Biomathematics
Volume 6

D. Smith N. Keyfitz

Mathematical Demography

Selected Papers

Springer-Verlag
Berlin · Heidelberg · New York

David Smith · Nathan Keyfitz

Mathematical Demography

Selected Papers

With 31 Figures

Springer-Verlag Berlin Heidelberg New York 1977

David P. Smith
Nathan Keyfitz
Harvard University, Center for Population Studies, 9 Bow Street,
Cambridge, Mass. 02138 (USA)

AMS Subject Classifications (1970)
92-01, 92-03, 92A15, 01A55, 01A60, 01A75

ISBN 3-540-07899-1 Springer-Verlag Berlin · Heidelberg · New York
ISBN 0-387-07899-1 Springer-Verlag New York · Heidelberg · Berlin

Library of Congress Cataloging in Publication Data Main entry under title: Mathematical demography. (Biomathematics; v. 6) Bibliography: p. Includes index. 1. Demography—Mathematical models—Addresses, essays, lectures. I. Smith, David P., 1944— II. Keyfitz, Nathan, 1913— HB885.M33 519.5 77-21464

Printed in Germany.
Typesetting and printing: Beltz, binding: Konrad Triltsch, Würzburg
2145/3120-543210

Biomathematics

Volume 6

For John and Harriet

Preface

This volume is an effort to bring together important contributions to the mathematical development of demography and to suggest briefly their historical context. We have tried to find who first thought of the several concepts and devices commonly used by demographers, what sort of problem he was facing to which the device or concept seemed the solution, and how his invention developed subsequently in the hands of others.

Historically, the book starts with a Roman table of life expectancies from the third century a.d. about which we know little, and with John Graunt's explorations in an area that was still popularly suspect when he wrote in 1662. These are followed by the astronomer Halley, who looked into the field long enough to invent the life table and to notice that Their Majesties would take a sizeable loss on the annuity scheme they had just launched; and by Euler, who was first to devise the formulas of stable population theory and to apply them to filling gaps in data. To these we add the handful of further contributions in the 19th century and many pieces from the explosion of contributions that began in this century with Lotka. We doubt that we have managed to trace everything back to its ultimate beginning, and suspect that our nominees in some cases have been anticipated by predecessors who will be turned up by other students.

The works we include form a living heritage in demography: Graunt; Halley; Euler; Lotka; Milne, who formalized life table construction; Lexis, who was preoccupied with the way members of a population are situated simultaneously in age and in time, and showed how a plane chart, now known as a Lexis diagram, can help analysis. Much less alive, and largely excluded here, are such notions as that of George King, that graduation of data for a life table was more accurate from pivotal death rates calculated at five-year intervals; John Graunt's belief that the right way to describe the dynamics of a population was as the *ratio* of births to deaths, without considering age; and devices that once reduced the labor of numerical calculation but are obsolete in a computer age. These and many other ideas that have proved to be dead ends and are now of merely antiquarian interest we tried to distinguish from those that were part of a chain of development that is still advancing. As far as we could discriminate our excerpts are confined to the latter.

To determine which works most deserve attention among the large number written has not been easy, and we have undoubtedly made mistakes both of inclusion and of omission. We were far from insisting on subtle mathematical ideas, but did look for the effective uses of mathematics that have come to be assimilated into population work. Articles that profess to deal with population but whose main interest was mathematics we tried to avoid, and we avoided them doubly if they were a mere import from some other subject that seemed unlikely ever to be naturalized in population analysis. Some ideas and techniques have a kind of *droit de la cité* in contemporary population study, and we hope these are the ones that predominate in our selections.

To find passages that were self-contained and suitable for contemporary reading was occasionally difficult. Writers often used symbols well known to their place and time, and their immediate readership had no need for definitions we would now miss. To this the earlier works add key formulas with no hint as to how they are derived. Where we expect readers to have trouble as a result, because we did, we include a brief explanation of what is being done.

The choice of excerpts from the classic articles and books rather than complete reprints in all cases was dictated partly by economy of publication, but this was not the only constraint. Benjamin Gompertz fairly compactly introduces his Law of mortality, but spends above fifty pages fitting it to life tables and working out its implications for annuity payments. Harro Bernardelli published the first article on the use of matrices in population projection in the *Journal of the Burma Research Society*, which is not a source that most of us would come across in our ordinary reading. He has top priority for inclusion, but he deals partly with problems of the Burmese economy under British colonial rule and with speculations on cyclic events that do not carry much interest for readers today. Leslie, whose reading in a sickbed had taken him deep into the mathematical properties of matrices, went into cogredient and contragredient transformations that are unlikely to have demographic application. We saw no need to burden the reader with these only to have him discover at the end that he would never need them.

In editing we did not strip down our authors to the point of losing the context of their contribution to our subject. We learned much of an incidental character in our reading and have tried to retain that richness. Where substantive omissions are made we note these for the reader's benefit.

Several topics that fall in the province of demography are not included, among them treatments of human spatial ecology, urbanization, and migration. Omission is partly due to space limitations, and partly to lack of confidence in our ability to decide what is basic in fields whose mathematical explorations are recent and expanding rapidly.

We expect from the reader at least some background in calculus and matrix algebra, and several papers will require an understanding of stochastic processes. The reader lacking a background in elementary mathematics will find the greater part of the book difficult.

Secondary accounts of much of what we present can be found in Keyfitz (1968), and stochastic processes are well handled in Feller (1968) and Chiang (1968). Our chief sources for the early histories given here are Hendriks (1852, 1853), Westergaard (1969), and Lorimer (1959).

Lecture Notes in Biomathematics

This series aims to report new developments in biomathematics research and teaching – quickly, informally and at a high level.

Volume 1: P. Waltman
Deterministic Threshold Models in the Theory of Epidemics

Volume 2:
Mathematical Problems in Biology
Victoria Conference
Editor: P. van den Driessche

Volume 3: D. Ludwig
Stochastic Population Theories
Notes by M. Levandowsky

Volume 4:
Physics and Mathematics of the Nervous System
Proceedings of a Summer School organized by the International Centre for Theoretical Physics, Trieste, and the Institute for Information Sciences, University of Tübingen, held at Trieste, August 21-31, 1973
Editors: M. Conrad, W. Güttinger, M. Dal Cin

Volume 5:
Mathematical Analysis of Decision Problems in Ecology
Proceedings of the NATO Conference held in Istanbul, Turkey, July 9-13, 1973
Editors: A. Charnes. W. R. Lynn

Volume 6: H. T. Banks
Modeling and Control in the Biomedical Sciences

Volume 7: M. C. Mackey
Ion Transport trough Biological Membranes
In Integrated Theoretical Approach

Volume 8: C. Delisi
Antigen Antibody Interactions

Volume 9: N. Dubin
A Stochastic Model for Immunological Feedback in Carcinogenesis: Analysis and Approximations

Volume 10: J. J. Tyson
The Belousov-Zhabotinskii Reaction

Volume 11:
Mathematical Models in Medicine
Workshop, Mainz, March 1976
Editors: J. Berger, W. Bühler, R. Repges, P. Tautu

Volume 12: A. V. Holden,
Models of the Stochastic Activity of Neurones

Volume 13:
Mathematical Models in Biological Discovery
Editors: D.L. Solomon, C.F. Walter

Volume 14: L. M. Ricciardi
Diffusion Processes and Related Topics in Biology
Notes taken by C. E. Smith

Volume 15: T. Nagylaki
Selection in One- and Two-Locus Systems

Volume 16: G. Sampath, S.K. Srinivasan
Stochastic Models for Spike Trains of Single Neurons

Volume 17: T. Maruyama
Stochastic Problems in Population Genetics

Volume 18:
Mathematics and the Life Sciences
Selected Lectures, Canadian Mathematical Congress, August 1975
Editor: D. E. Matthews

Springer-Verlag
Berlin
Heidelberg
New York

Table of Contents

The Life Table

Mathematical demography has its modern beginnings in the gradual development of correct procedures for forming life tables, and in a single remarkable paper by Leonard Euler (1760, paper 11 below) that introduced stable age distributions. One far older work of at least some quality survives: a third century a.d. table of life expectancies attributed to Ulpian (paper 1) that remained in use in northern Italy through the 18th century. The table and an accompanying discussion have been taken from C.F. Trenerry (1926).

The mathematician Girolamo Cardano took up the problem briefly in 1570, but without substantive results. Cardano made the assumption that a man who took great care in all things would have a certain life expectancy α, so that $\mathring{e}_x = \alpha - x$ for all ages x, and then asked what part of this might be forfeited by a relaxation of prudence. He proposed letting life expectancy fall by $\frac{1}{40}$ of its value during each year in which a man was reasonably careful but not fastidious: by the nature of the life expectancies, a man might be born with the prospect of living say 260 years and yet die at age 80, having every year thrown away by inattention a part of what remained to him (Cardano 1570, pp. 204—211). A modern interpretation would be that a cohort carries its life table with it. The result was not generalized to populations.

John Graunt's *Natural and Political Observations Upon the Bills of Mortality* (1662) is the first substantive demographic work to have been written. The book is occasionally curious but most often impressive, even from a perspective of three hundred years: Graunt culled a remarkable amount of information from the christening and death lists begun in the later plague period and usually understood its implications. Parts of the treatise are included here as paper 2.

A second work of great importance followed upon Graunt's, Edmund Halley's (1693) presentation of the Breslau (Wroclaw Poland) life table (paper 3). Halley had made an effort to obtain the Breslau lists in order to see what might be done with them, after learning of their apparent quality.

The methods of calculation Halley used in his life table were partly informal, as in his remarks on stationarity and in his unorthodox subtraction (where l_x is the number of survivors to exact age x in the life table from among l_0 births and L_x represents the number between ages x and $x+1$)

$$890[=l_1]-198=692[=L_6],$$

explained by the oblique statement: "198 do die in the *Five Years* between 1 and 6 compleat, taken at a *Medium*." [The terminology has created confusion down to the present century. Raymond Pearl (1922, p. 83), apparently reading the L_x terms that make up Halley's table as l_x, calculated life expectancy at birth as 33.5 years by the table instead of the correct 27.5. The mistake is carried over in Dublin, Lotka and Spiegelman (1949, p. 34)].

Johan DeWit (1671) preceded Halley in the correct calculation of annuities, using exact (l_x) as against Halley's approximate (L_x) denominators, and most of Halley's other *Uses* can be answered differently, but the quality of Halley's table and discussion much surpasses the few earlier works and several of the subsequent ones. His table is graphed below, alongside Ulpian's, Graunt's, DeWit's, and as references the middle level table for Crulai c. 1700 (Gautier and Henry 1958, pp. 163, 190) and one of the Coale and Demeny (1966) model life tables.

After Halley, the next impressive contribution was the series of life tables for annuitants and monastic orders by Antoine Deparcieux, printed in 1746. The accuracy of Deparcieux's data was sufficient for him to show that adult life expectancies had been increasing over the previous half century. Deparcieux calculated his $\mathring{e}_x$ values by the simple but adequate formula

$$\mathring{e}_x = \frac{\sum_{i=x}^{\omega} (l_i - l_{i+1})(i + 0.5 - x)}{l_x} = \frac{\sum_{i=x}^{\omega} l_i}{l_x} - 0.5 .$$

Of their utility he writes (1760, pp. 58—59): "Les vies moyennes [i.e., $\mathring{e}_x$] sont ce qui m'a paru de plus commode pour faire promptement & sans aucun calcul, la comparaison des différent ordres de mortalité qu'on a établis ... [Life expectancies are what have appeared to me most convenient for making promptly and without any calculation a comparison of different orders of mortality that one has established]."

Two later efforts merit attention here: Daniel Bernoulli (1766) introduced continuous analysis and suggested the force of mortality [$\mu(x)$] in an application of differential calculus to the analysis of smallpox rates. Later Émmanuel Étienne Duvillard (1806), in an article that also introduced the T_x column (defined as $T_x = \sum_{x}^{\omega} L_i$), applied Bernoulli's method to estimate the increase in life expectancy that would follow if smallpox were eliminated by Edward Jenner's vaccine. The calculus, which these and most modern work employ, dates to a seventy year period (1665—1736) between Isaac Newton's first investigations and the publication of his principal works. Westergaard (1969, pp. 92—93) comments however that it was not until the late nineteenth century that continuous analysis was widely enough understood for Bernoulli's work to be appreciated.

Joshua Milne in his excellent *Treatise on the Valuation of Annuities* (1815), which includes a careful analysis of life tables made prior to his, was first to suggest a formula by which l_x values could be calculated for real populations.

His is the well known expression

$$d_x = l_x\left(\frac{D_x}{P_x + \frac{1}{2}D_x}\right)$$

where d_x is the number dying between ages x and $x+1$ in the life table population, D_x represents calendar year deaths between these ages in an observed population, and $P_x + \frac{1}{2}D_x$ constructs an initial population analogous to l_x except for scale by adding half of the yearly deaths to the observed midyear population P_x.

Excerpts from Milne's discussions of the life table and of age-specific fertility rates (due to Henric Nicander (1800, pp. 323—324)) are given in paper 4. Milne's method for graduating data from grouped to single ages, not the best, has been omitted. His footnoted criticism of Thomas Simpson's (1742) work opens an area of discussion that is easily missed: Milne's life table was misread by William Sutton (1884)—whose clarification in 1874 of the construction of Richard Price's 1771 Northampton Table is a more competent work—but was immediately re-established by George King (1884). Like other fields, demography does not only move forward.

The fifth article in this section excerpts from George King (1902), whose notation is contemporary, his explanations of terms of the life table. From the middle of the 19th century William Farr standardized much of the life table, but he did not put his work in a form at all comparable to King's excellent textbook. [A recent addition to the life table, from C. L. Chiang (1960a), is the term $_na_x$. This is King's unremembered "average amount of existence between ages x and $x+n$, belonging to those who die between these ages," i. e.: $_na_x = \frac{_nL_x - n\,l_{x+n}}{_nd_x}$.]

Out of sequence, the Lexis (1875) diagram is introduced in paper 6. For most of a century it has been a standby of all analysis attempting to relate age and time. Among contemporary works, those of Roland Pressat (1969, 1972) exploit it most fully.

The important contributions to the life table in this century have been competent abridgement techniques for generating tables by five or ten year age groupings in place of single years of age. The Lowell Reed and Margaret Merrell (1939) article included here as paper 7 did much to establish the validity of abridgement techniques by its introduction of an attractive expression:

$$_nq_x = 1 - \exp[-n\,_nm_x - a\,n^3\,_nm_x^2]$$

for estimating $_nq_x$ from $_nm_x$ values where wide age groupings are used. In the expression, $_nm_x$ is the age-specific death rate in the life table population ages x to $x+n$ (that is, $_nm_x = {_nd_x}/{_nL_x}$) and $_nq_x$ the probability of dying within the interval for a person of exact age x ($_nq_x = {_nd_x}/l_x$). By empirical examination the authors found that the constant a required by the expression could be the same for all ages above infancy. Reed and Merrell examine two other approximations to

${}_nq_x$, the first of which:

$$_nq_x = \frac{n\,_nm_x}{1+\frac{n}{2}\,_nm_x}$$

can be derived from Milne's formula for d_x; the other due apparently to Farr (1864, pp. xxiii—xxiv), and evident earlier to Gompertz (1825, paper 30 below):

$$_nq_x = 1 - \exp[-n\,_nm_x]\,.$$

Following their article we include derivations for both expressions.

T. N. E. Greville (1943) was able through ingenious expansions to derive each of these equations by working with the definitions of ${}_nm_x$ and ${}_nq_x$ and to show that the Reed-Merrell formula incorporates Gompertz' Law that the force of mortality is an exponential function of age. The assumption is appropriate at older ages and inappropriate for infancy, and thus defines the age range over which the Reed-Merrell formula is applicable. In the same article Greville discusses approximations to ${}_nL_x$ values where, as before, age groupings are wide (paper 8).

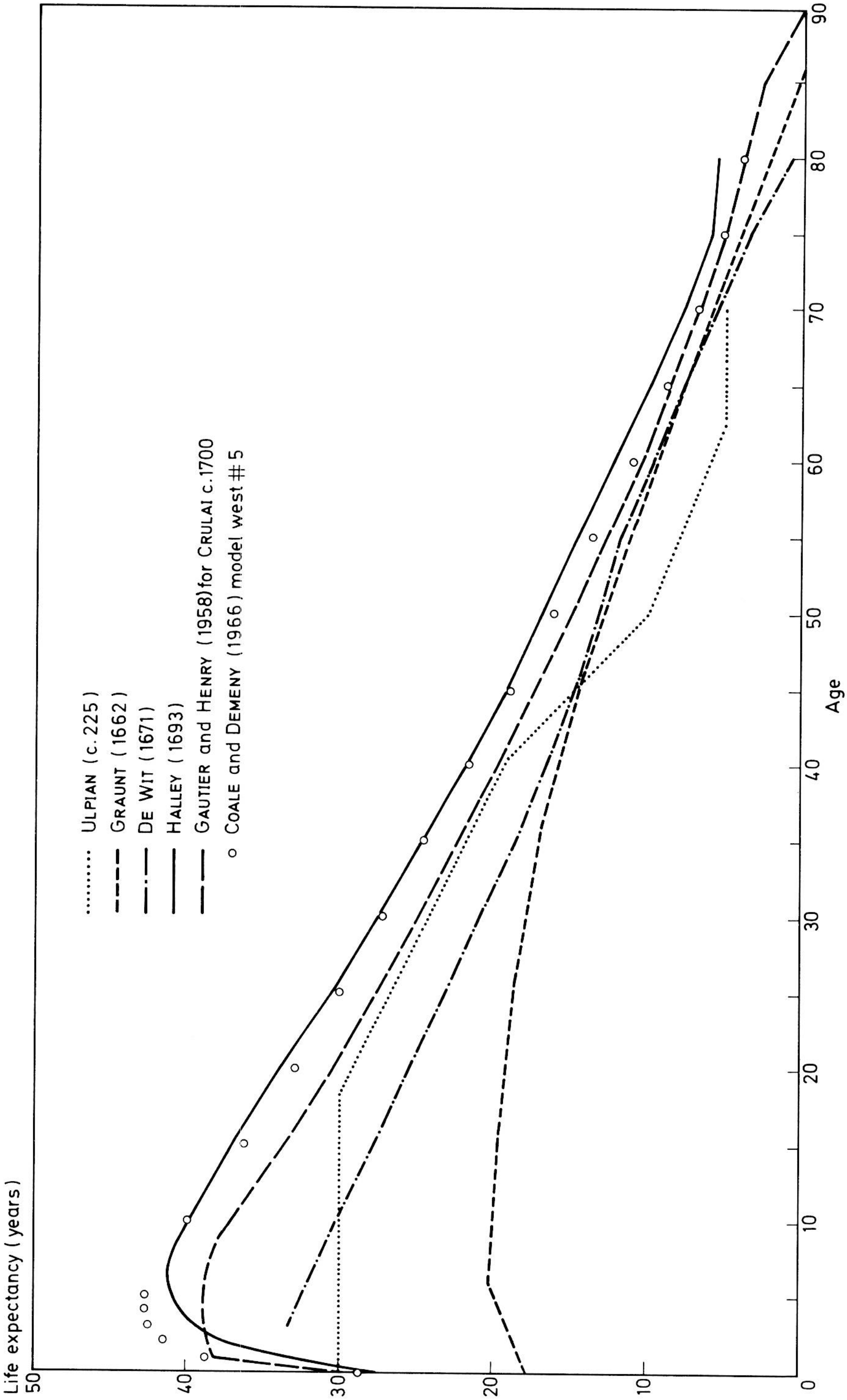
The methods used by Greville can be generalized to take advantage of the observed age structure of a population as well as its mortality schedule. In the simplest case this gives rise to the formula, due to Nathan Keyfitz and James Frauenthal (1975),

$$_nq_x = 1 - \exp\left[-n\,_nM_x + \frac{n}{48\,_nP_x}(_nP_{x+n} - {}_nP_{x-n})(_nM_{x+n} - {}_nM_{x-n})\right]$$

with as before the caveat that infancy requires separate consideration.

The chapter concludes with excerpts from the well known article by Edward Deevey (1947) in which he evaluates efforts that had been made up to that time to develop life tables for animal populations.

Sampling variances of life table terms are taken up in chapter 7.


Life expectancy (years)
Age
Ulpian (c. 225)
Graunt (1662)
De Wit (1671)
Halley (1693)
Gautier and Henry (1958) for Crulai c. 1700
Coale and Demeny (1966) model west # 5
0
10
20
30
40
50
60
70
80
90