DEMOGRAPHY Volume 7, Number 3 August 1970

STABLE AGE BY REGION DISTRIBUTIONS

Griffith M. Feeney
Department of Demography, University of California, Berkeley 94720

Abstract—If the pattern of fertility, mortality and interregional migration exhibited by the United States population during 1950–60 were to continue in the future, the proportions of persons in the various age groups and regions would fluctuate from decade to decade. These fluctuations would become less marked with time, however, and eventually all the proportions would stabilize at certain fixed values. This collection of values may be called a *stable age by region composition* corresponding to the given schedule of fertility, mortality and migration. The same phenomenon may be observed when individuals move between socioeconomic categories as, for example, socioeconomic status or educational attainment levels. The substantial differences between these various situations conceal remarkable similarities. In each case the continued operation of schedules of fertility, mortality and mobility between categories may result in a stable composition. The purpose of this paper is to shed some light on the nature of these stable compositions, on the interrelation between their various components, and on their relation to the patterns of fertility, mortality and mobility which generate them.

1. Generalized Leslie Matrices

In the discrete formulation of stable population theory the age distribution of a population at any given time is represented by a vector whose components give the numbers of persons in each age group. The fertility-mortality schedule exhibited by a population during any given time period is represented by a matrix whose entries describe the fertility and mortality of each age group. See Keyfitz (1968, Part II), for a development of these subjects and for references. Stable population theory may be extended to incorporate the regional distribution and interregional migration schedule of a population in the following manner. In place of a vector whose components are *numbers*, one considers a vector whose components are *vectors*, each vector giving the distribution by region of residence of persons in some age group. In place of a matrix whose entries are numbers, one considers a matrix whose entries are matrices describing both the fertility and mortality in each region and the migration between regions of each age group. This section elaborates on these statements and gives definitions and notation used in following sections.

The following identity applies to any population closed to migration. If x_1, x_2, . . . and y_1, y_2, . . . denote the numbers of persons in successive n year age groups at two points in time n years separated, then

$$1)\quad \begin{bmatrix} s'b_1 & s'b_2 & \cdot & \cdot & \cdot \\ s_1 & 0 & \cdot & \cdot & \cdot \\ 0 & s_2 & & & \\ \vdots & & \cdot & & \\ \vdots & & & \cdot & \cdot \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix},$$

where: a) s_i $(i = 1, 2, \ldots)$ denotes the proportion of persons in the ith n year age group at the beginning of the time period who survive to the end of the time period; b) b_i $(i = 1, 2, \ldots)$ denotes the number of births to the population during the time period to females in the ith n year age group at the beginning of the time period divided by the number of females in the ith n year age group at the beginning of the time period; and, c) s' is the proportion of births to the population during the time period who survive to the end of the time period. For notational convenience, let m_i denote $s'b_i$ $(i = 1, 2, \ldots)$. This convention will be maintained throughout this paper. The quantity m_i may be thought of as an 'effective' birth rate which includes a component for survival from birth during a time period to the end of the time period.

In the event that at the beginning of a time period a given population contains no persons in some n year age group, the survival and birth statistics for this age group will be undefined by the above definitions. There is evidently no reason to attempt to define the birth or death statistics in this case, however, for in the absence of persons in the age group one has no information on the fertility or mortality of this age group. The number n in the above definitions is arbitrary so long as females aged n years or less give no births during the time period.

Matrices of the form appearing in 1) may be called *Leslie* matrices. The definitions above are applicable to the female population only, to the male population only, or to the total population. In the first case the phrase "births to the population" indicates female births, in the second case male births, and in the last case all births. The same remarks apply to the definitions which follow.

Consider a population closed to migration and divided into N regions. The identity 1) may be extended as follows to incorporate regional distributions and interregional migration schedules of this population. If $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$ denote N-dimensional vectors giving the distribution by region of residence of persons in successive n year age groups at two points in time n years separated, then

$$2) \quad \begin{bmatrix} R_1'S'B_1 & R_2'S'B_2 & \cdot & \cdot & \cdot \\ R_1S_1 & 0 & \cdot & \cdot & \cdot \\ 0 & R_2S_2 & & & \\ \vdots & & \cdot & & \\ \vdots & & & \cdot & \cdot \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \\ \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \\ \end{bmatrix},$$

where: a) S_i, B_i $(i = 1, 2, \ldots)$ and S' denote $N \times N$ diagonal matrices with diagonal entries defined as for 1) for each of the N regions; b) R_i $(i = 1, 2, \ldots)$ denotes the $N \times N$ matrix whose jkth entry denotes the proportion of survivors at the end of the time period of persons in the ith n year age group and residing in the kth region at the beginning of the time period who reside in the jth region at the end of the time period; and c) R'_i $(i = 1, 2, \ldots)$ denotes the $N \times N$ matrix whose jkth entry denotes the proportion of survivors at the end of the time period of births to the population during the time period to females in the ith n year age group and residing in the kth region at the beginning of the time period who reside in the jth region at the end of the time period.

Matrices of the form appearing in 2) may be called *generalized Leslie* matrices.

In the event that at the beginning of a time period a given population contains no persons in some n year age group and region, the corresponding birth, survival and migration statistics will be undefined by the above definitions. In some cases this will indicate, just as for Leslie matrices, a lack of information on the fertility, mortality and migration of a particular age group and region. It may happen, however, that when applying these definitions to variables other than region of residence, certain combinations of categories and age groups are logically vacuous. For example, if a female population is classified by parity (number of live births), there will never be any females in the first five year age group and of the parity classification "three live births." When the combination of the ith n year age group and the kth category is logically vacuous, one may formally define the jkth entries $(j = 1, \ldots, N)$ of the matrix R_i by: the kkth entry of R_i is unity and the jkth entry $(j \neq k)$ is zero. This quite artificial convention has the useful consequence that the matrices R_i are undefined only through lack of information and are, when defined, always column stochastic. (A matrix is *column stochastic* if its various columns sum to unity.) The latter property plays a role in the proof of section 4 below. The same convention may be applied to the matrices R_i'.

2. RELATION TO MULTIREGIONAL MATRIX GROWTH OPERATORS

The pattern of nonzero entries in a generalized Leslie matrix is quite unlike that in Rogers' (1968, Chapter 2) multiregional matrix growth operators. To see the relation between the two, consider first the identity 1) of the preceding section. The components of the vectors in 1) give the numbers of persons in the various n year age groups in ascending order. However it is not necessary that they be given in this order. For example,

$$\begin{bmatrix} m_1 & m_2 & m_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & 0 & s_2 \\ m_3 & m_1 & m_2 \\ 0 & s_1 & 0 \end{bmatrix} \begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_3 \\ y_1 \\ y_2 \end{bmatrix}$$

are simply two different ways of writing the same three scalar identities, namely: $y_1 = m_1x_1 + m_2x_2 + m_3x_3$, $y_2 = s_1x_1$ and $y_3 = s_2x_2$. Note however that the reordering of the components of the vectors is accompanied by a change in the pattern of nonzero entries of the matrix. It is the same phenomenon that gives rise to the different patterns of nonzero entries in generalized Leslie matrices and multiregional matrix growth operators.

Since there is evidently no reason to use any ordering of the age groups other than ascending in 1) this notational slight of hand is of no particular interest in this case. However when age by region distributions are to be written as vectors the situation is quite different. In this case there are two "natural" ways to order the components, viz.: i) by listing the numbers of persons in successive age groups in the first region, followed by the numbers in successive age groups in the second region, and so on (this is the ordering used by Rogers); and, ii), by listing the numbers of persons in each region in the first age group, followed by the numbers in each region in the second age group, and so on (this is the ordering used here). The first ordering results in the pattern of nonzero entries exhibited by multiregional matrix growth operators, the second in the pattern exhibited by generalized Leslie matrices. As an example of these two orderings, compare the multiregional matrix growth operator in Figure

1 of Rogers (1966) and the corresponding generalized Leslie matrix in section 5 below.

These two orderings, then, yield two different ways of writing the same set of scalar equations. Since all that is involved is a change of notation, the distinction obviously has no direct demographic significance. However the pat*tern of nonzero entries in generalized* Leslie matrices lends itself particularly well to certain formal manipulations. The results of the following two sections depend crucially on the exploitation of this pattern.

3. STABLE AGE BY REGION DISTRIBUTIONS

What can be said about the successive age by region compositions of a system of regions (the totality of which is closed to migration) which exhibits the same fertility-mortality and interregional migration schedule for a succession of time periods? The answer to the analogous question for a single region is given in the following two assertions. If a population closed to migration exhibits the Leslie matrix

$$P = \begin{bmatrix} m_1 & m_2 & \cdot & \cdot & \cdot & m_{n-1} & m_n \\ s_1 & 0 & \cdot & \cdot & \cdot & & \\ 0 & s_2 & & & & & \\ \vdots & & \ddots & & & & \\ & & & & & s_{n-1} & \end{bmatrix}$$

for a succession of time periods, the age composition of this population tends (under quite general conditions) to a certain vector, x, called a "stable age composition." (For notational convenience, let p_i denote $s_i s_{i-1} \ldots s_1$ $(i = 1, 2, \ldots)$ and put $p_0 = 1$. This convention will be maintained throughout this paper. The quantity p_i may be thought of as a cumulative survival proportion for survival from the first n year age group to the $(i + 1)$st n year age group for a birth cohort which exhibits the survival proportion s_i as it passes from the ith to the $(i + 1)$st n year age group.) This stable age composition vector x has the property that, i), $Px = cx$, where c is a positive number, and, ii), x is a scalar multiple of the vector

$$\begin{bmatrix} 1 \\ p_1 c^{-1} \\ p_2 c^{-2} \\ \vdots \\ p_{n-1} c^{-(n-1)} \end{bmatrix}$$

where, iii), the number c satisfies the equation

$$c^n - \sum_{i=1}^{n} m_i p_{i-1} c^{n-i} = 0.$$

This section gives a generalization of the latter result to the stable age by region compositions corresponding to a generalized Leslie matrix. The question of necessary and sufficient conditions for convergence of successive population age by region compositions to a stable composition is not considered.

Suppose there exist a number c and a vector

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

with each component an N-dimensional vector, such that

$$1) \quad \begin{bmatrix} F_1 & F_2 & F_3 \\ G_1 & 0 & 0 \\ 0 & G_2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = c \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

where the matrix is a generalized Leslie matrix with $N \times N$ matric entries. (The replacement of e.g., $R_1'S\,'B_1$ by F_1 is a notational convenience. The specialization to the $3N \times 3N$ case makes the derivation easier to follow.) If 1) holds, then $F_1x_1 + F_2x_2 + F_3x_3 = cx_1$ and, 2),

$$x_2 = (c^{-1}G_1)x_1$$

$$x_3 = (c^{-2}G_2G_1)x_1$$

hence $F_1x_1 + F_2((c^{-1}G_1)x_1) + F_3\ ((c^{-2}G_2G_1)\ x_1 = cx_1$, or, 3),

$$(Ic^3 - F_1c^2 - F_2G_1c - F_3G_2G_1)x_1 = 0.$$

It is easy to reverse the argument and show that if 3) holds for some number c and vector x_1, then 1) holds with x_2 and x_3 defined by 2).

The same argument yields the following general proposition. Let

$$P = \begin{bmatrix} F_1 & F_2 & \cdot & \cdot & \cdot & F_{n-1} & F_n \\ G_1 & 0 & \cdot & \cdot & \cdot & & \\ 0 & G_2 & & & & & \\ \vdots & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & G_{n-1} & \end{bmatrix}$$

denote an arbitrary generalized Leslie matrix with $N \times N$ matric entries, and

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

denote an n-dimensional vector whose components are N-dimensional vectors. For notational convenience, let Q_i denote the matrix $G_iG_{i-1} \ldots G_1$ $(i = 1, 2, \ldots, n - 1)$ and put $Q_0 = I$ (the identity matrix). The matrix Q_i may be thought of as describing the cumulative survival-migration from the first to the $(i + 1)$st n year age group for a birth cohort which exhibits the survival-migration proportions $G_i = R_iS_i$ as it passes from the ith to the $(i + 1)$st n year age group.

Proposition. If x is a stable age by region distribution corresponding to the generalized Leslie matrix P, i.e., *if,* 1), *$Px = cx$ for some number c, then,* 2), *x is a scalar multiple of the vector*

$$\begin{bmatrix} x_1 \\ (Q_1c^{-1})x_1 \\ (Q_2c^{-2})x_1 \\ \vdots \\ (Q_{n-1}c^{-(n-1)})x_1 \end{bmatrix}$$

where, 3), *the number c and the vector x_1 satisfy the equation*

$$(Ic^n - \sum_{i=1}^{n} F_iQ_{i-1}c^{n-i})x_i = 0.$$

Conversely, if x_1 and c satisfy 3), *then the vector x given by* 2) *satisfies* 1).

4. CASE OF NO DIFFERENTIAL FERTILITY-MORTALITY

This section deals with the special case of a system of regions (the totality of which is closed to migration) which exhibit identical fertility-mortality schedules (cf. Rogers, 1968, Chapter 3). For some actual systems of regions, this situation may obtain, for practical purposes, exactly. For others it may serve as a useful first approximation. If there is no differential fertility-mortality between regions, it seems reasonable to expect that no variation in pattern or volume of interregional migration will have any effect of the rate of increase intrinsic to the fertility-mortality-migration schedule. This is indeed the case. Moreover the following stronger result holds: every characteristic value of the generalized Leslie matrix describing a system of regions each of which exhibits the same fertility-mortality schedule is a characteristic value of the Leslie matrix describing the fertility-mortality schedule common to the regions, and conversely. This section gives a precise statement and proof of these assertions.

A generalized Leslie matrix

$$\begin{bmatrix} R_1'S'B_1 & R_2'S'B_2 & \cdot & \cdot & \cdot \\ R_1S_1 & 0 & \cdot & \cdot & \cdot \\ 0 & R_2S_2 & & & \\ \vdots & & \ddots & & \end{bmatrix}$$

will be said to exhibit *no differential fertility-mortality by region* if there exist numbers s_i, b_i and s' such that $S_i = Is_i$, $B_i = Ib_i$ and $S' = s'I$ $(i = 1, 2, \ldots)$. As in the preceding section the $3N \times 3N$ case is considered, the result extending trivially to the general case.

Let M and m denote the matrices

$$\begin{bmatrix} R_1'm_1 & R_2'm_2 & R_3'm_3 \\ R_1s_1 & 0 & 0 \\ 0 & R_2s_2 & 0 \end{bmatrix}$$

$$\text{and} \quad \begin{bmatrix} m_1 & m_2 & m_3 \\ s_1 & 0 & 0 \\ 0 & s_2 & 0 \end{bmatrix}$$

respectively. The generalized Leslie matrix M exhibits no differential fertility-mortality by region. It will be shown that every characteristic value of the matrix M is a characteristic value of the matrix m and conversely.

Suppose first that c is a characteristic value of M, i.e., that for some nonzero vector x, $Mx = cx$. Then by the proposition of the preceding section there exists an N-dimensional vector x_1 such that

$$1) \quad c^3(Ix_1) - m_1p_0c^2(R_1'x_1) - m_2p_1c(R_2'R_1x_1) - m_3p_2(R_3'R_2R_1x_1) = 0,$$

and x_1 may be chosen so that its components sum to unity. Since the matrices in 1) are all column stochastic, the components of each of the vectors enclosed in parentheses in 1) sum to unity. Consequently if the N scalar equations in 1) are added the result is the single scalar equation $c^3 - m_1p_0c^2 - m_2p_1c - m_3p_2 = 0$, which shows that c is a characteristic value of the matrix m.

Suppose now that c is a characteristic value of the matrix m. Since the matrices in 1) are all column stochastic, the row vector obtained by summing the rows of the matrix

$$C = c^3I - m_1p_0c^2R_1' - m_2p_1cR_2'R_1 - m_3p_2R_3'R_2R_1$$

has each component equal to $c^3 - m_1p_0c^2 - m_2p_1c - m_3p_2$. But since c is a characteristic value of m, this quantity is zero, hence the matrix C is singular. It follows that there exists an N-dimensional vector $x_1 \neq 0$ such that $Cx_1 = 0$, and hence by the proposition of the preceding section c is a characteristic value of the matrix M.

5. NUMERICAL ILLUSTRATION

It is often useful to see how an abstract proposition manifests itself in a numerical example. With this in mind the proposition of section 3 above is applied here to the computation of a stable age by region composition corresponding to the generalized Leslie matrix

$$\begin{bmatrix} F_1 & F_2 \cdots F_8 & F_9 \\ G_1 & 0 \cdots & \\ 0 & G_2 & \\ \vdots & \ddots & \\ & & G_8 \end{bmatrix}$$

whose matric entries are given in Table 1. These entries were estimated by Rogers (1966) to represent the fertility, mortality and interregional migration experience during 1950–60 of the United States population aged less than 80 years in 1950. The regions considered are "California" and "the rest of the United States." International migration into and out of the United States is ignored, as is the migration of persons born to the population during the period. The choice

TABLE 1.—Estimated Matrix Entries of a Generalized Leslie Matrix for the United States Population: California and the Rest of the United States, First Nine Ten Year Age Groups, 1950–60

$G_1 = \begin{bmatrix} .8731 & .0279 \\ .1682 & .9903 \end{bmatrix}$	$G_5 = \begin{bmatrix} .6905 & .0300 \\ .0988 & .9176 \end{bmatrix}$	$F_2 = \begin{bmatrix} 0.3375 & 0 \\ 0 & 0.2257 \end{bmatrix}$
$G_2 = \begin{bmatrix} .9228 & .0399 \\ .1503 & .9523 \end{bmatrix}$	$G_6 = \begin{bmatrix} .7938 & .0145 \\ .0670 & .8499 \end{bmatrix}$	$F_3 = \begin{bmatrix} 1.1861 & 0 \\ 0 & 0.9746 \end{bmatrix}$
$G_3 = \begin{bmatrix} .7821 & .0450 \\ .2536 & .9843 \end{bmatrix}$	$G_7 = \begin{bmatrix} .6288 & .0137 \\ .0583 & .7171 \end{bmatrix}$	$F_4 = \begin{bmatrix} 0.4789 & 0 \\ 0 & 0.4252 \end{bmatrix}$
$G_4 = \begin{bmatrix} .7891 & .0316 \\ .1565 & .9594 \end{bmatrix}$	$G_8 = \begin{bmatrix} .3540 & .0126 \\ .0560 & .3921 \end{bmatrix}$	$F_5 = \begin{bmatrix} 0.0424 & 0 \\ 0 & 0.0364 \end{bmatrix}$

$$F_1 = F_6 = F_7 = F_8 = F_9 = 0$$

Source: Rogers, 1966, Figure 1, p. 542.

of this particular matrix has two advantages. First, it is simple, so the example does not become bogged down in a mass of numbers. Second, a stable age by region distribution corresponding to this matrix has been computed by Rogers (1966, Table 1, Column 5, Stability, interregional age structure) using an iterative procedure. Consequently the result obtained here by the application of the proposition of section 3 above may be compared to an independent computation.

The computation proceeds in several steps.

a) By 2 of the proposition, stable age by region distributions corresponding to the above generalized Leslie matrix are of the form

$$2) \quad \begin{bmatrix} x \\ (G_1c^{-1})x \\ (G_2G_1c^{-2})x \\ \vdots \\ (G_8G_7 \cdots G_1c^{-8})x \end{bmatrix}$$

where x is a certain vector (2-dimensional in this case, corresponding to the two regions considered), and c a certain number.

b) Hence if the vector x and the number c are known, the corresponding stable age by region distribution, 2, can be computed recursively as follows:

$$\begin{aligned} &x \\ &c^{-1}(G_1x) = (G_1c^{-1})x \\ &c^{-1}(G_2(c^{-1}G_1x)) = (G_2G_1c^{-2})x \\ &\vdots \end{aligned}$$

c) Suppose now that the number c is known, but that the vector x is unknown. By 3 of the proposition the vector x satisfies the matric equation

$$3) \quad \left(Ic^5 - \sum_{i=2}^{5} F_iQ_{i-1}c^{5-i}\right)x = 0.$$

Since c is known the numerical entries of the matrix in parentheses here can be computed and 3 becomes a homogeneous system

$$3') \qquad \begin{aligned} a_{11}x_1 + a_{12}x_2 &= 0 \\ a_{21}x_1 + a_{22}x_2 &= 0 \end{aligned}$$

of two linear equations in two unknowns, where a_{ij} $(i, j = 1, 2)$ denotes the ijth entry of the matrix in parentheses in 3 and x_j $(j = 1, 2)$ denotes the jth entry of the vector x in 3.

d) There remains the problem of obtaining the number c. For the purposes of this example, I simply assume the value of c known to be 1.2230. The system 3′ of linear equations then works out to be

$$\begin{aligned} 0.3536x_1 - 0.1923x_2 &= 0 \\ -0.7714x_1 + 0.4193x_2 &= 0 \end{aligned}$$

of which one solution is $x_1 = 0.3523$, $x_2 = 0.6477$. The computation described in b) above yields

$$x = \begin{bmatrix} .352300 \\ .647700 \end{bmatrix}$$

$$(G_1c^{-1})x = \begin{bmatrix} .266283 \\ .572914 \end{bmatrix}$$

$$(G_2G_1c^{-2})x = \begin{bmatrix} .219611 \\ .478829 \end{bmatrix}$$

$$\vdots$$

and so on. (Two extra digits have been carried to reduce the effect of accumulating rounding errors.) When the 18-dimensional vector so obtained is normalized, the result is the stable age by region composition given in Table 2.

TABLE 2.—A Stable Age by Region Composition Corresponding to the Generalized Leslie Matrix Given in Table 1 (proportion of persons in each age group and region)

	Region	
Age group	California	Rest of the United States
0 - 9	.0802	.1475
10 - 19	.0606	.1305
20 - 29	.0500	.1091
30 - 39	.0360	.0981
40 - 49	.0258	.0816
50 - 59	.0165	.0633
60 - 69	.0115	.0449
70 - 79	.0064	.0269
80 - 89	.0021	.0089

Note: See section 5 for explanation.

The stable age by region distribution given in Table 2 may be checked directly by: i) premultiplying the stable composition vector by the generalized Leslie matrix given in Table 1; ii) multiplying the stable composition vector by 1.2230; and iii) comparing the two results, which should be equal. The differences are at most plus or minus one unit in the fourth decimal place. The differences between the results in Table 2 and the results given by Rogers (1966) are as high as six units in the fourth decimal place. This discrepancy, which from a practical point of view is not large, can probably be explained by the different behaviour of rounding errors in the two methods.

ACKNOWLEDGMENTS

This paper is based in part on work done as a trainee at the Population Studies Center, University of Michigan, in a program supported by grants from The Ford Foundation. This work has been supported by training grants to the Department of Demography, University of California, Berkeley, from the National Center for Health Services Research and Development (8 T01 HS-00059) and The Ford Foundation.

REFERENCES

Keyfitz, Nathan. 1968. Introduction to the Mathematics of Population. Reading: Addison-Wesley.

Rogers, Andrei. 1966. The multiregional matrix growth operator and the stable interregional age structure. Demography 3:537–44.

———. 1968. Matrix Analysis of Interregional Population Growth and Distribution. Berkeley: University of California Press.