MATHEMATICAL MODELS IN DEMOGRAPHY AND ACTUARIAL MATHEMATICS

Robert Schoen
Hoffman Professor of Family Sociology and Demography, Pennsylvania State University, USA

Keywords: Life table, marriage squeeze, multistate population, population momentum, population projection, stable population, "two-sex" population model

Contents

Summary

Population models typically describe patterns of mortality, fertility, marriage, and/or migration, and depict how those demographic behaviors change the size and age structure of populations over time. Demographic behavior in fixed rate, one-sex populations is quite well understood by means of existing demographic and actuarial models, principally the stationary population (life table), stable population, and multistate models. Progress has been made on "two-sex" marriage and fertility models, though no consensus has been achieved. Significant work has been done on modeling populations with changing vital rates, but such dynamic modeling is still in a relatively early stage.

1. Introduction

Mathematical models of population date back at least to 1662, when John Graunt used parish records of birth and death to describe the demography of London, England and present the first example of a life table. Fundamentally, population models are elaborations of the basic differential equation

$$P' = m\,P \quad , \tag{1}$$

where *P* represents the number in the population of interest; *P'* is the time derivative of *P*, i.e. how the size of the population changes over time; and *m* is a force of transition, or an instantaneous rate (or probability, or risk) of change with respect to the demographic behavior of interest. The most common demographic behaviors are birth, death, marriage, divorce, and migration, though models have frequently been applied to many other topics, including educational enrollment, labor force status, and disability status.

Population models have the great advantage of being logically closed, that is population size and composition (or stock) at one time can be found from the population stock at an earlier time and the demographic events (or flows) that occurred between those time points. The classic expression of that closure is the so-called Vital Statistics Balancing Equation which, ignoring age and sex, can be written

$$P(t+1) = P(t) + B(t) - D(t) + I(t) - O(t), \tag{2}$$

where $P(t)$ is the size of the population of interest at the beginning of year *t*, $B(t)$ is the number of births during year *t*, $D(t)$ is the number of deaths during year *t*, $I(t)$ is the number of inmigrants during year *t*, and $O(t)$ is the number of outmigrants during year *t*. Fertility, mortality, and migration constitute the core demographic processes.

Most demographic models employ two significant simplifications. The first is that the models are deterministic, i.e. that they ignore chance fluctuations and focus on expected values of population behavior. For example, the same rates of death always have the same impact on survivorship, with no chance (or stochastic) variability. The second simplification is that within the categories specified, population heterogeneity is ignored. Thus in a model that only recognizes age and sex, all persons of the same age and sex are at equal risk of experiencing an event (i.e. of "transferring"). Population models are rooted in Markov processes, and embody the Markovian assumption that a person's risk of transfer depends only on current characteristics, with past experiences having no effect. While those assumptions are generally counterfactual, their practical importance varies greatly.

2. Life Table Models

2.1 Life Table Structure and Functions

The basic life table model follows a birth cohort to the death of its last member. More formally, there is only one living state, and the only recognized transfers are exits (decrements) from that state. At any exact age x, basic differential Eq. (1) can be written in the form

$$m(x) \ = \ -P'(x)/P(x)\,, \tag{3}$$

which defines $m(x)$, the force of mortality (or decrement) at exact age x, to be minus the derivative of the natural logarithm of the change in the number of persons at exact age x. Since the size of the life table cohort decreases monotonically, it is customary to use a minus sign so that the force [often written as $\mu(x)$] is positive. Let the initial size of the life table cohort, referred to as its radix, be designated by $l(0)$. Then Eq. (3) can be integrated to show that $l(x)$, the number of survivors to exact age x, is given by

$$l(x) = l(0)\exp\left[-\int_0^x m(y)dy\right] \tag{4}$$

Eq. (4) is the basic solution for a life table, as it transforms rates of decrement m into probabilities of survival $l(x)/\,l(0)$. As a practical matter, Eq. (4) can be implemented by choosing a simple functional form for $m(x)$. For single year age intervals, assuming that $m(x)$ is constant within intervals is generally acceptable.

An alternative approach, which has been termed the "General Algorithm" for life table construction, sets forth three sets of equations. The first set, the flow equations, describe the permissible flows into and out of states. In the basic life table, there is one flow equation which can be written

$$l(x+n) \ = \ l(x) \ - \ d(x,n) \tag{5}$$

where $d(x,n)$ is the number of deaths (or decrements) between exact ages x and $x+n$. Eq. (5) simply states that the number of survivors to exact age $x+n$ is the number of survivors to age x less the number who exit the table between exact ages x and $x+n$.

The second set of equations consists of orientation equations, which relate the observed rates of transfer to the model rates. It is generally convenient and desirable to equate those sets of rates, hence the basic orientation equation is

$$M(x,n) \ = \ m(x,n)\,, \tag{6}$$

where $M(x,n)$ and $m(x,n)$ refer, respectively, to the observed and model rates of transfer between exact ages x and $x+n$.

The third set of equations is the person-year set of equations, which relate the number of survivors to each exact age to the number of person-years lived in an interval (where one person year is one year lived by one person). The exact relationship is

$$L(x,n) \ = \ \int_0^n l(x+u)\,du\,, \tag{7}$$

where $L(x,n)$ is the number of person years lived between exact ages x and $x+n$. The General Algorithm shifts the task of making a life table from integrating the force of decrement function in Eq. (3) and then integrating the survivorship (l) function in Eq. (7), to simply integrating the survivorship function. The General Algorithm also readily generalizes to specify more complex models (e.g. multistate models). Constructing a life table from a set of data involves, implicitly or explicitly, making two choices. On is an assumption about how the observed rates relate to the life table rates; typically they are assumed to be the same. The second concerns the functional form of either the force of decrement or the survivorship curve. Many choices are possible. The simplest choice for the survivorship curve is a linear relationship, that is

$$L(x,n) \;=\; \tfrac{1}{2}\, n\, [\; l(x) \;+\; l(x+n)\;], \tag{8}$$

which is generally adequate for single years of age and for 5 year age intervals when the force of decrement is rising within the interval. Many other choices are possible and often desirable, especially for intervals longer than one year or when the force of decrement is large. The details of life table construction are discussed in the works cited in the bibliography. While there is no "ideal" solution, many adequate methods are available. Table 1 provides an example of a life table, based on the experience of California Females during the year 1970, and shows the principal life table functions. The table is "abridged", as age is shown for roughly every fifth year up to age 85. Age 1 is also shown as mortality in the first year of life is typically much higher than in the immediately following years. In the past, age 85 was the highest age commonly shown in life tables, though with improved survivorship and better data for the high ages, many tables now show survival to age 90. The next column is the survivorship column, which begins with the frequently used radix $l(0)$=100 000. Of that number, 32 338 females survive to exact age 85. Next is the $d(x,n)$ column, calculable from Eq. (5), which shows the age distribution of decrements. By definition, $l(85)=d(85,\infty)$.

The probability of dying between exact ages x and $x+n$, denoted $q(x,n)$, is shown in the fourth column, and is found from the relationship

$$q(x,n) \;=\; d(x,n) \;/\; l(x)\,. \tag{9}$$

The fifth column shows the life table decrement rates, $m(x,n)$. In terms of life table functions, they satisfy the relationship

$$m(x,n) \;=\; d(x,n) \;/\; L(x,n)\,. \tag{10}$$

The sixth column is often not presented, but can be quite useful. It shows Chiang's $a(x,n)$, the average number of years lived between exact ages x and $x+n$ by those decrementing (or dying) during the interval. Mathematically, it must satisfy the equation

$$L(x,n) \;=\; n\; l(x+n) \;+\; a(x,n)\; d(x,n) \tag{11}$$

as the total number of person-years lived in an interval is the sum of the years lived by those who survive the interval and those who do not.

Age at Start of age interval (x)	**Number alive at start of age interval $\ell(x)$**	**Number of deaths during age interval $d(x,n)$**	**Probability of dying during age interval $q(x,n)$**	**Average annual mortality rate during age interval $m(x,n)$**	**Years lived in interval by those dying in interval $a(x,n)$**	**Number of person-years lived during age interval $L(x,n)$**	**Number of person-years lived after age x, $T(x)$**	**Expected years of life at start of age interval $e(x)$**
0	100000	1568	0.015679	0.015905	0.097	98584	7566021	75.66
1	98432	271	0.002755	0.000690	1.498	393050	7467437	75.86
5	98161	135	0.001378	0.000276	2.489	490466	7074387	72.07
10	98026	143	0.001462	0.000293	2.818	489813	6583921	67.17
15	97882	342	0.003496	0.000700	2.655	488608	6094108	62.26
20	97540	395	0.004051	0.000812	2.549	486732	5605500	57.47
25	97145	437	0.004495	0.000901	2.606	484669	5118768	52.69
30	96706	566	0.005850	0.001173	2.638	482208	4634099	47.92
35	96143	825	0.008578	0.001723	2.670	478793	4151891	43.18
40	95318	1245	0.013060	0.002628	2.686	473709	3673098	38.54
45	94073	1937	0.020594	0.004159	2.663	465838	3199389	34.01
50	92136	2760	0.029961	0.006078	2.652	454195	2733551	29.67
55	89375	3945	0.044137	0.009016	2.629	437520	2279356	25.50
60	85430	5187	0.060717	0.012503	2.631	414864	1841836	21.56
65	80243	7207	0.089816	0.018760	2.635	384173	1426972	17.78
70	73036	9862	0.135030	0.028841	2.644	341941	1042799	14.28
75	63174	13998	0.221578	0.049580	2.604	282329	700858	11.09
80	49176	16838	0.342400	0.082311	2.546	204565	418529	8.51
85	32338	32338	1.000000	0.151139	6.616	213964	213964	6.62

Source: Adapted from R. Schoen and M. Collins (1973) Mortality By Cause: Life Tables for California 1950-1970, Sacramento: State of California, Department of Public Health, p25.Of 100,000 born alive

Table 1: Life Table for California Females, 1970.

The seventh column shows the number of person-years lived in every age interval. The last entry, $L(85,\infty)$= 213 964, gives the total number of person-years lived above age 85. The eighth column, $T(x)$, shows the number of person-years lived at and above each exact age x, and is the sum of the $L(x, n)$ column from age x through the highest age group in the table. The last column, $e(x)$, shows the expectation of life at age x, or the average number of years a person exact age x is expected to live. In terms of life table symbols

$$e(x) = T(x) \; / \; l(x) = \sum L_i / l(x), \tag{12}$$

where ΣL_i indicates the sum of the person-years lived at and above exact age x. For the highest age interval, $e(85)=a(85,\infty)$. The most commonly used life table summary measure is $e(0)$, the expectation of life at birth, here 75.66 years.

To this point, the life table has been viewed as following the experience of a birth cohort, and showing the implications of a set of decrement rates on survivorship. A second perspective needs to be recognized, a period perspective that gives rise to a stationary population. Assume that age ω is the highest age attained in the life table (i.e. no one survives to attain exact age ω+1). Now assume that the rates of decrement remain constant over time, and that for at least ω+1 consecutive years there are annual birth cohorts of $l(0)$ persons. The result is a stationary population, that is a population of constant size (specifically of $T(0)$ persons) and of unchanging age composition. The number of persons between the ages of x and $x+n$, at any time, is given by $L(x,n)$. The experience of the stationary population in one year captures that of the life table cohort over its entire lifespan. Each year in the stationary population $l(x)$ persons attain exact age x, there are $d(x,n)$ deaths between the ages of x and $x+n$ with a total of $l(0)$ deaths, and there are $L(x,n)$ person years lived between the ages of x and $x+n$ giving a total of $T(0)$ person-years lived.

Life tables have been widely used to reflect mortality patterns, and are now routinely produced by most national statistical organizations. The basic model can easily be generalized to recognize more than one cause of decrement. Cause-of-death life tables, nuptiality-mortality life tables, and numerous other types of multiple decrement models have been constructed, and describe the implications of competing risks. Cause-eliminated life tables (also known as associated single decrement life tables) have also been calculated. They can address the hypothetical question of how survivorship would change if one (or more) observed causes disappeared while the forces of decrement from the remaining causes remained unchanged.

Various means have been used to deal with problems related to population heterogeneity. Separate life tables are routinely prepared for males and females, and frequently recognize other characteristics such as geographical area or race/ethnicity. That approach, however, can quickly exhaust the data available for even a large population. Proportional hazard (or Cox) models have been used to incorporate more covariates, though usually with the assumption that risks differ by a constant factor over age. More sophisticated efforts have sought to explicitly model individual "frailty" (or susceptibility to death). They have been hampered by the complexity of the problem and the very limited empirical evidence available on the distribution of frailty, both between persons and over the life course.

2.2 Actuarial Science

Perhaps the leading use of the life table has been in the insurance and pension industries. Actuarial science uses the life table, and other models reflecting life contingencies, to determine insurance and pension risks, premiums, and benefits. In essence, actuarial methods combine the life table with functions related to an assumed fixed rate of interest.

Let $\imath$ denote the annual rate of interest such that \$1 at time t will increase to $1+\imath$ dollars in exactly one year. It is convenient to denote the present value of a dollar by v, where v represents the amount one must have at time t (under prevailing interest rate $\imath$) in order to have \$1 one year in the future. In symbols,

$$v = 1/\ (1+\imath)\,. \tag{13}$$

It immediately follows that v^n is the present value of \$1 n years in the future.

In its simplest form, a life annuity is a periodic series of payments, with the first payment of \$1 due in one year, and the payments continuing annually as long as the recipient is alive. Consider the present value of a life annuity payable to a person now exact age x. The present value of the payment due in one year is $v\ l(x+1)/l(x)$. The present value of the payment due in two years is $v^2 l(x+2)/l(x)$. It follows that a_x, the present value of a life annuity to a person exact age x is

$$a_x = \left[\sum_j v^j l(x+j)\right] / l(x) \tag{14}$$

where summation index j ranges from 1 to the highest age in the life table. [In a_x, age is indicated by a subscript. That is conventional in actuarial usage, and here serves to distinguish the life annuity from Chiang's a function, defined in Eq. (11).]

In its simplest form, a life insurance for a person exact age x is a payment of \$1 made at the end of the year in which that person dies. Combining an appropriate rate of interest and life table, and following the same logic as used in deriving Eq. (14), A_x, the present value of a life insurance for a person exact age x is

$$A_x = \left[\sum_j v^{j+1} d(x+j)\right] / l(x) \tag{15}$$

In practice, of course, Eqs. (13) - (15) are modified to deal with shorter intervals of time, more complex benefit options, and a variety of expense and other "loading" factors.

Insurance companies are quite conscious of “selection” factors, which motivate atypical persons to seek coverage. An extreme but not unlikely example is a person recently diagnosed with a fatal disease who seeks to buy life insurance. To gauge the size of those selection/population heterogeneity effects, actuaries have used “select and ultimate” life tables. For example, in the United States, the Society of Actuaries has produced life tables with a 15 year select period. That is, for the first 15 years of coverage, the risk of death is seen as depending on policy duration as well as age (and usually other characteristics such as sex). After 15 years, when the selection effect is deemed negligible, mortality rates depend on age alone. Tables that impose a second time varying parameter of that sort are known as semi-Markov models. They are common in actuarial work, but infrequently encountered in demographic analyses.

The life table, as a stationary population, is the basic model underlying benefit calculations. The classic article by C.L. Trowbridge on pension funding viewed the participants in a “mature” pension plan as constituting a stationary population. The “service” table commonly used by actuaries in benefit calculations is a multiple decrement life table where a steady stream of entrants to a benefit plan are followed over time subject to risks of death, withdrawal from employment, disability retirement, and “normal” retirement.

-
-
-

Bibliography

Caswell, H. (1989). *Matrix Population Models*, Sunderland MA: Sinauer. [This book provides an excellent overview of population models and their mathematical bases with an emphasis on biological applications]

Chiang, C.L. (1968). *An Introduction to Stochastic Processes in Biostatistics*, New York: Wiley. [The book is a rigorous mathematical statistical treatment of survival models, including multistate models.]

Coale, A.J. (1972). *The Growth and Structure of Human Populations: A Mathematical Investigation*, Princeton: Princeton University Press. [This monograph is a classic, but still the book with the most comprehensive treatment of populations with changing rates]

Coale, A.J. and P.Demeny. (1983). *Regional Model Life Tables and Stable Populations* (2d Ed). New York: Academic Press. [An extremely useful reference volume containing model life tables, their method of construction, and associated stable populations that bracket most of human demographic experience.]

Halli, S.S. and K.V. Rao. (1992). *Advanced Techniques of Population Analysis*, New York: Plenum. [A comprehensive look at demographic models, through multistate models from a fairly statistical perspective.]

Jordan, C.W. Jr. (1975). *Life Contingencies*, Chicago: Society of Actuaries. [An indispensable book on actuarial mathematics]

Keyfitz, N. (1977). *Introduction to the Mathematics of Population* (with revisions), Reading MA: Addison-Wesley. [A comprehensive book on basic population mathematics]

Keyfitz, N. and W. Flieger. (1990). *World Population Growth and Aging: Demographic Trends in the Late Twentieth Century*, Chicago: Univ. of Chicago. [A detailed exposition of discrete population models with applications that provide data on many contemporary populations.]

Land, K.C. and A. Rogers (Eds). (1982). *Multidimensional Mathematical Demography*, New York: Academic Press. [A collection of basic papers on multistate modeling]

Lee, R.D. and S. Tuljapurkar. (1994). "Stochastic population forecasts for the United States: Beyond high, medium, and low." *Journal of the American Statistical Assn* 89: 1175-89. [A sophisticated look at population projections and how to establish confidence intervals for them.]

Lotka, A.J. (1939). *Theorie Analytique des Associations Biologique*, Paris: Hermann. [The classic exposition of stable population theory by its developer.]

Namboodiri, K. and C.M. Suchindran. (1987). *Life Table Techniques and Their Applications*, Orlando FL: Academic Press. [A statistically oriented treatment of basic and multistate life table models]

Pollard, J.H. (1973). *Mathematical Models for the Growth of Human Populations*, Cambridge: Cambridge University Press. [A concise treatment of stable populations and related topics]

Preston, S.H., P. Heuveline, and M. Guillot. (2001). *Demography: Measuring and Modeling Population Processes*, Oxford: Blackwell Publishers. [An excellent, current demographic methods text with an introduction to fixed rate demographic models and the basic patterns of demographic behavior]

Rogers, A. (1975). *Introduction to Multiregional Mathematical Demography*, New York: Wiley. [An early exposition of multistate models with a detailed discussion of multistate stability.]

Schoen, R. (1988). *Modeling Multigroup Populations*, New York: Plenum. [A comprehensive treatment of fixed rate population models, with details on multistate life table construction and an emphasis on two-sex models]

Schoen, R. and Y.J. Kim. (1991). "Movement toward stability as a fundamental principle of population dynamics." *Demography* 28: 455-66. [An extension of the concept of population momentum that describes the process of population change at the margin.]

United Nations, Dept of International Economic and Social Affairs. (1983). *Manual X: Indirect Techniques for Demographic Estimation* by K. Hill, H. Zlotnik, and J. Trussell. Population Studies, No. 81. New York: United Nations. [An authoritative compendium of techniques for estimating demographic measures from incomplete data.]

Biographical Sketch

Robert Schoen is Hoffman Professor of Family Sociology and Demography at The Pennsylvania State University (USA). He received a Ph.D. degree in demography from the Department of Demography, University of California at Berkeley in 1972, and became an Associate of the Society of Actuaries in 1968. Prior to moving to the Department of Sociology and the Population Research Institute at Penn State in 1999, he held professorships in the Department of Sociology, University of Illinois at Urbana-Champaign and in the Department of Population Dynamics, Johns Hopkins University.

Professor Schoen's primary interests are in family formation and dissolution (including cohabitation and both marital and nonmarital fertility) and in demographic models and methods. His secondary interests are in ethnic demography, aging, and mortality. He is the author of *Modeling Multigroup Populations*, a 1988 monograph that examined the full range of fixed rate population models, with close attention to multistate and "two-sex" models. His 1997 *Population and Development Review* paper (with Y.J. Kim, C.A. Nathanson, J.M. Fields, and N.M. Astone), "Why Do Americans Want Children?", argued that fertility was sustained in low fertility societies by the social resource value of children. In a 2002 *Social Forces* paper (with N.M. Astone, K. Rothert, N.J. Standish, and Y.J. Kim), "Women's Employment, Marital Happiness, and Divorce," he argued that wives' employment did not destabilize marriages, but did facilitate divorce when both partners were not happy. His forthcoming work in *Demography* (with S.H. Jonsson), "Modeling Momentum in Gradual Demographic Transitions," develops a new form of Quadratic Hyperstable population model that relates exponentially changing fertility to a resultant exponentiated quadratic birth sequence. That model is then used to determine the population growth that accompanies gradual transitions from an initial stable population to an ultimate stationary population.