10}$$

is the age *structure* associated with the census vector K, the observations (5) and (9) suggest that $\lim S_t(M, k)$ and $\lim R_t(M, k)$ provide natural scale-invariant vector measures of the cumulative distance from the age structure k of the census K to the stable age structure associated with M.

In the next section of this note, we verify the limits (5) and (9) with a numerical example. Based on these limits, we propose indexes to measure the departure of an observed from a stable age structure. We compare numerically the behavior of these indexes with that of an index of dissimilarity Δ used by Keyfitz [8, p. 47] and Keyfitz and Flieger [9] as a measure of the departure of an observed from a stable age structure.

The final section of this note proves the properties of Z claimed in (5), (6), (7), the inequality (9), and some additional properties for nonnegative primitive matrices generally, without restriction to the form of population projection matrices.

2. Numerical illustration. The 1965 female population of the United States included $K_1 = 29{,}413$ thousand girls under 15; $K_2 = 20{,}886$ thousand women 15–29; and 18,040 thousand women 30–44 (Keyfitz [8, p. 56]). Using the same age categories

and a 15-year time unit, the projection matrix for this population, based on adjusted births, is (Keyfitz [8, p. 42])

$$M = \begin{pmatrix} 0.4271 & 0.8498 & 0.1273 \\ 0.9924 & 0 & 0 \\ 0 & 0.9826 & 0 \end{pmatrix}. \tag{11}$$

All subsequent calculations, based on these data, have been carried out with 10 significant figures and rounded only for reporting. In the stable age structure k^* of M, the three age classes would occur in the proportions $k_1^* = 0.4020$, $k_2^* = 0.3299$, $k_3^* = 0.2681$. The matrix B and growth rate λ are given correctly to the number of places shown by Keyfitz [8, p. 43; his Z_1 = our B, his λ_1 = our λ]. From (6),

$$Z = \begin{pmatrix} 0.8494 & 0.1607 & 0.0281 \\ 0.2191 & 0.7551 & -0.0272 \\ -0.2103 & 0.3075 & 0.9370 \end{pmatrix}. \tag{12}$$

Direct computation from (4) and (8) gives $S_t(M, K)$ and $R_t(M, K)$ for $t = 1, 2, \cdots, 18$ (Table 1). $S_t(M, K)$ shows a rapid damped oscillatory convergence to the limit $(Z - B)K$, which coincides with $S_{16}(M, K)$ to the accuracy shown. As $S_t(M, K)$ settles at $(Z - B)K$, $R_t(M, K)$ ceases to change.

We now consider these limits as a basis for measures of the distance between an observed age structure k and its stable limit k^*.

In a series of papers starting in 1955, O. Dudley Duncan and Beverly Duncan introduced into demography an index of dissimilarity Δ to measure the difference between two age structures k and k'. k and k' are nonnegative vectors, the elements of each of which sum to 1. Δ is the sum of the positive differences of the percentages in each age class, which is equivalent to

$$\Delta = (100/2) \sum_{i=1}^{n} |k_i - k_i'|. \tag{13}$$

This index was extensively used in studies of demographic segregation. For references to primary sources and applications, see Shryock and Siegel [12, I: 179, 232–233, 262]. Except for the factor 100 which converts from fractions to percentages, this index of dissimilarity is standard in probability theory as a measure of the difference between two probability measures (Loève [10, p. 367]).

Keyfitz [8, p. 47] and Keyfitz and Flieger [9] used Δ to compare, not two different observed age structures, but an observed age structure k and its limiting stable age structure k^*. For this purpose we suggest two alternative measures

$$D_1 = 50 \sum_{i=1}^{n} |([Z - B]k)_i|, \tag{14}$$

where $([Z - B]k)_i$ is the ith element of the n-vector $(Z - B)k$, and

$$D_2 = 50 \sum_{i=1}^{n} \lim_{t \to \infty} |R_t(M, k)|_i. \tag{15}$$

The arbitrary numerical constant 50 is introduced solely in order to adjust these two indexes to a magnitude close to that of Δ.

For the 1965 female population of the United States, using the three age groups given, $\Delta = 2.8355$ when k' is replaced by the stable age structure k^*, $D_1 = 2.5922$ and (taking $t = 25$ as a sufficient approximation to ∞) $D_2 = 9.0700$.

The reason for suggesting D_1 and D_2 becomes clearer if Δ, D_1 and D_2 are compared for the same M and k^*, but in place of k, a plausible hypothetical age structure k' with the proportions $k'_1 = 0.42$, $k'_2 = 0.30$, $k'_3 = 0.28$. Now Δ increases to 2.9905 but D_1 decreases to 1.5160 and D_2 decreases to 8.5280. The index Δ suggests that the observed $k = (.4304, .3056, .2640)$ is closer to the stable k^* than is the hypothetical k', while D_1 and D_2 suggest the opposite.

Δ measures the distance from k to k^* as the demographic crow flies, without regard to the trajectory of age structures through which the population must pass as it approaches stability. Thus for a given k and k^*, Δ would be the same regardless of the projection matrix, as long as the projection matrix had the same stable age structure k^*.

D_1 and D_2 are more like charts of the highway mileage or travel time between k and k^* since they cumulate over all time, with appropriate discounting, the signed or absolute discrepancies between $K(t)$ and the stable population. We suggest that for some purposes D_1 and D_2 may be demographically more meaningful indexes than Δ.

D_1 and D_2 appear to depend on the number n of age classes. For certain applications, such as comparisons across species, it might be desirable in the future to develop measures independent of n.

TABLE 1

Cumulated signed and absolute differences between projections of the 1965 *United States female population and the stable age distribution.*

Time t	$S_t(M, K)$ (in thousands) Age group			$R_t(M, K)$ (in thousands) Age group		
	0–14	15–29	30–44	0–14	15–29.	30–44
1	1589	−1947	−512	1589	1947	512
2	728	−643	−2094	2450	3251	2094
3	1173	−1350	−1035	2895	3957	3153
4	945	−984	−1609	3123	4323	3728
5	1061	−1171	−1312	3239	4510	4025
6	1002	−1076	−1464	3298	4605	4176
7	1032	−1125	−1386	3328	4653	4254
8	1017	−1100	−1426	3343	4678	4293
9	1025	−1112	−1406	3351	4691	4313
10	1021	−1106	−1416	3355	4697	4323
11	1023	−1109	−1411	3357	4700	4329
12	1022	−1108	−1413	3358	4702	4331
13	1022	−1109	−1412	3359	4703	4333
14	1022	−1108	−1413	3359	4703	4333
15	1022	−1108	−1412	3359	4703	4334
16	1022	−1108	−1413	3359	4704	4334
17	1022	−1108	−1413	3359	4704	4334
18	1022	−1108	−1413	3359	4704	4334

3. The fundamental matrix of a primitive matrix. Theorems 1 to 3 below generalize results known for the transition matrix of a regular Markov chain (Kemeny and Snell [7, pp. 75–77]) to any primitive nonnegative matrix T with dominant eigenvalue λ. Theorem 4 may be new even in the context of regular Markov chains. The results claimed in the first part of this note follow by specializing T to M.

From Seneta [11, p. 7] and Karlin and Taylor [6, p. 546], it is known that there exists a positive matrix B such that $\lim_{t\to\infty}(\lambda^{-1}T)^t - B = 0$. For any integer $t>0$, $B^t = B$. For some positive column vector x such that $Tx = \lambda x$ and for some positive row vector w^{T}, $w^{\mathrm{T}}x = 1$ and $xw^{\mathrm{T}} = B$. (Throughout, the superscript ^{T} means transpose.) Also $TB = BT = \lambda B$.

THEOREM 1. *$Z = (I + B - \lambda^{-1}T)^{-1}$ exists and is given by*

$$Z = I + \sum_{t=1}^{\infty} ([\lambda^{-1}T]^t - B). \tag{16}$$

The proof repeats the proof of Theorem 4.3.1 of Kemeny and Snell [7, p. 75] step for step, with $\lambda^{-1}T$ replacing their stochastic matrix P.

To obtain (5) from Theorem 1, subtract B from both sides of (16) and multiply both sides on the right by K.

By analogy with the convention in the theory of Markov chains, we call Z the fundamental matrix of the primitive matrix T. Z has properties which may prove useful.

THEOREM 2. *With T, x, Z, λ, and B as above,* (a) $TZ = ZT$, (b) $Zx = x$, (c) $I - Z = B - \lambda^{-1}TZ$.

The proof repeats the steps of the proof of Theorem 4.3.3 of Kemeny and Snell [7, pp. 75–76]. Contrary to Theorem 4.3.3(b) in [7], when $\lambda^{-1}T$ is not a stochastic matrix, the row sums of Z are not in general 1, as the example of Z in (12) shows.

If T is of rank 1, then $Z = I$. The proof is that if T is primitive and of rank 1 with positive eigenvector x corresponding to λ, then for some positive n-vector y, $T = xy^T$. Then $\lambda B = BT = (xw^T)(xy^T) = x(w^Tx)y^T = T$, which implies $B = \lambda^{-1}T$. Thus $Z = (I + B - \lambda^{-1}T)^{-1} = I^{-1} = I$.

Recall that the trace of a matrix is the sum of the elements on its main diagonal. Thus trace $(B) = 1$. Let $T_{ij}^{(m)}$ be the i, jth element of T^m.

THEOREM 3.

$$\lim_{t\to\infty} \sum_{m=0}^{t-1} \lambda^{-m} \operatorname{trace}(T^m) - t = \operatorname{trace}(Z) - 1. \tag{17}$$

Equation (7) *follows.*

To prove Theorem 3, let w_j be the jth element of w (recall that $B = xw^{\mathrm{T}}$). Then $(Z-B)_{jj} = z_{jj} - x_jw_j$. Summing over j gives $\sum_{j=1}^n (Z-B)_{jj} = \sum z_{jj} - 1$. But from Theorem 1, $\lim_{t\to\infty}\sum_{m=0}^{t-1}\lambda^{-m}T_{jj}^{(m)} - tx_jw_j = (Z-B)_{jj}$. Again summing over j and exchanging the limit with the finite sum gives (17). Eugene Seneta (personal communication) has another proof.

THEOREM 4. *There exist positive constants a and b such that*

$$\lambda^{-m}|T_{ij}^{(m)} - \lambda^m B_{ij}| \leqq a\, e^{-bm}, \qquad m = 0, 1, 2, \cdots. \tag{18}$$

Specifically, let $\lambda_2 \neq \lambda$ be an eigenvalue of T whose modulus $|\lambda_2|$ is exceeded by the modulus of no eigenvalue of T other than λ. If $|\lambda_2| > 0$, pick any ε such that $0 < \varepsilon < \log(\lambda/|\lambda_2|)$. Then for some $a > 0$, $b = \log(\lambda/|\lambda_2|) - \varepsilon > 0$ satisfies (18).

(If $|\lambda_2| = 0$, T is of rank 1 and $\lambda^{-1}T = B$, so the limit in (3) is attained as soon as $t = 1$. We therefore suppose $|\lambda_2| > 0$.)

To prove Theorem 4, let $f(m) = \lambda^{-m}|T_{ij}^{(m)} - \lambda^m B_{ij}|$. Then we claim there exist positive constants a and b such that (18) holds if there exists an integer $m_0 \geqq 0$ and positive constants a' and b such that $f(m) \leqq a'\, e^{-bm}$ for $m = m_0, m_0+1, \cdots$. To prove this, let $A = \max_{m=0,\cdots,m_0-1} f(m)$. A is finite and positive. Choose $a = \max(A\, e^{+bm_0}, a')$. Then $a\, e^{-bm} \geqq A \geqq f(m)$ for $m = 0, 1, \cdots, m_0 - 1$ and $a\, e^{-bm} \geqq a'\, e^{-bm}$ for $m =$

$m_0, m_0+1, \cdots$. Thus it suffices to find an exponentially decaying upper bound $a' e^{-bm}$ of $f(m)$ which holds for all but at most a finite number m_0 of terms.

Let λ_2 be the eigenvalue with largest algebraic multiplicity in the minimal polynomial of T among those eigenvalues with modulus equal to $|\lambda_2|$. We observe from the spectral decomposition of powers of a matrix (Gantmacher [3, I:104, his eq. (17)]; Seneta [11, p. 7]) that $T_{ij}^{(m)}$ is a polynomial of some terms of the form constant $\cdot\, m^{s_h}\lambda_h^m$ where λ_h runs over the set of eigenvalues of T ($\lambda_1 = \lambda$), and s_h runs from 0 up to one less than the multiplicity of λ_h as a root of the minimal polynomial of T ($s_1 = 0$ only). Here the constant coefficient may absorb some power of λ_h which is independent of m. Thus $T_{ij}^{(m)} = \lambda^m B_{ij} + cm^s\lambda_2^m +$ other terms of the form $c'(m')^{s'}\lambda_h^{m'}$, $h \geqq 2$, where if $|\lambda_h| = |\lambda_2|$ then $s' < s$. Thus $f(m) = \lambda^{-m}|T_{ij}^{(m)} - \lambda^m B_{ij}| \leqq |c|m^s(|\lambda_2|/\lambda)^m +$ other terms of smaller magnitude (when m is large enough) than $|c|m^s(|\lambda_2|/\lambda)^m$.

Now $m^s(|\lambda|/\lambda)^m \leqq a^* e^{-bm}$ if $s \log m + m \log(|\lambda_2|/\lambda) \leqq \log a^* - bm$ or $s \log m \leqq \log a^* - m(\log(|\lambda_2|/\lambda) + b)$. For $m > 1$, $\log m > 0$ and division by $\log m$ preserves the sense of the last inequality, giving $s \leqq (\log a^*)/(\log m) - m(\log(|\lambda_2|/\lambda) + b)/\log m$. Choosing $a^* = 1$ and $b = \log(\lambda/|\lambda_2|) - \varepsilon > 0$, the latest inequality certainly holds if $s \leqq \varepsilon m/\log m$, where $\varepsilon > 0$, or if $s/\varepsilon \leqq m/\log m$. It is well known that for any fixed $s/\varepsilon > 0$, there exists an integer $m_0 > 1$ such that $s/\varepsilon \leqq m/\log m$ for all $m \geqq m_0$. If necessary, m_0 can be adjusted upward to assure that all remaining terms in the spectral expansion of $T_{ij}^{(m)}$ are of smaller magnitude than $|c|m^s(|\lambda_2|/\lambda)^m$ for $m \geqq m_0$. Then choose $a' \geqq |c|$ to allow for these additional terms as well so that $f(m) \leqq a' e^{-bm}$ for $m \geqq m_0$. Given b, a', and m_0, we may find the required constants a and b.

From Theorem 4, (9) follows by summing (18) over m and multiplying by the census vector K.

Birkhoff [1, Chap. 16] proved that the limiting eigenvector is approached exponentially fast in the Hilbert projective pseudo-metric under the action of a *positive* matrix. He also gave (his Corollary 2, p. 385) an explicit upper bound for $|\lambda_h|/\lambda$, $h \geqq 2$. Golubitsky, Keeler and Rothschild [4] and Hajnal [5] greatly extended the conditions under which exponential convergence in the projective pseudo-metric holds. Theorem 4 and (9) prove complete exponential convergence of the absolute differences elementwise, rather than of the Hilbert projective pseudo-metric.

Acknowledgments. I thank Nathan Keyfitz, Martin Schwarz, Eugene Seneta, J. Laurie Snell and Anne Whittaker for help. The referee and editor made many valuable corrections and suggestions.

REFERENCES

[1] G. BIRKHOFF, *Lattice Theory*, 3rd ed., American Mathematical Society, Providence, RI, 1967.

[2] A. J. COALE, *The Growth and Structure of Human Populations: A Mathematical Investigation*, Princeton University Press, Princeton, NJ, 1972.

[3] F. R. GANTMACHER, *The Theory of Matrices*, vol. 1, Chelsea, New York, 1960.

[4] M. GOLUBITSKY, E. B. KEELER AND M. ROTHSCHILD, *Convergence of the age structure: Applications of the projective method*, Theor. Popul. Biol., 7 (1975), pp. 84–93.

[5] J. HAJNAL, *On products of non-negative matrices*, Math. Proc. Cambridge Philos. Soc., 79 (1976), pp. 521–530.

[6] S. KARLIN AND H. M. TAYLOR, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York, 1975.

[7] J. G. KEMENY AND J. L. SNELL, *Finite Markov Chains*, Springer, New York, 1976 (reprint of 1960 ed.).

[8] N. KEYFITZ, *Introduction to the Mathematics of Population*, Addison-Wesley, Reading, MA, 1968.

[9] N. KEYFITZ AND W. FLIEGER, *World Population*, University of Chicago Press, Chicago, 1968.

[10] M. LOÈVE, *Probability Theory*, Van Nostrand, Princeton, NJ, 1963.
[11] E. SENETA, *Non-negative Matrices: An Introduction to Theory and Applications*, Allen & Unwin, London, 1973.
[12] H. S. SHRYOCK AND J. S. SIEGEL, *The Methods and Materials of Demography*, U.S. Government Printing Office, Washington, DC, 1973.