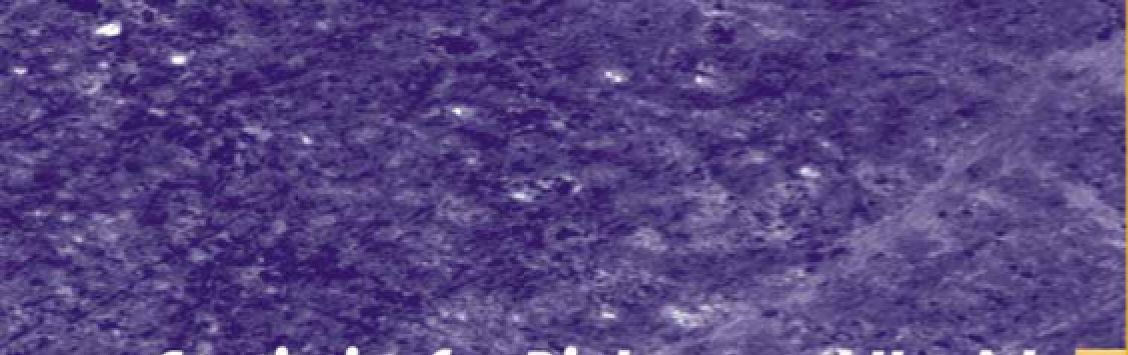


Statistics for Biology and Health

Nathan Keyfitz
Hal Caswell

Applied Mathematical Demography

Third Edition

Springer

Statistics for Biology and Health

Series Editors
M. Gail, K. Krickeberg, J. Samet, A. Tsiatis, W. Wong

Nathan Keyfitz
Hal Caswell

Applied Mathematical Demography

Third Edition

With 74 Illustrations

Springer

Nathan Keyfitz
Professor Emeritus
Department of Sociology
Harvard University
Cambridge, MA 02138
USA
nathankeyfitz@yahoo.com

Hal Caswell
Biology Department
Woods Hole Oceanographic Institution
Woods Hole, MA 02543-1049
USA
hcaswell@whoi.edu

Series Editors
M. Gail
National Cancer Institute
Rockville, MD 20892
USA

K. Krickeberg
Le Chatelet
F-63270 Manglieu
France

J. Samet
Department of Epidemiology
School of Public Health
Johns Hopkins University
615 Wolfe Street
Baltimore, MD 21205-2103
USA

A. Tsiatis
Department of Statistics
North Carolina State University
Raleigh, NC 27695
USA

W. Wong
Department of Statistics
Sequoia Hall
390 Serra Mall
Stanford University
Stanford, CA 94305-4065
USA

Library of Congress Cataloging-in-Publication Data
Keyfitz, Nathan, 1913–
Applied mathematical demography.—3rd ed. / Nathan Keyfitz, Hal Caswell.
p. cm. — (Statistics for biology and health)
Includes bibliographical references and index.
ISBN 0-387-22537-4 (alk. paper)
1. Demography—Mathematical models. 2. Population biology—Mathematical models.
3. Matrices. 4. Matrices—Data processing. I. Caswell, Hal. II. Title. III. Series.
HB849.51.K49 2005
304.6′01′51—dc22 2004052525

ISBN 0-387-22537-4 Printed on acid-free paper.

First edition © 1977 John Wiley & Sons, Inc.

© 2005, 1985 Springer Science+Business Media, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (EB)

9 8 7 6 5 4 3 2 1 SPIN 10951163

springeronline.com

To the students at Chicago, Berkeley, and Harvard who are responsible for any merit this book may have. NK

To Christine, with thanks for everything. HC

Preface to the Third Edition

When the second edition of this book was nearly exhausted, and orders were grinding to a halt, Springer urged me to do a third edition. At the age of 90 I said that was out of the question, but perhaps a co-author could be found. I spoke to several distinguished demographers, and the general drift of their replies was that the book was complete and well-rounded (my own opinion!) and there was nothing that could be usefully added.

We were monumentally wrong. We hadn't noticed the world of whales and birds and land animals, i.e., the world of biology. Hal Caswell has drastically broadened the perspective. Just as Alfred Lotka went far beyond the human species, so does Caswell. That wider perspective is fully incorporated in this third edition. It should not only be of interest to demographers, but to scholars of wide areas of biology. I can't thank Joel Cohen enough for realizing this and putting me on to Hal Caswell. His work on the third edition makes it a very different and much better book. Without Caswell, the book would have died with the second edition.

Cambridge, Massachusetts, USA NATHAN KEYFITZ

The request from Nathan Keyfitz that I collaborate on a new edition of *Applied Mathematical Demography* came out of the blue. The prospect was daunting. After all, when I began to study demographic analysis, Keyfitz's *Introduction to the Mathematics of Population* (1968) was the book that I turned to first. As for *Applied Mathematical Demography*, it had always seemed to me to embody a level of analytical insight that I could only envy.

But I accepted the invitation. Keyfitz's suggestion that we incorporate some of my earlier work on matrix population models seemed like a good way to complement the methods presented in the book, and to expand the range of their applications.

I am a demographer of plants and nonhuman animals. Some would call this an oxymoron, since the Greek root *demos* refers to people, and that's us. But there are precedents for taking more inclusive definitions of the Greek. Ecology and economics, for example, both come from the root *oikos*, referring to the household. Interpreting *demos* as referring to individuals, whether they are persons or not, lets demography apply across species. There is a long tradition of such crossover. Alfred J. Lotka is acknowledged as a founding father of both demography and ecology. Raymond Pearl used demographic methods to analyze the effects of toxic substances and crowding on fruit flies. Lee (1987) compared the density-dependence of the vital rates of human and non-human animals. Today, in studies of senescence, reproduction, and individual heterogeneity, the boundaries between animal and human studies are becoming increasingly blurred (Wachter et al. 1997, Carey 2003, Wachter and Bulatao 2003, Carey and Tuljapurkar 2003).

Matrix population models were developed in the 1940s by Patrick Leslie, an animal ecologist (Leslie 1945, 1948). They lay fallow until the mid-1960s when ecologists (Lefkovitch 1963, 1965) and human demographers (Keyfitz 1964, 1967, Rogers 1966) both rediscovered them. Both human demographers and ecologists needed to go beyond age-classified life table methods, because factors other than age affect the fates of individuals, regardless of species. In the 1970s this line of investigation was explored intensively by plant ecologists because the life cycles of plants, with their plasticity of development and multiple modes of reproduction, are particularly ill-suited to age-classified models. Now, stage-classified matrix population models are the most widely used framework for plant and animal demography, with applications in conservation biology, resource management, and pest control. All these ecological applications have parallels in human demography, in which the vital rates differ among individuals depending on age and other properties, and population dynamics depend on those vital rates and their variation in time and space. Modern mathematical software makes matrix methods not only theoretically appealing, but also practical tools for applied demographic analysis.

Much of the material on matrix population models in this book is extracted from the comprehensive treatment in *Matrix Population Models: Construction, Analysis, and Interpretation*, 2d edition (H. Caswell, 2001, Sinauer Associates, Sunderland, Massachusetts; www.sinauer.com). This book is referenced here as MPM.

We have made relatively few changes to the text of the second edition. Some topics have been rearranged, and we have added recent references, to permit a student access to current developments. We have purposely not removed many of the old references; they provide a valuable history of the development of the ideas, and are interesting in their own right. We have tried to unify the mathematical notation, we hope without confusing readers, whether more familiar with human demography or ecology.

On data in Applied Mathematical Demography

Applied Mathematical Demography is not—nor was in its first two editions—a book about demographic trends and patterns. Instead, it is a book about the theory underlying, and applied to, population dynamics. The empirical data shown here must be recognized for what they are: examples of how to apply analyses and interpret the results. Do not take them for a description of the state of the world.

For example, Section 10.1 analyzes the age distribution, especially the percentages under 15 and over 65, in three countries (Taiwan, the United States, and France) in 1965. Even when the first edition was published in 1977, this tabulation would hardly have qualified as even a cursory summary of age distributions around the world. Certainly the demography of each of these countries has changed in the last four decades. How would the comparison look now? How would other countries fit into the picture in the twenty-first century? The only way to find out would be to repeat the analysis with contemporary data. The same is true of all the other data appearing as examples here.

Asking students to conduct such analyses would be an excellent teaching tool. The exercise would not only help to clarify the theory; it would also hone two skills that, while critically important, are beyond the scope of this book. One is obtaining and evaluating demographic data. The other is implementing the calculations on the computer. The latter task has changed so much since 1985 (to say nothing of 1977) as to be unrecognizable. Mathematical software, readily available for personal computers, makes even complicated analyses easy. One of the best, especially for matrix calculations, is MATLAB®(The Mathworks, 3 Apple Hill Drive, Natick, MA 01760-2098, USA; www.mathworks.com), which was used for the matrix calculations in this book.

Acknowledgments

I thank the many excellent collaborators with whom I have had the good fortune to work. Special thanks to Joel Cohen, Christine Hunter, and Michael Neubert. I am grateful for support from the National Science Foundation (grants DEB-9973518 and DEB-0235692), the National Oceanic and Atmospheric Administration (grant NA03NMF4720491), the Environmental Protection Agency (grant R-82908901), the David and Lucille Packard Foundation (grant 2000-01740), and the Robert W. Morse Chair at the Woods Hole Oceanographic Institution. The writing process was aided by support from a Maclaurin Fellowship of the New Zealand Institute of Mathematics and its Applications, and by the hospitality of the Statis-

tics Department at the University of Auckland. I owe a special thanks to Sinauer Associates for their permission to use material from *Matrix Population Models*, where a much more detailed treatment of these models may be found.

Finally, my deepest thanks go to Nathan Keyfitz, for his contributions to mathematical demography over a remarkably long and productive career, and for the privilege of working with him on this book.

Woods Hole, Massachusetts, USA HAL CASWELL

Preface to the Second Edition (1985)

What follows is a new edition of the second in a series of three books providing an account of the mathematical development of demography. The first, *Introduction to the Mathematics of Population* (Addison-Wesley, 1968), gave the mathematical background. The second, the original of the present volume, was concerned with demography itself. The third in the sequence, *Mathematics Through Problems* (with John Beekman; Springer-Verlag, 1982), supplemented the first two with an ordered sequence of problems and answers.

Readers interested in the mathematics may consult the earlier book, republished with revisions by Addison-Wesley in 1977 and still in print. There is no overlap in subject matter between *Applied Mathematical Demography* and the *Introduction to the Mathematics of Population.* Three new chapters have been added, dealing with matters that have come recently into the demographic limelight: multi-state calculations, family demography, and heterogeneity.

NATHAN KEYFITZ

Preface to the First Edition (1977)

This book is concerned with commonsense questions about, for instance, the effect of a lowered death rate on the proportion of old people or the effect of abortions on the birth rate. The answers that it reaches are not always commonsense, and we will meet instances in which intuition has to be adjusted to accord with what the mathematics shows to be the case. Even when the intuitive answer gives the right direction of an effect, technical analysis is still needed to estimate its amount. We may see intuitively that the drop from an increasing to a stationary population will slow the promotion for the average person in a factory or office, but nothing short of an integral equation can show that each drop of 1 percent in the rate of increase will delay promotion to middle-level positions by 2.3 years.

The aim has been to find answers that will be serviceable to those working on population and related matters, whether or not they care to go deeply into the mathematics behind the answers. My earlier book, *Introduction to the Mathematics of Population*, had the opposite purpose of developing the theory, and mentioned applications mostly to illuminate the theory. Because of their different objectives there is virtually no overlap between the two books.

Population theory has developed at a sufficiently fast rate and in enough directions that no book of reasonable size can include all of its applications. A full development of theory ought to recognize not only age-specific rates of birth and death but also two sexes and two or more species. Age-specific rates can vary through time, and the theory can be stochastic in allowing to each individual member of the population his own separate risk, or deterministic in supposing that whatever probability applies to each individual is also the fraction of the population that succumbs to the risk. Thus population theory can be classified into the 16 categories shown in the accompanying table. By far the largest part of what is taken up in this book falls into the upper left category: it deals with one sex, usually female, and one species, man; it takes the age-specific rates of birth and death as fixed through time; it is deterministic rather than stochastic. This upper left-hand cell is conceptually the simplest of the 16, and it is mathematically the most tractable. But are these decisive arguments for its emphasis, given that real populations include two sexes; human populations interact with other species; birth and death rates change through time; and all life is stochastic?

The art of theory construction is to start with simple assumptions and then to introduce greater realism, which means more complexity, as required. On the path from simplicity to realism one must stop at a compromise point. My taste may not always be that of my readers; they may often say that a particular model I use is too simple, that they need to take into account factors that I neglect. This line of criticism is welcome, even though it leads to further and more difficult mathematics.

Classification of population theory

	Fixed rate		Changing rate	
	Deterministic	Stochastic	Deterministic	Stochastic
One sex				
One species	*			
Two or more species				
Two sexes				
One species				
Two or more species				

During 10 or more years of work on this book I have incurred more obligations than I can acknowledge or even remember. Students pointed out errors and obscurities; they helped in some cases by conspicuously failing to understand what I was saying and compelling me to think the matter through afresh. Colleagues looked at drafts and were generous with comments. Editors and referees of journals were helpful, especially Paul Demeny. No one is responsible for errors that remain but me.

Among these colleagues, students and correspondents who have been a source of ideas and a means of correcting errors, I recall especially William Alonso, Barbara Anderson, Brian Arthur, John C. Barrett, Ansley J. Coale, William Cochrane, Joel E. Cohen, Prithwis Das Gupta, Paul Demeny, Lloyd Demetrius, James Dobbins, Barry Edmonston, Jamie Eng, Thomas Espenshade, Noral Federici, Griffith Feeney, Gustav Feichtinger, Jair Fereira-Santos, James Frauenthal, A. G. Fredrickson, Robert Gardiner, Campbell Gibson, Noreen Goldman, Antonio Golini, David Goodman, Leo A. Goodman, Louis Henry, Jan Hoem, Barbara Keyfitz, S. Krishnamoorthy, Paul Kwong, Juan Carlos Lerda, John Lew, Gary Littman, Robert Lundy, James G. March, Robert Mare, George Masnick, John McDonald, David McFarland, Geoffrey McNicoll, Paul Meier, Jane Menken, Walter Meyer, George C. Myers, Frank Oechsli, Beresford Parlett, James Pick, Robert G. Potter, Jr., Samuel H. Preston, Thomas Pullum, Robert Retherford, Roger Revelle, Andrei Rogers, Normal Ryder, Paul Samuelson, Robert Sembiring, David P. Smith, Leroy O. Stone, Richard Stone, Michael Stoto, Michael Teitelbaum, Harold A. Thomas, Robert Traxler, James Trussell, Etienne Van de Walle, Kenneth Wachter, Frans Willekens and Harrison White, every one of whom has had some effect on this book. To be singled out especially for their help in the final stages are Noreen Goldman, David P. Smith, Gary Littman, S. Krishnamoorthy, and Michael Stoto, who read much of the manuscript and all of the page proof.

The most important acknowledgment is to my wife, who edited and typed the manuscript, some parts of it many times.

Cambridge, Massachusetts, USA NATHAN KEYFITZ
February 1977

Acknowledgments

An early version of Section 1.4 appeared as “How many people have lived on the earth?” *Demography* **3**: 581–582 (1966).

Section 1.10 was published as “A general condition for stability in demographic processes,” *Canadian Studies in Population* **1**: 29–35 (1974).

The life table method of Chapter 2 first appeared as “An improved life table method,” *Biometrics* **31**: 889–899 (1975) (with James Frauenthal).

Material in Chapter 3 is extracted from Chapters 2 and 3 of *Matrix Population Models: Construction, Analysis, and Interpretation*, 2d edition, Sinauer Associates, Sunderland, Massachusetts, USA. Used by permission.

Much of Chapter 4 was included in “Mortality comparisons: The male-female ratio,” *Genus* **31**: 1–34 (1975) (with Antonio Golini).

Section 5.8 is a condensed version of “Individual mobility in a stationary population,” *Population Studies* **27**: 335–352 (1973).

Chapter 7 contains material extracted from Chapter 4 of *Matrix Population Models: Construction, Analysis, and Interpretation*, 2d edition, Sinauer Associates, Sunderland, Massachusetts, USA. Used by permission.

Some of Section 8.5 appeared in “Migration as a means of population control,” *Population Studies* **25**: 63–72 (1971).

Section 8.6 is an abridged version of “On the momentum of population growth,” *Demography* **8**: 71–80 (1971).

Parts of Section 8.8 were published as "Age distribution and the stable equivalent," *Demography* **6**: 261–269 (1969).

Chapter 11 contains material from Chapter 5 of *Matrix Population Models: Construction, Analysis, and Interpretation*, 2d edition, Sinauer Associates, Sunderland, Massachusetts, USA. Used by permission.

Some parts of Chapter 12, as well as Section 14.3, were included in "On future population," *Journal of the American Statistical Association* **67**: 347–363 (1972).

Parts of Chapter 13 come from Chapters 9 and 10 of *Matrix Population Models: Construction, Analysis, and Interpretation*, 2d edition, Sinauer Associates, Sunderland, Massachusetts, USA. Used by permission.

Section 14.5 is an abridged version of "Backward population projection by a generalized inverse," *Theoretical Population Biology* **6**: 135–142 (1974) (with T. N. E. Greville).

The substance of Sections 14.10 and 14.11 was included in "Population waves," in *Population Dynamics* (edited by T. N. E. Greville), Academic Press, 1972, pp. 1–38.

Parts of Chapter 15 are from "Family formation and the frequency of various kinship relationships," *Theoretical Population Biology* **5**: 1–27 (1974) (with Leo A. Goodman and Thomas W. Pullum), and "Addendum: Family formation and the frequency of various kinship relationships," *Theoretical Population Biology* **8**: 376–381 (1975) (with Leo A. Goodman and Thomas W. Pullum).

An early statement of Section 16.1 appeared as "How birth control affects births," *Social Biology* **18**: 109–121 (1971).

Chapter 20 is reprinted with some modifications from "How do we know the facts of demography?" *Population and Development Review* **1**: 267–288 (1975).

Contents

A Word About Notation

Throughout the book, matrices are denoted by boldface capital letters ($\mathbf{A}$) and their elements by lower-case letters (a_{ij}). Vectors are denoted by boldface lower-case letters ($\mathbf{w}$) and their elements by lower-case letters (w_i). Vectors are column vectors by default. Occasionally, it is necessary to refer to the elements of subscripted matrices; in such cases, e.g., $a_{ij}^{(k)}$ is the (i,j) element of $\mathbf{A}_k$. The transpose of $\mathbf{A}$ is $\mathbf{A}^{\mathsf{T}}$. The complex conjugate transpose of $\mathbf{A}$ is $\mathbf{A}^*$. The determinant of $\mathbf{A}$ is denoted $\det(\mathbf{A})$ or $|\mathbf{A}|$. Unless otherwise specified, all logarithms are natural logarithms.

It has not been possible to develop a completely uniform set of symbols. Population size, in numbers of individuals, is usually denoted by N (as a scalar) or $\mathbf{n}$ (as a vector). The age distribution as a density function is usually denoted $p(a)$, so that, for example, the total population would be given by $N = \int_0^\infty p(a)da$. Double subscripts (e.g., ${}_5q_x$) are used to denote quantities defined by an age (the right subscript) and an age interval (the left subscript). The example just given is the probability that an individual aged x will die in the next 5-year age interval; i.e., between ages x and $x+5$ (see p. 30). There are exceptions, but symbols are always defined in the context of their use.

Equations are numbered within sections, so (12.2.3) or “equation 12.2.3” refers to the third numbered equation in Section 12.2.

1

Introduction: Population Without Age

Age is a characteristic variable of population analysis; much of the theory of this volume will be concerned with the effect of age distribution, either fixed or changing. But age is not the only characteristic variable, and part of this book will deal with methods capable of describing the effects of other variables besides, or in addition to, age. Abstraction is necessary in demographic as in other theory; is it possible to abstract even from age and still obtain results of value?

To represent a population as a number varying in time, and in disregard even of its age composition, is like treating the earth as a point in space—though too abstract for most purposes, it is useful for some. Partly as an introduction to the rest of the book, the nine sections of this chapter take up questions to which answers can be obtained without reference to age.

These topics vary in depth and subtlety. The first deals only with the use of logarithms in base 2, that is, the number of doublings to go from one total to a larger total. Such logarithms show dramatically how close birth rates must have been to death rates throughout human history. Next comes an aspect of bisexual reproduction, and some of the difficulty it causes; consideration of one sex at a time is a powerful simplifying device for understanding demographic mechanisms. Another question concerns a population made up of a number of subpopulations, each increasing at a *fixed* rate; we will see at what *increasing* rate the total population is expanding. Attention will be given also to the migration problem, from which the one-sex population recognizing age will emerge as a special case; that a proposition concerning age follows directly from one on migration illustrates the advantage of a formal treatment.

1.1 Definitions of Rate of Increase

Population change is expressed in terms of *rate of increase*, a notion different in an important respect from the rate of speed of a physical object such as an automobile. If a country stands at 1,000,000 population at the beginning of the year, and at 1,020,000 at the end of the year, then by analogy with the automobile its rate would be 20,000 persons per year. To become a demographic rate this has to be divided by the population, say at the start of the year; the population is growing at a rate of 20,000/1,000,000 or 0.02 per year. More commonly it is said to be growing at 2 percent per year, or 20 per thousand initial population per year.

The two kinds of rate are readily distinguished in symbols. The analogue of the physical rate x' in terms of population at time t and $t+1$, N_t and N_{t+1}, is

$$x' = N_{t+1} - N_t \qquad \text{or} \qquad N_{t+1} = N_t + x',$$

while the demographic rate x is

$$x = \frac{N_{t+1} - N_t}{N_t} \qquad \text{or} \qquad N_{t+1} = N_t(1+x).$$

Just as x' is similar to a rate of velocity, so x is similar to a rate of compound interest on a loan. Both are conveniently expressed in terms of a short time period, rather than 1 year, and in the limit as the time period goes to zero they become derivatives:

$$x' = \frac{dN(t)}{dt} \qquad \text{and} \qquad x = \frac{1}{N(t)}\,\frac{dN(t)}{dt}.$$

The rates x' and x give very different results if projected into the future. A population continuing to grow by 20,000 persons per year would have grown by 40,000 persons at the end of 2 years, and by 60,000 persons at the end of 3 years. One increasing at a rate of 0.02 would be in the ratio 1.02 at the end of 1 year, in the ratio 1.02^2 at the end of 2 years, and so on. The latter is called *geometric increase*, while the fixed increment is called *arithmetic increase.*

Thus the very notion of a rate of increase, as it is defined in the study of population, seems to imply geometric increase. In fact, it implies nothing for the future, being merely descriptive of the present and the past; rates need not be positive and can be zero or negative. Moreover, being positive today does not preclude being negative next year: ups and downs are the most familiar feature of the population record.

The future enters demographic work in a special way—as a means of understanding the present. To say that the population is increasing at 2 percent per year is to say that, if the rate continued, the total would amount to $1.02^{10} = 1.219$ times the present population in 10 years, and $1.02^{1000} = 398{,}000{,}000$ times the present population in 1000 years. If this tells anything, it says that the rate cannot continue, a statement about the future of a kind that will be studied in Chapter 8.

Still, the conditional growth rate whose hypothetical continuance is helping us to understand the present can be defined in many ways. In particular, in the example above, it could be an increase of 20,000 persons per year just as well as of 0.02 per year. Arithmetical and geometrical projections are equally easy to make; why should the latter be preferred?

The reason for preferring the model of geometric increase is simple: constancy of the elements of growth translates into geometric increase. If successive groups of women coming to maturity have children at the same ages, and if deaths likewise take place at the same ages, and if in- and out-migration patterns do not change, then the population will increase (or decrease) geometrically. Any fixed set of rates that continues over time, whether defined in terms of individuals, families, or age groups, ultimately results in increase at a constant ratio. That any fixed pattern of childbearing, along with a fixed age schedule of mortality, implies long-run geometric increase is what makes this kind of increase central in demography. (The stationary population and geometric decrease can be seen as special cases, in which the ratios are unity and less than unity, respectively.)

1.2 Doubling Time and Half-Life

The expression for geometric increase gives as the projection to time t,

$$N_t = N_0(1+x)^t, \tag{1.2.1}$$

where x is now the fraction of increase per unit of time. The unit of time may be a month, a year, or a decade, so long as x and t are expressed in the same unit. Any one of N_0, N_t, x, or t may be ascertained if the other three are given. When the quantity x is negative, the population is decreasing; and when t is negative, the formula projects backward in time. If x is the increase per year as a decimal fraction, then $100x$ is the increase as percentage and $1000x$ the increase per thousand population.

We are told that the population of a certain country is increasing at $100x$ percent per year, and would like to determine by mental arithmetic the population to which this rate of increase would lead if it persisted over a long interval. Translating the rate into doubling time is a convenience in grasping it demographically as well as arithmetically.

Since annual compounding looks easier to handle than compounding by any other period, we will try it first. We will see that it leads to an unnecessarily involved expression, whose complication we will then seek to remove.

If at the end of 1 year the population is $1+x$ times as great as it was at the beginning of the year, at the end of 2 years $(1+x)^2, \ldots,$ and at the end of n years $(1+x)^n$ times, then the doubling time is the value of n that satisfies the relation

$$(1+x)^n = 2. \tag{1.2.2}$$

To solve for n we take natural logarithms and divide both sides by $\log(1+x)$:

$$n = \frac{\log 2}{\log(1+x)} = \frac{0.693}{\log(1+x)}. \tag{1.2.3}$$

The Taylor series for the natural logarithm $\log(1+x)$ is

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots. \tag{1.2.4}$$

Entering the series for $\log(1+x)$ in (1.2.3) gives for doubling time

$$n = \frac{0.693}{\log(1+x)} = \frac{0.693}{x - (x^2/2) + (x^3/3) - \cdots}. \tag{1.2.5}$$

The right-hand expression can be simplified by disregarding terms beyond the first in the denominator. We could write $n = 0.693/x$, but for values of x between 0 and 0.04, which include the great majority of human populations, arithmetic experiment with (1.2.5) when x is compounded annually shows that it is, on the whole, slightly more precise to write $n = 0.70/x$ or, in terms of x expressed as a percentage (i.e., $100x$), to write $n = 70/100x$.

The expression

$$n = \frac{70}{100x} \tag{1.2.6}$$

is a simple and accurate approximation to (1.2.3). It tells us that a population increasing at 1 percent doubles in 70 years, and so forth. For Ecuador in 1965, $100x$ was estimated at 3.2 percent, so doubling time would be $n = 70/3.2 = 22$ years. The 1965 population was 5,109,000, and if it doubled in 22 years the 1987 population would be 10,218,000; this compares well with the more exact $(5{,}109{,}000)(1.032)^{22} = 10{,}216{,}000$. Similarly a further doubling in the next 22 years would give a population in the year 2009 of 20,436,000, as compared with $(5{,}109{,}000)(1.032)^{44} = 20{,}429{,}000$. The formula $n = 70/100x$ is plainly good enough for such hypothetical calculations.

The Period of Compounding. All this is based on a definition of x by which the ratio of the population at the end of the year to that at the beginning of the year is $1+x$. Interest calculations are said to be compounded annually on such a definition. A different definition of rate of increase simplifies this problem and others and without any approximation gets rid of the awkward series in the denominator of (1.2.5).

Suppose that x is compounded j times per year; then at the end of 1 year the population will have grown in the ratio $(1 + x/j)^j$. Is there a value j

more "natural" than 1? There is: infinity. When the rate x is compounded instantaneously, we will call it r. We then have

$$\lim_{j\to\infty}\left(1+\frac{r}{j}\right)^j = e^r,$$

where e is the base of natural logarithms and equals 2.71828. [To see how fast this converges try $(1 + r/100{,}000)^{100{,}000}$ for $r = 1$ and show that the results agree with e to five significant figures. To how many does $(1 + r/1{,}000{,}000)^{1{,}000{,}000}$ agree?]

Our calculations from this point will suppose such instantaneous compounding of population, unless otherwise stated; we will take it that a population increasing at rate r will equal e^r at the end of 1 year and e^{nr} at the end of n years. Thus, instead of saying that a population is increasing at 3.20 percent per year compounded annually, we will make the equivalent statement that it is increasing at 3.15 percent per year compounded continuously. Most formulas are thereby made simpler.

This device would equally simplify financial calculations. Tradition has caused these to be made on the basis of a variety of compounding periods—semiannually, quarterly, monthly, daily. If the rate compounded annually is 3.2 percent, the equivalent rates (percents) for other compounding periods are as follows:

Semiannually	3.1748
Quarterly	3.1623
Monthly	3.1540
Weekly	3.1508
Daily	3.1500
Continuously	3.1499

[Show that in general the rate compounded n times per year equivalent to the rate x compounded annually is

$$x_n = n\left[(1+x)^{1/n} - 1\right],$$

and verify the above numbers]. To quote interest rates, like rates of population increase, continuously avoids a good deal of unnecessary complication.

With r defined as an annual rate compounded continuously, the equation for the doubling time n becomes $e^{nr} = 2$, whose solution is just $n = 0.693/r$ or $69.3/100r$. For Mexico the rate of population increase is about 3.5 percent per year compounded momently. The population therefore doubles in $69.3/3.5 = 20$ years, quadruples in 40 years, . . . , and multiplies by $32 = 2^5$ in 100 years (strictly, 99.02 years).

For decreasing populations the same applies except that now we want halving time, and so we equate e^{nr} to 0.5, and obtain $n = -0.693/r$. If

$r = -0.01$, the $n = -0.693/(-0.01) = 69$ years, and this is the half-life. At the level of abstraction of the present section the continuous model of population is identical with that of radioactive substances.

The expressions are equally usable to find r, given n; if we know that a population has doubled in $n = 100$ years, its annual rate of increase compounded momently must be $r = 0.693/100 = 0.00693$. In general, $r = 0.693/n$.

1.2.1 Application to Human History

The simple apparatus of (1.2.6), or else doubling time $= 0.693/r$, can be applied to show how nearly equal birth rates have been to death rates for most of human history. Our race must have started with a sizable group; and even if there had been periodic censuses at the time, questions of definition would have made their interpretation controversial. Suppose for this example that one human couple living a million years ago has multiplied to equal the present human total of about 4 billions, and that it has grown uniformly by geometric increase over the entire period. Since $2^{10} = 1024$, multiplication by 1000 is about 10 doublings, multiplication by 1 billion is 30 doublings, and by 2 billion is 31 doublings. If just 31 doublings have occurred over the last million years, then each must have taken 1,000,000/31 or just over 32,000 years on the average.

But we saw that a population that takes n years to double is increasing at $r = 0.693/n$ per year. If the doubling time n is 32,000 years, the annual rate of increase is $0.693/32{,}000 = 0.000022$. In a population closed to migration (which the earth as a whole must be) the rate of increase must equal the birth rate (say b) less the death rate d: $r = b - d$, or $b = r + d$. Thus, with a doubling time of 32,000 years, if death rates were 40.000 per thousand population, then birth rates must have been 40.022 per thousand. Over the million years birth and death rates must on the average have been this close.

It is fanciful to average over a million years; consider now the shorter period from the time of Augustus, with a world population of about 250 million, to the mid-seventeenth century with 545 million people, a ratio of 2.18, or an average rate of increase of $\log 2.18/1650 = 0.00047$. If death rates were 40.00 per thousand, then birth rates must have been 40.47 per thousand on the average; again simple arithmetic gives an idea of how close, on the whole, birth and death rates have been over historical time.

To show further how exceptional is the situation in which death rates in many places have dropped to 20 per thousand while births stay at 40, let us see how short a time it would have taken to arrive at the world's present 4 billion population with birth rates that are 20 per thousand higher than death rates. Doubling would take 35 years, so 31 doublings would take $31 \times 35 = 1085$ years. A single couple starting two centuries before

William the Conqueror would have produced more than the present world population.

Lest anyone think that this argument is original here, consider this quote from Boldrini's *Demografia*, published in Milan in 1956. Given, he says, one couple and the rate of increase in population in the first half of the twentieth century of $r = 0.0082$, how long would it take to attain the world population of 2438 million? He finds 2567 years, or about the elapsed time from the founding of Rome (Boldrini 1956, p. 46). [Show that the numbers given work out to 2551 years if the rate 0.0082 is compounded continuously, or to 2562 years if it is compounded annually. Minor errors of this kind are easy to detect with a modern hand calculator.] Boldrini then goes on to give other variants of the same thought.

Long before Boldrini, Bortkiewicz (1911, pp. 75, 76) showed the usefulness of (1.2.2) and the following expressions for doubling time, and he in turn had precursors. Süssmilch (1788, pp. 291ff) was interested in doubling time and asked his contemporary, Leonard Euler, about reproduction under some extreme circumstances. Euler provided an account that comes close to modern stable theory.

1.2.2 Logarithms to Various Bases

The use of doubling time or half-life in such calculations is arbitrary; we might prefer to do the arithmetic by means of tripling time or one-third-life. Tripling time is equal to $n = (\log 3)/r = 1.099/r$. Instead of using 2 or 3 for the base of calculation, we might use 10; the time for tenfold increase is $n = (\log 10)/r = 2.303/r$. A population increasing at 3.5 percent per year will multiply by 10 in $n = 2.303/0.035 = 66$ years, by 100 in 132 years, and so on.

The number of doublings that correspond to a specified ratio of a final to an initial population is called the *logarithm of that ratio to base 2*; Thus the logarithm of 1000 to base 2 is the power to which 2 has to be raised to equal 1000. Since $2^{10} = 1024$, we can say approximately that the logarithm of 1000 to base 2 is 10; multiplication by 1000 is about equivalent to 10 doublings (more exactly, to 9.97 doublings). Then multiplication by a million is equivalent to 20 doublings, and by a billion to 30 doublings, so the logarithms of these numbers to base 2 are 20 and 30, respectively. For thinking about a steadily increasing population, doubling time is a more convenient unit than years or centuries, and logarithms to base 2 are correspondingly valuable.

To translate from number of doublings to number of triplings is to translate from logarithms to base 2 to logarithms to base 3. If x is the number of doublings and y the number of triplings, then

$$2^x = 3^y.$$

We take natural logarithms of both sides to obtain

$$x \log 2 = y \log 3,$$

which is the same as

$$y = \frac{\log 2}{\log 3} x = \frac{0.693}{1.099} x = 0.63x. \tag{1.2.7}$$

If x doublings constitute an increase in any given ratio R, then $0.63x$ triplings constitute the same increase. We found that to go from 2 persons to 4 billion persons requires 31 doublings. This is equivalent to $31 \times 0.63 = 20$ triplings.

In general, if x is the log of R to base a, and y is the log of R to base b, so that

$$a^x = b^y,$$

then taking logarithms to an arbitrary base gives

$$x \log a = y \log b,$$

whatever the base. If the base is b, then, since we know (from the fact that $b^1 = b$) that $\log_b b = 1$, we have

$$y = x \log_b a. \tag{1.2.8}$$

This says that to go from $\log_a$ to $\log_b$ we need merely multiply by $\log_b a$.

As a particular case of $y = x \log_b a$, we put $b = e$, and obtain

$$\log_a R = \frac{\log_e R}{\log_e a} \equiv \frac{\log R}{\log a}.$$

For purposes of calculation one usually starts with logarithms to base e, designated as log, and this formula changes to an arbitrary base a. For example,

$$\log_2 7 = \frac{\log 7}{\log 2} = \frac{1.946}{0.693} = 2.807.$$

It is convenient also to know that $\log_b a$ is the reciprocal of $\log_a b$. [Prove this.]

1.2.3 Prospective Possible Doublings

Doubling times provide a quick perspective on future limitations as well as on past history. We saw that of the 31 doublings from the hypothetical primeval couple about 4 have occurred since the time of Augustus. Filling all of the land surface of the planet Earth to the density of Manhattan would take us to only 10 more doublings from where we are. Spreading with equal density over the bottoms of the oceans as well would permit about 2 additional doublings. More realistic considerations suggest that the 31 doublings so far vouchsafed to man, with at most one or two more, are all he will ever have. The number 31, plus or minus one or two, may be thought of as a constant of nature.

1.3 One-Sex Versus Two-Sex Models: Descendants of the Pilgrim Fathers

Bisexual reproduction means that each of us has two parents, four grandparents, and eight great-grandparents, in general, 2^i great^{i-2}-grandparents. These are outside numbers; there must have been some inbreeding during the last 30 generations, since 2^{30} is 10^9 or 1 billion, and this has to be compared with the number of people on earth 30 generations ago. The number then living would not have sufficed to provide each of us with 2^{30} distinct great^{28}-grandparents, for going back 30 generations takes us to about the year A.D. 1200, when the world population was barely half a billion. Notwithstanding the ultimate unity of the human race, moreover, the ancestors of each of us since A.D. 1200 must have been a subgroup of humanity.

From this viewpoint ancestry is simpler to deal with than progeny, for each of us can have zero, one, two, or more children, but each of us must have had exactly two parents. Thought of as a branching process, each step backward in time is a simple bifurcation; forward in time each step can involve up to 10 or more forkings. But this simplicity of the backward process is more than offset by intermarriage.

Demography must transcend the detail of family trees, just as macroeconomics must aggregate the detail of individual transactions. The various sections of this book can be thought of as different ways of summarizing genealogies for the purpose of drawing conclusions on populations. A useful summarizing device is to consider only one sex at a time.

A large part of our work will treat only the female side of the population, or only the male side. This is not to imply that the other sex is not necessary, but rather to suppose that it exists in whatever numbers are required to produce the growth in the sex being followed. The present section is intended to show how this device gives clear answers to questions otherwise indeterminate. We will use the number of descendants of the Pilgrim fathers—not in itself an important demographic issue—as an extreme example of the uncertainty introduced when both sexes are considered simultaneously.

We saw that an increase of $100x$ percent per year implies a doubling time of about $70/100x$. Evidence from the total population of the United States and that of French Canada suggests that 2 percent per year, or a doubling in 35 years, is about right for an American group over the three and one-half centuries since the landing of the Pilgrims in 1620.

The population of the United States increased from 3,929,000 in 1790 to 204 million in 1970, or 51.9 times. This is 5.7 doublings, or say 6, and implies about $180/6 = 30$ years per doubling. The increase does include immigration, however, and 30 years is therefore too short a time for the doubling of a native group.

A self-contained population that probably had birth rates somewhat higher than those of the Pilgrims is the French Canadians, including the Franco-Americans. There were fewer than 10,000 original newcomers, say of average time of arrival 1700. By 1970 there were about 5,500,000 French Canadians; over the course of 270 years they had multiplied by 550, which is 9 doublings and hence also about 30 years per doubling. Their growth was almost entirely due to excess of births over deaths, with immigration making only a trifling contribution, but their birth rates continued to be high long after those of the Pilgrims' descendants had fallen.

This broadly supports a figure of about 35 years per doubling for the Pilgrims, which also happens to be convenient for the arithmetic of our example. Following the male line for each Pilgrim (except that for the few women Pilgrims we follow the female line) would give 10 doublings in the 350 years from 1620 to 1970. (The male line is sons of sons of sons, etc.; the female line, daughters of daughters of daughters. In this way of reckoning, the son of a daughter of Elder Brewster is not taken as a descendant.)

About half the Pilgrims are said to have died during the first hard winter; let us consider only the survivors, say 50 in number. Then each Pilgrim would have $2^{10} = 1024$ descendants through the line for his own sex, and all of them together would have 50×1024 or somewhat over 50,000 male descendants of male Pilgrims and female descendants of female Pilgrims. This would also approximate the present total if the Pilgrims had been a nuptially isolated group, which is to say their descendants were separated from the rest of the population and were always able to find spouses from their own numbers. This would have required an equal number of each sex from the beginning (or the appointment of a number of honorary female Pilgrims in the first generation).

More difficult is the question of descendants in view of the fact that they actually were not isolated. At the other extreme, if they could always have married spouses who were not descended from Pilgrims, and we count all their children rather than the line for one sex, the expansion of the Pilgrim-descended population would be at twice the rate of the preceding paragraph. Now each would have four descendants in the time supposed for two above, which is to say that time for doubling would average just one-half of 35 years. Thus each would have 2^{20} or over 1 million descendants in the 350 years to 1970, and the 50 Pilgrims surviving through the first winter would together have over 50 million descendants.

The number of persons now alive who can claim descent from the Pilgrims is thus at least 50,000 (if they were an isolated subpopulation, and the fact that the numbers of the two sexes were not equal at the beginning is disregarded) and at most 50,000,000 (if they completely avoided marrying one another). Without data on the extent of intermarriage we have no way of narrowing the range between 50,000 and 50,000,000 descendants. [What are some of the difficulties of collecting data on intermarriage?]

The purpose of this account being merely to illustrate the indeterminacy of a two-sex model, we have simplified the data. A more realistic analysis would start with the fact that just 23 of the Pilgrims founded continuing families, and would use historical records to ascertain with more precision just what their family sizes were in successive generations. Apparently the Mayflower Descendants Society counts about 15,000 members, all of whom have proved their descent, but this number is far short of the actual total of descendants.

That the one-sex problem gives the simple and unique answer of 50,000 on our assumptions, whereas the corresponding two-sex problem leaves us in the range of 50,000 to 50 million, is only one aspect of the difficulty. Another is the effect on marriages and births of adding a number of males to a population, as against the (presumably greater) effect of adding the same number of females. Hunting female rabbits affects reproduction more than hunting male rabbits—how much more depends on how actively the remaining males get around. Satisfactory answers to such questions are not easily found. They cannot be obtained without facts or assumptions regarding individual behavior of a more detailed kind than demography ordinarily introduces.

In most of the following chapters the population will be assumed closed to migration and will usually appear to consist only of females or only of males. We will not care how many individuals can trace their ancestry back to a given origin, but we will want to know how fast a closed population is increasing, and what determines its age distribution and other features. For that purpose we will consider the two sexes separately and avoid examining the availability of mates, a question whose difficulties it is the purpose of the present section to suggest.

The fact that the one-sex model gives simple answers to difficult questions, and that under a considerable range of circumstances these answers are realistic, makes it a positive achievement.

1.4 How Many People Have Lived on the Earth?

The number of people who have lived on the earth can be estimated by the births at two or more points in time and supposing uniform increase between these points. First we find the rate of increase r when births are given.

If in year t_1 the births were n_1, and in year t_2 (later than t_1) were n_2, the ratio of increase was n_2/n_1. In terms of r, the average annual rate of growth compounded momently, this same ratio of increase must be $e^{r(t_2-t_1)}$. Equating these two expressions,

$$\frac{n_2}{n_1} = e^{r(t_2-t_1)},$$

and taking logarithms of both sides and solving for r gives

$$r = \frac{\log n_2 - \log n_1}{t_2 - t_1}. \tag{1.4.1}$$

Now we have to find the total births over a long interval of time. If during the interval from t_1 to t_2 the rate of increase was at all times exactly r, in any intermediate year t the births were $n_1 e^{r(t-t_1)}$, and integrating this gives the total births as

$$\int_{t_1}^{t_2} n_1 e^{r(t-t_1)}\,dt = \frac{n_1}{r}\left[e^{r(t_2-t_1)} - 1\right] = \frac{n_2 - n_1}{r}, \tag{1.4.2}$$

using the fact that $e^{r(t_2-t_1)} = n_2/n_1$.

We want to express the births over the interval in a way that does not require explicit knowledge of r. To do so we substitute the value of r from (1.4.1) in (1.4.2) and so obtain for the births that occurred in the interval, which is the same as the total persons who lived, the simple result

$$\text{Persons who lived} = \frac{(n_2 - n_1)(t_2 - t_1)}{\log n_2 - \log n_1}. \tag{1.4.3}$$

A widely quoted article of the Population Reference Bureau (Cook 1962) estimated annual births at four points in the history of mankind as follows:

t	n
600,000 B.C.	1
6000 B.C.	250,000
A.D. 1650	25,000,000
A.D. 1962	110,000,000

For the first of the three intervals between these four points, that between 600,000 B.C. and 6000 B.C., the expression on the right of (1.4.3) gives persons who lived as

$$\frac{(n_2 - n_1)(t_2 - t_1)}{\log n_2 - \log n_1} = \frac{(250{,}000 - 1)(600{,}000 - 6000)}{12.4292 - 0} = 11.9 \times 10^9,$$

and for the other two intervals we have similarly 41.1×10^9 and 17.9×10^9, respectively. Adding the three intervals gives $11.9 + 41.1 + 17.9 = 70.9$ billion (not 77 billion, as the Population Reference Bureau calculated).

If we had more points at which population could be reasonably assessed, the estimate would turn out to be different. Assuming a longer stretch of time than 600,000 years would raise the number, and supposing arithmetic rather than geometric increase would raise it greatly. With the high mortality prevailing in most times and places, only about half of those born lived to maturity, so the number of *adults* who have ever lived is far less than the number of persons.

Deevey (1950) has looked into the past and present populations of the planet in more detail, and makes estimates for 12 points of time from 1 million years ago to the year 2000. He is not satisfied with the exponential curve, but applies three successive logarithmic curves, one for each of the three main phases of human evolution—toolmaking, agriculture, and scientific–industrial. He concludes that 110 billion will be the world total over all time to the year 2000.

Taking these and other (e.g., Winkler 1959, Fuchs 1951) calculations together suggests that the nearly 4 billion persons now alive* constitute between 4 and 6 percent of those who have ever lived, a proportion that would be somewhat smaller if we moved human origins back in time. The corresponding fraction for adults is greater, and the fraction of those with specific modern occupations who have ever lived, for instance engineers, much greater.

We now drop the homogeneity and suppose subpopulations having different rates of increase.

1.5 A Mixture of Populations Having Different Rates of Increase

A population of initial size Q growing at rate r numbers Qe^{rt} at time t, r being taken as fixed and the population as homogeneous. Now suppose heterogeneity—a number of subpopulations, of which the ith is initially Q_i growing at rate r_i, so that at time t the total number is $N(t) = \sum_i Q_i e^{r_i t}$. We will show that the total never stabilizes, that its rate of increase forever increases, and that the composition constantly changes.

The definition of rate of increase over a finite time δ may be written as

$$\frac{1}{N(t)} \frac{N(t+\delta) - N(t)}{\delta},$$

and in the limit as δ tends to zero this becomes

$$\frac{1}{N(t)} \frac{dN(t)}{dt}. \tag{1.5.1}$$

Thus at time t the derivative of the total population, $dN(t)/dt$, when divided by $N(t)$, provides us with the rate of increase, which for our mixture

---

*That number is now (2004) six billion; for discussions of the past and future of global human population growth, see Cohen (1995) and Bongaarts and Bulatao (2000).

of populations can be written as $\bar{r}(t)$. The derivative of $N(t) = \sum_i Q_i e^{r_i t}$ is $dN(t)/dt = \sum_i Q_i r_i e^{r_i t}$. Hence the rate of increase must be

$$\bar{r}(t) = \frac{1}{N(t)} \frac{dN(t)}{dt} = \frac{\sum_i Q_i r_i e^{r_i t}}{\sum_i Q_i e^{r_i t}},$$

which is the arithmetic mean of the r_i, each weighted by its population at time t.

The change in this mean rate of change, by the rule for the derivative of a ratio of two functions $u(t)$ and $v(t)$,

$$\frac{d(u/v)}{dt} = \frac{1}{v}\frac{du}{dt} - \frac{u}{v^2}\frac{dv}{dt},$$

is obtained from

$$\frac{du}{dt} = \sum_i Q_i r_i^2 e^{r_i t} \qquad \text{and} \qquad \frac{dv}{dt} = \sum_i Q_i r_i e^{r_i t}$$

as

$$\frac{d\bar{r}(t)}{dt} = \frac{\sum_i Q_i r_i^2 e^{r_i t}}{\sum_i Q_i e^{r_i t}} - \left(\frac{\sum_i Q_i r_i e^{r_i t}}{\sum_i Q_i e^{r_i t}}\right)^2. \tag{1.5.2}$$

Equation (1.5.2) informs us that the increase in the mean rate of increase at time t is equal to the difference between the mean square and the square of the mean of the rates weighted by the number of people at that time.

Variance is ordinarily defined as the mean square deviation from the mean: if a variable x takes values $x_1, x_2, \ldots, x_K$, then

$$\bar{x} = \sum x_i/K \qquad \text{and} \qquad \sigma^2 = \sum (x_i - \bar{x})^2/K.$$

But

$$\begin{aligned} \frac{\sum (x_i - \bar{x})^2}{K} &= \frac{\sum[(x_i - \bar{x})(x_i - \bar{x})]}{K} \\ &= \frac{\sum[x_i(x_i - \bar{x})] - \sum[\bar{x}(x_i - \bar{x})]}{K}, \end{aligned}$$

and the second summation vanishes to leave

$$\sigma^2 = \frac{\sum x_i^2}{K} - \bar{x}^2,$$

because $\sum(x_i - \bar{x}) = 0$ and $\sum x_i \bar{x}/K = \bar{x}^2$. The argument applies equally well if the x_i are weighted, and in particular it applies to our r_i weighted at any moment by their respective populations.

Hence (1.5.2) is the same as the variance among subpopulation rates of increase:

$$\frac{d\bar{r}(t)}{dt} = \sigma^2(t), \tag{1.5.3}$$

where $\sigma^2(t)$ is the variance among the r_i, each weighted according to its current subpopulation $Q_i e^{r_i t}$.

The nonnegative derivative of $\bar{r}(t)$ in (1.5.3) proves that $\bar{r}(t)$ is always increasing, unless the r_i are all the same, in which case the variance is zero, the derivative of $\bar{r}(t)$ is zero, and $\bar{r}(t)$ is constant.[†] But, though it keeps increasing when the rates of increase of the component subpopulations are not all the same, $\bar{r}(t)$ is bounded above; being an average of the several r_i, it can never be larger than the largest r_i. It will approach as close as we please, however, to the largest r_i; to prove this, note that with a finite number of subpopulations the one that is increasing fastest will come to have as high a ratio as we please to the one increasing second fastest, and indeed to the sum of all the other subpopulations. This example of instability contrasts with the stability described in Section 1.10, where the operation of fixed rates in homogeneous populations leads to fixed ratios among ages, regions, and other subgroups of the population.

The foregoing argument is more than adequate to prove that the sum of the several projections will be greater than the projection of the sum at the average rate of increase existing at the outset. An alternative proof requiring no calculus consists in reducing the proposition to the known fact that the arithmetic mean of distinct positive quantities, say $a > 0$ and $b > 0$, $a \neq b$, is greater than their geometric mean; in symbols, $(a+b)/2 > \sqrt{ab}$. The separate projection of two subpopulations amounts to $Q_1 e^{r_1 t} + Q_2 e^{r_2 t}$; the combined projection is

$$(Q_1 + Q_2) \exp\left[\left(\frac{Q_1 r_1 + Q_2 r_2}{Q_1 + Q_2}\right) t\right].$$

The sum of the separate projections is greater if

$$\frac{Q_1 e^{r_1 t} + Q_2 e^{r_2 t}}{Q_1 + Q_2} > \exp\left[\left(\frac{Q_1 r_1 + Q_2 r_2}{Q_1 + Q_2}\right) t\right].$$

But the first is the arithmetic mean of $e^{r_1 t}$ and $e^{r_2 t}$, weighted with Q_1 and Q_2, respectively, and the second is the corresponding geometric mean.

The general proof that an arithmetic mean is greater than a geometric mean can be developed in several ways, and Beckenbach and Bellman (1961) devote much of Chapter 4 of their book to it. That the inequality must hold for $Q_1 = Q_2 = 1$ we can see at the level of high school algebra. Plainly

$$(e^{r_1 t/2} - e^{r_2 t/2})^2 > 0, \quad r_1 \neq r_2,$$

and expanding the square gives

$$\frac{e^{r_1 t} + e^{r_2 t}}{2} > \exp\left[\left(\frac{r_1 + r_2}{2}\right) t\right].$$

[†] This is one way of deriving a version of Fisher's (1930) Fundamental Theorem of Natural Selection.

The argument can be extended to n subpopulations by induction; if Q_1 of the n are increasing at r_i, and so on, the proof then applies to integral weights Q_i; from there it can be carried to real weights Q_i. To present this in detail would carry us too far from demography, which is concerned only with the result that separate projection of each of the various elements of a heterogeneous population gives a total greater than is obtained by projection of the whole population at its average rate.

1.5.1 An Arithmetic Example for Two Subpopulations

To use round numbers for an example of how (1.5.3) operates, we take the United States population to be 200 million in 1970, increasing at 0.75 percent, and that of Mexico to be 50 million, increasing at 3.5 percent. The average rate is $[(200\times 0.75)+(50\times 3.5)]/250 = 1.3$ percent, and 250 million increasing at 1.3 percent for 50 years equals 478,885,000.

Let us now work out the sum of the two trajectories, using for each of the United States and Mexico, considered subpopulations, its own exponential (Table 1.5.1). After 50 years at these rates the populations would be 200,000,000 × $e^{50\times 0.0075}$ = 290,998,000 for the United States and 50,000,000 × $e^{50\times 0.035}$ = 287,730,000 for Mexico. The total is 578,728,000, or 100 million more than we had in the single calculation. The combined rate of increase now would be over 2.1 percent and rising, as column 6 of Table 1.5.1 shows. By 2020 the increase in the increase, $d\bar{r}(t)/dt$, would be $(0.021267 - 0.021076) = 0.000191$, as given in column 7.

According to (1.5.3), this ought to be the same as the weighted variance of the rates in 2020. The mean rate is $[(291)(0.0075)+(288)(0.035)]/(291+288) = 0.02118$, the weights being the entries in columns 1 and 2 for 2020. The variance is $[291(0.0075 - 0.02118)^2 + 288(0.035 - 0.02118)^2]/579 = 0.000189$, identical except for rounding with the entry in column 7 previously calculated.

In summary, we have calculated the rate of increase of the combined population by differencing successive years, showing the rate to increase from 0.021076 in 2019–20 to 0.021267 in 2020–21, an annual increase of 0.000191 in the rate of increase. The separately calculated weighted variance between the two component rates in 2020 is 0.00189, or virtually complete agreement.

Table 1.1. Calculation of rate of increase in combined population of the United States and Mexico if each continues at its own fixed rate (thousands of persons)

Year	United States increasing at $100r = 0.75$ (1)	Mexico increasing at $100r = 3.5$ (2)	Total $(3) = (1) + (2)$	Annual increase (4)	Mean population [average of successive years in (3)] (5)	Rate of increase $(6) = (4)/(5)$	Increase in increase [difference from (6)] (7)
1970	200,000	50,000	250,000				
2019	288,824	277,834	566,658				
				12,070	572,693	0.021076	
2020	290,998	287,730	578,728				0.000191
				12,440	584,948	0.021267	
2021	293,189	297,979	591,168				0.000188
				12,821	597,578	0.021455	
2022	295,396	308,593	603,989				

1.6 Rate of Increase Changing over Time

We started with a homogeneous population having a fixed rate of increase r and went on to a heterogeneous population composed of subpopulations each having a fixed rate of increase. The present section reverts to homogeneity, in which there is only one r at any given moment, but now the increase is time dependent, and to remind ourselves of this we call it $r(t)$. Our purpose is to determine the population after T years resulting from the variable increase $r(t)$.

If we divide time into short intervals dt, and for the first short interval suppose $r(t)$ to be fixed at r_0, and if at the start the population numbers N_0, after time dt it will number $N_0 e^{r_0\,dt}$, for we saw in Section 1.2 that the ratio of increase over a finite time at rate r compounded momently is equal to the exponential of the rate times the time. This can be applied to each of the short intervals dt into which, for this purpose, we divide the scale of time. Let the rate of increase be approximated by $r_0, r_1, r_2, \ldots,$ in those successive time intervals, each of length dt. Then we have an exponential for the ratio of increase over each interval, and the population at time T will be the product of these exponentials:

$$\begin{aligned} N(T) &= N_0 e^{r_0\,dt} e^{r_1\,dt} e^{r_2\,dt} \cdots \\ &= N_0 e^{r_0\,dt + r_1\,dt + r_2\,dt + \cdots}, \end{aligned}$$

so that in the limit, as dt tends to zero, the exponent tends to the integral of $r(t)$:

$$N(T) = N_0 \exp\left[\int_0^T r(t)\,dt\right]. \tag{1.6.1}$$

The above derivation of (1.6.1) proves *ab initio* a proposition of the integral calculus. Let us avoid this by starting with the definition of the rate of increase $r(t)$ as given in (1.5.1):

$$r(t) = \frac{1}{N(t)} \cdot \frac{dN(t)}{dt}. \tag{1.6.2}$$

The demographic definition can be treated as a differential equation. Its solution is obtained by separating the variables and integrating both sides of

$$r(t)\,dt = \frac{dN(t)}{N(t)}.$$

This gives

$$\int_0^T r(t)\,dt = \log N(t)\Big|_0^T,$$

or, on taking exponentials,

$$N(T) = N_0 \exp\left[\int_0^T r(t)\,dt\right],$$

as before.

The result can be checked by calculating $r(t)$ from the trajectory of population $N(T)$. Taking logarithms of both sides of (1.6.1) yields

$$\log N(T) = \log N_0 + \int_0^T r(t)\,dt,$$

and then, differentiating with respect to T, we have

$$\frac{1}{N(T)} \cdot \frac{dN(T)}{dT} = r(T), \tag{1.6.3}$$

which brings us back to the definition of $r(t)$ in (1.5.1) and (1.6.2).

A convenient way of writing (1.6.1) is in terms of $\bar{r}$, the arithmetic mean rate over the interval from zero to T:

$$N(T) = N_0 e^{\bar{r}T},$$

where

$$\bar{r} = \frac{1}{T}\int_0^T r(t)\,dt.$$

In words, the numerical effect on the population total of a varying rate of growth is the same as though the arithmetic average rate applied at each moment over the time in question.

Special Cases of Changing Rates. We may try various special functions for $r(t)$. If $r(t)$ is a constant, say r, then (1.6.1) reduces to $N(T) = N_0 e^{rT}$. If $r(t)$ is equal to k/t and we start at time 1, then

$$\begin{aligned} N(T) &= N_1 \exp\left(\int_1^T \frac{k}{t}\,dt\right) = N_1 \exp(k \log T) \\ &= N_1 T^k. \end{aligned}$$

With $k = 1$ this declining rate of increase would give us a linearly rising population.

As a numerical application of (1.6.1) suppose that the rate of increase starts at ρ in 1970 and declines to $\rho/2$ during the 30 years from 1970 to 2000, and that the decline is in a straight line. Then in any intermediate year t we will have

$$r(t) = \frac{\rho}{60}(2030 - t), \quad 1970 \leqslant t \leqslant 2000. \tag{1.6.4}$$

The proof that this $r(t)$ is the one specified is (a) it is linear in t; (b) for $t = 1970$ it equals ρ; (c) for $t = 2000$ it equals $\rho/2$. Entering (1.6.4) in

(1.6.1) gives for the population in year n, where $1970 \leqslant n \leqslant 2000$,

$$N_n = N_{1970} \exp \left[\frac{\rho}{60} \int_{1970}^{n} (2030 - t)\, dt \right],$$

and for $n = 2000$ this is equal to

$$N_{2000} = N_{1970} e^{22.5\rho}.$$

As a further example, suppose that a population in successive years increases at the rates 2, 3, 2.5, 2.7 and 2.3 percent. The arithmetic average of these rates is

$$\frac{2.0 + 3.0 + 2.5 + 2.7 + 2.3}{5} = 2.5 \text{ percent.}$$

Hence the population at the end of 5 years is

$$N_5 = N_0 e^{(5)(0.025)} = N_0 e^{0.125} = 1.133 N_0.$$

This is the same outcome as is obtained by calculating the population year by year with the given rates (all assumed to be compounded continuously):

$$N_1 = N_0 e^{0.02},\ N_2 = N_1 e^{0.03},\ \ldots,$$

and substituting successively,

$$N_5 = N_0 e^{0.02} e^{0.03} e^{0.025} e^{0.027} e^{0.023} = N_0 e^{0.125}.$$

Equation 1.6.1 is worth this extended study because of its important applications. In particular, when $r(t)$ is interpreted as $-\mu(a)$, $\mu(a)$ being mortality at age a, the result carries over to cohorts; a cohort is defined as a number of individuals born at a particular time and followed through life. Such a cohort is a peculiar population, in that it never receives new members after the initial moment, and its initial births die off during the 100 or so years of the cohort's duration. Equation 1.6.1 tells us that the survivors to age x, $l(x)$, of a cohort numbering l_0 at birth, are equal to

$$l(x) = l_0 \exp \left[- \int_0^x \mu(a)\, da \right]. \tag{1.6.5}$$

This expression is useful in discussion of the life table. The decrement $\mu(a)$ may represent death at age a, death from a particular cause, failing the ath grade at school, or break-up of a marriage in its ath year. Whatever the nature of the decrement, (1.6.5) translates the several hurdles into a probability of surviving the course to the xth hurdle.

1.7 Logistic Increase and Explosion

Verhulst (1838) and Pearl and Reed (1920) modified the exponential law of increase by supposing a fixed ceiling to population, set by nature or by the combined limits of nature and human techniques. Fixed growth, represented by the equation $dN(t)/dt = rN(t)$, of which the solution is $N(t) = N_0 e^{rt}$, can be modified by writing a further factor, say $1 - [N(t)/a]$, on the right-hand side to define the logistic function

$$\frac{dN(t)}{dt} = rN(t)\left[1 - \frac{N(t)}{a}\right].$$

The right-hand side equals zero when $N(t) = a$, so that must be where growth stops on the logistic curve. When the factors involving $N(t)$ are collected and then decomposed into partial fractions, we have

$$\left[\frac{1}{N(t)} + \frac{1}{a - N(t)}\right] dN(t) = r\,dt,$$

which can be readily integrated as $\log\{N(t)/[a - N(t)]\} = rt + c$, where c is a constant. Taking exponentials and then solving for $N(t)$, and changing the constant to $t_0 = -c/r$ in order to locate the midpoint of the curve at t_0, we obtain

$$N(t) = \frac{a}{1 + e^{-r(t-t_0)}}. \tag{1.7.1}$$

Pearl and others thought that fitting such a curve to a population time series would provide realistic short-term forecasts as well as estimates of the ultimate stationary population a. They were generally encouraged by predicting the 1930 U.S. Census count with an error that was probably less than the error of the census itself. But the 1940 Census was a disappointment—it fell far below the curve, and the logistic was dropped amid a barrage of criticism no less intense than the earlier enthusiasm. The fall in the birth rate over most of the last 20 years, however, makes the logistic look much better, and this, along with visible difficulties in the environment, has brought the logistic back into fashion in some circles. The fit to the United States population from 1800 to 1960 shows $a = 256.41$ million persons, a ceiling not sharply contradicted by current tendencies.

As Pearl expected, the logistic is hardly invariant with respect to the period over which it is fitted. He saw population as moving toward a ceiling at any given moment, but from time to time technical advance would create a new and higher ceiling so that progress would take the form of breaking into the higher logistic before the course of the earlier one was completed.

No detailed fitting is required merely to find the ceiling a. [If the population is N_1, N_2, and N_3 at times t_1, t_2, and $2t_2 - t_1$, respectively (i.e., at times that are equidistant), prove that the asymptote will be

$$a = \frac{1/N_1 + 1/N_3 - 2/N_2}{1/N_1N_3 - 1/N_2^2}, \tag{1.7.2}$$

and using as data the United States resident population for

1870	40 million
1920	106 million
1970	203 million

verify that the ultimate population will be 324 million.]

The logistic to the censuses from 1800 to 1910 inclusive shows an upper asymptote of 197 millions; if one accepts the theory behind the logistic, one would say that the conditions of life and technology had changed during the course of the twentieth century in such a way as to raise the population ceiling of the United States by 127 millions. Although this could be true, one would need considerable confidence in the appropriateness of the logistic to accept such a conclusion without more evidence.

Even more important than the selection of the time interval over which a fit is made is the nature of the curve chosen. Consider the United States population from 1870 to 1970, and fit to it the hyperbola (von Foerster et al. 1960)

$$N(t) = \frac{\alpha}{t_e - t}.$$

The hyperbola contains two constants, of which t_e is the time of population explosion, when $N(t_e) = \infty$. The time t_e is easily calculated from observations at two dates, t_1 and t_2, where the population is known to be N_1 at time t_1 and N_2 at time t_2; the reader may show that

$$t_e = \frac{N_2 t_2 - N_1 t_1}{N_2 - p_1}. \tag{1.7.3}$$

The 1870 resident population of the United States was 40 million and the 1970 population 203 million, and from (1.7.3) t_e works out to 1995.

Thus, based on the United States population from 1870 to 1970, the logistic shows that an upper asymptote of 324 millions would be approached (about the twenty-first century), whereas the hyperbola shows an explosion to infinity by 1995. Such experimenting suggests the hazards and perplexities of forecasting by the fitting of curves. [Show that no one need fear the explosion in hyperbolic form, for it would require toward the end an infinite birth rate.]

Can one discriminate among curves on the basis of their fits to past data? Winsor (1932) showed that the logistic is a better fit than the cumulated normal,

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx,$$

where μ and σ are constants to be determined by data from censuses. Either the logistic or the cumulated normal does better than the arc tangent curve, which has some resemblance to them. But in general past data are not very helpful in selecting from the considerable number of s-shaped curves that can be devised.

As a means for population forecasting, the logistic has become something of a museum piece. Any such overall approach, disregarding births, deaths, migration, and age distribution, is useful only in circumstances where resources decisively determine population without regard to differences among individuals. When, on the other hand, births, migration, and even deaths are socially determined, we are better off to attempt their separate forecasting, however uncertain this has been shown to be.

1.8 The Stalled Demographic Transition

In a famous article Notestein (1945) wrote about "the stage of transitional growth . . . in which the decline of both fertility and mortality is well established but in which the decline of mortality precedes that of fertility and produces rapid growth." His demographic transition refers to the uniformity of change from high to low birth and death rates among the countries of Europe and those overseas that had developed industrially. They showed first a decline in death rates, starting at the beginning of the nineteenth century or earlier, followed after a longer or shorter interval by a decline in birth rates. In France the fall in births was nearly simultaneous with that in deaths, whereas in England births did not begin to decline until about 1870, but all countries resembled one another to some degree (Flieger 1967). Our first question concerns the difference to the ultimate population of a given delay in the fall of the birth curve, a question of concern to the countries of Asia and Latin America whose deaths have now fallen but whose births remain high.

Suppose the deaths of a population go through a descending curve $d(t)$ and its births through $b(t)$, as in Figure 1.1. The initial and final conditions are both of zero increase; that is, the curves coincide at beginning and end. We seek the ratio of increase in the population between its initial and final stationary conditions.

Whatever the shape of the two descending curves of Figure 1.1, if they begin and end together the exponential of the area between them is the total increase over the time in question. For $\int_0^T [b(t) - d(t)]\, dt = \int_0^T r(t)\, dt = A$, say, according to (1.6.1) the ratio of increase in the population must be $\exp[\int_0^T r(t)\, dt]$, or simply e^A. This applies for any pair of monotonically descending curves that start at the same level and end at the same level.

In the special case where the birth and death curves of Figure 1.1 have the same shape as well, with $b(t)$ lagging L behind $d(t)$, and both dropping K over the transition, the area A equals KL, that is, the common difference between initial and final height, multiplied by the time by which the birth curve lags behind the death curve. For by dividing the interval between them into horizontal strips, equal in length to the lag L, it is plain that the

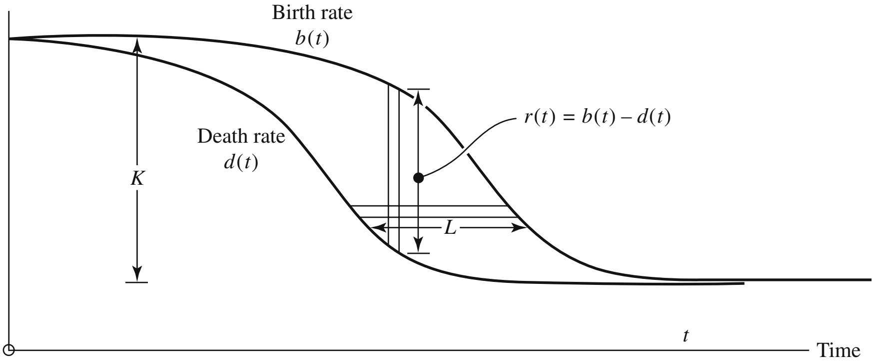


Figure 1.1. Stylized form of transition from high to low birth and death rates.

Table 1.2. Ratio of increase during the course of demographic transition as a function of the lag in fall of birth rate after fall of death rate

Lag (years) L	e^{KL} $(K = 0.02)$	e^{KL} $(K = 0.03)$
15	1.35	1.57
30	1.82	2.46
45	2.46	3.86
60	3.32	6.05
75	4.48	9.49

same strips can be arranged as a rectangle of length L and height K. The ratio of increase in the population of a country, $e^A = e^{KL}$, is a constant e^K taken to the power of the lag in the fall of birth rates.

If K, the common decline of births and deaths, is 0.02, and births follow deaths downward but with a lag L of 30 years, the population will increase in the ratio $e^{KL} = e^{(0.02)(30)} = 1.82$ before constancy is reestablished. If the lag is 60 years, the ratio will be the *square* of this, or 3.32; if 90 years, the cube, or 6.05. Note that this takes no account of a momentum effect due to age, which, as we will see in Section 8.6, can by itself add 60 percent or more to the ultimate population. The need for haste in lowering the birth rate in less developed countries is illustrated in Table 1.2, showing the effect of lag on the ultimate population, given $K = 0.02$ and $K = 0.03$.

This section has covered the general case of curves $b(t)$ and $d(t)$, similar to each nother but with $b(t)$ lagging behind $d(t)$. We now proceed from longitudinal to cross-sectional observations, and consider differential fertility insofar as it is a phenomenon of the demographic transition.

1.9 Differential Fertility Due to the Demographic Transition

Books on population treat the demographic transition in one chapter and in a quite different chapter deal with the differentials of fertility between social classes, educational groups, and religious denominations. The transition is thought of as applying to whole countries, and differential fertility as applying to groups within a country, in which, *grosso modo*, the better off, better educated, and urban have the fewest children. There may well be uncertainty as to how far these classical differentials are permanent, applying to all societies at all times, and how far they occur in a particular historical conjuncture, that in which all birth rates are falling but with different timing. Some observations show a positive relation—the richer the group, the higher its birth rate—both before the transition and after it has been passed and birth control techniques made effectively available to all strata. Much of the economic theory of fertility, summarized by Leibenstein (1974), Becker (1960), T. W. Schultz (1974), and T. P. Schultz (1973), has a bearing on this issue. The present section will examine a simple aspect of the problem: to what degree the different times of entry into the transition can account for the different levels of fertility among social classes at any one moment.

We will approximate the birth rate $b(t)$ by a straight line going from upper left to lower right in the range of interest; the slope or derivative $db\,(t)/dt$ is taken as negative. We suppose also that the slope is the same for all social groups, and that these are distinguished from one another only by their degree of horizontal displacement.

The sloping line on the right in Figure 1.2 is displaced from the population mean by Δt, say. It represents a social group whose fall in fertility takes place later by Δt than the average of all groups in the country in question. Suppose that because of its lag this group has fertility Δb higher than the average at the time when the country as a whole is passing through the midpoint of its drop from high to low rates. Then the derivative $db(t)/dt$ common to all the sloping lines serves to relate for the given subgroup of the population the departure of the birth rate from the mean at a given moment and the lag in time:

$$\Delta b = \frac{db(t)}{dt}\Delta t. \tag{1.9.1}$$

Now suppose many sloping lines, representing the several groups in the country, and square and average over these lines on both sides of (1.9.1) to find

$$\sigma_b^2 = \left[\frac{db(t)}{dt}\right]^2 \sigma_t^2, \tag{1.9.2}$$

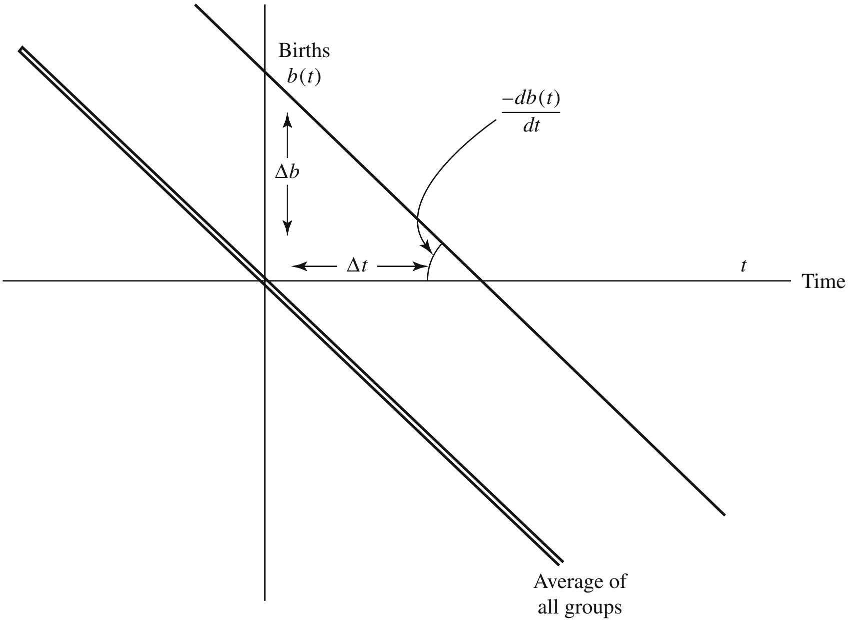


Figure 1.2. Simplified explanation of differential fertility in terms of demographic transition.

where σ_b^2 is the variance of birth rates at the time when the population average goes through the halfway point of decline, and σ_t^2 is the variance of the times at which the several groups go through this point. Take square roots of (1.9.2) to find

$$\sigma_b = \frac{db(t)}{dt}\sigma_t. \tag{1.9.3}$$

In this stylized version each statistically identifiable subpopulation has its own demographic transition and falls according to its own straight line, but with all the straight lines having the same slope. The result (1.9.3) relates differential fertility σ_b to variation in time of undergoing the transition σ_t.

1.10 Matrices in Demography

1.10.1 A Two-Subgroup Model

Within a population in which there exist statistically recognizable subpopulations—regions of a country, social classes, educational levels—we consider now not changes within, but transfers among, such groupings. Given the rate of growth in each subpopulation and the rates of transfer among subpopulations, it is possible to describe a trajectory by a set of differential or difference equations. For a simple special case, suppose two subpopulations of sizes $n_1(t)$ and $n_2(t)$ at time t. If growth of the subpopulations is accompanied by migration in both directions, change in the system can be described in terms of constants a_{ij}, $i = 1, 2, j = 1, 2$:

$$\begin{aligned} n_1(t+1) &= a_{11}n_1(t) + a_{12}n_2(t) \\ n_2(t+1) &= a_{21}n_1(t) + a_{22}n_2(t). \end{aligned} \tag{1.10.1}$$

Equations 1.10.1 are identical to the matrix equation

$$\begin{pmatrix} n_1 \\ n_2 \end{pmatrix}(t+1) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} n_1 \\ n_2 \end{pmatrix}(t), \tag{1.10.2}$$

which can be written compactly as

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t), \tag{1.10.3}$$

where $\mathbf{n}(t)$ is the population vector at time t, and $\mathbf{A}$ is the matrix of growth and transfer rates, referred to indifferently as a projection matrix. If the rates a_{11}, a_{21}, a_{12}, and a_{22} are fixed over time, (1.10.3) recurrently determines the population at any arbitrary time subsequent to t. The population at time $t+k$ equals that at time t successively operated on k times by $\mathbf{A}$:

$$\mathbf{n}(t+k) = \mathbf{A}(\ldots(\mathbf{A}(\mathbf{A}\mathbf{n}(t))\ldots) = \mathbf{A}^k\mathbf{n}(t). \tag{1.10.4}$$

If $a_{12} = a_{21} = 0$, and $a_{11} \neq a_{22}$, the argument of Section 1.5 applies; the two subpopulations never come into a finite, nonzero ratio to each another, and the subgroup with the higher rate keeps growing relative to the one with the lower rate. But if a_{12} and a_{21} are positive, then, no matter how different a_{11} and a_{22} may be, the two subpopulations will ultimately tend to increase at the same rate. This stability of the ratio of one population to the other will of course occur more quickly if a_{12} and a_{21} are large in relation to the difference between a_{11} and a_{22}.

The aspect of stability referred to above is that in which the ratio of the sizes of the subpopulations ultimately ceases to depend on time. When this occurs, it follows that each of the subgroups will increase at a rate not depending on time, that is to say, in geometric progression. For when population n_2 is c times as large as population n_1, for all times, then, from the first member of (1.10.1)

$$n_1(t+1) = a_{11}n_1(t) + a_{12}n_2(t) = (a_{11} + ca_{12})\, n_1(t),$$

which proves that $n_1(t)$ is a constant multiple of $n_2(t)$. This applies to any number of subgroups, and shows for the linear model of (1.10.1) that, if the ratio of the sizes of the subgroups to one another is constant, each is increasing geometrically. It can then be shown that all the groups are increasing at the same rate. Conversely, if the subgroups are increasing geometrically and at the same rate, they are in fixed ratios to one another.

When the process of which (1.10.1) is an example attains stability, not only are all its subpopulations increasing at the same rate and in fixed ratios to one another, but also those ratios are not in any way influenced by the starting ratios. It could be that $n_1(0)$ is 1 million times as large as $n_2(0)$, or that $n_2(0)$ is 1 million times as large as $n_1(0)$; the two cases will have the same ultimate ratio of $n_1(t)$ to $n_2(t)$. This property of the process, forgetting its past, is a third aspect of stability (see Section 7.2). Chapter 7 will show how to determine which projection matrices will result in stability, using life cycle graphs.

There is no moment when stability is suddenly attained. Stability is a limiting property by which a time can be found when the several subpopulations increase at rates that are arbitrarily close to one another.

Chapter 5 considers in much detail one case of stability—a population developing under fixed mortality and rate of increase. To prepare for it we need to transform raw mortality data into a life table, and this is the subject of Chapter 2.

2

The Life Table

The main part of this book starts where demography itself started, with the life table. The life table is couched in terms of probabilities for individuals, but for populations it is a deterministic model of mortality and survivorship. That it presents expected values and disregards random variation is contrary to the way nature works and, in particular, oversimplifies demographic mechanisms, yet a rich variety of useful results is based on it. A method is valuable in direct proportion to the substantive conclusions to which it leads, and in inverse proportion to its complexity.

The life table is mathematically simple and has a large substantive payoff. First, it answers questions concerning individuals: what is the probability that a man aged 30 will survive until he retires at 65, or that he will outlive his wife, aged 27? But it also answers questions concerning cohorts, groups of individuals born at the same time: what fraction of the births of this year will still be alive in the year 2000, or how many of them will live to the age of retirement? Third, it answers population questions: if births were constant from year to year in a closed population of constant mortality, what fraction of the population would be 65 and over?

2.1 Definition of Life Table Functions

The probability of surviving from birth to age x is designated as $l(x)$ for a continuous function of x and as l_x for discrete x. Life tables usually present the probability multiplied by 100,000, which is to say, on a *radix*, l_0, equal to 100,000. If l_x is a probability, then, strictly speaking, what life tables show is 100,000l_x, but it would be pedantic to repeat the 100,000 each time one refers to the l_x column. When l_x is interpreted as surviving members of the cohort, the radix is arbitrary; and setting $l_0 = 100{,}000$ enables one more easily to think of the column as numbers of persons reaching the given age.

The difference in number of survivors for successive ages, $l_x - l_{x+1}$, is designated as d_x; and, more generally, the difference for ages n years apart, $l_x - l_{x+n}$, is ${}_nd_x$. This divided by l_x is the probability of dying during the next n years for a person who has reached age x:

$${}_nq_x = \frac{l_x - l_{x+n}}{l_x} = \frac{{}_nd_x}{l_x}.$$

The total number of years lived during the next n years by those who have attained age x is

$${}_nL_x = \int_x^{x+n} l(a)\,da, \tag{2.1.1}$$

which is also the number of persons aged exactly x to $x+n$ (or x to $x+n-1$ at last birthday) in the *stationary* population. Setting n equal to infinity or (indifferently) to $\omega - x$, where ω is the highest age to which anyone lives, gives

$${}_\infty L_x = T_x = \int_x^{\omega} l(a)\,da.$$

The quantity T_x is the total number of years remaining to the cohort, when it reaches age x, until the last member dies at age ω. Dividing by l_x gives the average share in this total per person reaching age x:

$$\overset{o}{e}_x = \frac{T_x}{l_x} = \frac{\displaystyle\int_x^{\omega} l(a)\,da}{l_x}.$$

In terms of probabilities $\overset{o}{e}_x$ is the mean of the distribution of years to death for persons of age x and is called the *expectation of life.*

The age-specific death rate in the life table population is ${}_nm_x = {}_nd_x/{}_nL_x$. This may be compared with the probability ${}_nq_x = {}_nd_x/l_x$, and one can think of ${}_nq_x$ as something less than n times ${}_nm_x$. It is distinct from ${}_nM_x$, the observed death rate in a real population; ${}_nM_x < {}_nm_x$ for increasing populations at ages beyond about 10.

Theoretical statements on mortality are often expressed most simply in terms of the age-specific death rate in a narrow age interval dx, designated as $\mu(x)$:

$$\mu(x) = \operatorname*{limit}_{n\to 0} \frac{l_x - l_{x+n}}{\displaystyle\int_x^{x+n} l(a)\,da} = \frac{-1}{l(x)}\,\frac{dl(x)}{dx}.$$

[Show that the solution of this gives (1.6.5).] The age-specific death rate could also be written as

$$\mu(x) = \operatorname*{limit}_{n\to 0} {}_n m_x = \operatorname*{limit}_{n\to 0} \frac{{}_n d_x}{{}_n L_x},$$

if the mixing of continuous and discrete notation can be excused.

Mortality the Same for All Ages. A mathematical form that has often been used for short intervals of age, is one in which the force of mortality is constant. If $\mu(x) = \mu$, solving the differential equation that defines μ,

$$\mu = \frac{-1}{l(x)} \cdot \frac{dl(x)}{dx},$$

gives

$$l(x) = e^{-\mu x}.$$

The probability of living at least an additional n years after one has attained age x is

$$\frac{l_{x+n}}{l_x} = \frac{e^{-\mu(x+n)}}{e^{-\mu x}} = e^{-\mu n}.$$

The expectation of life is

$$\begin{aligned}\overset{o}{e}_x &= \frac{1}{l(x)} \int_0^\infty l(x+t)\,dt \\ &= \frac{1}{e^{-\mu x}} \int_0^\infty e^{-\mu(x+t)}\,dt = \frac{1}{\mu}.\end{aligned}$$

Both the probability of living an additional n years beyond age x and the expectation of life at age x are constants independent of x. The prospect ahead of any living person is the same, no matter how old he is, on this peculiar table. All the columns of the life table are in general derivable from any one column, and if any column is age-independent all the others referring to a person's prospects must be also.

Application of these definitions to other mathematically specified forms of $\mu(x)$ requires more difficult integration to find $l(x)$ and $\overset{o}{e}_x$. The function $\mu(a) = \mu_0/(\omega - a)$ is applied in Section 4.3. Here we proceed to methods for constructing a life table from empirical data.

2.2 Life Tables Based on Data

We read that for the United States during the year 1967, 122,672 men aged 65 to 69 years of age at last birthday died, and that the number of men of these ages alive on July 1, 1967, was 2,958,000. These may be called observations. We can divide the first figure by the second and find that the age-specific death rate for males 65 to 69, denoted as ${}_5M_{65}$. was 0.04147, again straightforward enough to be called an observation. How do these observations tell us the probability, $1 - (l_{70}/l_{65})$, that a man chosen at random from those aged exactly 65 will die before reaching age 70? This is the same as asking what fraction of a cohort of men aged 65 would be expected to die before reaching age 70, and its complement, l_{70}/l_{65}, the fraction of those aged 65 who survive to age 70. From time period data we want to know something about a cohort, men moving from age 65 to 70. This cohort is necessarily hypothetical, for with only one period of observation the best we can manage is to suppose mortality to be unchanging through time. No real cohort is likely to exhibit the regime that will be inferred from 1967 data.

The observations of a given period, the calendar year 1967, are, moreover, ambiguous because we know nothing about the distribution of exposure within each 5-year age interval, affected as it is by all the accidents of the birth curve and of migration.

2.2.1 Assuming Constant Probability of Dying with the Age Interval

If, in addition to supposing that the probability of dying is invariant with respect to time, we suppose it to be invariant with respect to age within the 5-year group, the probability l_x can be inferred from the rate ${}_5M_x$ in a straightforward manner. We saw in Section 1.6 that a population initially of size N_0 and increasing at constant rate r reaches a total N_0e^{rt} after t years. If it is decreasing at constant rate μ, as will be the case if it is closed and subject to a constant death rate μ, and it numbers l_x at the start, then at the end of 5 years it will be $l_{x+5} = l_xe^{-5\mu}$, as in the preceding section. Identifying ${}_5M_x$ with μ gives

$$\frac{l_{x+5}}{l_x} = e^{-5\,{}_5M_x} \tag{2.2.1}$$

as a first approximation to the desired ratio l_{x+5}/l_x. For our data on the United States in 1967, (2.2.1) is

$$\frac{l_{70}}{l_{65}} = e^{-5\,{}_5M_{65}} = e^{-5(0.04147)} = 0.81274.$$

This would be exact if either (*a*) the death rate were constant through the 5-year age interval, or (*b*) the population exposed to risk were constant through the 5-year age interval. Neither of these, however, applies in practice; in general, beyond age 10 the death rate is increasing through the interval and the population is diminishing.

2.2.2 The Basic Equation and a Conventional Solution

Suppose that the age distribution of the exposed population is given by $p(a)$ within the age interval x to $x+5$, and that at exact age a the age-specific death rate is $\mu(a)$. Then the observed rate ${}_5M_x$ can be identified with a ratio of integrals:

$$
{}_5M_x = \frac{\displaystyle\int_x^{x+5} p(a)\mu(a)\,da}{\displaystyle\int_x^{x+5} p(a)\,da}. \tag{2.2.2}
$$

Making a life table is a matter of inferring l_{x+5}/l_x from (2.2.2). This is equivalent to inferring the unweighted $\int_x^{x+5}\mu(a)\,da$ from the weighted average of the $\mu(a)$ in (2.2.2), for from (1.6.5) we know that $\int_x^{x+5}\mu(a)\,da = -\log(l_{x+5}/l_x)$, so the unweighted average of the $\mu(a)$ gives the $l(x)$ column (Weck 1947).

Basic equation 2.2.2 reminds us that the interpretation of the observed ${}_5M_x$ depends on the (unknown) distribution of population within the age interval x to $x+n$. We must somehow extract $\int_x^{x+5}\mu(a)\,da$ from (2.2.2), a task that appears hopeless when nothing is known but ${}_5M_x$. Yet every life table based on empirical data in 5-year groups implicitly infers $\int_x^{x+5}\mu(a)\,da$ from (2.2.2); this is achieved by making assumptions that somehow restrict $p(a)$.

One common solution of (2.2.2) is to assume $p(a) = l(a)$, and also to suppose that $l(x)$ is a straight line. The integral under the straight line is

$$
\int_x^{x+5} l(a)\,da = \tfrac{5}{2}(l_x + l_{x+5}). \tag{2.2.3}
$$

Since $\int_x^{x+5} l(a)\mu(a)\,da = l_x - l_{x+5}$, (2.2.2) may be written as

$$
{}_5M_x = \frac{l_x - l_{x+5}}{\frac{5}{2}(l_x + l_{x+5})}, \tag{2.2.4}
$$

from which the value of l_{x+5}/l_x may be obtained by dividing numerator and denominator on the right by l_x, and then solving a linear equation for l_{x+5}/l_x to find

$$
\frac{l_{x+5}}{l_x} = \frac{1 - 5{}_5M_x/2}{1 + 5{}_5M_x/2}. \tag{2.2.5}
$$

For United States males aged 65 to 69 this is 0.81213, obtained from ${}_5M_{65} = 0.04147$. This is closer to the corrected value than the 0.81274 of (2.2.1). From (2.2.5) the probability of dying is

$$
{}_5q_x = 1 - \frac{l_{x+5}}{l_x} = \frac{5{}_5M_x}{1 + 5{}_5M_x/2}. \tag{2.2.6}
$$

2.2.3 A Precise Life Table Without Iteration or Graduation

To improve on (2.2.5) or the exponential (2.2.1), we go a step further in extracting from (2.2.2) the quantity $\int_x^{x+5} \mu(a)\,da$. We would expect the answer to emerge as $\int_x^{x+5} \mu(a)\,da = 5({}_5M_x + C)$, where it will turn out that C is a correction easily obtained on the assumption that both the population $p(a)$ and the death rate $\mu(a)$ change linearly within the interval.

We calculate the correction to ${}_nM_x$ for a general interval of n years, starting at age x. First we change variables by writing $a = x + (n/2) + t$; (2.2.2) with n in place of 5 becomes

$$ {}_nM_x = \frac{\displaystyle\int_{-n/2}^{+n/2} p[x+(n/2)+t]\mu[x+(n/2)+t]\,dt}{\displaystyle\int_x^{x+n} p(a)\,da}. \tag{2.2.7} $$

Expanding each of the fractions in the numerator by Taylor's series to the term linear in t gives

$$ \begin{aligned} {}_nM_x &= \frac{\displaystyle\int_{-n/2}^{+n/2} [p(x+n/2)+tp'(x+n/2)][\mu(x+n/2)+t\mu'(x+n/2)]\,dt}{\displaystyle\int_x^{x+n} p(a)\,da} \\ &= \frac{\left[tp(x+n/2)\mu(x+n/2) + \dfrac{t^3}{3}p'(x+n/2)\mu'(x+n/2)\right]_{-n/2}^{+n/2}}{\displaystyle\int_x^{x+n} p(a)\,da} \\ &= \frac{np(x+n/2)\mu(x+n/2) + (n^3/12)p'(x+n/2)\mu'(x+n/2)}{\displaystyle\int_x^{x+n} p(a)\,da}. \end{aligned} \tag{2.2.8} $$

But this can be translated into known quantities as follows:

1. If $p(a)$ is a straight line in the interval between x and $x + n$, then $np(x + n/2)$ is the same as $\int_x^{x+n} p(a)\,da$ and cancels with the denominator.
2. If $\mu(a)$ is a straight line between x and $x + n$, then the midvalue $\mu(x + n/2)$ is $1/n$ of the integral we seek, i.e., is $(1/n)\int_n^{x+n} \mu(a)\,da$.
3. The integral $\int_x^{x+n} p(a)\,da$ is the observed population in the age interval, ${}_nN_x$.

Thus (2.2.8) becomes

$$ {}_nM_x = \frac{1}{n}\int_x^{x+n} \mu(a)\,da + \frac{n^3}{12\,{}_nN_x} p'\left(x+\frac{n}{2}\right)\mu'\left(x+\frac{n}{2}\right), \qquad (2.2.9) $$

which after transposing provides the desired $\int_x^{x+n}\mu(a)\,da$ as

$$ \int_x^{x+n} \mu(a)\,da = n_nM_x - \frac{n^4}{12_nN_x} p'\left(x+\frac{n}{2}\right)\mu'\left(x+\frac{n}{2}\right). \qquad (2.2.10) $$

All that remains for the application is to express the first derivatives on the right of (2.2.10) in terms of known quantities. One might make the natural assumption that the slope within the interval x to $x+n$ is given by the difference between neighboring intervals:

$$ \begin{aligned} p'\left(x+\frac{n}{2}\right) &= -\frac{{}_nN_{x-n} - {}_nN_{x+n}}{2n^2} \\ \mu'\left(x+\frac{n}{2}\right) &= \frac{{}_nM_{x+n} - {}_nM_{x-n}}{2n}, \end{aligned} \qquad (2.2.11) $$

and so the calculable value of l_{x+n}/l_x is

$$ \begin{aligned} \frac{l_{x+n}}{l_x} &= \exp\left(-\int_x^{x+n}\mu(a)\,da\right) \\ &= \exp\left[-n\,{}_nM_x - \frac{n}{48_nN_x}({}_nN_{x-n} - {}_nN_{x+n})({}_nM_{x+n} - {}_nM_{x-n})\right] \\ &= \exp\left[-n({}_nM_x + C)\right]. \end{aligned} \qquad (2.2.12) $$

[We are grateful to James Frauenthal for his correction of the original approximation to $p'(x+n/2)$ in (2.2.11), which had n rather than n^2 in the denominator.]

The expression for l_{x+n}/l_x in (2.2.12) is the outcome of the search for a life table that would accord with the data in the sense of having the same underlying $\mu(x)$ as the observations, yet be calculable in one simple step. It is equivalent to using the simple exponential $l_{x+n}/l_x = e^{-n_nM_x}$, but first raising ${}_nM_x$ by the quantity $C = ({}_nN_{x-n} - {}_nN_{x+n})({}_nM_{x+n} - {}_nM_{x-n})/48_nN_x$, a product that is positive whenever the population is declining with age and the death rate is rising.

The accuracy of (2.2.12) has been impressive in the tests so far done. Applying it to Swedish males, 1965, at ages 20 to 65, for example, we found that l_{65}/l_{20} differed by 0.00001 from the value obtained by interpolating to fifths of a year separately for deaths and population, constructing the life table in fifths of a year, and then reassembling into 5-year age groups. It also differed by about 0.00001 from the more elaborate iterative life table (Keyfitz 1968, Chapter 1).

Table 2.1. Example of life table calculation without iteration or graduation, United States males, 1972

$$\frac{l_{x+5}}{l_x} = \exp[-5\,{}_5M_x - \frac{5}{48\,{}_5N_x}({}_5N_{x-5} - {}_5N_{x+5})({}_5M_{x+5} - {}_5M_{x-5})] = \exp[-5({}_5M_x + C)]$$

Age	$\frac{{}_5N_x}{1000}$	$1000\,{}_5M_x$	Correction to ${}_5M_x$; $C = \frac{({}_5N_{x-5} - {}_5N_{x+5})({}_5M_{x+5} - {}_5M_{x-5})}{48\,{}_5N_x}$	$\frac{l_{x+5}}{l_x} = e^{-5({}_5M_x + C)}$	${}_5q_x = 1 - \frac{l_{x+5}}{l_x}$
x	(1)	(2)	(3)	(4)	(5)
35	5458	3.017			
40	5720	4.623	−0.0000058	0.97718	0.02282
45	5814	7.483	0.0000025	0.96326	0.03674
50	5616	11.367	0.0000388	0.94457	0.05543
55	4828	18.092	0.0000990	0.91306	0.08694
60	4192	27.483	0.0001667	0.87088	0.12912
65	3294	39.958	0.0003802	0.81735	0.18265
70	2330	59.770			

In an age of computers ease of calculation is less important than it once was, but nonetheless Table 2.1 is introduced to show the extreme simplicity of the arithmetic.

The present method can be adapted to the ages at the beginning and end of life. However, these ages involve data problems as well as rapidly changing mortality rates. The reader is referred to Shryock and Siegel (1971, Chapter 15), Wolfenden (1954), Keyfitz (1968, Chapter 1), or other sources for ages under 10 and over 80.

A further point due to Kenneth Wachter and Thomas Greville is that the derivative $\mu'(x + n/2)$ cannot strictly be estimated by $({}_nM_{x+n} - {}_nM_{x-n})/2n$, for the M's are *weighted* averages of the μ's and in a growing population will always be too low. We can escape the difficulty by a second iteration. When an approximate value has been found for the l_x, we in effect have an approximation to the unweighted $\int_{x-n}^{x} \mu(a)\,da = -\log(l_x/l_{x-n})$, and can enter this divided by n in place of the ${}_nM_{x-n}$. In short, we would substitute for ${}_nM_{x+n} - {}_nM_{x-n}$ the quantity $(1/n)\log(l_x l_{x+n}/l_{x-n} l_{x2n})$ obtained on the first iteration.

The numerical effect can be judged from the following values obtained from Table 2.1:

Age	${}_nM_{x+n} - {}_nM_{x-n}$	$\frac{1}{n}\log\left(\frac{l_x l_{x+n}}{l_{x-n} l_{x+2n}}\right)$
45	0.00674	0.00679
50	0.01061	0.01070
55	0.01612	0.01625
60	0.02187	0.02215

The differences ${}_nM_{x+n} - {}_nM_{x-n}$ are in all cases too low, but the largest discrepancy is about 1.25 percent. This means that our correction, itself of the order of 1 percent of ${}_nM_x$, would be raised by about 1 percent on the iteration. Few users will regard this correction of the correction as numerically important.

2.2.4 Greville and Reed–Merrell Methods Derived as Special Cases

The generality of (2.2.10) can be demonstrated by applying it to derive a well-known expression due to Greville (1943):

$$\int_0^n \mu(x+t)\,dt = n\,{}_nm_x + \frac{n^3}{12}\,{}_nm_x^2(\log {}_nm_x)', \tag{2.2.13}$$

in terms of ${}_nm_x$, the life table death rate defined as ${}_nd_x/{}_nL_x$, where the prime again signifies a derivative.

The demonstration starts by writing $l(x)$ for $p(x)$ in (2.2.10) and noting that $l'(x+n/2) = -l(x+n/2)\mu(x+n/2)$. Thus, when ${}_nM_x$ is replaced with ${}_nm_x$ and ${}_nN_x$ with ${}_nL_x$ (as though the data came from a stationary rather than an increasing population), (2.2.10) becomes

$$\begin{aligned}\int_0^n \mu(x+t)\,dt &= n\,{}_nm_x + \frac{n^4}{12\,{}_nL_x} l\left(x+\frac{n}{2}\right)\mu\left(x+\frac{n}{2}\right)\mu'\left(x+\frac{n}{2}\right)\\ &\approx n\,{}_nm_x + \frac{n^3}{12}\,{}_nm_x\,{}_nm_x',\end{aligned}$$

if we approximate $l(x+n/2)/{}_nL_x$ by $1/n$ and $\mu(x+n/2)$ by ${}_nm_x$. Multiplying and dividing the correction terms on the right of the last expression by ${}_nm_x$, and then using the fact that ${}_nm_x'/{}_nm_x = (\log {}_nm_x)'$, provides Greville's result (2.2.13).

Greville expressed his result as (2.2.13) to make use of the virtual constancy of $(\log {}_nm_x)'$ through most ages and for most life tables. If this is taken as $(\log {}_nm_x)' = 0.096$, then (2.2.10) becomes

$$\int_0^n \mu(x+t)\,dt = n\,{}_nm_x + 0.008n^3\,{}_nm_x^2,$$

so that the survival probability is

$$\frac{l_{x+n}}{l_x} = \exp\left[-\int_0^n \mu(x+t)\,dt\right] = \exp(-n\,{}_nm_x - 0.008n^3\,{}_nm_x^2), \tag{2.2.14}$$

which is the expression derived empirically by Reed and Merrell (1939). The tabulation of (2.2.14) included in the Reed–Merrell paper has been used more extensively than any other system for making a life table. It was indeed convenient in the days before computers were available, but it rests on two gross assumptions: (1) that the observed population provides the stationary age-specific rates of the life table; and (2) that the same form, $0.008n^3\,{}_nm_x^2$, serves to correct the ${}_nm_x$ for all ages and for any life table. The simple expression 2.2.12 avoids both these restrictions.

2.2.5 Bounds on Error

All of the above discussion would be superfluous if there were some *correct* way to make a life table. Unfortunately ignorance of the distribution of population and deaths within each 5-year age interval has to be compensated for by more or less arbitrary assumptions. For most populations single years are sought from the respondent, but the information is published only in 5-year intervals, a wise policy on the part of the statistical authorities in view of the inaccuracy of individual reporting. Whether the age intervals 5 to 9, 10 to 14, and so on, are the best is another matter; concentration on multiples of 5 causes these to understate the mean age in comparison with 3 to 7, 8 to 12, and so on.

The deaths and population can of course be interpolated so finely that the outcome is unique whatever the interpolation formula. Whole years or fifths of a year are sufficiently fine since the uncertainty of converting ${}_nM_x$ into ${}_nq_x$ decreases with the cube of the interval n. To be convinced of this, compare two formulas generalized from (2.2.1) and (2.2.6), respectively:

$${}_nq_x = 1 - e^{-n_nM_x} \qquad \text{and} \qquad {}_nq_x = \frac{n_nM_x}{1 + (n/2)({}_nM_x)}.$$

Expanding these in powers of n gives

$$1 - e^{-n\,{}_nM_x} = n_nM_x - \frac{n^2\,{}_nM_n^2}{2!} + \frac{n^3{}_nM_x^3}{3!} - \cdots$$

and

$$\frac{n_nM_x}{1 + (n/2)({}_nM_x)} = n_nM_x - \frac{n_n^2M_x^2}{2} + \frac{n^3{}_nM_x^3}{4} - \cdots.$$

The two agree up to the term in n^2; then the exponential is lower by the difference $n^3{}_nM_x^3/12$, disregarding higher-order terms. Thus the difference in single years is approximately 1/125 of the difference in 5-year age groups.

But interpolation cannot provide a uniquely correct table, since it depends on a choice of formula that is inevitably arbitrary. Iterative methods also make various assumptions; one such method (Keyfitz 1968, p. 19) supposes local stability—that the observed population has been increasing uniformly within 5-year age intervals. The method of the present chapter avoids both interpolation and iteration; its arbitrariness involves the derivatives for the correction to $l_{x+n}/l_x = e^{-n_nM_x}$ in (2.2.12), calculated by stretching a straight line between the age intervals below and above the one of interest, and experimenting has shown that different ways of calculating the first or higher derivatives make little difference.

Lacking knowledge of the true life table, can we at least set bounds on it? We can, establishing the higher bound by supposing the population (and corresponding deaths) within the age group x to $x + n$ to be all concentrated at the low end. This would mean that the observed death rate really refers to exact age x, so that what we call ${}_nM_x$ is not an average of the rates $\mu(t)$ from x to $x+n$, but is really $\mu(x)$. If, on the other hand, the population is all concentrated at the high end, what we observe as ${}_nM_x$ is really $\mu(x+n)$. These two opposite possibilities furnish the extreme bounds; the ${}_nM_x$, and hence the l_x column derived from it, could refer to a population $n/2$ years younger than stated, or to a population $n/2$ years older. No logic can demonstrate that either such freak situation is impossible, improbable though it may be in even a small population.

For the United States life table of 1967, $\overset{o}{e}_{15}$ for males was 54.22 years. On the above-mentioned argument this number could really represent anything from $\overset{o}{e}_{12\frac{1}{2}}$ to $\overset{o}{e}_{17\frac{1}{2}}$. Since at these ages $\overset{o}{e}_x$ is declining by about 4.75 years per 5 years of age, the true $\overset{o}{e}_{15}$ could lie approximately in a range from $54.22 - 4.75/2$ to $54.22 + 4.75/2$, or from 51.84 to 56.60. Such a range is too wide to be of much practical interest, and yet it is hard to see the logic by which one can narrow the possibilities. Not only is there no correct life table, but also there is not even a simple way of establishing a realistic range of error, analogous to the 0.95 confidence interval that is used where a probability model applies.

A lower bound to the error of the life table is obtained by supposing that individuals die independently at random, each with probability ${}_nq_x$ for his age x. The expression for this is easily derived (Keyfitz 1968, p. 341, is a secondary source with references). But such an error seems as far below the true error as that of this section is above.

2.3 Further Small Corrections

The method of calculating a life table expressed in (2.2.12) has proved highly satisfactory in practice, giving negligible departure from graduated life tables and from iterated tables, without requiring either graduation or iteration. However, it depends on solving basic equation 2.2.2 for the integral for $\mu(a)$, and therefore some readers may wish to look more closely at the rationale of (2.2.2).

Measure of Exposure. A diagram, due to Lexis (1875), that displays the population by age and time will help in this. Each individual at any moment is represented by a point; the collection of points for any individual is his life line through time; the end of the line is at the moment and age of his death.

Figure 2.1 shows the beginning and end of the year 1967, for which the observations are being analyzed, as horizontal lines, and ages 65 and 70 as vertical lines. In the rectangle $ABCD$, 122,672 males lines come to an end for the United States in 1967. We do not quite know how many lines are in the rectangle, but it was estimated that 2,958,000 crossed the horizontal line for July 1, 1967, and this number, ${}_5N_x$ in general, is commonly used to estimate exposure; it would be better to use person-years.

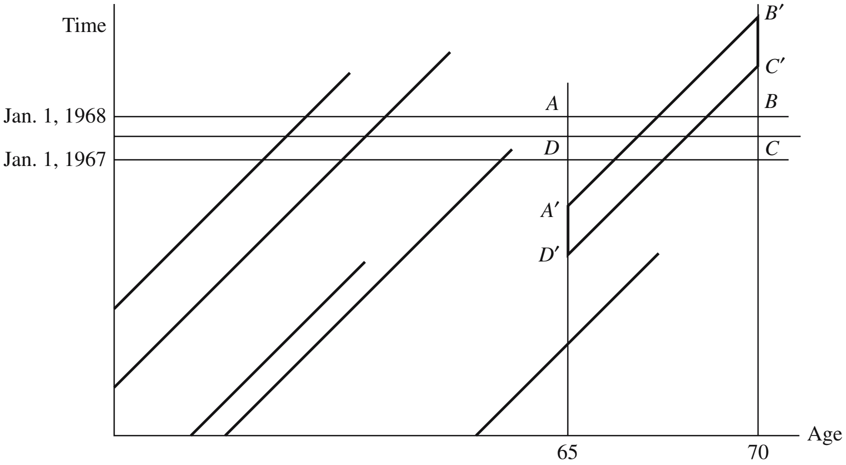


Figure 2.1. Living and dying population displayed on plane of age and time in Lexis diagram.

Table 2.2. Population and deaths, United States males, 1972, along with their first and second differences

Age x	Population $_5N_x$	$\Delta_5 N_x$	$\Delta_5^2 N_x$	Deaths $_5D_x$	$_5M_x = \frac{_5D_x}{_5N_x}$	$\Delta_5 M_x$	$\Delta_5^2 M_x$
50	5,616,000			63,838	0.011367		
		−788,000				0.006725	
55	4,828,000		152,000	87,348	0.018092		0.002666
		−636,000				0.009391	
60	4,192,000		−262,000	115,208	0.027483		0.003083
		−898,000				0.012474	
65	3,294,000		−66,000	131,620	0.039957		0.007339
		−964,000				0.019813	
70	2,330,000		218,000	139,264	0.059770		0.008784
		−746,000				0.028597	
75	1,584,000		84,000	139,974	0.088367		0.006392
		−662,000				0.034989	
80	922,000			113,734	0.123356		

Other directions of improvement, which the reader may wish to investigate, include the use of higher derivatives in the Taylor expansion of (2.2.8), and better ways of estimating the first derivative. Differences useful in approximating to the derivatives are provided in Table 2.2. Their effect on the correction C of Table 2.1 will be found to be slight.

2.4 Period and Cohort Tables

We have noted that, although in each age group the life table estimates how much a cohort will diminish by death, the probabilities for the several ages are chained together in a way that can represent only the period for which the data are gathered. The result describes what is sometimes referred to as a synthetic cohort.

To follow genuine cohorts we can chain together the survival probabilities of a sequence of periods. In 1919–21 the chance of survival for 5 years for a male just born was 0.87420 (Keyfitz and Flieger 1968, p. 142); in 1924–26 the chance of survival from age 5 to age 10 was 88,574/89,600 = 0.98855 (p. 144); in 1929–31 from age 10 to age 15 it was 88,814/89,587 = 0.99137 (p. 145); and so on. Chaining these together gives

$$l_0 = 100{,}000, \quad l_5 = 87{,}420, \quad l_{10} = 86{,}419, \quad l_{15} = 85{,}673 \quad \text{etc.},$$

as an estimate of survivorship for a real child born about 1917. Portrayal on a Lexis diagram suggests that the result is a reasonably good approximation, provided that mortality does not change abruptly over time. Refinements of the cohort calculation are being carried out by Michael Stoto, who has improved estimates of population and deaths in the lozenge $A'B'C'D'$ of Figure 2.1.

2.5 Financial Calculations

Demography has a part of its origin in actuarial calculations, just as probability sprang from gambling consultancies. The center of gravity of demography has shifted far from the insurance business, but at least the style of actuarial calculations is worth exhibiting here, perhaps as a small contribution toward bringing these disciplines closer together again.

2.5.1 Single-Payment Annuity and Insurance

The present value of a life annuity of 1 dollar per year, paid continuously starting from age x of the person, is equal to

$$\frac{1}{l_x}\int_0^{\omega-x} l(x+t)\,dt = \overset{o}{e}_x$$

dollars, if money carries no interest. It needs no mathematics to see that the expected number of dollars that a person will receive at the rate of 1 for each year of life is the same as the expected number of years he has to live. If money carries interest continuously compounded at annual rate i, a payment t years from now has a present value of e^{-it}; hence the present value of the annuity is

$$a_x = \frac{1}{l_x}\int_0^{\omega-x} e^{-it} l(x+t)\,dt \tag{2.5.1}$$

dollars. Similarly the present value of an assurance of 1 dollar on a life now aged x must be

$$A_x = \frac{1}{l_x}\int_0^{\omega-x} e^{-it} l(x+t)\mu(x+t)\,dt \tag{2.5.2}$$

dollars, which is the same as (2.5.1) except for the factor $\mu(x+t)$ in the integrand. [Show that, if interest is zero, $A_x = 1$, corresponding to the fact that dying is inevitable. Show also that, if $i > 0$, $A_x > \exp(-i\overset{o}{e}_x)$.]

2.5.2 Annual Premiums and Reserves

To find the annual premium P_x, note only that the annuity of premiums must cover the assurance; that is to say, the quantities $P_x a_x$ and A_x must be equal. Hence we have for the premium $P_x = A_x/a_x$ or, written out in full,

$$P_x = \frac{\displaystyle\int_0^{\omega-x} e^{-it} l(x+t)\mu(x+t)\,dt}{\displaystyle\int_0^{\omega-x} e^{-it} l(x+t)\,dt}. \tag{2.5.3}$$

In the early years of the policy, claims will be less than premiums, and a reserve will accumulate that will be drawn on in later years. After the policy has been in force for y years the present value of the claims will be A_{x+y}, and cannot be covered by the present value of the premiums $P_x a_{x+y}$, a smaller quantity for the ages at which mortality is rising; the difference $A_{x+y} - P_x a_{x+y}$, is the reserve prospectively needed.

2.6 Cause-Deleted Tables and Multiple Decrement

If a person dies of one cause, he cannot die at some later time of another; the literature speaks of competing causes of death. Since dying of a given cause avoids exposure to other causes, if we wish to know what the mortality from these other causes would be if the given cause were deleted, we need upward adjustment to the observed rates. Over a finite interval of time or age it would be incorrect to delete a cause simply by neglecting all deaths from this cause and calculating the life table from the remaining deaths; such a procedure would give too low mortality from the remaining causes. We need an estimate of exposed population that does not include persons dead of the deleted cause. To make the estimate it is customary to assume that the several causes act independently.

2.6.1 Dependence of Causes of Death

Think of a watch or other machine having parts and operating only as long as all the parts are functioning. Each part has its own life table; the chance that the ith part will operate for x years is $\bar{l}^{(i)}(x)$, calculated without reference to other parts. Then the chance that the watch will still be going

x years after its birth is

$$l(x) = \bar{l}^{(1)}(x) \cdot \bar{l}^{(2)}(x) \cdots \bar{l}^{(n)}(x), \tag{2.6.1}$$

a statement true if the mortality of each of the parts is unaffected by its incorporation in the watch. Then, if

$$-\frac{1}{\bar{l}^{(i)}(x)} \frac{d\bar{l}^{(i)}(x)}{dx} = -\frac{d \log \bar{l}^{(i)}(x)}{dx} = \mu^{(i)}(x),$$

it follows by taking logarithms and differentiating in (2.6.1) that

$$\mu(x) = \mu^{(1)}(x) + \mu^{(2)}(x) + \cdots + \mu^{(n)}(x). \tag{2.6.2}$$

The additivity of the instantaneous death rates follows from the multiplicativity of the survivorships.

We can think of many ways in which these conditions would not be fulfilled. The watch might keep going with one part defective but break down when two are defective. Or else the weakening of one part might put a strain on other parts; the watch would become "sick," and its death might ultimately be attributed to one of the parts that was so subjected to strain rather than to the part that became weak in the first place. There is no limit to the number of ways in which independence could be lost.

The parts of the watch might be said to resemble the parts of a person, and these would be independent if it could be said that he dies if his liver fails, or if his heart fails, and so on, and these were unrelated. The probability of each organ continuing to function might be thought of as the $\bar{l}^{(i)}(x)$ given in life tables showing the net probabilities of survival for the several causes. Unfortunately for this argument, human parts depend on one another more than do watch parts; separate life tables for the various organs, as though they were interchangeable, are to this degree artificial.

In addition to interrelations of parts, selection operates: the persons who die in skiing accidents are probably healthier on the average than the general population; therefore, if precautions that reduced the number of fatal accidents were introduced at ski resorts, the death rate from other causes would also be lowered. The nature of such dependencies is extremely difficult to establish. As long as we know only that A died of heart disease, B of kidney failure, and so on, nothing can be said about what would happen if one of the causes was reduced or eliminated. Faced with no evidence on the nature of dependencies, it is conventional to treat each cause as independent of the others. If parts (of a person or a watch) looked weak or had other clear symptoms before they failed altogether and brought the machine to a stop, and if such signs were present exactly y years before they would prove fatal, then something better could be done regarding dependencies. But even the most conscientious medical tests pre- and postmortem hardly provide such information.

It looks as though the analysis of interdependencies of the several organs will have to await further data. One kind of relevant data will become available when parts are commonly and safely replaced or interchanged among individuals. At present, however, such a possibility belongs to science fiction rather than to demography.

2.6.2 Method of Calculation

If all causes but the ith were removed, what would be the probability of surviving? To answer this without further data we suppose that the several causes act independently of one another, that is to say, that their forces of mortality are additive:

$$\mu_x^{(i)} + \mu_x^{(-i)} = \mu_x, \tag{2.6.3}$$

where $\mu^{(i)}$ is the mortality due to the ith cause, and $\mu^{(-i)}$ that due to all other causes. If the survival probability with only the ith cause acting is $\bar{l}_x^{(i)}$, we have from (1.6.5) and the additivity of the $\mu_x^{(i)}$, as a converse of the argument leading to (2.6.2),

$$l_x = \bar{l}_x^{(i)} \bar{l}_x^{(-i)}.$$

Thus essentially the same argument can start with the additivity of the $\mu_x^{(i)}$ and infer the multiplicativity of the $\bar{l}_x^{(i)}$, or else start with the latter as we did in (2.6.1).

Now expression 1.6.5 applies to the $\bar{l}_x^{(i)}$ just as it does to the l_x; hence, as pointed out by Jordan (1952, p. 258) and Chiang (1968, p. 246), we have

$$\begin{aligned}
-\log \frac{\bar{l}_{x+n}^{(i)}}{\bar{l}_x^{(i)}} &= \int_0^n \mu^{(i)}(x+t)\,dt \\
&= \left[\frac{\displaystyle\int_0^n \mu^{(i)}(x+t)\,dt}{\displaystyle\int_0^n \mu(x+t)\,dt} \right] \int_0^n \mu(x+t)\,dt \\
&= R \int_0^n \mu(x+t)\,dt,
\end{aligned}$$

say, and, on multiplying by -1 and taking exponentials of both sides,

$$\frac{\bar{l}_{x+n}^{(i)}}{\bar{l}_x^{(i)}} = \left(\frac{l_{x+n}}{l_x} \right)^R. \tag{2.6.4}$$

Once the ordinary life table from all causes together is available, all we need do is raise its survivorships to the powers represented by the R's; each is the ratio for an age group of the integral of the force of mortality for the ith cause to the corresponding integral for all causes.

Chiang (1968) takes R to be simply the ratio of the age-specific rates for the interval:

$$R = \frac{{}_nM_x^{(i)}}{{}_nM_x}. \tag{2.6.5}$$

We will use the technique of Section 2.2 to make a slight improvement on this. By expanding the $\mu^{(i)}(x+t)$ and $\mu(x+t)$ in a Taylor series about the midpoint of the n-year interval and then carrying through the integration, we find

$$R = \frac{\displaystyle\int_0^n \mu^{(i)}(x+t)\,dt}{\displaystyle\int_0^n \mu(x+t)\,dt} = \frac{\mu^{(i)}(x+n/2) + (n^2/24)\mu^{(i)\prime\prime}(x+n/2)}{\mu(x+n/2) + (n^2/24)\mu''(x+n/2)} \tag{2.6.6}$$

to second derivatives. Replacing the midperiod forces of mortality by the age-specific rates, using a symmetric estimate of the second derivatives, and then simplifying gives

$$R = \frac{{}_nM_x^{(i)}}{{}_nM_x}\left[1 + \frac{1}{24}\left(\frac{{}_nM_{x+n}^{(i)} + {}_nM_{x-n}^{(i)}}{{}_nM_x^{(i)}} - \frac{{}_nM_{x+n} + {}_nM_{x-n}}{{}_nM_x}\right)\right]. \tag{2.6.7}$$

The result would still hold with all M's replaced by D's where ${}_nD_x^{(i)}$ is the number of deaths observed from the ith cause between exact ages x and $x+n$.

In the way such calculations are used, the life table of interest is that in which one cause is deleted. For instance, on data for United States females in 1964, the complete expectation of life for females is $\overset{o}{e}_0 = 73.78$ years; if heart disease (CVR) is deleted, the $\overset{o}{e}{}_0^{(-\mathrm{CVR})} = 90.85$ years (Preston, Keyfitz, and Schoen 1972, p. 771). Deletion of cancer gives $\overset{o}{e}{}_0^{(-\mathrm{cancer})} = 76.34$ years. The gain by eradicating heart disease is 17.07 years; by eradicating cancer, 2.56 years.

2.6.3 Multiple Decrement

The probability that a person will die of a certain cause in the presence of other causes is presented in a multiple-decrement table. If the observed number dying of the given cause is ${}_nD_x^{(i)}$ and the life table number dying of that cause is ${}_nd_x^{(i)}$, we want to find how the life table deaths ${}_nd_x$ are distributed among the several causes, given the observed distribution among causes; given ${}_nd_x$, ${}_nD_x$, and ${}_nD_x^{(i)}$, we seek ${}_nd_x^{(i)}$. One way (Spiegelman 1968, p. 137) to make the calculation is

$${}_nd_x^{(i)} = \frac{{}_nD_x^{(i)}}{{}_nD_x}\,{}_nd_x, \tag{2.6.8}$$

but, as before, we try going one step in the refinement of this.

Table 2.3. Part of a multiple-decrement table applying to cardiovascular disease, comparing (2.6.9) with iterative and uncorrected methods, United States females, 1964

Age	${}_nd_x^{(i)}$	${}_nd_x^{(i)}$	${}_nd_x^{(i)} = {}_nd_x \dfrac{{}_nD_x^{(i)}}{{}_nD_x}$
x	Iterative	From (2.6.9)	From (2.6.8)
25	65	64.80	64.71
50	1,067	1,067.56	1,066.41
75	10,173	10,175.53	10,165.27

We have by definition

$$\frac{{}_nD_x^{(i)}}{{}_nD_x} = \frac{\displaystyle\int_0^n p(x+t)\mu^{(i)}(x+t)\,dt}{\displaystyle\int_0^n p(x+t)\mu(x+t)\,dt}$$

and applying the Taylor expansion used earlier results in

$$\begin{aligned} {}_nd_x^{(i)} = {}_nd_x \frac{{}_nD_x^{(i)}}{{}_nD_x} &\left\{1 + \frac{1}{48}\left(\frac{{}_nN_{x+n} - {}_nN_{x-n}}{{}_nN_x} + 2n{}_nM_x\right)\right. \\ &\times \left.\left[\frac{{}_nM_{x+n} - {}_nM_{x-n}}{{}_nM_x} - \frac{{}_nM_{x+n}^{(i)} - {}_nM_{x-n}^{(i)}}{{}_nM_x^{(i)}}\right]\right\}. \end{aligned} \tag{2.6.9}$$

Table 2.3 shows that the correction in (2.6.9), as compared with using the simple ${}_nd_x^{(i)} = {}_nd_x[{}_nD_x^{(i)}/{}_nD_x]$, is trifling. We seem to have reached a point at which it is usually immaterial whether the correction is made or not. Nonetheless cases will arise where within an age group one cause is declining and the others rising, and then correction 2.6.9 will bring improvement.

2.7 The Life Table as a Unifying Technique in Demography

The painstaking development of a method for inferring probabilities from observed rates is justified by the fact that the same problem arises in many fields. The ordinary life table, for which the data consist of the number of deaths and the exposed population, both by age, is only the best-known example. Figure 2.2 shows some other applications.

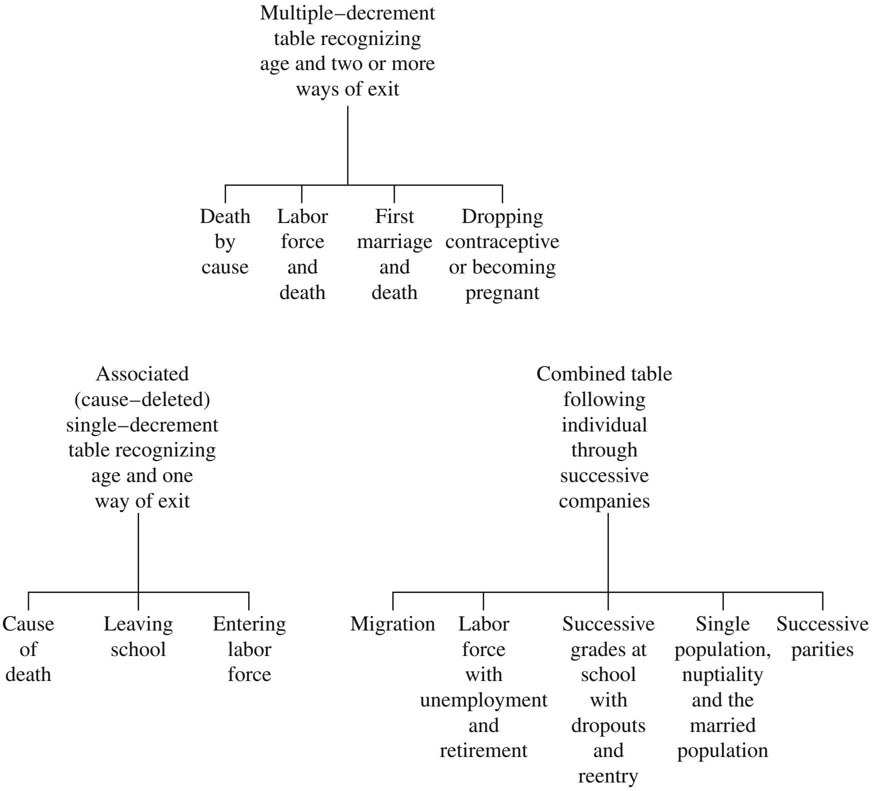


Figure 2.2. Some extensions of the life table.

3

The Matrix Model Framework

The life table, which formed the unifying technique for demographic analysis in Chapter 2, focuses on a very basic dichotomy (alive vs. dead) and follows individuals as they age. This chapter introduces a different, although related, unifying framework—that of the population projection matrix. It extends the dichotomy of the life table to consider individuals that differ according to many characteristics: age, sex, marital or employment status, maturity, etc. Over each interval of time, each type of individual has not only a probability of dying (on which the life table focuses) but also a probability of moving to another category (as when an unemployed person becomes employed) or of producing some number of new individuals (as by reproduction).

Given these probabilities, one can compute the expected population of each type of individual at the next time interval from that at the present. This calculation is called a *projection* of the population. The information needed for a projection can most easily be written down in the form of a matrix (called a population projection matrix) from which many characteristics of population dynamics can be calculated.

We will introduce the matrix model framework in the context of age-classified populations (here the relation to life table analysis is particularly close) and then proceed to populations classified by other factors. The ability of matrix models to classify individuals in many different ways is important to biologists because of the great diversity of plant and animal life cycles. Human demographers may not worry about individuals entering suspended animation, changing sex, shrinking instead of growing, or reproducing by breaking into pieces, but all these things (and more) occur among

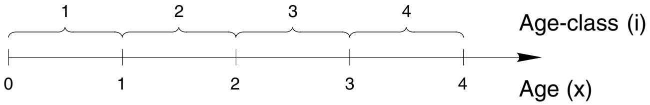


Figure 3.1. The relation between the continuous age variable x, used in the life table functions $m(x)$ and $l(x)$, and the discrete age classes i, used in the projection matrix parameters P_i and F_i.

nonhuman species. Humans too have many important attributes in addition to age. Attempts to fit these processes into the life table framework have a strained and unsatisfactory feel, but matrix methods are equally applicable to all of them (see Chapters 7 and 17, and AMD).

Matrix population models were introduced in the 1940s by Bernardelli (1941), Lewis (1942), and especially Leslie (1945, 1948), whose name is associated with them still. Leslie was an ecologist particularly interested in populations of small mammals (Crowcroft 1991). In addition to his work on matrix population models, he made the first life table calculation of the intrinsic rate of increase for any nonhuman species and made significant contributions to stochastic models and mark-recapture estimation. Matrix models were largely neglected until the mid-1960s, when both ecologists (Lefkovitch 1965) and human demographers (Lopez 1961, Keyfitz 1964, Rogers 1966) rediscovered them.

3.1 The Leslie Matrix

We begin with a model for an age-classified population. We divide the continuous variable *age*, which starts at 0, into a discrete set of *age classes*, which start at 1. The scheme is shown in Figure 3.1. Age class i corresponds to ages $i-1 \leq x \leq i$. According to this convention, the first age class is number 1. Some authors number the first age class as 0.

Our goal is to project the population from time t to time $t+1$. We assume that the unit of time is the same as the age class width. We call this unit the *projection interval*; its choice is one of the first steps in constructing a matrix model. Not surprisingly, a model that projects from year to year will differ from one that projects from month to month or decade to decade.

Suppose that the projection interval is one year, and that individuals are classified into three age classes (0–1, 1–2, and 2–3 years). The state of the population is described by a vector $\mathbf{n}(t)$, whose entries $n_i(t)$ give the numbers of individuals in each age class.

The individuals in age classes 2 and 3 at time $t+1$ are the survivors of the previous age classes at time t. That is,

$$\begin{aligned} n_2(t+1) &= P_1 n_1(t) \\ n_3(t+1) &= P_2 n_2(t), \end{aligned}$$

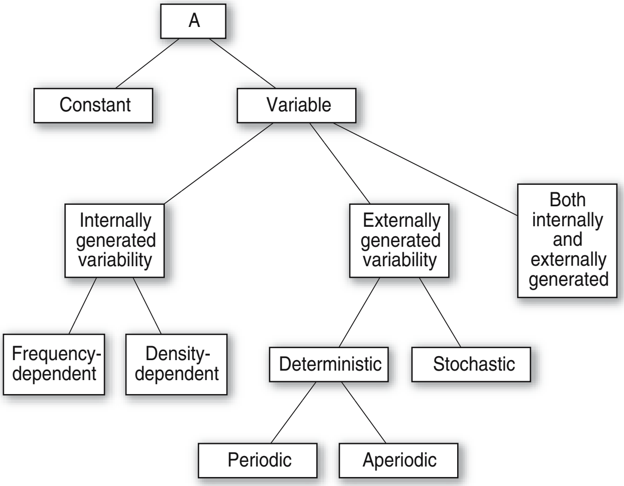


Figure 3.2. Classification of matrix population models depending on the kinds of variability included in the population projection matrix.

where P_i is the probability that an individual of age class i survives for one year.

The new members of age class 1 cannot be survivors of any other age-class; they must have originated from reproduction. Thus we write

$$n_1(t+1) = F_1 n_1(t) + F_2 n_2(t) + F_3 n_3(t), \tag{3.1.1}$$

where F_i is the per-capita *fertility*[*] of age class i, that is, the number of individuals in age class 1, at time $t+1$, per individual in age class i at time t.

These equations can be conveniently written in matrix form as

$$\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} (t+1) = \begin{pmatrix} F_1 & F_2 & F_3 \\ P_1 & 0 & 0 \\ 0 & P_2 & 0 \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} (t) \tag{3.1.2}$$

or, more compactly,

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t), \tag{3.1.3}$$

where $\mathbf{n}$ is the population vector and $\mathbf{A}$ is the population projection matrix. This special, age-classified version is often referred to as a *Leslie matrix.* It is nonnegative (since negative elements would imply negative numbers of

---

[*] We use *fertility* to describe actual reproductive performance and *fecundity* to denote the physiological potential for reproduction; in ecology the definitions are sometimes reversed.

individuals), with positive entries only in the first row (fertilities) and the subdiagonal (survival probabilities).

Matrix population models can be classified by the nature of the projection matrix **A** (see Figure 3.2). In the simplest case, the matrix is a constant. We will spend a lot of time on this case, in spite of the undeniable fact that survival and reproduction are not constant. The resulting model is a linear, time-invariant system of difference equations

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t). \tag{3.1.4}$$

If **A** is not constant, it may vary because of external factors independent of the population (e.g., weather), or because of changes in the internal state of the population itself. A variable external environment leads to a linear time-varying model

$$\mathbf{n}(t+1) = \mathbf{A}_t\mathbf{n}(t), \tag{3.1.5}$$

where each element of $\mathbf{A}_t$ may be a function of time. The variation may be deterministic or stochastic; if deterministic, it may be periodic or aperiodic.

Variation due to the population yields a nonlinear model

$$\mathbf{n}(t+1) = \mathbf{A}_{\mathbf{n}}\mathbf{n}(t), \tag{3.1.6}$$

where each entry of $\mathbf{A}_{\mathbf{n}}$ may be a function of the population vector **n**. Nonlinear models can be divided into density-dependent and frequency-dependent categories (two-sex models are generally frequency-dependent; MPM Chapter 17).

It is possible to combine environmental variation *and* density dependence or frequency dependence, to yield a system of inhomogeneous nonlinear equations

$$\mathbf{n}(t+1) = \mathbf{A}_{\mathbf{n},t}\mathbf{n}(t). \tag{3.1.7}$$

As you might expect, such models are difficult to analyze.

We will not explore time-varying or nonlinear models here, but they are treated at length in MPM.

3.2 Projection: The Simplest Form of Analysis

One of the advantages of matrix models is that they are easy to implement on a computer. Given an initial population $\mathbf{n}(0)$, you calculate the entries in $\mathbf{A}_{\mathbf{n}(0),0}$, taking into account whatever nonlinearities or time-variation may be operating. Use this matrix to produce $\mathbf{n}(1)$. Repeat the procedure:

$$\begin{aligned}
\mathbf{n}(1) &= \mathbf{A}_{\mathbf{n}(0),0}\mathbf{n}(0)\\
\mathbf{n}(2) &= \mathbf{A}_{\mathbf{n}(1),1}\mathbf{n}(1)\\
\mathbf{n}(3) &= \mathbf{A}_{\mathbf{n}(2),2}\mathbf{n}(2)
\end{aligned}$$

and so on, as long as desired. It is an easy and instructive exercise to program this repeated multiplication on a small computer (among other things, writing a program guarantees that you understand the rules for matrix multiplication). Here, we show the results of some simple projections of this sort.

Example 3.1 A linear, time-invariant model

Consider the projection matrix

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 5 \\ .3 & 0 & 0 \\ 0 & .5 & 0 \end{pmatrix}. \tag{3.2.1}$$

According to this matrix, the probability of surviving from the first age class to the second is $P_1 = 0.3$, and the probability of surviving from the second age class to the third is $P_2 = 0.5$. Individuals in the three age classes produce $F_1 = 0$, $F_2 = 1$, and $F_3 = 5$ surviving offspring per projection interval.

Figure 3.3 shows the results of applying this matrix to an initial population vector

$$\mathbf{n}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

For the first 15 time intervals, the abundances of the three age classes fluctuate irregularly, although there appears to be a slight upward trend. Looking at the results over a longer time scale and on a logarithmic abundance scale, reveals that each age class (and thus the total population) eventually grows exponentially at the same rate. The relative proportions of the three age classes eventually converge to constant values.

Example 3.2 Effects of initial conditions

The results in Figure 3.3 are specific to the initial population $\mathbf{n}(0)$. What if this initial population were changed? Figure 3.4 shows the result of 10 simulations, each with a different, randomly selected initial age distribution. All 10 eventually grow at the same rate and converge to the same age distribution. They do not, however, all achieve the same population size at any given time. One might suspect (given some foresight about the concept of reproductive value, to be introduced in Chapters 8 and 9) that this results from the timing of events in the life cycle. An initial condition biased toward individuals in age class 1 is at a disadvantage because these individuals must wait, with an attendant probability of mortality, one time step before they begin to reproduce and two time steps before they reach their maximum

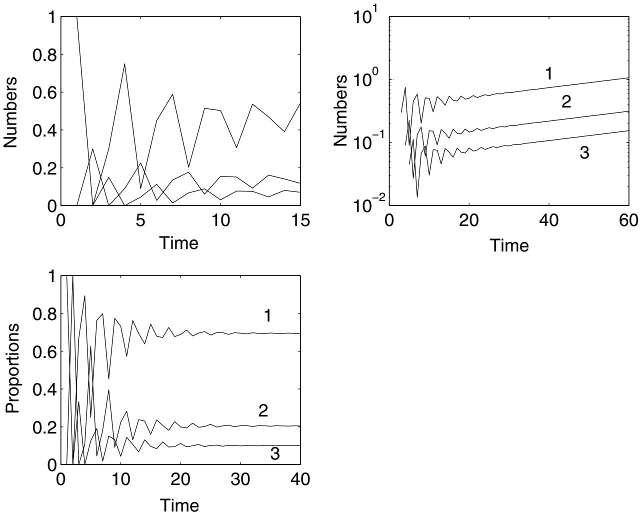


Figure 3.3. Projection of an initial population consisting of a single individual in age class 1, using the matrix in (3.2.1). Numbers on the line indicate age classes.

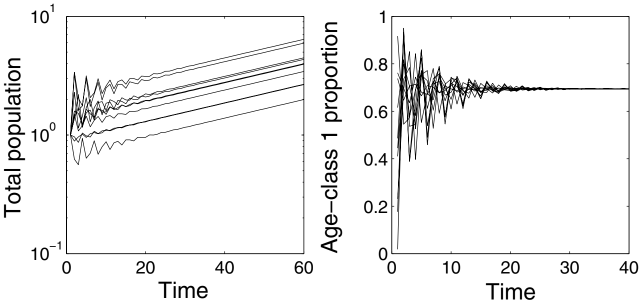


Figure 3.4. Projection of 10 randomly selected initial populations, all with the same total size, using the projection matrix of (3.2.1).

fertility. An initial condition biased toward age class 3 has an advantage because these individuals reproduce at their maximum rate immediately. Such a population has, and maintains, a head start over the others.

Example 3.3 Effects of perturbations

What happens if the entries in **A** are changed? Suppose that an environmental stress reduces the survival or reproduction of one age class by 10 percent, leaving all else unchanged. The results of these five perturbations are shown in Figure 3.5. Reducing survival or fertility

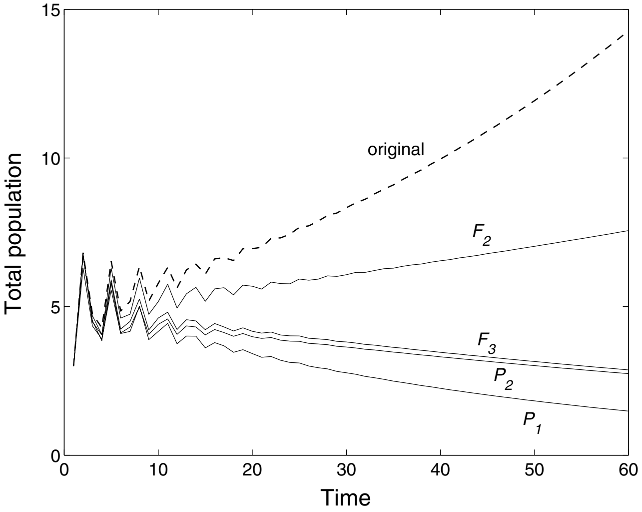


Figure 3.5. Projections of total population size, comparing the original projection matrix of Figure 3.3 (dotted line) with the results of reducing each of the vital rates by 10 percent.

reduces population growth rate (no surprise here), but *where* in the life cycle the change happens makes a big difference. A 10 percent reduction in F_2 reduces the growth rate but leaves it positive. The same reduction applied to F_3, P_2, or P_1 appears to drive the population to extinction. It appears that population growth is most sensitive to a proportional change in P_1.

3.2.1 A Set of Questions

The preceding examples have posed the four fundamental questions of demographic analysis.

Asymptotic analysis. A model describes a set of processes. Asymptotic analysis asks what happens if those processes operate for a very long time. What is the long-term behavior of the population? Does it grow or decline? Does it persist or go extinct? Converge to an equilibrium, oscillate, or do something more chaotic?

Ergodicity. The dynamics of a population depend not only on the model, but also on the initial conditions. A model (or the population it describes) is said to be *ergodic* if its asymptotic dynamics are independent of initial conditions. Ergodic results are useful because they imply that population patterns might reveal something about processes rather than initial conditions. Many scientists, as opposed to historians, have a strong belief that processes are important but initial conditions are historical accidents. On the other hand, if a model can be shown not to be ergodic, then it may be used to explain differences in dynamics in situations where the underlying processes are apparently the same.

Transient analysis. Short-term dynamics can be very different from asymptotic dynamics. Transient analysis, focusing on short-term behavior, may be more relevant than asymptotic analysis in characterizing the response of a population to perturbations.

Perturbation analysis. No matter how carefully it is constructed, a model always leaves things out, and the data from which parameter values are estimated are always imprecise. Any conclusions that depend on those *exact* values are immediately suspect. Furthermore, we usually want to extrapolate the results to other populations, other species, or other environments. Thus it is important to know how sensitive conclusions are to changes in the model. This kind of investigation is called *perturbation analysis* or *sensitivity analysis*. In the previous examples we applied it to the linear model to see how the growth rate changed when we modified survival and fertility.

Any study that does not address asymptotic analysis, transient analysis, ergodicity, and the results of perturbations has not completely explored its model.

3.3 The Leslie Matrix and the Life Table

The parameters in an age-classified matrix model are derived from the life table. We begin by distinguishing *birth-flow* populations, in which births occur continuously over the projection interval, and *birth-pulse* populations, in which reproduction is concentrated in a short breeding season within the interval (Caughley 1977). Humans are an example of a birth-flow population, whereas mammals, birds, and many other organisms in seasonal environments are more accurately described as birth-pulse populations. These patterns of reproduction produce different distributions of individuals within age classes, and lead to different approximations for the survival probabilities and fertilities.

3.3.1 Birth-Flow Populations

3.3.1.1 Birth-Flow Survival Probabilities

P_i is the probability that an individual in age class i will survive from t to $t+1$. This depends on the age of the individual within the age-class; the probability of survival from precise age x to $x+1$ is $l(x+1)/l(x)$. However, in forming age classes, we have given up all knowledge of age within the age class. Therefore, we approximate $l(x)$ within each age class by its average over the interval $i-1 \leq x \leq i$, so that

$$P_i = \frac{\int_i^{i+1} l(x)dx}{\int_{i-1}^{i} l(x)dx}. \tag{3.3.1}$$

In the notation of (2.1.1), this is

$$P_i = \frac{{}_1L_i}{{}_1L_{i-1}}, \tag{3.3.2}$$

which might be crudely approximated as

$$P_i \approx \frac{l(i)+l(i+1)}{l(i-1)+l(i)}. \tag{3.3.3}$$

Here, as is customary in matrix population models, time has been scaled in terms of the projection interval, so that the 5-year intervals of age and time in Chapter 2 appear here as 1 projection interval.

Other alternatives are possible. Assuming a constant force of mortality within the age interval suggests using the geometric rather than arithmetic mean as the approximation within the interval, in which case

$$P_i = \left(\frac{l(i)l(i+1)}{l(i-1)l(i)}\right)^{1/2} \tag{3.3.4}$$

$$= \left(\frac{l(i+1)}{l(i-1)}\right)^{1/2}. \tag{3.3.5}$$

A second alternative would be to calculate the probability of survival for each age and then average over the age interval:

$$P_i = \int_{i-1}^{i} \frac{l(x+1)}{l(x)}dx \tag{3.3.6}$$

$$\approx \frac{1}{2}\left(\frac{l(i)}{l(i-1)} + \frac{l(i+1)}{l(i)}\right). \tag{3.3.7}$$

Comparison of (3.3.3), (3.3.5), and (3.3.7) for several life tables suggests that the differences in the P_i are small (e.g., less than 2 percent in a life table for United States females using 5-year age classes), and probably irrelevant for most applications.

3.3.1.2 Birth-Flow Fertilities

The formulae for fertilities depend on the distribution of births and deaths within an age class (Leslie 1945, Keyfitz 1968). The F_i are defined by the first row of the projection matrix:

$$n_1(t+1) = \sum_i F_i n_i(t). \tag{3.3.8}$$

Let $B_{(t,t+1)}$ denote the total number of births in the interval $(t, t+1)$, and let $n(x,t)$ be the number of individuals aged $(x, x+dx)$ at time t (remember that x is a continuous variable). At time t, individuals of age x reproduce at the rate $m(x)n(x,t)$, where $m(x)dx$ is the expected number of female offspring produced by a female of age x in the interval $(x, x+dx)$. Integrating over time and age gives the total offspring production:

$$B_{(t,t+1)} = \int_0^\infty m(x) \int_t^{t+1} n(x,z)\,dz\,dx. \tag{3.3.9}$$

We, however, are ignorant of the detailed dynamics of $n(x,t)$ within the time interval, so we approximate $\int_t^{t+1} n(x,z)dz$ by the arithmetic mean of $n(x,t)$ and $n(x,t+1)$. With this approximation,

$$B_{(t,t+1)} \approx \int_0^\infty m(x) \left(\frac{n(x,t) + n(x,t+1)}{2} \right) dx. \tag{3.3.10}$$

Next, we approximate the continuous variables $m(x)$ and $n(x,t)$ by constant values (e.g., their means) m_i and $n_i(t)$ over the age interval $i-1 \leq x \leq i$. Given these approximations,

$$B_{(t,t+1)} \approx \frac{1}{2} \sum_{i=1}^\infty m_i \Big(n_i(t) + n_i(t+1) \Big). \tag{3.3.11}$$

However, $n_i(t+1) = P_{i-1} n_{i-1}(t)$ for $i \geq 2$. Substituting this expression into (3.3.11) and rearranging terms gives the number of births

$$B_{(t,t+1)} \approx \frac{1}{2} \sum_{i=1}^\infty \left(m_i + P_i m_{i+1} \right) n_i(t). \tag{3.3.12}$$

The number of births is not quite equal to $n_1(t+1)$; some of those offspring will not survive to time $t+1$. Those born just after t must survive almost an entire projection interval to be included in $n_1(t+1)$. Those born just before $t+1$ are at risk of mortality for only a brief time. An average individual must survive for one-half of the projection interval, the probability of which is $l(0.5)$. Thus

$$F_i = l(0.5) \left(\frac{m_i + P_i m_{i+1}}{2} \right). \tag{3.3.13}$$

If $l(0.5)$ is not known directly, it can be estimated using linear interpolation (Keyfitz 1968):

$$l(0.5) \approx \frac{l(0) + l(1)}{2}. \tag{3.3.14}$$

Most organisms, however, have relatively high neonatal mortality; in such cases logarithmic interpolation is more accurate. Suppose that $\mu(x) = \mu$ for $0 \leq x \leq 1$. Then $l(1) = l(0)\exp(-\mu)$ and

$$\begin{aligned} l(0.5) &= l(0)e^{-\mu/2} && (3.3.15)\\ &= l(0)\sqrt{l(1)}. && (3.3.16) \end{aligned}$$

If detailed survival information within the first age class is available, we can use $l(0.5) = \int_0^1 l(x)dx$ (Keyfitz 1968).

According to (3.3.13), the typical individual in age class i produces offspring at a rate that is the average of the maternity function for that age class and the subsequent age class, the latter weighted by the probability of survival to the subsequent age class. The offspring produced must survive for one-half time unit to be counted in the population at time $t+1$. An alternative interpretation of (3.3.13) is in terms of a typical individual in the center of age class i. It spends half of the projection interval producing offspring at the rate m_i, and, if it survives, passes into the next age class and produces offspring at the rate m_{i+1} for the rest of the interval.

3.3.2 Birth-Pulse Populations

In an ideal birth-pulse population, individuals reproduce on their birthday; thus the maternity function is a discontinuous series of delta functions (Figure 3.6). The age distribution at any time is also a discontinuous series of pulses (Figure 3.7). The age of these pulses depends on when the population is counted, relative to the time of the pulse of breeding. Let p $(0 < p < 1)$ denote the fraction of the time interval that elapses between the pulse of reproduction and the census. At census time, the age distribution consists of a series of pulses of individuals aged p, $1+p$, $2+p$, $\ldots$.

In many studies (e.g., birds, large mammals; see Caughley 1977) censuses are carried out either just before or just after breeding. These cases correspond to the limits as $p \to 1$ and $p \to 0$, and are called *prebreeding* and *postbreeding* censuses, respectively. The formulae for survival and fertility depend on which kind of census is assumed.

3.3.2.1 Birth-Pulse Survival Probabilities

In calculating birth-pulse survival probabilities, we no longer have to use approximate survival probabilities for "typical" individuals; every individual in age class i is identical, aged $i-1+p$. Thus

$$\begin{aligned} P_i &= P[\text{survival from age } i-1+p \text{ to } i+p] && (3.3.17)\\ &= \frac{l(i+p)}{l(i-1+p)}. && (3.3.18) \end{aligned}$$

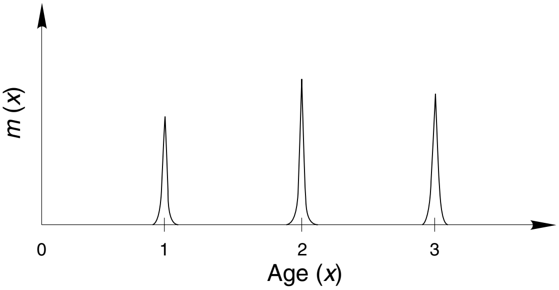


Figure 3.6. The maternity function for a birth-pulse population. Individuals reproduce only on their birthday.

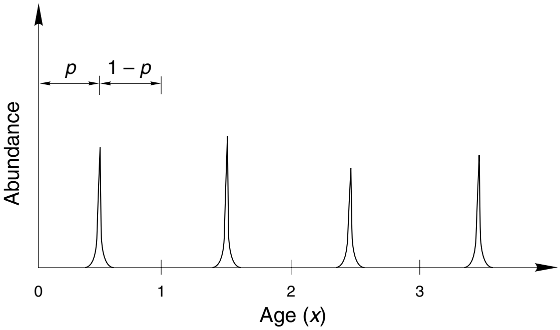


Figure 3.7. The age distribution at the time of census for a birth-pulse population. The time of the census is defined by p, the proportion of the time interval elapsing between the pulse of reproduction and the census.

Thus, as shown in Figure 3.8, the survival probabilities are calculated as

$$P_i = \begin{cases} \dfrac{l(i)}{l(i-1)} & \text{postbreeding } (p \to 0) \\[2ex] \dfrac{l(i+1)}{l(i)} & \text{prebreeding } (p \to 1) \end{cases} . \tag{3.3.19}$$

Mortality during the first age interval (say, the first year) appears in two different forms in prebreeding and postbreeding censuses. For a postbreeding census, $P_1 = l(1)/l(0)$, and includes first-year mortality. For a prebreeding census, $P_1 = l(2)/l(1)$, and excludes first-year mortality; in this case the missing mortality is incorporated into the fertility coefficients.

3.3.2.2 Birth-Pulse Fertilities

We begin by calculating the number of births in the interval $(t, t+1)$. These births occur when the individuals celebrate their next birthday, so

$$B_{(t,t+1)} = \sum_{i=1}^{\infty} n_i(t)\, m_i\, \phi_i, \tag{3.3.20}$$

where m_i is the reproductive output of an individual upon reaching its ith birthday and ϕ_i is the probability that an individual in age class i survives from the census to its next birthday. This probability can be approximated by assuming a constant force of mortality over the interval $(t, t+1)$. Since the probability of survival from t to $t+1$ is P_i, the probability of surviving for a fraction $1-p$ of a time unit is $\phi_i = P_i^{1-p}$. (Detailed information on seasonal mortality rates within the year could, of course, be used if it were available.)

Once reproduction has taken place, the offspring must survive a fraction p of a time unit in order to be counted in $n_1(t+1)$. That probability is given by $l(p)$; it may be estimated by interpolation, as $l(0.5)$ was for birth-flow populations [Equations (3.3.14) and (3.3.16)]. Thus the fertility coefficient for a birth-pulse population is given by

$$F_i = l(p) P_i^{1-p} m_i \tag{3.3.21}$$

$$= \begin{cases} P_i m_i & \text{postbreeding } (p \to 0) \\ l(1) m_i & \text{prebreeding } (p \to 1) \end{cases} . \tag{3.3.22}$$

Example 3.4 Parameterization of an age-classified matrix

Consider the following hypothetical life table for a population in which all individuals die by their fourth birthday.

x	$l(x)$
0	1.0
1	0.8
2	0.5
3	0.1
4	0.0

Application of (3.3.3) and (3.3.19) yields the following estimates for the P_i:

		Birth-pulse	
i	Birth-flow	$p \to 0$	$p \to 1$
1	$\frac{0.8+0.5}{1.0+0.8} = 0.722$	$\frac{0.8}{1.0} = 0.800$	$\frac{0.5}{0.8} = 0.625$
2	$\frac{0.5+0.1}{0.8+0.5} = 0.462$	$\frac{0.5}{0.8} = 0.625$	$\frac{0.1}{0.5} = 0.200$
3	$\frac{0.1+0.0}{0.5+0.1} = 0.167$	$\frac{0.1}{0.5} = 0.200$	$\frac{0.0}{0.1} = 0$
4	0	0	—

Notice that in the prebreeding census ($p \to 1$) there is no survival from the third to the fourth age class, because individuals in the fourth age class at the time of the census would be beginning to celebrate their fourth birthday, but the life table implies that none of them survives to do so. In the postbreeding ($p \to 0$) case, however, individuals in the fourth age class have just celebrated their *third* birthday; hence $P_3 > 0$.

Given the reproductive outputs for each age class, we can use (3.3.13) and (3.3.22) to calculate the fertility coefficients F_i:

			Birth-pulse	
i	m_i	Birth-flow	$p \to 0$	$p \to 1$
1	0	$0.9\left(\frac{0+2(0.722)}{2}\right) = 0.650$	$(0.8)(0) = 0$	$(0.8)(0) = 0$
2	2	$0.9\left(\frac{2+6(0.462)}{2}\right) = 2.052$	$(0.625)(2) = 1.250$	$(0.8)(2) = 1.6$
3	6	$0.9\left(\frac{6+3(0.167)}{2}\right) = 2.926$	$(0.200)(6) = 1.200$	$(0.8)(6) = 4.8$
4	3	$0.9\left(\frac{3+0(0.000)}{2}\right) = 1.350$	$(0)(3) = 0$	$(0.8)(3) = 2.4$

The parameterization of the matrix affects the results of the analysis. For example, the eventual rates of exponential growth (calculated using the methods of Section 7.1) implied by the three matrices in this example are

Birth-flow	1.793
Birth-pulse, $p \to 0$	1.221
Birth-pulse, $p \to 1$	1.221

The birth-pulse model yields the same growth rate regardless of the time of census, but the birth-flow and birth-pulse models predict quite different dynamics.

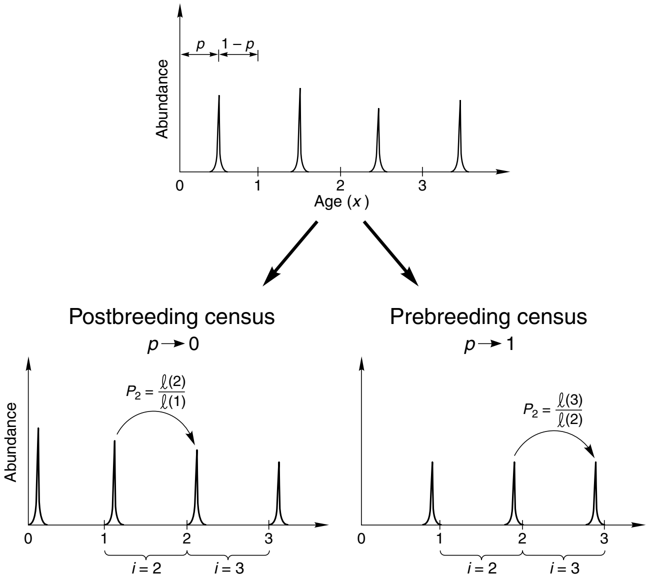


Figure 3.8. The calculation of survival probabilities for a birth-pulse population. The calculation of P_2 is shown; that of survival at other ages is similar. In a postbreeding census, individuals in age class 2 are just barely 1 year old; to survive for a year they must live to 2 years of age. In a prebreeding census, individuals in age class 2 are almost 2 years of age, and to survive a year they must live to reach 3 years of age.

3.4 Assumptions: Projection Versus Forecasting

Ecologists are, if anything, even more concerned than human demographers about the logic and assumptions underlying these analyses. So far, we have made three kinds of assumptions.

- We have assumed that it is appropriate to classify individuals by age. This is not an innocuous assumption; the demography of many organisms depends more on size or developmental stage than on age (for statistical analyses of the relative importance of age and other state variables see MPM Chapter 3).
- Because the model is discrete, it discards all information on the ages of individuals *within* age classes. When we choose age classes (or any other categories), we implicitly assume that this within-class information is irrelevant.
- The linear time-invariant model in Example 3.1 projects the population without changing the vital rates. This *seems* to assume that the fertilities and survival probabilities remain constant over time.

The third assumption seems absurd. The vital rates of most organisms vary conspicuously in time and space, and effects of density on population dynamics are well documented. How, then, can we justify analyses such as those we have just discussed? Worse yet, how can we justify the even more complex analyses of equally simple models that will follow in subsequent chapters?

The answer to this question is fundamental to the interpretation of demographic analyses. It turns on an important distinction between *projection* and *forecasting* or prediction (Keyfitz 1972a; see Chapter 12). A forecast predicts what *will* happen. A projection describes what *would* happen, given certain hypotheses. Grammatically speaking, forecasting uses the indicative mood and projection the subjunctive. For example, a projection matrix $\mathbf{A}$ implies an eventual population growth rate and structure. A population described by $\mathbf{A}$ would eventually grow at that rate with that structure, if nothing happened to change the vital rates. Ecological use of this purely analytical result is sometimes criticized as if it asserted that the environment *is* constant. However, one must assume a constant environment as a fact only if the model is being used as a forecast. No such assumption is required to interpret the growth rate and population structure as answers to the hypothetical question: how would the population behave if the present conditions were to be maintained indefinitely?

Population projections reveal something about present conditions (more precisely, about the relation between present conditions and the population experiencing them), *not* about the future behavior of the population. As Keyfitz (1972a) pointed out, one of the most powerful ways to study present conditions is to examine their consequences were they to remain as they are. A speedometer works the same way. A reading of 60 miles per hour *predicts* that, in one hour, the car will be found 60 miles in a straight line from its present location. As a forecast, this is almost always false. But as a projection, it provides valuable information about the present situation of the automobile.

So it is with demographic projections. They are particularly revealing because they integrate the impact of environmental conditions on vital rates throughout the life cycle. To know the survival probabilities and fertilities of every age class under a particular set of circumstances is to possess a great deal of biological information about those circumstances. This information is most valuable when coupled with a comparative approach, in which the vital rates are measured under two or more different conditions (e.g., Chapter 13). The use of demographic analysis in these studies does not depend on assumptions of density independence or environmental constancy.

3.5 State Variables and Alternatives to Age-Classification

Age-specific survival and fertility rates are not always sufficient to determine population dynamics. Even in human populations, for which age-classified demography was originally developed, factors other than age (sex, marital status, location) are known to affect the vital rates. In organisms with more complex life cycles, age is even less adequate, and demographic models should classify individuals by a more appropriate set of life cycle stages.

One of the first steps in an analysis is thus to choose a variable in terms of which to describe population structure. This choice can be understood in terms of the formal notion of "state" in dynamical system theory. The "state" of a system provides the information necessary to predict the response of the system. In Newtonian mechanics the state of a system is given by the positions and momenta of its component particles, because that information is sufficient to determine the response to any force. In ethology, the ideas of "motivation" or "drive" are used to describe the state of individual organisms, because they determine the response to a stimulus. Physiologists characterize individuals by their levels of energy reserves, lipid storage, hormones, metabolites, and so on. In demographic models the state of the population is usually given by the distribution of individuals among a set of categories (e.g., age classes). We will begin by examining the formal basis for these usually intuitive ideas.

3.5.1 State Variables in Population Models

Formal state theory was introduced into population ecology by Caswell et al. (1972), Boling (1973), and Metz (1977; see Metz and Diekmann 1986). Because demographic models connect individuals and populations, Metz and Diekmann (1986) recognized the need to begin with the state of the individual, which they called an *i-state.* Examples of i-state variables include age, size, maturity, developmental stage, and physiological condition. Papers by Hallam et al. (1990) and Gurney et al. (1990) and the book by Kooijman (1993) exemplify how much detail can be included in physiological i-states.

The i-state variable provides the information necessary to predict the response of an individual to its environment. However, we are interested in modelling the population and hence need a population state variable, or p-state variable. Metz and Diekmann (1986, Metz and de Roos 1992) gave two conditions sufficient to guarantee that the p-state variable can be written as a distribution of individuals among i-states (e.g., by an age distribution if individuals are characterized by age).

1. All individuals experience the same environment.

2. The effect of the population on the environment can be written as the sum of the contributions of the individuals.

This is an extremely important conclusion. It justifies the practice of writing p-states as vectors of abundances of individuals in various i-state categories, which is fundamental to all types of demographic models (Tuljapurkar and Caswell 1997).

The Metz–Diekmann conditions guarantee that we can test a would-be p-state variable by testing the adequacy of the corresponding i-state variable. This can be done by measuring the i-states for a set of individuals, documenting their responses to a common environment, and quantifying the information about individual fate provided by the i-state. Loglinear models (Section 17.6) are now widely used in ecology for this purpose (Caswell 1988, MPM Chapter 3).

The Metz–Diekmann conditions are often satisfied, at least approximately. For instance, in many demographic models, the effect of the population on the environment is not represented explicitly. Instead, the vital rates are written as functions of density, assuming some mechanism by which density is translated into an effect on the environment (e.g., consumption of resources), and hence back to the population. In such models, the second condition is satisfied automatically.

When the Metz–Diekmann conditions are violated, the state of the population cannot be described by the distribution of individuals among the i-states. Instead, the state of each individual must be accounted for explicitly. Such models are called i-state configuration models (Caswell and John 1992), individual-based models (DeAngelis and Gross 1992, Grimm et al. 1999), or, in the human demographic literature, microsimulation models (e.g., Ruggles 1993, Wachter 1997).

The conditions are violated when the vital rates are determined by interactions among specific individuals. The most common case involves sessile organisms where individuals interact only with their immediate neighbors. In such populations, each individual may experience a different environment, and the effects of an individual on the environment depend on where it is located and cannot be written as a sum of individual effects.

3.6 Age as a State Variable: When Does it Fail?

The adequacy of age as a state variable depends on how much information it provides about the demographically relevant aspects of individual development. Several circumstances combine to limit this information and make other state variables more suitable than age.

3.6.1 Size-Dependent Vital Rates and Plastic Growth

If the vital rates depend on body size and if growth is sufficiently plastic that individuals of the same age may differ appreciably in size, then age will provide little information about the fate of an individual. If the vital rates depend on developmental stage (e.g., instar) and stage duration varies among individuals, age will be a poor state variable. Such plasticity is widespread in plants, fish, and arthropods.

Size-dependent demography is not limited to species with indeterminate growth, however. Sauer and Slade (1987) have documented effects of body size on survival and reproduction in vertebrates, and they have used size-based demographic models for small mammals (Sauer and Slade 1985, 1986). Weight and body composition are known to affect the age at reproductive maturity in humans (Frisch 1984).

Size- or stage-dependent vital rates by themselves are not enough to render age inadequate as a state variable. If growth is so tightly regulated that age is a good predictor of size, then even if fecundity and mortality depend on size, age will work as an i-state variable. Some authors (e.g., Stearns and Koella 1986) have used this approach to develop models phrased in terms of size, but using age-classified demography.

3.6.2 Multiple Modes of Reproduction

Many organisms exhibit both sexual and vegetative or clonal reproduction (Jackson et al. 1985). Sexual and vegetative offspring of the same age may differ markedly in their demographic properties. Cook (1985), for example, summarizes data on several clonal plant species showing that the probability of successful establishment for vegetative offspring is from 3 to 30 times higher than the corresponding probability for offspring produced from seed. To the extent that individuals of the same age have different vital rates, age is an inadequate i-state variable for such species.

3.6.3 Population Subdivision and Multistate Demography

A third situation leading to inadequacy of age as an i-state variable results from population subdivision, when the subpopulations are exposed to different environments. Specification of an individual's state in such a population requires specification of its age *and* its environment. Spatial subdivision is an obvious example; if individuals migrate between regions characterized by different vital rates, the age × region distribution is required to specify the state of the population. There is a large literature on such "multiregional" models (e.g., Rogers 1975, 1985, 1995). These models also apply to populations with other kinds of heterogeneity (sex, marital status, parity, employment). The term "multistate demography" is used to describe models in which individuals are classified by multiple *i*-state variables (Land and Rogers 1982, Schoen 1988; see Chapter 17).

The choice of an appropriate state variable is a critical step in model construction. The concepts of *i*-state and *p*-state clarify the biological and statistical issues involved in that choice. They provide criteria for deciding, from studies of individuals, what variables need to be included in a population model.

3.7 The Life Cycle Graph

We turn now to constructing matrix population models for any discrete *i*-state variable; we will refer to these models as "stage-classified." Age-classified models are a special case of stage-classified models. We will see in Chapter 7 that one of the advantages of the matrix formulation is that results on asymptotic dynamics, ergodicity, transient analysis, and perturbation analysis apply equally to age- and stage-classified models.

We begin with a simple graphical description of the life cycle, called the *life cycle graph.* Figure 3.9 shows two examples. The construction proceeds as follows.

1. Choose a set of stages (i.e., *i*-states) in terms of which to describe the life cycle.
2. Choose the projection interval, defining the time step in the model.
3. Create a *node* for each stage; number the nodes from 1 through s. The order of the numbering is irrelevant, but it is often convenient to assign the number 1 to a stage representing "newborn" individuals. The symbol $\mathcal{N}_i$ denotes node i.
4. Put a directed line or *arc* from $\mathcal{N}_i$ to $\mathcal{N}_j$ if an individual in stage i at time t can contribute individuals (by development or reproduction) to stage j at time $t+1$. If an individual in stage i at time t can contribute to stage i at $t+1$ (e.g., by remaining in the same stage from one time to the next), put an arc from $\mathcal{N}_i$ to itself; such an arc is called a *self-loop.*
5. Label each arc by a coefficient; the coefficient a_{ij} on the arc from $\mathcal{N}_j$ to $\mathcal{N}_i$ gives the number of individuals in stage i at time $t+1$ per individual in stage j at time t. Thus

$$n_i(t+1) = \sum_{j=1}^{s} a_{ij} n_j(t).$$

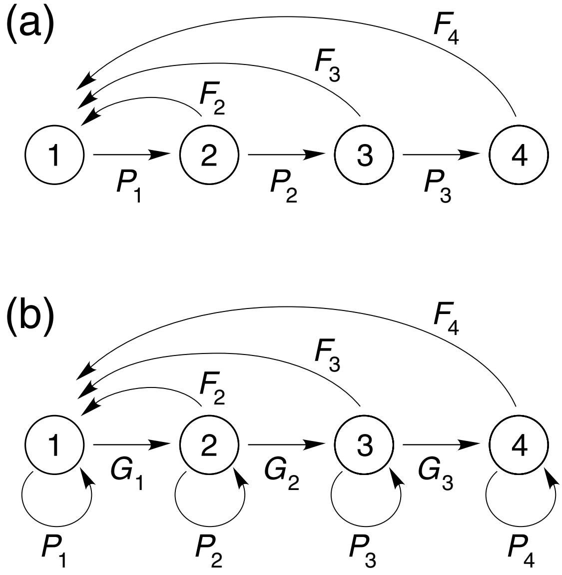


Figure 3.9. (a) A life cycle graph for an age-classified life cycle, in which the width of the age classes equals the projection interval. (b) A life cycle graph for the standard size-classified model. Nodes represent size classes, and an individual can grow no more than a single size class in the interval $(t, t+1)$.

> These coefficients may be transition probabilities or reproductive outputs.[†] Note the order of the subscripts!

Figure 3.9a shows the graph for an age-classified life cycle in which the age interval and projection interval are identical. Individuals survive, with probability P_i, to become one unit older, and reproduce, with fertility F_i, producing new individuals in age class 1. Figure 3.9b shows the graph for a size-classified life cycle. An individual in size class i may survive and grow to size class $i+1$ with probability G_i, or may survive and remain in size class i with probability P_i. Reproduction produces new individuals in the smallest size class. As drawn, this graph asserts that it is impossible for an individual to shrink, or to grow two or more size classes in a single time interval.

Figure 3.10 shows a life cycle graph for the killer whale *Orcinus orca* (Brault and Caswell 1993). The projection interval was one year, and individuals were classified as yearlings, juveniles (past their first year but not mature), mature females, and postreproductive females.

[†] The life cycle graph as defined here is known technically as the Coates graph of the matrix **A** (Chen 1976); its use in demographic analysis is due to Lewis (1972, 1976), Hubbell and Werner (1979), and Caswell (1982a,b). Some of the results that can be obtained from the graph will be seen in Chapter 9; see MPM Chapter 7 for details.

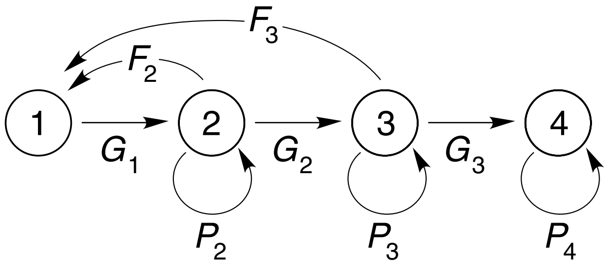


Figure 3.10. A life cycle graph for the killer whale (Brault and Caswell 1993). Nodes represent stages: $\mathcal{N}_1$ = yearlings, $\mathcal{N}_2$ = juveniles, $\mathcal{N}_3$ = mature females, and $\mathcal{N}_4$ = postreproductive females.

Three things about Figure 3.10 are worth noting. First, because stage 1 is an age class of the same length as the projection interval, individuals cannot remain in stage 1 from one time to the next. Thus $P_1 = 0$ and no self-loop has been drawn on $\mathcal{N}_1$. The loop could have been drawn and its coefficient set to zero without changing anything. Second, there is a postreproductive stage, which does not contribute individuals to any of the other stages. Third, it includes positive fertility for juvenile females. This arises because in a birth-flow model it is assumed that some juveniles may mature during the interval from t to $t+1$ and produce offspring at $t+1$.

3.8 The Matrix Model

The life cycle graph is more than a handy pictorial description of the life cycle. It is isomorphic to the population projection matrix $\mathbf{A}$ in the equation

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t),$$

where $\mathbf{n}(t)$ is now a vector of stage abundances. The rule for generating the matrix is simple: the matrix entry a_{ij} is the coefficient on the arc from $\mathcal{N}_j$ to $\mathcal{N}_i$ in the life cycle graph. Note the order of the subscripts.

The projection matrices corresponding to the two graphs in Figure 3.9 are

$$\mathbf{A}_a = \begin{pmatrix} 0 & F_2 & F_3 & F_4 \\ P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ 0 & 0 & P_3 & 0 \end{pmatrix} \tag{3.8.1}$$

$$\mathbf{A}_b = \begin{pmatrix} P_1 & F_2 & F_3 & F_4 \\ G_1 & P_2 & 0 & 0 \\ 0 & G_2 & P_3 & 0 \\ 0 & 0 & G_3 & P_4 \end{pmatrix}. \tag{3.8.2}$$

The age-classified graph (Figure 3.9a) yields a Leslie matrix. The matrix $\mathbf{A}_b$ resulting from the size-classified graph (Figure 3.9b) includes positive elements on the diagonal, corresponding to individuals remaining in the same size class, as well as on the subdiagonal and the first row.[‡] The matrix for the killer whale from the life cycle in Figure 3.10 is

$$\mathbf{A} = \begin{pmatrix} 0 & F_2 & F_3 & 0 \\ G_1 & P_2 & 0 & 0 \\ 0 & G_2 & P_3 & 0 \\ 0 & 0 & G_3 & P_4 \end{pmatrix}. \tag{3.8.3}$$

Deciding which properties of individuals to include in the life cycle calls on all the biological and sociological information available, and is always a compromise between recognizing the potential importance of even the tiniest differences and the acknowledging the limitations of the available data. But once the decision has been made and the projection matrix constructed, all manner of population analyses are possible. We return to these in Chapter 7.

[‡] This model, called the *standard size-classified model*, is widely used in ecology [e.g., Usher (1966), Pinero et al. (1984) for trees, Hughes and Jackson (1985) for corals, Heppel et al. (1996) for sea turtles, Doak et al. (1994) for tortoises, Warner and Hughes (1988) for fish].

4

Mortality Comparisons; The Male-Female Ratio

The United States in 1975 showed an expectation of life at birth for males of 68.5 years and for females of 76.4 years, a difference of 11.5 percent. But male death rates at most ages are at least 50 percent higher than female rates. The ratio of male to female rates, simply averaged over the ages, may show males 80 percent higher; the average with living population as weights may show males 70 percent higher; with deaths as weights males may be 50 percent higher. Our question is whether male mortality is 10, 50, 80, or 70 percent higher than female. The issue is raised in Sheps (1959) and Golini (1967).

The same difficulty arises when we compare two countries—the United States and Italy, or the United States and India—or two parts of the same country—South and North in Italy, or city and countryside. It arises also in comparisons through time: has mortality been improving in Europe and America during recent years, and if so by how much? Essentially the same issue can be expressed in terms of the conditional future: *if* mortality is reduced on an average of 5 percent at each age, what difference will this make to the expectation of life? The question is of special interest with respect to cause of death: if deaths from cancer are reduced by 10 percent on the average of the several ages, what will the effect be on expectation of life, provided that all other causes are unchanged? A more sophisticated form of the question is to suppose that deaths from cancer are reduced by 10 percent and to note the effect if cancer mortality is positively correlated to a given degree with other causes, that is, if those subject to cancer are somehow more susceptible also to other causes, but this is out of our scope.

Table 4.1. Ratio of male to female age-specific death rates for four countries

Age	United States, 1967	France, 1967	Greece, 1968	Mexico, 1966
–1	1.30	1.31	1.11	1.20
1–4	1.25	1.21	1.10	0.95
5–9	1.38	1.50	1.37	1.06
10–14	1.72	1.63	1.55	1.17
15–19	2.51	2.14	1.92	1.25
20–24	2.80	2.31	2.10	1.31
25–29	2.26	2.08	1.78	1.31
30–34	1.81	1.95	1.56	1.40
35–39	1.68	1.89	1.34	1.41
40–44	1.72	2.02	1.51	1.56
45–49	1.79	2.07	1.59	1.50
50–54	1.95	2.09	1.73	1.45
55–59	2.07	2.37	1.69	1.41
60–64	2.10	2.35	1.68	1.21
65–69	1.91	2.19	1.62	1.23
70–74	1.82	1.80	1.32	1.12
75–79	1.56	1.59	1.17	1.08
80–84	1.33	1.41	1.06	0.88
85+	1.08	1.24	1.15	0.75

Source: Based on Keyfitz and Flieger (1971).

In the simplest aspect of the problem a person saved from death through one cause at a particular time is rescued only to be exposed to the same cause and others in the subsequent period. To avoid death from heart disease at an age typical of the ages at which heart disease occurs can offer only a brief respite, since overall death rates increase rapidly in later life. Lower mortality of young and middle-aged females allows them to survive into ages at which mortality is very high for everyone. That the curve of death rates by age is concave upward and that populations are heterogeneous in risk are what give rise to the questions with which the present chapter deals.

4.0.1 Variation by Age in the Sex Ratio of Mortality

To bring the theory into contact with concrete data we show the ratios of male to female age-specific death rates for four countries in Table 4.1. The tendency is toward a characteristic pattern, with peaks around 0, 20, and 60, and troughs around 1, 35, and 85, the pattern being shown clearly by the United States and less clearly by Mexico.

4.1 The Multiplicity of Index Numbers

The most familiar comparison of mortality is by direct standardization: how many deaths would occur in a given population if it had the age distribution of the (standard) population with which it is being compared, and what is the ratio of this number to the deaths in the standard population? Kitagawa (1964) widened the perspective by assimilating the issues to those of price index numbers. Her formulae can be applied readily to the excess mortality of males. If the male population is p_x^m, the female population p_x^f, the male death rate μ_x^m, and the female death rate μ_x^f, the relative mortality at age x is μ_x^m/μ_x^f, and the index in which such relatives are weighted by *male population* is

$$I_p^m = \frac{\sum_x p_x^m(\mu_x^m/\mu_x^f)}{\sum_x p_x^m}, \tag{4.1.1}$$

and by *female population* is

$$I_p^f = \frac{\sum_x p_x^f(\mu_x^m/\mu_x^f)}{\sum_x p_x^f}. \tag{4.1.2}$$

The index weighted by *male deaths* is

$$I_d^m = \frac{\sum_x p_x^m \mu_x^m(\mu_x^m/\mu_x^f)}{\sum_x p_x^m \mu_x^m}, \tag{4.1.3}$$

and by *female deaths* is

$$I_d^f = \frac{\sum_x p_x^f \mu_x^f(\mu_x^m/\mu_x^f)}{\sum_x p_x^f \mu_x^f} = \frac{\sum_x p_x^f \mu_x^m}{\sum_x p_x^f \mu_x^f}. \tag{4.1.4}$$

To one of these four indices preferred in demography is I_d^f, the directly standardized rate (4.1.4). In economics I_d^f is called an aggregative index, in which the μ_x might be prices and the p_x quantities; f might be the base period and m the current period in a time comparison.

In addition to the four index numbers of (4.1.1) to (4.1.4) we can obtain four further numbers by using the harmonic means. For the male-population-weighted harmonic mean H_p^m we have

$$H_p^m = \frac{\sum p_x^m}{\sum p_x^m/(\mu_x^m/\mu_x^f)},$$

and similarly for the other three indices obtained by weighting by females and by deaths.

One might have thought that taking the male death rates as the denominator and then inverting the result would give additional indices. However, this merely reproduces in a different order the eight indices described above.

Geometric averaging does give different results (Schoen, 1970). The male-population-weighted index would be

$$G_p^m = \left(\prod (\mu_x^m/\mu_x^f)^{p_x^m}\right)^{1/\sum p_x^m}$$

and similarly for G_p^f and others.

A further variant is the exponential of the harmonic mean of the logarithms of the ratios,

$$HL_p^m = \exp\left[\frac{\sum p_x^m}{\sum p_x^m/\log(\mu_x^m/\mu_x^f)}\right],$$

which gives another four ways of expressing the relation of the given population to the standard.

We could obtain additional indices by cross-weighting and averaging. One instance is $I = \sqrt{(I_d^f)(I_d^m)}$, which may well give about as refined a comparison as any. England and Wales show considerably higher death rates than the United States at older ages and lower rates at younger ages. To find how the two countries stand on the whole we can first standardize on the United States 1960 census age distribution taking both sexes together, and calculate the intercountry value of (4.1.4), which turns out to be 1.218. Using the same formula but standardizing on the 1961 census of England and Wales, we find 1.231. The geometric mean of these quantities is 1.224.

Other compromises between the given and the standard age distributions are possible, as well as other kinds of averages. However, the conceptual points raised later in this section will lead to some skepticism about the usefulness of extreme refinement.

4.1.1 Weighted Index of Male to Female Mortality

The ratios of Table 4.1 (but to more decimal places) were combined in the four ways designated above as I_p^m, I_p^f, I_d^m, I_d^f, that is, weighting by male population, female population, male deaths, and female deaths, as in expressions 4.1.1 and 4.1.4. They were also combined to give the harmonic mean H, the geometric mean G, and the harmonic mean of the logarithm HL. The results are shown in Table 4.2 for six countries. Especially for low-mortality countries weighting by population gives a much higher excess mortality of males than weighting by deaths. The reason is that deaths are relatively heavier than population at the oldest ages, where the age-specific rates of females approach those of males.

Note that the arithmetic ratios are higher than the geometric, which in turn are higher than the harmonic. Moreover, the geometric ratios are equal to the means of the corresponding arithmetic and harmonic ratios:

$$G = \sqrt{(I)(H)}.$$

Table 4.2. Ratio of male to female mortality, United States, 1967, and five other countries

Index	United States, 1967	Austria, 1966–68	France, 1967	West Germany, 1967	Greece, 1968	Mexico, 1966
I_p^m	1.871	2.035	1.918	1.879	1.578	1.211
I_p^f	1.874	1.997	1.912	1.855	1.573	1.218
I_d^m	1.732	1.706	1.863	1.649	1.363	1.176
I_d^f	1.614	1.560	1.669	1.546	1.297	1.144
H_p^m	1.774	1.890	1.847	1.780	1.530	1.186
H_p^f	1.779	1.866	1.840	1.781	1.526	1.192
H_d^m	1.654	1.620	1.775	1.580	1.317	1.142
H_d^f	1.530	1.487	1.589	1.485	1.260	1.105
G_p^m	1.821	1.961	1.884	1.839	1.555	1.198
G_p^f	1.825	1.930	1.878	1.818	1.550	1.205
G_d^m	1.694	1.663	1.820	1.614	1.339	1.160
G_d^f	1.572	1.522	1.628	1.514	1.278	1.125
HL_p^m	1.630	1.697	1.703	1.687	1.371	1.374
HL_p^f	1.629	1.675	1.699	1.673	1.371	1.375
HL_d^m	1.407	1.444	1.631	1.445	1.185	1.960
HL_d^f	1.286	1.334	1.470	1.370	1.158	1.359
$\frac{e_0^f}{e_0^m}$	1.108	1.102	1.109	1.091	1.052	1.056

For the male-population-weighted case for the United States of 1967 (Table 4.2)

$$I_p^m = 1.871, \qquad H_p^m = 1.774,$$

of which means the geometric mean is 1.822, the same as G_p^m except for rounding.

The index numbers for the countries of low mortality in Table 4.2 typically show males to have 50 percent higher mortality than females. On the other hand, the expectation of life for females is nowhere much more than 10 percent higher than that for males. The ratio of the female to male $\overset{o}{e}_0$, that is, $\overset{o}{e}{}_0^f/\overset{o}{e}{}_0^m$, may be thought of as the ratio of male to female death rates in the respective stationary populations, and why it is so much lower than the other indices is investigated in Section 4.3.

4.1.2 Aggregative Indices Versus Averages of Relatives

Suppose one set up the criterion that an index ought to be affected in about the same amount by a change of one death at one age as at another. At least, it can be argued, the effect of an added death at one age ought not to be 100 times as great as the effect at another age. The addition of a single death to males at age x raises the death rate by $1/p_x^m$ and thus raises the numerator of the aggregative index I_d^f of (4.1.4) by p_x^f/p_x^m. Insofar as the ratio of males to females does not vary greatly in absolute numbers among age groups, the weight given to an extra male death in I_d^f is approximately invariant with age.

The same point can be made more generally by considering the effect of a small change $\Delta\mu_x^m$ in a male death rate on I_d^f; taking a small finite difference, we obtain

$$\Delta I_d^f = \frac{p_x^f \Delta\mu_x^m}{\sum_a p_a^f \mu_a^f}.$$

The denominator of this expression is the same regardless of the age x at which the change occurs; the numerator $p_x^f \Delta\mu_x^m$ is of the order of magnitude of the additional absolute number of deaths.

Quite different is the effect on the weighted relatives I_p^m of (4.1.1), where the effect of an increment $\Delta\mu_x^m$ in the male death rate is

$$\Delta I_p^m = \frac{p_x^m(\Delta\mu_x^m/\mu_x^f)}{\sum_a p_a^m},$$

which depends on the reciprocal of the female death rate at the age in question. Now the numerator contains $p_x^m(\Delta\mu_x^m)$, which is the number of additional male deaths, but it also contains the reciprocal of μ_x^f, which can be anywhere from 1/0.001 to 1/0.1 or even fall outside this range. That a small change in male deaths can have an effect on the comparison 10 times as great at age 20 as at age 70 argues against the index I_p^m.

4.2 Should We Index Death Rates or Survivorships?

Superimposed on the great variation in the above indices is an even greater variation arising from the choice between death rates and survival rates for the comparison. Consider the United States versus England and Wales, each about the time of the last census, and take males aged 70 to 75 as an example. The probabilities of dying are as follows:

U.S.: ${}_5q_{70} = 0.2526$ versus England and Wales: ${}_5q_{70} = 0.2915$,

so that England and Wales, 1960–62, would have $0.2915/0.2526 = 1.154$ times as high a probability of dying as the United States in 1959–61. The corresponding probabilities of surviving are ${}_5p_{70} = 1 - {}_5q_{70}$ or

U.S.: ${}_5p_{70} = 0.7474$ versus England and Wales: ${}_5p_{70} = 0.7085$, so that the United States has 1.055 times the survival probability of England and Wales. We do not know from this whether to say that the United States is healthier (for the age group in question) than England and Wales by 5.5 percent, or that England and Wales are more hazardous than the United States by 15.4 percent.

Without further sharpening of the model, as Sheps (1959) insists, we must abandon the statement of a percentage excess in either direction and be satisfied with the difference. The most that can be said is that the difference in the probability of dying is $0.2915 - 0.2526 = 0.0389$, this being identical with the difference $0.7474 - 0.7085$ in the probability of surviving. To say more invites the inconsistencies of the preceding paragraph.

But suppose it were the case (presented here only for illustration) that Englishmen of this age die of all the causes Americans die of plus some causes special to them. American mortality, represented by the probability of dying within a year, is q, say, and English mortality for those who do not die of the American causes is δ. Then the English probability of dying is $q + (1 - q)\delta$, and the ratio of this to q is $1 + [(1 - q)/q]\delta$, an expression that involves q and δ intertwined in complex fashion. But the American chance of survival is $1 - q$, and the English is $(1 - q)(1 - \delta)$, so the ratio of probabilities of survival is $(1 - q)(1 - \delta)/(1 - q) = 1 - \delta$. This involves δ and omits q altogether. Its complement is simply δ, the pure additional English mortality, *obtainable easily from the ratio of survivorships but not in any direct way from the probabilities of dying.*

Insofar as we adhere to this model, we ought to make an index consisting of the weighted complements of the age-specific death rates, and then take the complement of the resultant index. But the argument depends on the Englishman's being subject to a source of extra mortality that enters only if he escapes the American sources. Sheps (1959) raises the point in regard to smokers and nonsmokers; the former do have a clear added hazard if they escape all the causes of death to which the latter are liable. It is not obvious that such a model applies to the whole of intercountry or intersex comparisons, but it could apply in part. There could be a climatic or other special hazard in England; there could be some added hazard through heart disease (or, at younger ages, accident) for males in addition to the dangers to which women are subject.

Let us take the Sheps model one stage further and suppose some limited overlap, that is to say, common causes of mortality as well as causes special to each group. Let both men and women be subject to q; and men in addition to δ_m and women in addition to δ_f, the additional causes being restricted to individuals who do not die of the common causes. These are the parts of the model assumed to underlie the observed survival probabilities for men (P_m) and for women (P_f). Then we have $P_m = (1 - q)(1 - \delta_m)$

and $P_f = (1-q)(1-\delta_f)$; the ratio of survivorships is

$$\frac{P_f}{P_m} = \frac{(1-q)(1-\delta_f)}{(1-q)(1-\delta_m)} = \frac{1-\delta_f}{1-\delta_m}.$$

This is not bad: the ratio of survivorships taken from the raw data gives exactly the ratio of the chances of escaping the special female and male hazards, respectively.

Compare this with the ratio of the probabilities of dying, the complements of the P's, now taken as male to female:

$$\frac{1-P_m}{1-P_f} = \frac{1-(1-q)(1-\delta_m)}{1-(1-q)(1-\delta_f)} = \frac{\delta_m + q(1-\delta_m)}{\delta_f + q(1-\delta_f)}.$$

The interpretation of this is straightforward only if $q = 0$, that is, if there are no common causes. If, however, there are common causes, the ratio of observed death rates does not provide the ratio δ_m/δ_f of the causes special to the sexes, but a biased estimate of this ratio. Where q may be appreciable, the terms in q in numerator and denominator obscure the δ_m/δ_f that it is natural to seek.

Uncertainty as to the overlap of causes, and hence uncertainty as to whether one ought to be comparing the chances of survival or of death, constitute a major obstacle to precise comparison of overall rates. In nonexperimental comparisons, this consideration serves as a caution against excessive refinement in index numbers of the type presented in (4.1.1) to (4.1.4).

4.3 Effect on $\overset{o}{e}_0$ of Change in $\mu(x)$

The index number problem applies as much to changes through time as to comparisons across space. Given a general initial age schedule of mortality, suppose a certain kind of change in that schedule and see what the effect is on the expectation of life. We do this first with a constant increase δ in $\mu(x)$ at all ages, so that $\mu(x)$ becomes $\mu(x) + \delta$.

When the fixed quantity δ is added to mortality at every age, the probability of surviving to age x becomes $\exp[-\int_0^x [\mu(a)+\delta]\,da] = e^{-\delta x} l(x)$, that is, is altered in the ratio $e^{-\delta x}$; if the odd probability was $l(x)$, the new one is $e^{-\delta x} l(x)$. Also, the new expectation of life (distinguished by *) is the integral of this through the whole of life:

$$\overset{o}{e}{}_0^* = \int_0^\omega e^{-\delta x} l(x)\,dx. \tag{4.3.1}$$

To find the effect on the expectation of life of the addition δ to the age-specific death rates we seek the derivative $d\overset{o}{e}{}_0^*/d\delta$:

$$\frac{d\overset{o}{e}{}_0^*}{d\delta} = -\int_0^\omega xe^{-\delta x} l(x)\,dx = -\bar{x}\,\overset{o}{e}_0,$$

evaluated at $\delta = 0$, if $\bar{x}$ is the mean age of the stationary population. In finite terms for δ small,

$$\Delta\overset{o}{e}_0 \approx -\bar{x}\,\overset{o}{e}_0\delta \qquad \text{or} \qquad \frac{\Delta\overset{o}{e}_0}{\overset{o}{e}_0} \approx -\bar{x}\delta. \tag{4.3.2}$$

The relative change in the expectation of life equals minus the change in the death rate times the mean age in the life table population. Thus, if the expectation of life is 70 years, 0.001 is subtracted from mortality $\mu(x)$ at every age, and the mean age in the life table population is 35 years, the fraction added to the expectation is approximately (35)(0.001) = 0.035 or, in absolute amount, (0.035)(70) = 2.45 years.

4.3.1 A Proportional Difference Uniform at All Ages

However, although the effect of a fixed difference in death rates is expressible in simple form, we are more likely to be interested in the effect of a given proportional difference. Ratios of age-specific mortality rates are hardly constant between any two groups, but for developed countries the ratio of male to female mortality ranges from about 1.10 to about 2.80 (Table 4.1); certainly the ratios are closer to constancy than the differences (not shown), which vary from about 0.002 to 0.0400. The largest ratio is 2 to 3 times the smallest; the largest difference is 200 times the smallest.

Suppose now that the death rate $\mu(x)$ is multiplied by $1 + \delta$, so that $\mu^*(x) = (1+\delta)\mu(x)$. Then the new probability of surviving to age x becomes

$$l_x^* = \exp\left[-\int_0^x \mu^*(a)\,da\right] = \exp\left[-\int_0^x (1+\delta)\mu(a)\,da\right] = l_x^{1+\delta},$$

and the new expectation of life is

$$\overset{o}{e}{}_0^* = \int_0^\omega l(a)^{1+\delta}\,da. \tag{4.3.3}$$

The application of this to the special function $\mu(x) = \mu_0/(\omega - x)$ is satisfactorily simple. Integrating and taking the exponential gives $l(x) = [1 - (x/\omega)]^{\mu_0}$. Integrating this in turn gives $\overset{o}{e}_x = (\omega - x)/(\mu_0 + 1)$. Then we have

$$\frac{\overset{o}{e}{}_0^*}{\overset{o}{e}_0} = \frac{\displaystyle\int_0^\omega [l(x)]^{1+\delta}\,dx}{\displaystyle\int_0^\omega l(x)\,dx} = \frac{\mu_0 + 1}{\mu_0(1+\delta) + 1}. \tag{4.3.4}$$

If mortality at all ages rises by δ, then $\overset{o}{e}{}_0^*$ declines but by a lesser amount, as (4.3.4) shows. For males compared with females δ might be 0.43, and μ_0 might be 0.30. Then by (4.3.4) the ratio $\overset{o}{e}_0/\overset{o}{e}{}_0^*$ equals $1.43/1.30 = 1.10$. Integrating the reciprocal of the expectation of life as the life table death rate, an excess of 43 percent in all age-specific death rates translates into an excess of 10 percent in the life table overall death rate, roughly consistent with Table 4.2.

But this is on a hyperbolic curve for $\mu(x)$; in general the proportionate change of $\mu(x)$ does not so easily translate into a change of $\overset{o}{e}_0$, and we need the flexible differential calculus to establish a constant useful for describing life tables. To find the effect of a small change δ on the expectation of life, we seek the derivative of the right-hand side of (4.3.3) with respect to δ and find

$$\frac{d\overset{o}{e}{}_0^*}{d\delta} = \int_0^\omega [\log l(a)] l(a)^{1+\delta}\, da, \tag{4.3.5}$$

a quantity that cannot be positive, since $l(a)$ cannot be greater than unity. In the neighborhood of $\delta = 0$ we have

$$\frac{\Delta \overset{o}{e}_0}{\overset{o}{e}_0} \approx \frac{\displaystyle\int_0^\omega [\log l(a)] l(a)\, da}{\displaystyle\int_0^\omega l(a)\, da}\delta = -H\delta, \tag{4.3.6}$$

say, where H is minus the mean value of $\log l(a)$, weighted by $l(a)$. (The measure H is called entropy or information in other contexts, and has been applied in population biology by Demetrius (1974, 1983); see Tuljapurkar (1982) for its relation to the rate of convergence to the stable age distribution.)

The ratio of integrals in (4.3.6) is necessarily negative, so that H is positive. We can imagine H as low as zero, if all mortality were concentrated at one age. If, for instance, everyone lives until 70 and then dies, $l(a)$ will be unity for all ages up to 70, and its logarithm will be zero. At the other extreme, if mortality μ is the same at all ages, we will have $l(x) = e^{-\mu x}$, $\overset{o}{e}_x = 1/\mu$, a constant, and

$$\begin{aligned}\frac{\Delta \overset{o}{e}_0}{\overset{o}{e}_0} &\approx \frac{\displaystyle\int_0^\omega [\log 1(a)] l(a)\, da}{\displaystyle\int_0^\omega l(a)\, da}\delta \\ &= \frac{\displaystyle\int_0^\infty -\mu a e^{-\mu a}\, da}{1/\mu}\delta = -\delta.\end{aligned}$$

In this case $H = 1$, and the proportional change in the death rates translates into the same change in the expectation of life, but of course in the

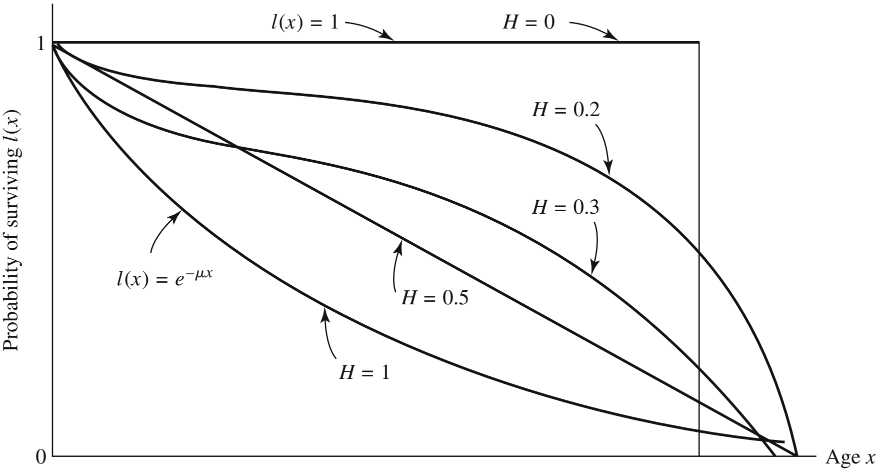


Figure 4.1. Extreme cases of survivorship curves $l(x)$, and three intermediate cases, showing how H becomes smaller as survivorship moves toward a rectangular form.

opposite direction. With $H = 1$, when the death rates at all ages increase by 1 percent, the expectation of life diminishes by 1 percent. [Show that for $l(x)$ a straight line, $H = 0.5$.]

Figure 4.1 shows the $l(x)$ for the two extreme cases and for intermediate ones, with corresponding values of H. Deevey (1950) gives a number of curves for animal species that resemble the several curves of Figure 4.1, with fruit flies near $H = 0$, oysters below $H = 1$, and hydra near $H = \frac{1}{2}$.

4.3.2 Observed Values of the Constant H

For countries of Europe and America with expectations of life around 70 years, H is now of the order of 0.2 for males and 0.15 for females, down from the 0.3 to 0.4 of about 30 years earlier. Apparently H is a convenient summary of the degree of concavity in an $l(x)$ column; as mortality improves, a larger fraction of deaths occurs in the 60s and 70s of age, and the drop in the value of H measures this tendency; with improvement in mortality everyone dies at about the same age, and a proportional improvement in mortality at all ages makes less and less difference in the expectation of life.

When we are told that, of two countries, or two sexes, one has an expectation of life in a certain ratio to the other, we can approximate to the ratio of age-specific death rates, supposing this ratio to be $1 + \delta$, the same at all ages. Taking

$$\frac{\overset{o}{e}{}_0^*}{\overset{o}{e}{}_0} = \frac{\displaystyle\int_0^\omega l(x)^{1+\delta}\,dx}{\displaystyle\int_0^\omega l(x)\,dx}$$

as a function of δ, say $f(\delta)$, and expanding as $f(\delta) = f(0) + \delta f'(0)$, where we know that $f'(0) = -H$, gives

$$\frac{\overset{o}{e}{}_0^*}{\overset{o}{e}{}_0} \approx 1 - \delta H.$$

If, for example, $\overset{o}{e}{}_0^*/\overset{o}{e}{}_0 = 1.10$, we can say that

$$1.10 = \frac{\displaystyle\int_0^\omega l(x)^{1+\delta}\,dx}{\displaystyle\int_0^\omega l(x)\,dx} \approx 1 - \delta H;$$

and if H is 0.20, we have the equation for δ:

$$1.10 = 1 - (\delta)(0.20),$$

or

$$\delta = \frac{1 - 1.10}{0.20} = -0.50.$$

The population with 10 percent greater expectation of life has death rates 50 percent lower. This approach can also be tried when the ratio of death rates is not uniform, and to this situation we proceed.

4.3.3 An Aspect of the Index Number Problem

The theory developed above seeks to find the effect on expectation of life of a uniform proportional excess in the force of mortality $\mu(x)$. What we have in practice, however, is different proportional increases at the several ages. One way of answering the question of what is the real average increase—for example, the real percentage excess of male over female mortality—is to calculate the proportional increase that, applied uniformly to all ages, will have the same effect on the expectation of life as the observed set of increases. In application to the excess mortality for the United States, 1967, which showed (Table 4.1) a ratio of 1.30 for ages under 1, 1.25 for those 1 to 4, etc., we would ask what uniform excess of male over female mortality would provide the same ratio of expectations of life as observed. Since we saw that raising $\mu(x)$ in the ratio $1 + \delta$ raises $l(x)$ to the power $1 + \delta$, we would need to solve the equation

$$\frac{\overset{o}{e}{}_0^m}{\overset{o}{e}{}_0^f} = \frac{\displaystyle\int_0^\omega l(x)^{1+\delta}\,dx}{\displaystyle\int_0^\omega l(x)\,dx} \tag{4.3.7}$$

for the unknown δ. This can be done directly by computer.

The equation can also be solved approximately in terms of our parameter H, along with other derivatives of the numerator of the right-hand side of the equation. Rewriting $l(x)^{1+\delta}$ in (4.3.7) as $l(x)\exp[\delta\log l(x)]$ and expanding the exponential term in a Taylor series changes the equation to

$$\frac{\overset{o}{e}{}_0^m}{\overset{o}{e}{}_0^f} = 1 - H\delta + \frac{H_2\delta^2}{2!} - \frac{H_3\delta^3}{3!} + \cdots, \tag{4.3.8}$$

where

$$H_i = \frac{\displaystyle\int_0^\omega [-\log l(x)]^i l(x)\,dx}{\displaystyle\int_0^\omega l(x)\,dx}, \quad i = 2, 3, \ldots.$$

Approximating with the linear term only, we have for Italy, 1964, whose $\overset{o}{e}{}_0^m/\overset{o}{e}{}_0^f = 1.081$,

$$\delta = \frac{\left(\overset{o}{e}{}_0^m/\overset{o}{e}{}_0^f\right) - 1}{-H} = \frac{(1/1.081) - 1}{-(0.207 + 0.163)/2},$$

then using H from Table 4.3. (It seems best to average H for males and females.) The result is

$$\delta = \frac{0.0749}{0.185} = 0.405,$$

or $1 + \delta = 1.405$. This compares with $I_d^f = 1.441$, or an average of I_d^f and I_d^m of 1.490. [For greater accuracy the reader may wish to experiment with further terms of (4.3.8).]

Thus, given the expectations of life of two populations, along with an average H, we can say what uniform excess of mortality of one over the other accounts for the ratio of expectations. This is the converse of the way in which H was originally derived—as the ratio of expectations of life corresponding to a given uniform ratio of death rates.

For the particular hyperbolic form in which

$$\mu(a) = \frac{\mu_0}{\omega - a} \tag{4.3.9}$$

$$l(a) = \left(1 - \frac{a}{\omega}\right)^{\mu_0} \tag{4.3.10}$$

Table 4.3. Values of $\overset{o}{e}_0$ and parameter $H = -\int_0^\omega [\log l(a)] l(a)\,da / \int_0^\omega l(a)\,da$ for males and females, United States, 1919–21 to 1959–61

	Male		Female	
	$\overset{o}{e}_0$	H	$\overset{o}{e}_0$	H
1919–21	54.59	0.3804	56.41	0.3547
1924–26	56.34	0.3401	59.01	0.3113
1929–31	57.27	0.3272	60.67	0.2942
1934–36	58.53	0.3105	62.58	0.2725
1939–41	61.14	0.2747	65.58	0.2361
1944–46	62.26	0.2632	68.11	0.2087
1949–51	65.28	0.2260	70.86	0.1823
1954–56	66.45	0.2134	72.61	0.1660
1959–51	66.84	0.2083	73.40	0.1594

Source: Computed from data in Keyfitz and Flieger (1968).

$$\overset{o}{e}_a = \frac{\omega - a}{\mu_0 + 1} \tag{4.3.11}$$

$$H = \frac{\mu_0}{\mu_0 + 1} \tag{4.3.12}$$

H tells us considerably more than is indicated by its being the first derivative. It happens also to give without approximation the value of the increase in $\overset{o}{e}_0$ (or $\overset{o}{e}_x$ for that matter) when $\mu(a)$ goes to $\mu(a)(1+\delta)$, and when δ may be large. For the ratio of expectations, we have

$$\frac{\overset{o}{e}{}_0^m}{\overset{o}{e}{}_0^f} == \frac{1}{1+\delta H},$$

under the hyperbolic mortality function assumed. In the example given, with $\overset{o}{e}{}_0^m / \overset{o}{e}{}_0^f = 1/1.081$, $H = 0.185$; hence the equation for δ is

$$\frac{1}{1.081} = \frac{1}{1+\delta(0.185)},$$

or $\delta = 0.081/0.185 = 0.438$. Thus the fact that Italian females of 1964 have 8.1 percent longer expected life than males is the equivalent of their having 43.8 percent as high mortality.

Although this holds exactly only for the special graduation $\mu(a) = \mu_0/(\omega - a)$, it is probably an improvement when δ is substantial for any life table. Hence the rule for finding the new $\overset{o}{e}_0$ when all rates rise by δ is to divide the old $\overset{o}{e}_0$ by $1 + \delta H$, rather than multiplying by $1 - \delta H$ as in the first approximation to (4.3.8).

4.3.4 Fractional Change in Mortality Due to a Given Cause

Of at least equal interest is to find the effect, not of a change in mortality in general, but of a change due to a given cause. We might want to investigate, for example, what would happen to the expectation of life if 10 percent of the cancer deaths were eliminated at each age, and we would need for this a cause-specified analogue of the H above. If the chance of surviving against the risk of cancer deaths alone is $l^{(i)}(x)$, and that against the risk of all other deaths is $l^{(-i)}(x)$, the probability that a person will still be alive by age x, $l(x)$, under independence is the product

$$l(x) = l^{(i)}(x)l^{(-i)}(x).$$

If we let 100δ represent the percentage change in the death rate due to cancer, the new expectation of life involves $l^{(i)}(x)$ raised to the $(1+\delta)$th power:

$$\begin{aligned}\overset{o}{e}{}_0^{*} &= \int_0^{\omega} l^{*}(x)\,dx = \int_0^{\omega} l^{(i)l+\delta}(x)l^{(-i)}(x)\,dx \\ &= \int_0^{\omega} l^{(i)\delta}(x)l(x)\,dx.\end{aligned}$$

Again resorting to the derivative, we have

$$\frac{d\overset{o}{e}{}_0^{*}}{d\delta} = \int_0^{\omega} [\log l^{(i)}(x)]l^{(i)\delta}(x)l(x)\,dx.$$

At the point where $\delta = 0$ we can drop the middle factor in the integrand and have for the derivative

$$\frac{d\overset{o}{e}_0}{d\delta} = \int_0^{\omega} [\log l^{(i)}(x)]l(x)\,dx. \tag{4.3.13}$$

If $H^{(i)}$ is defined as

$$H^{(i)} = \frac{-\displaystyle\int_0^{\omega} [\log l^{(i)}(x)]l(x)\,dx}{\displaystyle\int_0^{\omega} l(x)\,dx}, \tag{4.3.14}$$

then δ small but finite gives

$$\frac{\Delta\overset{o}{e}_0}{\overset{o}{e}_0} \approx -H^{(i)}\delta. \tag{4.3.15}$$

From (4.3.14) we see that $H^{(i)}$, like H, is minus a weighted average of logarithms, the weights being the $l(x)$ column in both cases. The logarithm of $l_x^{(i)}$ is closer to zero than the logarithm of l_x, so the linear approximation (4.3.15) is better than (4.3.6).

The several causes of death show characteristic values of $H^{(i)}$, and these values are worth studying for what they tell us about the effect on expectation of life of eradication of a small part of each cause. Table 4.4 presents $H^{(i)}$ values for 12 causes of death, for males and females, in the United States and Italy, 1930–31 and 1964. Taking, for example, cardiovascular renal diseases (cause 4 in the list assembled by Preston, Keyfitz, and Schoen, 1972), we see that for United States males in 1964 a drop of 1 percent in CVR deaths uniformly at all ages would result in an increase of 0.0840 percent in the expectation of life, that is, about one-twelfth as much.

Note the additivity of the $H^{(i)}$, in the sense that $\sum_i H^{(i)} = H$. For from the fact that $l(x) = \Pi l^{(i)}(x)$, with $l_0^{(i)} = 1$, it follows that the integral H defined in (4.3.6) must be the sum of the integrals $H^{(i)}$ defined in (4.3.14).

4.3.5 Comparison of $H^{(i)}$ with $\overset{o}{e}{}_0^{(-i)} - \overset{o}{e}_0$

A commonly used way of ascertaining the seriousness of a cause of death is to determine by how much the expectation of life at age zero would be increased if the cause in question were eliminated: letting $\overset{o}{e}{}_0^{(-i)}$ represent the expectation of life with this cause eliminated, we find that the increase in expectation of life equals $\overset{o}{e}{}_0^{(-i)} - \overset{o}{e}_0$. One trouble with subtracting $\overset{o}{e}_0$ from $\overset{o}{e}{}_0^{(-i)}$ is that the latter supposes the total elimination of the ith cause. What would happen with the total elimination of a cause is of less immediate interest than what would happen with the elimination of 1 percent, say, of that cause, and the latter is readily calculated from $H^{(i)}$.

Table 4.5 shows several facts about four causes of death for United States males and females, 1964. These include the years of life that would be added if the respective causes were eradicated, $\overset{o}{e}{}_0^{(-i)} - \overset{o}{e}_0$. Thus eliminating neoplasms would add 2.265 years for males. Against that we have the fact that $H^{(\text{neoplasms})}$ is 0.0302. A 1 percent drop in neoplasms at all ages would raise the expectation of life by $-0.01 \times H^{(\text{neoplasms})} \times \overset{o}{e}_0 = 0.01 \times 0.0302 \times 66.905 =$ 0.0202 year. Since H is applicable only to small uniform percentage changes at all ages, it is not strictly proper to multiply it by $\overset{o}{e}_0$ to find the result of completely eliminating the given cause. If such a multiplication is carried out for neoplasms, the result is 2.021 years, somewhat less than the $\overset{o}{e}{}_0^{(-i)} - \overset{o}{e}_0 = 2.265$ of Table 4.5. For some of the causes the agreement is closer than this. But results for cardiovascular renal diseases are much farther off, both for males and for females; the effect of eliminating 1 percent is to add far less than 1 percent of the years that would be gained by complete eradication.

The fact that $H^{(i)}$ is additive is a clear convenience—the reduction of all causes by δ would increase $\overset{o}{e}_0$ by an amount $H\delta$ equal to the sum of the effects on $\overset{o}{e}_0$ of the eliminating of the several causes of death $H^{(i)}\delta$:

$$H\delta = \sum H^{(i)}\delta.$$

Table 4.4. Values of $H^{(i)}$, giving effect on $\overset{o}{e}_0$ of small fractional decrease in each of 12 causes of death, United States and Italy, 1930–31 and 1964, males and females

	United States				Italy			
	1930		1964		1931		1964	
Cause of disease	Male	Female	Male	Female	Male	Female	Male	Female
Respiratory tuberculosis	0.0194	0.0190	0.0012	0.0005	0.0266	0.0286	0.0045	0.0014
Other infectious and parasitic diseases	0.0253	0.0211	0.0018	0.0015	0.0403	0.0391	0.0033	0.0031
Neoplasms	0.0186	0.0289	0.0302	0.0308	0.0139	0.0170	0.0347	0.0300
Cardiovascular renal diseases	0.0636	0.0622	0.0840	0.0650	0.0469	0.0529	0.0612	0.0564
Influenza, pneumonia, bronchitis	0.0370	0.0308	0.0078	0.0059	0.0800	0.0715	0.0154	0.0122
Diarrheal disorders	0.0141	0.0116	0.0012	0.0011	0.0640	0.0638	0.0044	0.0040
Certain degenerative diseases	0.0252	0.0261	0.0090	0.0080	0.0167	0.0141	0.0120	0.0080
Maternal causes	0.0000	0.0089	0.0000	0.0007	0.0000	0.0049	0.0000	0.0012
Certain diseases of infancy	0.0311	0.0239	0.0170	0.0125	0.0301	0.0259	0.0250	0.0199
Motor vehicle accidents	0.0120	0.0043	0.0129	0.0049	0.0000	0.0000	0.0118	0.0025
Other violence	0.0327	0.0116	0.0192	0.0077	0.0219	0.0071	0.0109	0.0042
Other and unknown causes	0.0482	0.0448	0.0232	0.0184	0.0541	0.0478	0.0245	0.0200
Total	0.3272	0.2932	0.2073	0.1571	0.3945	0.3726	0.2075	0.1631

Source: Calculated by Dr. Damiani of the University of Rome, from data in Preston, Keyfitz, and Schoen (1972).

Table 4.5. Effects of individual causes of death, United States, 1964; two methods compared

Comparison	Neoplasms	Cardiovascular renal diseases	Certain degenerative diseases	Motor vehicle accidents
Males: $\overset{o}{e}_0 = 66.905$				
Crude rate	0.00170	0.00572	0.00046	0.00036
Added years of life if eliminated: $\overset{o}{e}_0^{(-i)} - \overset{o}{e}_0$	2.265	13.299	0.627	0.874
$H^{(i)}$	0.0302	0.0840	0.0090	0.0129
$H^{(i)}\,\overset{o}{e}_0$	2.021	5.620	0.602	0.863
Females: $\overset{o}{e}_0 = 73.777$				
Crude rate	0.00139	0.00448	0.00037	0.00013
Added years of life if eliminated: $\overset{o}{e}_0^{(-i)} - \overset{o}{e}_0$	2.558	17.068	0.637	0.366
$H^{(i)}$	0.0308	0.0650	0.0080	0.0049
$H^{(i)}\,\overset{o}{e}_0$	2.272	4.796	0.590	0.362

Source: Preston, Keyfitz, and Schoen (1972, pp. 768–771) for $\overset{o}{e}_0^{(-i)} - \overset{o}{e}_0$; Table 4.4 for $H^{(i)}$.

No such statement can be made, however, about $\overset{o}{e}_0^{(-i)}$; the elimination of all causes would make the duration of life infinite, whereas the sum of the increases due to the elimination of 12 groups of causes would be only about 22 years for United States males in 1964. The total of $\overset{o}{e}_0^{(-i)} - \overset{o}{e}_0$ over all causes depends on what breakdown of causes is recognized, whereas $\sum H_i$ is invariant with respect to the grouping of causes.

4.3.6 *Interrelations of the Several Causes*

The quantity H can be thought of as a second parameter alongside $\overset{o}{e}_0$ of the curve of survivorship. It measures the convexity of the $l(x)$ function; and though in principle it could vary independently of $\overset{o}{e}_0$, among observed human populations it seems closely related to $\overset{o}{e}_0$. Extension of life is due mostly to mortality declines at younger ages, and much less to changes beyond age 70; therefore H tends to diminish over time. A similar feature of life tables arises by a mechanism described in the following piece of fantasy.

4.4 Everybody Dies Prematurely

A man of 65 is struck by a car while crossing the street and so deprived of the 13 years of expectation of life that he is credited with by the United States 1973 life table for males. We extend this notion, recognized in legal decisions and in common sense, and think of any death, whether from accident, heart disease, or cancer, as an "accident" that deprives the person involved of the remainder of his expectation of life. To calculate the number of years of which people are, on the average, deprived by virtue of the particular circumstances that caused their deaths, we suppose that everyone is saved from death once, but that thereafter he is unprotected and is, say, subject to the mortality shown for persons of his age in the United States 1973 (Krakowski 1972, Cohen 1973).

In life table notation $-dl(a)$ persons die between ages a and $a + da$ at last birthday. These are deprived of $\overset{o}{e}(a)$ years each. Hence the average deprivation is

$$\text{Dep} = -\frac{1}{l(x)}\int_x^{\omega} \overset{o}{e}(a)\,dl(a) = \frac{1}{l(x)}\int_x^{\omega}\int_a^{\omega} l(t)\mu(a)\,dt\,da. \qquad (4.4.1)$$

Consider the very special life table in which the chance of dying is constant at all ages, say μ. Then $l(a) = e^{-\mu a}$, and $\overset{o}{e}(a) = 1/\mu$, a constant for all ages. In this population the average prospective deprivation for those aged x is

$$\text{Dep} = -\frac{1}{e^{-\mu x}}\int_x^{\infty} \frac{1}{\mu}\,d(e^{-\mu a}) = \frac{1}{\mu}.$$

Since $\overset{o}{e}_0$ also equals $1/\mu$, this says that in the special case of equal death rates throughout life the deprivation involves a length of time equal to one's initial expectation of life. We can therefore write $\text{Dep}/\overset{o}{e}_0 = 1$, meaning that anyone who avoids death has a completely fresh start on this peculiar schedule of mortality.

At the opposite extreme, suppose that everyone dies at exact age ω. Then the expectation of life at ω is zero, and death deprives no one of any expectation. Thus the amount of deprivation depends on the shape of l_x. If $\mu(x) = \mu$, a constant, then $\text{Dep} = 1/\mu$ and $\text{Dep}/\overset{o}{e}_0 = 1$; at the other extreme, if everyone dies at the same age, $\text{Dep}/\overset{o}{e}_0 = 0$; any real life table would seem likely to fall between these two. At the extreme values $\text{Dep}/\overset{o}{e}_0$ is identical with H but is not quite the same in general. However, $\text{Dep}/\overset{o}{e}_0$ is identical with H for the hyperbolic life table function introduced in (4.3.9)–(4.3.12). [Prove that $H = \mu_0/(\mu_0 + 1) = \text{Dep}/\overset{o}{e}_0$ for this case.]

If $\mu_0 = 0.30$, as suggested above for females in a contemporary population, the deprivation is $0.30/1.30 = 0.23$ of the expectation of life at age zero. This is the fractional extra expectation if one could be excused her first death. For males it would be higher: $0.43/1.43 = 0.30$ on the same hypothetical table.

4.4.1 Average Expectation of Life

Similar features of the survivorship curve can be obtained from the average expectation of life in the stationary population.

$$E = \frac{\displaystyle\int_0^\infty l(a)\,\overset{o}{e}(a)\,da}{\displaystyle\int_0^\infty l(a)\,da}.$$

Again, with $\mu(a) = \mu$, a constant, this is equal to $\overset{o}{e}_0$; if everyone lives to age ω, it equals $\overset{o}{e}_0/2$; for intermediate degrees of convexity it varies between these numbers.

Evaluating E for they hyperbolic life table function (4.3.9) gives its ratio to $\overset{o}{e}_0$ as

$$\frac{E}{\overset{o}{e}_0} = \frac{\mu_0 + 1}{\mu_0 + 2}.$$

This contrasts with

$$\frac{\text{Dep}}{\overset{o}{e}_0} = H = \frac{\mu_0}{\mu_0 + 1}.$$

Thus the average expectation of life in the life table population is less sensitive to the shape of the $l(x)$ curve than is H.

The manifestations of convexity in the life table function have by no means been exhausted in the foregoing discussion. As mortality at younger ages declines in relation to that at older ages, the mean age in the stationary population rises. Additional indices could be obtained from this consideration and others. [Experiment with the joint expectation of two lives, $\int_0^\omega [l(a)]^2\,da$ as a ratio to $\overset{o}{e}_0$.]

4.4.2 Oldest Person in Group

Griffith Feeney presents a problem that shows another aspect of the convexity of the survivorship function. A certain tribe has the custom of making its oldest member king, and he remains king until his death. If the tribe is stationary and has B births each year, (a) what is the expected length of time that a given child just born will be king, calculated at the moment of his birth; (b) what is the probability that he will become king; and (c) what is the expected tenure at the time of appointment as king? The following solution is due to S. Krishnamoorthy and Noreen Goldman.

The probability that a given baby will become king is the same as the probability that he will be alive when all members of the tribe older than he are dead. Under the rule, those younger than he can be disregarded, since none can be king while he is alive. In a stationary population with

B births the number of persons aged x at any moment is $Bl(x)$, and the chance that all of these will be dead by time t is

$$\left[\frac{l(x) - l(x+t)}{l(x)}\right]^{Bl(x)} .$$

The chance for all ages x is the product of this over x, where we suppose deaths to be independent. The product can be made to mean something for the continuous case by taking its logarithm and integrating:

$$B \int_0^{\omega} l(x) \log \left[\frac{l(x) - l(x+t)}{l(x)}\right] dx.$$

The exponential of the above is the probability that all are dead by time t.

(a) The chance that a given baby is still alive at time and age t is $l(t)$. Hence the expectation at birth of his tenure as oldest person is

$$E = \int_0^{\omega} \exp \left\{ B \int_0^{\omega} l(x) \log \left[\frac{l(x) - l(x+t)}{l(x)}\right] dx \right\} l(t)\, dt.$$

(b) The probability, say P, of his attaining the post is the same with $\mu(t)$ in the outer integrand.

(c) The expected tenure once he is appointed is E/P.

To generalize the above from a stationary age distribution to an arbitrary age distribution in which $Bp(a)\,da$ persons are between ages a and $a + da$, B still being births, replace the first $l(x)$ in the expression for E by $p(x)$, and do the same for P.

4.4.3 Effect of a Health Improvement

Shepard and Zeckhauser (1975) emphasize that populations are heterogeneous in regard to mortality risks, and that the several causes of death are not independent. The usual ways of calculating the gain through a medical or safety improvement that disregard heterogeneity and dependence can grossly overstate the benefits to be attained. This applies whether the improvement be mobile coronary care units, prophylactic hysterectomies, seatbelts in automobiles, or the equipping of airplanes with sensing devices that warn the pilot when the plane comes too close to the ground or to an obstacle (see Chapter 19 for a discussion of heterogeneity).

Professionals have long been aware that to calculate lives saved is too simple, since a young man saved from death in a motor vehicle accident has a wholly different expectation thereafter from a man of 85 with a coronary attack saved by prehospital attention. In no case can death be prevented; at most it is deferred, and the question is for how long. One way of recognizing this is to measure the effect of a particular improvement by years added. For instance (Preston, Keyfitz, and Schoen 1972, p. 769 and Table 4.5), at rates for the United States in 1964 the elimination of cancer among males would have increased the expectation of life by 2.26 years, and of motor vehicle accidents by 0.87 year; although the number of deaths due to cancer was almost 5 times that from accidents, the years added are only in the ratio of 2.6 to 1. The values of $H^{(i)}$ are in the ratio 2.3 to 1.

Although this apparently sophisticated method takes full account of age, it can still exaggerate the benefit of a health improvement insofar as there is heterogeneity in the population in respects other than age. Populations are rarely homogeneous with regard to a risk. Some individuals are hereditarily subject to cancer; some travel more than others in airplanes. The special danger of exaggerating a benefit occurs if those subject to the risk that is diminished are subject also to greater than average risks on other accounts.

Thus allowance for age is not by itself adequate if there are other agelike differentials of mortality. Suppose that among people of a given age some are much healthier than others, and a particular improvement saves the lives of those who are in the worst general condition. Then disregard of this aspect would exaggerate the benefit of the improvement, in the same way that disregard of age would overstate for an improvement that affected old people.

5

Fixed Regime of Mortality and Fertility: The Uses of Stable Theory

A stable age distribution exists when age-specific birth and death rates have been constant over a considerable past period. The stable model is an advance in realism over the stationary population of the life table representing the special case of stability in which births are equal to deaths; although the stable model is restricted, the restriction will turn out to be acceptable for a number of purposes. Stable theory tells what age distribution is implied by a given and fixed regime of age-specific rates of birth and death; conversely, it permits in some instances inferring birth and death rates from an observed age distribution. It tells the cost of old-age pensions as a function of the rate of increase in population. For a given life table and rate of increase, average time of promotion in organizations with some degree of seniority can be calculated. These and other applications of stable theory are the subject of the present chapter. Chapter 15 will show how a regime of mortality and fertility implies kinship numbers. Stable theory can be generalized in various directions by modifying the assumption of a fixed regime, as will be seen in Chapter 14.

Lotka (1939, p. 18) applied the term *Malthusian* to a population with a given life table and an arbitrary rate of increase. He used the term *stable* for the case where the rate of increase is calculated from given and fixed age-specific birth rates, and in this sense stable populations are first treated in the present book in Chapter 6. Writers subsequent to Lotka, however, have not preserved this distinction.

5.1 Stable Theory

Suppose that the chance of living to age x is $l(x)$, and $l(x)$ is a function of age but not of time. A population whose births number B, uniformly spread through each year, where B does not change, and which is closed to migration will contain just $Bl(x)\,dx$ individuals between ages x and $x + dx$ at any given time, where $l(x)$ is normed to the radix $l_0 = 1$. It will contain $B_5L_x = B\int_0^5 l(x+t)\,dt$ individuals between exact ages x and $x+5$. This theory could be applied to both sexes together, using a survivorship function $l(x)$ applicable to whatever mix of men and women is taken to be present, but the usual practice is to consider one sex at a time.

The stationary population produced by this assumption of fixed annual births and a fixed life table for each sex is generalized by supposing births to follow the exponential Be^{rt}. It will turn out that even slow growth, say $r = 0.005$, affects the age distribution at any one time considerably. To recognize steady growth requires only a slight complication of the argument needed for stationarity and fits observed ages better.

Consider the female (or male) part of a large population closed to migration and subject to a fixed life table, with births increasing exponentially. These conditions are sufficient to produce a stable age distribution, in which the number of persons living in each age group, as well as the deaths in each age group and the total population, are all increasing exponentially in the same ratio.

If the probability of living to age x is $l(x)$, and the births at time t are B_0e^{rt}, then, to find the expected number of individuals between ages x and $x + dx$ we have to go back in time x to $x + dx$ years, when the number of births was $B_0e^{r(t-x)}\,dx$. The fraction of these births that survive to time t must be $l(x)$; therefore the absolute number of persons aged x to $x+dx$ at time t is

$$B_0e^{r(t-x)}l(x)\,dx.$$

The integral of this quantity is the total population at time t, and dividing by this total gives the fraction of the population aged x to $x + dx$ at time t, say, $c(x)\,dx$:

$$c(x)\,dx = \frac{e^{-rx}l(x)\,dx}{\displaystyle\int_0^{\omega} e^{-rx}l(x)\,dx} = be^{-rx}l(x)\,dx, \tag{5.1.1}$$

where B_0e^{rt} has been cancelled out from numerator and denominator, and b has been written for $1/\int_0^{\omega} e^{-rx}l(x)\,dx$.

The result is important enough to be worth deriving by an alternative, more intuitive means. Still supposing the fixed survivorship schedule $l(x)$, a birth rate of b per unit population existing now, and a current rate of increase r compounded each moment, it follows that the fraction of the births of x years ago that are now alive (and therefore aged x) must be

$l(x)$. But if the births are increasing in the ratio e^r each year, that is, at the annual rate r compounded each moment, then, the births now being in the ratio b to the population, the births x years ago must have been be^{-rx} *as a fraction of the present population.* And of those births the fraction $l(x)$ would be expected to be still alive. Then the expected number of persons alive now and aged x to $x + dx$ must be

$$be^{-rx}l(x)\,dx,$$

still reckoned per one of the present population. This compact derivation of the stable age distribution is essentially due to Euler (1760), and it has been rediscovered many times since.

Expression (5.1.1) provides the required age distribution $c(x)\,dx = be^{-rx}l(x)\,dx$ for the fraction of the population between ages x and $x + dx$. The argument rests on the supposition that the life table $l(x)$ is fixed in time for all x, and that the births and hence the population are growing exponentially, which is equivalent to assuming that birth as well as death rates have remained the same for a long time in the past. Because $c(x)\,dx$ is the fraction of the population aged x to $x + dx$, its total $\int_0^\omega c(x)\,dx$ must be unity; hence

$$\int_0^\omega be^{-rx}l(x)\,dx = 1.$$

This may be treated as an equation in b to obtain what was given in (5.1.1) as a definition:

$$b = \frac{1}{\displaystyle\int_0^\omega e^{-rx}l(x)\,dx}. \tag{5.1.2}$$

On (5.1.1) the "radix" or norm, $c(0)$, is b. It could, if we wished, be taken as $c(0) = 1$, as is done with the life table. If $c(0)$ is taken as 1, we replace the above argument by one calculating the present population aged x *per current birth*; it will be seen to be $e^{-rx}l(x)$, or a total at all ages of $\int_0^\omega e^{-rx}l(x)\,dx$; hence the birth rate must be $1/\int_0^\omega e^{-rx}l(x)\,dx$.

The several stable age distributions produced by a given life table are shown in Table 5.1, where the Coale and Demeny model West female life table with $\overset{o}{e}_0 = 65$ is combined with $r = -0.010$, 0, 0.010, 0.020, 0.030, 0.040. As one goes from columns at the left to columns at the right, the fall in the proportions becomes steeper and the values at the youngest ages higher; the proportion 5 to 9 years of age, for example, is 5.18 percent for $1000r = -10$ and 16.12 percent for $1000r = 40$. The fraction 65 and over falls from 19.46 to 2.25 percent. Note that the values for ages 20 to 40 first rise and then fall, and that the percent aged 15 to 44 goes from 37.30 to a maximum of 43.51 at $1000r = 20$ and then declines to 40.16. [Find a theoretical expression for the interval at which the proportion 15 to 44 at last birthday peaks.]

Table 5.1. Percentage age distributions $100\,{}_5C_x$ based on life table with $\overset{o}{e}_0 = 65$ and various values of $1000r$

Age	Values of $1000r$					
x	-10.00	0.00	10.00	20.00	30.00	40.00
0	5.01	7.26	9.98	13.11	16.49	20.02
5	5.18	7.14	9.35	11.67	13.97	16.12
10	5.42	7.10	8.84	10.50	11.95	13.12
15	5.66	7.05	8.35	9.44	10.22	10.67
20	5.89	6.98	7.87	8.46	8.71	8.65
25	6.12	6.90	7.40	7.56	7.41	7.00
30	6.34	6.80	6.94	6.74	6.28	5.65
35	6.55	6.69	6.49	6.00	5.32	4.55
40	6.75	6.55	6.04	5.32	4.48	3.65
45	6.91	6.38	5.60	4.68	3.76	2.91
50	6.99	6.15	5.13	4.08	3.12	2.29
55	6.97	5.83	4.63	3.50	2.54	1.78
60	6.76	5.38	4.06	2.92	2.02	1.34
65	6.27	4.74	3.40	2.33	1.53	0.97
70	5.37	3.87	2.64	1.72	1.08	0.65
75	4.06	2.78	1.81	1.12	0.66	0.38
80+	3.76	2.41	1.46	0.85	0.47	0.25
Total	100.00	100.00	100.00	100.00	100.00	100.00
Average age	42.20	36.96	31.97	27.46	23.54	20.25
Population $100\,{}_{30}C_{15}$	37.30	40.98	43.09	43.51	42.42	40.16
Ratio $\frac{{}_5C_0}{{}_{30}C_{15}}$	0.134	0.177	0.232	0.301	0.389	0.498
Ratio $\frac{{}_{10}C_5}{{}_\infty C_5}$	0.112	0.154	0.202	0.255	0.310	0.366
Dependency $\frac{{}_{15}C_0 + {}_\infty C_{65}}{{}_{50}C_{15}}$	0.719	0.685	0.711	0.793	0.929	1.121

Source: Coale and Demeny (1966), p. 62.

5.1.1 A Discrete Form

Although the stable age distribution is easier to think about in the continuous version, application requires a discrete form. We need to translate $c(x)\,dx = be^{-rx}l(x)\,dx$ into 5-year age groups to match the population data as usually provided. Integrating both sides of (5.1.1) between exact ages x to $x+5$ gives for ${}_5C_x = \int_0^5 c(x+t)\,dt$,

$$\begin{aligned} {}_5C_x &= b\int_0^5 e^{-r(x+t)}l(x+t)\,dt \\ &\approx be^{-r(x+2.5)}\int_0^5 e^{-r(t-2.5)}l(x+t)\,dt. \end{aligned} \tag{5.1.3}$$

The integral here is very close to $\int_0^5 l(x+t)\,dt$, tabulated in presentations of the life table as ${}_5L_x$. To this approximation

$${}_5C_x \approx be^{-r(x+2.5)}{}_5L_x. \tag{5.1.4}$$

[Work out an expression for the error of (5.1.4).]

Suppose that we are satisfied to integrate by putting a cubic through l_{x-5}, l_x, l_{x+5}, l_{x+10}, which bounds an area between x and $x+5$ equal to

$$ {}_5L_x = \tfrac{65}{24}(l_x + l_{x+5}) - \tfrac{5}{24}(l_{x-5} + l_{x+10}). \tag{5.1.5}$$

[Derive this expression.] In application to Mexico, 1970, females aged 60 to 64, with $l_0 = 100{,}000$,

$$ {}_5L_{60} = \tfrac{65}{24}(68{,}745 + 62{,}304) - \tfrac{5}{24}(73{,}836 + 53{,}058) = 328{,}488.$$

Taking $r = 0.03395$ gives

$$e^{-62.5r}\,{}_5L_{60} = 39{,}355$$

as the stable number per 100,000 current births.

A more precise way to evaluate (5.1.3) is to integrate, not the function $l(x)$, but the function $e^{-rx}l(x)$. We multiply l_{60} by e^{-60r}, l_{65} by e^{-65r}, and so on. Doing this first and then integrating by the same cubic formula (5.1.5) gives the stable number 39,449, as opposed to 39,355 from (5.1.4). The difference between the two methods is usually less than 1 percent.

5.2 Population Growth Estimated from One Census

Perhaps the most important fact bearing on the future of a population is its rate of natural increase, and yet for most countries of the world the obvious way of obtaining this rate—subtracting registered deaths from registered births—does not offer acceptable precision. The task of installing a modern registration system in developing countries is difficult, because people have little use for birth certificates, which become important only under modern conditions. Proof of date of birth and of citizenship is not required by immobile peasants, either for themselves or for their children, so these people do not respond dependably to a law from a distant capital requiring them to go to the town or village registrar each time a child is born. If complete vital statistics were the only possible source of information, the problem of rapid population increase might well be solved before it could be measured.

Hence indirect methods are called for. Extensive collections of possible procedures are given in Coale and Demeny (1967) and Brass (1975). The nature of one group of such methods is examined in this and the next four sections.

As in Section 2.3, each individual in the population is represented by a line in the Lexis plane of age and time, starting at the moment of birth on the time axis and proceeding at a 45° angle as simultaneously the person ages and time moves forward. A census provides a count by age that

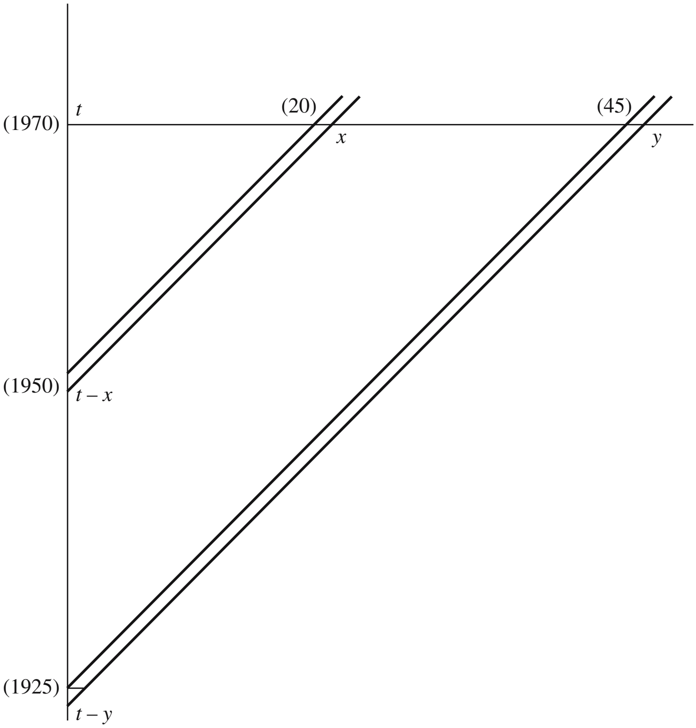


Figure 5.1. Relation of census count at ages x and y to birth cohorts x and y years earlier.

includes all life lines crossing a horizontal line at the date of the census. Figure 5.1 shows a census taken in 1970, and marks the intervals for individuals aged 20 and 45 at last birthday as examples. Each cohort born in a particular year is a band of 1 year: measured vertically 1 calendar year, measured horizontally 1 year of age. Our problem is, for example, to make observations of the number of survivors to ages 20 and 45 in 1970 tell us how fast births were increasing between 1925 and 1950. The density at any point can be portrayed as an altitude over the age–time plane; a census gives the profile on a time section.

Expressed in more general terms, the problem is as follows: given the number of individuals N_x at age x and N_y at age y from a census taken in the year t, find the rate at which the births were increasing between years $t - y$ and $t - x$, where y is greater than x. Call the birth function for the year preceding time t $B(t)$; if the population is closed to migration, we have

$$\begin{aligned} B(t-y)L_y &= N_y, \\ B(t-x)L_x &= N_x, \end{aligned} \tag{5.2.1}$$

where L_x and L_y are the fractions of each cohort that attain the census moment t in question; the equations are exact if L_x and L_y reflect the mortalities of the two cohorts, usually different.

Along with the exactitude of equations (5.2.1) goes their insolubility on the basis of the data available, which are only N_y and N_x from the census at time t. They contain two unknown survivorships, as well as the unknown function $B(t)$. To suppose that the survivorships follow the same life table, even without knowing what that table is, simplifies the problem considerably, for we can then disregard the first x years of life and be concerned only with the mortality between ages x and y. This is apparent on dividing the second member of the pair by the first, which leaves us with one equation containing only L_y/L_x:

$$\frac{B(t-y)}{B(t-x)}\left(\frac{L_y}{L_x}\right)=\frac{N_y}{N_x}. \tag{5.2.2}$$

But we must not forget in application that this disregard of survivorship at ages less than x is permissible only if the two cohorts have been subject to the same mortality up to that age, a point that will be reconsidered in Chapter 14.

We still have trouble; even if the life tables for the two cohorts are the same and are known, the most that (5.2.2) can do is to trace out the birth function, given N_x/L_x and N_y/L_y for the several combinations of x and y. We would like to use it to find a rate of increase—one number that holds through time. Let us take it that the birth function is an exponential, say $B(t)=B_0e^{rt}$. This is true in the circumstance that age-specific birth and death rates have been unchanged, not only over the interval between times $t-y$ and $t-x$, but also before $t-y$. (If the births were irregular before that date, so would be the age distribution, and this would affect the birth rate between $t-y$ and $t-x$, even with a fixed regime of birth and death rates in that interval.) Equation (5.2.2) now reduces to

$$e^{(y-x)r}=\frac{N_x/L_x}{N_y/L_y} \tag{5.2.3}$$

and contains only two unknowns, r and the survivorship L_y/L_x.

If the deaths for the population are known for past times, they can be made into a life table, which ideally would be a cohort life table applicable to the common mortality of the two cohorts in question. If the deaths are known for a current period, the current life table will be applicable to the cohorts if mortality has not been changing. If even current death data are lacking, there is no recourse but to select a life table from another source that will be somehow appropriate. This could be a table for a neighboring country with presumed similar mortality, but more often one resorts to a model or reference table. Let us suppose that by some means, however arbitrary, a value of L_y/L_x has been obtained. Then (5.2.3) is an equation

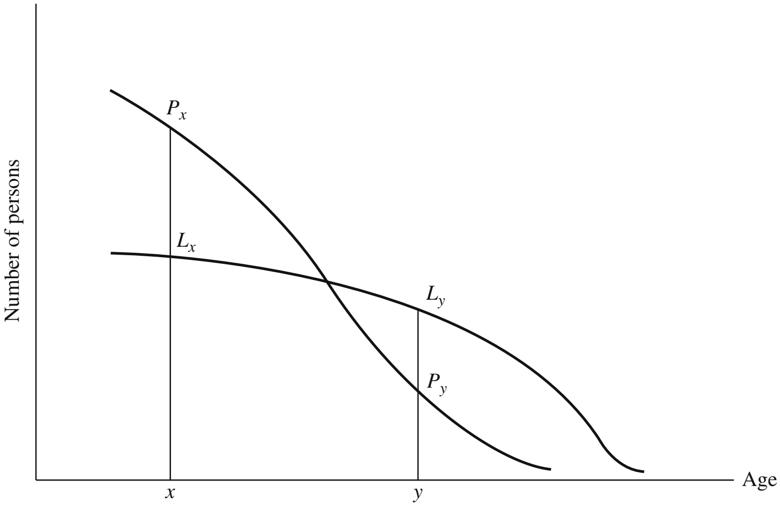


Figure 5.2. Given and stationary populations used to measure increase.

for r alone, which readily yields the solution

$$r = \frac{1}{y-x} \log \left(\frac{N_x/L_x}{N_y/L_y} \right) . \tag{5.2.4}$$

This is the simplest of all the ways of using a current age distribution to provide estimates of the rate of increase.

Figure 5.2, drawn for an increasing population, may help toward an intuitive understanding of (5.2.4). The life table curve is flat, and the stably increasing population less so; the formula in effect measures the relative steepness of the increasing population in terms of the steepness of the life table.

Since one value of the rate of increase r is obtained for each pair of ages, potentially thousands of values of r are available from one census. Each of these can be interpreted as an estimate of the (supposedly fixed) rate of increase, or more realistically as the rate of increase of births between the two cohorts.

The foregoing argument, whose several assertions may be followed on the Lexis diagram, has in effect been a listing of the assumptions required to bring us from the pair of equations (5.2.1), whose solution in general is hopeless, to (5.2.4), which constitutes an exact answer on the assumptions given. The database is a census that obtained present characteristics of persons, in particular their ages. From one point of view, each person enumerated in a census is asked to report on a vital event—his own birth—that took place a long time earlier. This is a special case of retrospective information. Unfortunately, knowledge of this retrospective event can genuinely be lost, say among a people who do not celebrate birthdays or otherwise keep their ages in mind. In such an instance N_x and N_y are irrecoverable.

Table 5.2. Values of r for Peru, 1963, computed for selected values of $\overset{o}{e}_0$ and the four model life tables of Coale and Demeny (1966), using ${}_5N_{25}$ and ${}_5N_{35}$ for females

$\overset{o}{e}_0$	West	North	East	South
22.5	0.01015	0.01229	0.01473	0.01482
32.5	0.01708	0.01836	0.01997	0.02028
42.5	0.02187	0.02264	0.02381	0.02415
52.5	0.02543	0.02580	0.02676	0.02721
62.5	0.02848	0.02842	0.02922	0.02958
72.5	0.03100	0.03055	0.03114	0.03125

5.2.1 Effect of Choice of Model Life Table

Let us try a small experiment to find the difference that the choice of life table makes for the rate of increase in births. For Peru, 1963, the female population included ${}_5N_{25} = 405{,}179$ women aged 25 to 29 at last birthday and ${}_5N_{35} = 293{,}494$ aged 35 to 39. For several life expectancies, considerably above and below the 62.5 years surmised, (5.2.4) was applied to these data, employing the four regional model life tables constructed by Coale and Demeny (1966). For life expectancy 62.5 years Table 5.2 shows the differences across the four life tables to be satisfactorily small, with the highest rate only about 4 percent above the lowest. The variation is greater for smaller $\overset{o}{e}_0$.

Within each of the columns, however, the computed r varies considerably among life expectancies. Assuming high mortality seems to cause L_y/L_x to absorb some of the r. In the West column the value for Peru at $\overset{o}{e}_0 = 62.5$ is 0.02848; a guess of $\overset{o}{e}_0 = 52.5$ would bring this down to 0.02543, which is 11 percent lower. Life tables with $\overset{o}{e}_0$ between 52.5 and 72.5 list r as between about 0.025 and 0.031.

Data drawn from the Togo, 1961, female population, in which there were 80,746 women aged 25 to 29 at last birthday and 51,975 aged 35 to 39, were used to examine the variation for a high-mortality population. Table 5.3 presents computations of r for each of the four regional life tables at life expectancy 40.0 years. Although rates steadily increase as one proceeds from West to South, the range of values is again small, less than 8 percent between the largest and the smallest.

More experimenting is needed to ascertain the sensitivity of the inferred r to the life table chosen.

Table 5.3. Values of r for Togo, 1961, computed for $\overset{o}{e}_0 = 40.0$ on the regional model life tables, using ${}_5N_{25}$ and ${}_5N_{35}$ for females

$\overset{o}{e}_0$	West	North	East	South
40.0	0.03259	0.03349	0.03475	0.03512

5.2.2 Theory for the Error Arising from Use of an Improper Life Table

Meanwhile the effect of a wrong life table can be calculated under the simplified condition that the table is wrong by a given amount, say δ, for the age-specific rates $\mu(a)$ at all ages: the true $\mu^*(a)$ equals the guessed $\mu(a)$ plus δ. If $\mu^*(a) = \mu(a) + \delta$, then $l^*(a) = l(a)e^{-\delta a}$, so that, in a slight rearrangement of (5.2.4),

$$\begin{aligned} r^* &= \frac{1}{y-x} \log \left[\frac{N_x/N_y}{L_x^*/L_y^*} \right] \approx \frac{1}{y-x} \log \left[\frac{N_x/N_y}{(L_x/L_y)e^{\delta(y-x)}} \right] \\ &= d\frac{1}{y-x} \left\{ \log \left[\frac{N_x/N_y}{L_x/L_y} \right] - \log[e^{\delta(y-x)}] \right\} = r - \delta. \end{aligned} \tag{5.2.5}$$

Thus taking a life table based on mortality rates too low by δ for both cohorts involved causes the estimate of increase to be too high by δ.

A more realistic alternative is to suppose the true $\mu^*(a) = \mu(a)(1+\delta)$, so that $l^*(a) = l(a)^{1+\delta}$. Then for the true rate of increase r^* in terms of r we have

$$\begin{aligned} r^* &= \frac{1}{y-x} \log \left(\frac{N_x/N_y}{L_x^*/L_y^*} \right) \approx \frac{1}{y-x} \log \left[\frac{N_x/N_y}{(L_x/L_y)^{1+\delta}} \right] \\ &= r - \frac{\delta \log(L_x/L_y)}{y-x}. \end{aligned} \tag{5.2.6}$$

Also, as was shown in Section 4.3, the effect of the same error on $\overset{o}{e}_0$ is, for δ small,

$$\overset{o}{e}{}_0^* = \overset{o}{e}_0(1 - H\delta),$$

where H is about 0.20 for modern populations. This gives

$$\delta = \frac{1 - (\overset{o}{e}{}_0^*/\overset{o}{e}_0)}{H}, \tag{5.2.7}$$

or, for $\overset{o}{e}{}_0^* = 65$, $\overset{o}{e}_0 = 60$ in the problem with $x = 25$, $y = 35$ of Table 5.3,

$$\delta = \frac{1 - (65/60)}{0.20} = -0.4;$$

with $\log(L_{25}/L_{35}) = 0.03$, r^* of (5.2.6) comes out to

$$r^* = r - \frac{(-0.4)(0.03)}{10} = r + 0.0012.$$

Therefore using a life table with $\overset{o}{e}_0$ 5 years too low produces a rate of increase that is 0.0012 too low, an error of $0.0012/0.03 = 4$ percent. This is verified by doing the calculation in the West series with $\overset{o}{e}_0 = 60$ and $\overset{o}{e}_0 = 65$; the values of r obtained differ by about 0.0012.

Stable theory can be applied with two censuses, which permit more satisfactory selection of the life table to be used. It combines with and is supplemented by other devices, of which extensive selections are given in Coale and Demeny (1967) and Brass (1975). The purposes of the foregoing treatment of the simplest method were to show how its assumptions can be clarified through use of the Lexis diagram and, on the other hand, to study in a preliminary way its degree of robustness when, as is usually the case, the assumptions do not apply.

5.3 Mean Age in the Stable Population

An increasing population must be on the average younger than a stationary population. This is due to children being born more recently than their elders, so that with constant birth rates the increasing population must have a larger proportion of children. If the current birth rate is b, the proportion of children under age α is from (5.1.1),

$${}_{\alpha}C_0 = \frac{\displaystyle\int_0^{\alpha} e^{-ra} l(a)\, da}{\displaystyle\int_0^{\omega} e^{-ra} l(a)\, da}.$$

(Here, the symbol C stands for the theoretical stable population, N for the observed population, though the distinction is not always easy to preserve in attempts to assimilate observations to theory.) To trace this out as a function of r for given α and $l(x)$ we note that

$$\frac{d({}_{\alpha}C_0)}{dr} = {}_{\alpha}C_0(m - m_1),$$

where m is the mean age of the entire (stationary) population, and m_1 the mean age of the part of it less than α years old. [Prove this formula.] Then expanding ${}_{\alpha}C_0$ as a function of r about $r = 0$ gives

$${}_{\alpha}C_0 \approx \frac{{}_{\alpha}L_0}{\overset{o}{e}_0}[1 + r(m - m_1)],$$

${}_{\alpha}L_0/\overset{o}{e}_0$ being the value of ${}_{\alpha}C_0$ when $r = 0$, that is, the proportion under α years of age in the life table. This linear first approximation is suitable for small values of r.

By how many years is the mean age of the increasing population younger than that of the stationary population with the same life table? We shall see that the difference in mean ages is approximately equal to the rate of increase of the growing population times the variance of ages of the stationary one. Average age, as observed or as in the stable model, will be designated by $\bar{x}$, and where $\bar{x}$ is low we will speak of the population as young.

From (5.1.1) it follows that the average age in the stable population is

$$\bar{x} = \int_0^{\omega} xc(x)\,dx = \frac{\displaystyle\int_0^{\omega} xe^{-rx}l(x)\,dx}{\displaystyle\int_0^{\omega} e^{-rx}l(x)\,dx};$$

and on expanding the exponentials and integrating the separate terms in both numerator and denominator, we find that

$$\bar{x} = \frac{L_1 - rL_2 + (r^2L_3/2!) - (r^3L_4/3!) + \cdots}{L_0 - rL_1 + (r^2L_2/2!) - (r^3L_3/3!) + \cdots}, \tag{5.3.1}$$

where $L_i = \int_0^{\omega} x^i l(x)\,dx$ is the numerator of the ith moment about zero of the stationary population of the life table. Ordinary long division on the right-hand side of (5.3.1) gives for $\bar{x}$ the series $L_1/L_0 - [L_2/L_0 - (L_1/L_0)^2]r + \cdots$, which may be written as

$$\bar{x} \approx \frac{L_1}{L_0} - \sigma^2 r \tag{5.3.2}$$

to the term in r, where L_1/L_0 is the mean and σ^2 the variance of the age distribution in the stationary population, both necessarily positive. Hence $\bar{x}$ is less than L_1/L_0 as long as r is greater than zero, supposing that the subsequent terms are small enough not to interfere.

An exact infinite series for $\bar{x}$ would be

$$\bar{x} = \frac{L_1}{L_0} - \sigma^2 r + \frac{\kappa_3 r^2}{2!} - \frac{\kappa_4 r^3}{3!} + \cdots, \tag{5.3.3}$$

where κ_3 is the third cumulant and κ_4 the fourth cumulant of the distribution with density $l(x)/L_0$ (Lotka 1939, p. 22). The third cumulant is the same as the third moment about the mean:

$$\kappa_3 = \mu_3 = \int_0^{\omega} \left(x - \frac{L_1}{L_0}\right)^3 \frac{l(x)\,dx}{L_0},$$

and is associated with skewness. The fourth cumulant κ_4 is the fourth moment about the mean less 3 times the square of the variance, and is a measure of kurtosis or peakedness, though its visual interpretation is not straightforward.

Let us identify the terms of (5.3.3) for Irish males of 1968, using data given in Keyfitz and Flieger (1971, p. 446). The cumulants of the life table are as follows:

$$\begin{aligned} L_1/L_0 &= 36.678 \\ \sigma^2 &= 495.59 \\ \kappa_3 &= 2243.9 \\ \kappa_4 &= -240{,}862 \\ r &= 0.01889. \end{aligned}$$

Taking four terms in (5.3.3) approximates $\bar{x}$ by

$$\bar{x} \approx 36.678 - 9.362 + 0.400 + 0.271 = 27.987.$$

This approximation to $\bar{x}$, the mean of the stable age distribution, 27.953, is excellent. But Irish births have fluctuated considerably, and the mean of the observed age distribution is 31.806, about 4 years higher than the $\bar{x}$ calculated above.

For Colombia females of 1964 (Keyfitz and Flieger 1968, p. 191), the first approximation (5.3.2) to the mean $\bar{x}$ of the stable age distribution is

$$\begin{aligned} \bar{x} = \frac{L_1}{L_0} - \sigma^2 r &= 37.360 - (539.04)(0.0283) \\ &= 37.36 - 15.25 = 22.11. \end{aligned}$$

In comparison, the stable age distribution shows a mean of 24.13 and the observed female age distribution a mean of 22.32. More instances would be needed to obtain a realistic estimate of the error of (5.3.2) as an approximation to the means of the stable and the observed age distributions, respectively. Apparently the error increases with mortality and with r.

A growing population tends to be younger than the corresponding stationary population also within any particular interval of ages. An argument identical to the one above may be applied to any age interval—we did not require the limits to be 0 to ω in deriving $\bar{x} \approx (L_1/L_0) - \sigma^2 r$. Thus children under 15 years of age are slightly younger, and people over 85 are younger, in the growing than in the corresponding stationary population. With the obvious redefinition of its constants, (5.3.2) is not restricted to the whole of life, but is applicable to any age interval.

Demographic Calculations Need Not Start at Age Zero. Various subgroups can be defined in such fashion that theory developed for whole populations applies to them. We have already dropped down to one sex and refer to the "female population," a useful abstraction but an abstraction nonetheless, since females cannot go through the process of birth and increase without males. The "population over age 15" could similarly be described as though it were self-contained. All of stable theory would apply with at most an alteration in the definitions of symbols. Now "births" would be the individuals passing through their fifteenth birthdays; survivorship would be not l_x/l_0, which we have been calling l_x, but rather l_x/l_{15}, $x > 15$; the stable fraction between x and $x + dx$ would be

$$c'(x)\,dx = b'e^{-rx}\frac{l(x)}{l_{15}}\,dx,$$

where $b' = l_{15}/\int_{15}^{\omega} e^{-rx}l(x)\,dx$, and evidently

$$c'(x)\,dx = \frac{e^{-rx}l(x)\,dx}{\displaystyle\int_{15}^{\omega} e^{-ra}l(a)\,da}.$$

By a further extension ages beyond 35, say, could be masked out and an expression developed for the stable population aged 15 to 34 at last birthday. It would similarly be feasible to recognize ages before birth, and even, where data on fetal mortality are available, to go back to conception.

5.3.1 Use of Population Mean Age

Relationship (5.3.2) may be applied in the opposite direction—to tell the rate of increase of a population when we know its mean age and the mean and variance of the life table applicable to it. This way of estimating the rate of increase of a population uses the data somewhat differently than does (5.2.4).

Transposing (5.3.2) into the equivalent

$$r \approx \frac{(L_1/L_0) - \bar{x}}{\sigma^2} \tag{5.3.4}$$

provides an approximation to the rate of population increase r, given the life table mean age L_1/L_0, the variance of the life table age distribution σ^2, and the observed mean age $\bar{x}$. Entering the figures for Colombia, 1964, already used, we find that

$$r \approx \frac{(L_1/L_0) - \bar{x}}{\sigma^2} = \frac{37.36 - 22.32}{539.04} = 0.0279,$$

as against the observed $r = 0.0283$.

When the additional parameters κ_3 (skewness) and κ_4 (kurtosis) are reliably enough available to be taken into account in estimating r, we can transform (5.3.3) into

$$r^* = \frac{(L_1/L_0) - \bar{x}}{\sigma^2 - (\kappa_3 r/2) + (\kappa_4 r^2/6)}. \tag{5.3.5}$$

In this iterative form we would start with an arbitrary r on the right-hand side, obtain the improved r^*, enter it on the right-hand side, obtain a further improvement, and so continue. The iteration is a way of solving the cubic equation in the unknown r represented by truncation of (5.3.3) after the term in r^3. Three or four cycles of iteration suffice for convergence to considerably more decimal places than are demographically meaningful.

Note that the method requires, not complete knowledge of the life table number-living column, but only its mean and other cumulants. In the absence of a life table we would guess the L_1/L_0 and σ^2 of (5.3.4), perhaps calculating them from a set of model tables. The less dependent the outcome on the life table chosen, the more useful is the method. We are encouraged to find that the mean ages of the life table $l(x)$ vary only about one-fifth as much as their $\overset{o}{e}_0$, the expectation of life at age zero. For instance, the Coale and Demeny (1966, p. 54) model West table for females with $\overset{o}{e}_0 = 55$ shows a mean L_1/L_0 of 34.95, and that with $\overset{o}{e}_0 = 60$ has $L_1/L_0 = 35.96$.

Countries and regions lacking birth data are likely to have inaccurate censuses, and we want to rely on censuses only at their strongest points. The art of this work is to take account of what ages are well enumerated and what ones poorly enumerated, as well as to use the model with the weakest assumptions. Literally thousands of ways of using the age distribution may be devised, and ideally one should choose the method least sensitive to (*a*) the accuracy of enumeration of ages; (*b*) the appropriateness of the life table, which often has to be selected arbitrarily; (*c*) the assumption of stability; and (*d*) possible in- and out-migration. To be able to choose from among a large stock of methods is an advantage in many instances, and the following sections continue our partial inventory of this stock.

5.4 Rate of Increase Estimated from the Fraction Under Age 25

Suppose that a population is underenumerated at ages 0 to 4 because infants are omitted, at 10 to 14 because some children of this age are entered as 5 to 9, and at 15 to 19 because young adults are mobile and the enumerator sometimes fails to find them. Suppose correspondingly that the numbers 5 to 9 and 20 to 24 are overstated (Coale and Demeny, 1967, p. 17), and that these errors offset to some degree those mentioned in the preceding sentence, so that the proportion under age 25, say α, is given correctly. We would like to use nothing but α from a census, along with a suitable life table, to estimate the rate of increase r.

Referring to the same stable age distribution (5.1.1), we can construct an equation in which both the observed α and the unknown r appear. The proportion of the population under age 25 is $\int_0^{25} be^{-ra}l(a)\,da$, where $\int_0^{\omega} be^{-ra}l(a)\,da = 1$. The ratio of the first of these integrals to the second can be equated to α and the birth rate b canceled out:

$$\alpha = \frac{\displaystyle\int_0^{25} e^{-ra} l(a)\, da}{\displaystyle\int_0^{\omega} e^{-ra} l(a)\, da}. \tag{5.4.1}$$

The problem now is to solve (5.4.1) for the unknown r, supposing the life table to be given.

One method is to leaf through a collection of model tables (e.g., Coale and Demeny 1966) designed for such purposes, and find among those for the given mortality the one with an α that matches the given α. The r of that stable population is the solution to (5.4.1).

An alternative is an iterative formula to solve (5.4.1) for r, given $l(a)$: multiply the numerator and denominator of the expression on the right-hand side of (5.4.1) by e^{10r}, then multiply both sides of (5.4.1) by e^{-10r}/α, next take logarithms of the reciprocals of both sides, and finally divide by 10 to obtain

$$r^* = \tfrac{1}{10} \log \left[\frac{\alpha \displaystyle\int_0^{\omega} e^{-r(a-20)} l(a)\, da}{\displaystyle\int_0^{25} e^{-r(a-10)} l(a)\, da} \right] \tag{5.4.2}$$

(Keyfitz and Flieger 1971, p. 28). This equation is algebraically identical to (5.4.1) except that the r on the left has been starred to suggest its use in iteration. Starting with an arbitrary r on the right, calculating r^*, entering this on the right in place of r, and repeating, is a process that ultimately converges to $r^* = r$, and whatever r satisfies this is also bound to satisfy (5.4.1). The multiplication of (5.4.1) by e^{-10r}/α (rather than e^{-15r}/α, say) is arbitrary, and experiment shows that e^{-10r} provides fast convergence. [Use the theory of functional iteration (e.g., Scarborough, 1958, p. 209) to find the optimum multiplier in place of e^{-10r}/α.] This is an example of how knowledge of the approximate magnitude of the quantities concerned enables demographers to devise iterative procedures that are simpler than more general ones such as the Newton–Raphson method.

To apply (5.4.2) we replace the integrals by finite expressions. The usual rough but serviceable approximation is in 5-year intervals. For example, the interval from $a = 0$ to $a = 5$ contributes to the integral in the numerator of the logarithm in (5.4.2) the amount

$$\begin{aligned} \int_0^5 e^{-r(a-20)} l(a)\, da &\approx e^{-r(2.5-20)} {}_5L_0 \\ &= e^{17.5r} {}_5L_0, \end{aligned}$$

and similarly for the other 5-year age intervals. Then the iterative formula becomes

$$r^* = \tfrac{1}{10} \log \frac{\alpha \Sigma_{i=0}^{\omega-5} e^{-r(i-17.5)} {}_5L_i}{\Sigma_{i=0}^{20} e^{-r(i-7.5)} {}_5L_i} \tag{5.4.3}$$

in terms of the ${}_5L_i$ tabulated in published life tables (except that most life tables are on radix $l_0 = 100{,}000$; that is, in terms of the symbols here used they tabulate $10^5 {}_5L_i$). Summations in (5.4.3) are over i's that are multiples of 5. With zero as the arbitrary initial value of r, (5.4.3) converges to six decimal places in from 3 to 10 cycles in most instances.

Method (5.4.3) is easily adapted to a split in the distribution at an age other than 25, and it can be generalized beyond this to other fractions, for example, the fraction of the population between ages 10 and 30, as indicated by examples in Section 5.6.

5.5 Birth Rate as Well as Rate of Increase Estimated for a Stable Population

We saw in (5.1.1) that in a closed population where birth and death rates have been constant for a long time the proportion of individuals between ages x and $x+dx$ is $c(x)\,dx = be^{-rx}l(x)\,dx$. Similarly the proportion of the population between ages y and $y+dy$ is $c(y)\,dy = be^{-ry}l(y)\,dy$. To adapt to 5-year age intervals we integrated each equation over 5 years to construct

$$\begin{aligned} {}_5C_x &= be^{-r(x+2.5)} {}_5L_x \\ {}_5C_y &= be^{-r(y+2.5)} {}_5L_y. \end{aligned} \tag{5.5.1}$$

Dividing and taking logarithms, we obtained in Section 5.1

$$r = \frac{1}{y-x} \log \left(\frac{{}_5C_x / {}_5L_x}{{}_5C_y / {}_5L_y} \right). \tag{5.5.2}$$

Now we go on to find the value of b by eliminating r from the pair (5.5.1):

$$b = \frac{\left(\dfrac{{}_5C_x}{{}_5L_x} \right)^{(y+2.5)/(y-x)}}{\left(\dfrac{{}_5C_y}{{}_5L_y} \right)^{(x+2.5)/(y-x)}}. \tag{5.5.3}$$

We derived these formulae by defining ${}_5C_x$ as an integral over a range of ages in the stable population; we would apply the formulae by entering for ${}_5C_x$ the observed fraction, from a census or estimate, over that range.

An alternative form with 10-year intervals is

$$b = \exp\left[\frac{\dfrac{\log({}_{10}C_x/{}_{10}L_x)}{x+5} - \dfrac{\log({}_{10}C_y/{}_{10}L_y)}{y+5}}{\dfrac{1}{x+5} - \dfrac{1}{y+5}}\right].$$

Entering the numbers for Colombian females, 1965, in thousands, for ages 5 to 15 and 20 to 30 (Keyfitz and Flieger 1971, p. 366, reproduced in Table 5.6 below) gives

$$b = \exp\left\{\frac{\dfrac{\log[(2575/9125)/8.753]}{10} - \dfrac{\log[(1401/9125)/8.550]}{25}}{\frac{1}{10} - \frac{1}{25}}\right\} = 0.0476,$$

or 47.6 births per thousand population.

There is no need to confine the calculation to two ages (Bourgeois-Pichat, 1958). Taking logarithms in (5.5.1) provides a linear equation in r and $\log b$ for each age:

$$\log b - (x + 2\tfrac{1}{2})r = \log\left(\frac{{}_5C_x}{{}_5L_x}\right). \tag{5.5.4}$$

Again in application we replace the stable proportions ${}_5C_x$ by the observed proportions P. For Colombian females, 1965, the needed quantities in the six age intervals in the range 0 to 29 are shown in Table 5.4. The least square line relating $x + 2\frac{1}{2}$ and $\log({}_5P_x/{}_5L_x)$ is

$$-3.126 - (x + 2\tfrac{1}{2})(0.0352) = \log\left(\frac{{}_5P_x}{{}_5L_x}\right),$$

and this identifies $\log b$ with -3.126 and r with 0.0352; hence b is estimated at $\exp(-3.126) = 0.0439$, r at 0.0352, and d, the death rate, at the difference, 0.0087.

The ages to be used are at our disposal; which should we choose? With only two age intervals x and y, as in (5.5.2), if x and y are equal, neither (5.5.2) nor (5.5.3) tells us anything, and if they are close to each other the answers will be sensitive to random errors in the population count identified with ${}_5C_x$ and ${}_5C_y$. On the other hand, choosing x and y far apart increases the risk of straddling a substantial change in the birth rates that determine the cohort sizes, that is, a departure from stability. If the probability of a change in birth rates between x and y years ago is proportional to $y - x$, the length of the interval between them, then for given random errors in ages there will be an optimum choice of the interval $y - x$, perhaps in the neighborhood of 25 years.

Suppose a single change in birth rates in the past, at a known or suspected date D years ago. Making both x and y less than D, or both x and y greater than D, is clearly advisable. If both x and y are less than D,

Table 5.4. Data for estimating b and r for Colombia, females, 1965

$x + 2\frac{1}{2}$	Observed $_5P_x$	Life table $_5L_x$	$\log\left(\frac{_5P_x}{_5L_x}\right)$
$2\frac{1}{2}$	0.1718	4.54231	−3.275
$7\frac{1}{2}$	0.1558	4.39360	−3.339
$12\frac{1}{2}$	0.1263	4.35979	−3.542
$17\frac{1}{2}$	0.1048	4.33516	−3.722
$22\frac{1}{2}$	0.0841	4.29919	−3.934
$27\frac{1}{2}$	0.0695	4.25043	−4.113

Source: Keyfitz and Flieger (1971, p. 366).

the more recent rate of increase will be estimated; if they are greater, the earlier rate.

If nothing is known about shifts in the past birth rate, and we want an average of all rates that have prevailed during the lifetimes of persons now alive, we might calculate all possible values of r and average them. The average would be best weighted, with a maximum weight for $y - x = 25$ years or thereabouts, and with lower weights as we moved toward $y-x = 0$ or $y - x = 50$. The population above 70 provides little information on rate of increase. Beyond 70 ages are inaccurately reported; the life table is less precisely known; the survivors have passed through a variety of mortality conditions; and their number is often too small to disregard sampling error as we have done throughout.

5.6 Comparison of the Several Ways of Using the Age Distribution

To summarize and extend the methods described above, still supposing stability, we extract from an observed age distribution the quantity $_mN_x/_nN_y$, that is, the ratio of the observed persons x to $x + m$ years of age to those y to $y+n$. This ratio may be equated to the corresponding quantity estimated from a life table and the rate of increase, which is approximately

$$\frac{e^{-(x+m/2)r}/e^{-(y+n/2)r}}{_nL_y/_mL_x},$$

and the equation provides the value for r:

$$r = \frac{1}{y + (n/2) - x - (m/2)} \log\left[\frac{_mN_x/_mL_x}{_nN_y/_nL_y}\right]. \tag{5.6.1}$$

Table 5.5. Rate of natural increase as estimated from pairs of age ranges, Colombia, males, 1965*

x to $(x+m-1)$	All ages	y to $(y+n-1)$				
		5–ω	10–ω	0–69	5–69	10–69
0–14	0.03547	0.03486	0.03567	0.03569	0.03502	0.03589
0–24	0.03286	0.03294	0.03434	0.03290	0.03298	0.03452
0–34	0.03256	0.03278	0.03450	0.03256	0.03281	0.03473
0–44	0.03236	0.03274	0.03482	0.03230	0.03278	0.03512
5–19	0.03513	0.03439	0.03558	0.03548	0.03457	0.03587
5–29	0.03159	0.03231	0.03446	0.03135	0.03229	0.03472
5–39	0.03088	0.03245	0.03518	0.02939	0.03244	0.03563
5–49	0.03289	0.03318	0.03641	0.03270	0.03348	0.03723

*Programmed by Geoffrey McNicoll.

Where m and n are more than 5 years this approximation will not serve and an iterative form is needed. Such a form was used to calculate the entries in Table 5.5 (Keyfitz and Flieger 1971).

A special case is pairs of age intervals that approximately represent ratios of female children to mothers, where now the y group is 30 years wide. In Table 5.6, ${}_5N_0/{}_{30}N_{15}$ gives $r = 0.03066$. Again the estimates are mostly in the range 3.0 to 3.5 percent, and all are greater than the 0.02814 provided by registrations. The method has been named after William R. Thompson, one of the pioneers of American demography.

Using age intervals x to $x+4$ and $x+15$ to $x+44$, so that the individuals in the first group are mostly the children of those in the second group, protects in some degree against errors due to variation in the rate of increase of the successive birth cohorts; that is, it protects against the inappropriateness of the stability assumption. To see this we note that Thompson's index,

$$\mathrm{Th} = \frac{{}_5N_x/{}_5L_x}{{}_{30}N_{x+15}/{}_{30}L_{x+15}}, \tag{5.6.2}$$

which is in this special case the contents of the square brackets in (5.6.1), is an approximation to $e^{27.5r}$. In fact $r = (1/27.5)\log \mathrm{Th}$ is one way of writing (5.6.1) for $m = 5$, $n = 30$. In terms of the 5-year age groups of the mother generation, (5.6.2) breaks down to

$$\mathrm{Th} = \frac{\dfrac{({}_5N_{x+15}F_{x+15} + {}_5N_{x+20}F_{x+20} + \cdots)}{({}_5N_{x+15} + {}_5N_{x+20} + \cdots)}}{\dfrac{{}_5L_x}{{}_{30}L_{x+15}}}, \tag{5.6.3}$$

where F_{x+15} and so on are the age-specific rates of childbearing, in 5-year intervals, counting for this purpose only children that live to the end of the 5-year period. In effect the N's of (5.6.3) are weights on the F's.

Table 5.6. Estimates of rates of increase for Colombia, 1965, females* Part (a)

x	${}_5N_x$	${}_5L_x$
0	1,567,251	454,231
5	1,421,920	439,360
10	1,152,736	435,979
15	956,462	433,516
20	767,534	429,919
25	633,851	425,043
30	545,307	419,014
35	495,077	411,199
40	369,047	401,498
45	309,618	389,415
50	263,402	374,099
55	169,199	351,971
60	181,209	323,465
65	100,134	281,903
70	83,009	230,533
75	45,586	173,813
80	34,442	116,072
85+	28,852	106,954
Total	9,124,636	

1. Local stability (5.5.2) with 5-year age intervals

Age range		r
5–9,	15–19	0.03831
10–14,	20–24	0.03927
15–19,	25–29	0.03917
20–24,	30–34	0.03161

2. Bourgeois-Pichat regression (5.5.4) based on 30-year age intervals

Age range	b	r
0–29	0.04388	0.03520
5–34	0.04616	0.03706
10–39	0.04210	0.03255
15–44	0.04153	0.03255
20–49	0.04024	0.03185
25–54	0.03973	0.03143
30–59	0.05287	0.03848
35–64	0.04377	0.03367
40–69	0.04636	0.03535

Table continues in part (b)

Table 5.6. (contd.) Estimates of rates of increase for Colombia, 1965, females*
Part (b)

3. Rate of increase from Thompson's Index (5.6.2) by iterative method

Age range		r
0–4,	15–44	0.03066
5–9,	20–49	0.03451
10–14,	25–54	0.03271
15–19,	30–59	0.03147

4. Rate of increase from pairs of wide age ranges using iterated version of (5.6.1)

Age range	All ages	5+	10+	0–69	5–69	10–69
0–14	0.03415	0.03345	0.03395	0.03430	0.03351	0.03407
0–24	0.03330	0.03269	0.03346	0.03340	0.03269	0.03356
0–34	0.03378	0.03281	0.03370	0.03403	0.03284	0.03385
0–44	0.03356	0.03243	0.03362	0.03392	0.03240	0.03378
5–19	0.03615	0.03441	0.03484	0.03675	0.03462	0.03508
5–29	0.03557	0.03347	0.03431	0.03662	0.03361	0.03455
5–39	†	0.03390	0.03480	†	0.03424	0.03521
5–49	†	0.03423	0.03524	†	0.03508	0.03587
10–24	0.03190	0.03157	0.03312	0.03163	0.03136	0.03323
10–34	†	0.03164	0.03382	0.03084	0.03125	0.03414
10–44	0.03385	0.02770	0.03353	0.03349	0.02204	0.03398
10–54	0.03438	0.04228	0.03237	0.03406	0.03873	0.03243
15–29	0.04068	0.02606	0.03105	0.03883	0.02327	0.03073
15–39	0.03504	†	0.03161	0.03461	0.03994	0.03116
15–49	0.03413	0.03719	0.02928	0.03393	0.03618	0.03561
15–59	0.03411	0.03613	0.03622	0.03395	0.03555	0.03454

*Programmed by Geoffrey McNicoll.

†Indeterminate.

Source: Keyfitz and Flieger (1971, p. 366).

Under stability ${}_5N_{x+20}$ will be $e^{-5r}({}_5L_{x+20}/{}_5L_{x+15})$ times as numerous as ${}_5N_{x+15}$; if it is not, the numerator and denominator of the top section of (5.6.3) are affected in the same direction.

In Table 5.6 the various methods described above are shown in application to Colombian females. No single number is trustworthy, but the collection shows a considerable degree of agreement, especially where wide intervals are involved. The agreement with results for males as given in Table 5.5 is also worth noting. [Show that (5.6.1) with the data in Table 5.6 gives 0.03042 for the first entry under Thompson, and devise an iterative form that will give the more accurate 0.03066.]

Fortunately these formulae are most applicable to populations of rapid and fairly steady increase, which happen to be just the ones least provided with usable vital statistics, though they usually have recent censuses.

Further work remains to be done on robustness, however, in the face of inaccurate data and inappropriate assumptions.

5.6.1 Incomplete Population and Deaths

William Brass has suggested (personal communication) that one can make use of an incomplete census, along with incomplete death statistics, to ascertain r, the rate of increase of a population, provided only that the omissions at the several ages are nearly constant and the population is more or less stable. Although such assumptions may seem strong, the calculation itself checks them. Several values of r may be produced by the calculation; and if they are not in agreement, we know that the assumptions are not met and reject the method for the case in question.

The Brass suggestion requires an expression for $c(x)$, the fraction of persons at age x, in terms of the population and deaths cumulated from age x to the end of life, ${}_\infty N_x$ and ${}_\infty D_x$, respectively. An integration by parts in the expression for deaths over age x as a fraction of total population will establish the required relation. The deaths from age x to the end of life as a fraction of total population are as follows:

$$\begin{aligned}\frac{{}_\infty D_x}{N} &= \int_x^\omega be^{-ra}l(a)\mu(a)\,da\\ &= -b\int_x^\omega e^{-ra}\,dl(a)\\ &= -b\left[e^{-ra}l(a)|_x^\omega + r\int_x^\omega e^{-ra}l(a)\,da\right]\\ &= be^{-rx}l(x) - \frac{r{}_\infty N_x}{N}\\ &= c(x) - \frac{r{}_\infty N_x}{N}.\end{aligned}$$

Transposing gives the required relation of $c(x)$ to the cumulative population and deaths:

$$c(x) = \frac{r{}_\infty N_x}{N} + \frac{{}_\infty D_x}{N}. \qquad (5.6.4)$$

This is interpretable as a flow equation, in which the entrants $c(x)$ into the age group x and over are equal to the net natural increase $r{}_\infty N_x/N$ of that age group plus the deaths ${}_\infty D_x/N$, all in terms of ratios to the total population.

Equation (5.6.4) may be used to calculate r, given the age distribution and the deaths, but it involves a gratuitous error insofar as the deaths suffer from more (or less) incompleteness than the population. Brass avoids this by multiplying and dividing the term on the right, ${}_\infty D_x/N$, by $d = D/N$, the death rate, and then treating d as unknown:

$$c(x) = \frac{r_\infty N_x}{N} + \frac{d_\infty D_x}{D}. \tag{5.6.5}$$

Now the fact that the population and deaths are incomplete by some unknown fraction at all ages makes no difference. Any two ages, say x and y, provide a pair of equations solvable for r and d. With more than two ages one can fit by least squares, and use the magnitude of departures from the fitted r and d to check the assumptions of stability and of constant proportional completeness in population and deaths.

The logic of populations permits modification of this without further algebra—for instance, if the population and deaths under age 5 are thought to be especially uncertain; in that case one would reinterpret all the elements of (5.6.5) on the basis of the population and deaths at 5 years of age and over: $c(x)$ would be the ratio of those aged x to the population 5 and over, D would be the deaths for ages 5 and over, and N would be the population aged 5 and over. To find the rate of increase of a long past time one might consider only the part of the population 60 and over, interpret $c(x)$ as the ratio of those aged x to the number aged 60 and over, and D and N as deaths and population, respectively, over 60, and put x equal to 60 and to 70 to obtain two equations.

Taking Mexico, 1966, females as an example, we have as the equation for $x = 0$, using all ages,

$$\frac{900}{22{,}029} = r + d,$$

and, for $x = 60$

$$\frac{81.8}{22{,}029} = r\frac{1161}{22{,}029} + d\frac{58{,}923}{198{,}893}.$$

Solving these together gives $r = 0.0345$, $d = 0.0064$, and $b = d{+}r = 0.0409$.

Now suppose that we disregard the first 60 years of life, and use $x = 60$; then

$$\frac{81.8}{1161} = r + d,$$

and, for $x = 70$, still considering population 60 and over,

$$\frac{43.4}{1161} = r\frac{448}{1161} + d\frac{41{,}148}{58{,}923},$$

which solve to $r = 0.0378$ and $d = 0.0326$. In this formulation the death rate d does not have the usual meaning, since its denominator is population 60 and over, but r is an estimate of the usual rate of increase.

5.6.2 Estimates from Two Censuses

When two censuses, say 10 years apart, are available, much more can be done; not only can assumptions be weakened, but also increased redundancy permits a better estimate of error.

If the population is closed, and whether or not it is stable, survivorship rates can be found by comparing the number in a cohort as given by the two censuses. For instance, the chance of surviving 10 years for those aged 10 to 14 is the ratio of the number 20 to 24 at the later census to that 10 to 14 at the earlier one; if ${}_5N_{10}^{(t)}$ is the population aged 10 to 14 at time t, then

$$\frac{{}_5L_{20}}{{}_5L_{10}} = \frac{{}_5N_{20}^{(t+10)}}{{}_5N_{10}^{(t)}}.$$

Having a set of survivorships for pairs of ages 10 years apart, one could interpolate to construct the life table. Unfortunately the countries less developed statistically, for which reliable mortality registrations are lacking, are not likely to have accurate censuses, and inaccuracies in the census reporting of ages produce a very irregular table. Therefore we must seek a way to use the survivorships that is less sensitive to their irregularities.

One such way is to match the survivorships to model life tables. Each of the 10-year survivorships would permit picking a member of the series of life tables. With perfect data and a series of model tables appropriate to the underlying mortality of a closed population, the identical life table would be picked by the several ages. In practice, however, different life tables will be picked, and if they do not differ too greatly they can be averaged.

Once the life table is available, chosen by the observations rather than imported more or less arbitrarily as with a single census, one can use either census for estimating r by the methods outlined above. Again the degree of agreement among the several estimates of r will inform us as to how far to trust the results.

A further and obvious resource for calculating r is the intercensus increase. In some cases the accuracy of enumeration is inadequate to provide survivorships between the censuses for the several cohorts, but is sufficient to show how fast the total population is increasing; r would be obtained from the ratio of census totals as $r = \frac{1}{10}\log\left[N^{(t+10)}/N^{(t)}\right]$. The value of r so obtained, taken in conjunction with either age distribution, can provide a life table. Arriaga (1968) used an observed age distribution along with a value of r in his study of mortality in Latin America; from (5.1.1) he obtained $l(x)$ as $e^{rx}c(x)/c(0)$.

5.7 Sensitivity Analysis

The stable model was used in Sections 5.3 to 5.6 to estimate rates of increase of populations lacking in vital statistics. It can also provide answers to substantive questions concerned, not with data, but with the way populations—or at least their parameters—behave.

5.7.1 Mean Age as a Function of Rate of Increase

We have prepared the way for one example with (5.3.2), showing how the mean age of a population is related to its rate of increase:

$$\bar{x} \approx \frac{L_1}{L_0} - \sigma^2 r.$$

Insofar as this applies, it tells what a change in r will do to the population mean age $\bar{x}$. If in our population r were higher by Δr, then $\bar{x}$ would be lower by $\sigma^2\,\Delta r$, on the understanding that the life table remained unchanged. Essentially the same fact can be expressed by the derivative:

$$\frac{d\bar{x}}{dr} \approx -\sigma^2, \tag{5.7.1}$$

and derivatives are useful in cases where the device by which (5.3.2) was established is not applicable.

For Canada, 1968, σ^2, the variance of women's ages in the life table population, was 573.7. Hence we can say from (5.7.1) that a fall of 0.010 in r would raise the mean age by 5.737 years, the life table remaining unchanged. This touches genuine causation, although it does not say what would happen in the real world, where many other changes would occur at the same time. Like any other causal law, it is capable only of telling what will happen conditionally, provided that its action is not overlaid, or reversed altogether, by other changes such as migration.

One added qualification is that the effect does not immediately follow the cause in time. In this section and others presenting sensitivity analysis it is two ultimate stable conditions that are compared, a device known in economics as *comparative statics.*

5.7.2 Pension Cost

Consider a pension scheme in which those over exact age 65 are paid a pension equal to salary, and salary is unity and the same for all individuals and at all ages between 20 and 65. If N is the number of persons of all ages in the population, the total payments by the fund each year will be $N\int_{65}^{\omega} be^{-rx}l(x)\,dx$, and the total receipts of the fund each year will be $Ng\int_{20}^{65} be^{-rx}l(x)\,dx$, where g is the premium as a fraction of salary. For the fund to remain in balance these two quantities must be equal, so we have for the premium

$$g = \frac{\displaystyle\int_{65}^{\omega} e^{-rx} l(x)\,dx}{\displaystyle\int_{20}^{65} e^{-rx} l(x)\,dx}, \tag{5.7.2}$$

still as a fraction of the uniform annual salary of individuals aged 20 to 65. Adapting this to salaries and pensions that vary with age, to salaries that vary at each given age, to cases where increasing numbers of persons are unable to work as they become older, and to other practical circumstances complicates the formula somewhat but entails no new principle.

A formula enables us to see how demographic conditions affect pension arrangements. For the scheme whose premium is given by (5.7.2) we might ask what difference it makes if mortality improves beyond age 65, if the age at pension is 60 rather than 65, and if the population increases more slowly or becomes stationary. The stable model permits pencil-and-paper experiments to answer such questions; let us consider the last one briefly.

To see what difference it makes if r changes we calculate dg/dr from (5.7.2). Taking logarithms and using the fact that

$$\frac{d \log g}{dr} = \frac{1}{g}\frac{dg}{dr}$$

gives immediately, with fixed life table,

$$\frac{d \log g}{dr} = -\frac{\displaystyle\int_{65}^{\omega} x e^{-rx} l(x)\,dx}{\displaystyle\int_{65}^{\omega} e^{-rx} l(x)\,dx} + \frac{\displaystyle\int_{20}^{65} x e^{-rx} l(x)\,dx}{\displaystyle\int_{20}^{65} e^{-rx} l(x)\,dx}.$$

The first term on the right is minus the mean age of the pensioners, say M and the second is the mean age of the contributors, say m. Hence

$$\frac{1}{g}\frac{dg}{dr} = -(M - m). \tag{5.7.3}$$

According to (5.7.3) a small finite change Δr in r ultimately causes a relative change $\Delta g/g$ in the premium, i.e.,

$$\frac{\Delta g}{g} \approx -(M - m)\,\Delta r. \tag{5.7.4}$$

In words, the fractional change in the premium equals minus the difference of mean ages in the pensioned and working groups times the absolute change in rate of increase. If the mean M is about 70 years and m about 40 years, $M - m$ is about 30, and we conclude that a decrease in r by 1 percentage point, say from 0.020 to 0.010, increases the fraction of salary to be paid as premium by $-(30)(-0.01) = 0.30$. Practically without data we have found that a 30 percent increase in premiums for a contributory nonreserve pension scheme is associated with a fall in the population growth rate of 0.01. We could sharpen the 30 percent by using observed numbers for $M - m$, but 30 years is close enough to suggest how much such financial calculations depend on the rate of increase of populations. As Thomas Espenshade pointed out to me, the lower birth rate as r becomes smaller releases women into the labor force and so, if jobs are available, helps to pay the premium for the increased fraction of aged persons. With some additional data the effect of this factor and others may be calculated.

5.7.3 Fraction of Old People

The objective of the preceding pages can be attained more precisely by employing derivatives and then treating the Taylor expansion as a differential equation. Suppose we would like to know how the fraction of the population that is old depends on the rate of increase, and assume that death rates do not change. The fraction $f(r)$ over age 65 is

$$f(r) = \frac{\displaystyle\int_{65}^{\omega} e^{-ra} l(a)\, da}{\displaystyle\int_{0}^{\omega} e^{-ra} l(a)\, da}. \tag{5.7.5}$$

Taking logarithms and differentiating both sides gives

$$\frac{1}{f(r)} \frac{df(r)}{dr} = \kappa_1 - k_1, \tag{5.7.6}$$

where κ_1 is the mean age of the entire population and k_1 of the part of it 65 years of age and over. Integrating (5.7.6) and taking exponentials gives

$$f(r) = f_0 e^{(\kappa_1 - k_1)r},$$

where f_0 is the life table proportion 65 and over.

The result is obtainable from the term in r of the difference of two cumulant generating functions, one for the distribution of those 65 and over, the other for all ages:

$$\begin{aligned} \log f(r) = \log f_0 &+ \left(-rk_1 + \frac{r^2}{2!}k_2 - \frac{r^3}{3!}k_3 + \cdots \right) \\ &- \left(-r\kappa_1 + \frac{r^2}{2!}\kappa_2 - \frac{r^3}{3!}\kappa_3 + \cdots \right), \end{aligned}$$

where the k's are the cumulants of the persons 65 and over and the κ's of everyone. Taking exponentials up to the term in r^2 gives

$$f(r) = f_0 \exp\left[r(\kappa_1 - k_1) - \frac{r^2}{2!}(\kappa_2 - k_2) \right]. \tag{5.7.7}$$

Table 5.7. First and second approximations to the percent over 65 years of age from (5.7.7), along with the total as taken from Table 5.1

	Approximation from (4.7.7)		As totaled from
r	To term in r	To term in r^2	Table 4.1
−0.01	19.9	19.6	19.46
0	13.8	13.8	13.80
0.01	9.5	9.3	9.31
0.02	6.5	6.1	6.02
0.03	4.5	3.8	3.74
0.04	3.1	2.2	2.25

Experimenting with Table 5.1, and using round numbers for the means $k_1 = 72$, $\kappa_1 = 35$, so that $\kappa_1 - k_1 = -37$, and for the variances $k_2 = 30$, $\kappa_2 = 450$, so that $\kappa_2 - k_2 = 420$, we have the fraction 65 and over as $0.138e^{-37r}$ to a first approximation, and as

$$f(r) = 0.138e^{-37r-210r^2}$$

to a second approximation. Table 5.7 shows that the second approximation at least is very close. Such discrepancies as exist between the last column and the preceding column are due to the crude estimates of the first and second cumulants and to the omission of later cumulants.

It bears repeating that (5.7.7) gives only the ultimate difference between the original situation and the one with incremented r; it says nothing about the path by which the second is reached from the first. When the birth rate declines with unchanged $l(x)$, it requires more than one generation under the new conditions for the age distribution to stabilize. The projection techniques of Chapter 12 are capable of showing the actual path of change in a quantity such as our premium g when birth rates change in any specified manner. The results are inevitably more complicated than those of stable theory.

5.8 The Degree to Which Promotion Within Organizations Depends on Population Increase

In a stationary population the progress of individuals through whatever hierarchy they belong to, that is, their social and economic mobility, will on the average be slower than in an increasing population. To show the effect in its pure form we will disregard the differences in merit, luck, inheritance, and influence among individuals whereby some move up faster and others more slowly; our interest is in *average* rates of promotion. In other words, we seek the rate of promotion insofar as it is affected by one rate of population increase as opposed to another (Waugh 1971, Bartholomew 1967).

The concept of promotion stands out most clearly in a hierarchical organization that observes some degree of seniority in promotion, and that permits a certain fraction of its employees to be foremen and above, or colonels and above, or associate professors and above. Most organizations do not have exact ratios for the number of those above to those below these ranks, but the nature of their work and budgets is such that de facto limits exist on the numbers at higher ranks. We will take as fixed and given the ratio k of those above the given rank to those below, and we want to find the age x at which a person on the average reaches the status expressed by k; we seek x as a function of r, the rate of increase. Suppose as in (5.1.1) that the population is stable, so that the fraction of individuals between ages a and $a+da$ is $be^{-ra}l(a)\,da$. Then the fraction of the population above the rank in question is the integral of this from x to β, and the fraction of the population below is the integral from α to x, where α is the age of recruitment and β the age of compulsory retirement.

The condition is described by an equation in which k is identified with a ratio of integrals:

$$k = \frac{\displaystyle\int_x^\beta e^{-ra}l(a)\,da}{\displaystyle\int_a^x e^{-ra}l(a)\,da}. \tag{5.8.1}$$

Once the life table, along with k, α, and β, is known, (5.8.1) implicitly gives the age of promotion x as a function of r.

Equation (5.8.1) cannot be solved in closed form for x, but the theory of implicit functions enables us to approximate the derivative of x with respect to r. If the ratio of integrals in (5.8.1) is called u/v, then (5.8.1) is the same as $\phi = u - kv = 0$. With a few steps of elementary calculus (details are given in Keyfitz 1973) we find

$$\frac{dx}{dr} = -\frac{\partial\phi/\partial r}{\partial\phi/\partial x} = -\frac{1}{1+k}\left[\frac{\displaystyle\int_x^\beta e^{-ra}l(a)\,da}{e^{-rx}l(x)}\right](M-m), \tag{5.8.2}$$

where M is the mean age of the group between x and β, and m of the group between α and x. Evidently dx/dr is negative; as r increases, the age of promotion declines.

To obtain an idea of how much it declines, suppose the age of recruitment α to be 20 and the age of retirement β to be 65. Then $M - m$ cannot be very different from half of the interval from 20 to 65, that is, 22.5. Consider a rank somewhat above the middle, where the person holding it has one employee above him for each two below; that is, $k = \frac{1}{2}$. With a small positive rate of increase r the expression in the square brackets of (5.8.2)

Table 5.8. Age x of passing through position k for $k = 1$ up to $k = 0.2$, for $r = 0.00$ to $r = 0.04$, based on male life table for the United States, 1968

Value of	Value of r				
k	0.00	0.01	0.02	0.03	0.04
1.0	40.86	38.55	36.39	34.43	32.71
0.8	43.26	40.91	38.64	36.54	34.66
0.6	46.31	43.97	41.64	39.41	37.35
0.4	50.32	48.14	45.86	43.56	41.33
0.2	55.93	54.27	52.36	50.27	48.08

will be about 15. Then

$$\frac{dx}{dr} \approx -\frac{1}{1+\frac{1}{2}}(15)(22.5) = -225.$$

This means that, when r falls from 0.02 to stationarity, the age of promotion to a rank two-thirds of the way up the hierarchy rises by $-225 \times (-0.02) = 4.5$ years.

For a more exact result we can solve (5.8.1) numerically for x with $r = 0.02$ and $r = 0.00$. This requires interpolation in the 5-year intervals for which such data are given, including solution of a cubic equation (Keyfitz 1973). Some of the results are shown as Table 5.8 for $l(x)$ from the United States, 1968, male life table. With $k = 0.6$ we find $x = 46.31$ for $r = 0.00$ and $x = 41.64$ for $r = 0.02$, a difference of 4.67 years, or very nearly what the derivative showed in the preceding paragraph. With $k = 1$ the difference from $r = 0.04$ to $r = 0.00$ is $40.86 - 32.71 = 8.15$ years.

The same calculation has been made on the male life table of Sweden, 1783–87, with $\overset{o}{e}_0 = 33.6$ years, representing about the highest observed mortality for which a life table is available. Such high mortality ought at least to have the advantage of speeding the promotion of the survivors, and indeed, at a given r, they show promotion to middle positions about 2 to 3 years younger than the ages in the United States, 1968, table. Thus the possible effect of high mortality in favoring promotion is 2 to 3 years; the possible effect of rapid population growth is 8 years. The reason for this is that the range of recorded human population growth is from $-\frac{1}{2}$ percent through stationary to 4 percent per year. The range of human mortality averaged over ages 20 to 65 is much narrower, say $\frac{1}{2}$ to 2 percent (Table 5.9) or one-third as much.

A Simplification. The theory can be made simpler with only a small sacrifice of realism by supposing constant mortality between ages α and β (McFarland, personal communication). Mortality rises over the age interval, but we are encouraged to suppose constancy by the small effect that mortality seems to have in any case. If the death rate is fixed at μ, the

Table 5.9. Average mortality μ for males between ages 20 and 65, where $e^{-45\mu} = l_{65}/l_{20}$, compared with intrinsic rate on female-dominant model

Country	1000μ (males)	$1000r$ (females)
Austria, 1966–68	8.1	7.8
Canada, 1966–68	7.3	7.4
Ceylon, 1967	6.9	24.6
France, 1967	8.2	8.3
Malaysia (West), 1966	9.5	30.9
Mauritius, 1966	11.3	30.5
Mexico, 1966	10.1	34.6
Norway, 1967	5.7	10.4
Spain, 1967	6.5	10.1
Sweden, 1783–87	20.2	1.4
United States, 1968	9.2	5.9
Venezuela, 1965	8.8	37.6

probability of surviving to age a is $l(a) = e^{-\mu a}$, and the equation becomes

$$k = \frac{\displaystyle\int_x^\beta e^{-ra}e^{-\mu a}\,da}{\displaystyle\int_\alpha^x e^{-ra}e^{-\mu a}\,da}.$$

Now the right-hand side of this expression lends itself to integrations in closed form, and we have

$$k = \frac{e^{-(r+\mu)x} - e^{-(r+\mu)\beta}}{e^{-(r+\mu)\alpha} - e^{-(r+\mu)x}},$$

which may be solved for x to give

$$x = -\frac{1}{r+\mu}\log\left[\frac{ke^{-(r+\mu)\alpha} + e^{-(r+\mu)\beta}}{k+1}\right], \tag{5.8.3}$$

an expression that is exactly the same function of μ as it is of r. Hence dx/dr is equal to $dx/d\mu$; population increase and high mortality help promotion to exactly the same degree.

5.8.1 The Chain Letter Principle

The analogy of the promotion process to a chain letter is easily drawn. In a chain letter (say of four names in each transmittal) the recipient (Ego) sends the person at the top of the list he has received a sum, say 1 dollar, crosses this name off the list, adds his own name at the bottom, reproduces the letter, and sends it on to four other persons. If four copies are sent on at each stage, in the short time required for five successive letters to be mailed and delivered Ego will receive $4^4 = 256$ dollars. With 10 names he will receive 10^{10} dollars. Each individual may be thought of as an ancestor of those whose names are below his on the list, and the payments are made by descendants to ancestors.

There is no fallacy in the operation of a chain letter, provided that recipients follow instructions and new people continue to be available to be brought into the scheme. Once the available population has been brought in, so that names have to be repeated, however, the scheme breaks down. And this is certain to happen long before 10^{10} names are in the process.

Payments by descendants to ancestors, say sons to great-grandfathers, in analogy to the chain letter with four names, have a very different way of cumulating than payments by ancestors to descendants, as in the ordinary passing of property from father to son. With a growing population, equal inheritance will fractionate property and give successively less to each generation. Large families and growth of population are a drawback in a regime of inherited private property, just as clearly as they are an advantage in the chain letter or promotion. Norman Ryder has emphasized this point in his lectures.

Whether the institutions of a community are such that benefits flow upward, as in promotion or the chain letter, or flow downward, as in inheritance, is crucial to attitudes on collective growth and to the size of individual families. It is arguable that the peasants' preoccupation with holding property intact in nineteenth century France was associated with low fertility; a peasant obtained security in his old age by having *few* children and thereby avoiding division of his land. In Java, on the other hand, wage labor is a main means of livelihood, and the landless laborer wants *many* children, whose combined contributions will permit him to subsist in old age. This supposes appropriate discipline in the children, just as does the chain letter.

The American pay-as-you-go social security system resembles that of Java in that the working generation supports the retired one, with the difference that in the United States the unit is no longer the family but the entire country. The U.S. system works well when there are many children to support few old people, as was the case a generation back; it runs into trouble when many old people have to be supported by few children, as will be the case early in the twenty-first century. Table 14.4 suggests the consequences for pay-as-you-go social security of the present decline in U. S. births.

It would be useful to inventory the situations in which something—money, supervisory positions, prestige—flows up from the younger generation to the older, and examine empirically corresponding attitudes on reproduction; the hypothesis of this section is that they would be found to be pronatalist. On the other hand, in situations where the flow is from the older to the younger, of which inheritance of property is the most familiar, restrictive attitudes toward reproduction would be expected.

Growth is popular in comparison to mortality: having children is preferable to dying. Moreover, a given amount of growth helps as much as a given amount of mortality; and since growth varies 3 times as much as mortality among observed populations, it can accelerate promotion about 3 times as much as even the survivors can expect in a high-death community. Cessation of growth has costs and is inevitable, and the question is what generation ought to incur the cost.

6

Birth and Population Increase from the Life Table

The stable age distribution of Chapter 5 supposed births growing at fixed rate r and subject to probability $l(a)$ of survivorship to age a. Various conclusions were reached by noting how the stable age distribution was affected by r, treated as an arbitrarily disposable parameter. Now we regard rate r not as arbitrary but as determined by the joint action of birth and death rates.

The argument rests on birth rates $m(a)$, where $m(a)\,da$ is the probability that a woman who has reached age a will give birth to a girl child before age $a + da$, or alternatively that each woman will produce exactly $m(a)\,da$ of a child in interval da. A model in which exactly $m(a)\,da$ of a child is produced by each woman in any small interval of time da departs considerably from the reality of human reproduction, where children are born, not continuously, but in units of one child each, with occasional multiple births, and random variation is conspicuous. Moreover, we will deal with the female part only of the population, and disregard the contribution of males to reproduction. (Or, less often, with the male part, but not with both together.) Such fictional treatment permits some useful conclusions to be drawn and ought to be exploited to the limit before resorting to more realistic (and necessarily more complex) models. Although unusable for families and other small populations, the continuous deterministic model provides useful answers to many questions concerning mean values in large populations.

6.1 The Characteristic Equation

To find an equation for the ultimate rate of increase r by a continuation of the means developed in Chapter 5, suppose that for a long time the chance of living to age a is $l(a)$, and that births increase at fixed rate r. Then the age distribution becomes stable, a proposition proved many times (Sharpe and Lotka 1911, Lotka 1939, Parlett 1970, Lopez 1961, McFarland 1969) and illustrated in Section 1.10. When stability has been attained, the fraction of population at ages a to $a + da$ is $be^{-ra}l(a)\,da$, as shown in the argument leading to (5.1.1). Since births are calculated by multiplying exposed population by $m(a)$, the fraction $be^{-ra}l(a)\,da$ of the population must give rise to $be^{-ra}l(a)m(a)\,da$ births per unit of total population. The integral of this last expression, from the youngest fertile age α to the highest β, is $\int_\alpha^\beta be^{-ra}l(a)m(a)\,da$, the overall birth rate.

But b is the overall birth rate in the population with age distribution $be^{-ra}l(a)\,da$, and it may therefore be equated to the integral derived in the preceding paragraph. This provides the equation

$$\int_\alpha^\beta be^{-ra}l(a)m(a)\,da = b, \tag{6.1.1}$$

or, on dividing by b,

$$\int_\alpha^\beta e^{-ra}l(a)m(a)\,da = 1, \tag{6.1.2}$$

obtained in more general fashion by Lotka (1939, p. 65; see Section 7.5 of the present chapter) and called the characteristic equation for the unknown r. Equation 6.1.2 tells us the ultimate rate of increase implied by $l(a)$ and $m(a)$ after they have been acting long enough for stability to be attained.

For a general net maternity function $l(a)m(a)$, equation (6.1.2) has an infinite number of roots, of which only the real root answers the questions treated in this chapter. The uniqueness of the real root follows from the fact that $\int_\alpha^\beta e^{-ra}l(a)m(a)\,da$, say $\psi(r)$, has a negative first derivative [prove this] and so is a monotonically decreasing function of r, which takes the value ∞ for $r = -\infty$ and the value 0 for $r = \infty$ (Figure 6.1). Thus $\psi(r)$ can only once cross any given horizontal line in the half-plane $\psi(r) > 0$, including the line one unit above the r axis, and hence $\psi(r) = 1$ can have only one real root. The function $\psi(r)$ crosses the vertical axis at

$$\psi(0) = \int_\alpha^\beta l(a)m(a)\,da = R_0,$$

the net reproductive rate.

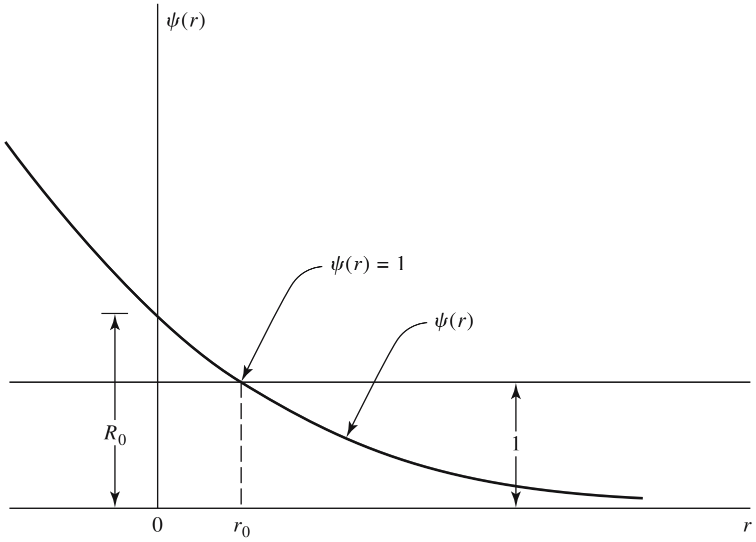


Figure 6.1. Values of $\psi(r)$, showing $\psi(0) = R_0$, and root r_0 of $\psi(r) = 1$.

6.1.1 Why Stress the Female Model?

Although the argument leading to (6.1.2) is applicable to either sex, it is most often applied to females because:

1. Data on age of parent at the birth of a child are more often collected by the registration authorities for women than for men.
2. Even when the authorities present births by age of father as well as by age of mother, more cases of unstated age of father occur, especially for illegitimate births.
3. Women have their children within a narrower and more sharply defined range of ages than do men, say 15 to 50 as against 15 to 80.
4. Both the spacing and number of children are less subject to variation among women; a woman can have children only at intervals of 1 or 2 years, so she is physiologically limited to a score or so of children, whereas a man can have hundreds. Under monogamy this asymmetry of the sexes is of no consequence, but every society allows some departures from strict monogamy, even if only to permit the remarriage of widowers.

The reasons for the emphasis on females apply over a wide range of population analysis. Both understanding and prediction must be based on whatever constancies we can find; that potential variation is less in births to women than in births to men makes the female model more useful.

On the other hand, when the age distributions and life tables for males as well as for females are available, the rate of increase can be estimated from either sex, and two estimates provide more information than one. When only one is to be had, we must rest on the fact that usually in the short run, and always in the long run, the males and females of any population increase at nearly the same rate.

6.1.2 An Iterative Method for Calculating *r*

Means of calculating this root are given in several places (Lotka 1939, p. 70; Coale 1957; Keyfitz 1968, p. 108) and need not be repeated here. But an iterative method is worth sketching because it is suggestive for other equations.

If T is an arbitrary age chosen near the mean age of childbearing, multiply both sides of (6.1.2) by e^{Tr}:

$$\int_{\alpha}^{\beta} e^{-(a-T)r} l(a)m(a)\,da = e^{Tr}, \tag{6.1.3}$$

then take logarithms of both sides, and divide by T to obtain

$$r^* = \frac{1}{T} \log \int_{\alpha}^{\beta} e^{-(a-T)r} l(a)m(a)\,da. \tag{6.1.4}$$

The left-hand side of (6.1.4) has been marked with an asterisk to suggest the use of the equation for iteration. The unknown r enters on the right, but in a way that does not greatly influence the integral. The extent to which r influences the right-hand side of (6.1.4) determines the speed of convergence, and is measured by the absolute value of the derivative dr^*/dr, a fact demonstrated formally in books on numerical analysis, such as that of Scarborough (1958, p. 209). If $|dr^*/dr|$ is less than 1 in the neighborhood of the solution, convergence will occur; if it is much less than 1, convergence will be rapid. From (6.1.4) dr^*/dr is seen to have a factor $1 - \kappa/T$, where κ is the mean age of childbearing in the stable population:

$$\kappa = \int_{\alpha}^{\beta} a e^{-ra} l(a)m(a)\,da.$$

(Since r satisfies (6.1.2), the denominator of the expression for mean age κ is unity.) The derivative dr^*/dr seems to be telling us that if we can put the arbitrary T equal to some number near κ, rapid convergence will result. In fact we do not know κ before we find r, but κ varies little among populations and T can be put at 27.5 years.

No direct use can be made of a continuous form like (6.1.4)—it must be converted to the discrete form for calculations. If T is chosen as 27.5, then with the customary approximation

$$\int_{x}^{x+5} e^{-ra} l(a)m(a)\,da = e^{-(x+2\frac{1}{2})r} {}_5L_x F_x,$$

Table 6.1. Second iteration of intrinsic rate of natural increase for the United States, females, 1968

x	${}_5L_x$	$1000F'_x$	${}_5L_xF'_x$	$e^{-(x+2\frac{1}{2}-27\frac{1}{2})(0.0056)}$
15	4.867	67.1	0.327	1.058
20	4.851	167.4	0.812	1.028
25	4.831	140.3	0.678	1.000
30	4.807	74.9	0.360	0.972
35	4.770	35.6	0.170	0.946
40	4.714	10.2	0.048	0.919
Total			2.395	

Source: For ${}_5L_x$, *Vital Statistics of the United States* (1968, Vol. 2, Section 5, pp. 5–7). For F'_x, *Statistical Abstract of the United States* (1973, p. 52).

where ${}_5L_x = \int_x^{x+5} l(a)\,da$, and F_x is the observed age-specific fertility rate for ages x to $x+4$ at last birthday. We translate (6.1.4) into calculable terms as

$$\begin{aligned} r^* = \frac{1}{27.5}\log(&e^{10r}{}_5L_{15}F_{15} + e^{5r}{}_5L_{20}F_{20} + {}_5L_{25}F_{25} \\ &+ e^{-5r}{}_5L_{30}F_{30} + e^{-10r}{}_5L_{35}F_{35} + e^{-15r}{}_5L_{40}F_{40}). \end{aligned} \tag{6.1.5}$$

The expression in parenthesis contains six terms if we include births under 15 with the 15 to 19 group and those over 45 with the 40 to 44 group. Data for United States females in 1968 are shown in Table 6.1, where F'_x, the birth rate of children of both sexes, is converted to F_x, the female birth rate, by multiplying by the fraction of births that were girls in 1968, that is, 0.487. The calculation of (6.1.5) starting with $r = 0$ gives $r^* = \log(2.395 \times 0.487)/27.5 = 0.00560$ and $r^{**} = 0.00587$ in two successive iterations.

The method is applicable to complex as well as to real roots. With complex roots we must check to ensure that the root being approached is the one sought—the real root being unique, no such precaution is needed. If all complex roots are sought, one can sweep out each as it is found, after the ith, by dividing out the polynomial in e^{-5r}; this will prevent the iteration from converging to any root more than once.

6.1.3 The Intrinsic Rate for Various Kinds of Data

The usual application of (6.1.2) is to the births and deaths of a particular period, 1 or 3 or 5 calendar years, as when the *Statistical Abstract of the United States* (1973) tells us that the intrinsic rate of increase was 0.6 percent for the United States in 1968, with 0.4 percent for White and 1.6 percent for Nonwhite components.

Male Period Intrinsic Rates. The above numbers apply to females; a corresponding calculation can be made by imputing male births to fathers, and

Table 6.2. Male and female intrinsic rates during the 1960s

	1000r	
Country and year	Male	Female
United States		
1966	12.82	9.70
1964	17.48	15.70
1963	18.68	17.10
1962	19.81	18.83
1959–61	20.92	20.70
Chile, 1964	23.55	21.05
Hungary, 1964	−2.95	−7.28
England and Wales, 1960–62	9.55	9.77

Source: Keyfitz and Flieger (1968).

in general it produces a different answer. Sometimes the answer is substantially different, as for the United States, 1964, data, where females show 1000r as 15.70, and males as 17.48. The higher male rate is largely due to fathers being older than mothers on the whole, while numbers of persons in the main childbearing ages during the 1960s reflected the rising births during the late 1930s and 1940s. If the men of an age to be fathers come from earlier smaller cohorts, their rates will be higher for the given number of births that actually took place, and it is largely this that is reported by the difference between the male and female intrinsic rates.

Table 6.2 shows some male and female rates; the differences are considerable and larger than can be explained by misstatement of age. For the United States in particular the steady increase in the gap between male and female rates seems due to the relative shortage of men of the age to be fathers in the wake of the baby boom.

Cohort Intrinsic Rate. The same equation (6.1.2) can help interpret the death and birth rates in a cohort. By following deaths and births to the individuals born in a given year or 5 years, we obtain a cohort life table and a set of age-specific birth rates that together determine r. Applied to the women who were 0 to 4 years of age in 1920 in the United States, and who therefore were 5 to 9 years of age in 1925, and so forth, they yield an intrinsic rate of 1.5 per thousand (Keyfitz and Flieger, 1968, p. 601). This means that the age-specific rates of birth and death were such that, if they persisted, the cohorts (and incidentally the period cross section) would ultimately increase at 1.5 per thousand. This value of 1000r is low because the group in question passed through its principal ages of childbearing in the late 1930s and early 1940s. Period and cohort influences are confounded in the concrete historical record.

Intrinsic Rate for One Family. The concept of intrinsic rate can be applied to a particular family. If a girl child is born at age a_1 of the mother and another girl at age a_2, r is found from

$$e^{-ra_1}l(a_1) + e^{-ra_2}l(a_2) = 1, \tag{6.1.6}$$

which is easily solved for r by the method of (6.1.4) or otherwise. The meaning of r is that a population in which all women married and bore girl children at ages a_1 and a_2 would increase at this rate; the r so calculated is an interpretation of the fertility behavior of the particular family. Nonmarriage or sterility within marriage can be accommodated by using $l(a)$ as the probability of surviving and being married and fertile by age a.

Whatever the group to which the characteristic equation 6.1.2 applies, one of its uses is to find how the rate of increase r depends on various features of the net maternity function. We will see how this is done for an increment in the birth rate $m(a)$ at a single age x. More frequently referred to, however, are the relations between the intrinsic rate and the several moments of the net maternity function. These relations are easily found once we express the characteristic equation in terms of the moments.

6.2 A Variant Form of the Characteristic Equation

It happens that $\psi(r)$, the left-hand side of (6.1.2), when divided by R_0 is the moment-generating function of the normalized net maternity function $l(a)m(a)/R_0$:

$$\frac{\psi(r)}{R_0} = \int_\alpha^\beta e^{-ra}\frac{l(a)m(a)}{R_0}\,da.$$

The sense in which $\psi(r)/R_0$ generates the moments is that it is equivalent to the infinite series

$$\frac{\psi(r)}{R_0} = 1 - \frac{rR_1}{R_0} + \frac{r^2}{2!}\frac{R_2}{R_0} - \cdots,$$

where $R_i = \int_\alpha^\beta a^i l(a)m(a)\,da$, so that R_i/R_0 is the ith moment about zero of the distribution $l(a)m(a)/R_0$. But the moments about zero are large; and even though the series converges, an impractical number of terms would be required for a good approximation. A rapidly converging series is obtained by taking the logarithm of $\psi(r)/R_0$, and so generating the functions of the moments called *cumulants* (Kendall and Stuart 1958, Vol. 1, p. 70) that we have already encountered in Section 5.3:

$$\log\frac{\psi(r)}{R_0} = -r\mu + \frac{r^2\sigma^2}{2!} - \frac{r^3\kappa_3}{3!} + \frac{r^4\kappa_4}{4!} - \cdots. \tag{6.2.1}$$

The cumulants expressed in terms of the R_i are as follows:

$$\mu = \frac{R_1}{R_0}$$

$$\sigma^2 = \frac{R_2}{R_0} - \left(\frac{R_1}{R_0}\right)^2$$
$$\kappa_3 = \frac{R_3}{R_0} - 3\frac{R_2 R_1}{R_0^2} + 2\left(\frac{R_1}{R_0}\right)^3$$
$$\vdots$$

and in terms of the moments about the mean,

$$\mu_i = \frac{\int_\alpha^\beta (a-\mu)^i l(a)m(a)\,da}{R_0},$$

they are

$$\begin{aligned}
\kappa_1 &= \mu = \mu_1 \\
\kappa_2 &= \sigma^2 = \mu_2 \\
\kappa_3 &= \mu_3 \\
\kappa_4 &= \mu_4 - 3\mu_2^2 \\
\kappa_5 &= \mu_5 - 10\mu_3\mu_2 \\
&\cdots .
\end{aligned}$$

Now the characteristic equation is $\psi(r) = 1$; that is, $\log[\psi(r)/R_0] = -\log R_0$; or, written out in terms of the cumulants, from (6.2.1) the equivalent to (6.1.2) is

$$\log R_0 = r\mu - \frac{r^2\sigma^2}{2!} + \frac{r^3\kappa_3}{3!} - \frac{r^4\kappa_4}{4!} + \cdots . \tag{6.2.2}$$

The successive terms diminish rapidly in value, even where the cumulants increase. For the United States in 1967 ($1000r = 7.13$), we have the following:

	Cumulant	Term of (6.2.2)
R_0	1.205	0.186
μ	26.28	0.187
σ^2	34.98	−0.000889
κ_3	127.5	0.0000077
κ_4	−120.7	0.000000013

The rate of convergence is slower for populations of higher r, and much slower for complex roots of (6.2.2).

6.3 Perturbation Analysis of the Intrinsic Rate

A main use of the foregoing theory is to compare populations that differ in some birth or death parameter, and note by how much they differ in intrinsic rate. For the corresponding calculations for stage-structured matrix models, see Section 13.1.

6.3.1 How the Intrinsic Rate Varies with the Moments

We are now in a position to find how the intrinsic rate varies with the several moments. To find, for example, how r in (6.2.2) varies with σ^2 we might first think of solving explicitly for r and noting how σ^2 appears in the solution. Since an explicit solution is impossible, we are fortunate that it is not necessary: the theory of implicit functions will again provide the derivative of r with respect to σ^2. If σ^2 is subject to a small increment, say $\Delta\sigma^2$, and the corresponding change in r is Δr, then, writing $\phi(r, \sigma^2)$ for the right-hand side of (6.2.2), we have in terms of partial derivatives

$$\Delta\phi(r, \sigma^2) = \frac{\partial\phi}{\partial r}\Delta r + \frac{\partial\phi}{\partial\sigma^2}\Delta\sigma^2.$$

Insofar as $\phi(r, \sigma^2)$ is a constant (namely, $\log R_0$) for values of r satisfying the characteristic equation, its change $\Delta\phi(r, \sigma^2)$ must be zero; therefore

$$\frac{\partial\phi}{\partial r}\Delta r + \frac{\partial\phi}{\partial\sigma^2}\Delta\sigma^2 = 0,$$

and solving for $\Delta r/\Delta\sigma^2$ results in

$$\frac{\Delta r}{\Delta\sigma^2} = -\frac{(\partial\phi/\partial\sigma^2)}{\partial\phi/\partial r}. \tag{6.3.1}$$

In the limit as the increment tends to zero this gives the derivative of r with respect to σ^2 in terms of the two partials, with all other moments constant.

The partials are readily calculated from (6.2.2) as

$$\frac{\partial\phi}{\partial r} = \mu - r\sigma^2 + \frac{r^2\kappa_3}{2} - \cdots \tag{6.3.2}$$

$$\frac{\partial\phi}{\partial\sigma^2} = -\frac{r^2}{2} \tag{6.3.3}$$

Therefore

$$\frac{dr}{d\sigma^2} \approx \frac{r^2}{2(\mu - r\sigma^2)}, \tag{6.3.4}$$

if we truncate after the variance, equivalent to fitting a normal curve to the net maternity function $l(a)m(a)/R_0$.

This tells us that if two increasing populations are identical in all moments, except that one has a larger variance in age of childbearing than the other, the one with the larger variance will have the higher rate of increase. If the ages of childbearing are more spread out, apparently the gain through some children being born earlier more than offsets the loss through those born later. [What if R_0 is allowed to vary?]

Easier to understand is the relation between r and μ, the mean age of childbearing. The same technique as before gives for the required derivative

$$\frac{dr}{d\mu} = \frac{-r}{\mu - r\sigma^2}. \tag{6.3.5}$$

The term $r\sigma^2$ in the denominator is relatively small; therefore it seems that for an increasing population $dr/d\mu$ is negative, and the larger μ is the smaller r is, again with everything else the same. Larger μ implies slower turnover.

At one time comparison of the United States and Canada showed a contrast between larger families in Canada, and thus a higher R_0, and younger marriage and childbearing in the United States, and thus a smaller μ. The smaller μ reversed the effect of the higher R_0, and the net outcome was a higher r for the United States.

Finally, and most obviously, the intrinsic rate is positively related to R_0. Neglecting all but the first two terms on the right-hand side of (6.2.2), we have

$$\begin{aligned}\frac{dr}{dR_0} &= -\frac{(\partial\psi/\partial R_0)}{\partial\psi/\partial r}\\ &= \frac{1}{R_0(\mu - r\sigma^2)}.\end{aligned}$$

If r is small, the intrinsic rate rises with R_0 as long as $\mu - r\sigma^2$ is positive, that is to say, always.

One kind of change can be compared with another and a set of equivalencies found. What increase in R_0 exactly offsets an increase of $\Delta\mu$ in μ and leaves the rate of increase unchanged? Should effort, say in India, go into raising the age of marriage or into disseminating birth control within marriage. India should of course do both, but the question still remains of where the marginal effort should go. With $R_0 = 1.77$, and $\mu = 25$, if raising the average age of childbearing by 3 years is easier than lowering age-specific birth rates by 5.9 percent, the effort should concentrate on age at marriage. In general if a change from R_0 to R_0^* is to be equivalent to one from μ to μ^* then if σ^2 does not change much, the equation

$$\frac{\log R_0^*}{\mu} \approx \frac{\log R_0}{\mu^*}$$

holds to a close approximation. But this is true only if we can be sure that a higher rate of childbearing will not take place within the delayed marriages, and that illegitimacy will not be substantially increased.

Such statements can be made for any of the cumulants. The first three derivatives from (6.2.2), along with their values for the case of $R_0 = 2$, $\mu = 27$, and $\sigma^2 = 40$, so that $r = 0.02618$, disregarding terms beyond $r^2\sigma^2/2$, are as follows:

$$\begin{aligned} \frac{dr}{dR_0} &= \frac{1}{R_0(\mu - r\sigma^2)} = 0.01927 \\ \frac{dr}{d\mu} &= \frac{-r}{\mu - r\sigma^2} = -0.001009 \\ \frac{dr}{d\sigma^2} &= \frac{r^2}{2(\mu - r\sigma^2)} = 0.0000132. \end{aligned} \tag{6.3.6}$$

Thus a small increase ΔR_0 increases r by $0.01927\Delta R_0$ and so forth. In proportions the derivatives are

$$\frac{dr}{dR_0} : \frac{dr}{d\mu} : \frac{dr}{d\sigma^2} = \frac{1}{R_0} : -r : \frac{r^2}{2}.$$

The influence of successive cumulants of the net maternity function alternates in sign and decreases with the powers of r. As a matter of curiosity we can extend the series, and find the ratios

$$\frac{dr}{d\kappa_3} : \frac{dr}{d\kappa_4} : \frac{dr}{d\kappa_5} : \cdots = -\frac{r^3}{3!} : \frac{r^4}{4!} : -\frac{r^5}{5!} : \cdots .$$

6.3.2 Change in Births at One Age

Other calculations of the relation of rate of increase to aspects of the net maternity function depend on the form (6.1.2) of the characteristic equation. For example, we would like to know whether the use of contraception by young women will have more effect on the rate of increase than its use by older women. Suppose that for the 1-year interval around x the value of $m(x)$ is changed to $m(x) + \Delta m(x)$; our problem is to find the value Δr by which this modifies r.

The new value $r + \Delta r$ is obtained from the characteristic equation 6.1.2 in the form

$$\int_\alpha^\beta e^{-(r+\Delta r)a} l(a)[m(a) + \Delta m(x)]\, da = 1, \tag{6.3.7}$$

where we have modified $m(a)$ by adding to it the quantity $\Delta m(x)$ for the 1 year x. [This unorthodox notation will not cause any trouble if we think of $\Delta m(x)$ as a function of a, defined to be zero everywhere except in the interval $x - \frac{1}{2}$ to $x + \frac{1}{2}$, where it is $\Delta m(x)$.] The left-hand side of (6.3.7) consists of two additive parts, one an integral, and the other approximately $e^{-(r+\Delta r)x}l(x)\Delta m(x)$. If for $e^{-(r+\Delta r)a}$ we write $e^{-ra}(1-a\Delta r)$ within the integral, two integrals emerge, of which $\int_{\alpha}^{\beta} e^{-ra}l(a)m(a)\,da$ equals unity by our original equation (6.1.2), and $\int_{\alpha}^{\beta} ae^{-ra}l(a)m(a)\,da$ (without denominator) equals κ, the mean age of childbearing in the stable population. If $\Delta m(x)$ is small enough that the term involving $\Delta r\Delta m(x)$ may be ignored, we obtain

$$\Delta r \approx \frac{e^{-rx}l(x)}{\kappa}\Delta m(x). \tag{6.3.8}$$

Thus the intrinsic rate r is changed by $e^{-rx}l(x)/\kappa$ times the change in the age-specific birth rate $m(x)$. Note that the coefficient of $\Delta m(x)$ is proportional to the number of women in the stable age distribution. Thus the relative effect of changes in rates at the several ages is proportional to the number of women at these ages. Such sensitivity analysis is a way of exploiting models to obtain conditional statements of cause and effect in other instances where the result is less obvious.

An equivalent method of working out the effect of change in a birth or death rate involves the use of implicit functions. The interested reader can work out other examples. Here we proceed to a more general case in which not single ages but groups of ages are considered.

6.4 Arbitrary Pattern of Birth Rate Decline

As the birth rate declines in the United States or any other country, it falls more rapidly at some ages than at others. For modernizing populations the initial fall has been greatest at the oldest ages of childbearing, as has been noted alike for the United States and for Taiwan. The ages that drop are partially related to the means of population control used: sterilization applies mostly to the older ages of childbearing; the pill, to younger ages (at least while it is a novelty); abortion, to all ages. The intrauterine device (IUD) is not much used by women until they have had a child, suggesting an aggregate impact on ages intermediate between the pill and sterilization. Possible cuts $g(a)$ that might be taken out of the birth function $m(a)$ by these three methods of birth control are suggested in Figure 6.2. The fertility function that remains is $m(a)-kg(a)$, where k is a constant.

We will first consider an arbitrary function $g(a)$ and see the effect of removing it times some constant k from the birth function $m(a)$. Our analysis will concentrate on two special cases:

1. The effect on r of deducting $kg(a)$ from $m(a)$, where k is small, as though one were trying to examine the direction and pace of the first move toward fertility reduction.

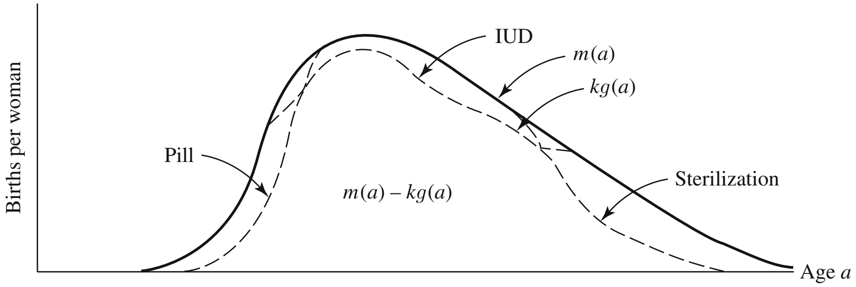


Figure 6.2. Area $kg(a)$ removed from net maternity function by the several methods of birth control.

2. The final condition of stationarity that will result from subtraction of $kg(a)$ from $m(a)$, where now k is large enough to produce stationarity.

The first case is simple because k is small, and the second is simple because the final condition is stationary; intermediate values of k are more difficult to handle and will not be discussed here. From the first case we will know in what direction the system starts to move, and from the second where it ends; by comparing these two we obtain at least a suggestion of how the movement is modified as k goes through intermediate finite values.

6.4.1 Effect of Small Arbitrary Change in Birth Function

Suppose an arbitrary function $g(a)$, positive or negative, such as unity for 1 year of age and zero elsewhere, or equal to $m(a)$, the birth function, or some other variant. We consider deductions from $m(a)$ equal to $kg(a)$, that is to say, $kg(a)$ is a bite of arbitrary shape taken out of $m(a)$. For purposes of this part of the argument k is small enough that we can neglect second-order terms like $(\Delta r)^2$ and $k\,\Delta r$, where Δr is the difference that the deduction of $kg(a)$ from $m(a)$ makes to r.

To find the value of Δr that corresponds to $kg(a)$ we have to solve the equation

$$\int_\alpha^\beta e^{-(r+\Delta r)a} l(a)[m(a) - kg(a)]\,da = 1, \tag{6.4.1}$$

where r is defined by (6.1.2). Expanding the exponential $e^{-\Delta r a} \approx 1 - \Delta r a$ in (6.4.1), ignoring the term involving $k\,\Delta r$, and solving for Δr gives

$$\Delta r \approx -\frac{\displaystyle\int_\alpha^\beta e^{-ra} l(a) g(a)\,da}{\kappa} k, \tag{6.4.2}$$

where κ is the mean age of childbearing in the stable population.

If $g(a) = m(a)$ for all a, (6.4.2) becomes

$$\Delta r = -\frac{k}{\kappa} \qquad \text{or} \qquad k = -\kappa\,\Delta r.$$

The question might be what fractional change in age-specific rates will bring the rate of increase down by Δr; the answer here given is that a subtraction of $100\kappa\,\Delta r$ percent at each age does it. A variety of results, including (6.3.8), can be obtained as special cases of (6.4.2). The method is still that of comparative statics: birth rates are taken as $m(a)$ for one population and as $m(a) - kg(a)$ for another, and the two stable conditions compared.

6.4.2 Amount of Change Needed for Drop to Bare Replacement

For bare replacement the net reproduction rate, the expected number of girl children by which a girl child born now will be replaced,

$$R_0 = \int_\alpha^\beta l(a)m(a)\,da,$$

must equal unity. If R_0 is to equal unity on the age-specific birth rates $m(a) - kg(a)$, we must have

$$\int_\alpha^\beta l(a)[m(a) - kg(a)]\,da = 1,$$

where the solution for k will tell just how much of the change shaped like $g(a)$ in the age pattern is required for the reduction to replacement. The answer is evidently

$$k = \frac{R_0 - 1}{\displaystyle\int_\alpha^\beta l(a)g(a)\,da}.$$

For $g(a) = m(a)$ this is $k = (R_0 - 1)/R_0$, as accords with intuition. For applications of this logic in epidemiology, where the reproductive rate of concern is that of the pathogen, see p. 194.

6.4.3 Effect of Uniformly Lower Death Rates

Suppose two populations of which one has the force of mortality $\mu(a)$ and the other $\mu^*(a) = \mu(a) + \delta$, where δ is the same for all ages. Then the population with $\mu(a) + \delta$ will have survivorship

$$l^*(x) = \exp\left\{-\int_0^x [\mu(a) + \delta]\,da\right\} = e^{-\delta x} l(x).$$

Its characteristic equation will be

$$\int_{\alpha}^{\beta} e^{-r^* a} l^*(a) m(a)\, da = 1$$

or

$$\int_{\alpha}^{\beta} e^{-(r^*+\delta)a} l(a) m(a)\, da = 1.$$

But this is identical with the characteristic equation for $l(a)m(a)$, as given in (6.1.2), except that it has $r^* + \delta$ rather than r. And, since the characteristic equation has a unique real root, it follows that $r^* + \delta = r$, or $r^* = r - \delta$. In words, the solution with the incremented mortality is the original solution less the increment of mortality.

From this it follows that age distribution is unaffected by a constant increment of mortality. This is so because the age distribution is proportional to $e^{-rx} l(x)\, dx$; and if we change $l(x)$ to $l^*(x) = e^{-\delta x} l(x)$, so that r goes to $r^* = r - \delta$, then $e^{-rx} l(x)\, dx$ goes to $e^{-(r-\delta)x} e^{-\delta x} l(x)$, that is, it is unchanged on cancellation of $e^{-\delta x}$.

That a change in death rates that is the same at all ages has no effect on age distribution is true more generally. Suppose an age distribution, however irregular, and a decline of 0.001 in mortality rates at all ages. Then exactly one person in a thousand who would have died on the former regime now survives, age by age. This increases the number at every age by exactly 0.001, that is, multiplies it by 1.001, and multiplying every age by 1.001 can have no effect on the age distribution (Coale 1956).

6.5 Drop in Births Required to Offset a Drop in Deaths

Let us find the fraction by which existing fertility must decline so as to offset an absolute decrease in mortality equal to k at every separate age. In symbols, suppose that for women who have reached age a the chance of dying between ages a and $a + da$ drops from $\mu(a)\, da$ to $[\mu(a) - k]\, da$, and this applies with the same k at all ages; by what uniform fraction f at all ages would women have to lower their birth rates in order for the intrinsic rate r to remain the same? If the drop in births takes place uniformly at all ages, the old probability of having a child between ages a and $a + da$ being $m(a)\, da$, and the new probability being $(1 - f)m(a)\, da$, we can determine the unknown f.

The effect of lowering the death rate by the amount k at each age is equivalent to increasing the survivorship $l(a)$ by the factor e^{ka}, for by definition the new $l^*(a)$ must be, as before,

$$l^*(a) = \exp\left\{-\int_0^a [\mu(t) - k]\,dt\right\}, \tag{6.5.1}$$

and carrying out the integration gives $l^*(a) = e^{ka}l(a)$. Alongside the old characteristic equation 6.1.2 we have a new one:

$$\int_\alpha^\beta e^{-ra}e^{ka}l(a)(1-f)m(a)\,da = 1. \tag{6.5.2}$$

We will replace e^{ka} by $1 + ka$, permissible if k is small, take advantage of (6.1.2) to cancel $\int_\alpha^\beta e^{-ra}l(a)m(a)\,da$ on the left of (6.5.2) against the 1 on the right, and solve for f as

$$f = \frac{k\kappa}{1 + k\kappa}, \tag{6.5.3}$$

where κ is again the mean age of childbearing in the stable population. Approximately, the fraction that the births would have to be reduced to offset a fall of k in the mortality of every age is $k\kappa$. For a population whose κ is 27 years, an absolute fall of 0.001 in the death rate at each separate age would require a fall of the fraction $(0.001)(27)/[1 + (0.001)(27)] = 0.0263$ in the birth rates at each age to offset it.

A variant of the same question, left to the reader, asks how much must be cut off the $m(a)$ curve at its upper end to offset a drop k in the age-specific death rates. We go on to find the drop in fertility that would offset the largest possible decrease in mortality—its complete elimination.

6.5.1 *The Drop in Fertility That Would Offset a Drop in Mortality to Zero*

The question has been raised how a drop in mortality to zero could be offset by a change in fertility at all ages of women. In the simplest form, what value of f, the fractional fall in fertility at all ages, is associated with an intrinsic rate r that would remain unchanged if mortality to the end of reproduction were to drop to zero? Putting $l(a) = 1$ and changing $m(a)$ to $m(a)(1 - f)$ in (6.1.2) provides the equation that f must satisfy:

$$\int_\alpha^\beta e^{-ra}m(a)(1-f)\,da = 1, \tag{6.5.4}$$

while r is still subject to (6.1.2). The value of f from (6.5.4) is

$$f = 1 - \frac{1}{\displaystyle\int_\alpha^\beta e^{-ra}m(a)\,da}. \tag{6.5.5}$$

A fertility drop by the fraction f at all ages would preserve r, the ultimate rate of increase, against the hypothetical drop to zero of mortality up to age β. A very slightly different problem is to find the fraction f that would preserve the generation ratio of increase R_0, the net reproduction rate, defined as

$$R_0 = \int_\alpha^\beta l(a)m(a)\,da.$$

The required fraction is the solution in f of the equation

$$\int_\alpha^\beta m(a)(1-f)\,da = R_0, \tag{6.5.6}$$

or

$$f = 1 - \frac{R_0}{\displaystyle\int_\alpha^\beta m(a)\,da} = 1 - \frac{R_0}{G_0}, \tag{6.5.7}$$

the integral in the denominator being the gross reproduction rate G_0 (Coale 1973b).

Applying this to developed countries shows that a 3 or 4 percent decline in births would offset the drop to zero mortality. In 1967 the United States G_0 was 1.26 and the R_0 was 1.21 (Keyfitz and Flieger 1971, p. 361). For less developed countries where mortality is somewhat higher, a 5 to 15 percent drop in births would suffice. No one need be concerned that a further fall in the Mexican death rate, for example, will add seriously to the demographic problem of that country. Mexico's G_0 in 1966 was 3.17, its R_0 2.71, and its birth rate $1000b = 43.96$ (Keyfitz and Flieger 1971, p. 345). A drop to $1000b = 37.58$ would offset a transition to zero mortality.

The conclusion that further decline in mortality need not provoke any substantial rise in the rate of population growth probably applies to the world as a whole. Insofar as mortality up to the end of childbearing is already very low, further diminution in it can have only a small effect on the rate of increase, and this small effect is likely to be offset for most countries by the decline in fertility during the 1970s. The parts of the world where this conclusion does not yet apply are mostly in tropical Africa.

6.5.2 Diseases of Infancy Versus Heart Disease: Their Effects on Population Increase

The death rate from cardiovascular renal diseases for females in the United States in 1964 was 448 per 100,000 population, and that from certain diseases of infancy (as defined) was 26 (Preston, Keyfitz, and Schoen, 1972, p. 770). Standardized, the numbers were 383 and 26, still a ratio of nearly 15 to 1. Eliminating certain diseases of infancy adds only 0.930 year to $\overset{o}{e}_0$, whereas eliminating CVR adds 17.068 years, or 18 times as much. All indications are that CVR has at least 15 times the effect on mortality of certain diseases of infancy.

But no such proportion appears in the effects on overall population increase of eliminating these diseases. Consider first the effect on the net reproduction rate. Using the life table calculated as though CVR were eliminated raises the net reproduction rate by 0.002. Eliminating certain diseases of infancy without altering the other causes raises the net reproduction rate by 0.018. Heart disease may be 15 times as prevalent as diseases of infancy, but its long-term effect on increase is only about one-ninth as great.

6.6 Moments of the Dying Population in Terms of Those of the Living, and Conversely

If we know the age distribution of the living population of the life table, we ought to be able to find the age distribution of the dying. Stated in life table symbols, the problem is simply that of finding the distribution d_x in terms of l_x, and the answer is $d_x = l_x - l_{x+1}$.

Translating moments of the living into those of the dying is straightforward; given the moments of $l(x)$, we proceed to find the moments of $l(x)\mu(x)$, where $\mu(x) = -\frac{1}{l(x)}\frac{dl(x)}{dx}$. The result will enable us to find the mean age, variance, and other parameters of the living from those of the dying, and vice versa.

One device is to relate the cumulant-generating function of the ages of the dying to that of the ages of the living. For a general distribution function $f(x)$, define $\psi(r)$ as the transform

$$\psi(r) = \log\left[\int_0^\omega e^{-rx} f(x)\,dx\right],$$

and define the cumulants of $f(x)$ as the coefficients of the powers of r in the expansion

$$\psi(r) = -r\kappa_1 + \frac{r^2}{2!}\kappa_2 - \frac{r^3}{3!}\kappa_3 + \cdots. \tag{6.6.1}$$

For the special case where $f(x)$ is proportional to $l(x)$, the transform is

$$\log\left[\frac{\displaystyle\int_0^\omega e^{-rx} l(x)\,dx}{\displaystyle\int_0^\omega l(x)\,dx}\right],$$

which will be called $\psi_L(r)$; the κ_L are the cumulants of the distribution $l(x)/\int_0^\omega l(x)\,dx$. In reference to our distribution of population the mean age is μ_L, the variance σ_L^2, the skewness κ_{3L}, and the kurtosis κ_{4L}. For deaths the mean is $\kappa_1 = \overset{o}{e}_0$, the variance $\kappa_2 = \sigma^2$, the skewness κ_3, and the kurtosis κ_4.

The exponential of the cumulant-generating function of the ages of the dying is

$$\exp[\psi(r)] = \int_0^\omega e^{-rx} l(x)\mu(x)\,dx = -\int_0^\omega e^{-rx}\,dl(x),$$

and on integration by parts we have

$$\begin{aligned}-\int_0^\omega e^{-rx}\,dl(x) &= -e^{-rx}l(x)|_0^\omega - r\int_0^\omega e^{-rx}l(x)\,dx \\ &= 1 - r\int_0^\omega e^{-rx}l(x)\,dx.\end{aligned}$$

But the integral here is $\int_0^\omega l(x)\,dx$ times the exponential of the cumulant-generating function of the living. Hence there follows the identity

$$\exp[\psi(r)] = 1 - r\overset{o}{e}_0\exp[\psi_L(r)], \tag{6.6.2}$$

where $\psi_L(r) = \log[\int_0^\omega e^{-rx}l(x)\,dx/\int_0^\omega l(x)\,dx]$, referring to the living, is distinguished from $\psi(r) = \log[\int_0^\omega e^{-rx}l(x)\mu(x)\,dx]$, referring to the dying. Note that $\int_0^\omega l(x)\mu(x)\,dx = 1$; therefore we do not need a denominator for $\psi(r)$, generating the cumulants of the dying, but do require the denominator $\int_0^\omega l(x)\,dx = \overset{o}{e}_0$ for $\psi_L(r)$.

The basic result is (6.6.2); and, subject to conditions of convergence that do not seem to give trouble in practice, expanding $\psi(r)$ and $\psi_L(r)$ within it enables us to find one set of cumulants in terms of the other:

$$\begin{aligned}&\exp\left(-\overset{o}{e}_0 r + \frac{\sigma^2 r^2}{2} - \frac{\kappa_3 r^3}{6} + \cdots\right) = \\ &\qquad 1 - r\overset{o}{e}_0\exp\left(-\mu_L r + \frac{\sigma_L^2 r^2}{2} - \frac{\kappa_{3L} r^3}{6} + \cdots\right),\end{aligned}$$

or on expanding the exponentials

$$\begin{aligned}&1 - \overset{o}{e}_0 r + (\sigma^2 + \overset{o}{e}{}_0^2)\frac{r^2}{2} - \cdots = \\ &\qquad 1 - r\overset{o}{e}_0\left[1 - \mu_L r + (\sigma_L^2 + \mu_L^2)\frac{r^2}{2} - \cdots\right].\end{aligned}$$

Equating powers of r, we get

$$\overset{o}{e}_0 = \overset{o}{e}_0$$

$$\frac{\sigma^2 + \overset{o}{e}{}_0^2}{2} = \overset{o}{e}_0\mu_L$$

$$\frac{\kappa_3 + 3\sigma^2\overset{o}{e}_0 + \overset{o}{e}{}_0^3}{6} = \frac{\overset{o}{e}_0(\sigma_L^2 + \mu_L^2)}{2}$$

and so on.

Thus $\sigma^2 = 2\overset{o}{e}_0\mu_L - \overset{o}{e}{}_0^2$, identically, and we have without approximation the variance in the age of dying in terms of the mean age of the living and $\overset{o}{e}_0$, all for the life table population. For example, in the case of Colombian females, 1964, where $\overset{o}{e}_0 = 61.563$ and $\mu_L = 37.360$, the variance in age at death is

$$\sigma^2 = 2(61.563)(37.360) - (61.563)^2 = 810.0.$$

Thus the four cumulants of ages in a stationary population given for some 200 populations in Keyfitz and Flieger (1968) provide five cumulants of ages of the dying for those same populations.

6.6.1 Expectation of Life as a Function of Crude Birth and Death Rates

If we know only the crude death rate of a population we can say little about its expectation of life; a crude death rate of 10 can apply to a country with a life expectation as high as 75 years or as low as 60 years, in the former case with a rate of increase of 0.005 and in the latter of 0.030. If we are also told the crude birth rate or the rate of natural increase, we can narrow this range considerably, as McCann (1973) has shown.

We know that d, the crude death rate in the stable population, is

$$d = \int_0^\omega be^{-ra}l(a)\mu(a)\,da,$$

so that, dividing by b and noting that the right-hand side is the exponential of the cumulant-generating function of the distribution of deaths by age expressed in terms of $-r$, we have

$$\frac{d}{b} = \int_0^\omega e^{-ra}l(a)\mu(a)\,da,$$

or $\log d - \log b = \psi(r)$, and this is the same as

$$\log b - \log d = r\overset{o}{e}_0 - \frac{r^2\sigma^2}{2} + \frac{r^3\kappa_3}{6} - \cdots,$$

since in the stationary population the mean age of dying is the expectation of life $\overset{o}{e}_0$. Solving for $\overset{o}{e}_0$ gives

$$\overset{o}{e}_0 = \frac{\log b - \log d}{b-d} + r\left(\frac{\sigma^2}{2} - \frac{r\kappa_3}{6} + \frac{r^2\kappa_4}{24} - \cdots\right), \tag{6.6.3}$$

where $r = b - d$. Now the expression in parentheses on the right depends on the life table in a fairly systematic way. With a set of model life tables we can iterate to the value of $\overset{o}{e}_0$, given b and d.

A different approach to the same problem is to start with the identity

$$\frac{1}{b} = \int_0^\omega e^{-rx}l(x)\,dx,$$

divide by $\overset{o}{e}_0 = \int_0^\omega l(x)\,dx$, take logarithms, and then use the fact that the right-hand side is the cumulant-generating function of the living:

$$\log(b\,\overset{o}{e}_0) = \mu_L r - \frac{\sigma_L^2 r^2}{2} + \frac{\kappa_{3L} r^3}{6} - \cdots,$$

where now the cumulants are of the distribution of the living in the stationary population rather than of the dying. Hence we have

$$\overset{o}{e}_0 = \frac{1}{b}\exp\left(\mu_L r - \frac{\sigma_L^2 r^2}{2} + \cdots\right). \tag{6.6.4}$$

Applying (6.6.3) to Colombian females, 1964, and dropping terms involving third and higher cumulants, we obtain

$$\begin{aligned}\overset{o}{e}_0 &= \frac{\log 0.03840 - \log 0.01006}{0.02835} + \frac{(0.02835)(810.0)}{2} \\ &= 47.25 + 11.48 = 58.73\end{aligned}$$

against the life table value of 61.56. From (6.6.4), dropping terms, we have

$$\begin{aligned}\overset{o}{e}_0 &= \frac{e^{\mu_L r}}{b} = \frac{e^{(37.36)(0.02835)}}{0.03840} \\ &= 75.10,\end{aligned}$$

which is a much poorer approximation. But if we take one more term in the exponential of (6.6.4) with $\sigma_L^2 = 539.34$, this becomes

$$\overset{o}{e}_0 = 60.47,$$

which is the best of the three.

These results use the mean and variance of the life table, for which we are looking. In practice one would not be able to attain them in a single calculation, but would find them from the $\overset{o}{e}_0$ of the previous iteration. The method can choose an $\overset{o}{e}_0$ from a one-parameter set of life tables.

Which of (6.6.3) and (6.6.4) is preferable? The decision depends on which set of cumulants (that for the living or that for the dying) is more nearly invariant with respect to $\overset{o}{e}_0$, since that determines the number of iterations required in a given series of model life tables. The answer also depends on which is more robust with respect to the choice of the series of model tables, a line of investigation that must be left for another time. (James McCann and Samuel Preston contributed much of the foregoing argument.)

7

Birth and Population Increase from Matrix Population Models

7.1 Solution of the Projection Equation

Chapter 6 combined birth and death in the life table framework to arrive at an analysis of population growth. In this chapter we approach the same question using the matrix population model

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t), \tag{7.1.1}$$

where $\mathbf{n}(t)$ is the population vector and $\mathbf{A}$ is a stage-classified projection matrix. Instead of using repeated multiplication to generate numerical projections, as we did in Chapter 3, we turn now to the analytical solution of (7.1.1). This solution provides the basic tools for demographic analysis. We will present two derivations, because it may be helpful to see two ways of getting to the same result (cf. Caswell 1997).

Both derivations use the eigenvalues and eigenvectors of $\mathbf{A}$. We recall that a vector $\mathbf{w}$ is a (right) eigenvector of $\mathbf{A}$ and the scalar λ is the corresponding eigenvalue if

$$\mathbf{A}\mathbf{w} = \lambda\mathbf{w}. \tag{7.1.2}$$

Equation (7.1.2) implies that

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{w} = \mathbf{0}, \tag{7.1.3}$$

where $\mathbf{I}$ is an identity matrix and $\mathbf{0}$ is a vector of zeros. A nonzero solution for $\mathbf{w}$ exists only if $(\mathbf{A} - \lambda\mathbf{I})$ is singular; i.e., if

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0. \tag{7.1.4}$$

This is the characteristic equation, corresponding to the continuous-time equation (6.1.2).

Associated with each eigenvalue there is also a left eigenvector $\mathbf{v}$ that satisfies

$$\mathbf{v}^*\mathbf{A} = \lambda\mathbf{v}^*, \tag{7.1.5}$$

where v^* is the complex conjugate transpose of $\mathbf{v}$.

If $\mathbf{A}$ is an $s \times s$ matrix, the characteristic equation is an sth-order polynomial and $\mathbf{A}$ will have s eigenvalue–eigenvector pairs.

$$\mathbf{A}\mathbf{w}_i = \lambda_i\mathbf{w}_i \tag{7.1.6}$$

$$\mathbf{v}_i^*\mathbf{A} = \lambda\mathbf{v}_i^* \tag{7.1.7}$$

each of which is a solution to the characteristic equation. The solution to (7.1.1) depends on them all. We will assume that the eigenvectors are linearly independent; a sufficient condition for this is that all the eigenvalues are distinct.

Example 7.1 Characteristic equation for age-classified models

The characteristic equation for age-classified models can be derived from a small example. Consider a Leslie matrix with four age classes; its characteristic equation is

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det\begin{pmatrix} F_1 - \lambda & F_2 & F_3 & F_4 \\ P_1 & -\lambda & 0 & 0 \\ 0 & P_2 & -\lambda & 0 \\ 0 & 0 & P_3 & -\lambda \end{pmatrix} = 0. \tag{7.1.8}$$

Expanding this determinant along the first column gives

$$\begin{aligned} 0 &= (F_1 - \lambda)\det\begin{pmatrix} -\lambda & 0 & 0 \\ P_2 & -\lambda & 0 \\ 0 & P_3 & -\lambda \end{pmatrix} - P_1\det\begin{pmatrix} F_2 & F_3 & F_4 \\ P_2 & -\lambda & 0 \\ 0 & P_3 & -\lambda \end{pmatrix} \\ &= (F_1 - \lambda)(-\lambda)\det\begin{pmatrix} -\lambda & 0 \\ P_3 & -\lambda \end{pmatrix} - P_1F_2\det\begin{pmatrix} -\lambda & 0 \\ P_3 & -\lambda \end{pmatrix} \\ &\quad + P_1P_2\det\begin{pmatrix} F_3 & F_4 \\ P_3 & -\lambda \end{pmatrix} \\ &= \lambda^4 - F_1\lambda^3 - P_1F_2\lambda^2 - P_1P_2F_3\lambda - P_1P_2P_3F_4. \end{aligned} \tag{7.1.9}$$

Dividing both sides by λ^4 and rearranging puts the equation into a more familiar form:

$$1 = F_1\lambda^{-1} + P_1F_2\lambda^{-2} + P_1P_2F_3\lambda^{-3} + P_1P_2P_3F_4\lambda^{-4} \tag{7.1.10}$$

or, in general,

$$1 = \sum_i \left(\prod_{j=1}^{i-1} P_j \right) F_i \lambda^{-i}. \tag{7.1.11}$$

This is a discrete form of Lotka's integral equation for the instantaneous population growth rate r

$$1 = \int_0^\infty l(x)m(x)e^{-rx}dx$$

with $\prod_{j=1}^{i-1} P_j$ corresponding to $l(x)$, F_i corresponding to $m(x)$, and λ^{-i} corresponding to e^{-rx}.

7.1.1 Derivation 1

We turn now to the solution of (7.1.1) for general matrices. We write the initial population $\mathbf{n}_0$ as a linear combination of the right eigenvectors $\mathbf{w}_i$ of $\mathbf{A}$:

$$\mathbf{n}_0 = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_s\mathbf{w}_s \tag{7.1.12}$$

for some set of coefficients c_i yet to be determined. The linear independence of the eigenvectors guarantees that we can write any $\mathbf{n}_0$ in this form.

We find the coefficients c_i by writing (7.1.12) as

$$\begin{aligned} \mathbf{n}_0 &= \begin{pmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_s \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_s \end{pmatrix} && (7.1.13) \\ &= \mathbf{Wc} && (7.1.14) \end{aligned}$$

where $\mathbf{W}$ is a matrix whose columns are the eigenvectors $\mathbf{w}_i$, and $\mathbf{c}$ is a vector whose elements are the c_i. Thus

$$\mathbf{c} = \mathbf{W}^{-1}\mathbf{n}_0. \tag{7.1.15}$$

Now multiply $\mathbf{n}_0$ by $\mathbf{A}$ to obtain $\mathbf{n}(1)$:

$$\begin{aligned} \mathbf{n}(1) &= \mathbf{An}_0 \\ &= \sum_i c_i \mathbf{Aw}_i \\ &= \sum_i c_i \lambda_i \mathbf{w}_i. && (7.1.16) \end{aligned}$$

If we multiply by $\mathbf{A}$ again, we get $\mathbf{n}(2)$:

$$\begin{aligned} \mathbf{n}(2) &= \mathbf{An}(1) \\ &= \sum_i c_i \lambda_i \mathbf{Aw}_i \end{aligned}$$

$$= \sum_i c_i \lambda_i^2 \mathbf{w}_i. \tag{7.1.17}$$

It should not be hard to convince yourself that continuing this process yields the solution

$$\mathbf{n}(t) = \sum_i c_i \lambda_i^t \mathbf{w}_i. \tag{7.1.18}$$

The solution to (7.1.1) is a weighted sum of s exponentials, the weights determined by the initial conditions.

7.1.2 Derivation 2

Iterating (7.1.1) leads directly to the solution

$$\mathbf{n}(t) = \mathbf{A}^t \mathbf{n}_0. \tag{7.1.19}$$

The dynamics of the population are determined by the behavior of $\mathbf{A}^t$. This is a function of the matrix $\mathbf{A}$, and evaluating functions of matrices is an important problem in matrix algebra. If $\mathbf{A}$ were a diagonal matrix the answer would be easy:

$$\mathbf{A}^t = \begin{pmatrix} a_{11}^t & 0 & \cdots & 0 \\ 0 & a_{22}^t & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & a_{ss}^t \end{pmatrix}. \tag{7.1.20}$$

$\mathbf{A}$ is not diagonal, but if it has linearly independent eigenvectors, it is similar to a diagonal matrix $\boldsymbol{\Lambda}$ whose diagonal entries are the eigenvalues λ_i. That is, there exists a nonsingular matrix $\mathbf{W}$ such that

$$\mathbf{W}^{-1}\mathbf{A}\mathbf{W} = \boldsymbol{\Lambda} \tag{7.1.21}$$

or, equivalently,

$$\mathbf{A} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^{-1}. \tag{7.1.22}$$

Thus

$$\begin{aligned} \mathbf{A}^2 &= \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^{-1}\mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^{-1} \\ &= \mathbf{W}\boldsymbol{\Lambda}^2\mathbf{W}^{-1} \end{aligned} \tag{7.1.23}$$

and, in general,

$$\mathbf{A}^t = \mathbf{W}\boldsymbol{\Lambda}^t\mathbf{W}^{-1} \tag{7.1.24}$$

$$= \mathbf{W} \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \lambda_s^t \end{pmatrix} \mathbf{W}^{-1}. \tag{7.1.25}$$

Equation (7.1.21) implies that $\mathbf{AW} = \mathbf{W\Lambda}$; thus the columns of $\mathbf{W}$ are the right eigenvectors $\mathbf{w}_i$ of $\mathbf{A}$:

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_s \end{pmatrix} \tag{7.1.26}$$

[i.e., the same as (7.1.14)]. It also implies that $\mathbf{W}^{-1}\mathbf{A} = \mathbf{\Lambda W}^{-1}$. Thus the rows of $\mathbf{W}^{-1}$ are the complex conjugates of the left eigenvectors $\mathbf{v}_i$ of $\mathbf{A}$:

$$\mathbf{W}^{-1} = \overline{\mathbf{V}} = \begin{pmatrix} \mathbf{v}_1^* \\ \mathbf{v}_2^* \\ \vdots \\ \mathbf{v}_s^* \end{pmatrix} \tag{7.1.27}$$

So, we can rewrite (13.3.4) as

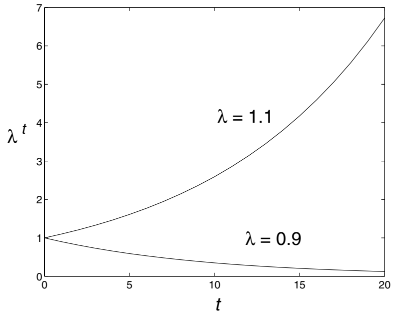
$$\begin{aligned} \mathbf{n}(t) &= \mathbf{A}^t\mathbf{n}_0 && (7.1.28)\\ &= \mathbf{W\Lambda}^t\overline{\mathbf{V}}\mathbf{n}_0 && (7.1.29)\\ &= \sum_i \lambda_i^t \mathbf{w}_i\mathbf{v}_i^*\mathbf{n}_0, && (7.1.30) \end{aligned}$$

where $\mathbf{v}_i^*$ is the complex conjugate transpose of the left eigenvector corresponding to λ_i. The product $\mathbf{w}_i\mathbf{v}_i^*$ is a matrix; these are sometimes called the *constituent matrices* of $\mathbf{A}$. The product $\mathbf{v}_i^*\mathbf{n}_0$ is a scalar; it is the same as the coefficient c_i given by (7.1.15); see also Chapter 9.

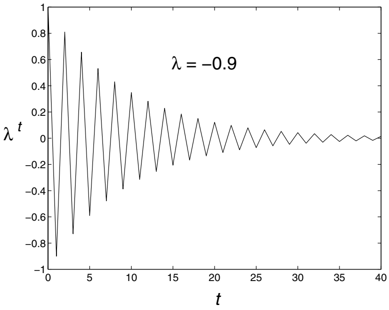
7.1.3 Effects of the Eigenvalues

No matter how you derive it, the long-term behavior of $\mathbf{n}(t)$, given by (7.1.18) and (7.1.30), depends on the eigenvalues λ_i as they are raised to higher and higher powers. The eigenvalues may be real or complex. Their contributions to the solution can be summarized as follows.

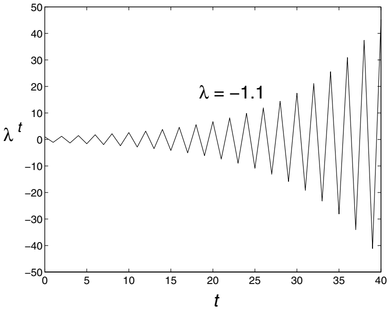
- If λ_i is positive, λ_i^t produces exponential growth if $\lambda > 1$ and exponential decay if $\lambda < 1$.



- If $-1 < \lambda_i < 0$, then λ_i^t exhibits damped oscillations with a period equal to 2.



- If $\lambda_i < -1$, then λ_i^t produces diverging oscillations with period 2.



- Complex eigenvalues produce oscillations. Suppose that $\lambda = a + bi$, and write it in polar coordinates,

$$\lambda = |\lambda|(\cos\theta + i\sin\theta),$$

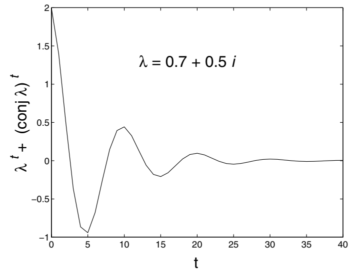
where $|\lambda| = \sqrt{a^2 + b^2}$ is the magnitude of λ and $\theta = \tan^{-1}(b/a)$ is the angle formed by λ in the complex plane. Raising λ to the tth power yields

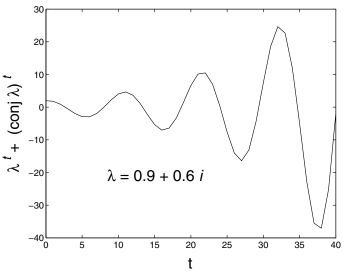
$$\lambda^t = |\lambda|^t(\cos\theta t + i\sin\theta t). \tag{7.1.31}$$

Complex eigenvalues come in complex conjugate pairs, so $\bar{\lambda} = a - bi$ will also be an eigenvalue. The solution to the projection equation will thus contain terms of the form

$$\lambda^t + \bar{\lambda}^t = |\lambda|^t 2\cos\theta t. \tag{7.1.32}$$

Thus, as a complex eigenvalue is raised to higher and higher powers, its magnitude $|\lambda|^t$ increases or decreases exponentially, depending on whether $|\lambda|$ is greater or less than 1. Its angle in the complex plane increases by θ each time step, completing an oscillation with a period of $2\pi/\theta$.





Regardless of whether λ_i is real or complex, the boundary between population increase and population decrease comes at $|\lambda_i| = 1$.

Example 7.2 An age-classified population

Keyfitz and Flieger (1971) give an age-classified matrix, with 5-year age classes and a projection interval of 5 years, for the United States population in 1966. The entries are

i	F_i	P_i
1	0	0.99670
2	0.00102	0.99837
3	0.08515	0.99780
4	0.30574	0.99672
5	0.40002	0.99607
6	0.28061	0.99472
7	0.15260	0.99240
8	0.06420	0.98867
9	0.01483	0.98274
10	0.00089	

The eigenvalues of this matrix (in decreasing order of absolute magnitude), their magnitudes, and the angle θ (as a fraction of π) defined in the complex plane by each are

λ_i	$\lvert\lambda_i\rvert$	θ/π
1.0498	1.0498	0.0000
$0.3112 + 0.7442i$	0.8067	0.3739
$0.3112 - 0.7442i$	0.8067	-0.3739
$-0.3939 + 0.3658i$	0.5375	-0.7618
$-0.3939 - 0.3658i$	0.5375	0.7618
$0.0115 + 0.5221i$	0.5223	0.4930
$0.0115 - 0.5221i$	0.5223	-0.4930
$-0.4112 + 0.1204i$	0.4284	-0.9093
$-0.4112 - 0.1204i$	0.4284	0.9093
-0.0852	0.0852	1.0000

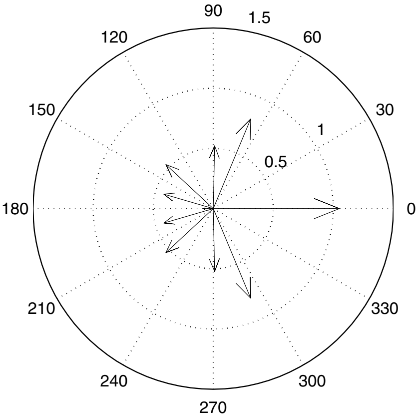


Figure 7.1. The eigenvalue spectrum for the projection matrix for the United States population in 1966, plotted in the complex plane. The dominant eigenvalue is $\lambda_1 = 1.0498$.

> Figure 7.1 shows the eigenvalue spectrum plotted in the complex plane. (The computation of eigenvalues and eigenvectors is discussed in Section 7.4.)
> The largest eigenvalue is real and positive. The second largest is a complex conjugate pair at angles of about $\theta = \pm 0.37\pi$ in the complex plane. This pair would generate oscillations with a period $2\pi/\theta \approx 5.4$ projection intervals. Since the projection interval is 5 years, the oscillation produced by this pair of eigenvalues has a period of 27 years.

7.2 The Strong Ergodic Theorem

A population is said to be *ergodic* if its eventual behavior is independent of its initial state (Cohen 1979a). Ergodic theorems will appear repeatedly throughout the following chapters. Most linear time-invariant matrix models are ergodic in a very strong sense of the word. Our understanding of ergodicity relies on a powerful theorem about the eigenvalues of nonnegative matrices. A matrix is *nonnegative* if all its elements are greater than or equal to zero, and *positive* if all its elements are strictly greater than zero. All population projection matrices are nonnegative, because negative entries imply negative organisms, but they are not usually positive.

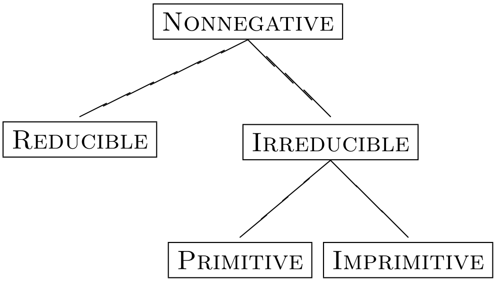


Figure 7.2. The properties of nonnegative matrices.

7.2.1 The Perron–Frobenius theorem

A set of results known collectively as the Perron–Frobenius theorem describes the eigenvalues of nonnegative matrices. Two additional properties of the matrix are important for this theorem: irreducibility and primitivity. Nonnegative matrices can be divided into reducible and irreducible matrices; irreducible matrices in turn are divided into primitive and imprimitive matrices (Figure 7.2).

To define these terms, we need to define some structures in the life cycle graph. A *path* from $\mathcal{N}_i$ to $\mathcal{N}_j$ is a sequence of arcs, traversed in the direction of the arrows, beginning at $\mathcal{N}_i$ and ending at $\mathcal{N}_j$, and passing through no node more than once. A *loop* is a path from a node to itself. The length of a path or a loop is the number of arcs it contains. A *self-loop* has length 1. Thus, for example, the life cycle graph in Figure 3.9(a) contains three loops (of length 2, 3, and 4) and no self-loops. The sequence $\mathcal{N}_1 \to \mathcal{N}_2 \to \mathcal{N}_3$ is a path of length 2 from $\mathcal{N}_1$ to $\mathcal{N}_3$. The sequence $\mathcal{N}_1 \to \mathcal{N}_2 \to \mathcal{N}_1 \to \mathcal{N}_2 \to \mathcal{N}_3$ is not a path, because it passes through $\mathcal{N}_2$ twice.

7.2.1.1 Irreducibility

A nonnegative matrix is *irreducible* if and only if its life cycle graph contains a path from every node to every other node (Rosenblatt 1957, Berman and Plemmons 1994). Such a graph is said to be strongly connected. A reducible life cycle contains at least one stage that cannot contribute, by any developmental path, to some other stage or stages. A reducible matrix can always be rearranged, by renumbering the stages, into a normal form:

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{B} & \mathbf{0} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right), \tag{7.2.1}$$

where the square submatrices $\mathbf{B}$ and $\mathbf{D}$ are either irreducible or can themselves be divided to eventually yield a series of irreducible diagonal blocks (Gantmacher 1959, Berman and Plemmons 1994).

Most life cycle graphs are irreducible. One common exception (Figure 7.3a) occurs in life cycles with postreproductive age classes, which cannot contribute to any younger age class (e.g., the killer whale life cy-

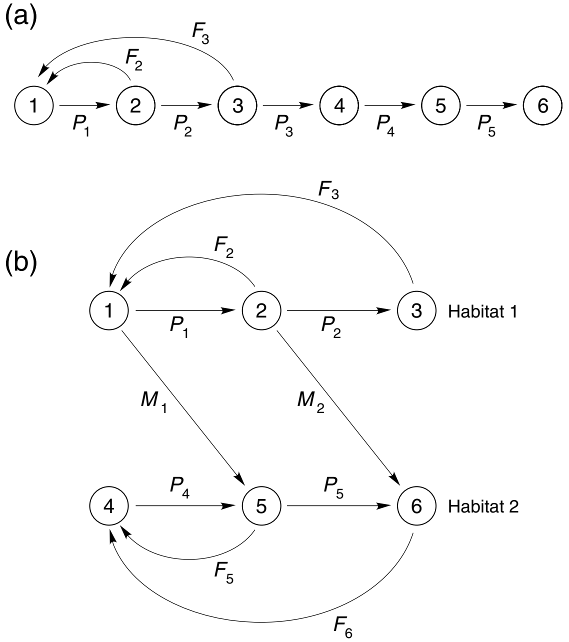


Figure 7.3. (a) An age-classified population with postreproductive age classes. The corresponding matrix is reducible, because postreproductive age classes cannot contribute to any younger age class. (b) A source-sink metapopulation; the resulting matrix is reducible because individuals in Habitat 2 make no contribution to any stage in Habitat 1.

cle in Figure 3.10). Another exception might arise in spatially structured populations with one-way dispersal patterns (a "source-sink" model). Figure 7.3b shows an example; Habitat 1 contributes individuals to Habitat 2, but not vice versa.

7.2.1.2 Primitivity

An irreducible nonnegative matrix $\mathbf{A}$ is *primitive* if it becomes positive when raised to sufficiently high powers, i.e., if $\mathbf{A}^k$ is strictly positive for some $k > 0$. A reducible matrix cannot be primitive, because when the matrix in (7.2.1) is raised to powers, the upper-right block remains zero.

Primitivity can be evaluated from the life cycle graph; a graph is primitive if it is irreducible and the greatest common divisor of the lengths of its loops is 1 (Rosenblatt 1957, Berman and Plemmons 1994). An imprimitive matrix is said to be *cyclic*, and to have an *index of imprimitivity d* equal to the greatest common divisor of the loop lengths in the life cycle graph. The age-classified graph in Figure 3.9(a) has loops of length 2, 3, and 4; the greatest common divisor of these lengths is 1, so the corresponding matrix is primitive.

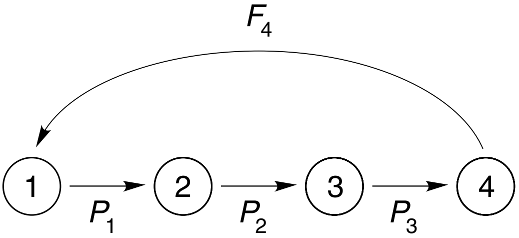


Figure 7.4. An imprimitive life cycle graph for an age-classified population with one fixed age at reproduction. The graph contains only one loop, of length 4.

A sufficient condition for primitivity of an irreducible age-classified matrix is the existence of any two adjacent age classes with positive fertility (Sykes 1969, Demetrius 1971). For irreducible stage-classified models, a sufficient condition for primitivity is the presence of at least one self-loop.

Most population projection matrices are primitive. The only significant exceptions are age-classified matrices with a single reproductive age class, which might be appropriate for semelparous species such as monocarpic perennial bamboos (Janzen 1976), periodical cicadas, or Pacific salmon. The life cycle graph for such an organism (Figure 7.4) contains only one loop of length d, where d is the age of reproduction. Imprimitive matrices also arise in models of annual plants (Chapter 13 of MPM).

7.2.1.3 Evaluating Irreducibility and Primitivity Numerically

It is easy to evaluate irreducibility and primitivity of small matrices by inspection of the life cycle graph, but large matrices can be difficult. Horn and Johnson (1985, pp. 507–520) summarize several theorems that provide numerical methods to evaluate irreducibility and primitivity. Suppose that $\mathbf{A}$ is a nonnegative $s \times s$ matrix. Then

- $\mathbf{A}$ is irreducible if and only if $(\mathbf{I}+\mathbf{A})^{s-1}$ is positive.
- Let c denote the length of the shortest loop in the life cycle graph of $\mathbf{A}$. Then $\mathbf{A}$ is primitive if and only if $\mathbf{A}^{s+c(s-2)}$ is positive. Since the exponent increases with c, this result can be applied letting c be the length of *any* loop in the life cycle graph.
- $\mathbf{A}$ is primitive if and only if $\mathbf{A}^{s^2-2s+2}$ is positive.

Thus, for example, a 20×20 matrix can be checked for irreducibility by seeing if $(\mathbf{I}+\mathbf{A})^{19}$ is positive. If the graph of $\mathbf{A}$ contains a loop of length 4, $\mathbf{A}$ is primitive if and only if $\mathbf{A}^{92}$ is positive. Without knowing the lengths of any loops, the primitivity of $\mathbf{A}$ can be checked by seeing if $\mathbf{A}^{342}$ is positive. The calculation of powers in these results is made easier by repeated squaring of the matrix, i.e., calculating $\mathbf{A}^2$, $\mathbf{A}^4$, $\mathbf{A}^8$, etc. The primitivity of a 20×20 matrix can be evaluated by calculating $\mathbf{A}^{512}$ using only nine matrix multiplications.

Irreducibility and primitivity are determined by the arrangement of zero and nonzero entries in the matrix; they are independent of the values of those entries. Thus it is sometimes convenient to evaluate them using the adjacency matrix of $\mathbf{A}$, which has zeros in the same locations as the zeros in $\mathbf{A}$, and ones in those locations where $\mathbf{A}$ has positive elements. To avoid overflow or underflow when raising matrices to high powers, rescale the matrix to its adjacency matrix after each multiplication.

7.2.1.4 The Perron–Frobenius theorem

The Perron–Frobenius theorem[*] describes the eigenvalues and eigenvectors of a nonnegative matrix $\mathbf{A}$. Its most important conclusion is that there generally exists one eigenvalue that is greater than or equal to any of the others in magnitude. Without loss of generality, we will call this eigenvalue λ_1; it is called the *dominant eigenvalue* of $\mathbf{A}$.

Primitive matrices: If $\mathbf{A}$ is primitive, then there exists a real, positive eigenvalue λ_1 that is a simple root of the characteristic equation. This eigenvalue is strictly greater in magnitude than any other eigenvalue. The right and left eigenvectors $\mathbf{w}_1$ and $\mathbf{v}_1$ corresponding to λ_1 are real and strictly positive. There may be other real eigenvalues besides λ_1, but λ_1 is the only eigenvalue with nonnegative eigenvectors.

Irreducible but imprimitive matrices: If the matrix $\mathbf{A}$ is irreducible but imprimitive, with index of imprimitivity d, then there exists a real positive eigenvalue λ_1 which is a simple root of the characteristic equation. The associated right and left eigenvectors $\mathbf{w}_1$ and $\mathbf{v}_1$ are positive.

The dominant eigenvalue λ_1 is greater than or equal in magnitude to any of the other eigenvalues; i.e.,

$$\lambda_1 \geq |\lambda_i| \qquad i > 1$$

but the spectrum of $\mathbf{A}$ contains d eigenvalues equal in magnitude to λ_1. One is λ_1 itself; the others are the $d-1$ complex eigenvalues

$$\lambda_1 e^{2k\pi i/d} \qquad k = 1, 2, \ldots, d-1.$$

For example, if

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 5 \\ 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \end{pmatrix} \tag{7.2.2}$$

---

[*] For proofs, see Gantmacher (1959), Seneta (1981), or Horn and Johnson (1985).

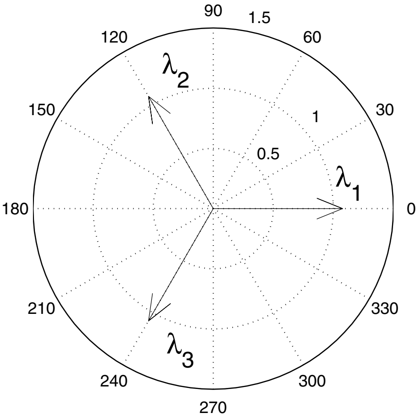


Figure 7.5. The eigenvalue spectrum of the 3×3 imprimitive age-classified matrix (7.2.2) with index of imprimitivity $d = 3$.

the index of imprimitivity $d = 3$, and the eigenvalues are $\lambda_1 = 1.0772$, $\lambda_2 = -0.5389 + 0.9329i$, and $\lambda_3 = \overline{\lambda_2}$. Figure 7.5 plots these three eigenvalues in the complex plane.

Reducible matrices: If $\mathbf{A}$ is reducible, there exists a real eigenvalue $\lambda_1 \geq 0$ with corresponding right and left eigenvectors $\mathbf{w}_1 \geq 0$ and $\mathbf{v}_1 \geq 0$. This eigenvalue $\lambda_1 \geq |\lambda_i|$ for $i > 1$.

The results of the Perron–Frobenius Theorem are summarized in Figure 7.6.

7.2.2 Population Growth Rate

The dominant eigenvalue λ_1 determines the ergodic properties of population growth. Consider $\mathbf{n}(t)$ from (7.1.18):

$$\mathbf{n}(t) = c_1\lambda_1^t\mathbf{w}_1 + c_2\lambda_2^t\mathbf{w}_2 + c_3\lambda_3^t\mathbf{w}_3 + \cdots,$$

where the eigenvalues are numbered in order of decreasing magnitude. If λ_1 is strictly greater in magnitude than all the other eigenvalues, it will eventually dominate all the other terms in (7.1.18). Regardless of the initial population, the other exponential terms will eventually become negligible and the population will grow at a rate given by λ_1 and with a structure proportional to $\mathbf{w}_1$. Dividing both sides by λ_1^t yields

$$\frac{\mathbf{n}(t)}{\lambda_1^t} = c_1\mathbf{w}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^t\mathbf{w}_2 + c_3\left(\frac{\lambda_3}{\lambda_1}\right)^t\mathbf{w}_3 + \cdots. \qquad (7.2.3)$$

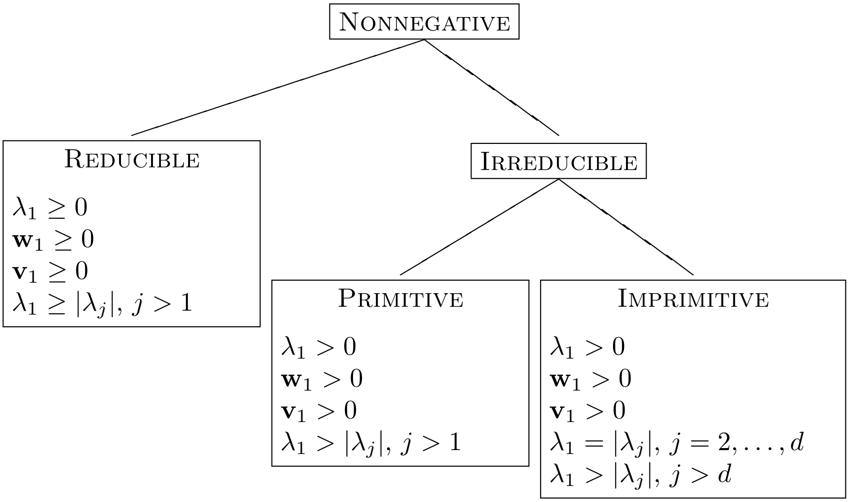


Figure 7.6. The relations among nonnegative, irreducible, and primitive matrices, and a summary of the results of the Perron–Frobenius theorem.

If $\lambda_1 > |\lambda_i|$ for $i \geq 2$, then taking the limit as $t \to \infty$ yields

$$\lim_{t \to \infty} \frac{\mathbf{n}(t)}{\lambda_1^t} = c_1 \mathbf{w}_1. \tag{7.2.4}$$

This result is known as the strong ergodic theorem (Cohen 1979a); it shows that, if $\mathbf{A}$ is primitive, the long-term dynamics of the population are described by the population growth rate λ_1 and the stable population structure $\mathbf{w}_1$. The growth rate λ_1 is related to the intrinsic rate of increase r obtained from Lotka's equation by $\lambda_1 = e^r$ or $r = \ln \lambda_1$.

The stable population structure is given by $\mathbf{w}_1$. Since eigenvectors are determined only up to a scalar constant, $\mathbf{w}_1$ can be scaled as desired, for example, so that its elements sum to 1 and represent proportions, or so that they sum to 100 and represent percentages. Writing the coefficients $c_i = \mathbf{v}_i^* \mathbf{n}_0$, as in (7.1.30), requires that whatever scaling is chosen satisfies $\mathbf{v}_i^* \mathbf{w}_i = 1$.

7.2.2.1 The Stable Age Distribution

We can find the stable age distribution directly from the Leslie matrix. Writing $\mathbf{Aw} = \lambda \mathbf{w}$, rows 2–$s$ give

$$\begin{aligned} P_1 w_1 &= \lambda w_2 \\ P_2 w_2 &= \lambda w_3 \\ P_3 w_3 &= \lambda w_4 \\ &\vdots \end{aligned}$$

$$P_{s-1}w_{s-1} = \lambda w_s.$$

Since $\mathbf{w}$ can be scaled at will, let $w_1 = 1$, and then solve for successive values. This gives the stable age distribution, with abundances of each age class measured relative to the abundance of the first:

$$\begin{aligned} w_1 &= 1 \\ w_2 &= P_1\lambda^{-1} \\ w_3 &= P_1P_2\lambda^{-2} \\ &\vdots \\ w_s &= P_1P_2\cdots P_{s-1}\lambda^{-s+1}, \end{aligned} \tag{7.2.5}$$

which is directly analogous to (5.1.1), $\prod_i P_i$ corresponding to $l(x)$ and λ^{-i} corresponding to e^{-rx}.

7.2.3 Imprimitive Matrices

An imprimitive matrix $\mathbf{A}$ has d eigenvalues with the same absolute magnitude, where d is the index of imprimitivity. Only one of these eigenvalues (λ_1) is real and positive; the others form angles in the complex plane of $\theta = 2\pi/d,\ 4\pi/d, \ldots, (d-1)2\pi/d$, and are thus either complex or, if $d = 2$, negative. The common magnitude of this set of d eigenvalues is strictly greater than the magnitude of any of the remaining eigenvalues, so as $t \to \infty$ only the d leading eigenvalues have any influence on population dynamics.

Cull and Vogt (1973, 1974, 1976) and Svirezhev and Logofet (1983) discuss the resulting dynamics in detail. Because of the complex eigenvalues, the stage distribution does not converge, but instead oscillates with a period d, as does the total population size. Suppose that $d = 3$, and consider (7.2.3). The eigenvalues λ_2 and λ_3 are now complex, and $|\lambda_2| = |\lambda_3| = \lambda_1$. Using (7.1.31) for λ_i^t, the limit (7.2.4) is replaced by

$$\frac{\mathbf{n}(t)}{\lambda_1^t} \to c_1\mathbf{w}_1 + c_2(\cos\theta t + i\sin\theta t)\mathbf{w}_2 + c_3(\cos\theta t - i\sin\theta t)\mathbf{w}_3 \tag{7.2.6}$$

as $t \to \infty$. Since $\mathbf{w}_2$ and $\mathbf{w}_3$ and c_2 and c_3 are complex conjugates, the imaginary parts of (7.2.6) cancel out, so that $\mathbf{n}(t)$ is real, as it should be.

From (7.2.6) it follows that $\mathbf{w}_1$ is still a stable stage distribution in the sense that, if $\mathbf{n}(0)$ is proportional to $\mathbf{w}_1$, so that $c_2 = c_3 = 0$, the population will remain at that structure for all time. However, $\mathbf{w}_1$ is *not* stable in the sense that an initial population not proportional to $\mathbf{w}_1$ will converge to it. Instead, the limit in (7.2.6) is periodic, with period d.

Cull and Vogt (1973) show that a running average of $\mathbf{n}(t)$, with the average taken over the period of the oscillation, converges to $\mathbf{w}_1$ and grows at a rate λ_1:

$$\lim_{t\to\infty} \frac{1}{d}\sum_{j=1}^{d} \frac{\mathbf{n}(t+j)}{\lambda_1^{t+j}} = c_1\mathbf{w}_1 \tag{7.2.7}$$

The oscillatory asymptotic dynamics of imprimitive matrices makes intuitive sense in the context of the life cycle graph. The existence of an index of imprimitivity $d > 1$ means that there is an inherent cyclicity in the life cycle; all loops are multiples of some common loop length. This cyclicity is reflected in the dynamics of the population.

7.2.4 Reducible Matrices

The dynamics generated by a reducible matrix $\mathbf{A}$ are affected by initial conditions. Rewrite $\mathbf{A}$ in normal form as

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{B}_1 & \mathbf{0} \\ \hline \mathbf{B}_{21} & \mathbf{B}_2 \end{array}\right). \tag{7.2.8}$$

If $\mathbf{B}_1$ and $\mathbf{B}_2$ are reducible, they are further subdivided until we finally arrive at

$$\mathbf{A} = \begin{pmatrix} \mathbf{B}_1 & & & & \\ \mathbf{B}_{21} & \mathbf{B}_2 & & & \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_3 & & \\ \vdots & & & \ddots & \\ \mathbf{B}_{m1} & \mathbf{B}_{m2} & \cdots & \cdots & \mathbf{B}_m \end{pmatrix}, \tag{7.2.9}$$

where all the diagonal blocks *are* irreducible. Except for possible permutations of the block matrices, this decomposition is unique.

Let S_i denote the set of stages in the submatrix $\mathbf{B}_i$. The stages in S_i communicate with each other (because $\mathbf{B}_i$ is irreducible), and may communicate with stages in $S_{i+1}, \ldots, S_m$, but cannot communicate with the stages in $S_1, \ldots, S_{i-1}$. Since the dynamics of the stages in S_1 are independent of the stages in any of the other sets, the irreducible matrix $\mathbf{B}_1$ can be analyzed by itself. Sometimes this makes biological sense, as the following example shows.

Example 7.3 A population with postreproductive age classes

The matrix for a population with postreproductive age classes (Figure 7.3a) can be put into normal form as

$$\left(\begin{array}{c|c} \mathbf{B}_1 & \mathbf{0} \\ \hline \mathbf{B}_{21} & \mathbf{B}_2 \end{array}\right) = \left(\begin{array}{ccc|cc} 0 & F_2 & F_3 & 0 & 0 \\ P_1 & 0 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 & 0 \\ \hline 0 & 0 & P_3 & 0 & 0 \\ 0 & 0 & 0 & P_4 & 0 \end{array}\right). \tag{7.2.10}$$

In this case, $S_1 = \{\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3\}$ and $S_2 = \{\mathcal{N}_4, \mathcal{N}_5\}$. The irreducible submatrix $\mathbf{B}_1$ is the population projection matrix for the reproductive age classes, while the submatrix $\mathbf{B}_2$ contains the survival probabilities of the postreproductive age classes. An analysis of $\mathbf{B}_1$ alone will give the growth rate of the reproductive part of the population. This rate is unaffected by the postreproductive age classes, and the strong ergodic theorem guarantees that the age distribution of this part of the population will converge to stability.

Human demographers routinely truncate projections of the female population at the end of the reproductive period. The 10×10 matrix for the United States population in Example 7.2 is based on 5-year age classes; it actually represents the submatrix $\mathbf{B}_1$, but the vital rates of postreproductive females (over 50 years old) have no impact on the eigenvalues.

If you choose instead to analyze the entire matrix $\mathbf{A}$, the long-term dynamics depend on the initial conditions. The dynamics of the population in Example 7.3, for example, depend very much on whether the initial population consists only of postreproductive females, or includes some reproductive individuals.

More general cases can be more complicated. The sets of states S_i defined by the partition of the life cycle generate a set of *invariant subspaces* (Gantmacher 1959):

$$R_1 = S_1 + \cdots + S_m \tag{7.2.11}$$

$$R_2 = S_2 + \cdots + S_m \tag{7.2.12}$$

$$\vdots \tag{7.2.13}$$

$$R_{m-1} = S_{m-1} + S_m \tag{7.2.14}$$

$$R_m = S_m. \tag{7.2.15}$$

The subspaces are "invariant" because if the population lies in R_i at time t, it also lies in R_i at time $t+1$. The invariant subspaces are nested, so $R_m \subset R_{m-1} \cdots \subset R_1$.

The existence of more than one invariant subspace means that the population is not, strictly speaking, ergodic. The long-term dynamics are not independent of initial conditions. A trajectory whose initial condition is in R_j but not in any of the larger subspaces $R_1, \ldots, R_{j-1}$ is trapped in R_j. It may or may not grow at the rate given by the dominant eigenvalue of $\mathbf{A}$. In Example 7.3 a population beginning with only postreproductive individuals certainly does not do so.

An initial condition that lies in R_1, but not in any of the smaller subspaces $R_2, \ldots, R_m$, contains individuals lying in each of the sets of states $S_1, \ldots, S_m$. The trajectory resulting from this initial condition will eventually grow at the rate given by the dominant eigenvalue of $\mathbf{A}$.

In practice, when confronted by a reducible matrix, examine the life cycle structure to see which subspaces are of most biological interest. In a population with postreproductive age classes, it is obvious that the reproductive part of the population dominates the dynamics, and that submatrix is of main interest. In a metapopulation model the entire matrix is of interest, and you must take care in interpreting the eigenvalues and eigenvectors.

7.3 Transient Dynamics and Convergence

The dominant eigenvalue of $\mathbf{A}$ gives the asymptotic population growth rate. The population would grow at this rate if present environmental conditions were maintained indefinitely. That is unlikely to happen, so it is often useful to calculate measures of the short-term or transient dynamics. The simplest approach is numerical projection, which shows exactly what happens to the population from any specific initial condition.

7.3.1 The Damping Ratio and Convergence

The solution (7.1.18) contains information on the rate of convergence to the stable population structure and the oscillations produced by the subdominant eigenvalues during convergence (e.g., Lefkovitch 1971, Usher 1976, Horst 1977, Longstaff 1984, Rago and Goodyear 1987). The asymptotic rate of convergence is governed by the eigenvalue(s) with the second largest magnitude. From (7.2.3) it is clear that, all else being equal, convergence will be more rapid the larger λ_1 is relative to the other eigenvalues. This leads to the definition of the *damping ratio*

$$\rho = \lambda_1 / |\lambda_2|. \tag{7.3.1}$$

From (7.2.3) it follows that

$$\lim_{t \to \infty} \left(\frac{\mathbf{n}(t)}{\lambda^t} - c_1 \mathbf{w}_1 - c_2 \rho^{-t} \mathbf{w}_2 \right) = 0.$$

Thus, for large t

$$\begin{aligned} \left\| \frac{\mathbf{n}(t)}{\lambda_1^t} - c_1 \mathbf{w}_1 \right\| &\leq k \rho^{-t} \\ &= k e^{-t \log \rho} \end{aligned} \tag{7.3.2}$$

for some constant k. That is, convergence to the stable structure is asymptotically exponential, at a rate at least as fast as $\log \rho$. Convergence could be faster; for example, an initial population for which $c_2 = 0$ would converge at a rate at least as fast as $\log(\lambda_1 / |\lambda_3|)$. And, to take the extreme case, an initial condition that is proportional to $\mathbf{w}_1$ has already converged.

The time t_x required for the contribution of λ_1 to become x times as great as that of λ_2 can be calculated as

$$\left(\frac{\lambda_1}{|\lambda_2|}\right)^{t_x} = x \tag{7.3.3}$$

from which

$$t_x = \log(x)/\log(\rho). \tag{7.3.4}$$

For a method that considers all the subdominant eigenvalues, see Horst (1977) and Rago and Goodyear (1987).

7.3.1.1 Entropy and Convergence

Demetrius (e.g., 1974, 1983) introduced a quantity he calls population entropy, H, which is related to the rate of convergence. The derivation of H is based on subtle connections between the dynamics of the age structure in a matrix population model and the dynamics of the probability distribution of “genealogies,” that is, of pathways that an individual, its ancestors, and its descendants may take through the life cycle graph, and is claimed to have important evolutionary implications.

The result, for an age-classified model, is defined in terms of the discrete net fertility function

$$\phi_i = P_1 P_2 \cdots P_{i-1} F_i \tag{7.3.5}$$

Since the characteristic equation (7.1.11) is

$$\sum_i \phi_i \lambda^{-i} = 1, \tag{7.3.6}$$

the quantity $\phi_i \lambda^{-i}$ can be treated as a probability distribution, and one can calculate the entropy

$$H = -\sum_i \phi_i \lambda^{-i} \log(\phi_i \lambda^{-i}). \tag{7.3.7}$$

Demetrius actually uses

$$H = -\frac{1}{T} \sum_i \phi_i \lambda^{-i} \log(\phi_i \lambda^{-i}), \tag{7.3.8}$$

where $T = \sum_i i \phi_i \lambda^{-i}$ is a measure of mean generation length.

H measures the extent to which reproduction is spread through the life cycle. For a semelparous life cycle, $H = 0$, while H is maximized by spreading expected reproduction out evenly over the life cycle. For a related index measuring the curvature of the survivorship curve, see Section 4.3.

Tuljapurkar (1982, 1993) has studied the relations between population entropy and measures of convergence based on the eigenvalues. He shows that H gives a lower bound on the rate of convergence to the stable age distribution. In the second paper, he discusses the case of an iteroparous but imprimitive matrix, in which the entropy is nonzero, but the population does not converge to the stable structure. In these cases, H measures the rate at which the age distribution converges to the d-dimensional subspace of the state space within which the oscillating population structure must lie (Tuljapurkar 1993).

7.3.1.2 Factors Influencing the Damping Ratio

The factors determining ρ in age-classified populations have been studied by Coale (1972) for humans and Taylor (1979) for insects. Both studies used Lotka's continuous time model (7.5.1) for the birth series $B(t)$. They measured the damping ratio by $r_1 - u_2$ where u_2 is the real part of the second root of Lotka's equation; this measure is equivalent to $\log \rho$.

Human Populations

Coale's (1972) calculations are based on the net fertility function $\phi(x) = l(x)m(x)$, which gives the age-specific expected reproductive output of a newborn individual. He assumed that $\phi(x)$ is approximately symmetrical, with mean μ_1 (μ_1 is a measure of mean generation time; see Section 11.3.5). He measured the concentration of reproduction near μ_1 by the proportion of $\phi(x)$ in the interval $3\mu_1/4 \leq x \leq 5\mu_1/4$, and measured the asymmetry of $\phi(x)$ by the ratio of the median to the mean. Based on 47 human life tables, he concluded that $\log \rho$ was negatively correlated with both the concentration and the asymmetry of fertility. Thus, populations in which reproduction is spread out symmetrically over a wide range of ages should converge to the stable age distribution more rapidly than those with tightly restricted ages at reproduction, or skewed distributions of age at reproduction. See Section 14.6 for a more analysis.

Insect Populations

Taylor (1979) carried out an extensive investigation of $r_1 - u_2$ in insect populations. He used an age-classified model after converting calendar time to units of "degree-days" to compensate for the temperature dependence of development rate in insects. He used t_{20}, the time required for contribution of the second root to decline to 5 percent of that of the dominant root [see (7.3.4)], to measure of the time required for convergence to the stable age distribution.

Based on statistical models of the $l(x)$ and $m(x)$ functions, Taylor concluded that the time to convergence was most strongly affected by the age at first reproduction (earlier reproduction leads to more rapid convergence) and the variance in $m(x)$ (greater variance leads to more rapid convergence). The time to convergence was nearly independent of survivorship and the amount of reproduction.

Taylor also calculated t_{20} from data for 36 populations of 30 species of insects and mites. He found values of t_{20} ranging from 280 degree-days (for the aphid *Myzus persicae*) to 115,120 degree-days (for each of two species of moths). Typical figures for the duration of a growing season are on the order of 1000–3000 degree-days, which would allow 40–70 percent of the populations Taylor examined to converge to within 5 percent of their stable age distributions.

Of course, this conclusion assumes constant vital rates during the course of the growing season. As Taylor notes, variation in the vital rates would slow down the process of convergence.

Taylor concludes that "the greater part of insect species existing in seasonal environments never experience, or spend a small proportion of their time in, a stable age distribution" (Taylor 1979, p. 527). This conclusion may be too strong. There is nothing magical about a 5 percent contribution of the second root as representing convergence. If a 10 percent threshold is used instead, 55–75 percent of the populations would have time to converge in a typical growing season. A 20 percent threshold would allow 65–80 percent of the populations to converge.

Carey (1983) provides some additional insight into the process of convergence. He collected information on the stage distribution (egg, immature, and mature) of a tetranychid mite on cotton plants throughout a growing season [mites were among the most rapidly converging species in Taylor's (1979) tabulation]. He compared these distributions with the stable age distribution calculated on the basis of laboratory life table experiments. The laboratory data predicted stable age distributions for increasing, stationary, and declining populations differing only in their fertilities. He found that age distributions of the increasing, stationary, and declining phases of the field population tended to converge to the predicted stable values. This indicates that even if populations do not reach a constant stable age distribution, the patterns of convergence to and deviation from that distribution may provide useful biological information.

Multi-regional and Age-Size Models.

In age-size or multi-regional models individuals are characterized by multiple criteria. In this context, one can ask whether the age distribution converges more or less rapidly than the size or region distribution (Liaw 1980, Law and Edley 1990). The answer seems to depend on the details of the model.

Liaw (1980) found that the age distribution in a multi-regional model for the human population of Canada converged more rapidly than the region distribution. He interpreted this in terms of the magnitudes of the subdominant eigenvalues of the matrix. Keyfitz (1980) likened the phenomenon to the convergence of temperature in a set of interconnected rooms; the higher rate of mixing homogenizes temperature within each room before the temperature differences between rooms can decay. This interpretation depends, of course, on the fact that migration between regions is generally small.

Law and Edley (1990) draw the opposite conclusion for age-size matrices. They conclude that the size distribution should converge more rapidly than the age distribution. This is partly a function of their assumption that all births take place in the smallest size class. Since all individuals are born at the same size, each cohort proceeds through the life cycle graph with the same distribution of size-at-age. Thus, once the individuals in the initial population have died out, the size distribution no longer changes. This will not be true in a multi-state model in which individuals may be born into more than one size class, region, or other state.

7.3.2 The Period of Oscillation

When complex eigenvalues are raised to powers, they produce oscillations in the stage distribution, the period of which is given by

$$\begin{aligned} \mathcal{P}_i &= \frac{2\pi}{\theta_i} && (7.3.9) \\ &= \frac{2\pi}{\tan^{-1}\left(\frac{\Im(\lambda_i)}{\Re(\lambda_i)}\right)}, && (7.3.10) \end{aligned}$$

where θ_i is the angle formed by λ_i in the complex plane and $\Re(\lambda_i)$ and $\Im(\lambda_i)$ are the real and imaginary parts of λ_i, respectively.

The longest-lasting of the oscillatory components is that associated with λ_2. In age-classified models, $\mathcal{P}_2$ is approximately equal to the mean age of childbearing in the stable population (Lotka 1945, Coale 1972). Thus we would expect that perturbations to the stable age distribution would be followed by damped oscillations with a period about equal to the generation time.

In complex life cycles, $\mathcal{P}_2$ cannot be identified with the mean age of reproduction, but it still measures the period of the oscillations contributed by the most important subdominant eigenvalue. Resonance between these oscillations and environmental fluctuations is an important factor in the dynamics of populations in periodic environments (Nisbet and Gurney 1982, Tuljapurkar 1985).

7.3.3 Measuring the Distance to the Stable Stage Distribution

It is useful to be able to measure the distance between two stage distributions, or between an observed stage distribution and the stable stage distribution. We will discuss two such measures here; a third (the Hilbert projective pseudometric) is useful in the study of convergence in variable environments (Golubitsky et al. 1975; MPM, Chapter 13), but has received little actual use in demography.

We wish to measure the distance between $\mathbf{n}(t)$ and the stable population $\mathbf{w}$. Without loss of generality, $\mathbf{w}$ can be scaled so that $\sum_i w_i = 1$, and we can transform $\mathbf{n}(t)$ into $\mathbf{x}(t) = \mathbf{n}(t)/\sum n_i(t)$, so that both vectors describe the proportions of the population in the different stages.

7.3.3.1 Keyfitz's Δ

Keyfitz (1968, p. 47) proposed a measure that is equivalent to

$$\Delta(\mathbf{x}, \mathbf{w}) = \frac{1}{2} \sum_i |x_i - w_i|, \tag{7.3.11}$$

which is a standard measure of the distance between probability vectors. Its maximum value is 1 and its minimum is 0 when the vectors are identical.

7.3.3.2 Cohen's Cumulative Distance

Keyfitz's Δ measures the distance between $\mathbf{n}$ and $\mathbf{w}$ independent of the path by which $\mathbf{n}$ would actually converge to $\mathbf{w}$. Cohen (1979b) proposed two indices that measure the distance between the two vectors along the pathway by which convergence takes place. We know from (7.2.4) that $\mathbf{n}(t)/\lambda^t$ will converge to $c_1\mathbf{w}_1$. Suppose that $\mathbf{n}(0) = \mathbf{n}_0$. Cohen's indices are obtained by accumulating the differences between $\mathbf{n}(t)/\lambda^t$ and $c_1\mathbf{w}_1$:

$$\mathbf{s}(\mathbf{A}, \mathbf{n}_0, t) = \sum_{i=0}^{t} \left(\frac{\mathbf{n}(i)}{\lambda^i} - c_1\mathbf{w}_1 \right) \tag{7.3.12}$$

$$\mathbf{r}(\mathbf{A}, \mathbf{n}_0, t) = \sum_{i=0}^{t} \left| \frac{\mathbf{n}(i)}{\lambda^i} - c_1\mathbf{w}_1 \right|. \tag{7.3.13}$$

The vector $\mathbf{s}(t)$ accumulates the difference between $\mathbf{n}(t)/\lambda^t$ and $c_1\mathbf{w}_1$, whereas $\mathbf{r}(t)$ accumulates the absolute value of those differences.

As a measure of the cumulative distance between an initial population $\mathbf{n}_0$ and its eventual limiting distribution, Cohen proposes calculating the limit as $t \to \infty$ of $\mathbf{s}(t)$ and $\mathbf{r}(t)$, and then adding the absolute values of the elements of the vectors.

$$D_1 = \sum_i \lim_{t\to\infty} |s_i(\mathbf{A}, \mathbf{n}_0, t)| \tag{7.3.14}$$

$$D_2 = \sum_i \lim_{t\to\infty} |r_i(\mathbf{A}, \mathbf{n}_0, t)|. \tag{7.3.15}$$

Cohen actually proposes scaling these indices by multiplying by a constant (50 in his example) to make their magnitude comparable to that of Δ, but this is not necessary.

Cohen gives an analytical expression for the limit in (7.3.14). Let $\mathbf{B} = \mathbf{w}\mathbf{v}'$, and let $\mathbf{Z} = (\mathbf{I} + \mathbf{B} - \mathbf{A}/\lambda)^{-1}$. Then

$$\lim_{t\to\infty} \mathbf{s}(\mathbf{A}, \mathbf{n}_0, t) = (\mathbf{Z} - \mathbf{B})\mathbf{n}_0. \tag{7.3.16}$$

The limit in D_2, however, must be calculated numerically. When $\mathbf{n}_0 = \mathbf{w}$, $c_1 = 1$ and $D_1 = D_2 = 0$. There is no well-defined upper bound for either distance.

Example 7.4 Calculation of D_1 and D_2

Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 0.4271 & 0.8498 & 0.1273 \\ 0.9924 & 0 & 0 \\ 0 & 0.9826 & 0 \end{pmatrix} \tag{7.3.17}$$

(for the 1965 U.S. population with 15-year age intervals, not that it matters here). The stable age distribution and reproductive value, scaled so that $bov_1^*\mathbf{w}_1 = 1$, are

$$\mathbf{w}_1 = \begin{pmatrix} 0.4020 \\ 0.3299 \\ 0.2681 \end{pmatrix} \qquad \mathbf{v}_1 = \begin{pmatrix} 1.4487 \\ 1.1419 \\ 0.1525 \end{pmatrix} \tag{7.3.18}$$

and the observed age distribution, which we take as the initial population, is

$$\mathbf{n}_0 = \begin{pmatrix} 0.4304 \\ 0.3056 \\ 0.2640 \end{pmatrix}. \tag{7.3.19}$$

Thus $c_1 = \mathbf{v}_1'\mathbf{n}_0 = 1.0127$.

The cumulative deviation vectors $\mathbf{s}\left(\mathbf{A}, \mathbf{n}_0, t\right)$ and $\mathbf{r}\left(\mathbf{A}, \mathbf{n}_0, t\right)$ are

t	$\mathbf{r}\left(\mathbf{A}, \mathbf{n}_0, t\right)^{\mathsf{T}}$			$\mathbf{s}\left(\mathbf{A}, \mathbf{n}_0, t\right)^{\mathsf{T}}$		
0	0.0232	0.0285	0.0075	0.0232	−0.0285	−0.0075
1	0.0358	0.0476	0.0306	0.0106	−0.0094	−0.0306
2	0.0424	0.0579	0.0461	0.0172	−0.0198	−0.0151
3	0.0457	0.0633	0.0545	0.0138	−0.0144	−0.0235
4	0.0474	0.0660	0.0589	0.0155	−0.0171	−0.0192
5	0.0483	0.0674	0.0611	0.0147	−0.0157	−0.0214
6	0.0487	0.0681	0.0622	0.0151	−0.0165	−0.0203
7	0.0489	0.0685	0.0628	0.0149	−0.0161	−0.0209
8	0.0490	0.0686	0.0631	0.0150	−0.0163	−0.0206
9	0.0491	0.0687	0.0633	0.0149	−0.0162	−0.0207
10	0.0491	0.0688	0.0633	0.0150	−0.0162	−0.0206

By $t = 10$, both $\mathbf{r}(\cdot)$ and $\mathbf{s}(\cdot)$ have converged to four decimal places. The distance indices are $D_1 = 0.0518$ and $D_2 = 0.1814$.

Unlike Δ, both D_1 and D_2 are functions not only of $\mathbf{n}_0$, but also of the projection matrix $\mathbf{A}$, which determines the pathway by which $\mathbf{n}(t)$ will converge. Consider the following three age-classified projection matrices:

$$\mathbf{A}_1 = \begin{pmatrix} 0.3063 & 0.6094 & 0.0913 \\ 0.9924 & 0 & 0 \\ 0 & 0.9826 & 0 \end{pmatrix}$$

$$\mathbf{A}_2 = \begin{pmatrix} 0 & 0.8784 & 0.1316 \\ 0.9924 & 0 & 0 \\ 0 & 0.9826 & 0 \end{pmatrix}$$

$$\mathbf{A}_3 = \begin{pmatrix} 0 & 0.0641 & 0.9603 \\ 0.9924 & 0 & 0 \\ 0 & 0.9826 & 0 \end{pmatrix}.$$

These matrices all have the same dominant eigenvalue ($\lambda_1 = 1$) and the same stable age distribution,

$$\mathbf{w} = \begin{pmatrix} 0.3370 \\ 0.3344 \\ 0.3286 \end{pmatrix}, \tag{7.3.20}$$

but $\mathbf{A}_2$ and $\mathbf{A}_3$ have reproduction concentrated in the last two and the last age class, respectively.

Suppose that all three populations begin with the common initial vector (7.3.19). The distance between $\mathbf{n}_0$ and $\mathbf{w}$ as measured by Δ is $\Delta = 0.093$, regardless of which projection matrix is used. However, the values of D_1 and D_2 vary dramatically depending on the pattern of reproduction:

Matrix	D_1	D_2
$\mathbf{A}_1$	0.1995	0.6058
$\mathbf{A}_2$	0.1652	1.5646
$\mathbf{A}_3$	0.1066	5.6866

When measured by D_1, the cumulative distance from $\mathbf{n}_0$ to $\mathbf{w}_1$ decreases as reproduction is concentrated in a single age class. When measured by D_2, the cumulative distance increases as reproduction is concentrated. This reflects the increasing oscillations in the age distribution as reproduction is concentrated. Since $\mathbf{s}\left(\mathbf{A}, \mathbf{n}_0, t\right)$ accumulates positive and negative deviations from the stable distribution, oscillations tend to cancel each other out. On the other hand, $\mathbf{r}\left(\mathbf{A}, \mathbf{n}_0, t\right)$ accumulates the absolute value of the deviations, and the oscillations are reflected in larger values of D_2.

7.3.4 Population Momentum

Consider a population growing subject to a set of vital rates, and imagine that you instantly and permanently change those rates to a set of stationary rates; i.e., rates that yield $\lambda_1 = 1$. The population will eventually stop growing, but unless its structure at the moment of the change happens to match the stable stage distribution of the new vital rates, it will not stop immediately. Hence, its final size will differ from its size when the vital rates were changed. Keyfitz (1971b; see Sections 8.6 and 12.3) introduced the term *population momentum* to describe this difference.

Let $t = 0$ denote the instant when the vital rates are changed; the population vector is $\mathbf{n}(0) = \mathbf{n}_0$. The momentum M is

$$M = \lim_{t\to\infty} \frac{\|\mathbf{n}(t)\|}{\|\mathbf{n}_0\|}, \tag{7.3.21}$$

where $\|\mathbf{n}\| = \sum_i |n_i|$ is the total population size. When $M > 1$ the population stabilizes at a size larger than its size at $t = 0$. When $M < 1$, the population shrinks before stabilizing.

Denote the old and new projection matrices by $\mathbf{A}_{\text{old}}$ and $\mathbf{A}_{\text{new}}$, with eigenvalues $\lambda_i^{(\text{old})}$ and $\lambda_i^{(\text{new})}$. Assume that both matrices are primitive. Since $\lambda_1^{(\text{new})} = 1$, we know from (7.1.30) that

$$\begin{aligned}\lim_{t\to\infty} \frac{\mathbf{n}(t)}{\lambda_1^{(\text{new})}} &= \lim_{t\to\infty} \mathbf{n}(t)\\ &= \left(\mathbf{v}_1^{(\text{new})*}\mathbf{n}_0\right)\mathbf{w}_1^{(\text{new})}.\end{aligned}$$

Thus

$$M = \frac{\mathbf{e}^{\mathsf{T}}\left(\mathbf{v}_1^{(\text{new})*}\mathbf{n}_0\right)\mathbf{w}_1^{(\text{new})}}{\mathbf{e}^{\mathsf{T}}\mathbf{n}_0}, \tag{7.3.22}$$

where $\mathbf{e}$ is a vector of 1s.

For age-classified models M can be calculated explicitly in terms of the birth rate, life expectancy, mean age of childbearing, and net reproductive rate (Section 8.6).

The momentum of *declining* populations may be of interest in conservation biology. Consider a declining endangered species (i.e., $\lambda_1^{(\text{old})} < 1$), and suppose (optimistically) that a new management strategy instantaneously brings the vital rates up to replacement level. If $M < 1$, the population will continue to decline before stabilizing, and the extent of this decline might be an important management consideration. It turns out that stage-classified models can yield results that are surprising from an age-classified perspective.

Example 7.5 Population momentum in *Calathea ovandensis*

Calathea ovandensis is a perennial herb that grows in the understory of tropical forests. Horvitz and Schemske (1995) reported size-classified matrices for four sites, over four years, in a Mexican rain forest. Individuals were classified into eight stages: seeds, seedlings, juveniles, pre-adults, and small, medium, large, and extra-large adults. Only the adult stages reproduce.

Here we focus on two of the 16 matrices reported by Horvitz and Schemske (Plot 3, 1982–1983 and Plot 3, 1983–1984; we will call them plots 3-82 and 3-83). The growth rates and net reproductive rates[†] are

	λ_1	R_0
Plot 3-82	1.1572	13.7360
Plot 3-83	0.8876	0.5771

The conditions in plot 3-82 support rapid population growth with a high value of R_0. Suppose that, concerned about a population explosion of *C. ovandensis*, we implemented a pest control strategy that instantly reduced fertility to replacement level by dividing each fertility by R_0. The momentum, calculated from (7.3.22), is $M = 0.107$. Rather than continuing to grow, the population would shrink to about 10 percent of its current size before stabilizing.

The conditions in plot 3-83 would, if maintained, lead to a population decline of almost 12 percent per year and eventual extinction. Suppose that, concerned about this endangered species, we introduced a conservation strategy that instantly increased fertility to replacement level. The resulting momentum is $M = 5.067$. The population would grow to five times its present size before stabilizing.

These results differ from those familiar in age-classified models because of the stable stage distributions before and after the change in the vital rates. These distributions are

Plot 3-82			Plot 3-83		
$\mathbf{w}_1^{(\text{old})}$	$\mathbf{w}_1^{(\text{new})}$	$\frac{\text{old}}{\text{new}}$	$\mathbf{w}_1^{(\text{old})}$	$\mathbf{w}_1^{(\text{new})}$	$\frac{\text{old}}{\text{new}}$
0.875	0.737	1.186	0.942	0.971	0.970
0.104	0.102	1.025	0.031	0.028	1.093
0.009	0.010	0.885	0.004	0.001	5.482
0.002	0.003	0.693	0.007	0.000	36.820
0.002	0.008	0.288	0.015	0.000	54.864
0.000	0.002	0.241	0.002	0.000	62.810
0.002	0.039	0.055	0	0	—
0.005	0.099	0.047	0	0	—

In plot 3-82, the old stable stage distribution has too few individuals in the adult stages (5–8). As a result, the population actually declines as it converges to stationarity. In contrast, in plot 3-83, the old stable stage distribution has too many individuals in the adult stages. The

---

[†] See Section 11.3.4 for calculation of R_0 from stage-classified models.

reproduction by these individuals generates an increase in population size once the vital rates change to stationarity.

Dividing fertilities by R_0 is not the only way to achieve a set of stationary vital rates. Changes in survival or transition probabilities could also be used, and might have different consequences for momentum. It is unlikely that the vital rates could be changed instantaneously to replacement level. Li and Tuljapurkar (1997) have considered the more general problem of an arbitrary change in the vital rates, over some specified period of time, from the old to the new stationary rates.

7.4 Computation of Eigenvalues and Eigenvectors

The numerical calculation of eigenvalues and eigenvectors is not an easy problem. In their compilation of numerical recipes, Press et al. (1986) call it "one of the few subjects covered in this book for which we do *not* recommend that you avoid canned routines." See Wilkinson (1965) and Wilkinson and Renisch (1971) for the details if you really want to try it yourself.

Fortunately, there are now a number of computer packages available for matrix calculations. By far the best is MATLAB.[‡] It is available for most platforms and operating systems, permits both interactive calculations and the writing of complex programs, contains powerful graphics capabilities, and is widely used in the mathematics, physics, and engineering communities. The availability of this software and the power of microcomputers have revolutionized demographic calculations involving matrices.[§]

7.4.1 The Power Method

The dominant eigenvalue of $\mathbf{A}$, and the corresponding eigenvectors, can be easily computed by the power method. If $\mathbf{A}$ is primitive,

$$\lim_{t\to\infty} \frac{\mathbf{A}^t}{\lambda_1^t} = \mathbf{w}_1\mathbf{v}_1^{\mathsf{T}}. \tag{7.4.1}$$

For large t, the columns of $\mathbf{A}^t$ are eventually all proportional to $\mathbf{w}_1$; the rows are all proportional to $\mathbf{v}_1$. The eigenvalue λ_1 can be calculated as the ratio of any entry of $\mathbf{A}^{t+1}$ to the corresponding entry of $\mathbf{A}^t$.

The entries of $\mathbf{A}^t$ may overflow or underflow the computer for large t. This can be avoided by rescaling $\mathbf{A}^t$ at each time step (e.g., divide $\mathbf{A}^t$ by

---

[‡] The MathWorks, Inc., 3 Apple Hill Drive, Natick MA, 01760, U.S.A.; http://www.mathworks.com.

[§] For example, as of this writing, computing all the eigenvalues and eigenvectors of a 100×100 matrix in MATLAB takes less than 0.1 seconds on an ordinary laptop computer. A 500×500 matrix takes only 6.5 seconds.

its maximum entry, so that all values remain between 0 and 1). Wilkinson (1965, Chapter 9) discusses the extension of the power method to find the subdominant eigenvalues and their eigenvectors.

7.5 Mathematical Formulations of the Basic Equation of Population

So far, this book has dealt with what in retrospect can be seen as special cases of a general analysis. Chapter 2 treated the pure death process of the life table, in which births, if considered at all, were just equal to deaths. Chapter 5 used the life table along with an arbitrary rate of increase r to constitute a stable age distribution. Chapters 6 and 7 ascertained the same r from equations incorporating age-specific birth rates as well as the life table.

The general analysis of which all these are parts relates the entire population of each generation or period to the preceding. This can be done in the form of an integral equation, a matrix multiplication, a difference equation, and a partial differential equation.[¶] Because all these methods distinguish individuals on the basis of some i-state variables, they are referred to in general as *structured population models*; for a discussion with many ecological examples see the chapters in Tuljapurkar and Caswell (1997).

These mathematical forms look very different from one another, though they are ultimately equivalent. The purpose of presenting them is not to exhibit the mathematical virtuosity of their several authors, but to take advantage of the fact that some applications are easier with one approach, others with another. This book is not the place to treat the mathematics in detail; that has been done elsewhere (Coale 1972, Pollard 1973, Keyfitz 1968, Rhodes 1940, Tuljapurkar and Caswell 1997). Here the several approaches will be presented, and some observations made on their relation to one another and to the applications that are the subject of this book.

7.5.1 The Lotka Integral Equation

Historically the earliest formulation was in terms of an integral equation solved by Sharpe and Lotka (1911), whose unknown is the trajectory of births and which, under the name of the renewal equation, has become famous in many other contexts. Births $B(t)$ at time t are the outcome of the births a years earlier, where a ranges from about 15 to about 50, say from α to β in general. Newborns of a years earlier, numbering $B(t-a)$,

---

[¶] There is also a large literature using delay-differential equations for the same purpose (Nisbet 1997, Gurney and Nisbet 1998) and a small but growing literature on systems of integrodifference equations (Neubert and Caswell 2000, Easterling et al. 2000).

have a probability $l(a)$ of surviving to time t: those that do survive have a probability $m(a)\,da$ of themselves giving birth in the time interval a to $a+da$; the total of these over ages α to β is $\int_\alpha^\beta B(t-a)l(a)m(a)\,da$, which ultimately must equal current births $B(t)$. But any such system has to have an initial condition to start it off; in the Lotka equation the initial condition is $G(t)$, the number of births due to the women already born at the start of the process. The function $G(t)$ is zero for $t \geqslant \beta$, when all the females alive at $t = 0$ have passed beyond childbearing. Entering the known $G(t)$ and $l(a)m(a)$ in the Lotka integral equation,

$$B(t) = \int_\alpha^\beta B(t-a)l(a)m(a)\,da + G(t), \tag{7.5.1}$$

determines the trajectory $B(t)$.

The method used by Lotka to solve (7.5.1) for $B(t)$ was first to deal with the homogeneous form, in which $G(t)$ is disregarded, and to try $B(t) = e^{rt}$. Entering this value for $B(t)$ and corresponding $B(t-a) = e^{r(t-a)}$ in (7.5.1) without $G(t)$ gives the characteristic equation (6.1.2), which has for the general net maternity function $l(a)m(a)$ an infinite number of roots. This is a more satisfactory way of deriving (6.1.2) than was used above. The left-hand side of (6.1.2), say $\psi(r)$, is large for negative r and diminishes toward zero as r becomes large and positive; we saw that only one of the roots can be real. Suppose that the roots in order of magnitude of their real parts are r_1, r_2, $r_3, \ldots$. The complex roots, such as the pair r_2, r_3, must be pairs of complex conjugates, say $x + iy$, $x - iy$, where x and y are real; if they are not conjugates, they could not be assembled into the real equation (6.1.2).

The homogeneous form [i.e., omitting $G(t)$ from (7.5.1)] is linear, and hence if $e^{r_1 t}$ is a solution so is $Q_1 e^{r_1 t}$, where Q_1 is an arbitrary constant. If a number of such terms are solutions, so is their sum. These considerations provide the general solution to the homogeneous form

$$B(t) = Q_1 e^{r_1 t} + Q_2 e^{r_2 t} + \cdots, \tag{7.5.2}$$

where the r's are the roots of (6.1.2) and the Q's are arbitrary constants. The solution to the nonhomogeneous form (7.5.1) containing $G(t)$ is obtained by selecting values of Q that accord with the births $G(t)$ to the initial population. Lotka showed that these are

$$Q_s = \frac{\displaystyle\int_0^\beta e^{-r_s t} G(t)\,dt}{\displaystyle\int_0^\beta a e^{-r_s a} l(a)m(a)\,da} \qquad s = 1, 2, \ldots. \tag{7.5.3}$$

The constant Q_1 attached to the real term will be important in Chapter 6, where we seek the effect on the trajectory of adding one person aged x.

A less awkward way to solve (7.5.1) (Feller 1941, Keyfitz 1968, p. 128) is by taking the Laplace transform of the members. That of $B(t)$, for example, is

$$B^*(r) = \int_0^\infty e^{-rt} B(t)\,dt$$

and similarly for $G(t)$ and $l(a)m(a) = \phi(a)$, say. The integral on the right of (7.5.1) is a convolution, that is to say, the sum of the arguments of the two functions $B(t-a)$ and $l(a)m(a) = \phi(a)$ in the integrand does not involve a; the transform of a convolution is the product of the transforms of the two functions. The equation in the transforms (distinguished by asterisks) is thus

$$B^*(r) = G^*(r) + B^*(r)\phi^*(r),$$

and is readily solved for the transform $B^*(r)$ of the unknown function $B(t)$:

$$B^*(r) = \frac{G^*(r)}{1-\phi^*(r)}.$$

To invert the right-hand side, expand in partial fractions, using the factors of $1-\phi^*(r)$, obtained from the roots r_1, $r_2, \ldots$, of $\phi^*(r) = 1$, which is the same as (6.1.2). Call the expansion of $B^*(r)$

$$B^*(r) = \frac{Q_1}{r-r_1} + \frac{Q_2}{r-r_2} + \cdots.$$

The Laplace transform of $e^{r_1 t}$ is

$$\int_0^\infty e^{-rt} e^{r_1 t}\,dt = \frac{1}{r-r_1},$$

as appears from integration, so the inverse transform of $1/(r-r_1)$ is $e^{r_1 t}$. Using this fact permits writing the solution in the form of (7.5.2), and the coefficients Q_s are the same as before, that is, (7.5.3). Specializing (7.5.3) to $s = 1$ gives the constant Q_1 for the first term of the solution as

$$Q_1 = \frac{\displaystyle\int_0^\beta e^{-rt} G(t)\,dt}{\kappa}, \tag{7.5.4}$$

where again κ is the mean age of childbearing in the stable population. The roots $r_2, r_3, \ldots$, all have real parts less than r_1, so the terms involving them become less important with time, and the birth curve $B(t)$ asymptotically approaches $Q_1 e^{r_1 t}$. This statement holds under general conditions on $\phi(a)$ analogous to those of Section 1.10.

What the solution (7.5.2) amounts to is a real term that is an exponential, plus waves around this exponential, of which the demographically interesting one corresponds to the pair of complex roots of largest absolute value, has the wavelength of the generation, and accounts for the echo effect: other things being constant, a baby boom in one generation is followed by a secondary baby boom in the next generation.

7.5.2 The Leslie Matrix

The age-classified Leslie matrix, from which the stage-classified models in this chapter evolved, can be traced back to Whelpton's (1936) presentation of what he called the components method of population projection, in which an age distribution in 5-year age groups is "survived" along cohort lines, and births less early childhood deaths are added in each cycle of projection. The method was also used by Cannan (1895) and Bowley (1924) after him.

The dominant eigenvalue of a Leslie matrix, using a k-year projection interval, corresponds to e^{kr_1} of the Lotka formulation with time measured in years. With finite age groups, in the usual finite approximation, the population grows somewhat faster on the Leslie than on the Lotka projection; for instance, over a 5-year period, using Mexican data for 1966, λ_1 was 1.1899 whereas e^{5r_1} was 1.1891. For a low-increase country like the United States the difference does not show by the fourth decimal place, and it vanishes altogether when the projection interval is made small. In fact, as the interval becomes small, the two models become identical (Keyfitz 1968, Chapter 8).

7.5.3 The Difference Equation

A third way of looking at the population trajectory, developed like the matrix during the 1930s and 1940s, by Thompson (1931), Dobbernack and Tietz (1940, p. 239), Lotka (1948, p. 192), and Cole (1954, p. 112), is in terms of a difference equation. A secondary treatment is found in Keyfitz (1968, p. 130). Although the approach is now mainly of historical interest, it is still occasionally used (e.g., Croxall et al. 1990).

Consider one girl baby together with the series expected to be generated by her at 5-year intervals, say u_0, u_1, $u_2, \ldots,$ where $u_0 = 1$. The simplest way to describe the model is to bunch the person-years lived into points at 5-year intervals. The u_i will at first decrease, corresponding to the probability that the girl will die during the time before she begins to bear children; then they will start to increase, and they will increase further with the approach to u_8 just 40 years later, when her children start to bear. The series generated by the girl now alive includes the probability that she will live for 15 years and then have a child, say a probability of f_3, that she will live 20 years and then have a child, f_4, and so on. After n periods of 5 years the girl's descendants (including herself if still alive) are u_n. The u_n must be equal to the chance of her living that long and having a child then, $f_n u_0$, plus the chance of her having lived $n - 1$ periods and having a child that would now have given rise to u_1, and so on. In sum,

$$u_n = f_n u_0 + f_{n-1} u_1 + f_{n-2} u_2 + \cdots \tag{7.5.5}$$

for all n. In this series the f are known quantities corresponding to the $l(a)m(a)$ of the Lotka form, or to the $P_i F_i$ of the matrix; the u series is the unknown function corresponding to $B(t-a)$.

Just as the integral equation was solved by the Laplace transform, so (7.5.5) is solved by generating functions. Multiply (7.5.5) for u_n by s^n, add all such equations from $n = 0$ to $n = \infty$, and then express the result in terms of

$$U(s) = u_0 + u_1 s + u_2 s^2 + \cdots,$$
$$F(s) = f_0 + f_1 s + f_2 s^2 + \cdots,$$

so that the set of equations 7.5.5 amounts to

$$U(s) = 1 + U(s)F(s),$$

which is readily solved to give

$$U(s) = \frac{1}{1 - F(s)}. \tag{7.5.6}$$

Dividing out the right-hand side would only generate again the series with which we started. But first finding the roots of $1 - F(s) = 0$, a further form of the characteristic equation, using the roots to break down the right-hand side of (7.5.6) into partial fractions of the type $d_i/(s - s_i)$, and then expanding each of the partial fractions, we get a series in powers of s that corresponds to (7.5.2) in the Lotka solution.

The descriptions read as though the Lotka model is backward looking, since it finds the relation between the present generation and the preceding one, whereas the Leslie model and the set of difference equations are forward looking, portraying the continuance of present rates into the future. These differences of direction are superficial, however, and the mathematical outcomes are essentially identical.

7.5.4 The McKendrick–von Foerster Equations

A fourth approach, due to McKendrick (1926, Kermack and McKendrick 1927) and von Foerster (1959), uses partial differential equations, with both age and time continuous variables. It is most easily visualized in terms of the Lexis diagram (Figure 2.1), redrawn in the present context as Figure 7.7. If $P(a,t)$ is the population (say of females) at age a and time t, the female population of age $a + \Delta a$ at time $t + \Delta t$ is $P(a + \Delta a, t + \Delta t)$. If $\Delta a = \Delta t$ the latter includes the same individuals as were counted in $P(a,t)$, only subject to deductions for mortality (as well as for emigration if one wishes, but the present section excludes migration). The equation of change involving

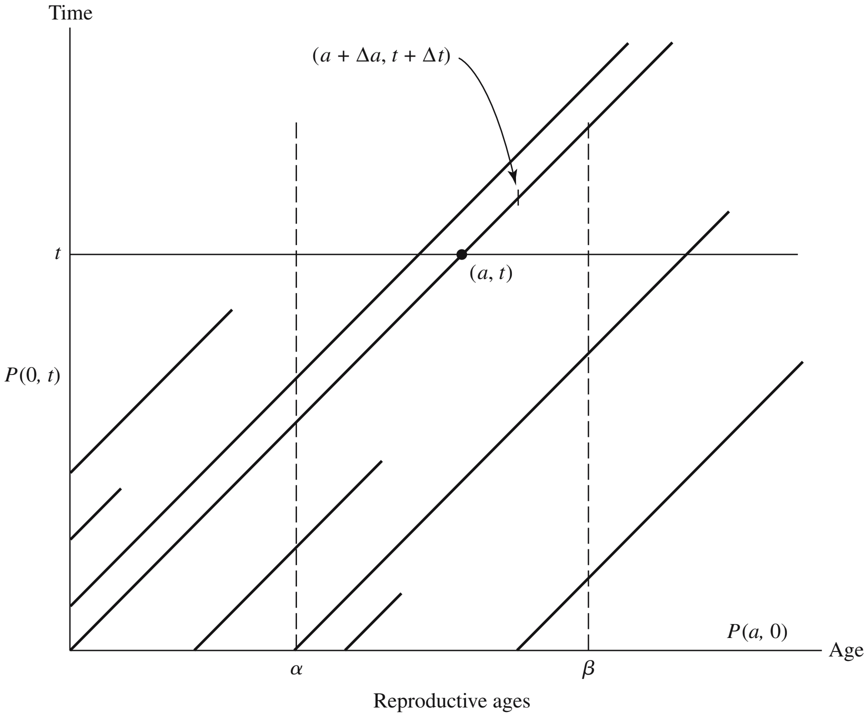


Figure 7.7. Lexis diagram showing boundary distributions $P(a, 0)$ along age axis given at the outset, and $P(0, t)$ along time axis generated by (7.5.8) and (7.5.9); (7.5.8) fills out the interior of the quadrant.

mortality $\mu(a)$, a function of age but not of time, is

$$P(a + \Delta a, t + \Delta t) = P(a, t) - \mu(a)P(a, t)\,\Delta t. \tag{7.5.7}$$

Expanding $P(a + \Delta a, t + \Delta t)$ by Taylor's theorem for two independent variables and canceling $P(a, t)$ from both sides leaves

$$\frac{\partial P(a,t)}{\partial t}\Delta t + \frac{\partial P(a,t)}{\partial a}\Delta a = -\mu(a)P(a,t)\,\Delta t.$$

Dividing by Δa, which is equal to Δt, we have

$$\frac{\partial P(a,t)}{\partial t} + \frac{\partial P(a,t)}{\partial a} = -\mu(a)P(a,t). \tag{7.5.8}$$

This is the McKendrick–von Foerster partial differential equation for one sex, for all values of $0 < a < \omega$, where ω is the oldest age to which anyone lives.

Births enter as a boundary condition at age zero:

$$P(0,t) = \int_{\alpha}^{\beta} P(a,t)m(a)\,da, \tag{7.5.9}$$

where α is the youngest age of childbearing and β the oldest, and $m(a)$ is the age-specific birth rate, supposed to be invariant with respect to time.

This derivation assumes that a represents age, but these models have been extended to classification by size or physiological condition by Sinko and Streifer (1967, 1969), Oster (1976), and van Sickle (1977a,b). For density-dependent versions see Gurtin and MacCamy (1974), Oster and Takahashi (1974), and Oster (1976). The monograph of Metz and Diekmann (1986) provides a complete but mathematically challenging treatment; de Roos (1997) provides the best introduction for beginners.

In the general model, let $g(a,t)$ be the "growth rate", that is, the rate at which individuals increase in whatever characteristic a represents. Then (7.5.8) becomes

$$\frac{\partial P(a,t)}{\partial t} + \frac{\partial g(a,t)P(a,t)}{\partial a} = -\mu(a)P(a,t). \tag{7.5.10}$$

For the special age-classified case, $g(a) = 1$.

7.5.5 A Common Structure

The common structure of the four presentations may not stand out conspicuously, but each has the following.

1. An initial age distribution.
2. Provision for mortality; in the first three methods this is the survival probability $l(x)$ or L_x of the life table, while in the von Foerster method it is the age-specific death rate $\mu(a)$; if age and time move forward together we have a cohort, and the partial differential equation whose solution is
$$l(x) = \exp\left[-\int_0^x \mu(a)\,da\right].$$
3. Provision for reproduction, presented as the first row of the matrix **A**, and as the boundary condition (7.5.9) in the partial differential equation.
4. A characteristic equation, whose real or positive root is the intrinsic rate or its exponential, obtained from the birth boundary condition in the partial differential equation.
5. Right eigenvectors, of which the positive one is the stable age distribution.
6. Left eigenvectors, of which the positive one contains the reproductive values of the several ages or stages (Chapters 8 and 9).

In principle any property of the deterministic population process under a fixed regime is obtainable from any of the four forms, and numerical differences disappear if the interval of time and age is small enough.

8

Reproductive Value from the Life Table

When a woman of reproductive age is sterilized and so has no further children, the community's subsequent births are reduced. When a woman dies or otherwise leaves the community, all subsequent times are again affected. Our formal argument need make no distinction between emigration and death, between leaving the country under study for life and leaving this world altogether. A single theory answers questions about the numerical effect of sterilization, of mortality, and of emigration, all supposed to be taking place at a particular age x. By means of the theory we will be able to compare the demographic results of eradicating a disease that affects the death rate at young ages, say malaria, as against another that affects the death rate at older ages, say heart disease.

A seemingly different question is what would happen to a rapidly increasing population if its couples reduced their childbearing to bare replacement immediately. The period net reproduction rate R_0, the number of girl children expected to be born to a girl child just born, would equal 1 from then on, and ultimately the population would be stationary. But the history of high fertility has built up an age distribution favorable to childbearing, and the ultimate stationary total will be much higher than the total at the time when the birth rate dropped to bare replacement. The amount by which it will be higher is calculable, and by the same function—reproductive value—that is used for problems of migration and changed mortality.

8.1 Concept of Reproductive Value

Without having these particular problems in mind, Fisher (1930, p. 27) developed a fanciful image of population dynamics that turns out to provide solutions to them. He regarded the birth of a child as the lending to him of a life, and the birth of that child's offspring as the subsequent repayment of the debt. Apply this to the female part of the population, in which the chance of a girl living to age a is $l(a)$, and the chance of her having a girl between ages a and $a + da$ is $m(a)\,da$, so that the expected number of children in the small interval of age specified is $l(a)m(a)\,da$. This quantity added through the whole of life is what was defined as the net reproduction rate R_0 in Section 6.1:

$$R_0 = \int_{\alpha}^{\beta} l(a)m(a)\,da,$$

where α is the youngest age of childbearing and β the oldest. The quantity R_0 is the expected number of girl children by which a girl child will be replaced; for the population it is the ratio of the number in one generation to the number in the preceding generation, according to the given $l(a)$ and $m(a)$ (see Chapter 9 for the generalization to stage-classified models).

Fisher's image discounts the future, at a rate of interest equal to the intrinsic rate r of Section 6.1. The value of 1 dollar, or one child, discounted back through a years at annual rate r compounded momently is e^{-ra}; therefore the value of $l(a)m(a)\,da$ children is $e^{-ra}l(a)m(a)\,da$, as in the financial calculations of Section 2.5. The present value of the repayment of the debt is the integral of this last quantity through the ages to the end of reproduction. Thus the debt that the girl incurs at birth is 1, and the discounted repayment is the integral $\int_{\alpha}^{\beta} e^{-ra}l(a)m(a)\,da$. If loan and discounted repayment are to be equal, we must have

$$1 = \int_{\alpha}^{\beta} e^{-ra}l(a)m(a)\,da,$$

and this is the same as the characteristic equation (Lotka 1939, p. 65, and (6.1.2)), from which the r implied by a net maternity function $l(a)m(a)$ is calculated. The equation can now be seen in a new light: the equating of loan and discounted repayment is what determines r, r being interpretable either as the rate of interest on a loan or as Lotka's intrinsic rate of natural increase.

The loan-and-repayment interpretation of the characteristic equation suggests calculating how much of the debt is outstanding by the time the girl has reached age $x < \beta$. This is the same as the expected number of subsequent children discounted back to age x. Her expected births in the interval a to $a + da$, $a > x$, are $[l(a)/l(x)]m(a)$; and if these births are discounted back $a - x$ years, her debt outstanding at age x is

$$v(x) = \int_x^\beta e^{-r(a-x)} \frac{l(a)}{l(x)} m(a)\, da$$

or, as Fisher (1930) wrote,

$$v(x) = \frac{1}{e^{-rx} l(x)} \int_x^\beta e^{-ra} l(a) m(a)\, da, \tag{8.1.1}$$

where $v(x)$ will be called reproductive value at age x. Evidently $v(0) = 1$, and, for $x > \beta$, $v(x) = 0$.

For his studies in genetics Fisher needed to know the extent to which persons of given age (say x), on the average contribute to the births of future generations. This seemingly different question is answered by a function proportional to $v(x)$; its value can be established at $v(x)/\kappa$, where, as in Section 6.1,

$$\kappa = \int_\alpha^\beta a e^{-ra} l(a) m(a)\, da; \tag{8.1.2}$$

that is, κ is the mean age of childbearing in the stable population. The basic proposition is that the addition of a girl or woman aged x to the population at time zero adds an expected $v(x)e^{rt}/\kappa$ baby girls at time t, always supposing the continuance of the same regime of fertility and mortality. The simplest derivation of this takes off from the real term of solution 7.5.2 to the Lotka renewal equation. A self-contained version is provided in Section 8.9.

8.1.1 Reproductive Value from the Lotka Integral Equation

One Woman Aged x. The continuous model of Section 7.5 provides the curve of descendants of an arbitrary initial age distribution, and its asymptotic trajectory is the real term $Q_1 e^{r_1 t}$ of (7.5.2), the value of Q_1 being given by (7.5.4). For a distribution consisting of one woman aged x, disregarding questions of continuity and of random variation, we find that the children expected at time t to $t + dt$ are $[l(x+t)/l(x)]m(x+t)\, dt$, which is therefore the function $G(t)$. Entering it in (7.5.4), that is, in $Q_1 = \int_0^\beta e^{-rt} G(t)\, dt/\kappa$, we have

$$Q_1 = \frac{\displaystyle\int_0^{\beta - x} e^{-rt} [l(x+t)/l(x)] m(x+t)\, dt}{\kappa}, \tag{8.1.3}$$

which except for the divisor κ is identical to $v(x)$ of (8.1.1), giving the discounted value of the expected future births to a woman aged x. In the special case of a baby just born, $x = 0$; and, by virtue of (8.1.3) and the characteristic equation (6.1.2), $Q_1 = 1/\kappa$.

Stable Age Distribution. The same constant Q_1 can be readily evaluated for a population of unity having the stable age distribution $be^{-ra}l(a)$. We can guess in advance from the nature of stability that the asymptotic population will be e^{rt} and the births be^{rt}, so Q_1 must equal b for this case.

The proof seems simplest if we start by calculating the total expected reproductive value of a stable population:

$$\int_0^\beta be^{-rx}l(x)v(x)\,dx,$$

then cancel the $e^{-rx}l(x)$ with the denominator of $v(x)$ of (8.1.1) to find

$$b\int_0^\beta \int_x^\beta e^{-ra}l(a)m(a)\,da\,dx,$$

and finally integrate by parts to obtain

$$b\int_0^\beta ae^{-ra}l(a)m(a)\,da = b\kappa$$

as the reproductive value of a population of unity having a stable age distribution. The constant Q_1 is this total reproductive value divided by κ, that is, $b\kappa/\kappa = b$, as suggested by intuition.

Arbitrary Age Distribution. A more general statement can be made. Let $p(x)$ be the age distribution as a density function; i.e., so that the number of individuals between ages a and $a+5$, say, is ${}_5N_a = \int_a^{a+5} p(x)dx$. Whatever the initial age distribution $p(x)$ of a closed population acted on by fixed rates of birth and death, its births have an asymptotic trajectory $Q_1e^{r_1t}$ where Q_1, defined by (7.5.4), is equal to $\int_0^\beta p(x)v(x)\,dx/\kappa$, that is, the sum of reproductive value in the population divided by the mean age of childbearing.

To see this, note that the total reproductive value of $p(x)$ is

$$V = \int_0^\beta p(x)v(x)\,dx = \int_0^\beta p(x)\int_0^{\beta-x} e^{-rt}\frac{l(x+t)}{l(x)}m(x+t)\,dt\,dx. \quad (8.1.4)$$

But this is the same as the numerator of Q_1 in (7.5.4), where $G(t)$ is the number of children expected to be generated by the initial population $p(x)$. For the number of those children born at time t will be

$$G(t) = \int_0^\beta p(x)\frac{l(x+t)}{l(x)}m(x+t)\,dx, \quad (8.1.5)$$

and multiplying by e^{-rt} and then integrating over t gives the double integral in (8.1.4). This demonstrates that $Q_1 = V/\kappa$, where V is the number of women, each weighted by the $v(x)$ for her age x. Once again, $1/\kappa$ of the present value of the balance outstanding by age x in the hypothetical loan is equal to the contribution of a woman aged x to the ultimate trajectory, and both the loan and the trajectory are additive for a group of women of arbitrary ages.

The foregoing proof depends on the solution of the integral equation. The proof in Section 8.9, on the other hand, stands on its own feet. The same result can be derived using matrix formulations, without assuming age-classification (Chapter 9).

Once we know the effect on the birth trajectory of adding one girl and assume a fixed birth rate b, we can obtain the effect on the population trajectory by dividing by b. This *is* obvious, for since the birth rate b is B/N, births divided by population, the population must be $N = B/b$, births divided by the birth rate. Hence the effect of adding a girl or woman aged x is to add $v(x)e^{rt}/\kappa$ to ultimate births and $v(x)e^{rt}/\kappa b$ to ultimate population.

To obtain some intuitive feeling for the reason why the effect of one child just born on the ultimate birth trajectory is to raise it by $v_0 e^{rt}/\kappa = e^{rt}/\kappa$, rather than just e^{rt} or some other value, suppose that all children are born at the same maternal age and that this age is κ. Then the birth of an additional girl child now will result in R_0 girl children in κ years, R_0^2 children in 2κ years, and R_0^n in the nth generation, where R_0 is, as before, the net reproduction rate; that is to say, a child born now outlines a birth curve (Fig. 8.1) rising in the ratio of R_0 every κ years, but with births occurring only at κ-year intervals. In other words, the curve outlined gives the number of births per κ years resulting from one birth at the outset; it is reduced to births per year by dividing by $\kappa : e^{rt}/\kappa$. This argument is at best heuristic; the result applies much more generally than to the primitive model in which all births occur at the same maternal age.

8.1.2 Numerical Calculation

The expression for $v(x)$ in (8.1.1) applies to exact age x, and an approximation analogous to that customarily made for the stable age distribution is

$$v_x \approx \frac{e^{-2\frac{1}{2}r}{}_5L_xF_x + e^{-7\frac{1}{2}r}{}_5L_{x+5}F_{x+5} + \cdots}{l_x}. \tag{8.1.6}$$

This, with numerator and denominator multiplied by e^{-rx}, is shown in Table 8.1 for Mauritius females, 1966. Figure 8.2 shows the curves of $v(x)$ for Mauritius, the United States, and Hungary, taken from Keyfitz and Flieger (1971, pp. 315, 361, and 443).

For the average reproductive value for the age interval x to $x+4$ at last birthday the recurrence formula

$${}_5V_x = \tfrac{5}{2}F_x + \frac{e^{-5r}{}_5L_{x+5}}{{}_5L_x}\left(\tfrac{5}{2}F_{x+5} + {}_5V_{x+5}\right) \tag{8.1.7}$$

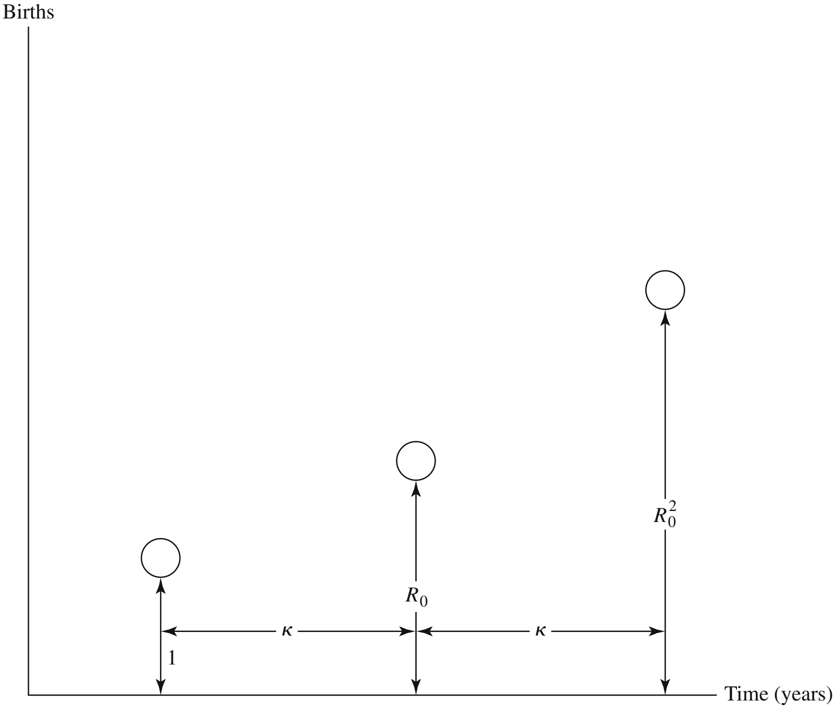


Figure 8.1. Effect of one birth if all children are born at age κ of mother.

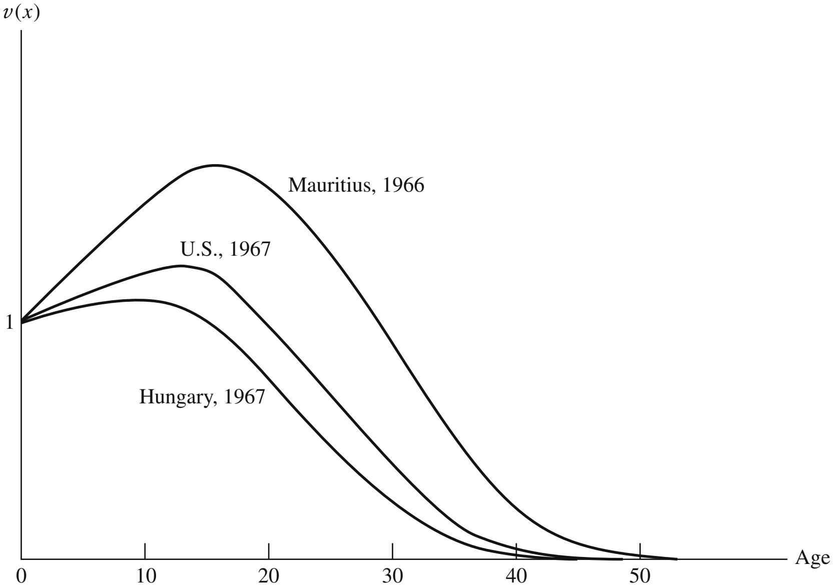


Figure 8.2. Curves of reproductive value for females of three countries.

Table 8.1. Calculation of reproductive values for females of exact ages 0, 5, 10, . . . , 50, Mauritius, 1966; $r = 0.0305371$

Age x	${}_5L_xF_x$ (1)	$e^{-(x+2\frac{1}{2})r}$ (2)	$\sum_{y=x}^{\beta-5} e^{-(y+2\frac{1}{2})r}\,{}_5L_yF_y$ (3)	$e^{-rx}l_x$ (4)	v_x (5) = (3)/(4)
0		0.92650	1	1	1
5		0.79531	1	0.78463	1.2745
10	0.0014	0.68269	1	0.66822	1.4965
15	0.1858	0.58602	0.99907	0.57152	1.7481
20	0.6236	0.50304	0.89019	0.48700	1.8279
25	0.6061	0.43181	0.57649	0.41212	1.3988
30	0.4730	0.37067	0.31477	0.34841	0.9034
35	0.3239	0.31818	0.13944	0.29392	0.4744
40	0.1201	0.27312	0.03639	0.24719	0.1472
45	0.0146	0.23445	0.00358	0.20816	0.0172
50	0.0008	0.20125	0.00016	0.17341	0.0009

Source: Net maternity function from Keyfitz and Flieger (1971, p. 315).

Table 8.2. Values of ${}_5V_x$, the Fisher reproductive value of females aged x to $x+4$ at last birthday, and ${}_5V_x/(b\kappa)$, the coefficient of the amount by which population at time t is raised by one added person aged x to $x+4$ at time zero, Mauritius, 1966

Age	${}_5V_x$	${}_5V_x/(b\kappa)$
0–4	1.159	1.092
5–9	1.381	1.301
10–14	1.618	1.524
15–19	1.783	1.679
20–24	1.611	1.517
25–29	1.151	1.084
30–34	0.690	0.650
35–39	0.312	0.294
40–44	0.083	0.078
45–49	0.009	0.008

$v_0 = 1$

$$v_0/(b\kappa) = \frac{1}{(0.03889)(27.30)} = 0.942$$

Source: Keyfitz and Flieger (1971), p. 315.

provides a reasonable approximation. However, the ${}_5V_x$ of Table 8.2 was calculated, not in this way, but by the method (easier if a computer is available) of finding the left eigenvector of the projection matrix in Sections 7.1 and 7.5; see Section 9.1 for the connection between reproductive value and eigenvectors.

Evidently total reproductive value of a population of arbitrary age distribution acted on by a fixed regime increases at rate r in the short as well as the long run. Such a statement is conspicuously not true for the size of the total population, whose increase in the short run depends on its age distribution. Both births and population acted on by a fixed regime *ultimately* go into an exponential trajectory with parameter r; the total of reproductive values *immediately* follows an exponential trajectory (Fisher 1930, p. 30).

The above like other pieces of theory in this book can be justified only by its ability to answer demographic questions. The following section deals with the first of a series of such questions.

8.2 Ultimate Effect of Small Out-Migration Occurring in a Given Year

When people leave a crowded island like Barbados or Java, they make life somewhat easier for those who remain behind, assuming that the rates of mortality and fertility do not change as a result of their departure.

The age at which they leave determines the effect. Departures of persons who are already past the ages of reproduction cannot influence the ultimate population trajectory; the effect of their leaving is only the subtraction of the person-years they themselves will live from the time of departure to death.

A one-time departure of a person of reproductive age or below will lower the expected population trajectory, but cannot change its rate of climb as long as the age-specific rates of birth and death remain unchanged. In symbols, if the ultimate trajectory is Qe^{rt}, a one-time departure of an individual or a group under age β can lower Q but will not alter r. It follows from the theory of Section 8.1 that a female of age x leaving reduces the female births at time t by $v(x)e^{rt}/\kappa$ and the female population at time t by $v(x)e^{rt}/(b\kappa)$, where we take t to be large. Thus the change in Q for population due to the departure of one female aged x is $\Delta Q = -v(x)/(b\kappa)$.

We are still on the one-sex model and suppose female dominance, that is, that births are determined by the number of females at the several ages and not by the number of males. This would be true if males were in the majority or polygyny prevailed or artificial insemination were applied. The extension of the ideas of the present chapter to a genuine two-sex model depends on behavioral variables not readily incorporable in demographic theory.

The effect of a one-time bulge in births follows readily. With ΔB extra births in a given year the birth trajectory would be raised $e^{rt}\,\Delta B/\kappa$, and the population trajectory would be raised this amount divided by the birth rate b.

Does a female of random age affect the ultimate population more or less than a girl baby? The former, entering at time zero, raises the population at time t an expected e^{rt}, while the latter raises it by $e^{rt}/(b\kappa)$. The mean age of childbearing κ is never very far from 27, and the reciprocal of 27 is 0.037. For low-fertility populations b is considerably less than $1/\kappa \approx 0.037$; hence a baby has more effect than a female between zero and ω randomly chosen from the stable age distribution. For high-fertility populations, on the other hand, b is greater than $1/\kappa$ and a baby has less expected effect than a randomly selected female. Thus for Mexico the departure of a woman of random age has more effect than averting one birth; for the United States averting a birth has more effect.

The same technique can be used to find a variety of equivalents. By what amount, for example, would births have to drop in a particular year to offset an immigration of 1000 women aged 15 to 19 in the same year? The population at distant time t resulting from 1000 women aged 15 to 19 is $1000{}_5V_{15}e^{rt}/(b\kappa)$. The population from B births at time t is $Bv_0e^{rt}/(b\kappa)$. Equating these two expressions, we obtain

$$B = \frac{1000{}_5V_{15}}{v_0}$$

as the required equivalent number of births. From the Mauritius information in Table 8.2 we have, since ${}_5V_x$ is normed to $v_0 = 1$,

$$B = 1000{}_5V_{15} = 1783.$$

In any one year (or other period) a drop of 1783 female births would be required to offset the immigration of 1000 women aged 15 to 19 at last birthday.

8.3 Effect of Continuing Birth Control and Sterilization

Suppose that a few women each year resort to birth control when they are of age a, and this occurs year after year, so that the birth rate $m(a)$ is permanently lowered for age a, but all other age-specific birth rates remain unaltered. If the change in the age-specific birth rate in the single year of age a is $\Delta m(a)$, a quantity that will carry a minus sign for decrease in $m(a)$, the change in the intrinsic rate of the population is determined by finding the derivative $dr/dm(a)$ in the characteristic equation $\int_\alpha^\beta e^{-rx}l(x)m(x)\,dx = 1$ as in Section 6.3, and for finite increments Δr and $\Delta m(a)$ is approximately

$$\Delta r \approx -\frac{e^{-ra}l(a)\,\Delta m(a)}{\kappa}, \tag{8.3.1}$$

the same as (6.3.8). The result depends on $\Delta m(a)$ being small enough so that e^{-ra}, as well as κ, is substantially unaffected. Subject to this same condition, we can find the combined effect of small increments at two different ages, say a and $a+1$. The effect on r will be approximately the sum of the Δr for $\Delta m(a)$ and that for $\Delta m(a+1)$, and similarly for any other groups of ages. This type of perturbation analysis of the rate of increase is expanded on in Chapter 13.

Now suppose a permanent change in $m(a)$ for all a from age x onward, so that the new birth function is $m(a)$, $a < x$, and $(1-f)m(a)$, $a \geqslant x$, f being a small positive or negative fraction. This could be the result of sterilization becoming the custom at age x, or of the fraction f of women at age x turning to conventional birth control in order to avoid all further children. If f is small we can enter $-fm(a)$ for $\Delta m(a)$ in the preceding display, and find the total effect Δr by adding the Δr's for the several ages:

$$\Delta r = -\frac{f\displaystyle\int_x^\beta e^{-ra}l(a)m(a)\,da}{\kappa}. \tag{8.3.2}$$

The integral here will be recognized as the same one that turned up in $v(x)$ of (8.1.1). Entering $v(x)$ makes this

$$\Delta r = -\frac{fe^{-rx}l(x)v(x)}{\kappa}.$$

In words, the decrease by the fraction f of fertility rates for all ages above x lowers the intrinsic rate by $v(x)$ multiplied by $fe^{-rx}l(x)/\kappa$. Remembering that $be^{-rx}l(x)$ is the fraction of the population at age x, where at this point it is convenient to make x integral and have it represent exact ages $x-\frac{1}{2}$ to $x+\frac{1}{2}$, we can say that the decrease in r is $f/(b\kappa)$ times the fraction of the population aged x, times the reproductive value at age x. More simply, the integral in (8.3.2) is the fraction of current mothers aged x and over, so (8.3.2) tells us that the effect on r is equal to the fraction f dropping out of childbearing, times the fraction of babies born to women aged x and over, divided by the mean age of childbearing. Designating by b_x the fraction of births occurring to mothers aged x and over, (8.3.2) can be written

$$\Delta r = -\frac{fb_x}{\kappa}.$$

Conventional birth control, sterilization, or mortality, if they take place year after year can lower births to women over age x by a small fraction f, and if they do the rate of increase r is reduced by f times the fraction b_x of children born to women aged x and older, divided by the mean age of childbearing.

The preceding discussion also covers the consequences of a fall in the death rate. Suppose that the rate at ages $x-\frac{1}{2}$ to $x+\frac{1}{2}$ goes from $\mu(x)$ to $\mu(x)+\Delta\mu(x)$ and remains at that level, or (what is practically the same) that $\Delta\mu(x)/\delta$ is permanently added to the density $\mu(x)$ over a narrow age interval δ. Then all the results of this section apply. The derivation first finds the effect of $\Delta\mu(x)/\delta$ on $l(x)$, using the approximate formula

$e^{-\Delta\mu(x)/\delta} \approx 1 - \Delta\mu(x)/\delta$. Thereafter the derivation is the same as for (8.3.1), since $l(a)$ and $m(a)$ enter symmetrically into the characteristic equation.

8.4 Large Change in Regime

So far only small changes have been discussed. We now ask the same question in reference to an arbitrary, possibly large, change: if birth control is applied by women aged x and above, what fraction of births must they avoid in order to change the rate of increase from r to $r + \Delta r$?

Suppose that in every cohort women aged x and higher apply birth control to the point where they reduce their age-specific rates by the fraction f of what they were before; sterilization of f of the women reaching age x would have this effect. The original intrinsic rate of increase was found by solving for r in the characteristic equation. The equation for the new rate of increase $r + \Delta r$ breaks down into two parts:

$$\int_\alpha^\beta \exp\Big[-(r+\Delta r)a\Big]\,l(a)m(a)\,da \\ -f\int_x^\beta \exp\Big[-(r+\Delta r)a\Big]\,l(a)m(a)\,da = 1, \qquad (8.4.1)$$

where we suppose $a \leqslant x \leqslant \beta$. Equation 8.4.1 could be solved for x if f and $r + \Delta r$ were given, or for $r + \Delta r$ if x and f were given. A simple explicit solution is available for f, the fraction of decrease above the given age x that will suffice to change the intrinsic rate from r to $r + \Delta r$:

$$f = \frac{\displaystyle\int_\alpha^\beta \exp\Big[-(r+\Delta r)a\Big]\,l(a)m(a)\,da - 1}{\displaystyle\int_x^\beta \exp\Big[-(r+\Delta r)a\Big]\,l(a)m(a)\,da}. \qquad (8.4.2)$$

Result 8.4.2 depends in no way on Δr being small. [Find its limiting value when Δr is small.]

The numerator of (8.4.2) is bound to be positive for $\Delta r < 0$, corresponding to the birth control formulation in which f is defined as positive and birth rates go from $m(a)$ to $(1 - f)m(a)$. In the special case where the desired $r + \Delta r = 0$ we would have the simpler form

$$f = \frac{\displaystyle\int_\alpha^\beta l(a)m(a)\,da - 1}{\displaystyle\int_x^\beta l(a)m(a)\,da} = \frac{R_0 - 1}{\displaystyle\int_x^\beta l(a)m(a)\,da}. \qquad (8.4.3)$$

The f of (8.4.3) is the fraction by which women aged x and over must reduce fertility to bring the rate of population increase r down to zero. The age x is arbitrary but is required to stay within certain limits if $0 < f < 1$. For data for Colombia, 1965, one observes that no reduction of fertility in women 30 and over could bring stationarity if ages under 30 retained existing rates, for we have $R_0 = 2.267$ and $\int_{30}^{50} l(a)m(a)\,da = 1.001$, and hence a drop to $R_0 = 1$ would not occur even if all fertility above age 30 disappeared.*

One would have thought that a girl child would contribute the same amount to the ultimate trajectory irrespective of the age of her mother; all babies start at age zero, after all. The expression $\Delta r = e^{-ra}l(a)\,\Delta m(a)/\kappa$ in (8.3.1) is consistent with this view, for it says that the effect of a small change $\Delta m(a)$ in the age-specific birth rate is proportional to $e^{-ra}l(a)$, that is, proportional to the number of women of that age in the stable population; this has to be right, in that a given change in the birth rate will alter the number of babies in proportion to the number of women to whom the change is applied. The expression for Δr in (8.3.1) supposes that $\Delta m(a)$ is small enough not to affect κ, the mean age of childbearing.

But for the ultimate effect of a large change that takes place generation after generation, it does make a difference whether women are young or old when they have their children. Avoiding births at age 40 is not as effective as avoiding them at age 20, because of the more rapid turnover of a population in which births occur to younger mothers. This is taken into account in (8.4.2) and (8.4.3).

8.5 Emigration as a Policy Applied Year After Year

Each year some inhabitants of Java go to Sumatra under an official transmigration program that has been government policy for two-thirds of a century. The authorities have always recognized that the amount of relief provided to Java depends on the age of the migrants at the time of their out-migration, and that young couples are the ideal ones to go, but they have tended to exaggerate the effect. Widjojo (1970) shows realistic population projections under alternative assumptions about the rate of movement, from which the consequences of different policies can be seen.

---

*The net reproductive rate has come to play a central role in modelling epidemic diseases, treated as a problem in pathogen demography. In this context, R_0 is the expected number of secondary cases caused by a single infected individual over its entire infectious period. Whether R_0 is greater or less than 1, when the population consists entirely of susceptible hosts, determines whether the disease will spread or die out. Calculations essentially identical to those used here to determine the amount by which fertility must be reduced in order to stop population growth are used to calculate the level of vaccination that must be imposed to stop the spread of a disease. See Diekmann et al. (1990), Anderson and May (1991), and Diekmann and Heesterbeek (2000).

In this section we will examine one aspect of policy only: the effect of the age of the migrants on the ultimate rate of increase of the population.

We can express (8.4.2) in terms of a generalization of reproductive value. In this general reproductive value, say $v_{x,\bar{r}}$, future children are discounted, not at the intrinsic rate r of the observed population, but at the rate $\bar{r}$ at which the emigration policy is to aim:

$$v_{x,\bar{r}} = \int_0^{\beta-x} e^{-\bar{r}t} \frac{l(x+t)}{l(x)} m(x+t)\,dt.$$

Then the alternative form of (8.4.2) is

$$f_x = \frac{v_{0,\bar{r}} - 1}{e^{-\bar{r}x} l_x v_{x,\bar{r}}}. \tag{8.5.1}$$

The argument of this section pivots on the simple result 8.5.1. If $\bar{r} = 0$, we obtain the fraction f_x emigrating out of each cohort for stationarity. In general, (8.5.1) serves to show how much emigration is required to attain the demographic objective represented by a rate of increase $\bar{r}$, given the continuance of the life table $l(a)$ and the birth rates $m(a)$.

To apply (8.5.1) we need only the net maternity function $l(a)m(a)$. For Mauritius, 1966, this is given in Table 6.1 in 5-year age intervals. The intrinsic rate of Mauritius is estimated at 30.54 per thousand. How much emigration will be required for the modest goal of bringing it down to 20 per thousand? If the emigrants are $x = 25$ years of age, (8.5.1) tells us that with $\bar{r} = 0.020$ a fraction $f_{25} = 0.417$ of each cohort must leave on reaching this age. If the emigrants are $x = 20$ years of age, the proportion that will have to leave is smaller, 0.279.

Thus, to bring about a drop from the actual increase of 30.54 per thousand to one of 20.00 per thousand, the departure of 41.7 percent of each cohort will be required if the emigrants leave at age 25, and of 27.9 percent if they leave at age 20. Emigration is not the easiest means of population control.

To find the amount of emigration that will hold the ultimate rate of increase down to zero we need the value of f_x when $\bar{r}$ is replaced by zero in (8.5.1). The integral in the numerator is then R_0, the net reproduction rate, and the integral in the denominator is the part of R_0 beyond age x. Hence we have again (8.4.3),

$$f_x = \frac{R_0 - 1}{\displaystyle\int_x^{\beta} l(a)m(a)\,da} \tag{8.5.2}$$

as the fraction of the age x that must emigrate per year to hold the ultimate population stationary, x again being low enough for f_x not to exceed unity. To see (8.5.2) independently of its derivation as a special case of (8.5.1) we note that to bring the net reproduction rate down to 1 we need to lose $R_0 - 1$ births per woman from each birth cohort. The number of births per woman lost by removing a proportion f of women at age x is

$$f_x \int_x^\beta l(a)m(a)\,da.$$

Equating this to $R_0 - 1$ yields (8.5.2).

8.6 The Momentum of Population Growth

The authorities of some underdeveloped countries fear that once birth control is introduced their populations will immediately stop increasing. Such fears are misplaced, partly because diffusion takes time, and even when birth control is available it is not immediately used. But let us leave aside this behavioral aspect, and consider only the momentum of population growth that arises because the age distribution of a rapidly increasing population is favorable to increase. The concept has been introduced in Section 7.3.4; here we take advantage of the age-classification to explore what determines population momentum.

Suppose that all couples adopt birth control immediately and drop their births to a level that permits bare replacement. With U. S. mortality rates fertile couples need on the average (Section 16.3) 2.36 children to give a net reproduction rate R_0 of unity. An average of 2.36 children covers the loss of those dying before maturity, the fact that not everyone finds a mate, and some sterility among couples.

We saw that without any change in birth rates the ultimate birth trajectory due to $p(x)\,dx$ persons at age x to $x + dx$ would be $e^{rt}p(x)v(x)\,dx/\kappa$, and for the whole population distributed as $p(x)$ would be $e^{rt}\int_0^\beta p(x)v(x)\,dx/\kappa$. For calculating the effect of the fall to bare replacement we want the trajectory based on the existing age distribution $p(x)$, but with a function $v^*(x)$, corresponding to an intrinsic rate $r = 0$. We can arrange this, without changing any other feature of the age incidence of childbearing, by replacing $m(x)$ by $m^*(x) = m(a)/R_0$, which will ensure that $R_0^* = 1$ and $r^* = 0$. Then the ultimate stationary number of births must be

$$\int_0^\beta p(x)v^*(x)\,dx/\kappa, \tag{8.6.1}$$

where κ becomes μ, the mean age of childbearing in the stationary population because $v^* = 0$:

$$\frac{v^*(x)}{\kappa} = \frac{1}{\mu l(x)} \int_x^\beta \frac{l(a)m(a)\,da}{R_0}.$$

Ascertaining the ultimate stationary total population requires dividing by b, the stationary birth rate, which is the same as multiplying by $\overset{o}{e}_0$, the expectation of life at age zero.

Expression 8.6.1 is readily usable. If we have a table of the net maternity function in 5-year age intervals up to age 49 and the initial age distribution, then, by cumulating the net maternity function to obtain ${}_5V_x^*$ and multiplying 10 pairs of ${}_5N_x$ and ${}_5V_x^*$, we have the ultimate stationary population

$$\overset{o}{e}_0 \Sigma_0^{\beta-5}\, {}_5N_x\, {}_5V_x^*/\mu, \tag{8.6.2}$$

where

$${}_5V_x^* = \frac{(5/{}_5L_x)(\frac{1}{2}{}_5L_xF_x + {}_5L_{x+5}F_{x+5} + \cdots)}{R_0}.$$

This calculation will give the same result as a full population projection with the new $m^*(x)$.

If the initial age distribution $p(x)$ can be taken as stable, the result is even simpler. Entering $p(x) = p_0 be^{-rx}l(x)$ in (8.6.1), where r is the intrinsic rate before the drop to zero increase, canceling out $l(x)$ in numerator and denominator, and multiplying by $\overset{o}{e}_0$ to produce the stationary population rather than stationary births, we obtain

$$(1/p_0)\,\overset{o}{e}_0 \int_0^\beta p(x)v^*(x)\,dx = \frac{b\overset{o}{e}_0}{\mu}\int_0^\beta \int_x^\beta e^{-rx}l(a)\frac{m(a)}{R_0}\,da\,dx \tag{8.6.3}$$

as the ratio of the ultimate stationary population to the population at the time when the fall occurs.

The double integral is evaluated by writing b_x for $\int_x^\beta l(a)m(a)\,da/R_0$ and integrating by parts in (8.6.3) to obtain

$$\frac{b\overset{o}{e}_0}{\mu}\int_0^\beta e^{-rx}b_x\,dx = \frac{b\overset{o}{e}_0}{\mu}\left[\frac{e^{-rx}}{-r}b_x\bigg|_0^\beta - \frac{1}{r}\int_0^\beta e^{-rx}\frac{l(x)m(x)}{R_0}\,dx\right].$$

We find that the right-hand side reduces to

$$\frac{b\overset{o}{e}_0}{r\mu}\left(\frac{R_0-1}{R_0}\right) \tag{8.6.4}$$

on applying the fact that $b_0 = 1$ and $\int_0^\beta e^{-rx}l(x)m(x)\,dx = 1$. Expression 8.6.4 gives the ratio of the ultimate population to population just before the fall to zero increase and is the main result of this section.

For Ecuador, 1965, the data are $1000b = 44.82$, $\overset{o}{e}_0 = 60.16$, $1000r = 33.31$, $\mu = 29.41$, and $R_0 = 2.59$. These make expression 8.6.4 equal to 1.69. By simple projection or by (8.6.2), which does not depend on the stable assumption, we would have a ratio of the ultimate stationary to the present population of 1.67. This experiment and others show that the degree of stability in many underdeveloped countries makes (8.6.4) realistic.

Table 8.3. Values of $b\overset{o}{e}_0/\sqrt{R_0}$, the approximate ratio of the ultimate to the present population if the birth rate falls immediately from $b = 0.045$ to that needed for bare replacement, $1/\overset{o}{e}_0$

Initial	$\overset{o}{e}_0$		
R_0	40	50	60
1.5	1.47	1.84	2.20
2.0	1.27	1.59	1.91
2.5	1.14	1.42	1.71

James Frauenthal has pointed out to me that $(b\overset{o}{e}_0/r\mu)[(R_0 - 1)/R_0]$ of (8.6.4) is very nearly $b\overset{o}{e}_0/\sqrt{R_0}$. For R_0 is approximately $e^{r\mu}$, and hence

$$\frac{b\overset{o}{e}_0}{r\mu}\left(\frac{R_0 - 1}{R_0}\right) = \frac{b\overset{o}{e}_0}{\sqrt{R_0}}\left[\frac{e^{r\mu/2} - e^{-r\mu/2}}{r\mu}\right]$$

$$= \frac{b\overset{o}{e}_0}{\sqrt{R_0}}\left(1 + \frac{r^2\mu^2}{24} + \frac{r^4\mu^4}{960} + \cdots\right)$$

on expanding both the exponentials in powers of $r\mu$. The product $r\mu$ is of the order of unity, so that $r^2\mu^2/24$ must be close to 0.05. The example of Ecuador, 1965, gives $b\overset{o}{e}_0/\sqrt{R_0} = 1.68$ as compared with 1.69 for (8.6.4).

To obtain an intuitive meaning of this, note that the absolute number of births just after the fall must be $1/R_0$ times the births just before the fall. Births will subsequently rise and then drop in waves of diminishing amplitude, and it seems likely that the curve will oscillate about the mean of the absolute numbers before and after the fall. If the geometric mean of 1 and $1/R_0$ applies, the ultimate number of births will be $1/\sqrt{R_0}$ times the births before the fall. In that case the ultimate population will be $\overset{o}{e}_0/\sqrt{R_0}$ times the births before the fall, or $b\overset{o}{e}_0/\sqrt{R_0}$ times the population before the fall.

In words, the approximation $b\overset{o}{e}_0/\sqrt{R_0}$ says that the momentum factor is proportional to the birth rate and the expectation of life, and inversely proportional to the square root of the net reproduction rate. Table 8.3 suggests to what degree the factor depends on $\overset{o}{e}_0$ and to what degree on R_0 for an initial birth rate of $1000b = 45$. The conclusion is that with an immediate fall in fertility to bare replacement Ecuador and demographically similar countries would increase by about 50 percent or more before attaining stationarity. Note that (8.6.4) or $b\overset{o}{e}_0/\sqrt{R_0}$ is a good approximation to the degree in which the age distribution before the fall is stable. [Using model tables or otherwise, comment on the consistency of the pattern $b = 0.045$, $\overset{o}{e}_0 = 60$, $R_0 = 1.5$ that gives rise to the ratio 2.20 in Table 8.3.]

Table 8.4. Deaths from malaria and heart disease, Philippines, 1959 and 1960

Age x to $x+n$	Malaria, Cause B–16, 1959	Degenerative heart disease, Cause B–26, 1959	${}_nV_x$ Reproductive value for females, Philippines, 1960
All ages	913	918	
–5	251	12	1.21
5–14	156	7	1.64
15–24	133	37	2.00
25–44	186	198	0.76
45–64	138	322	0
65+	45	333	0
Unknown	4	9	
Total reproductive value for deaths of stated age	967	250	

Source: *United Nations Demographic Yearbook* (1961, p. 498); Keyfitz and Flieger (1971, p. 411).

8.7 Eliminating Heart Disease Would Make Very Little Difference to Population Increase, Whereas Eradication of Malaria Makes a Great Deal of Difference

Age distributions of deaths from malaria and heart disease are shown in Table 8.4 for the Philippines, 1959. Evidently malaria affects the young ages, whereas heart disease is negligible before middle life. Although the two causes are responsible for about equal numbers of deaths, malaria has a much greater effect on the chance that a child will survive to reproductive age and on the number of women living through reproduction.

Finding the effect on the population trajectory of eliminating deaths in any one year requires that each death at age x be evaluated as $v(x)$, that is to say, we need the sum $\int_0^\beta p(x)v(x)\,dx$, where now $p(x)\,dx$ is the population removed by death at ages x to $x+dx$. (The constants b and κ will not affect the relative positions of the two causes.) The broad age groups and lumping of the two sexes in Table 8.4 prevent us from attaining high accuracy. Table 8.4 shows unweighted arithmetic averages of $v(x)$ for the age groups required. The value of the malaria deaths, if they were female, would be $(251)(1.21)+(156)(1.64)+(133)(2.00)+(186)(0.76)=967$; that of the heart disease deaths similarly calculated would be 250. In practice men and women influence mortality in different degrees, and no easy way to allow for this suggests itself.

But the complexities that a two-sex model would introduce would not greatly affect the present conclusion: although absolute numbers of deaths from heart disease are about equal to those from malaria, malaria has nearly 4 times the effect on subsequent population.

8.8 The Stable Equivalent

The stable equivalent Q, associated with long-run projections, helps to interpret an observed past age distribution from the viewpoint of reproductive potential, and so bridges the present chapter and the preceding one dealing with reproductive value. It is the natural companion of the intrinsic rate of natural increase r. The rate r tells us *how fast* the population would ultimately increase at present age-specific rates; Q tells us *at what level* the ultimate population curve would stand.

8.8.1 Population Projection and the Stable Approximation Thereto

Most of this chapter has used the continuous renewal equation model for age-classified populations. Here we shift perspective to the discrete population projection matrix method. We are given an observed (from a mathematical viewpoint an arbitrary) age distribution for one sex, which is arranged as a vertical vector $\mathbf{n}(0)$, together with a set of age-specific birth and death rates arranged in the form of a matrix $\mathbf{A}$. If a 5-year projection interval, and 5-year age groups to age 85 to 89 at last birthday are recognized, $\mathbf{A}$ has 18×18 elements and $\mathbf{n}(0)$ has 18×1. The first row of $\mathbf{A}$ provides for fertility, and the subdiagonal for survivorship; this is, in fact, the Leslie matrix of Section 3.1. The age distribution projected through $5t$ years is

$$\mathbf{n}(t) = \mathbf{A}^t\mathbf{n}(0). \tag{8.8.1}$$

An approximation to this projection, called asymptotic because it is approached as closely as one wishes with sufficiently large t, is

$$\mathbf{n}(t) \approx \mathbf{q}e^{5rt}, \tag{8.8.2}$$

where the vector $\mathbf{q}$ is the stable equivalent of the age distribution.

To calculate $\mathbf{q}$ choose a large t and equate the right-hand sides of (8.8.1) and (8.8.2). If the population were of stable age distribution from the start, and contained $\mathbf{q}$ individuals at the several ages, by time $5t$ it would grow to $\mathbf{q}e^{5rt}$. In fact, we know that it is of age distribution $\mathbf{n}(0)$, and when projected it grows to $\mathbf{A}^t\mathbf{n}(0)$ by time $5t$. The matrix equation for the

Table 8.5. Female population total by conventional projection and by contribution of dominant root, starting from United States, 1960* (000s)

Year	t	Leslie projection with fixed 1960 rates	Contribution of positive term Qe^{rt}
1960	0	91,348	76,840
1970	10	106,220	94,986
1980	20	125,669	117,416
1990	30	150,129	145,144
2000	40	181,464	179,419
2010	50	222,196	221,789
2020	60	273,949	274,164
2030	70	338,990	338,907
2040	80	418,996	418,939

*Right-hand column is $Qe^{rt} = 76{,}840e^{0.0212t}$.

calculation of $\mathbf{q}$ is thus

$$\mathbf{q}e^{5rt} = \mathbf{A}^t\mathbf{n}(0)$$

or

$$\mathbf{q} = \frac{\mathbf{A}^t\mathbf{n}(0)}{e^{5rt}}. \tag{8.8.3}$$

One way of describing (8.8.3) is to say that $\mathbf{n}(0)$, the initial population is projected *forward* t periods by the matrix $\mathbf{A}$ and *backward* an equal length of time by the real root r, that is, by dividing by e^{5rt}. The quantity $\mathbf{q}e^{rt}$ corresponds to the real term in the solution of the Lotka equation (7.5.2), but is more complete in providing the several ages of the population rather than births alone. The total of all ages, written as $Q = \sum q_i$, is shown in Table 8.5 for United States females, starting with the 1960 age distribution and projected by 1960 births and life table.

The intrinsic rate of natural increase for the regime of 1959–61 being $r = 0.0212$, and the stable equivalent of the initial population being $Q = 76{,}840{,}000$, the future female population t years after 1960, if age-specific rates remained fixed and the stable model applied, would be $76{,}840{,}000e^{0.0212t}$. Table 8.5 compares this at 10-year intervals with the full projection, which implicitly uses all terms in the right-hand side of (7.5.2). By the year 2000 the discrepancy is down to 1.1 percent.

However, between 1960 and 1965 some of the postwar cohorts moved into childbearing ages, and the age distribution became more favorable, to the point where the stable equivalent and the observed total practically coincided, both being just under 99 million (Table 8.6). At the same time a drastic decline in the birth rates occurred, so that the intrinsic rate fell

Table 8.6. Female population P_0 and stable equivalent number Q, United States, 1919–21 to 1965, adjusted births

Year	Observed female population (000s) $N(0)$	Stable equivalent (000s) Q
1919–21	52,283	55,519
1924–26	57,016	61,442
1929–31	60,757	72,304
1934–36	63,141	78,879
1939–41	65,811	77,279
1944–46	69,875	72,016
1949–51	76,216	68,376
1954–56	83,248	69,535
1960	91,348	76,840
1965	98,703	98,645

Source: Keyfitz and Flieger (1968).

to $r = 0.01267$. Hence the future from the 1965 vantage point was

$$98{,}645{,}000e^{0.01267(t-5)}, \tag{8.8.4}$$

if t is still measured from 1960.

8.8.2 Application of the Stable Equivalent Q

Table 8.6 shows Q to be considerably above the observed female population $N(0) = \sum n_i(0)$ for the United States during the 1930s, and below it in the 1950s. This reflects the tendency for there to be proportionately more women of the age of motherhood in the population for some years after a fall in the birth rate. The crude birth rate usually lags behind the intrinsic birth rate after an upturn or downturn in fertility. The stable equivalent Q measures the favorability of the age distribution to reproduction, given the current regime of mortality and fertility.

In Table 8.7 historical data on Q are presented for four other countries. Again a high Q relative to population after a fall in birth rates appears for England and Wales between 1901 and 1941, and for Australia and Canada before 1941. The Netherlands also shows this feature, but to a more moderate degree.

8.8.3 Relation Between Q and Reproductive Value V

Reproductive value, the discounted future girl children that will be born to a woman, has a close relation to Q. [Prove that Q, like V but unlike $N(t)$, has the property of increasing at a constant rate under a fixed regime of

Table 8.7. Observed female population and stable equivalent, historical data for four countries

Country and year	Female population (000s) $N(0)$	Stable equivalent (000s) Q	Ratio $Q/N(0)$
Australia			
1911	2,152	2,395	1.11
1921	2,683	3,013	1.12
1933	3,263	4,267	1.31
1947	3,782	3,501	0.93
1957	4,758	4,215	0.89
1960	5,083	4,494	0.88
1965	5,632	5,659	1.00
Canada			
1931	5,001	5,706	1.14
1941	5,608	6,356	1.13
1951	6,751	6,431	0.95
1961	8,794	8,120	0.92
1965	9,479	9,839	1.04
England and Wales			
1861	10,324	10,802	1.05
1871	11,695	11,966	1.02
1881	13,373	13,608	1.02
1891	14,989	15,805	1.05
1901	16,845	19,047	1.13
1911	18,655	22,014	1.18
1921	19,816	22,229	1.12
1931	20,839	27,321	1.31
1941	21,515	27,522	1.28
1946	21,979	20,511	0.93
1951	22,751	22,741	1.00
1956	23,150	21,577	0.93
1961	23,820	19,764	0.83
Netherlands			
1901	2,615	2,647	1.01
1910	2,960	3,064	1.03
1920	3,419	3,615	1.06
1930	3,954	4,386	1.11
1940	4,437	4,983	1.12
1945	4,619	4,551	0.99
1950	5,074	5,077	1.00
1955	5,395	5,405	1.00
1960	5,766	5,615	0.97
1965	6,081	5,942	0.98

Source: Keyfitz and Flieger (1968).

Table 8.8. Observed female population, stable equivalent, and reproductive value (000s)

Country and Year	Observed female population $N(0)$	Stable equivalent Q	Ratio of stable to observed $Q/N(0)$	Reproductive value in units of girl babies V
Austria, 1964	3,845	3,187	0.83	1,665
Czechoslovakia, 1964	7,198	7,312	1.02	3,253
Denmark, 1964	2,380	2,326	0.98	1,091
Fiji Islands, 1964	219	229	1.04	229
Finland, 1964	2,370	2,540	1.07	1,227
Germany (East), 1964	9,257	7,871	0.85	3,499
Germany (West), 1964	30,980	27,755	0.90	13,124
Netherlands, 1964	6,081	5,942	0.98	3,665
Norway, 1964	1,854	1,649	0.89	914
Puerto Rico, 1964	1,309	1,375	1.05	1,050
Roumania, 1964	9,665	13,250	1.37	4,088
Switzerland, 1964	2,940	2,861	0.97	1,431

Source: Keyfitz and Flieger (1968).

mortality and fertility. The proof involves the fact that $(\mathbf{A}^t/e^{5rt})\mathbf{n}(0)$ is invariant with respect to t as long as t is large; in particular, $\mathbf{A}^t/e^{5rt}$ is the same when $t+1$ is written for t (Section 8.1).]

In fact, V is a simple multiple of Q. In the continuous representation, V is exactly equal to Q multiplied by the intrinsic birth rate b and by the mean age of childbearing in the stable population κ, two constants obtainable from the age-specific rates and having nothing to do with the observed age distribution. The reader may prove this statement by rearranging the double integral contained in $\int_0^\beta N(x)v(x)\,dx$, where $v(x)$ is defined as in (8.1.1), and showing it to be the same as the numerator of the first coefficient Q in the solution (7.5.4) to the Lotka equation. In the present notation this will prove

$$Q = \frac{V}{b\kappa}. \tag{8.8.5}$$

Goodman (1968) shows this result to apply in the discrete case. Values of $N(0)$, Q, and V are given in Table 8.8 for a number of countries.

A question arises of the degree to which Q, the stable equivalent, is sensitive to the particular set of age-specific birth and death rates used in its calculation. The first row of Table 8.9 shows Q for the age distribution of 1960, worked out according to the 1960 and 1965 patterns of mortality and fertility as embodied in the two $\mathbf{A}$'s; the second row shows the corresponding Q's for the 1965 age distribution. The values obtained for Q depend greatly on the set of age-specific rates applied as $\mathbf{A}$. But if we study only the *change* in Q in the United States between 1960 and 1965, it turns out that the increase is 11.14 percent on the 1960 $\mathbf{A}$ and 10.84 percent on the

Table 8.9. Stable equivalents Q for United States females in 1960 and 1965, each calculated with two different matrices **A** (thousands)

	Calculated with matrix **A** of	
	1960	1965
Age distribution:		
1960	76,912	89,001
1965	85,478	98,645
Percent increase	11.14	10.84

Table 8.10. Stable equivalents Q for the United States and Mexico (thousands)

	Calculated with matrix **A** of	
	United States, 1962	Mexico, 1962
Age distribution:		
United States	82,933	63,395
Mexico	23,388	18,863
Ratio, Mexico to United States	0.282	0.272

1965 **A**. This way of making a comparison (applied between France and Italy) is due to Vincent (1945), who noted the virtual invariance with respect to the mortality and fertility patterns used. We are entitled to say that the age distribution of women in the United States became about 11 percent more favorable to reproduction during the 5 years in question, and the statement is true almost without regard to the fertility and mortality patterns used in making this assessment.

As an example of a place comparison, Table 8.10 shows Q values for Mexico and the United States, both for 1962. The Q for Mexico is 0.282 of that for the United States when the **A** of the latter is used; it is 0.272 when the **A** of the former is used. Jeffrey Evans has programmed place comparisons among five countries which demonstrate the same invariance.

Section 12.3 below uses the stable equivalent to compare the effect on age distribution of eliminating cancer with that of eliminating heart disease.

8.8.4 A More General Stable Equivalent

Age is merely a special case of the stable equivalent. Any model that possesses the ergodic property, that is, that tends asymptotically to a distribution unaffected by the initial distribution, is equally capable of analysis by the methods given above. In fact (8.8.3) remains unchanged; only now the matrix **A** and the vector $\mathbf{n}(0)$ provide for the two sexes, regions, married and single populations, and any other groups recognized in the model. For details see Chapter 9 and some applications in Keyfitz (1969).

8.9 Reproductive Value as a Contribution to Future Births

Section 8.1 appeals to intuition to make it appear likely that the effect of adding one girl or woman aged x to the population is to raise the number of births t years hence, where t is large, in proportion to $v(x)e^{rt}$, $v(x)$ being defined as

$$v(x) = \frac{\displaystyle\int_x^{\beta} e^{-ra} l(a) m(a)\, da}{e^{-rx} l(x)}.$$

This result can be derived from the Lotka equation of Section 7.5, but here we examine a demonstration that is self contained, using the familiar device of calculating the situation at time t from two successive moments near the present. For purposes of this section $v(x)$ will be defined afresh, in terms of the ultimate birth trajectory.

Suppose that a woman aged x at time zero contributes $v(x)e^{rt}$ to the births at subsequent time t, where $v(x)$ is to be determined and t is large. This means that her disappearance would lower the ultimate birth trajectory by $v(x)e^{rt}$. We assume that age-specific birth and death rates are fixed, so that her descendants will ultimately increase in geometric proportion and be unaffected by other members of the population.

The woman aged x can, in the next short period of time and age, say Δ, have a child, and whether or not she has a child can survive to the next age, $x + \Delta$. The chance of her having a child is $m(x)\Delta$, and the chance of her surviving is $l(x + \Delta)/l(x)$. By having a child she would contribute $v(0)e^{r(t-\Delta)}$ to the births at time t, and by surviving she would convert herself into a woman of reproductive value $v(x + \Delta)$ and so contribute $v(x + \Delta)e^{r(t-\Delta)}$. If the progression of childbearing and aging at the given rates over the time Δ is not to affect the ultimate birth trajectory, we can equate the two expressions for later births:

$$v(x)e^{rt} = \left[m(x)v(0)\Delta + \frac{l(x+\Delta)}{l(x)} v(x+\Delta) \right] e^{r(t-\Delta)}. \tag{8.9.1}$$

If we multiply both sides of (8.9.1) by

$$\frac{1}{\Delta} \frac{l(x)}{v(0)} e^{-rx} e^{-rt},$$

we obtain

$$\begin{aligned}\frac{1}{\Delta}l(x)\frac{v(x)}{v(0)}e^{-rx} &= m(x)l(x)e^{-rx}e^{-r\Delta}\\ &\quad + \frac{1}{\Delta}l(x+\Delta)\frac{v(x+\Delta)}{v(0)}e^{-r(x+\Delta)}.\end{aligned} \tag{8.9.2}$$

Subtracting the rightmost term from both sides and letting $\Delta \to 0$, we have directly

$$-\frac{d}{dx}l(x)\frac{v(x)}{v(0)}e^{-rx} = m(x)l(x)e^{-rx},$$

and integrating gives

$$e^{-rx}l(x)\frac{v(x)}{v(0)} = \int_x^\beta e^{-ra}l(a)m(a)\,da, \tag{8.9.3}$$

so that, if $v(0)$ is set equal to unity, $v(x)$ again comes out as shown in (8.1.1). No constant of integration is needed, since both sides are unity for $x = 0$. Equation (8.9.3) establishes $v(x)$ to within a multiplicative constant.

Let us find the constant $v(0)$ that corresponds to the ultimate effect of adding one female to the population.

If the initial age distribution is stable, we know that the population at time t must be e^{rt} for each person initially present, and hence the births at time t are be^{rt}. Equating the two values for time t, we have

$$be^{rt} = \int_0^\beta be^{-rx}l(x)v(x)\,dx\,e^{rt}; \tag{8.9.4}$$

since from (8.9.3) $v(x)$ may be written as

$$\frac{v(0)}{e^{-rx}l(x)}\int_x^\beta e^{-ra}l(a)m(a)\,da,$$

the $be^{-rx}l(x)$ within the integral of (8.9.4), as well as e^{rt} outside the integral, cancels, and we obtain the following equation for $v(0)$:

$$\frac{1}{v(0)} = \int_0^\beta\int_x^\beta e^{-ra}l(a)m(a)\,da\,dx. \tag{8.9.5}$$

The integral on the right-hand side is evaluated by integration by parts and turns out to be κ, the mean age of childbearing in the stable population. This proves that for the $v(x)$ function of this section, $v(0) = 1/\kappa$, and that the $v(x)$ function of the main body of the chapter, defined in (8.1.1), gives the ultimate birth trajectory due to a woman aged x as $e^{rt}v(x)/\kappa$.

9

Reproductive Value from Matrix Models

The concept of reproductive value is not limited to age-structured populations. It also applies to matrix population models for stage-structured populations, where it appears as an eigenvector of the projection matrix.

9.1 Reproductive Value as an Eigenvector

We begin by returning to the projection equation

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t) \qquad \mathbf{n}(0) = \mathbf{n}_0 \tag{9.1.1}$$

the solution of which (Section 7.1) can be written

$$\mathbf{n}(t) = \sum_i c_i \lambda_i^t \mathbf{w}_i, \tag{9.1.2}$$

where λ_i and $\mathbf{w}_i$ are the eigenvalues and right eigenvectors of $\mathbf{A}$ and the scalar constants c_i are determined by the initial conditions $\mathbf{n}_0$;

$$\mathbf{c} = \mathbf{W}^{-1}\mathbf{n}_0 \tag{9.1.3}$$

$$= \overline{\mathbf{V}}\mathbf{n}_0. \tag{9.1.4}$$

The matrix $\mathbf{W}$ has the right eigenvectors $\mathbf{w}_i$ as its columns; $\mathbf{W}^{-1}$ has as its rows the complex conjugate transposes of the left eigenvectors $\mathbf{v}_i$. Thus

$$c_i = \mathbf{v}_i^* \mathbf{n}_0 \tag{9.1.5}$$

with $\mathbf{w}_i$ and $\mathbf{v}_i$ scaled so that $\mathbf{v}_i^* \mathbf{w}_i = 1$.

If $\mathbf{A}$ is primitive, then

$$\lim_{t\to\infty} \frac{\mathbf{n}(t)}{\lambda_1^t} = c_1\mathbf{w}_1. \tag{9.1.6}$$

The growth rate and stable population structure are independent of $\mathbf{n}_0$, but the size of the population at any (large) time t depends on $\mathbf{n}_0$, through the constant c_1. From (9.1.5), c_1 is a weighted sum of the initial population, with weights equal to the elements of $\mathbf{v}_1$.

Thus, if we take "the contribution of stage i to long-term population size" as a reasonable measure of the "value of stage i," the left eigenvector $\mathbf{v}_1$ gives the relative reproductive values of the stages (Goodman 1968, Keyfitz 1968). We must insert the qualifier "relative" because eigenvectors can be scaled by any nonzero constant. The result $c_1 = \mathbf{v}_1^*\mathbf{n}_0$ holds when $\mathbf{v}_1^*\mathbf{w}_1 = 1$, but any other scaling can be accounted for by setting $c_1 = \mathbf{v}_1^*\mathbf{n}_0/\mathbf{v}_1^*\mathbf{w}_1$, and eventual population size is still proportional to $\mathbf{v}_1^*\mathbf{n}_0$. It is customary to scale $\mathbf{v}_1$ so that its first entry is 1.

Regardless of the scaling imposed on $\mathbf{v}_1$, the total reproductive value of a population, $V(t) = \mathbf{v}_1^*\mathbf{n}(t)$, increases exponentially at the rate λ_1, regardless of the stage distribution:

$$V(t+1) = \mathbf{v}_1^*\mathbf{n}(t+1) \tag{9.1.7}$$
$$= \mathbf{v}_1^*\mathbf{A}\mathbf{n}(t) \tag{9.1.8}$$
$$= \lambda\mathbf{v}_1^*\mathbf{n}(t). \tag{9.1.9}$$

9.1.1 The Effect of Adding a Single Individual

Suppose that we add a single individual of stage j to the initial population $\mathbf{n}_0$. Let $\mathbf{e}_j$ be a vector with zeros everywhere except for a 1 in the jth entry. If we drop the subscripts on λ_1, $\mathbf{w}_1$ and $\mathbf{v}_1$, we have

$$\lim_{t\to\infty} \frac{\mathbf{A}^t\left(\mathbf{n}_0 + \mathbf{e}_j\right)}{\lambda^t} = \mathbf{v}^*\left(\mathbf{n}_0 + \mathbf{e}_j\right)\mathbf{w} \tag{9.1.10}$$
$$= \mathbf{v}^*\mathbf{n}_0\mathbf{w} + v_j\mathbf{w}. \tag{9.1.11}$$

The total population is $\mathbf{v}^*\mathbf{n}_0\|\mathbf{w}\| + v_j\|\mathbf{w}\|$, which differs from (9.1.6) by $v_j\|\mathbf{w}\|$. That is, adding a single individual in stage j increases asymptotic population size by an amount proportional to the reproductive value of stage j.

Reproductive Value and Extinction.

Any population is subject to stochastic fluctuations because the vital rates are probabilities applied to discrete individuals (demographic stochasticity). These fluctuations lead to a nonzero probability of extinction, even when $\lambda > 1$. This probability can be calculated for unstructured populations from the Galton–Watson branching process (see Section 16.4). The corresponding probability for structured population is calculated from the multi-type branching process (Pollard 1973, MPM Chapter 15). In several empirical examples (MPM Section 15.4.5), it has been shown that the probability of non-extinction of a population descended from a single founder is directly proportional to the reproductive value of that founder. This suggests, though it does not prove, that the reproductive value of an individual influences not only long-term population size but also short-term risk of extinction.

9.1.2 Age-Specific Reproductive Value

We can write down the reproductive value for the age-classified case directly from the equations defining the eigenvector:

$$\mathbf{v}^{\mathsf{T}}\mathbf{A} = \lambda \mathbf{v}^{\mathsf{T}},$$

where we have dropped the subscript, and are assuming that $\mathbf{v}$ is real. Suppose there are four age classes, as in Figure 3.9a, and set $v_1 = 1$. Then

$$\begin{pmatrix} 1 & v_2 & v_3 & v_4 \end{pmatrix} \begin{pmatrix} F_1 & F_2 & F_3 & F_4 \\ P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ 0 & 0 & P_3 & 0 \end{pmatrix} = \lambda \begin{pmatrix} 1 & v_2 & v_3 & v_4 \end{pmatrix} \tag{9.1.12}$$

or, writing each equation out

$$F_1 + v_2 P_1 = \lambda \tag{9.1.13}$$

$$F_2 + v_3 P_2 = \lambda v_2 \tag{9.1.14}$$

$$F_3 + v_4 P_3 = \lambda v_3 \tag{9.1.15}$$

$$F_4 = \lambda v_4. \tag{9.1.16}$$

From the last equation

$$v_4 = F_4 \lambda^{-1}. \tag{9.1.17}$$

Substituting this into the next-to-last equation gives

$$v_3 = F_3 \lambda^{-1} + P_3 F_4 \lambda^{-2} \tag{9.1.18}$$

and then

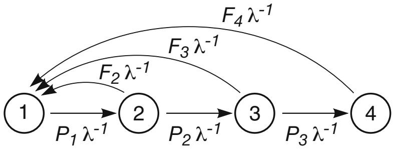
$$v_2 = F_2 \lambda^{-1} + P_2 F_3 \lambda^{-2} + P_2 P_3 F_4 \lambda^{-3}. \tag{9.1.19}$$

Finally, substituting this into the first equation gives

$$1 = F_1 \lambda^{-1} + P_1 F_2 \lambda^{-2} + P_1 P_2 F_3 \lambda^{-3} + P_1 P_2 P_3 F_4 \lambda^{-4} \tag{9.1.20}$$

which is the characteristic equation (see Example 7.1). In general the age-specific reproductive value is

$$v_i = \sum_{j=i}^{s} \left(\prod_{h=i}^{j-1} P_h \right) F_j \lambda^{i-j-1}, \tag{9.1.21}$$



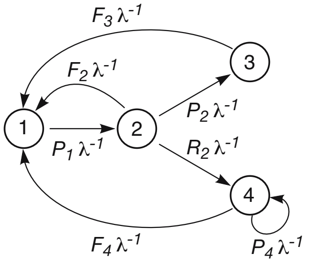


Figure 9.1. The transformed graphs for two life cycles. Above: an age-structured model with four age classes. Below: a hypothetical life cycle in which individuals of stage $\mathcal{N}_2$ have two developmental choices, in one of which ($\mathcal{N}_3$) they reproduce only once and in the other of which ($\mathcal{N}_4$) they survive with probability P_4 and reproduce repeatedly.

which is the discrete version of Fisher's formula (8.1.1).

9.1.3 Stage-Specific Reproductive Value and the Life Cycle Graph

We can understand the correspondence of the left eigenvector and reproductive value in stage-structured models by writing down the eigenvector directly from the life cycle graph (Caswell 1982a; see Chapter 7 of MPM for details). Begin by transforming the life cycle graph by replacing each coefficient a_{ij} with $a_{ij}\lambda^{-1}$. (This is known as the z-transform of the graph; in our context, however, the variable usually denoted by z will be the eigenvalue of $\mathbf{A}$, so we denote it as λ.)

Figure 9.1 shows the transformation of the life cycle graph for the age-classified model. Comparing this graph with (9.1.17)–(9.1.20), we see that v_i is the sum, over all pathways from $\mathcal{N}_i$ to $\mathcal{N}_1$, of the product over each pathway of the transformed life cycle graph coefficients. There are, for example, two pathways from $\mathcal{N}_3$ to $\mathcal{N}_1$. The products of the transformed coefficients on these pathways are $F_3\lambda^{-1}$ and $P_3F_4\lambda^{-2}$; the sum of these is v_3 in (9.1.18).

In other words, v_i measures the expected future reproductive contribution from stage $\mathcal{N}_i$, discounted by the population growth rate and the time required for the contribution; e.g.,

$$v_2 = \underbrace{F_2\lambda^{-1}}_{\text{1 step}} + \underbrace{P_2F_3\lambda^{-2}}_{\text{2 steps}} + \underbrace{P_2P_3F_4\lambda^{-3}}_{\text{3 steps}}. \tag{9.1.22}$$

This algorithm gives the left eigenvector for a wide class of life cycles.[*]

In the (imaginary) stage-classified life cycle of Figure 9.1b, an individual in $\mathcal{N}_2$ may proceed to a stage ($\mathcal{N}_3$) in which it only reproduces once or to a stage ($\mathcal{N}_4$) in which it survives indefinitely with a probability P_4. The resulting reproductive value vector, obtained by summing contributions from each stage back to the first, is

$$v_1 = 1 \tag{9.1.23}$$

$$v_2 = F_2\lambda^{-1} + P_2F_3\lambda^{-2} + \frac{R_2F_4\lambda^{-2}}{1 - P_4\lambda^{-1}} \tag{9.1.24}$$

$$v_3 = F_3\lambda^{-1} \tag{9.1.25}$$

$$v_4 = \frac{F_4\lambda^{-1}}{1 - P_4\lambda^{-1}}. \tag{9.1.26}$$

Each of these values is clearly a measure of future contribution to births, discounted by the population growth rate.[†]

Residual Reproductive Value.

Equation (8.9.1) decomposed reproductive value at age x into two components, one from reproduction at age x and the other from survival to, and reproduction at, later ages. These components were called *current reproduction* and *residual reproductive value* by Williams (1966). In the age-classified case (9.1.12), e.g.,

$$v_2 = F_2\lambda^{-1} + P_2\lambda^{-1}v_3, \tag{9.1.27}$$

where the first term is current reproduction and the second is residual reproductive value. In the stage-classified example, an individual in $\mathcal{N}_2$ has two possible fates, so

$$v_2 = F_2\lambda{-}1 + P_2\lambda^{-1}v_3 + R_2\lambda^{-1}v_4. \tag{9.1.28}$$

The first term is current reproduction and the second two terms together constitute residual reproductive value.

---

[*] Including any life cycle in which all loops other than self-loops pass through $\mathcal{N}_1$. If there are more complicated loop structures, additional terms are required (Caswell 1982a, MPM Chapter 7).

[†] The terms

$$\frac{1}{1 - P_4\lambda 1} = 1 + P_4\lambda^{-1} + P_4^2\lambda^{-2} + \ldots$$

created by the self-loop on $\mathcal{N}_4$ reflect the probability that the individual will remain in $\mathcal{N}_4$ for $1, 2, \ldots$ time steps.

9.2 The Stable Equivalent Population

The stable equivalent population of Section 8.8 applies to any classification of individuals (Keyfitz 1969). An initial population $\mathbf{n}_0$ with an arbitrary stage distribution will asymptotically produce an exponentially growing population of the same size as an initial population of size Q with the stable stage distribution.

To calculate Q, we scale $\mathbf{w}$ so that $\|\mathbf{w}\| = 1$, and $\mathbf{v}$ so that $\mathbf{v}^*\mathbf{w} = 1$. The population starting at $\mathbf{n}_0$ will eventually grow as

$$\lim_{t \to \infty} \frac{\mathbf{n}(t)}{\lambda^t} = \mathbf{v}^*\mathbf{n}_0\mathbf{w} \tag{9.2.1}$$

while that starting from $\mathbf{n}(0) = Q\mathbf{w}$ will grow as

$$\lim_{t \to \infty} \frac{\mathbf{n}(t)}{\lambda^t} = Q\left(\mathbf{v}^*\mathbf{w}\right)\mathbf{w}. \tag{9.2.2}$$

Equating the two gives

$$Q = \mathbf{v}^*\mathbf{n}_0. \tag{9.2.3}$$

That is, the stable equivalent is just the total reproductive value of the initial population, when scaled so that $\|\mathbf{w}\| = 1$ and $\mathbf{v}^*\mathbf{w} = 1$.

We note in passing that the models considered here and in Chapter 8 describe constant environments. Tuljapurkar and Lee (1997) have extended the stable equivalent concept to models in which the vital rates fluctuate stochastically in time.

Example 9.1 Stable equivalent for the killer whale

Killer whales (*Orcinus orca*) live in stable social groups called pods. A life cycle is shown in Figure 3.10, and a set of vital rates estimated from an intensively studied population of 18 pods in coastal waters of Washington and British Columbia is shown in Example 11.1. The right and left eigenvectors, appropriately scaled, are

$$\mathbf{w} = \begin{pmatrix} 0.037 \\ 0.316 \\ 0.323 \\ 0.324 \end{pmatrix} \qquad \mathbf{v} = \begin{pmatrix} 1.142 \\ 1.198 \\ 1.794 \\ 0 \end{pmatrix}. \tag{9.2.4}$$

Each pod has its own observed structure, and Table 9.1 compares the stable equivalent and the observed population of each. In contrast to the comparison of the stable and observed population sizes of 12 countries in Table 8.8, which were within a few percent of each other, among killer whale pods the stable equivalent ranges from 22 percent smaller to 71 percent larger than the observed population. When $Q < N$, the population is biased toward individuals of low reproductive value, and vice versa.

Table 9.1. The observed female population ($N = \|\mathbf{n}\|$) and the stable equivalent population Q for each of 18 pods of resident killer whales (*Orcinus orca*) in Washington and British Columbia.

N	Q	$\frac{Q}{N}$
12.93	10.59	0.82
10.30	8.05	0.78
26.23	28.10	1.07
5.77	5.77	1.00
4.20	6.33	1.51
7.73	9.48	1.23
1.23	2.11	1.71
5.10	4.74	0.93
5.63	6.69	1.19
11.53	12.27	1.06
5.13	7.33	1.43
3.07	4.82	1.57
2.50	3.78	1.51
2.30	1.99	0.87
6.37	9.12	1.43
6.13	9.00	1.47
2.93	4.22	1.44
8.47	11.71	1.38

The influence of $\mathbf{n}_0$ on eventual population size (and probability of extinction) is of more than academic interest in conservation biology. Invasions of introduced animals and plants create huge environmental and economic problems around the world. Studies of the determinants of invasion success of birds and mammals in New Zealand (which, because of its isolation, has been particularly vulnerable to invasions) have shown a correlation between the size of the introduced population and the success of the invasion (Veltman et al. 1996, Forsyth and Duncan 2001). The stable equivalent of the introduced population might be even more relevant.

The effect of initial population also arises in attempts to reintroduce threatened species to areas from which they have been exterminated. This is an increasingly frequent task; at this writing, 132 such projects involving 63 species are underway in New Zealand alone. Many of these involve introductions of individuals to offshore islands from which introduced predators have been eliminated. All else being equal, it might be useful to try to maximize the stable equivalent population size in such introductions.

9.2.1 Other Scalings of the Eigenvectors

Since the time of Fisher (1930) it has been customary to scale reproductive value so that $v_1 = 1$, as was done in Section 8.8. But if $\mathbf{w}$ is also scaled to sum to 1, this means that Q must be modified to

$$Q = \frac{\mathbf{v}^* \mathbf{n}_0}{\mathbf{v}^* \mathbf{w}}. \tag{9.2.5}$$

For age-classified matrix models, it can be shown that, with this scaling of $\mathbf{v}$ and $\mathbf{w}$, the denominator

$$\mathbf{v}^* \mathbf{w} = \lambda^{-1} B \overline{A}, \tag{9.2.6}$$

where B is the finite birth rate and $\overline{A}$ is the mean age of childbearing in the stable population [the equivalent of κ in (8.1.2)]. Thus (9.2.5) is the analogue of the continuous-time result (8.8.5); see also Goodman (1968).

It does not appear that the interpretation of $\mathbf{v}^* \mathbf{w}$ in terms of birth rate and generation time holds for general stage classifications, so in the general case it is easier to compute Q by scaling $\|\mathbf{w}\| = 1$ and $\mathbf{v}^* \mathbf{w} = 1$ and sacrificing $v_1 = 1$.

10

Understanding Population Characteristics

To understand a phenomenon we must break it down into simple elements and then put these elements together again in such a way as to reconstruct the phenomenon. This was the method Descartes proposed for study of the physical world, and it can be used to make intelligible the population characteristics presented as census and other data. Such characteristics as age, sex, marital status, birthplace, occupation, and industry can be treated by the Cartesian method, though not all with equal effectiveness.

The transition from one year of age to the next against the hazard of death, and the population rate of increase, are elements that help to explain age distribution. The transition from one school grade to the next in the face of the hazard of dropping out has a bearing both on the grade distribution of pupils presently in the schools and on the distribution by years of schooling completed in the population at large. The transition into the labor force and that into retirement go some distance toward accounting for the distribution of the population by labor force status. These transitions can be incorporated into models capable of approximating the present distribution of ages, schooling, and labor force participation.

Explanation involves many levels arranged in an infinite regression. Why the age distribution or the sex ratio is as it is, is traceable through deeper and deeper stages, involving fetal mortality and its causes and the causes of death in young and older people. The present argument can go only one short step in this regression.

10.1 Accounting for Age Distribution

10.1.1 Young and Old Populations

In 1976, the United States was celebrating its 200th birthday; France was older, having gained its independence from Roman colonialism about the fifth century and having become a unified nation in the seventeenth century; Taiwan was much younger than either, having been established after World War II. In the United States the percentage of children under 15 years of age was 30.9, while in France it was 24.6; Taiwan had a larger proportion of children, 45.2 percent (all 1965 figures). For these three countries and for many others, the older the country as a political entity the smaller is the fraction of its population under 15 years of age. Yet no one could take seriously an assertion relating political to demographic age. The correlation can only be called spurious since we have no reason in logic to think that the fraction of children is related to political youth or age. Chapter 17 shows other weaknesses of a purely empirical approach to demography.

Let us here drop political age and call a country (demographically) young if it has a large fraction of children and a small fraction of old people. "Young" and "old" in terms of this definition will be explainable by the life table and rate of increase. Of 800 age distributions for various countries and times that are available for examination, that of Honduras in 1965 is the youngest, with 50.8 percent of its population reported as under 15 years of age. The average age of Honduran males was 19.8 years; of United States males, 30.8 years; of Swedish males, 36.1 years (again all for 1965). Demographic youth or age can have direct and traceable consequences. Other things being equal, if a country has many children to support, it will be occupied in building houses and schools for them and will have fewer resources for building factories to increase its future income. This issue will reappear in Section 17.6.

Our first attempt at explanation will again be Euler's stable age distribution (5.1.1), by which, under a fixed regime of mortality and fertility including the probability $l(a)$ of surviving from birth to age a and a rate of increase r, the population between ages a and $a + da$ is $e^{-ra}l(a)\,da$ per current birth, and as a proportion this is divided by its integral over the range zero to ω.

A comparison of (5.1.1) with observed proportions for two age groups and three countries appears in Table 10.1, all for females. Of the difference in the under 15 group between France and Taiwan ($45.2 - 23.6 = 21.6$ percent) about two-thirds ($40.2 - 26.5 = 13.7$ percent) is accounted for on the stable model. The stable model accounts for a similar fraction of the differences in the numbers over 65.

The stable model we have constructed involves nothing but the $l(a)$, which depend on the present life table, and r, the rate of increase. Since it is based only on current information on birth and death, and does not take

Table 10.1. Percentage of females under 15 and over 65 for three countries in 1965

	Percentage under 15		Percentage over 65	
Country	Observed	Stable	Observed	Stable
Taiwan	45.2	40.2	3.1	4.5
United States	29.8	28.4	10.4	10.6
France	23.6	26.5	15.3	12.2

account of migration or of past wars and epidemics, we should be impressed by its capacity to explain so large a part of the differences among real age distributions.

10.1.2 Age Distribution as a Function of Rate of Increase

We saw earlier how the fraction of any age in a stable population is expressible in terms of the rate of increase of the population and the mean ages of the relevant subgroup. A special case, the fraction of population 65 years of age and over was treated in Section 5.7. In general terms the fraction between any pair of ages α and β is equal to

$$ {}_{\beta-\alpha}C_{\alpha} = \frac{\displaystyle\int_{\alpha}^{\beta} e^{-ra} l(a)\, da}{\displaystyle\int_{0}^{\omega} e^{-ra} l(a)\, da}. \tag{10.1.1} $$

By a proposition cited in Section 6.2, $\log[\int_0^{\omega} e^{-ra} f(a)\, da]$ generates the cumulants of the density distribution $f(a)$, the first cumulant being the mean, and the second and third cumulants being moments about the mean. We may define $f(a)$ as $l(a)/\int_0^{\infty} l(x)\, dx$ for the denominator, and the same for the numerator between α and β, outside of which $f(a) = 0$. Normalizing (10.1.1) by dividing by $\int_{\alpha}^{\beta} l(a)\, da / \int_0^{\omega} l(a)\, da$, taking logarithms of both sides, and expanding the two cumulant-generation functions of $-r$ gives

$$ \begin{aligned} \log {}_{\beta-\alpha}C_{\alpha} = \log\left(\frac{{}_{\beta-\alpha}L_{\alpha}}{\overset{o}{e}_0}\right) - \Bigg[k_1 r - \frac{k_2 r^2}{2!} + \frac{k_3 r^3}{3!} - \cdots \\ - \left(\kappa_1 r - \frac{\kappa_2 r^2}{2!} + \frac{\kappa_3 r^3}{3!} - \cdots\right)\Bigg], \end{aligned} \tag{10.1.2} $$

where the k's are the cumulants of the life table distribution between α and β, and the κ's are the cumulants over the whole range of the age distribution.

Hence to a first approximation, and on taking derivatives and then entering finite increments,

$$ \Delta_{\beta-\alpha}C_{\alpha} \approx (\kappa_1 - k_1)_{\beta-\alpha}C_{\alpha}\, \Delta r, \tag{10.1.3} $$

as was established for special cases in Sections 5.3 and 5.7.

Putting $\alpha = 0$ and $\beta = 15$ shows that the proportion under age 15 goes up as r goes up, since k_1, the mean age of those under 15, is bound to be less than κ_1, the mean age of the whole population. Similar considerations apply to other age intervals. The relation can be studied as the exponential of a quadratic or cubic wherever r is large enough to make further terms important.

The stable model does not always fit. Among 800 populations tested, that of England and Wales, 1881, comes closest; the Netherlands, 1901, is second. But even there agreement is not perfect. Females aged 0 to 14 in the actual population of England and Wales, 1881, comprised 35.5 percent; in the fitted stable model, 35.0 percent. Discrepancies between the model and the observed populations invariably demand our attention.

Since in Table 10.1 the 1965 rates of death were the source of the life table $l(a)$, and the 1965 intrinsic rate of increase was the source of r, one reason why the model can be at variance with reality is that different rates of death and increase applied in earlier years. This indeed appears to be the main source of discrepancy. In Taiwan birth rates have been falling; the 1965 birth rates are too low to represent the preceding 15 years. A calculation based on the stable model with 1959–61 birth and death rates shows 43.6 percent under 15, a value closer to the observed one.

For intervals α to β straddling the mean age, k and κ will not be very different; (10.1.3) shows that differences in r do not greatly affect the middle ages. By a similar argument drastic declines in the later ages are associated with increases in r.

10.1.3 Neutral and Nonneutral Change in Mortality

How do populations that have been subject to different death rates differ in age? If the difference is the same amount at all ages, it has no effect on the age distribution. If birth rates are the same and mortality at all ages has been higher by exactly 0.01 in one population than in the other, the age distributions will be identical (Coale, 1968).

This seems to be a paradox. A rise in mortality that prevents people from living to as old an age as they once did ought to make the population consist to a greater degree of young people, one would think. It certainly makes the prospect of attaining old age less than it was before for an individual; why does it not do the same for the community? Here, as in many other situations, the same rule does not apply to populations as to individuals.

When mortality is higher with fertility unchanged, the rate of increase is just enough lower to compensate as far as age distribution is concerned for the change in the life table. The canceling out is a property of exponential growth dovetailing with the exponential effect on $l(x)$ of an addition to $\mu(x)$.

If k (e.g., 0.01) is added to the mortality rate $\mu(x)$ at each age, so that it becomes $\mu^*(x) = \mu(x) + k$, the probability of surviving to age x is $l^*(x) = e^{-kx}l(x)$ instead of $l(x)$.

But if the age-specific birth rates are unchanged, the rate of increase of the population diminishes, and by exactly the same amount, k. The proof of this statement is that the new rate of increase r^* must satisfy the characteristic equation

$$\int_{\alpha}^{\beta} e^{-r^* x} l^*(x) m(x)\, dx = 1$$

or

$$\int_{\alpha}^{\beta} e^{-r^* x} e^{-kx} l(x) m(x)\, dx = 1$$

or

$$\int_{\alpha}^{\beta} e^{-(r^* + k)x} l(x) m(x)\, dx = 1.$$

Since the original r satisfied the same equation that $r^* + k$ now satisfies, by virtue of the uniqueness of the real root we have $r = r^* + k$, so that $r^* = r - k$; the new rate is exactly k less than the old rate.

Combining the preceding two paragraphs gives, for the new stable population on a radix of one current birth per year,

$$e^{-r^* x} l^*(x) = e^{-(r-k)x} e^{-kx} l(x) = e^{-rx} l(x);$$

hence the number of persons at age x per current birth is the same under the new regime as under the old.

Changes in mortality that have occurred historically can be distinguished according to whether they tended to be at younger or older ages, that is, whether or not they tended to lower the mean age of the stable population if the birth rate remained the same. We find that the improvements in England and Wales from 1861 to 1891 were on balance at younger ages, those between 1891 and 1911 were at older ages; and those from 1911 to 1931 again affected younger ages to a greater degree (Keyfitz 1968, p. 191).

10.1.4 Accounting for Observed Ages

The first part of the analysis of age distributions here outlined is concerned with the regime of mortality and natural increase actually existing; the second, with trends in the regime during the lifetimes of the current population; the third, with extraordinary historical events, especially those that cause short-term fluctuations in birth and death. These events include wars, which reduce the numbers of males in their twenties; immigration, which usually consists of young adults; and famines, which especially affect young children. In this third part of the analysis theory does not help much; we must look at the record to see what happened.

These results are a special case of perturbation analysis of the right eigenvector of the population projection operator. A general formula is available for matrix population models (Section 13.3), but its results rarely lend themselves to interpretation in the way the approach here does.

10.1.5 Are Birth or Death Rates the Major Influence on Age Distribution?

The stable model, which can reconstruct an age distribution from a life table and rate of increase, enables us to compare two populations, say those of the United States and Madagascar, of which one is old and the other young, and to explain the difference between them. We know that the United States has lower birth rates as well as lower death rates; the question is, how much of the difference in ages is due to the difference in births, and how much to the difference in deaths?

Without the Cartesian decomposition, through the stable or some other model, no answer can be given to such a question. The rates of birth and death of both countries are what they are; their age distributions are what they are. We cannot even think of an experimental treatment of real populations that would answer the question, let alone perform one. But in a model we can vary one factor and see how the age distribution changes, and then vary the other and again see how the distribution changes. If the two kinds of change add up to the total change, we have decomposed the total; our model has paid off in providing without cost an experiment that tells the relative effects of birth and death on age when everything else remains unchanged.

Table 10.2 provides examples of the experimental treatment. For the stable model, with the female age-specific rates of birth and death observed in the United States in 1967, it shows 24.5 percent under 15 years of age (second item in second row). If we now alter the death rates to those of Madagascar, 1966, leaving the birth rates as they were, we find 22.0 percent under age 15. The higher death rates of Madagascar lower the proportion under 15 by 2.5 percentage points.

Now, retaining the death rates of the United States but entering in the stable model the birth rates of Madagascar, we find the proportion under age 15 rising from the original 24.5 percent to 48.6 percent. The higher birth rates of Madagascar raise the proportion under 15 by $48.6 - 24.5 = 24.1$ percentage points.

The two preceding paragraphs show that birth rates have about 10 times as much effect as do death rates on the proportion of a population under 15 years of age. As a check we take the differences in the other direction, starting with the same 45.2 percent for Madagascar births and deaths, subtracting 48.6 to find the effect of death rates at -3.4, and subtracting 22.0 to find the effect of birth rates at 23.2, so now births are 7 times as influential as deaths. The discrepancy is $(45.2-22.0)-(48.6-24.5) = -0.9$.

Table 10.2. Features of age distribution and rate of increase obtained by combinations of female birth and death rates from five countries, stable model

	Age-specific birth rates of:				
Age-specific death rates of:	Venezuela, 1965	United States, 1967	Madagascar, 1966	England and Wales, 1968	Sweden, 1803–07
		Percent under 15: $100{}_{15}C_0$			
Venezuela	47.7	23.9	47.8	23.6	34.2
United States	48.5	24.5	48.6	24.2	34.8
Madagascar	45.0	22.0	45.2	21.8	32.1
England and Wales	48.5	24.5	48.6	24.2	34.8
Sweden	43.6	21.0	43.8	20.8	31.3
		Dependency ratio percent: $100({}_{15}C_0 + {}_{\infty}C_{65})/{}_{50}C_{15}$			
Venezuela	102.1	58.8	102.4	58.7	70.3
United States	105.4	61.1	105.6	60.9	72.5
Madagascar	91.3	51.5	91.8	51.3	62.8
England and Wales	105.2	60.3	105.5	60.1	72.1
Sweden	85.6	46.7	86.2	46.6	58.9
		Percent 65 and over: $100{}_{\infty}C_{65}$			
Venezuela	2.8	13.1	2.8	13.3	7.1
United States	2.8	13.5	2.8	13.7	7.3
Madagascar	2.7	12.0	2.7	12.2	6.5
England and Wales	2.8	13.1	2.7	13.3	7.1
Sweden	2.5	10.9	2.5	11.0	5.8
		Intrinsic rate of natural increase per 1000: $1000r$			
Venezuela	38.5	4.9	38.6	4.5	19.5
United States	40.7	7.6	40.8	6.7	21.5
Madagascar	22.3	−11.3	22.5	−11.6	3.8
England and Wales	41.0	7.4	41.1	7.0	21.8
Sweden	24.3	−9.4	24.6	−9.6	6.4

This quantity is the interaction between birth and death: it is the difference between the effect of births in the presence of Madagascar deaths and the effect of births in the presence of United States deaths. It measures the uncertainty in the decomposition, and is small enough in this case not to affect our assertion that the age difference is due chiefly to the birth difference.

This may seem obvious for the proportion under age 15, which represents the births of the last 15 years and therefore ought to be closely related to the birth rate. More surprising is the outcome of the same analysis for the fraction of the population 65 years of age and older. Now we have for United States births and deaths 13.5 percent; a drop of only 1.5 to 12.0 for Madagascar with the change in deaths alone, and a drop of 10.7 to 2.8 with the change in births alone. Again births are the main factor, accounting for 8 times as much of the change as do deaths.

Similar calculations may be performed on Table 10.2 for the dependency ratio $({}_{15}C_0 + {}_{\infty}C_{65})/{}_{50}C_{15}$, where ${}_nC_x$ is the population between exact ages x and $x+n$, as well as for the intrinsic rate r. Variation along rows (due to births with deaths held constant) is everywhere greater than variation down columns.

One cannot but be puzzled on noting that historically the birth rate has changed little in the countries of Asia and Africa since the time when they had a much lower proportion of young people than they now have. With little change over time in births, and much change in deaths, rates of increase and age distribution have changed drastically; how is it that Table 10.2 shows births as the cause? The answer is that cross-sectional analysis need have no relation to longitudinal (see Section 12.4). It is the fact that people are always born at age zero, whereas they die at all ages, that gives the birth rate more leverage on age distribution in a cross-sectional analysis such as that of Table 10.2; the historical trend is mostly determined by the acceleration of population increase due to falling mortality at young ages.

10.2 Why There Are More Women Than Men at Older Ages in Modern Populations

On the whole the population of the United States contains more females than males, but the difference is by no means uniform through the several ages. Males are in excess to age 20, but from then on there are more females, with large proportional differences after age 65 (Figure 10.1). Can theory, combined with known facts of mortality, account for this difference between male and female age distributions? Specifically, can theory tell why in 1967 there are 2,236,000 men aged 70 to 74 against 2,941,000 women, a sex ratio of 76 males per 100 females? Why has the ratio of males to females in the United States declined steadily during the course of the present century? Essentially the same method of study that worked for age distribution will help to account for the varieties of sex ratios in observed populations.

Our search, as before, will start with the current age-specific rates of birth and death, now pertaining to the two sexes separately, from which a stable model may be constructed. What is left unexplained by the stable model based on current rates may be referred to trends in the rates in recent years; what is then still unexplained can be pursued by the study of migration and sudden changes in death rates due to wars, epidemics, and other historical events.

To convert the one-sex model used earlier for the analysis of ages into the (very primitive) two-sex model needed now, one item of data is required: the sex ratio at birth, say s. If s boys are born for each girl, and B girls are

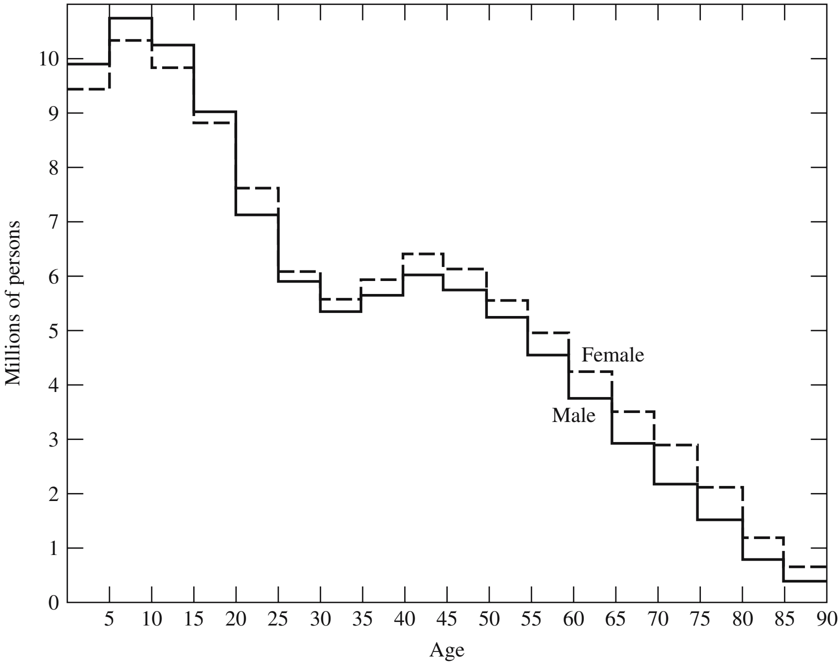


Figure 10.1. Male and female populations by age, United States, 1967.

born in the current year, then the number of boys born is sB. By the same argument as was used for (5.1.1) we can write down the sex–age distribution if the regime of mortality and fertility is fixed and population is increasing in a geometrical sequence with a ratio e^{5r} of increase over any 5-year period. Births, as well as each age–sex group, are all supposed to be increasing in this ratio; hence, if currently there are B female births per year, on an approximation to the stable population argument of Section 5.1 there will be $B{}_5L_0e^{-2\frac{1}{2}r}$ females aged 0 to 4, $B{}_5L_5e^{-7\frac{1}{2}r}$ aged 5 to 9, and similarly for later ages. We are here satisfied with the common approximation

$$\int_x^{x+5} e^{-ra}l(a)\,da \approx e^{-r(x+2\frac{1}{2})}\int_x^{x+5} l(a)\,da = {}_5L_xe^{-r(x+2\frac{1}{2})}.$$

The corresponding male births are sB, and male survivors at the end of the first 5-year period are $Bs{}_5L_0^*e^{-2\frac{1}{2}r}$, the same as for females but using s and the male life table function ${}_5L_0^*$. Thus the number of females is

$$B{}_5L_xe^{-(2\frac{1}{2}+x)r},$$

and of males is

$$Bs{}_5L_x^*e^{-(2\frac{1}{2}+x)r},$$

both in the age group x to $x+4$ at last birthday, where we suppose the two sexes to be increasing at the same rate r.

Then the sex ratio in the age group is

$$\frac{Bs_5L_x^*e^{-(2\frac{1}{2}+x)r}}{B_5L_xe^{-(2\frac{1}{2}+x)r}} = \frac{s_5L_x^*}{_5L_x},$$

in which everything has canceled but s and the life table survivors.

In application to the United States in 1967, we have $s = 1.050$, $_5L_{70} = 329{,}111$, and $_5L_{70}^* = 225{,}346$, so the stable model gives the sex ratio at age 70 as

$$\frac{(1.050)(225{,}346)}{329{,}111} = 0.72,$$

compared with 0.76 observed. The difference arises because the current life tables have a higher ratio of male to female mortality than the cohort in question experienced; immigration and other factors also operate.

10.3 Age at Marriage

A curve to describe proportion married with age, backed by a convincing rationale, is provided by Hernes (1972). In such an age-graded population it hardly occurs to individuals to marry when most of their contemporaries are unmarried; but as marriage gets under way after about age 18, each sees the number of his or her unmarried friends diminishing and begins to feel left out—experiences, indeed, an increasing pressure to marry. Hence, argues Hernes, we can take the rate of transfer into the married state as proportional to the fraction married. But it must also be proportional to the fraction not yet married; the marrying can come only from the stock of the unmarried. The result is a logistic curve of which Savage (1973) provides a generalization.

10.3.1 A Sum of Random Intervals Model

Coale (1971) found by experimenting with a number of observed distributions of age at marriage that all fitted a standard curve, provided only that the curve was adjusted for (a) the age at which girls become marriageable, (b) the proportion ever marrying by the end of life, and (c) horizontal scale. If the distributions are presented as risk functions, that is, age-specific marriage rates for the still-single population but excluding those who never marry, they take the form for first marriages

$$r(x) = 0.174e^{-4.411e^{-0.309x}}, \tag{10.3.1}$$

where x is the age of the person, measured from the origin for the particular population. Calculation yields the following values:

x	$r(x)$
0	0.0021
1	0.0068
2	0.0161
3	0.0304
4	0.0483
5	0.0679
10	0.1424
20	0.1724

This zero-parameter or three-parameter curve, depending on how one looks at it, gives close fit to data but, as Coale points out, lacks an obvious behavioral rationale.

Feeney (personal communication) suggested that the marriage curve may be composed of a random age of entry followed by a random delay. Coale and McNeil (1972) carried this thought further and developed a distribution involving three delays.

If the probability that X falls between x and $x + dx$ is $f_X(x)\,dx$, and the probability that Y falls between y and $y + dy$ is $f_Y(y)\,dy$; the joint probability is $f_X(x)f_Y(y)\,dx\,dy$. To find the distribution $f(z)$ of $Z = X + Y$ we need to integrate this over all x and y such that $x + y = z$, that is,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx. \tag{10.3.2}$$

This expression for $f_Z(z)$ is called a convolution, like (7.5.1), because the sum of the arguments of the functions in its integrand does not involve x, the variable of integration. It is readily applied to the sum of two negative exponential distributions, $f_X(x) = r_1e^{-r_1x}$ and $f_Y(y) = r_2e^{-r_2y}$, to obtain the distribution of $Z = X + Y$ as

$$f_Z(z) = \frac{r_1r_2}{r_1 - r_2}(e^{-r_2z} - e^{-r_1z}). \tag{10.3.3}$$

It may also be applied to more than two exponentials and to a normal curve followed by three exponentials, as long as independence holds.

This is how Coale and McNeil (1972) proceed to account for the distribution of marriage ages. They suppose a normal distribution of ages at which girls become marriageable, followed by three delays, each exponentially distributed. Fitting to data for French couples, they find the mean age of entry into a state of marriageability as 16.6 years, the mean interval between then and meeting the future husband as 4.02 years, the mean interval from acquaintance to engagement as 1.53 years, and the mean interval from engagement to marriage as 0.93 year. The closeness of fit of the convolution to the observed data for this and other populations confirms their behavioral model.

10.3.2 Small Marriage Circles

So much for marriage partners in a large population, where random variation in the time of individual marriage is recognized, but random variation in the number of men and women of marrying age can have only a minor effect. If in fact each individual, as he or she comes of age, seeks a marriage partner from a relatively small circle, some part of nonmarriage can be due to random differences in the numbers of men and women in these circles. Henry (1969) has addressed himself to this aspect of the problem.

He asks us to think of a circle containing candidates for marriage, whether or not they see themselves in that light. A circle may be an office or other workplace, a social club, or a neighborhood. Suppose that it contains $2n$ members, all of such age and disposition as to be candidates for marriage to one another. Let them be randomly distributed by sex; that is, the probability of a member being male is $1/2$, and of being female $1/2$. The expected number of male candidates would be n, and of females likewise n, so that in any realization of our hypothetical circle in which these expected values held all $2n$ candidates would marry. The task is to find the expected number of candidates that fail to find mates over all realizations of the circle. The following solution of Henry's problem is due to McFarland (1970).

If the chance of each individual being male is $1/2$, the chance that of the $2n$ members the numbers of males is $n+k$ and of females $n-k$ is

$$\binom{2n}{n+k}\left(\frac{1}{2}\right)^{2n}. \tag{10.3.4}$$

In this situation the nonmarrying males will number $2k = n+k-(n-k)$, $k > 0$. Hence the expected number of males not marrying must be $n+k-(n-k)$ times probability 10.3.4, added through all possible constitutions of the circle that have males in excess of females:

$$\sum_{1}^{n}[(n+k)-(n-k)]\frac{(2n)!}{(n+k)!(n-k)!}\left(\frac{1}{2}\right)^{2n} =$$

$$\left(\frac{1}{2}\right)^{2n}(2n)!\left[\sum_{1}^{n}\frac{1}{(n+k-1)!(n-k)!}-\sum_{1}^{n-1}\frac{1}{(n+k)!(n-k-1)!}\right]. \tag{10.3.5}$$

We have made the upper limit in the right-hand summation $k = n-1$, because $k = n$ gives a zero term owing to the factor $n-k$. Now note that the term for $k = 2$ in the left summation cancels with that for $k = 1$ on the right, both being equal to $1/(n+1)!(n-2)!$. All other terms similarly cancel, except that for $k = 1$ on the left, which is equal to $1/n!(n-1)!$. Hence the required expected number of males left over is

$$\left(\frac{1}{2}\right)^{2n}(2n)!\frac{1}{(n)!(n-1)!} = n\left(\frac{1}{2}\right)^{2n}\frac{(2n)!}{n!n!}. \tag{10.3.6}$$

The expected number of females left over must be the same by symmetry, and hence for both sexes we have an expected

$$2n\left(\frac{1}{2}\right)^{2n}\frac{(2n)!}{n!n!}$$

persons unmarried or, as a fraction of the $2n$ individuals in the circle,

$$\left(\frac{1}{2}\right)^{2n}\frac{(2n)!}{n!n!}. \tag{10.3.7}$$

McDonald (1965) finds the same result (10.3.7) for the matching of male and female worms in the spread of certain infestations. For no readily understandable reason, the expected fraction of marriage candidates (or worms) unmatched is the same as the probability that the group is equally divided between males and females.

Stirling's approximation to the factorial is $n! = n^n e^{-n}\sqrt{2\pi n}\,[1+(1/12n)]$, omitting terms from $1/n^2$ onward. Entering this in (10.3.7) gives

$$\left(\frac{1}{2}\right)^{2n}\frac{(2n)^{2n}e^{-2n}\sqrt{4\pi n}\,\left[1+(1/24n)\right]}{\left\{n^n e^{-n}\sqrt{2\pi n}\,\left[1+(1/12n)\right]\right\}^2},$$

which after cancellation becomes

$$\frac{1}{\sqrt{\pi n}}\,\frac{1+1/(24n)}{\left(1+1/(12n)\right)^2}\approx\frac{1-1/(8n)}{\sqrt{\pi n}}. \tag{10.3.8}$$

Thus, if a marriage circle containing $2n$ individuals is considered, the expected fraction of excess of one sex or the other, that is to say, the average fraction who cannot marry within the circle, is slightly under $1/\sqrt{\pi n}$. For a group of 20 individuals this is $1/\sqrt{\pi n} = 0.178$, reduced to 0.176 by the factor $1 - 1/(8n)$ (Table 10.3 and Figure 10.2).

This solution may be extended to a circle having a disequilibrium of the sexes, in the sense that the probability of a random individual being male is not 1/2 but some other number; in this case, another device is required to sum the series.

10.3.3 How Many Households Are Implied by Birth, Death, and Marriage Rates?

Age-specific rates of birth and death, along with the fraction of women married at each age, imply an average size of nuclear family (Goodman, Keyfitz, and Pullum 1974).

For the general case, call the age distribution of women $p(a)$, not necessarily stable, so that $\int_0^\omega p(a)\,da = N^f$, and the fraction married at age a is $q(a)$; here the total number of married women is $\int_0^\omega p(a)q(a)\,da$. It is reasonable to attach a separate family to each married woman, where for this purpose only those currently living with a spouse will be defined as

Table 10.3. Values of $(2n)!/(2^{2n}n!n!)$ and Stirling approximations

Size of group $2n$	$(2n)!/(2^{2n}n!n!)$	$1/\sqrt{\pi n}$	$\left(1-\dfrac{1}{8n}\right)\Big/\sqrt{\pi n}$
2	0.5	0.5642	0.4937
4	0.375	0.3989	0.3740
6	0.3125	0.3257	0.3122
8	0.2734	0.2821	0.2733
10	0.2461	0.2523	0.2460
12	0.2256	0.2303	0.2255
14	0.2095	0.2132	0.2094
16	0.1964	0.1995	0.1964
18	0.1855	0.1881	0.1855
20	0.1762	0.1784	0.1762
40	0.1254	0.1262	0.1254
80	0.08893	0.08921	0.08893
120	0.07268	0.07284	0.07268
160	0.06298	0.06308	0.06298
200	0.05635	0.05642	0.05635

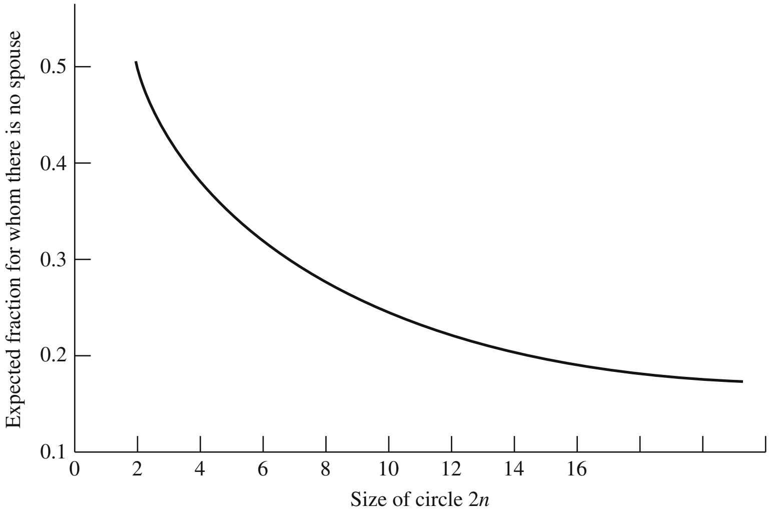


Figure 10.2. Diminishing fraction of expected number without spouse as marriage circle increases.

married. Then the total number of families is also $\int_0^\omega p(a)q(a)\,da$, and the mean number of persons per family is

$$\frac{2N^f}{\displaystyle\int_0^\omega p(a)q(a)\,da} \tag{10.3.9}$$

if the numbers of males and females in the population are equal.

To go from this to the mean size of family implied by age-specific rates in a stable population, we replace $p(a)/N^f$ by $be^{-ra}l(a)$, as in Section 5.1. Then we have for the mean number of persons per family

$$\frac{2}{\displaystyle\int_0^{\omega} be^{-ra}l(a)q(a)\,da}. \tag{10.3.10}$$

This is different in concept from the average we would derive from the census itself by counting persons in families and dividing by the number of families. Our (10.3.10) neglects persons living not in families but in prisons and other institutions, and implicitly somehow allocates illegitimate children to married women, but this is a minor divergence from the census. The main difference is that (10.3.10) gives what is implied by the current rates rather than by rates that have existed over the lifetimes of the people counted in a census. To reconcile these two kinds of average would be a difficult but worthwhile piece of research.

What about the size of the extended family, supposing that a separate extended family is formed by a married woman if her mother is dead? We need merely enter in (10.3.10) the probability that a woman aged a does not have a living mother, a subject to be developed in Chapter 15. If the probability that a woman aged a has a living mother is $M_1(a)$ [an expression for this is given later as (15.1.4)], the chance that she has not is $1 - M_1(a)$, and from (10.3.9) extended families will average

$$\frac{2N^f}{\displaystyle\int_0^{\omega} [1 - M_1(a)]\,p(a)q(a)\,da} \text{ persons,} \tag{10.3.11}$$

or in the stable case

$$\frac{2}{\displaystyle\int_0^{\omega} [1 - M_1(a)]\,be^{-ra}l(a)q(a)\,da} \text{ persons.} \tag{10.3.12}$$

10.3.4 Intrinsic Rates of Natural Increase: Age, Parity, and Nuptiality

The intrinsic rate of growth, as defined by Lotka and discussed in Sections 6.1 and 7.2, is that which ultimately results from the continuance of observed age-specific rates of birth and death. The intrinsic rate depends in no way on the observed age distribution, but is a means of interpreting the age-specific rates. And yet the choice of age-specific rates was in a sense arbitrary; we could have used other classifications—say rural–urban-specific rates, classified by age or not. Any characteristic that shows differentials of birth and death rates is a candidate for providing an intrinsic rate, although not all characteristics are equally logical candidates.

Oechsli (1971) has calculated rates intrinsic for age, parity, and nuptiality, and for certain combinations of these. The age–parity rate, for instance, tells what the increase of the population would be once its age–parity distribution came to be that resulting without disturbance from the observed set of rates in the several age–parity groups.

To see the meaning of various directions of adjustment consider the following five rates (per 1000 population) for United States females in 1960:

Crude	Crude rate of natural increase	14.7
A	Age intrinsic rate	20.8
A–P	Age–parity intrinsic rate	23.0
A–N	Age–nuptiality intrinsic rate	18.4
A–N–P	Age–nuptiality–parity intrinsic rate	19.6

The rise when we go from the crude rate in the first line to the age intrinsic rate in the second means that the observed age distribution was unfavorable to increase—that there was a smaller proportion of women of childbearing age than are present in the stable condition at 1960 age-specific rates of birth and death. This is a well-known result of the small cohorts born in the 1930s.

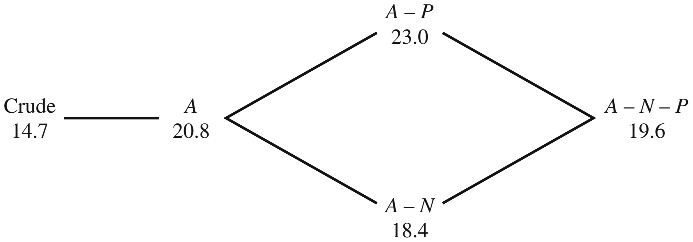
When we go from the age intrinsic to the age–parity intrinsic rate, we find a further increase, which can mean only that within age groups the distribution of observed parities was unfavorable to reproduction. The highest-bearing parities are the low ones; apparently the stable condition on 1960 age-parity specific rates would have relatively more individuals of low parity than the actual situation of 1960.

On the other hand, when we go from the age intrinsic rate to the age–nuptiality intrinsic rate (from 20.8 to 18.4), we find a decline; this means that the proportion of married women in the population of 1960 was greater (in the main ages of childbearing) than the proportion of married women in the stable condition. If the marriage rates, the age-specific nuptial fertility rates, and the age-specific mortality rates of 1960 were allowed to work themselves out, they would result in a lower overall birth rate than would the age-specific fertility and mortality rates of 1960, disregarding marriage.

The fall from the age–parity to the age–nuptiality–parity says that separation of the married women in the age–parity analysis brings the rate down, and the effect is so drastic that we find *A–N–P* below the simple age intrinsic rate *A*. In short, the effect of nuptiality more than offsets the effect of simple parity. Karmel (1950) suggested in his comment on Whelpton (1946) that the parity correction by itself could be deceptive. Suppose from one year to the next a rise in marriages, while age–parity rates at each marriage duration remained the same. The age–nuptiality–parity intrinsic rate would be unchanged, whereas both the age intrinsic and the age–parity intrinsic rates would rise. But the latter would rise more, for the higher first births would be divided by zero-parity women rather than, as in the age-specific rate, being diluted by division by all women.

The rise from the 18.4 in A–N to the 19.6 in A–N–P means that within the married group, age by age, the parity distribution of the observed population was less favorable to fertility than the parity distribution of the stable condition.

All of this can be shown on a diagram. The lozenge form brings out the contrast between the effects of parity and of nuptiality (A = age; P = parity; N = nuptiality):



10.3.5 The Life Cycle

According to Glick's calculation (1955), the following changes in median ages of various events in American families took place between 1890 and 1950:

	Median age of husband		Median age of wife	
Event	1950	1890	1950	1890
First marriage	22.8	26.1	20.1	22.0
Birth of last child	28.8	36.0	26.1	31.9
Marriage of last child	50.3	59.4	47.6	55.3
Death of one spouse	64.1	57.4	61.4	53.3
Death of other spouse	71.6	66.4	77.2	67.7

Whereas in 1890 a minority of couples survived until the last child was married, by 1950 most survived, on the average by about 14 years. A new demographic element came into existence: husband and wife living together after their last child had left home—the empty nest. This corresponds to a new pattern of spatial movement as well; a couple may start out in an apartment in the city, move to a suburb when their first child is born, and remain there until their last child has left home. If the husband is middle class, his income has usually continued to climb, and his earnings are higher than when the children were at home; moreover, his wife may well take up some new kind of paid work or resume an earlier career. The couple no longer need their suburban home and schools and are in a position to move into town and pay a substantial rent. Collectively such couples have affected the skylines of many American cities as new high-rise, high-rent apartments spring up to accommodate the new demographic element.

The change in life cycle is due to improved mortality, especially for women, along with earlier marriage in both the older and the younger generation—a 1-year fall in the age of marriage subtracts 2 years from the age at which the last child is married, everything else remaining the same. Moreover the birth rate has fallen so that children are fewer; and, especially if the wife is eager to resume her career, the children that couples do have are more closely spaced. To all this is added an increased tendency for children to leave home before marriage.

We thus know to what factors the changes are due, but we do not know how much of a particular change is due to each factor. Yet it need not be difficult to break down any part of the difference. In 1890, for husbands, the first death of spouse occurred 2 years before the last child was married; in 1950 it was 14 years after. One would like to know how much of the change was due to improved mortality, how much to earlier marriage, and so on. These effects would be ascertained by the technique exhibited in Table 10.2. The first step would be to set up a model, including death rates, marriage rates, home-leaving rates, birth rates, all by age. Then the model with the rates for 1890 would be used to generate the 1890 column of Glick's table, or an approximation to it. The same would be done for the 1950 column. Then the various elements of the models would be interchanged; for example, the marriage rates for 1950 would be combined with the other elements for 1890 to obtain an estimate of the effect of changed marriage ages. The result would be somewhat more elaborate to interpret than Table 10.2, since it would have several orders of interaction, but no new principle would be required. See Chapter 13 for other ways to decompose effects of factors and their interactions on the rate of increase.

This demographic theory of the life cycle needs to be further developed, and then integrated with the theory of savings and consumption over the life cycle that has become a standard part of macroeconomics. Stage-classified matrix population models would certainly be applicable to it.

10.3.6 Married and Divorced

The numbers of married and divorced persons counted in a census, like those of the several ages and sexes, are related to the country's previous history of marriage and divorce. It is necessary to consider the time series of marriages of preceding years, along with other series on the dissolution of marriages by death of one spouse or divorce. The present is in principle explainable by and reconcilable with the past. If the reconciliation and explanation are in a small local area, one would have to take account of immigration and emigration as well as marriages, deaths, and divorces, but this complication will be disregarded here. The discussion that follows is due to Preston (1975b).

Still for simplicity, disregard the age of the individuals concerned, and take account only of duration of marriage. (First marriages fall in a relatively narrow band of ages, and the theory will apply to them more strictly than to all marriages.) Call $\mu^m(x)\,dx$ the chance that a marriage that has been in existence x years will break up during the time x to $x+dx$ through the death of the husband, $\mu^f(x)$ that it will break up through the death of the wife, and $\mu^d(x)$ that it will break up through the divorce of the couple. Then the argument that started in Section 1.6, of which a special case was the survivorship of an individual in the face of death (Section 2.1), is here applicable to the survivorship of a couple in the face of the death of either member or of divorce. It is as though a group of marriages, initially N_0 in number, had to run the gauntlets of death and divorce, and only a certain fraction survived in each time period. The surviving couples at the end of t years would be

$$N(t) = N_0 \exp\left\{-\int_0^t \left[\mu^m(x) + \mu^f(x) + \mu^d(x)\right]\,dx\right\}. \tag{10.3.13}$$

The expected number of divorces between time t and $t + dt$ would be $N(t)\mu^d(t)\,dt$, assuming independence, and disregarding remarriage.

These causes of dissolution compete with each other. Higher mortality lowers the divorce rate, everything else remaining the same. The reason for this is to be seen in the expression for the number of marriages that will eventually end in divorce:

$$D = \int_0^\infty N(t)\mu^d(t)\,dt.$$

This, divided by the number N_0 of marriages that started the cohort off, gives the fraction expected to end in divorce. If, as between two populations otherwise similar, mortality is higher at any age in one, $N(t)$ will be lower in that one for all subsequent ages, so D will also be lower. This is analogous to the way that deaths from cancer in the multiple-decrement table of Section 2.6 depend on deaths from other causes.

The Current Divorce–Marriage Ratio. The ratio of current divorces to current marriages, say for the year 1973, is a convenient measure of the extent of divorce. This measure is often used, and almost as often criticized on the ground that current divorces come from the marriages of a period going back a number of years, and hence the denominator of the ratio ought to be these marriages. The ideal is a schedule of divorce rates according to duration of marriage, so that duration-specific divorce rates can be calculated. But when these are not available and we resort to the current divorce–marriage ratio, we can apply an interpretation of this ratio that has been provided by Samuel Preston.

Let $D(t)$ be the curve of divorces with time, and $M(t)$ the curve of marriages. Again let the (unknown) continuous duration-specific force of divorce be $\mu^d(x)$ for marriages that have lasted x years, and call the probability of survival of the marriage for x years against the contingencies of both death and divorce $p(x)$. Then the current number of divorces $D(t)$ must be

$$D(t) = \int_0^\infty M(t-x)p(x)\mu^d(x)\,dx, \tag{10.3.14}$$

and under a fixed rate of increase r the current number of marriages is

$$M(t) = M_0 e^{rt}. \tag{10.3.15}$$

Therefore, on entering $M(t-x) = M_0 e^{r(t-x)}$ in the expression (10.3.14) for $D(t)$, we have

$$D(t) = \int_0^\infty M_0 e^{rt} e^{-rx} p(x)\mu^d(x)\,dx. \tag{10.3.16}$$

From (10.3.15) and (10.3.16)

$$\frac{D(t)}{M(t)} = \int_0^\infty e^{-rx} p(x)\mu^d(x)\,dx. \tag{10.3.17}$$

How much does this ratio depend on the rate of increase of marriages r? Taking the logarithm and then differentiating both sides of (10.3.17) gives the simple result

$$\frac{d}{dr}\left\{\log\left[\frac{D(t)}{M(t)}\right]\right\} = -A_D, \tag{10.3.18}$$

where A_D is the mean number of years of marriage at the time of divorce, calculated for the part of the population that ultimately does divorce. One could confine the argument to first marriage and first divorce, or else to first marriage and any divorce, but to take all marriages and all divorces accords with the data most commonly available.

Integrating (10.3.18) gives

$$\log\left[\frac{D(t)}{M(t)}\right] = \log\left[\frac{D_0(t)}{M_0(t)}\right] - rA_D,$$

where $D_0(t)/M_0(t)$ is the ratio in the stationary population. Taking exponentials provides

$$\frac{D(t)}{M(t)} = \frac{D_0(t)}{M_0(t)} e^{-rA_D}, \tag{10.3.19}$$

which would have been obtained more directly as an approximation to (10.3.17).

For populations that have not experienced sudden changes in the marriage rate (10.3.19) enables us to infer the ratio in the stationary population, $D_0(t)/M_0(t)$, which is the same as the probability that a marriage will end in divorce. The mean number of years married for those who divorce was constant through the 1960s at about 7 (*Statistical Abstract for the United States*, 1972, p. 3). Suppose a rate of increase of about 1 percent per year. Then Preston's result (10.3.19) can be used to calculate the probability that a random marriage will end in divorce, assuming values for r and A_D and applying them to the observed $D(t)/M(t)$:

$$\frac{D_0(t)}{M_0(t)} = \frac{D(t)}{M(t)} e^{rA_D} = \frac{D(t)}{M(t)} e^{(0.001)(7)} = 1.073 \frac{D(t)}{M(t)}. \tag{10.3.20}$$

This result is directly applicable only if birth, marriage, divorce, and death rates have all been fixed in the past. But even under instability in these rates (10.3.20) is useful, for the birth rates really do not matter; what is necessary is that marriages have been increasing in the past at a reasonably steady and known rate.

10.4 The Foreign-Born and Internal Migrants

The 1970 census gave 9,619,000 foreign-born in the United States, a decline from previous censuses and considerably less than the 11,595,000 reported in 1940. If the flow of immigrants (net of those leaving) had been constant and at constant ages, improvements in mortality would have led to a steady increase in the number alive to be counted in the census; the reason for the decline must have been a greater decline in entrants than the increase in fraction surviving. But the preceding decade's immigration records had shown 3,322,000 entrants, the peak to that time in a steady rise from a low of 528,000 immigrants in 1931–40. The number of immigrants entering the United States rose sharply in the postwar period; why, then, were fewer and fewer foreign-born counted in successive censuses?

The answer requires a longer view backward than the postwar period. The 40 years from 1931 to 1970 showed entrants numbering 7,400,000 against 22,326,000 for the preceding 40 years. This contrast is what was reflected in the declining immigrant population; the years since the 1930s had witnessed the dying off of the immigrants from a time earlier in the century. Immigrants are typically young adults who have on the average 40 or more years to live after their arrival.

Any statistical reconciliation is rendered more difficult yet by the considerable volume of illegal immigrants, who are certainly not included in the annual official numbers of entrants; some unknown fraction are picked up by the census.

10.4.1 A Matrix Analysis

The next stage in analysis, beyond the accounting for present numbers of foreign-born in terms of the preceding stream of immigration, is to project the consequences for future population of the present rates of immigration or of some hypothetical change in them. This is easily done using matrix population models (Chapter 7) that classify individuals by location instead of (or in addition to) age. Such models go beyond description and reconciliation of reports into at least a primitive kind of mechanism: the effect of fixed (or changing) rates on the pattern of residents from abroad, where "abroad" may mean another country, state, or county.

Rogers (1968; see also 1995) provides an exposition of the theory and examples of the migration process for the United States. As an example he considers 1955–60 migration and natural increase between California and the rest of the United States with the matrix

$$\mathbf{A} = \begin{pmatrix} 1.0215 & 0.0127 \\ 0.0627 & 1.0667 \end{pmatrix}.$$

He starts the process with the initial 1955 vector in thousands:

$$\mathbf{n}(1955) = \begin{pmatrix} 12{,}988 \\ 152{,}082 \end{pmatrix}$$

and finds a 1960 distribution between California and the rest of the United States equal to

$$\mathbf{n}(1960) = \begin{pmatrix} 1.0215 & 0.0127 \\ 0.0627 & 1.0667 \end{pmatrix} \begin{pmatrix} 12{,}988 \\ 152{,}082 \end{pmatrix} = \begin{pmatrix} 15{,}199 \\ 163{,}040 \end{pmatrix}.$$

Continuation of the process, or use of eigenvalues, gives the ultimate ratio of California to United States population if these rates continue.

10.4.2 Migration and Age

Among other simplifications the models of the preceding paragraphs suffer from the omission of age. Rogers goes on to show (1968, p. 10) that one can make each of the a_{ij} itself a matrix, with nonzero elements in the first row and subdiagonal, and so he combines the migration and age effects in a single large matrix. With 50 states and 18 ages for females the matrix would be 900 × 900.

Arrangement of the elements within $\mathbf{A}$ and $\mathbf{n}(t)$ is arbitrary, as long as the same order is used in both. One could put the states in alphabetical order or in order of their 1970 populations, and even the ages could be rearranged with no loss other than inconvenience in remembering the sequence. Moreover the ages need not be a subdivision of the states; the states could just as well be a subdivision of the ages. In fact Feeney (1970) shows the advantage of considering the ages separately, making a block of the interstate migration for ages 0 to 4, then another block for ages 5 to 9, and so on. The block matrix for ages is then largely empty, and its properties are readily worked out. More recently, Hunter and Caswell (2004) have shown how to easily construct multiregional models with either classification, using an operator called the vec-permutation matrix. Their approach maintains a useful block-diagonal structure for the projection matrices.

The literature on migration and settlement is too extensive to refer to here in any complete way, let alone to summarize. Classic papers on the mathematics are due to Blumen, Kogan, and McCarthy (1955), Goodman (1961b), and McGinnis (1968). Alonso (1978), Lowry (1964), and Stone (1971) provide theory on rural–urban movements, and an annotated bibliography is due to Price and Sikes (1975). The economics of migration is treated by David (1974) and Todaro (1969), and geographical aspects by Berry (1973) and Zelinsky (1971). Rogers (1984) explores important aspects of the age-dependence of migration.

Of interest as potentially unifying migration theory and integrating it with multiple decrement life tables, tables of the work force, and tables of marital status, is the work of Rogers (1975, 1995) and Rogers and Willekens (1976). Their matrix formulation simplifies and generalizes the work of Schoen (1975). For other approaches using matrix formulations, see Lebreton and Gonzalez-Davila (1993), Lebreton (1996), and Hunter and Caswell (2004). Alonso (1978; see de Vries et al. 2000) has developed a general framework that holds promise of being able to handle nonlinearities that are important in the real world.

10.5 Human Stocks and Flows

The study of population involves relating stocks to flows, as Richard Stone (1972) reminds us. Censuses report the stocks, according to the categories of age, marital status, education, occupation, and other characteristics, while vital statistics, numbers of graduates provided by schools, and similar current data report the flow. In contrast to economic data, which are more plentiful with respect to flows, social statistics are richer and higher in quality with respect to stocks. Relating the two is important in all fields.

Underlying all social statistics are individuals passing through the several states of their life cycles. They begin by being born, as reported in a flow series, spend some years at home, go into the educational system at the elementary level, sooner or later move to the intermediate and higher levels, eventually enter the labor force, where they may stay briefly or long, perhaps go back to school, and ultimately retire, possibly after several entries into and withdrawals from the labor force. At the same time they pass from a household of orientation, are married and form a household of procreation, have children, divorce, remarry, become widowed, and go through other domestic transitions. If it were useful, we could consider each economic activity state classified by the several domestic states, but it is less demanding of data to avoid this and to use two models in which the two collections of states are separately treated.

The two sequences outlined above through which individuals go are referred to by Stone as the *active* sequence, relating the individual to production and the market, including schooling and retirement, and the *passive* sequence, including the successive types of household of which he is part, the housing conditions, and his neighborhood and location. A further sequence is that of *health and medical care*, as the individual passes through the stages of being well and sick, encounters successive medical practitioners, is hospitalized and released. Some individuals go also through a sequence of *delinquency* states, starting with aberrant behavior and its consequences as they pass through the hands of police, courts, and prisons. These can all be treated as cases of *stage-classification* (Chapter 3), and approached via matrix models.

For each of the states in each of the four sequences mentioned above, a precise definitional boundary is required for purposes of measurement. For education it happens to be relatively easy to draw a sharp boundary around formal full-time schooling, and this is where the boundary is usually placed for statistical purposes; the important component of home and part-time learning is neglected. A conflict between precision of measurement and relevance of concept exists in any empirical science. As another example, it is sharper to restrict statistically economic production to that which goes through the market, locating self-transport to place of work and other do-it-yourself contributions outside the boundary of production.

To introduce the notation and ideas with an uncomplicated example, consider just two states, alive and dead, and the corresponding flows of birth and death. The statement is an adaptation of the Leslie (1945) model to a stationary population. Suppose a matrix $\mathbf{S}$ in which the subdiagonal elements are the probabilities of surviving from one year of age to the next, so that the first is, in the notation of Chapter 2, L_1/L_0, the second L_2/L_1, and so on; all elements but the subdiagonal are zero. The new population that enters each year through birth or migration is a vertical vector $\mathbf{b}$. Suppose also that $\mathbf{b}$ consists of the number L_0 in its first position and zeros elsewhere. Then, if the population at time t is (t), also a vertical vector, we have

$$\begin{aligned}
\mathbf{n}(1) &= \mathbf{S}\mathbf{n}(0) + \mathbf{b} \\
\mathbf{n}(2) &= \mathbf{S}\mathbf{n}(1) + \mathbf{b} = \mathbf{S}^2\mathbf{n}(0) + \mathbf{S}\mathbf{b} + \mathbf{b} \\
&\vdots
\end{aligned}$$

After the process has been going on for a long time, the term in $\mathbf{S}^t\mathbf{n}(0)$ will have vanished; in fact all powers of $\mathbf{S}$ beyond the $(m-1)$th, if it is an $m \times m$ matrix, will consist entirely of zeros. Similarly, if t is m or greater, such terms as $\mathbf{S}^t\mathbf{b}$ will be zero, so we can add them in with impunity. This gives us the series

$$\mathbf{n}(t) = \mathbf{b} + \mathbf{S}\mathbf{b} + \mathbf{S}^2\mathbf{b} + \cdots \qquad t > m \tag{10.5.1}$$

$$= [\mathbf{I} + \mathbf{S} + \mathbf{S}^2 + \cdots]\mathbf{b}, \tag{10.5.2}$$

which can be thought of as infinite. The sum of a geometric series of matrices is obtained in the same way as the sum of a series of scalars of which the common ratio is less than unity. Let the sum be **N**:

$$\mathbf{N} = \mathbf{I} + \mathbf{S} + \mathbf{S}^2 + \cdots.$$

Multiplying **N** by **S** gives the same series, except that the **I** is missing; hence we have

$$\mathbf{N} = \mathbf{N}\mathbf{S} + \mathbf{I},$$

or, on subtracting **NS** from both sides and factoring out **N**,

$$\mathbf{N}[\mathbf{I} - \mathbf{S}] = \mathbf{I},$$

and, on multiplying on the right by the inverse $[\mathbf{I} - \mathbf{S}]^{-1}$,

$$\mathbf{N} = [\mathbf{I} - \mathbf{S}]^{-1}.$$

Entering this result in (10.5.2) gives for the population distribution by age at time t

$$\mathbf{n}(t) = [\mathbf{I} - \mathbf{S}]^{-1}\mathbf{b}. \tag{10.5.3}$$

The factor multiplying **b**, which is called the fundamental matrix of the absorbing Markov chain implied by the matrix **S**, will appear again in Chapter 11.

For our particular matrix **S** giving survivorships as the transitions between successive ages, the determinant of $[\mathbf{I} - \mathbf{S}]$ is readily seen to be unity, and the inverse is the transpose of the cofactors of the elements of $\mathbf{I} - \mathbf{S}$; that is to say, for the jth element of the ith row of the inverse we enter the determinant obtained by deleting the ith column and jth row of $\mathbf{I} - \mathbf{S}$. This procedure provides the fundamental matrix of the survivorship matrix:

$$[\mathbf{I} - \mathbf{S}]^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ \dfrac{L_1}{L_0} & 1 & 0 & 0 & \cdots \\ \dfrac{L_2}{L_0} & \dfrac{L_2}{L_1} & 1 & 0 & \cdots \\ \dfrac{L_3}{L_0} & \dfrac{L_3}{L_1} & \dfrac{L_3}{L_2} & 1 & \cdots \\ \vdots & & & & \end{pmatrix}. \tag{10.5.4}$$

The elements of the first column provide the probability of attaining the successive ages, starting at age zero; the elements of the second column give the same probabilities, starting at age one, and so on for the several columns. The totals of the columns are the expectations of further life, starting in the first age, the second age, and so on. Expressed formally, if $\mathbf{e}$ is the column vector consisting of ones then the ith element of $\mathbf{e}^{\mathsf{T}}[\mathbf{I}-\mathbf{S}]^{-1}$, is the expectation of further life for a person who has attained the ith age. (In contrast to most demographic and actuarial work, this way of developing the subject takes individuals in midstate rather than completed ages. Unless the states were age intervals much shorter than 1 year, calling the midstates $\frac{1}{2}$, $1\frac{1}{2}, \ldots$ intervals would not make the expectations quite equal numerically to those calculated in the usual life table. The difference is of no consequence for the present purpose.)

Merely to arrive at a somewhat inaccurate version of the life table, the matrix analysis above is cumbersome as well as superfluous. Its advantage is that it gives stocks (of people) in terms of flows, not only for death but also for many other conditions, and that it permits a connection to powerful results on absorbing Markov chains (Chapter 11).

We were able conveniently to sum the powers of $\mathbf{S}$ because $\mathbf{S}$ vanishes beyond a certain power. That does not happen in general, but in many situations the elements of the powers become smaller and smaller in something like a diminishing geometric progression. They do this when the matrix $\mathbf{S}$ is part of an absorbing chain; our matrix $\mathbf{S}$ above is part of an absorbing chain in the sense that everyone is absorbed by death sooner or later in the course of his progress through the several ages. See Chapter 11 for a complete description of these methods.

10.6 The Demography of Organizations

This book deals for the most part with large populations. It presents theory suited to the understanding of national populations, like those of the United States or Mexico. But this is an inappropriate restriction; every city ward, corporation, school, hospital, or seaside resort has its demography. The operations of communities and institutions depend on a flow of people, whose entries are analogous to births and their departures to deaths. For institutions that have relatively well-determined posts, differentiated by age and sex like a national population and in addition by skills and other role requirements, a thoroughgoing demography would be an extensive subject, and the pages that follow are barely an introduction to it. They draw from contributions due to Bartholomew (1982), Coleman (1973), March (1975), and Waugh (1971). Their claim to a place in this chapter is the help they offer in understanding distributions of income and similar variables.

We tackle the sheer size aspect first in political, then in economic, organizations.

10.6.1 Loss of Power

One aspect of the number of persons in an organization has been called by Coleman "loss of power," the diminishing influence of the individual as the organization grows in size. This appears clearly in voting, where the individual has a deciding voice only when his associates are evenly divided. The probability of such even division can be shown to be inversely proportional to the square root of the size of the voting group. If there are 100,000 individuals aside from Ego, the chance that they will split evenly is approximately one-tenth as great as if there were 1000 individuals. This inverse square root law is easily derived.

With $2n+1$ members voting, the chance that Ego's vote will be decisive is the same as the chance that the other $2n$ individuals will split n for and n against whatever motion is under debate. The number of ways in which this can happen is the number of ways of putting $2n$ objects equally into two boxes, that is, $(2n)!/n!n!$. The total number of combinations of the $2n$ votes is 2^{2n}, since each could vote "yes" or "no." With equal chances or each vote the probability that the $2n$ votes will be evenly split is

$$p = \left(\frac{1}{2}\right)^{2n} \frac{(2n)!}{n!n!}.$$

The solution happens to be formally identical with that for the fraction unmarried in (10.3.7).

Entering as before the Stirling first approximation to the factorial $n! = n^n e^{-n}\sqrt{2\pi n}$, we find that everything cancels except $1/\sqrt{\pi n}$. With $n = 3$, $1/\sqrt{\pi n}$ is 0.3257 against the exact 0.3125; with $n = 5$, $1/\sqrt{\pi n}$ is 0.2523 against the exact 0.2461. For higher values of n the error of the approximation is negligible, as Table 10.3 shows. This establishes the proposition that a person's power goes down as the square root of the size of the group within which he is operating provided he is the last one to vote; in the symmetrical condition before the voting starts each member can have no more and no less than $1/n$ of the power.

10.6.2 Organizing Political Success

As Coleman points out, one means by which an individual can apply social skill to attain power is by restructuring the electorate. In real life an electorate, whether it be a local boys' club, a town, a nation, or the Congress of the United States, is made up of cliques, factions, committees, parties, and other subgroups. In the smallest cell the probability $1/\sqrt{\pi n}$ provides an appreciable chance to the individual. Ego may well gain a sense of power through having the decisive vote in the clique with relative frequency, even though the decisions concern small matters such as how the clique as a whole will vote in a larger contest. If the voting is to be by groups, like an electoral college, Ego's chance of being the deciding voter is little different from that in individual voting. But evidently a person gains power nearly in the ratio of $\sqrt{2} = 1.41$ by being a member of a coalition containing half the members plus one. He gains if he becomes a member of a coalition of any size that votes together while others vote individually. Coleman proves that the size of the coalition that maximizes the person's power is considerably less than the 51 percent that gives absolute control.

10.6.3 Economic Hierarchies

Economic organizations often fall naturally into hierarchies by virtue of the technical requirements of production and decision making. To simplify a typical organization, assume that there is one head of it, he has s persons reporting directly to him, and each of these has s persons reporting to him in turn, where s can be a number like 5 and is called the span of control (Cramer 1972). If there are n levels then counted from the top downward the personnel of the organization will be distributed as follows:

Level	1	2	3	$\cdots$	n
Number of employees	1	s	s^2	$\cdots$	s^{n-1}

and the total personnel, say at time t, will be $N(t) = (s^n - 1)/(s - 1)$. Then, if a retirement takes place at the ith level, the chance of promotion for each individual at the $(i+1)$th level is $1/s^i$, which is a very small number for the lower echelons of the hierarchy. The theory here does not have any simple relation to that relating promotion and age of Sections 5.8 and 14.10.

Note that, as long as s is greater than 1, there are more persons in the lowest level than in all higher levels together. If seniority in promotion is maintained, the halfway point between bottom and top will always be within the lowest level. A person in an organization with $s = 5$ will be 80 percent of the way to the top by the time he is promoted out of the lowest level, when the scale is in terms of individuals. But when the scale is in terms of levels, he will be only $1/n$ of the way to the top.

Suppose that the organization starts with N_0 employees, in k layers; then

$$N_0 = 1 + s + s^2 + \cdots + s^{k-1} = \frac{s^k - 1}{s - 1};$$

and if it is growing at rate r, its size at time t can be expressed as $N(t) = N_0 e^{rt}$. Equating this to the above expression $N(t) = (s^n - 1)/(s - 1)$ gives

$$\frac{s^n - 1}{s - 1} = \left(\frac{s^k - 1}{s - 1}\right) e^{rt}. \tag{10.6.1}$$

Even for a small population the error in substituting s^n for $s^n - 1$ and s^k for $s^k - 1$ is negligible. The substitution permits solving (10.6.1) for n:

$$s^n = s^k e^{rt},$$

so that

$$n = k + \frac{rt}{\log s}, \tag{10.6.2}$$

where k, the initial number of levels, depends only on the initial population and the span of control. The conclusion is that the creation of new levels goes on at the steady rate of $r/(\log s)$ per unit of time or, equivalently, that each $(\log s)/r$ units of time a new level is created.

A typical entrant who makes a career in a large organization virtually ensures that he will not rise far above the bottom. With a span s of 5, even spectacular success, rising above 0.999 of fellow employees, will put him only at 4.4 levels from the bottom, or in the fifth layer. This is of course modified for a two-class organization, like the traditional British army, in which one entered either as an enlisted man, with zero chance of rising above sergeant, or as an officer, with appreciable chance of rising to the highest ranks.

The boundaries of an organization are not always well defined; in fact, they may be determined by the viewpoint of the individual. An orderly in a hospital is near the bottom of the paid hierarchy; but if he thinks of patients as part of the hospital, he adds a substantial layer below him. A teaching assistant in a college is at the bottom or not, depending on whether students are included in the hierarchy. A faculty member may be high in her department, but low in the collection of departments in the country that make up her professional peer group.

Suppose that the rules of an organization call for absolute seniority, so that everyone has a turn at the top, if only for a short period. In a stationary condition the fraction of his service that he could have at the top is

$$\frac{1}{N} = \frac{s-1}{s^n - 1}.$$

If $s = 5$, $n = 9$, he could be boss for $4/(5^9 - 1) = 1/488{,}000$ of his career, which works out to 11 office minutes in the 45 years from age 20 to age 65.

But if the situation is hopeless for the mass of people in an organization, by the same logic the person who has an exceptional endowment of luck, brains, or influence can very quickly rise to the top. If the organization has 488,000 employees, and everyone in it above the bottom level has 5 people reporting to him, there are only 9 layers. A person starting at the bottom need be promoted only once every 3 years to become head of General Motors before he is 50.

11

Markov Chains for Individual Life Histories

The passage of an individual through its life cycle, from birth to death, is marked by various notable events. Some, like maturation or mating or reproduction, may be optional. Others, like death, are inevitable. A matrix model contains a great deal of information (or, equivalently, makes many strong assertions) about these events, their probabilities, and the sequences in which they tend to occur. This chapter shows how to extract some of the implications of these assertions. For example, what is the probability of surviving from birth to maturity in a model where individuals may reach maturity by many pathways? What is the average time required to mature? What is the probability of experiencing an event (e.g., a disease, or attack by a predator) before maturing, or before dying? Such questions can be answered by describing the individual life cycle as a Markov chain. The results provide ways to calculate age-specific parameters from stage-classified models.

11.1 $\mathbf{A} = \mathbf{T} + \mathbf{F}$: Decomposing the Matrix

Most life cycles permit a distinction between the transitions of living individuals and the production of new individuals. Thus the projection matrix can be written

$$\mathbf{A} = \mathbf{T} + \mathbf{F}, \tag{11.1.1}$$

where $\mathbf{T}$ describes transitions and $\mathbf{F}$ describes reproduction. The element f_{ij} of $\mathbf{F}$ is the expected number of type i offspring produced by an individual in stage j. The element t_{ij} of $\mathbf{T}$ is the probability that an individual in stage j at time t is alive and in stage i at time $t+1$. This decomposition has been applied by Cochran and Ellner (1992) to density-independent models and Cushing (1988, 1997, Cushing and Yicang 1994) to density-dependent models.

11.1.1 The Life Cycle as a Markov Chain

We will describe the movement of an individual through its life cycle as a Markov chain. The matrix $\mathbf{T}$ describes part of this movement, but it must be augmented by adding an extra state, "dead." If there are s stages, the result is an $s+1$ dimensional Markov chain with transition matrix*

$$\mathbf{P} = \left(\begin{array}{c|c} \mathbf{T} & \mathbf{0} \\ \hline \mathbf{m} & 1 \end{array}\right) \text{ of dimension } \left(\begin{array}{c|c} s \times s & s \times 1 \\ \hline 1 \times s & 1 \times 1 \end{array}\right). \tag{11.1.2}$$

Here, $m_j = 1 - \sum_i t_{ij}$ is the probability of death for stage j. State $s+1$ (death) is an absorbing state; once an individual enters this state, it never leaves.

There is a sizeable literature on Markov chain models for movement of individuals among social and occupational classes (e.g., Bartholomew 1982), some of which has been applied to demography (see Section 10.5). Feichtinger (1971, 1973) applied these methods to age-classified models with multiple absorbing states, first marriage models, and models for multiple classifications (age and parity, age and marital status). Cochran and Ellner (1992) independently used Markov chains in an analysis of stage-classified models. They also considered cases where $\mathbf{F}$ is divided into separate matrices representing birth and fission.

The states in an absorbing Markov chain can be divided into two sets: a set $\mathcal{T}$ of transient states and a set $\mathcal{A}$ of absorbing states; in this model

$$\begin{aligned} \mathcal{T} &= \{1, 2, \ldots, s\} \\ \mathcal{A} &= \{s+1\}. \end{aligned}$$

We will denote the number of elements in the set $\mathcal{A}$ by $|\mathcal{A}|$ (this is the *cardinality* of $\mathcal{A}$). At this point, $|\mathcal{A}| = 1$, but later we will add more absorbing states. We assume (it is implicit in calling the set $\mathcal{T}$ transient) that there is a pathway from each of the states in $\mathcal{T}$ to one of the states in $\mathcal{A}$. That is, there are no immortal stages in the life cycle. If there were, they would appear as another set of absorbing states.

---

*To emphasize the parallel with population projection matrices, $\mathbf{P}$ has been written as a column-stochastic matrix. Many texts on Markov chains would write it as a row-stochastic matrix, equivalent to the transpose of the matrix listed here. Either way is OK; just be careful to get the proper orientation of formulae presented in such texts.

Let $\mathbf{x}$ be a column vector giving the probability distribution of states, where $0 \le x_i \le 1$ and $\sum_i x_i = 1$. Then

$$\mathbf{x}(t+1) = \mathbf{P}\mathbf{x}(t) \tag{11.1.3}$$

and

$$\mathbf{x}(t) = \mathbf{P}^t\mathbf{x}(0), \tag{11.1.4}$$

where

$$\mathbf{P}^t = \left(\begin{array}{c|c} \mathbf{T}^t & 0 \\ \hline \mathbf{m}\sum_{n=0}^{t-1}\mathbf{T}^n & 1 \end{array}\right). \tag{11.1.5}$$

Our assumption that it is possible to reach the state "dead" from every state in $\mathcal{T}$ guarantees that the dominant eigenvalue of $\mathbf{T}$ is strictly less than 1; thus $\lim_{t\to\infty}\mathbf{T}^t = \mathbf{0}$ (Iosifescu 1980, Theorem 2.2). Thus (11.1.3) leads to the basic result on absorbing Markov chains: No matter what the initial probability distribution, the probability of the system being in any state other than an absorbing state eventually approaches zero.

Example 11.1 Markov-chain decomposition for killer whales

The killer whale life cycle graph is given in Figure 3.10. Using vital rates estimated for the resident population in coastal waters of Washington and British Columbia (Brault and Caswell 1993), the decomposition of $\mathbf{A}$ yields

$$\mathbf{T} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.9775 & 0.9111 & 0 & 0 \\ 0 & 0.0736 & 0.9534 & 0 \\ 0 & 0 & 0.0452 & 0.9804 \end{pmatrix} \tag{11.1.6}$$

$$\mathbf{F} = \begin{pmatrix} 0 & 0.0043 & 0.1132 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{11.1.7}$$

Because only one kind of offspring is produced, the fertility matrix $\mathbf{F}$ contains only a single nonzero row. The transition matrix is

$$\mathbf{P} = \left(\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 0 \\ 0.9775 & 0.9111 & 0 & 0 & 0 \\ 0 & 0.0736 & 0.9534 & 0 & 0 \\ 0 & 0 & 0.0452 & 0.9804 & 0 \\ \hline 0.0225 & 0.0153 & 0.0014 & 0.0196 & 1.0000 \end{array}\right). \tag{11.1.8}$$

Example 11.2 From school to work.

Some of the considerations just mentioned can be illustrated by a model of the school population and its emergence into the labor force

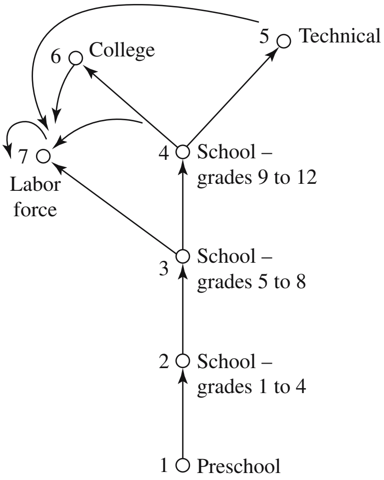


Figure 11.1. Graph of progress through school system and points of emergence into the labor force, along with corresponding matrix **A**, disregarding mortality and repeating; projection interval = 4 years.

(Keyfitz 1977). Consider only the flows portrayed in the graph of Figure 11.1. We ignore mortality, but in this model stage $\mathcal{N}_7$, the labor force, is an absorbing state. The transient states are $\mathcal{N}_1$–$\mathcal{N}_6$. In this case, there is no reproductive process, and the transition matrix is

$$\mathbf{P} = \left(\begin{array}{cccccc|c} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{43} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{54} & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{64} & 0 & 0 & 0 \\ \hline 0 & 0 & 1-p_{43} & 1-p_{54}-p_{64} & 1 & 1 & 1 \end{array}\right) \qquad (11.1.9)$$

11.1.2 The Analysis of Absorbing Chains

So far we have only stated the obvious: if every individual *can* die, every individual eventually *will* die. Now we want to know what happens to individuals en route to that end. The results summarized here are standard in Markov chain theory; see Iosifescu (1980, Chap. 3) or Kemeny and Snell (1976, Chap. 3) for full treatments.

Notation alert. If **X** is any matrix, then $\mathbf{X}_{\text{dg}}$ is the matrix with the diagonal of **X** on its diagonal, and zeros elsewhere. The row sums of **X** are **Xe**, and the column sums are $\mathbf{e}^{\mathsf{T}}\mathbf{X}$, where **e** is a column vector of ones.

11.1.2.1 The Fundamental Matrix

The (i, j) entry of the matrix $\mathbf{T}^t$ is the probability that an individual in stage j at time 0 will be in stage i at time t, for $i, j \in \mathcal{T}$. Since absorption is certain, these probabilities eventually decay to zero. But an individual will visit various transient states, various numbers of times, before absorption happens. Let ν_{ij} be the number of visits to transient state i before absorption, given that the individual starts in state j. The expected values of the ν_{ij} are given by a matrix

$$\begin{aligned} \begin{pmatrix} E(\nu_{ij}) \end{pmatrix} &= \mathbf{I} + \mathbf{T} + \mathbf{T}^2 + \cdots && (11.1.10) \\ &= (\mathbf{I} - \mathbf{T})^{-1} && (11.1.11) \\ &\equiv \mathbf{N}. && (11.1.12) \end{aligned}$$

The matrix $\mathbf{N}$ is called the *fundamental matrix* of the Markov chain.

The second moments of the ν_{ij} are

$$\begin{pmatrix} E\left(\nu_{ij}^2\right) \end{pmatrix} = (2\mathbf{N}_{\mathrm{dg}} - \mathbf{I})\,\mathbf{N} \qquad (11.1.13)$$

(Iosifescu 1980). Thus the variance of the number of visits to each state is given by the matrix

$$\begin{pmatrix} V(\nu_{ij}) \end{pmatrix} = (2\mathbf{N}_{\mathrm{dg}} - \mathbf{I})\,\mathbf{N} - \mathbf{N} \circ \mathbf{N}, \qquad (11.1.14)$$

where $\circ$ denotes the Hadamard, or element-by-element, product.

The fundamental matrix provides information about the time to absorption (i.e., to death). Let the time to absorption, starting in transient state i, be η_i. The mean of η_j is the sum of column j of the fundamental matrix $\mathbf{N}$ (i.e., the number of visits to all the transient states)

$$\begin{pmatrix} E(\eta_1) & \cdots & E(\eta_s) \end{pmatrix} = \mathbf{e}^{\mathsf{T}}\mathbf{N}. \qquad (11.1.15)$$

Iosifescu shows that the second moments of the η_j are given by

$$\begin{pmatrix} E\left(\eta_1^2\right) & \cdots & E\left(\eta_s^2\right) \end{pmatrix} = \mathbf{e}^{\mathsf{T}}\mathbf{N}\,(2\mathbf{N} - \mathbf{I})\,. \qquad (11.1.16)$$

Thus the variance of time to absorption can be written

$$\begin{pmatrix} V(\eta_1) & \cdots & V(\eta_s) \end{pmatrix} = \mathbf{e}^{\mathsf{T}}\left(2\mathbf{N}^2 - \mathbf{N}\right) - \mathbf{e}^{\mathsf{T}}\mathbf{N} \circ \mathbf{e}^{\mathsf{T}}\mathbf{N}. \qquad (11.1.17)$$

We can calculate not only the mean but the complete probability distribution of the η_j. The entries of $\mathbf{T}^t$ give the probabilities of being in each of the transient states at time t. Thus the sum of column j of $\mathbf{T}^t$ is the probability that absorption *has not* occurred by time t, starting from state j. Thus

$$\begin{pmatrix} P[\eta_1 > t] & \cdots & P[\eta_s > t] \end{pmatrix} = \mathbf{e}^{\mathsf{T}}\mathbf{T}^t. \qquad (11.1.18)$$

Since $P[\eta_i = t] = P[\eta_i > t-1] - P[\eta_i > t]$, we obtain the probability distribution for the times to absorption

$$\begin{pmatrix} P[\eta_1 = t] & \dots & P[\eta_s = t] \end{pmatrix} = \mathbf{e}^{\mathsf{T}}(\mathbf{T}^{t-1} - \mathbf{T}^t). \tag{11.1.19}$$

11.1.2.2 Probabilities of Absorption

So far we have assumed only a single absorbing state (death). It is sometimes useful to consider models with multiple absorbing states, in which case

$$\mathbf{P} = \left(\begin{array}{c|c} \mathbf{T} & \mathbf{0} \\ \hline \mathbf{M} & \mathbf{I} \end{array}\right), \tag{11.1.20}$$

where $\mathbf{M}$ is a matrix whose entries m_{kj} give the probability of moving from state $j \in \mathcal{T}$ to the kth absorbing state in $\mathcal{A}$. Every individual will end up in an absorbing state, but *which* absorbing state is uncertain. We would like to calculate the probabilities

$$b_{kj} = P[\text{absorption in } k | \text{starting in } j] \quad 1 \le k \le |\mathcal{A}|,\ j \in \mathcal{T}. \tag{11.1.21}$$

We know that if the system starts in transient state j it spends an average of n_{ij} time steps in transient state i (where $i \in \mathcal{T}$). At each of these time steps, the probability of moving to the kth absorbing state is m_{ki}. Thus

$$b_{kj} = \sum_{i \in \mathcal{T}} m_{ki} n_{ij} \qquad 1 \le k \le |\mathcal{A}| \tag{11.1.22}$$

or, in matrix notation

$$\mathbf{B} = \mathbf{MN}. \tag{11.1.23}$$

11.1.2.3 Probabilities Conditional on Absorption

If more than one absorbing state exists, we can calculate probabilities conditional on which of those states an individual ends up in. For example, we might describe mortality and emigration as absorbing states. We could compare the properties (e.g., mean time before leaving the population) of individuals that emigrate versus those that die before emigrating. We do this by creating a new transition matrix, conditional on absorption in a particular state.

Suppose we condition on absorption in the kth absorbing state. The state space of the resulting Markov chain is the set $\mathcal{T}$ of transient states plus the kth absorbing state. The transition matrix for this new conditional Markov chain is

$$\mathbf{P}^{(c)} = \left(\begin{array}{c|c} \mathbf{T}^{(c)} & 0 \\ \hline \mathbf{m}^{(c)} & 1 \end{array}\right), \tag{11.1.24}$$

where the superscript c denotes the conditional chain. Bayes' theorem[†] implies that the conditional probabilities can be written in terms of the matrix $\mathbf{M}$ and the matrix $\mathbf{B}$ calculated from $\mathbf{P}$:

$$t_{ij}^{(c)} = \frac{t_{ij} b_{ki}}{b_{kj}} \tag{11.1.25}$$

$$m_j^{(c)} = m_{kj}/b_{kj} \tag{11.1.26}$$

or, in matrix notation,

$$\mathbf{T}^{(c)} = \mathbf{D}\mathbf{T}\mathbf{D}^{-1}, \tag{11.1.27}$$

where, letting $\mathbf{b}_{k\cdot}$ denote the kth row of $\mathbf{B}$,

$$\mathbf{D} = \operatorname{diag}(\mathbf{b}_{k\cdot}) = \begin{pmatrix} b_{k1} & & \\ & b_{k2} & \\ & & \ddots \end{pmatrix}. \tag{11.1.28}$$

The fundamental matrix for the conditional chain is

$$\mathbf{N}^{(c)} = \mathbf{D}\mathbf{N}\mathbf{D}^{-1}. \tag{11.1.29}$$

It can be used like any fundamental matrix.

We turn now to several important applications of these results.

11.2 Lifetime Event Probabilities

At any stage in its life cycle, an individual is embarking on a developmental process that will eventually end in death. At each step of that process, the individual experiences the risk of various events (e.g., contracting a disease, or attack by a predator). The probability that an event occurs at least once in the individual's lifetime may be of interest in a variety of contexts. For example:

- The plant *Lomatium grayi* is attacked by a moth called *Depressaria multifidae. Lomatium* grows through a series of size classes, alternating between vegetative and flowering condition. The probability of attack by *Depressaria* is stage-dependent. Thompson and Moody (1985) wanted to know the probability that a new seedling will be attacked by the moth at least once during its lifetime. They interpret this probability as a measure of the "apparency" of the plant to its herbivore. It has been suggested that apparency should determine the type of chemical defenses that a plant deploys against a herbivore (Feeney 1976).

[†] For any two events X and Y, $P[X|Y] = P[X \text{ and } Y]/P[Y]$, provided that $P[Y] > 0$.

- The U.S. Department of Justice has calculated the probability that a 12-year-old American will experience certain kinds of violent crime in his or her lifetime (Koppel 1987). The probabilities were 0.83 for all types of violent crime, 0.99 for personal theft, 0.74 for assault, 0.30 for robbery, and 0.08 for rape (the latter figure for women). These lifetime probabilities are presented as better measures of the true risk of these crimes than annual incidence statistics ("If the earth revolved around the sun in 180 days, all our annual crime rates would be halved, but we would not be safer." Koppel 1987)
- Bone fractures are a significant health risk to postmenopausal women. Cummings et al. (1989) found that the lifetime probability of hip fracture for a 50-year-old U.S. woman was 0.156. The corresponding probabilities for wrist fractures and atraumatic vertebral fractures were 0.15 and 0.32. Combining these probabilities with estimates of the mortality due to these injuries, they concluded that the lifetime risk of death from hip fracture was comparable to that from breast cancer. Such comparisons might be useful in comparing the risks and benefits of therapies designed to reduce osteoporosis.

In each of these cases, the lifetime probability is affected by both the risk of the event and the demography. The lifetime probability of hip fracture could be lessened by reducing the risk of falling (e.g., by studying t'ai chi chuan; Wolf et al. 1996) or by increasing mortality, so that fewer women survive to ages where falls are more frequent. The former strategy is obviously preferable, but in more complex life cycles the choices might be less plain. *Lomatium grayi* might, for example, want to consider whether it should try to prevent *Depressaria* attack, or adjust its life cycle to spend less time in stages particularly vulnerable to attack.

Chiang (1968) shows how to calculate lifetime risks from age-classified life tables. Here we compute these probabilities from stage-classified models using the Markov chain description of the life cycle. We begin with the basic structure (11.1.2) and add an additional absorbing state, "event-before-death." Because we are interested only in the probability of experiencing the event, what happens to an individual *after* the event is irrelevant (of course, it may be very relevant for other questions). Thus we can simply leave individuals who have experienced the event in "event-before-death" and not worry about them further.

Let α_i be the probability that an individual in stage i experiences the event in the interval $(t, t+1]$. Carry out the following procedure.

1. Create a new matrix $\mathbf{T}'$ describing transitions of individuals that neither die nor experience the event. Its entries are

$$t'_{ij} = (1-\alpha_j)\, t_{ij} \qquad i,j \in \mathcal{T}. \tag{11.2.1}$$

2. Create a new matrix $\mathbf{M}'$ containing the probabilities of transition from each transient state to the two absorbing states
$$m'_{ij} = \begin{cases} (1-\alpha_j)\,m_j & i=1,\ j \in \mathcal{T} \\ \alpha_j & i=2,\ j \in \mathcal{T} \end{cases} \tag{11.2.2}$$
The new transition matrix is
$$\mathbf{P}' = \left(\begin{array}{c|c} \mathbf{T}' & \mathbf{0} \\ \hline \mathbf{M}' & \mathbf{I} \end{array} \right). \tag{11.2.3}$$
3. Use the methods of Section 11.1.2.2 to compute the probability of absorption in each of the two states.
$$\mathbf{B} = \mathbf{M}'\left(\mathbf{I} - \mathbf{T}'\right)^{-1}. \tag{11.2.4}$$
The lifetime event probability is the probability of absorption in "event-before-death," which is given by the second row of $\mathbf{B}$.

11.3 Age-Specific Traits From Stage-Specific Models

In a stage-classified model, the ages of individuals are not known, but it is possible to calculate age-specific survivorship and fertility, mean age at first reproduction, net reproductive rate, generation time, and age-within-stage distributions directly from the Markov chain describing the life cycle. We will focus on two examples. One is the stage-structured model for killer whales (Brault and Caswell 1993; see Example 11.1), because an independent age-classified analysis exists for comparison (Olesiuk et al. 1990).

The second example is a stage-structured model for teasel (*Dipsacus sylvestris*). Teasel is a weedy plant, introduced to North America from Europe in the late nineteenth century. Seeds germinate to form flattened rosettes, which grow until they reach a critical size, at which point they flower and die. Seeds may remain dormant for up to a few years. A variety of evidence pointed to size as the critical *i*-state variable, so Werner and Caswell (1977, Caswell and Werner 1978) developed size-classified models whose entries corresponded to the life cycle graph of Figure 11.2. Because the projection interval was one year, reproduction in year t can produce small, medium, or large rosettes in year $t+1$. The production of more than one kind of offspring is not common in human demographic models, but it might appear in analyses of infant mortality or of family development; it is not uncommon in ecological applications, and this example will show how to account for it in calculations.

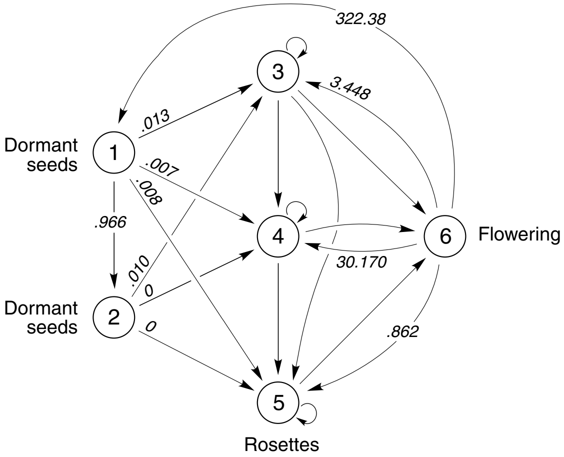


Figure 11.2. A life cycle graph for teasel (*Dipsacus sylvestris*). Stages: $\mathcal{N}_1$ = first-year dormant seeds, $\mathcal{N}_2$ = second-year dormant seeds, $\mathcal{N}_3$ = small rosettes, $\mathcal{N}_4$ = medium rosettes, $\mathcal{N}_5$ = rosettes, and $\mathcal{N}_6$ = flowering plants.

Example 11.3 Markov chain decomposition for teasel

The population projection matrix for teasel is composed of transition and fertility matrices

$$\mathbf{T} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.966 & 0 & 0 & 0 & 0 & 0 \\ 0.013 & 0.010 & 0.125 & 0 & 0 & 0 \\ 0.007 & 0 & 0.125 & 0.238 & 0 & 0 \\ 0.008 & 0 & 0.038 & 0.245 & 0.167 & 0 \\ 0 & 0 & 0 & 0.023 & 0.750 & 0 \end{pmatrix} \tag{11.3.1}$$

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 322.388 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.448 \\ 0 & 0 & 0 & 0 & 0 & 30.170 \\ 0 & 0 & 0 & 0 & 0 & 0.862 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{11.3.2}$$

The sixth column of **T** is zero because flowering plants all die after reproduction. The nonzero entries in the sixth column of **F** give the production of different kinds of offspring (first-year dormant seeds and small, medium, and large rosettes). The resulting Markov chain transition matrix is

$$\mathbf{P} = \left(\begin{array}{cccccc|c} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.966 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.013 & 0.010 & 0.125 & 0 & 0 & 0 & 0 \\ 0.007 & 0 & 0.125 & 0.238 & 0 & 0 & 0 \\ 0.008 & 0 & 0.038 & 0.245 & 0.167 & 0 & 0 \\ 0 & 0 & 0 & 0.023 & 0.750 & 0 & 0 \\ \hline 0.006 & 0.990 & 0.712 & 0.494 & 0.083 & 1.000 & 1.000 \end{array}\right).$$

11.3.1 Age-Specific Survival

Basic information on age-specific survival—the mean and variance of the time spent in each stage, and the mean and variance of the time to death—is obtained from the fundamental matrix **N**. For killer whales, the fundamental matrix is

$$\mathbf{N} = \begin{pmatrix} 1.00 & 0 & 0 & 0 \\ 10.99 & 11.25 & 0 & 0 \\ 17.37 & 17.77 & 21.46 & 0 \\ 40.05 & 40.97 & 49.49 & 51.02 \end{pmatrix}. \qquad (11.3.3)$$

Thus a yearling killer whale spends, on the average, about 11 years as an immature, 17 years as a reproductive adult, and 40 years as a postreproductive female. A mature adult, in contrast, spends an average of 21 years in that stage, and almost 50 years as a postreproductive. These averages are larger than those for a yearling, because some yearlings spend zero years as an adult or postreproductive, having died first.

The variance in the amount of time spent in each stage, computed from (11.1.14), is

$$\left(\ V(\nu_{ij})\ \right) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 115.5 & 115.3 & 0 & 0 \\ 426.4 & 429.1 & 439.0 & 0 \\ 2442.7 & 2461.1 & 2551.2 & 2552.1 \end{pmatrix}. \qquad (11.3.4)$$

One way to interpret these variances is to compute the coefficient of variation (the ratio of the standard deviation to the mean) of the time spent in each stage, which is

$$\left(\ CV(\nu_{ij})\ \right) = \begin{pmatrix} 0 & - & - & - \\ 0.98 & 0.95 & - & - \\ 1.19 & 1.17 & 0.98 & - \\ 1.23 & 1.21 & 1.02 & 0.99 \end{pmatrix}, \qquad (11.3.5)$$

where "–" denotes coefficients of variation that are undefined because both the mean and the standard deviation are zero. There is considerable variation in the duration of each stage except the first, with coefficients of variation on the order of 1.

The mean and variance of the time η_j to death, given that the individual starts in stage j are given by (11.1.15) and (11.1.17) with results

$$\left(\begin{array}{c} E(\eta_j) \end{array} \right) = \left(\begin{array}{cccc} 69.411 & 69.985 & 70.947 & 51.02 \end{array} \right) \tag{11.3.6}$$

$$\left(\begin{array}{c} V(\eta_j) \end{array} \right) = \left(\begin{array}{cccc} 3341.4 & 3308.1 & 2990.3 & 2552.1 \end{array} \right). \tag{11.3.7}$$

The mean age at death is the *life expectancy* or expection of life; the life expectancy of a newborn individual is 69.4 years, with a standard deviation of 57.8 years. (You might want to add one year to this number to get the average age at death, calculated from birth instead of from the yearling stage.)

The mean age at death provides only part of the information about survival. We can also calculate the age-specific survivorship (i.e., the probability of survival to age x) for individuals in each stage from the transition matrix $\mathbf{T}$, using (11.1.18).

The survivorship conditional on starting in stage 1 (newborn) is shown in Figure 11.3; this corresponds more or less[‡] to the classical survivorship function $l(x)$ starting at birth. The age-specific survivorship function reported by Olesiuk et al. (1990) is shown for comparison. Survivorship from the stage-classified model does not capture all the changes in age-specific mortality. No surprise there; it is defined by only four stage-specific survival probabilities. It does, however, capture the average survivorship from birth up to 60 or 70 years of age. Survivorship of the oldest individuals is much higher in the stage-classified model than in the age-classified model, because the former ignores senescence of post-reproductive females.

11.3.2 Age-Specific Fertility

Age-specific fertility is described by a matrix $\mathbf{\Phi}(x)$, whose entries $\phi_{ij}(x)$ are the mean number of type i offspring produced at age x by an individual starting in stage j at age 0. If stage j corresponds in any sense to a "newborn" individual, then $\phi_{ij}(x)$ is the age-specific fertility in the usual sense of the word.

The fertility matrix is calculated from $\mathbf{T}$ and $\mathbf{F}$. The (i,j) entry of $\mathbf{T}^x$ is the probability of being in stage i at time x, conditional on starting in stage j. Dividing each column of $\mathbf{T}^x$ by its sum gives the distribution of individuals among stages at time x conditional on survival to time x:

$$\mathbf{T}^x \left(\text{diag}\left(\mathbf{e}^{\mathsf{T}} \mathbf{T}^x \right) \right)^{-1}. \tag{11.3.8}$$

[‡] Individuals in stage 1 (yearlings) do not correspond exactly to "newborn" individuals in a life table, but more closely to an individual somewhere between 1/2 and 1 year old. It is probably not worth trying to make the correspondence too exact.

Left-multiplying this matrix by the fertility matrix $\mathbf{F}$ gives the matrix of fertilities at time x

$$\mathbf{\Phi}(x) = \mathbf{F}\mathbf{T}^x \left(\text{diag}\left(\mathbf{e}^{\mathsf{T}}\mathbf{T}^x\right)\right)^{-1}. \tag{11.3.9}$$

In the case of the killer whale, $\mathbf{F}$ contains only a single nonzero row, and hence so does $\mathbf{\Phi}(x)$. Thus at $x = 1$ the fertility is given by

$$\mathbf{\Phi}(1) = \mathbf{F}\mathbf{T}^1 \left(\text{diag}\left(\mathbf{e}^{\mathsf{T}}\mathbf{T}^1\right)\right)^{-1} \tag{11.3.10}$$

$$= \begin{pmatrix} 0.0043 & 0.0124 & 0.1081 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{11.3.11}$$

At $x = 20$, fertility has increased to

$$\mathbf{\Phi}(20) = \begin{pmatrix} 0.0573 & 0.0569 & 0.0503 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{11.3.12}$$

The entry $\phi_{11}(x)$ corresponds to the classical notion of age-specific fertility. The other entries of the first row give fertility at "age" x of an individual starting life as an immature, mature, or postreproductive female, respectively. Figure 11.4 shows $\phi_{11}(x)$ as a function of age. An age-specific fertility estimate is also shown.[§] The stage-classified model captures the overall shape of the age-specific function, but produces some fertility earlier, and a lot of fertility later, than does the age-classified analysis. This result is to be expected, since the stage-classification spreads out the passage of individuals through the reproductive stage.

When there are multiple types of offspring, there is no exact analogue of age-specific fertility. Consider the case of teasel (Example 11.3) where there are four types of offspring. At $x = 0$, the fertility matrix is

$$\mathbf{\Phi}(0) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 322.38 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.45 \\ 0 & 0 & 0 & 0 & 0 & 30.17 \\ 0 & 0 & 0 & 0 & 0 & 0.86 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{11.3.13}$$

[§] Obtained by multiplying the $m(x)$ figures in Table 14 of Olesiuk et al. (1990) by their estimate (0.57) of survival to age 1/2.

This is just **F**, because only an individual in stage $\mathcal{N}_6$ (flowering plants) can produce any offspring at $x = 0$. At $x = 3$, we obtain

$$\boldsymbol{\Phi}(3) = \begin{pmatrix} 107.95 & 75.43 & 158.62 & 204.02 & 263.68 & \text{NaN} \\ 0 & 0 & 0 & 0 & 0 & \text{NaN} \\ 1.15 & 0.81 & 1.70 & 2.18 & 2.82 & \text{NaN} \\ 10.10 & 7.06 & 14.84 & 19.09 & 24.68 & \text{NaN} \\ 0.29 & 0.20 & 0.42 & 0.55 & 0.71 & \text{NaN} \\ 0 & 0 & 0 & 0 & 0 & \text{NaN} \end{pmatrix}, \qquad (11.3.14)$$

where NaN ("not a number") is the result of dividing 0 by 0 in MATLAB.

Equation (11.3.9) will not work in this case, because the sixth column of $\mathbf{T}^x$ is always zero. Thus diag $(\mathbf{e}^\mathsf{T}\mathbf{T}^x)$ always has a zero in the $(6, 6)$ position. Although it is singular, MATLAB will happily compute its inverse, but it returns a matrix of NaNs, because the 0/0 division propagates throughout the inversion process. A workable alternative is to compute the inverse matrix as

```
diag(1./sum(T^x))
```

This produces a matrix $\boldsymbol{\Phi}(x)$ with NaN entries only in the last column.

An individual that starts life as a dormant seed ($\mathcal{N}_1$) will, at age 3, produce an average of 108 dormant seeds, 1.15 small rosettes, 10.1 medium rosettes, and 0.29 large rosettes, provided that it survives to age 3. These multiple types of offspring can be analyzed independently, or can be summed to give total numbers of offspring. Cochran and Ellner (1992) suggested weighting the values by the reproductive value **v** calculated from the stage-classified matrix **A** and then summing them.

The fact that an individual can be "born" in any of four different stages also complicates matters. At age $x = 3$ an individual born as a dormant seed has less than half the reproductive output of an individual of the same age that was born as a large rosette. Figure 11.5 shows the summed fertility, as a function of age, for individuals born as dormant seeds and as small, medium, and large rosettes. Fertility climbs and remains high indefinitely because this is the fertility of surviving individuals. Very few individuals survive more than a few years.

11.3.3 Age at First Reproduction

In an age-classified model, the age at first reproduction is simply the first age x for which the maternity function $m(x)$ is nonzero. It is not a good idea to apply this approach to $\Phi(x)$ calculated from a stage-classified model. We have just seen that the killer whale model predicts that some individuals will reproduce at $x = 1$. A more reasonable measure would be the *mean* time from birth (however that is defined) to the first entry into a reproductive stage (Cochran and Ellner 1992). There may be several such stages, and individuals may reach them at different ages and by different pathways. We

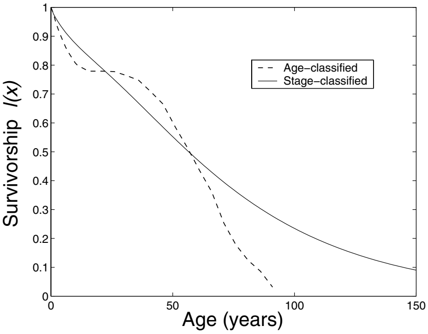


Figure 11.3. A comparison of the age-specific survivorship function for killer whales derived from the stage-classified model of Brault and Caswell (1993) and from the age-classified analysis of Olesiuk et al. (1990).

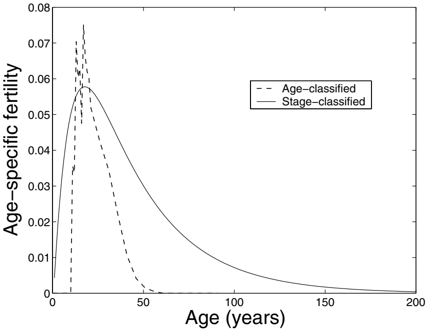


Figure 11.4. Age-specific fertility for killer whales calculated from the stage-classified matrix model and from the age-specific maternity function $m(x)$ of Olesiuk et al. (1990).

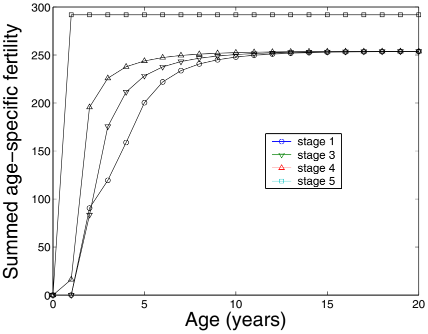


Figure 11.5. Age-specific fertility, summed over all offspring types, for teasel plants "born" in stages $\mathcal{N}_1$ (dormant seeds), $\mathcal{N}_3$ (small rosettes), $\mathcal{N}_4$ (medium rosettes), and $\mathcal{N}_5$ (large rosettes).

want the mean, over all these pathways, calculated from individuals that do not die first, i.e., the mean conditional on reaching the reproductive stage before death. This conditional mean is calculated by making the set of reproductive stages absorbing, creating a new chain conditional on absorption there, and calculating the mean time to absorption from that chain.

1. Decide on a set of "reproductive" states; call this set $\mathcal{R}$.
2. Create a new absorbing state, "reproduced-before-dying."
3. Create a new transition matrix $\mathbf{P}'$ in which an individual that enters any state in $\mathcal{R}$ spends one time step there and then is absorbed in "reproduced-before-dying":

   $$\mathbf{P}' = \left(\begin{array}{c|c} \mathbf{T}' & \mathbf{0} \\ \hline \mathbf{M}' & \mathbf{I} \end{array} \right), \tag{11.3.15}$$

   where

   $$t'_{ij} = \left\{ \begin{array}{ll} t_{ij} & j \notin \mathcal{R} \\ 0 & j \in \mathcal{R} \end{array} \right. \tag{11.3.16}$$

   $$m'_{1j} = \left\{ \begin{array}{ll} m_j & j \notin \mathcal{R} \\ 0 & j \in \mathcal{R} \end{array} \right. \tag{11.3.17}$$

   $$m'_{2j} = \left\{ \begin{array}{ll} 0 & j \notin \mathcal{R} \\ 1 & j \in \mathcal{R} \end{array} \right. . \tag{11.3.18}$$

   We are interested only in the time required to reach $\mathcal{R}$, not in what happens afterward. Thus reproductive individuals can be left in the "reproduced-before-dying" state after they reach $\mathcal{R}$.
4. Calculate the probability of absorption in "reproduced-before-dying" using (11.1.23),

   $$\mathbf{B}' = \mathbf{M}' \left(\mathbf{I} - \mathbf{T}' \right)^{-1}. \tag{11.3.19}$$

   The second row, $\mathbf{b}_{2\cdot}$, of $\mathbf{B}'$ gives the probabilities of reproducing before death.
5. Create a new Markov chain conditional on absorption in "reproduced-before-dying," using (11.1.25) and (11.1.26):

   $$\mathbf{P}^{(c)} = \left(\begin{array}{c|c} \mathbf{T}^{(c)} & 0 \\ \hline \mathbf{m}^{(c)} & 1 \end{array} \right), \tag{11.3.20}$$

   where

   $$\mathbf{T}^{(c)} = \mathrm{diag}\left(\mathbf{b}'_{2\cdot} \right) \mathbf{T}' \, \mathrm{diag}\left(\mathbf{b}'_{2\cdot} \right)^{-1}. \tag{11.3.21}$$

6. Generate the mean times to absorption for this conditional chain using (11.1.15):

$$\left(\; E\left(\eta_i^{(c)}\right) \; \right) = \mathbf{e}^{\mathsf{T}} \left(\mathbf{I} - \mathbf{T}^{(c)}\right)^{-1}. \tag{11.3.22}$$

For example, in the case of the killer whale model, $\mathcal{R} = \{3\}$ and the transition matrix is [cf. (11.1.8]

$$\mathbf{P}' = \left(\begin{array}{rrrr|rr} 0 & 0 & 0 & 0 & 0 & 0 \\ 0.9775 & 0.9111 & 0 & 0 & 0 & 0 \\ 0 & 0.0736 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9804 & 0 & 0 \\ \hline 0.0225 & 0.0153 & 0 & 0.0196 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right) \tag{11.3.23}$$

and

$$\mathbf{B}' = \left(\begin{array}{rrrr} 0.1907 & 0.1721 & 0 & 1.0000 \\ 0.8093 & 0.8279 & 1.0000 & 0 \end{array} \right) \tag{11.3.24}$$

Thus, the probability of reaching maturity before death for an individual starting in stage 1 is 0.8093. Notice that the probability of reproducing before death is zero for an individual starting as a postreproductive female (i.e., $b'_{24} = 0$). This means that the conditional transition probabilities for stage 4 are undefined (you cannot condition on an event of probability zero) and that calculating $\left(\operatorname{diag}\left(\mathbf{b}'_{2\cdot}\right)\right)^{-1}$ involves 1/0.

This is not a major problem; it can be solved by inserting an arbitrary nonzero value (we will use 1.0) for b'_{24} and then ignoring the conditional probabilities for stage 4, since we know that individuals in stage 4 can never reach stage 3. The resulting conditional transition matrix is

$$\mathbf{T}^{(c)} = \left(\begin{array}{rrrc} 0 & 0 & 0 & - \\ 1.0 & 0.9111 & 0 & - \\ 0 & 0.0889 & 0 & - \\ 0 & 0 & 0 & - \end{array} \right). \tag{11.3.25}$$

The expected time to absorption in "reproduced-before-dying," conditional on eventual absorption in that state, is then

$$\left(\; E(\eta_i^{(c)}) \; \right) = \mathbf{e}^{\mathsf{T}} \left(\mathbf{I} - \mathbf{T}^{(c)}\right)^{-1} \tag{11.3.26}$$

$$= \left(\; 13.25 \quad 12.25 \quad 1.00 \quad - \; \right). \tag{11.3.27}$$

The mean age at maturity is $E\left(\eta_1^{(c)}\right) = 13.25$ years.[¶] This agrees well with Olesiuk et al.'s (1990) age-classified estimate of 13.1 years for the mean age at first birth.

11.3.4 Net Reproductive Rate

The net reproductive rate R_0 is the mean number of offspring by which a newborn individual will be replaced by the end of its life, and thus the rate by which the population increases from one generation to the next. Since the probability of surviving from birth to age x is $l(x)$ and the rate of reproduction at that age is $m(x)$, this expectation is given by

$$R_0 = \int_0^\infty l(x) m(x)\, dx \tag{11.3.28}$$

(Section 6.1).

The discrete equivalent, calculated from an age-classified matrix, is

$$\begin{aligned} R_0 &= F_1 + P_1 F_2 + P_1 P_2 F_3 + \cdots & (11.3.29) \\ &= \sum_i F_i \prod_{j=1}^{i-1} P_j. & (11.3.30) \end{aligned}$$

In age-classified models,

$$\begin{aligned} R_0 &< 1 \Leftrightarrow \lambda_1 < 1 \\ R_0 &= 1 \Leftrightarrow \lambda_1 = 1 \\ R_0 &> 1 \Leftrightarrow \lambda_1 > 1. \end{aligned} \tag{11.3.31}$$

We would like a stage-classified net reproductive rate with all the properties of the age-classified rate; it should give the per-generation growth rate, should relate to expected lifetime reproductive output, and should determine whether λ_1 is less than, equal to, or greater than one. This problem has been attacked independently by Cushing (1988, 1997, Cushing and Yicang 1994) and Cochran and Ellner (1992).

The fundamental matrix $\mathbf{N}$ gives the expected number of time steps spent in each transient state, and the fertility matrix $\mathbf{F}$ gives the expected number of offspring of each type produced per time step. Thus the matrix

$$\mathbf{R} = \mathbf{F}\mathbf{N} \tag{11.3.32}$$

[¶] Some care is needed in translating between time to absorption and age at entering a stage. We should subtract 1 from $\eta_1^{(c)}$, because absorption in "reproduced-before-dying" happens one time step after entering $\mathcal{R}$. But then we should add 1, because absorption is measured from stage 1, not from birth.

has entries r_{ij} that give the expected lifetime production[‖] of type-i offspring of an individual starting life in stage j. In the case of the killer whale, the matrix $\mathbf{R}$ is

$$\mathbf{R} = \begin{pmatrix} 2.0131 & 2.0595 & 2.4292 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{11.3.33}$$

A yearling female killer whale can expect to produce 2.01 female offspring during its life. Thus r_{11} is an attractive candidate for R_0, but the situation is more complicated when there is more than one type of offspring. For teasel,

$$\mathbf{R} = \begin{pmatrix} 3.66 & 0.27 & 27.33 & 103.06 & 290.26 & 322.39 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.04 & 0.00 & 0.29 & 1.10 & 3.10 & 3.45 \\ 0.34 & 0.03 & 2.56 & 9.64 & 27.16 & 30.17 \\ 0.01 & 0.00 & 0.07 & 0.28 & 0.78 & 0.86 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{11.3.34}$$

A plant starting life as a dormant seed (stage 1) can expect to produce 3.66 dormant seeds, 0.04 small rosettes, 0.34 medium rosettes, and 0.01 large rosettes over its lifetime. A plant starting life as a large rosette (stage 5), however, can expect to produce 290 dormant seeds, 3.1 small rosettes, 27.16 medium rosettes, and 0.78 large rosettes over its lifetime.

To account for multiple types of offspring, consider the following calculation, inspired by Cushing and Yicang (1994). Let $\mathbf{y}(0)$ be a vector giving the composition of an initial generation at $t = 0$. The fates of these individuals at $t = 1, 2, \ldots$ are given by $\mathbf{Ty}(0)$, $\mathbf{T}^2\mathbf{y}(0)$, ... At each time, the surviving individuals produce offspring according to F; thus, offspring production by this generation is $\mathbf{Fy}(0)$, $\mathbf{FTy}(0)$, $\mathbf{FT}^2\mathbf{y}(0)$, ...

Summing this offspring production over the life of the generation gives the next generation, $\mathbf{y}(1)$:

$$\begin{aligned} \mathbf{y}(1) &= \mathbf{Fy}(0) + \mathbf{FTy}(0) + \mathbf{FT}^2\mathbf{y}(0) + \cdots && (11.3.35) \\ &= \mathbf{F}\left(\mathbf{I} + \mathbf{T} + \mathbf{T}^2 + \cdots\right)\mathbf{y}(0) && (11.3.36) \\ &= \mathbf{F}\left(\mathbf{I} - \mathbf{T}\right)^{-1}\mathbf{y}(0) && (11.3.37) \\ &= \mathbf{FNy}(0) && (11.3.38) \\ &= \mathbf{Ry}(0). && (11.3.39) \end{aligned}$$

Thus, $\mathbf{R}$ projects the population from one generation to the next. If $\mathbf{R}$ has a dominant eigenvalue, that eigenvalue will give the rate of growth of the

---

[‖] Cushing and Yicang (1994) define $\mathbf{R} = \mathbf{NF}$, which does not give expected offspring production. However, as the eigenvalues of $\mathbf{FN}$ and of $\mathbf{NF}$ are the same, none of Cushing and Yicang's uses of R_0 are affected.

population from one generation to the next.[**] Thus we conclude that

$$R_0 = \text{dominant eigenvalue of } \mathbf{R}. \tag{11.3.40}$$

When, as in the killer whale, there is only one type of offspring, **R** has only a single nonzero row, and the dominant eigenvalue is just r_{11}.

We have shown that R_0 calculated as the dominant eigenvalue of **R** is the per-generation growth rate. When there is only one offspring type, it is also the expected number of offspring produced by an individual during its lifetime. Cushing and Yicang (1994, Theorem 3) proved that R_0 calculated from (11.3.40) also corresponds to λ as in (11.3.31).

11.3.5 Generation Time

There are several measures of generation time in age-classified models (Coale 1972):

1. The time T required for the population to increase by a factor of R_0, which satisfies $\lambda_1^T = R_0$. Thus
$$T = \frac{\log R_0}{\log \lambda_1}; \tag{11.3.41}$$
2. The mean age μ_1 of the parents of the offspring produced by a cohort over its lifetime. This is given by the mean of the net fertility schedule $l(x)m(x)$
$$\mu_1 = \frac{\int_0^\infty x l(x)m(x)\,dx}{\int_0^\infty l(x)m(x)\,dx}, \tag{11.3.42}$$

[**] Actually, this requires not only a dominant eigenvalue, but also convergence to the corresponding eigenvector. In the case of the population projection matrix **A**, the strong ergodic theorem guarantees this convergence. The alert reader may have noticed that in Section 7.2 we assumed that **A** was similar to a diagonal matrix, and used that fact in our proof of the ergodic theorem. A sufficient, but not necessary, condition for this is that the eigenvalues be distinct, which they usually are for population projection matrices. The eigenvalues of **R**, however, are usually *not* distinct; both the killer whale and teasel examples have one positive eigenvalue and a repeated eigenvalue of zero. The following result gives both necessary and sufficient condition for diagonalizability (Horn and Johnson 1985, Corollary 3.3.8, p.145). Let the distinct eigenvalues of **R** be $\rho_1, \ldots, \rho_m$. The matrix **R** is diagonalizable if and only if

$$q(\mathbf{R}) = (\mathbf{R} - \rho_1 \mathbf{I})(\mathbf{R} - \rho_2 \mathbf{I}) \cdots (\mathbf{R} - \rho_m \mathbf{I}) = 0,$$

which is true for both the killer whale and teasel models. Even if **R** is not diagonalizable, the Jordan canonical form can be used to show that the population will converge to the dominant eigenvector.

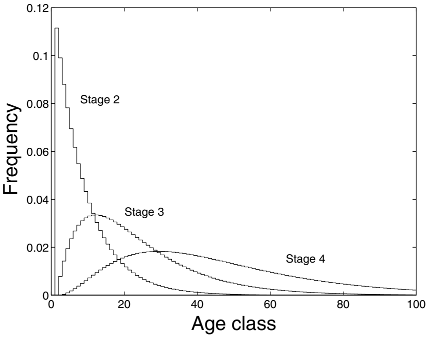


Figure 11.6. The stable age-within-stage distributions for killer whales, for stages $\mathcal{N}_2$–$\mathcal{N}_4$. Stage 1 (yearlings) is composed entirely of individuals in age class 1, and is not shown.

or, in discrete time

$$\mu_1 = \frac{\sum_i i \left(\prod_{j=1}^{i-1} P_j \right) F_i}{\sum_i \left(\prod_{j=1}^{i-1} P_j \right) F_i}, \tag{11.3.43}$$

where we define $\prod_{j=1}^{0} P_j = 1$.

3. The mean age $\bar{A}$ of the parents of the offspring produced by a population at the stable age distribution:

$$\bar{A} = \frac{\int_0^\infty x e^{-rx} l(x) m(x)\, dx}{\int_0^\infty e^{-rx} l(x) m(x)\, dx}, \tag{11.3.44}$$

or, in discrete time

$$\bar{A} = \frac{\sum_i i \lambda^{-i} \left(\prod_{j=1}^{i-1} P_j \right) F_i}{\sum_i \lambda^{-i} \left(\prod_{j=1}^{i-1} P_j \right) F_i}. \tag{11.3.45}$$

The denominator in both of these equations is the characteristic equation, and is thus equal to 1. Note that μ_1 can be obtained from (11.3.45) by setting $\lambda = 1$. Thus, in a stationary population, $\mu_1 = \bar{A}$.

In human populations $T \approx (\mu_1 + \bar{A})/2$, and all three measures are very similar (Coale 1972). In species with higher mortality rates and/or rates of increase farther from 1, the differences among these indices of generation time are likely to be greater.

The discrete versions of T, μ_1, and $\bar{A}$ can be calculated directly from a Leslie matrix for an age-classified population. The methods of this chapter let us calculate them from stage-classified models. The time required to increase by a factor R_0 can be calculated directly as $T = \log R_0 / \log \lambda_1$, where R_0 is calculated according to Section 11.3.4. For example, in the killer whale example, $R_0 = 2.0131$ and $\lambda_1 = 1.0254$, so $T = 27.85$ years, which compares well with the value of 24.8 years calculated by Olesiuk et al. (1990) from an age-classified model.

Both $\bar{A}$ and μ_1 can be calculated using the age-specific survival and fertility values obtained from the stage-classified matrix[††] (see Sections 11.3.1 and 11.3.2). In the case of the killer whales, this calculation yields $\bar{A} = 23.67$ and $\mu_1 = 32.18$. Since the population is increasing, the stable age distribution is skewed to younger ages than is a cohort; thus $\bar{A} < \mu_1$. Also note that $(\bar{A} + \mu_1)/2 = 27.93$, which is close to $T = 27.85$, as noted by Coale (1972) for age-classified models for humans.

11.3.6 Age-Within-Stage Distributions

The individuals within a stage may have arrived there by many different pathways, and thus be of many different ages. We can calculate the stable age distribution within each stage and the stable stage distribution within each age class.

Let λ be the dominant eigenvalue of $\mathbf{A} = \mathbf{T} + \mathbf{F}$ and let $\mathbf{w}$ be the corresponding right eigenvector. Suppose that the population has been growing at the rate λ for a long time, so that the age and stage distributions have stabilized. We observe the population at some time t. Because the stage distribution is stable, the births at t are proportional to $\mathbf{Fw}$. The births at time $t - a$ were proportional to $\lambda^{-a}\mathbf{Fw}$. Those individuals are now in age class $a + 1$, with a stage distribution proportional to $\lambda^{-a}\mathbf{T}^a\mathbf{Fw}$.

Create a rectangular array

$$\mathbf{X} = \begin{pmatrix} \mathbf{Fw} & \lambda^{-1}\mathbf{TFw} & \lambda^{-2}\mathbf{T}^2\mathbf{Fw} & \cdots \end{pmatrix}.$$

The columns of $\mathbf{X}$ correspond to ages, the rows to stages. Thus row i of $\mathbf{X}$, normalized to sum to 1, gives the stable age distribution within stage i. Column j, normalized to sum to 1, gives the stable stage distribution within age class j.

In the killer whale example, $\lambda = 1.0254$ and

$$\mathbf{w} = \begin{pmatrix} 0.0370 \\ 0.3161 \\ 0.3229 \\ 0.3240 \end{pmatrix}.$$

[††] These measures are not well defined when there are multiple types of newborn individuals. Cochran and Ellner (1992) suggest a way to calculate them using fertility weighted by reproductive value, but it is not clear to me that the resulting model actually corresponds to that underlying the definitions of $\bar{A}$ and μ_1.

The first few columns of **X** are given by

	Age class					
Stage	1	2	3	4	5	⋯
1	0.0379	0	0	0	0	⋯
2	0	0.0361	0.0321	0.0285	0.0253	⋯
3	0	0	0.0026	0.0047	0.0064	⋯
4	0	0	0	0.0001	0.0003	⋯

Scaling so that the rows sum to one gives

	Age class					
Stage	1	2	3	4	5	⋯
1	1.0000	0	0	0	0	⋯
2	0	0.1115	0.0991	0.0880	0.0782	⋯
3	0	0	0.0078	0.0143	0.0195	⋯
4	0	0	0	0.0004	0.0010	⋯

Thus all the stage-1 (yearling) individuals are in age class 1, which is to be expected since they spend only a single time step there. The distributions for stages $\mathcal{N}_2$–$\mathcal{N}_4$ are shown in Figure 11.6. They are not unreasonable, and where they seem a little strange (e.g., mature adults still appreciably frequent at ages greater than 50, postreproductive females beginning to appear before age 10) the discrepancies are obvious consequences of the stage classification. Because there are only three stages before the postreproductive stage, *some* individuals will become postreproductive after only three iterations. Similarly, because there is a self-loop on stage 3, *some* individuals will remain there indefinitely.

Boucher (1997) used essentially identical calculations to compare the stable age-within-stage distributions from 10 models (turtles, trees, killer whales, herbaceous plants, corals). He found that, as in Figure 11.6, the age-distribution within later stages becomes lower, more symmetric, and flatter. The skewness and the kurtosis of the age-within-stage distribution decreased, roughly exponentially, from early to later stages.

12

Projection and Forecasting

All statistical facts refer to the past. The United States Census of April 1970 counted 203 million of us, but no one knew this until the following November, and the details of the count were published over the course of years. The census differs only in degree from stock market prices, which are hours old before they appear in the daily press. There are no exceptions—not even statistics of intentions—to the rule that all data are to some degree obsolete by the time they reach us.

On the other hand, all use of data refers to the future; the business concern proposing to set up a branch in a certain part of the country consults the census but is interested in what it tells only as an indication of what will come in the future. The branch plant, or a school or hospital, may take 3 years to build and will be in existence over the following 30 years; whether the decision to build was wise depends on circumstances between 3 and 33 years hence, including the number and distribution of people over that period.

12.1 Forecasting: Both Unavoidable and Impossible. Past Data, Present Action, and Future Conditions of Payoff

That all data refer to the past and all use of data to the future implies a line between past and future drawn at “now.” Without continuities that make possible extrapolation across that line statistical data would be useless, indeed the very possibility of purposeful behavior would be in doubt.

The separation between a past census and the future in which action will be implemented is not only the instance of *now*, but also a finite period of time that includes the interval from enumeration of the census to publication of its results, the interval between their publication and the use of them in making a decision, and the interval between the decision and its implementation. The slab of time separating past data on population and the start of operation of a factory or school or telephone exchange decided upon by projecting these data can easily be a decade or more. Prediction often consists in examining data extending several decades back into the past and inferring from them what will happen several decades in the future, with one decade of blind spot separating past and future.

Since our knowledge of population mechanisms is weak, moreover, predictions or forecasts, more appropriately and modestly called projections, must involve some element of sheer extrapolation, and this extrapolation is from a narrow database. Below the observations is an historical drift in underlying conditions that makes the distant past irrelevant to the future. If in the nineteenth century fluctuations in population were caused largely by epidemics, food scarcities, and other factors, and if now these factors are under better control but even larger changes in population are caused by parental decisions to defer or anticipate births, then carrying the series back through the nineteenth century will not be of much help in present extrapolations to the future. Thus, even supposing the continuity that makes forecasting possible in principle, the volume of past data enabling us to make a particular forecast is limited. Moreover, this intrinsic scarcity of relevant data is in addition to the shortcomings of past statistical collections. For such reasons some of those who are most knowledgeable refuse to take any part in forecasting.

Yet ultimately the refusal to forecast is absurd, for the future is implicitly contained in all decisions. The very act of setting up a school on one side of town rather than the other, of widening a road between two towns, or of extending a telephone exchange is in itself a bet that population will increase in a certain way; not doing these things is a bet that population will not increase. In the aggregate implicit bets, known as investments, amount to billions of dollars each year. The question is only who will make the forecast and how he will do it—in particular, whether he will proceed intuitively or use publicly described methods. As Cannan (1895) said in the very first paper using the components method of forecasting, “The real question is not whether we shall abstain altogether from estimating the growth of population, but whether we shall be content with estimates which have been formed without adequate consideration of all the data available, and can be shown to be founded on a wrong principle.”

In any concrete investment decision, the bet on population is combined with a bet on purchasing power, on preferences for one kind of goods rather than another, on technology as it affects alternative methods of production. The component of the bet that is our interest in this book, population, is somehow incorporated into a package of bets.

12.1.1 Heavy Stakes on Simultaneous Lotteries

A school construction program, for example, cannot be based on population forecasts alone; it requires participation rates as well—what fraction of the school-age population will want to attend school. These two elements both change over time. During the 1960s, for example, United States population and participation rates both rose among youths of college age. Those 20 to 24 years of age numbered 10,330,000 in 1960 and 15,594,000 in 1970, a rise of 51 percent; the fraction enrolled at school went from 13.1 to 21.5 percent, a rise of 64 percent over the same decade. The ratio for the population being 1.51, and the ratio for participation 1.64, the combined effect of these two factors was the product of 1.51 and 1.64 or 2.48, equal to the ratio of 1970 to 1960 enrollment. Planning a school for any future period is in effect taking out a package of at least two lottery tickets.

Sometimes a movement of one of the factors counteracts a movement of the other, and in that case one turns out to be better off with the package bet than one would have been with either component alone. New college enrollments have started to decline in some parts of the country, while entrants are still the cohorts of the 1950s when births were constant or rising. Thus so far the two offset each other. But if decline in the participation rate continues through the 1980s, when the college-age cohorts will decline, the drop in attendance will be rapid.

It is often said that population forecasts should be made in conjunction with forecasts of all the other variables with which population interacts. The state of the economy is certainly related to population: marriages, and hence first births, will be numerous in good times, and will be few when incomes are low and unemployment high. This is the case in advanced countries; in poor countries births will fall with the process of economic development, and the sooner the rise in income the sooner will be the fall in births, at least on the theory that has been dominant. In both rich and poor countries shortages of land, energy, minerals, food, and other resources will, through different mechanisms, restrict population growth.

Our present capacity to discern such mechanisms is embarrassingly limited, for reasons briefly explored in Section 17.5. At best we can suggest which factors may be related to which other ones, but even the direction of effect, let alone quantitative knowledge of the relations, is still largely beyond us. That is why most of the work reported in this chapter, like most population projection as practiced, concentrates on demographic variables alone.

12.1.2 Projection as Distinct from Prediction

The preceding chapters, mostly based on stable theory, have also been concerned with projection in a population closed to migration, but always with *long-term* projection. They answered a variety of forms of such questions as: what will be the *ultimate* age distribution if present birth and death rates continue, or what difference will it make to the ultimate rate of increase if the birth rate to women over 35 drops permanently to half its present value, everything else remaining unchanged? The long-run answers are usually simpler than the *medium-term* ones of population projection.

The questions asked in the preceding chapters were so clearly of the form, "What will happen if ... ?" that there was no need to stress the conditional character of the statements referring to the future. Medium-term population projection offers the appearance of prediction, but the detailed reference to the future need not alter the conditional character of the numbers produced. Section 12.2 begins the treatment of projection, while Section 12.5 concerns prediction or forecasting, that is, statements intended to apply to a real rather than to a hypothetical future, but in fact the two topics are inseparable.

A projection is bound to be correct, except when arithmetic errors make the numbers constituting its output inconsistent with the assumptions stated to be its input. On the other hand, a forecast is nearly certain to be wrong if it consists of a single number; if it consists of a range with a probability attached, it can be correct in the sense that the range straddles the subsequent outcome the stated fraction of times in repeated forecasting. A probability can—indeed, should—be attached to a forecast, whereas a probability is meaningless for a (hypothetical) projection.

The distinction between projection and forecasting (Keyfitz 1972a), has become critically important in population ecology, where environmental fluctuations and density-dependence make forecasting even more difficult, but projection even more useful. In particular, a projection can be interpreted as providing information about the current situation rather than future population ("conditions here and now are such that if they were somehow held constant, the population would grow at this rate, with this structure, etc."). Such conclusions are particularly valuable in comparative studies (see Section 13.4).

Despite the apparently sharp distinction, projections put forward under explicit assumptions are commonly interpreted as forecasts applying to the real future. Does the intention of the authors or the practice of the users determine whether projection or forecasting has occurred in a particular instance? Insofar as the assumptions of a projection are realistic, it is indeed a forecast. Those making projections do not regard all assumptions as equally worth developing in numerical terms and presenting to the public, but at any given moment they select a set regarded as realistic enough to be of interest to their readers. When they change to a different set, users suppose that the old assumptions have become unrealistic in view of current demographic events. More will be said later about the dangers and excitements of forecasting; the section immediately following deals with projection.

12.2 The Technique of Projection

Projection in demography is calculating survivors down cohort lines of those living at a given point in time, calculating births in each successive period, and adding a suitable allowance for migration. Of the various ways of looking at population dynamics, the one most convenient for the present purpose is the matrix approach (Chapter 7). Such calculations were used, before Leslie (1945) expressed them in matrix form, by Cannan (1895), Bowley (1924), and especially Whelpton (1936).

Survivorship. Within the supposedly homogeneous subpopulation the first step is to convert the death rates that are assumed to apply at each period in the future into a life table, or else directly assume a life table for each future period. The column used for the purpose is the integral

$$ {}_5L_x = \int_0^5 l(x+t)\,dt, $$

which in general varies over time as well as over age.

If the population at the jumping-off point of the projection is ${}_5N_x^{(0)}$ for ages $x = 0, 5, 10, \ldots$, the population 5 years older and 5 years later may be approximated as

$$ {}_5N_{x+5}^{(5)} = {}_5N_x^{(0)} \left(\frac{{}_5L_{x+5}}{{}_5L_x} \right). \tag{12.2.1} $$

For changing rates a new life table would be assumed operative at each future 5-year interval and applied as in (12.2.1). By this means the population would be "survived" forward until the cohorts of the initial period were all extinguished.

Expression (12.2.1) is not exact even if the life table is appropriate, unless the distribution within the age group $(x, x+5)$ happens to be exactly proportional to the stationary population $l(x+t)$, $0 \leqslant t \leqslant 5$. If the population is increasing steadily at rate r, the distribution within the age interval will be proportional to $e^{-rt}l(x+t)$; and the greater r is, the more the population will be concentrated at the left end of the age interval. Being on the average younger within each age interval, it will average somewhat higher survivorship from one age interval to the next in the part of the span of life where mortality is rising with age. [Using the Taylor expansion method of Section 2.2 find a formula for the small addition to (12.2.1) to allow for

Table 12.1. Population and life table data for estimating survivorship of females 70–74 years of age, England and Wales, 1968

Age	${}_5N_x$	${}_5L_x$	l_x
65	1,280,500	391,456	82,172
70	1,033,400	339,660	73,827
75	753,600	265,710	61,324

Source: Keyfitz and Flieger (1971), pp. 152, 154.

this, and calculate its amount from the data of Table 12.1. A different form of correction is given as (11.1.16) in Keyfitz (1968, p. 249).]

Reproduction. To estimate the births of girl children to women in each 5-year age interval we need a set of age-specific birth rates, say F_x, for ages x to $x+4$ at last birthday (for brevity omitting the prescript). The F_x might be obtained from past experience, over 1 calendar year, say, of observed live female births ${}_5B_x$ to women of exact age $(x, x+5)$. If midperiod population corresponding to these births numbered ${}_5N_x$, then $F_x = {}_5B_x/{}_5N_x$.

In the first time period we begin with ${}_5N_x^{(0)}$ women aged x to $x+4$ at last birthday and, applying (12.2.1), end with ${}_5N_{x-5}^{(0)}({}_5L_x/{}_5L_{x-5})$ women in this age class. Within the age class, an estimate of the mean surviving women (or the average exposure per year during the 5 years) is the average of these two numbers. Multiplying this average by 5 yields the total female exposure to conception during the 5 years. Multiplication of the total exposure by F_x yields the expected births to women aged x to $x+4$ $(l_0 = 1)$:

$$\frac{5}{2}\left[{}_5N_x^{(0)} + {}_5N_{x-5}^{(0)}\left(\frac{{}_5L_x}{{}_5L_{x-5}}\right)\right]F_x, \quad x = \alpha, \alpha+5, \ldots, \beta-5. \tag{12.2.2}$$

However, since we are interested in the population aged 0 to 5 rather than the number of births, we must multiply (12.2.2) by ${}_5L_0/5$ and sum over all the fertile ages between α and β:

$${}_5N_0^{(5)} = \frac{1}{2}\sum_{\alpha}^{\beta-5}\left[{}_5N_x^{(0)} + {}_5N_{x-5}^{(0)}\left(\frac{{}_5L_x}{{}_5L_{x-5}}\right)\right]{}_5L_0F_x. \tag{12.2.3}$$

This corresponds to the calculation of $\mathbf{n}_1(t+1)$ using an age-classified population projection matrix with fertility defined as in Section 3.3.1. There, we used this same tactic of calculating the number of women exposed to the risk of childbearing as the average of the numbers at the beginning and at the end of the period. This is a slight overstatement of exposure if the increase is at a steady rate, that is, rising geometrically and so concave upward. The correction for this has been presented elsewhere (Keyfitz 1968, p. 252) and will here be disregarded.

Extension to All Ages and Both Sexes. If convenience in arranging the worksheets were thereby served, we could first confine the calculation to ages under β, the end of reproduction. The survivorship and birth operations described above can be repeated for an indefinite sequence of 5-year cycles in disregard of the population beyond reproduction. Older ages can be filled in by repeated application of (12.2.1).

Males, like females beyond age β, can be dealt with after the projections for the female ages under age β are completed. Pending parental control of the sex of offspring, we can suppose a fixed ratio of males to females among births. If this ratio is taken as s, ordinarily a number close to 1.05, then multiplying the girl births by s, or the number of girls under 5 years of age by $s_5L_0^*/_5L_0$ (where $_5L_0^*$ refers to the male life table, $_5L_0$ to the female), will give the corresponding number of boys. The male part of the projection can then be filled out by repeated application of (12.2.1), using a life table for males.

The procedure described to this point can alternatively be arranged as matrix multiplication to secure a numerically identical result. The number of females in successive age groups is recorded in the top half of a vertical vector $\mathbf{n}$ and the number of males in the bottom half. The vector, containing 36 elements if 5-year age intervals to 85–89 are recognized for each sex, is premultiplied by a matrix $\mathbf{M}$ whose nonzero elements are in its subdiagonal and its first and nineteenth rows. Population after t cycles of projection at constant rates will be $\mathbf{n}(t) = \mathbf{M}^t\mathbf{n}(0)$. If fertility ends at age $\beta = 50$, the upper left-hand corner of the matrix will be a 10×10 submatrix containing birth elements in its first row, and constituting the self-acting portion of the larger matrix (Table 12.2). Every part of the preceding description can be readily altered to provide for different rates from one period to another, and in practical work rates are assumed to change.

The upper left-hand 10×10 submatrix of $\mathbf{M}$ shown in Table 12.2 can be called $\mathbf{A}$ and analyzed with the methods of Chapter 7. By drawing a graph, we can show $\mathbf{A}$ to be irreducible in that its positive elements are so arranged as to permit passage from any position to any other. Its primitivity—that there is at least one power k such that all elements of $\mathbf{A}^k$ are positive for that k—follows as long as any two relatively prime ages have nonzero fertility. Alternatively, the entire matrix $\mathbf{M}$ can be analyzed, as long as the irreducible nature of the matrix is recognized, so that the appropriate eigenvalues and eigenvectors are used to draw conclusions about future population (Section 7.2.2).

This description corresponds to the usual execution of the projection in being female dominant: all births are imputed to females. The same theory is applicable to a process in which births are imputed to males. With the prevailing method of reproduction a male and a female are required for each birth, and with either male or female dominance the other sex is implicitly taken to be present in whatever numbers are required. Two-sex models in which neither sex is dominant are nonlinear, with reproductive rates

Table 12.2. Female-dominant projection matrix **A** with 36 rows and 36 columns in 5-year age intervals; $l_0 = l_0^* = 1$; male life table distinguished by asterisk

$$\left(\begin{array}{ccccc|ccc}
0 & 0 & \frac{{}_5L_0}{2}\left(\frac{{}_5L_{15}}{{}_5L_{10}}F_{15}\right) & \frac{{}_5L_0}{2}\left(F_{15}+\frac{{}_5L_{20}}{{}_5L_{15}}F_{20}\right) & \cdots & 0 & 0 & \cdots \\
\frac{{}_5L_5}{{}_5L_0} & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots \\
0 & \frac{{}_5L_{10}}{{}_5L_5} & 0 & 0 & \cdots & 0 & 0 & \cdots \\
0 & 0 & \frac{{}_5L_{15}}{{}_5L_{10}} & 0 & \cdots & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots \\
\hline
0 & 0 & s\frac{{}_5L_0^*}{2}\left(\frac{{}_5L_{15}}{{}_5L_{10}}F_{15}\right) & s\frac{{}_5L_0^*}{2}\left(F_{15}+\frac{{}_5L_{20}}{{}_5L_{15}}F_{20}\right) & \cdots & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & \cdots & \frac{{}_5L_5^*}{{}_5L_0^*} & 0 & \cdots \\
0 & 0 & 0 & 0 & \cdots & 0 & \frac{{}_5L_{10}^*}{{}_5L_5^*} & \cdots \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & \ddots \\
\vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots
\end{array}\right)$$

depending on the relative frequencies of males and females of the various ages or stages. For a review of such models and the dynamics they can produce, see MPM Chapter 17).

12.2.1 Age Versus Other Variables

As we saw in Chapter 3, the concentration on age in the projection described above, though it follows a long and honorable tradition, may be considered arbitrary from a wider viewpoint. We know that mortality and fertility rates depend on age, but they also depend (though less decisively) on rural and urban residence, education, social class, and many other variables.

A projection matrix that recognizes rural and urban residence but disregards age can easily be devised:

$$\mathbf{An}(0) = \begin{pmatrix} a_{rr} & a_{ur} \\ a_{ru} & a_{uu} \end{pmatrix} \begin{pmatrix} n_r \\ n_u \end{pmatrix} (0), \qquad (12.2.4)$$

where a_{ru} is the fraction moving from rural to urban, and a_{ur} that from urban to rural. If a_{ur} and a_{ru} were zero, a_{rr} and a_{uu} would be the change ratios of rural and urban parts. By repeated application of (12.2.4), the population at the end of t periods of 5 years is $\mathbf{A}^t\mathbf{n}(0)$. The procedure would be preferred if a move to an urban area caused a change in natural increase greater than the change in going from one age group to another.

Recognition of rural and urban residence does not preclude recognition of age. We can make area the primary classification and show the breakdown by age within each area, as does Rogers (1968, p. 13), or else we can take age first and show within each age the rural and urban areas, as does Feeney (1970). Both of these approaches have been used in recent multiregional matrix models (Rogers 1995, Hastings 1992, Bravo et al. 1995, 1997, Sanz and Bravo 1999, 2000, Lebreton 1996, Lebreton and Gonzalez-Davila 1993). Hunter and Caswell (2004) have presented an approach that makes it easy to move from one to the other.

The usual model of projection recognizes age and sex, sometimes region, less often rural and urban parts, and in a few countries race. But age and sex have been the universal dimensions of population projection since the 1940s; no other is accepted as indispensable.

The concentration on age and sex could represent a cultural lag. It might be that at certain times this breakdown is important, at other times less so. To see just how much improvement results from recognizing age and sex, as against not recognizing them, the experiment of Table 12.3 was carried out. For each of 20 populations a 15-year projection was made by age and sex, using the age-sex-specific rates of birth and death for the initial moment—1 year, 3 years, or 5 years, but always symmetrically arranged around the midpoint of the stated jumping-off year. Except for the constancy of the rates assumed and the omission of migration, this is the usual projection by periods and is designated as (*b*) in the table.

Contrasted with this is the simpler projection in which no account is taken of age and sex, but the crude or observed rates of birth and death are made to provide a rate of increase, and that rate is applied to the total population $N_t = \|\mathbf{n}(t)\|$ of the jumping-off moment. This method, which has not been used by demographers in good standing for the last 70 years or more, is shown as (*a*) in the table.

The observed population 15 years later is the last column of the table, and projections (*a*) and (*b*) may be compared with it. Our intention, before seeing the outcome, was to estimate quantitatively how much improvement is due to the incorporation of age and sex, by seeing how much closer column (*b*) was to the subsequent performance. That column (*a*) would turn out equally well was never anticipated. In fact, in the slight majority of the instances (marked with an asterisk) the simple projection (*a*) comes closer to the performance than (*b*), which recognizes age and sex.

There are obvious qualifications in regard to the result. An experiment involving 20 cases is too small, especially since they are not entirely independent of one another; when the European birth rate was falling, somewhat the same influences acted on age distribution in all countries. (However, a similar experiment on an additional 20 cases, using a 10-year prediction span, gave a similar result.) The use of fixed rather than changing rates of birth and death is a handicap equally to methods (*a*) and (*b*). The same is true for the absence of migration, omitted because data are

Table 12.3. Projections for 20 populations over a 15-year time period, (a) neglecting age and sex, and (b) taking account of age and sex, in both cases neglecting migration. The projections are compared with observed populations 15 years later; the closer of (a) and (b) is starred

Time (t)	Population (000s) $N_t/1000$	Rate of natural increase per 1000 population $1000r$	Projected 15 Years		Observed time $t+1$ $N_{t+15}/1000$
			Without age and sex $N_te^{15r}/1000$ (a)	By age and sex Total of $\mathbf{A}^3\mathbf{n}(t)/1000$ (b)	
United States					
1920	106,630	10.93	125,627*	123,665	127,252
1945	131,976	11.64	157,153*	150,910	179,990
Japan, 1951	84,260	15.71	106,651	106,607*	98,859
Belgium, 1935	8,279	3.01	8,661*	8,346	8,646
France, 1876	36,824	3.54	38,832	38,447*	38,139
Hungary, 1951	9,423	8.76	10,746	10,427*	10,179
Netherlands, 1935	8,387	11.70	9,996*	9,889	10,114
Norway, 1951	3,296	10.08	3,834	3,674*	3,753
Sweden					
1780	2,104	8.59	2,393	2,381*	2,274
1805	2,418	6.39	2,661*	2,666	2,573
1830	2,876	6.70	3,180	3,216*	3,296
1855	3,625	9.23	4,163*	4,144	4,164
1880	4,572	11.96	5,470*	5,485	4,896
1905	5,278	10.68	6,195*	6,265	5,876
1930	6,131	3.18	6,431*	6,409	6,630
1945	6,636	9.12	7,609*	7,287	7,499
Switzerland, 1951	4,749	7.39	5,306*	5,118	5,917
England and Wales					
1941	38,743	1.41	39,571*	38,259	44,667
1946	40,595	7.36	45,333*	43,273	46,166
Australia, 1951	8,422	13.61	10,329*	9,887	11,550

more difficult to come by, and also because it presents special problems: should one use the absolute numbers of the jumping-off year or the rates, and how would they be applied by age and sex?

The conclusion from the experiment is that in some circumstances the effect of age is important, and in others it is not. The experiments summarized in Table 12.3 are not the only ones in which simple models seem to outperform models including more demographic detail (Keyfitz 1981, 1982; see the review by Rogers 1995b). Such results hardly constitute justification for abandoning age, especially since we usually need the forecast by age, but we should not be under the illusion that projection by age and sex is a powerful technique for discerning the future.

To find which of the other dimensions can provide real gains in accuracy of forecast will require additional experimenting similar to that of Table 12.3. We ought to be trying to project marriages, and then applying rates of nuptial fertility; perhaps not age within marriage but duration of marriage is the superior variable; perhaps first births, second births, and so on are more easily extrapolated separately than are births of all orders together, and so the model ought to be order-specific.

Whatever the variables recognized, their change from period to period is easily provided for. Each 5-year period would have its own projection matrix, so that after t periods we would have not $\mathbf{A}^t\mathbf{n}(0)$, but instead, if $\mathbf{A}_{i-1}$ is the matrix for the ith period,

$$\mathbf{n}(t) = \mathbf{A}_{t-1} \cdots \mathbf{A}_2\mathbf{A}_1\mathbf{A}_0\mathbf{n}(0). \tag{12.2.5}$$

Several different cases of time variation must be distinguished, because different analytical methods are available for each: periodic variation (MPM Chapter 13), stochastic variation (MPM Chapter 14), and variation due to nonlinear feedback between the population and its own vital rates (MPM Chapter 16). Whereas the changing matrix can be symbolically incorporated into a formula without difficulty, and analyzed with some effort, the attempt to determine the elements of the matrix, that is, to say what future mortality and fertility will be, is the most difficult problem of demography, and will be the subject of Section 12.4.

12.2.2 Projection in a Heterogeneous Population

When a heterogeneous population is projected without recognition of its subpopulations, the result is always lower than what would have been found by applying the same projection process to the subpopulations and adding. This was shown to be true in Section 1.5, without recognizing age. It is true also for projections by age, but to prove it in general for an arbitrary initial age vector is difficult.

As an example that will suggest the quantitative effect, the United States population in 1966 without breakdown by color, as projected with a life table made from the deaths in that year and the fertility rates implied by the 1966 births, results in a total population for 1981 of 230,477,000. Recognizing the two separate groups of Whites and Nonwhites, whose 1966 populations, births, and deaths add exactly to the totals for the United States, constructing life tables and age-specific birth rates for the two groups separately, and then projecting each by means of its own life table and birth rates, gives 1981 Whites as 199,287,000 and Nonwhites as 31,441,000, which add to 230,728,000. The deficiency of the projection not recognizing color is 251,000 or, about 1 part in 1000, after 15 years. After 100 years the two separate projections add to 8 percent more than the projection without breakdown by color.

12.3 Applications of Projection

Projection leads itself to making “if . . . then . . . ” statements; indeed it consists of such statements. If England and Wales have the same age-specific birth and death rates as in 1968 for the following 25 years, the total population will grow from the 48,593,000 estimated for 1968 to 54,869,000 in 1993. The increase will average 0.0049 or 4.9 per 1000 per year, less than the intrinsic rate r of 1968, which was 6.9 per 1000. The slower rate of the projection is due to an unfavorable 1968 age distribution.

The United States 1967 population, projected at 1967 rates, amounts to 230,109,000 by 1982, an average rate of increase of slightly over 1 percent per year. In this case the rate of increase is raised by an exceptionally large proportion of women in childbearing ages; the intrinsic rate r of 1967 was only 0.74 percent.

12.3.1 Population Dynamics with One Cause of Death Eliminated

A transparently simple application of projection involves the effect on population growth of eliminating a particular cause of death, say cancer. We want the effect of this change separate from the effects of all other changes that may take place. Excluding the effect of the birth rate or of income incurs no difficulties, but excluding changes in other causes of death raises a problem if these causes would become more (or less) serious as a direct result of eliminating cancer. Since we know little about the interrelations of the several causes, we assume that age by age the others remain unchanged when cancer is eliminated. (Section 2.6)

A life table was calculated omitting all mortality from cancer, and another omitting all mortality from heart disease, using the technique of Section 2.6 and data for the United States, 1964. The unaltered table and each of these two modifications were applied to the 1964 female population of the United States to separate out the effects of the two causes on the rate of increase and on the age distribution of the ultimate stable population. The resulting rates and age distributions are shown in Table 12.4, always supposing that the birth rates of 1964 continue to prevail.

The intrinsic rate of natural increase is affected by less than 1 part in 200 by omission of heart disease: it rises from 0.01573 to 0.01579. Not many women are struck down by heart disease at ages before the end of childbearing. Cancer deaths are fewer in total, but they affect younger ages in somewhat more cases and so their omission produces a slightly greater rise in the rate of increase.

To represent the effect on age distribution we have used the stable equivalent $\mathbf{q}$, corresponding to an age distribution, or an age-sex distribution $\mathbf{n}(0)$, obtained in (8.8.3) as

$$\mathbf{q} = \frac{\mathbf{A}^t \mathbf{n}(0)}{e^{5rt}}, \tag{12.3.1}$$

where t is large, say 100 cycles of 5 years, and e^{5r} is the ratio of increase per cycle as t becomes large. The stable equivalent is the number of persons in

Table 12.4. Ultimate age distribution and intrinsic rates corresponding to United States 1964 population, with (*a*) all causes of death, (*b*) all causes except heart disease, and (*c*) all causes except cancer (thousands of persons)

	(*a*) All causes of death		(*b*) Death from heart disease removed		(*c*) Death from cancer removed	
	Stable equivalent		Stable equivalent		Stable equivalent	
Age	Males	Females	Males	Females	Males	Females
0–4	10,535	10,125	10,537	10,127	10,537	10,126
5–9	9,698	9,329	9,699	9,330	9,700	9,330
10–14	8,943	8,609	8,942	8,608	8,944	8,609
15–19	8,229	7,941	8,227	7,939	8,230	7,940
20–24	7,545	7,317	7,543	7,315	7,546	7,315
25–29	6,910	6,737	6,911	6,737	6,914	6,736
30–34	6,326	6,195	6,332	6,198	6,332	6,197
35–39	5,774	5,682	5,795	5,693	5,786	5,693
40–44	5,238	5,193	5,290	5,216	5,260	5,219
45–49	4,700	4,719	4,808	4,762	4,740	4,768
50–54	4,141	4,249	4,335	4,326	4,212	4,332
55–59	3,547	3,780	3,864	3,906	3,662	3,902
60–64	2,921	3,299	3,390	3,503	3,088	3,463
65–69	2,268	2,788	2,907	3,109	2,485	2,992
70–74	1,629	2,242	2,429	2,721	1,872	2,474
75+	2,003	3,491	7,833	11,856	2,596	4,168
Total	90,407	91,696	98,842	101,346	91,904	93,264
Total both sexes	182,103		200,188		185,168	
Females $\frac{15\text{–}44}{\text{Total}} \times 100$		42.60		38.57		41.92
Female mean age		31.32		36.79		32.03
Intrinsic rates						
Birth		0.02348		0.02124		0.02308
Death		0.00774		0.00546		0.00726
Natural increase		0.01573		0.01579		0.01582
Sex ratio						
All ages	0.986		0.975		0.985	
Ages 75+	0.574		0.661		0.623	

each category of age and sex that could be projected to (distant) time t by multiplication by e^{rt} at all ages, with assurance that the same numbers at time t would result as from projection of the observed population by the usual components method described in Section 12.2.

The stable equivalent shown in Table 12.4 is little affected at younger ages by the omission of heart disease or cancer, and even by age 50 the increase is less than 5 percent for males and 2 percent for females. At the very oldest ages of 75 and over, however, the effect of eliminating heart disease is dramatic: a nearly fourfold rise for men and well over a threefold increase for women.

The net result is a rise in the average age of females from 31.32 to 36.79 years with the elimination of heart disease, and to 32.03 with the elimination of cancer. The percentage of females in the childbearing ages 15 to 44 falls from 43.21 to 39.56 with the elimination of heart disease, and to 42.54 with the elimination of cancer.

The main limitation of the result given in Table 12.4 is the certainty that if heart disease were eliminated other causes would also be affected. They would be increased insofar as heart disease selects weaker individuals, more subject to other ailments than randomly selected members of the population, an issue raised in Section 4.4.

12.3.2 Effect of Immediate Drop to Replacement Fertility

Continuing with the effort to make population projection perform experiments, we ask another question: "What would be the ultimate stationary population if birth rates were to drop immediately and permanently to bare replacement level?" Though a drop to bare replacement would sooner or later terminate growth, a large concentration of women at childbearing ages exists as a result of growth in the past, and the ultimate level would be considerably higher than the present population.

The question may be answered by a population projection, applying the observed life table and observed age-specific birth rates divided by R_0, the net reproduction rate. We refrain from trying to guess what mortality is going to do, nor do we suppose changes in the age distribution of childbearing. The object is to see the outcome of a drop to replacement fertility if everything else remains unchanged.

Taking the birth and death rates for the year listed, making a life table from the death rates, dividing the birth rates by R_0, and projecting some 500 years into the future gives the results shown in Table 12.5 for six countries. Thus Colombia showed 17.993 million people in 1965; the projection climbed to 29.786 million, a rise of 65.5 percent. Such a projection takes full account of the initial age distribution.

The results can be compared with the corresponding calculation on the stable age distribution. We saw in Section 8.6 that the ratio of the ultimate stationary to the present population, on the assumption of an immediate drop to bare replacement fertility, is

$$\frac{b\overset{o}{e}_0}{r\mu}\left(\frac{R_0 - 1}{R_0}\right), \tag{12.3.2}$$

where the birth rate b, the rate of natural increase r, the net reproduction rate R_0, and the mean age μ of childbearing in the stationary population are all calculated before the fall. Applying (12.3.2) to Colombian females of 1965 gives 1.59, that is, an ultimate rise of 59 percent over the 1965 population. This is somewhat less than the 66 percent rise on the projection, so presumably the observed 1965 Colombian sex–age distribution departed

Table 12.5. Current and ultimate stationary populations, on assumption that birth rates drop immediately to stationary level, for six countries

Population	Number (000s) Current	Number (000s) Ultimate	Percent increase to ultimate
Chile, 1965	8,584	12,916	50.5
Colombia, 1965	17,993	29,786	65.5
Ecuador, 1965	5,109	8,518	66.7
Italy, 1966	53,128	62,189	17.1
Peru, 1963	14,713	23,080	56.9
United States			
1966	195,857	259,490	32.5
1967	197,863	267,096	35.0

from the stable one in a direction favorable to reproduction. However, the departure could be an artifact due to defects in the statistics. In the face of inaccurate primary data it is an advantage to have available both the projection and the stable formula (12.3.2).

A number of such experiments are described by Frejka (1973). He sees what the ultimate stationary population will be with an immediate drop and with a drop spread over 25 years. He does not keep other factors constant at their base values, but supposes that mortality will fall according to a certain curve. This has the advantage of greater realism if indeed the projected fall of mortality occurs. On the other hand, it has the drawback of requiring us to keep in mind a particular fall in mortality rather than just constancy. Such a conflict between realism and simplicity is typical of demographic projection.

We have all thought how convenient it would be if we knew the condition at some distant future and could interpolate between the jumping-off point and then. Sooner or later the population must cease to grow; and if we could only guess the date at which it would become stationary, most of the problem of forecasting would be solved, for the interpolation could not involve the gross errors to which extrapolation is subject. Frejka's treatment is an attempt along this line.

12.4 The Search for Constancies

Any forecasting method, whether naive or sophisticated, depends on some function being approximately constant. For the short term it may be sufficient to suppose that the absolute number of births of the current year will be repeated over the next few years, and the same for deaths and migration. More convincing is to suppose that the crude rates will be constant for the next few years. Insofar as age distribution is changing, the rates that are supposed fixed had better be age specific. Insofar as a visible trend appears in the age-specific rates, one may wish to take year-to-year differences of the age-specific rates as fixed; if ${}_5M_x^t$ is the death rate at time t for the age interval x to $x+4$, then ${}_5M_x^t - {}_5M_x^{t-1}$ being constant will give ${}_5M_x^{t+1} = 2{}_5M_x^t - {}_5M_x^{t-1}$. If the past trend looks to be exponential, the logarithm of the numbers in previous years can be supposed to have fixed first differences.

If the trend in total population can be assumed to be logistic, what is fixed is the ratio of the difference of reciprocals in successive years; with N_t standing for the population in year t,

$$\frac{1/N_t - 1/N_{t-1}}{1/N_{t-1} - 1/N_{t-2}}$$

is constant.

Better adapted ways of choosing the entities that are to be held constant are provided by relational methods, which deserve a more detailed treatment.

12.4.1 Relational Methods

To project mortality or fertility or nuptiality—indeed, any component that varies with age—we need to determine not only overall trends but also what is going to happen to the several ages. We could suppose that all ages are affected in the same way, either adding to all age-specific rates some constant for each year that goes by, or multiplying them by some constant. But actual changes are far from being either constant increments or constant multiples applicable to all ages. We could meet this objection by extrapolating each age separately, but that would produce gratuitous errors, as well as inconvenient discontinuities; with most data age distributions would become less and less smooth as we went forward in time on separate projections for the several ages.

We could observe how past changes have occurred, for instance, how successive mortality tables for Sweden differ, and use the Swedish trend as showing the future for countries that are not yet down to the Swedish level of mortality. Similar possibilities exist for fertility and nuptiality. This method has been used, but it presents difficulties. The problem is to find some way of relating the future to the past that involves a small number of parameters, say two, and that is a sufficiently good fit to all times for the two parameters extrapolated to produce acceptable curves for future dates.

Mathematical curves would serve this purpose if any could be found that fitted closely enough. If the age distribution of childbearing could be satisfactorily fitted by a normal curve, for example, we could extrapolate the total, mean age, and variance of ages to find the age distribution of childbearing at future times. But Lotka's (1939, p. 70) normal curve does not fit past data well enough to justify its use for prediction, nor does the incomplete gamma function applied by Wicksell (1931) or Hadwiger's (1940) more elaborate exponential. (Keyfitz 1968, Chapter 6, is a secondary source on these.) Murphy and Nagnur (1972) tried a Gompertz curve, long used for mortality, with some success.

Success in specifying algebraic or transcendental curves has been incomplete enough that most workers resort to sets of model tables, for fertility as well as for mortality. A set of useful model fertility tables has been published by Coale and Trussell (1974), which do for this component what the Coale and Demeny (1966) tables did for mortality.

An alternative is the relational method evolved by Brass (1974) over the course of some years of experimenting. Instead of aiming at the desirable but unavailable mathematical curve, or attempting to sum up the entire range of empirical materials in a set of model tables, the relational method embraces changes through time in a flexible description involving two (or at most three) parameters. Let us see how the Brass method fits mortality; fertility and nuptiality may be treated in a similar way.

Mortality. In the first application, to mortality, Brass found that to go from one life table to another was easy if both were translated into logits, that is, their $l(x)$ columns were transformed by

$$Y(x) = \tfrac{1}{2}\log\left[\frac{1-l(x)}{l(x)}\right].$$

If the early table is distinguished by a subscript s, the relation is the linear

$$Y(x) = \alpha + \beta Y_s(x),$$

or, written out in full,

$$\tfrac{1}{2}\log\left[\frac{1-l(x)}{l(x)}\right] = \alpha + \frac{\beta}{2}\log\left[\frac{1-l_s(x)}{l_s(x)}\right].$$

In forecasting one would take a life table for time t_s as $l_s(x)$ and fit to it the table for later time t_1, finding the constants α_1 and β_1, say. A new α_2 and β_2 would be found on fitting the table for t_2 to t_1, and α_3 and β_3 on fitting the table for t_3 to t_2. The several fittings would indicate a trend in α and in β that could then be extrapolated into the future. The extrapolated α and β for future dates, along with the base life table $l_s(x)$, would carry the life table to the dates in question. Studying mortality trends of the past half-century in England and Wales, Brass considers it possible that in the future β may well come close to unity, and α decrease by 0.5 every 40 years. The trend in α and in β must be worked out for each population on which the method is tried. The method imports nothing—neither a mathematical curve nor model tables—from the experience of other populations than the one under consideration: for England and Wales β and the first difference of α are constant, but other functions of α and β would be used in other populations.

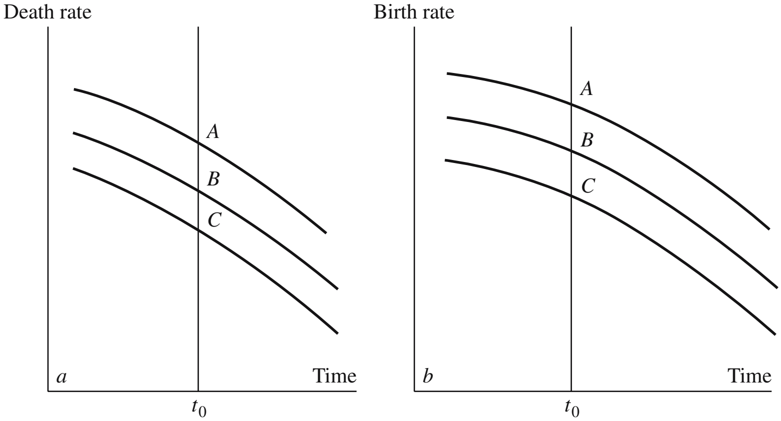


Figure 12.1. Demographic transition for three provinces or other groups designated as A, B, and C.

12.4.2 Are Longitudinal Relations Demonstrated by Cross-Sectional Data?

In seeking a constancy of some function of the population variables through time one is tempted to transfer a relation found in cross-sectional data. Constancy from place to place at a given time ought to be evidence of constancy through time. The provinces of India whose mortality is low seem also to be the ones whose fertility is low; does it not follow that, as mortality falls in the years ahead for the whole of India, so also will fertility? And—only a short step further—the low mortality may be argued to *cause* the low fertility, among other reasons because parents aim to have a certain number of children who survive to maturity. These statements may be true, and if they are they would be very helpful in forecasting. But other evidence than the cross-sectional relation is required to prove them, as can be seen from the very simple cases of Figures 12.1 to 12.3.

Figure 12.1 is a stylized illustration of the problem for three provinces or other population groups, A, B, and C, with deaths in Figure 12.1a and births in Figure 12.1b. From the birth and death rates of the provinces at t_0, i.e., taking vertical cross sections of the two families of curves at t_0, we can obtain a cross-sectional correlation of birth rates and death rates, and it would be positive; the provinces are in the same sequence A, B, and C. On the other hand, we obtain a longitudinal correlation by comparing birth and death rates through time for any particular province. In Figure 12.1 the longitudinal, like the cross-sectional, correlations of birth and death rates are positive.

But we could draw the curves differently and obtain the opposite result, as has been done in Figure 12.2, where deaths are declining and births are rising. The correlation of birth and death rates through time for any of the

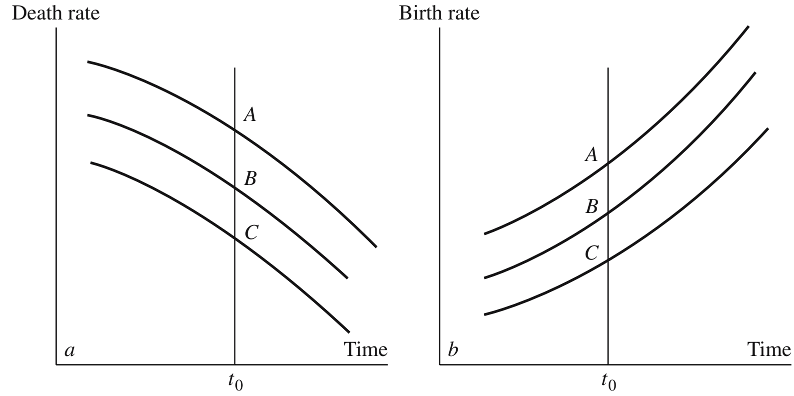


Figure 12.2. Opposite to demographic transition for three provinces A, B, and C.

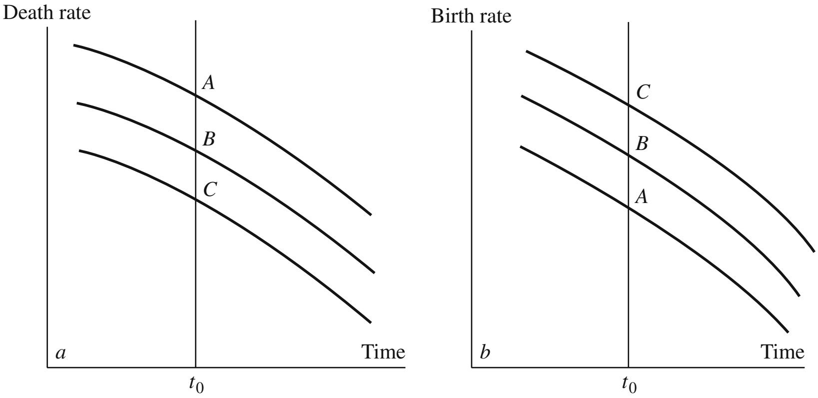


Figure 12.3. Demographic transition for three provinces with sequence reversed.

provinces A, B, or C would be negative, though that for the cross section at any given time would again be positive.

Is the cross-sectional correlation even *necessary* to the correlation in time? Examine Figure 12.3, drawn in the same way as Figure 12.1, except that in Figure 12.3*b* the births have been labeled differently. In this diagram the cross-sectional correlation is negative, while the correlation through time is positive. There is nothing impossible a priori about this arrangement. The cross-sectional correlation is neither necessary nor sufficient for the longitudinal one (Janowitz 1971). This is part of the elusiveness of the inference of future trends from past conditions.

12.5 Features of Forecasting and Forecasting Error

No one expects botanists, geographers, or anthropologists to make declarations regarding the future.* Physicists do make predictions, but within an experimental situation closely protected against interference from outside events in the real world. Demographers, on the other hand, are expected to predict future population as it will actually occur. And they respond as best they can; much of the published literature in demography, and even more of the unpublished, whether written by amateurs or by professionals, consist in statements about the future.

We saw that a projection over t units of time can be written as $\mathbf{A}^t\mathbf{n}$, where $\mathbf{A}$ is a projection matrix and $\mathbf{n}$ the initial population vector. If we know the elements of $\mathbf{A}$ and of $\mathbf{n}$, the estimate of the future consists in the multiplication. We may incorporate in $\mathbf{A}$ and $\mathbf{n}$ several regions, rural and urban populations, ethnic and racial groups, or other divisions (Chapter 7). We may admit that the transition probabilities are a function of time and say (equation 12.2.5) that at time t the population is

$$\mathbf{n}(t) = \mathbf{A}_{t-1}\mathbf{A}_{t-2}\cdots\mathbf{A}_0\mathbf{n}(0).$$

The preceding paragraph, like Section 12.2, is complicated enough to have the effect of concealing the logical status of what we are doing. Age and other distinctions do matter, but their numerical effects are often small in comparison with the prediction error within any category. Let us drop the categories and think only of total population projected over one period, say of 30 years; thus we are back to multiplying scalars rather than matrices and vectors. For example, the population of the United States was counted in 1970 at 203 million; what will be its number in the year 2000? The ultimate in simplicity is to say that, *if* the population increases by 33 percent, it will number 270 million. Most projections snow us with breakdowns by age, sex, race, and region, and take a number of time intervals, all of them valuable, but in their combination causing us to lose perspective on the problem. By peeling off the breakdown, we arrive at the essence of calculation of future population in its two aspects of projection and prediction.

Projection is where the 33 percent is hypothetical. All projection consists of such statements as the following: "*If* (which we do not assert) the population grows at 33 percent in 30 years, then by the year 2000 it will have increased to 203 × 1.33 = 270 million." The projection consists in performing the multiplication, is conditional on the 33 percent, and is as unassailable as the laws of arithmetic. No projection risks being in error; it

---

*Actually, this statement has become less and less true over time, as problems in conservation and resource management have required ecologists to try harder and harder to predict the future growth of plant and animal populations. This has led them to confront exactly the same problems of forecasting and forecasting error as human demographers.

cannot be proved or disproved by what the population actually turns out to be in the year 2000.

Stated in this stripped-down form, the result is wholly uninteresting. At most it does the reader the service of performing the multiplication, and in an age of computers no one would pay much to be told the bare arithmetical fact that $203 \times 1.33 = 270$. The demographer's service surely starts with his comment on the realism of the 1.33, and with that he enters the field of forecasting.

Forecasting is where the 33 percent is taken as the real prospective increase: "The population *will* grow 33 percent in 30 years." Unless this is to be soothsaying, we now need evidence for the 33 percent. Such evidence may be one or another kind of extrapolation; in the simplest case the argument would be that the next 30 years will show an increase of 33 percent because the last 30 years showed such an increase. A more sophisticated version would assert, not that the ratio will be constant, but that the rate of change in the ratio will be constant. In the most general statement the forecast asserts that some more complex function, possibly involving economic, ecological, psychological, or other variables sufficient to determine the future population numbers, will remain constant into the future, and the task of forecasting consists in the search for such a functional invariance.

One circumstance that can make prediction interesting is what makes a poker game interesting: the fact that the cards, first hidden as the future is hidden, are later turned face upward on the table, and so test the accuracy of the forecast. In economic forecasting a frequent span is 6 months or 1 year; such a period is not too long for the predictor to see whether she guessed right, or at least more accurately than someone else, and to learn from the experience. But 30 years, which is a typical span for demographic forecasting, is too long a wait for learning to occur. To wait 30 years to see the cards removes the interest inherent in the rapid alternation of forecast–verification–forecast–verification that holds us at the poker table.

12.5.1 Extrapolation Versus Mechanism

Any forecasting that takes itself seriously tries to base itself on the underlying mechanisms, rather than merely extrapolating trends. Yet the distinction is not always sharp. The components projection on which much stress is laid is indeed a mechanism, but it still depends on extrapolation of birth, death, and migration rates.

A mechanism that lends itself to prediction is Easterlin's model described in Section 14.11, in which age-specific birth rates depend negatively on the size of the cohort of an age to marry and bear children at the given time. In effect it predicts waves of twice the length of generation, so that in the United States birth rates would be low for the next decade or more, and only in the 1990s would they pick up again. [Of course, even if the mechanism is operative exactly as described, the intensity of its operation, measured by the constant γ of (14.11.1), would still have to be ascertained by extrapolation.]

Other mechanisms are less useful for forecasting. That working women have fewer children than nonworking women is true without telling us whether seeking work is primary and lack of children the result, or whether selection is involved: women who do not have children take jobs. But suppose the first case: that the fall in births during the 1960s was due to more and more women preferring work outside the home to child raising. Perhaps they needed the income; perhaps outside work came to carry greater prestige. Certainly women's liberation will reduce childbearing, at least until the day when men take over the unpaid work of child raising. To use this in prediction we would need some way of knowing in advance about the shifting preferences for work and income versus the satisfactions of motherhood. Will the uptrend in the former continue over the next decade and so maintain the downtrend in births?

The difficulty of answering this question suggests that the mechanism in question is an ad hoc explanation of the past, to be retained as long as women increasingly enter the labor force and the birth rate falls, but to be quickly dropped and replaced by some other mechanism as soon as births turn upward. One can imagine, after births do start to rise, articles showing how inevitable is the reassertion of the durable values of motherhood as opposed to ephemeral economic interest. Whether such explanations are true or false, if they come after the fact they are too late to forecast turning points.

In this difficult situation it is natural to resort to asking women what their childbearing intentions are—just as people are queried about their house-buying intentions, and firms their investment plans. Such data help but only in short-term forecasting. Most of the children born more than 5 years from now will be to mothers who are presently still in their teens, unmarried, in no position to provide a realistic statement of their futures as mothers (Ryder and Westoff 1967, Siegel and Akers 1969). For the shorter term, intentions may predict well; Westoff, Potter, and Sagi (1963) found that about the best predictor of whether a couple with two children will have a third is their own statement.

Even for estimating turning points in the birth curve less than 5 years ahead, however, the statements of wives have not always been borne out. Women may declare an intention that accords with present rates of childbearing, but once the future becomes the present they are influenced by whatever the fashion is at the time. It is too much to say that current childbearing influences stated intentions more than intentions anticipate future childbearing, but some hint of this does appear in the time series of intentions, on the one hand, and performance, on the other.

We have some evidence (Masnick and McFalls 1976) that women's attitudes toward childbearing are formed (or manifest themselves) early in their married careers. If they start married life in a time of low fertility when contraception is practiced rigorously, they tend to continue to practice it. Such a fact, if the manner and degree of its operation could be established, would facilitate forecasting.

12.5.2 Shape of the Projection Fan

If we know the past, and must make estimates of the future, the population trajectory starts out as a single, somewhat jagged line, and at the moment where past and future meet fans out into a set of relatively smooth lines representing the several possibilities. Disregarding for the moment the errors, incompleteness, and delays in statistics on the past, the question is how the single line or curve, representing more or less certain past knowledge, fans out into the future to make the horn shape familiar in graphical representations of projection.

If the lines of the fan were straight, each would require only one parameter—say, the angle made with the time axis—for its full description. Such straight lines imply that uncertainty increases in equal increments with time from the present—that if we know in 1970 that the population for 1980 will be in a certain range with a certain probability, the population for 1990 will be in twice that range, both expressed in numbers of persons (Figure 12.4*a*).

If, on the other hand, one could say that for all future time the (fixed) rate of increase r would be within the range r_1 to r_2, the horn would take the shape of exponentials as in Figure 12.4*b*. Drawn on semilog paper, this would be back to the form of Figure 12.4*a*.

Only slightly more difficult is thinking that the future rate of increase would be $r(t)$ at time t, drawn from a specified probability distribution. Since, as in (1.6.1), population at time t is $N_t = N_0 \exp[\int_0^t r(\tau)\, d\tau]$, what is operative is the sum of $r(t)$. If the increases at the several times in the future are drawn independently at random, the variance of $\int_0^t r(\tau)\, d\tau$ is proportional to the time t, the standard deviation to $\sqrt{t}$. The range in this case would be a curve like $e^{\sqrt{t}}$, in suitable units. Or under other circumstances the curve might be of the form $\exp(\sqrt{\sqrt{t}})$ (Figure 12.4*c*). A priori the possibilities are unlimited.

But now recall that we can predict survivorship better than births. As time moves on, more and more of the population will have been born since the jumping-off point, and a smaller and smaller fraction will consist of survivors from that point. As a larger fraction passes into the group known less well, the fan widens more rapidly than is shown in any of the three diagrams of Figure 12.4.

In symbols, suppose that at a given future time f of the population will have been born since the jumping-off point and will be known with uncertainty represented by the standard deviation σ, and $1 - f$ of the population with uncertainty $k\sigma$, where $k < 1$. Then the accuracy with

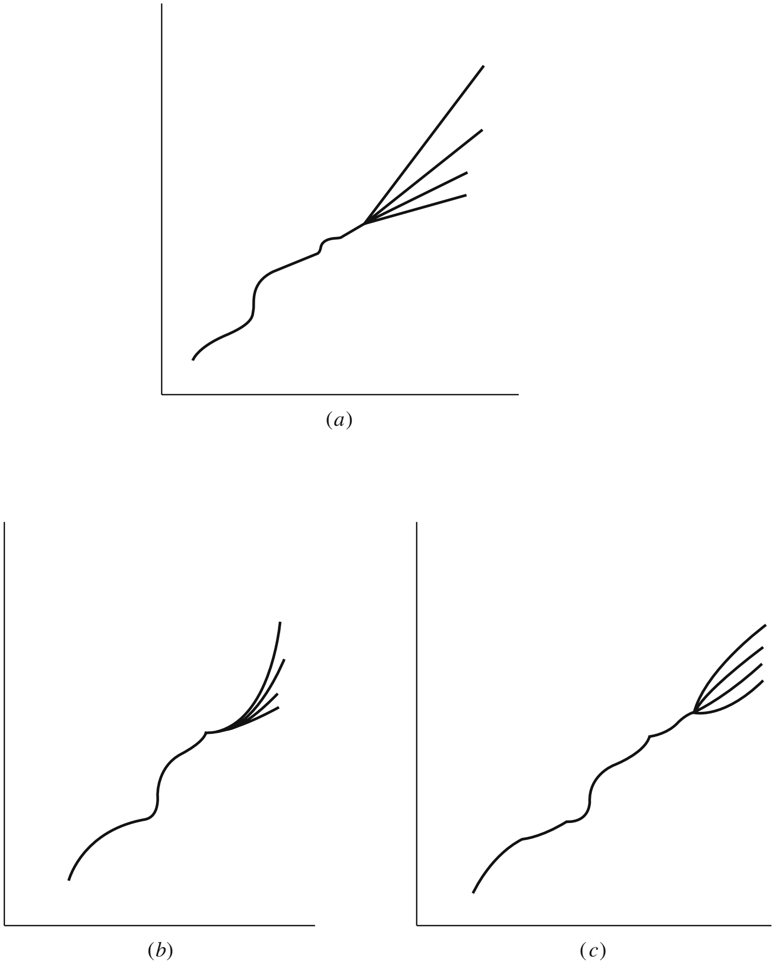


Figure 12.4. Shape of the fan under various conditions. (*a*) Ignorance of population increasing with time in equal increments. (*b*) Ignorance of (fixed) rate of increase. (*c*) Future subject to diminishing variation.

which the whole population is known has the standard deviation

$$\sigma^* = \sigma\sqrt{f^2 + (1-f)^2k^2}.$$

The spread can be expressed directly in terms of survivors and births. For *ex ante* evaluation the population t years hence may be written as

$$N_t = \int_0^t B(\tau)l(t-\tau)\,d\tau + \int_0^\omega p(a)\frac{l(a+t)}{l(a)}\,da + I_t - E_t,$$

where

$B(\tau) =$ the births at time τ,
$l(t-\tau) =$ the probability of survival of a birth to age $t-\tau$, that is, to time t,
$p(a)\,da =$ the number of persons age a to $a+da$ at the jumping-off point where $t=0$,
$I_t =$ the cumulative number of immigrants by time t, as projected from immigration statistics by a survivorship function,
$E_t =$ the cumulative number of emigrants, also as projected to time t.

The quantities in the equation are subject to different amounts of error—current population $p(a)$ is best known, prospective survivorship $l(a{+}t)/l(a)$ next best, future births $B(\tau)$ only poorly. Emigrants present more difficulties than immigrants in current statistics as well as in the future. Separate estimates need to be made of the precision of each of these components. As time goes on from the jumping-off point, the dependence of the forecast on future births increases, and hence the precision decreases. The fan of projections will correspondingly open out more widely than in straight lines or exponentials.

These considerations apply to projections using matrix population models with arbitrary stage classification. If the projection matrix $\mathbf{A}_t$ at time t is determined by an stationary ergodic stochastic process, and if the matrices are sufficiently well behaved (a stochastic analogue of the primitivity condition for convergence), then the log of population size is asymptotically normally distributed with a variance that increases linearly with time at a rate we can denote by σ^2 (Furstenberg and Kesten 1960, Tuljapurkar and Orzack 1980). Thus the projection fan in this case also increases as $\exp(\sigma\sqrt{t})$, but matrix methods are required to estimate σ^2 (see Tuljapurkar 1990 and MPM Chapter 14).

These results are asymptotic, and apply in stationary stochastic environments, whereas projections are often short-term and in conditions where the statistical properties of the vital rates are not constant. For an approach to such cases, see Lee and Tuljapurkar (1994).

12.6 The Components of Forecasting Error *Ex Ante*

The errors that arise in estimates of the future are classifiable under five headings. Arrangement is in ascending order of difficulty of estimation, and probably also of magnitude.

1. There is random variation from the deterministic model on which the estimate was made. Corresponding to the deterministic model used for forecasting, one can construct many stochastic models, all with expected values equal to the forecast. The most obvious of these stochastic models allows the same probability to act on the several individuals independently.

   In application to a single item, say births numbering B, the coefficient of variation (standard deviation divided by expectation) can be calculated as a function of the numbers involved, for example the number of births B. For large closed populations, assuming statistical independence between individuals gives a coefficient of variation equal to $1/\sqrt{B}$. When births are in millions, the error is trifling; when even as few as 10,000 births the coefficient of variation is only 1 percent. This component of forecasting error is important only for populations as small as a neighborhood or a village.

   This type of variation is called *demographic stochasticity*, to distinguish it from *environmental stochasticity* caused by stochastic temporal fluctuations in the vital rates. Including demographic stochasticity in matrix population models leads to a multi-type branching process, an extension of the model described in Section 16.4. The connection between matrix population models and multi-type branching processes was first studied by Pollard (1966, 1973) for age-classified models, and was generalized to stage-classified models in Chapter 15 of MPM.

2. The constants in the models are not known exactly. The projection may suppose the age-specific rates of birth and death, first differences of these rates over time, ultimate family size, or something else to be fixed into the future. If whatever is taken as constant really is constant but is improperly measured or assessed, the future will to that extent be incorrectly estimated.

   The extent of this uncertainty can be evaluated by associating confidence intervals with projections, based on the sampling distributions of the parameters on which they are based. This was once difficult or impossible, but the development of bootstrap resampling methods (Efron and Tibshirani 1993) now make it a fairly straightforward task in even complicated matrix population models (MPM Chapter 12). Estimation errors may be amplified in a forecast, but they still add little to the other, deeper errors inherent in forecasting and considered below.

3. The quantities supposed constant, which constitute the parameters of the model, really vary in time. Only a finite past record, a short sample of time periods, is available for estimating their paths. Often the sample of time periods is a single past year or decade. It is

hardly a properly drawn sample, yet the forecaster has no other way of estimating the future values of the needed parameters.

4. Probably more important than any of the foregoing, the model itself is incorrectly specified. The main fault is not the bad estimates of parameters, but an altogether inappropriate function supposed to hold for all time. For example, the period age-specific rates of birth are held constant, when what is really constant is their first differences, or else not period rates at all but the average number of children born to a cohort. The finite past experience that can be brought to bear has little power to discriminate among models, as can be seen by experimenting with the logistic and explosion models of Section 1.7. It is not easy to compel the past to inform us which of several models that fit indifferently well will give the best forecast, yet the futures given by these models can be very different.
   Recent statistical developments, combining information theory and maximum likelihood estimation, make it possible to objectively assign weights to different competing models, and take account of the uncertainty of model specification in projections (see Burnham and Anderson 2002 for a particularly clear account, with many applications to population biology). This approach is rapidly becoming standard in some branches of population ecology, and has great potential for human demography as well.

5. Finally, even perfect use of exact facts regarding a homogeneous past may be frustrated by the future being genuinely different. The uncertainty about the future is superimposed on projections of past data at the disposal of the demographer, and such uncertainty does not lend itself to estimation in advance. In some countries and epochs this fifth component of error will be small, but the twentieth century is not such an epoch. This component is wholly distinct from sampling error and is the characteristic special difficulty of prediction.

Of the five components of error the first four are, at least in principle, accessible to statistical analysis, and the fifth is not accessible by any means. A rough technique may be devised for handling the fourth component or, more strictly, the first four combined. This provides a lower bound for the overall error *ex ante*, that is to say, made at the same time as the estimate itself. This is contrasted below with the error *ex post*, calculable after the event predicted, the predictand is known, and which does include all five sources of error.

In the course of estimating the future many separate decisions have to be made, and none of these follows uniquely from any accepted principles. Extrapolation of death rates can proceed from a shorter or longer experience; it can be done separately for the several age–sex groups or for all groups together, supposing that mortality will improve according to the pattern of a series of model tables. If each time such a decision is made one or more other reasonable possibilities are also followed through, a large number of combinations of these alternatives will be available. Extrapolation from short versus long experience of deaths, and from separate ages versus all ages together following a series of model tables, and similarly for births, alone gives 16 combinations and therefore 16 forecasts. The variation among these provides an *ex ante* estimate of error.

Alternatively, if the calculation is more intuitive than these mechanical extrapolations, and a 2 percent per year decline in fertility is proposed, the estimator can ask himself whether 1 percent and 3 percent are also reasonable, or whether he would bet 19 to 1 odds that the range is limited to $1\frac{1}{2}$ to $2\frac{1}{2}$ percent. Similarly in regard to mortality and migration. The aggregation of these sources of error provides upper and lower bounds for the future population. The aggregation may be based on the assumption that the several errors act independently, that is to say, their squares are added, and at the end the square root taken, or more conservatively the errors may themselves be added.

12.6.1 The Length of the Experience Base

Every forecast works from some period in the past. As a minimum it takes the population of an initial year as its jumping-off point and assumes that the age-specific rates of birth, death, and migration of the initial year remain constant into the future. To derive a trend in such age-specific rates requires at least 2 past years; better estimates of trend can be obtained from 5, 10, or more years, using least squares or other fitting. The longer base will be of value also in providing an estimate of error of the fit. If the past were homogeneous, we could never have too much of it, and would incorporate the whole available statistical record in the forecast.

Since we always suspect heterogeneity in the past, under what circumstances is it desirable to use a long experience base and when is a short base to be preferred (Anderson 1971)? If sudden or sharp changes occur from time to time, or a gradual drift in relationships is taking place, the experience base should be confined to the most recent time and be relatively short, in the hope that the part of the past used is homogeneous with the future. A prediction for the United States today would not be greatly aided by what happened before World War I, even if we had good data for that long past time.

The forecaster who would provide a distribution needs data even more badly than the one content to produce a single figure. He is even more tempted to go back in time to increase his stock, and also more concerned about the danger that his results will be distorted by changes in the system the further back he goes.

To ascertain what part of the past record is relevant to the future being predicted is a subsidiary but not unimportant subject of investigation.

12.7 *Ex Post* Evaluation of Point Estimates

To check on the *ex ante* estimate of error of a particular forecasting approach, one can apply the approach to the past, noting, say, the differences between the population in 1970 as predicted in 1935 and the observed 1970 population now available. Experiments on the experience base should lead to improved estimates of future population and their variance.

After all justifiable use of the past has been made, concentrated on a point estimate for a future moment, and after the moment arrives and a census is taken, one would think that then, at least, we would be in a position to judge the estimate unambiguously. But further problems arise; surprisingly, even the possession of the actual population count, with which the forecast made earlier can be compared to provide an *ex post* assessment of its accuracy, leaves some residual difficulties. Statement of the accuracy of prediction attained is not unambiguous even after the event, at least if we want to compare two predictions made at different times for the same future date, and more generally to compare estimates over different time intervals and in different countries.

For all such purposes a benchmark of some kind is needed, an extension of the "persistence" (forecasting that tomorrow's weather will be the same as today's) that meteorologists use to see how well their elaborate methods have served, and that Mosteller et al. (1949, p. 297) have applied to election polls. The benchmark may be a naive estimate of the kind discussed above, a standard form of projection by age and sex with fixed rates, or some other, as will be suggested below.

The Scripps medium projection made in 1935 for the 1970 population of the conterminous United States was 155 million. The actuality as counted in the 1970 census was 203 million for the same area. Was 48 million a large error? The estimate was 76 percent of the true figure, or 24 percent short, but was this good or bad? If one could predict the national income 35 years in advance, within 24 percent, he would be doing well; on the other hand, 24 percent error in predicting next year's federal tax collections would be decidedly poor. The only way to assess the quality of a point forecast is to compare it with some standard way of forecasting.

If the quality of a prediction has plainly to be assessed in relation to some other prediction produced by a standard method that can be taken as a base or benchmark, it is much less obvious how this base should be selected. One benchmark corresponds to persistence in weather forecasting: assume that the population will not change from the 1935 level. The 1935 level was 127 million, so the true increase over the whole period from 1935 to 1970 was the difference between this and the realization: $203 - 127 = 76$ million. This was the error to which the use of the 1935 figure for 1970 would have been subject.

Somewhat more formally, the general expression for the quality of a forecast as measured *ex post* is

$$\text{Quality of forecast} = \frac{\text{Forecast} - \text{Benchmark}}{\text{Realization} - \text{Benchmark}}. \tag{12.7.1}$$

In effect, our first effort was to use a benchmark of zero, giving for 1970

$$\text{Quality} = \frac{155 - 0}{203 - 0} = 0.76, \tag{12.7.2}$$

the complement of which, 24 percent, is one way of describing the error.

The second effort described above, is to use the 1935 observed population of 127 million as a benchmark, and we find, again for 1970,

$$\text{Quality} = \frac{155 - 127}{203 - 127} = \frac{28}{76} = 0.37. \tag{12.7.3}$$

The more precise benchmark provides a more rigorous evaluation, as though we estimate not the 1970 population but the increase from 1935.

We now proceed to a yet tighter measure of quality. The forecast will not be credited with the fraction of the increase from the jumping-off point that it includes, but only the part of it not covered by a projection at 1935 rates of birth and death. Such a projection from 1935 gives 148 million for 1970, so we have

$$\begin{aligned} \text{Quality} &= \frac{\text{Forecast} - \text{Projection at rates of 1935}}{\text{Realization} - \text{Projection at rates of 1935}} \\ &= \frac{155 - 148}{203 - 148} = \frac{7}{55} = 0.13. \end{aligned} \tag{12.7.4}$$

This last measure of quality in effect recognizes that whoever made the forecasts in 1935 knew the 1935 population, birth rates, and death rates. He could have projected with these rates, but used his knowledge and intuition to improve on them. We will credit him with good judgment only to the degree that he bettered that fixed-rate forecast.

The same expression (12.7.1) can be applied to assess 1950 and 1960 as well as 1970, and Table 12.6 shows the results. Note that (12.7.2) seems to give worse results as one proceeds in time, and so does (12.7.3). Only (12.7.4) constitutes a measure of quality that is virtually the same for 1950, 1960, and 1970.

Population 10 to 30 years ahead is better predicted than income, demand for housing, interest rates, unemployment, prices, technology, weather, earthquakes, or any other common subject of forecasting. But we cannot take much satisfaction in this superiority. Our sophisticated techniques are effective only in the degree to which they produce better results than naive methods. Thus one direction of evaluation of population projections is how they compare with simple straight-line extrapolation and other naive methods (Agnew, 1972).

Table 12.6. Three retrospective or *ex post* evaluations of the 1935 Scripps Medium Projection to 1950, 1960, and 1970

Year	Forecast as percent of census count (8.7.2)	Estimated increase from 1935 as percent of census increase (8.7.3)	Departure from fixed-rate benchmark as percent of census departure (8.7.4)
1950	94	61	12
1960	84	43	12
1970	76	37	13

Table 12.7. Annual rate of future increase implied by official United States projections (percents)

Period	C	D	E	F
1972–1980	1.26	1.13	0.88	0.76
1972–1990	1.35	1.19	0.92	0.75
1972–2000	1.30	1.12	0.84	0.65

Source: *Statistical Abstract of the United States* (1973, p. 6.)

The benchmark is a simple-minded, conventional, stylized standard. It is intended to judge how good refined methods are, and to compare one refined method with another. The refined methods can use intuitive elements that do not have to be justified; the benchmark to judge them must be entirely objective, so that it can be constructed after the event, the predictand, is known. If the benchmark is made too good, the forecasts being judged by it look too bad. Some correspondents have argued that the benchmark ought to be nothing but the actual population at the jumping-off point, but this seems to let the forecaster off too lightly, since anyone can improve on such a benchmark. In principle the benchmark should incorporate the minimum knowledge that the forecaster should have taken into account, and so it records how much more than the minimum skill he possessed.

12.7.1 Future Percentage Increase

To regard projections in terms of the percentage increase they imply accords with the wider range of uncertainty in prospective population as we go forward in time. Table 12.7 shows the angle of climb of the curve expressed in terms of the average percentage increase from 1972, implied by the four official projections, C, D, E, and F, for the United States.

The differences between 1972–80, 1972–90, and 1972–2000 are small compared with the differences between projections C, D, E, and F, and one can summarize the whole table by its bottom line, which says that the projected rate of increase is between 0.65 and 1.30 percent. For the increase of the whole period the C estimate comes out just double the F estimate.

12.8 A Division of Labor

Who ought to estimate future population? Is it true that in a free-enterprise society the risks of decision are taken by the entrepreneur, that these risks include the danger of being wrong regarding the future population, and that if we leave things alone there will be a natural selection of entrepreneurs favoring those who have, among other talents, a superior ability to forecast population (or who can hire better demographers)? In this scheme there is no room for official forecasting of population.

Yet such an austere view has never been held in regard to knowledge in general, whose officially sponsored diffusion is, rather, part of the infrastructure on which a free-enterprise system functions. If there is some information regarding the future that can serve everyone and improve many decisions, then let it be assembled, published, and paid for by the public authorities. This is so notwithstanding the fact that mistakes are made in forecasting the future; the test is whether forecasts add more to the quality of decisions than is spent in making them.

Among other difficulties, all users do not want the same forecast. One user needs a conservative estimate of the future—a figure that she is reasonably sure will be exceeded. Another needs a high figure; he wants to guard especially against any possibility that the estimate he uses will be exceeded by the realization. How can we separate the task of forecasting from the adaptation of the forecast to a particular use? The loss function does this in a precise way.

12.8.1 The Loss Function Permits a Three-Way Division of Labor

Only the user can know how much he stands to lose through a projection being wrong in one direction or the other, or by a certain amount. His loss function may be strongly asymmetric. Suppose that the future population of a town is wanted for deciding the capacity of a new water reservoir. If the reservoir turns out to be too small, another will soon have to be built, and this will be expensive. If it turns out to be too large, it will be underutilized, and the loss is only the marginal cost of the excess capacity, which may be small.

The logic for arriving at exactly the right prediction $\hat{x}$ for a particular use is based on developments in statistics going back to the 1930s, and in its application to demography is set forth by Muhsam (1956). Suppose that the user has studied his own application enough to be able to judge how much loss he would suffer with a particular departure of the estimate $\hat{x}$ from the true population x (not knowable until the future date has arrived), and suppose that he can express this in the form of a loss function $L(\hat{x}, x)$ or, less generally but more simply, $L(\hat{x} - x)$.

He is given the population forecast in the form of a probability distribution $P(x)$ by the official agency mentioned above. Not being able to control x, the user integrates it out to provide his total expected loss:

$$R(\hat{x}) = \int_{-\infty}^{+\infty} L(\hat{x} - x)P(x)\,dx. \tag{12.8.1}$$

It remains only to choose $\hat{x}$ so as to minimize the total expected loss $R(\hat{x})$; the value of $\hat{x}$ required is the solution of the equation $dR(\hat{x})/d\hat{x} = 0$, provided that the functions $P(x)$ and $L(\hat{x} - x)$ are well behaved.

For a highly simplified numerical example, suppose the official estimate to be discrete and to have the following form:

Forecast x persons	Probability $P(x)$
50,000	0.25
60,000	0.50
70,000	0.25

Now suppose that the user's loss function is as follows:

Forecast error $\hat{x} - x$	Loss $L(\hat{x} - x)$
−20,000	$2,000,000
−10,000	500,000
0	0
10,000	100,000
20,000	200,000

Thus, if he uses $\hat{x} = 50,000$ and the performance turns out to be $x = 70,000$, he loses $2,000,000, and so on.

We can make a two-way table of loss, $L(x - \hat{x})$:

$\hat{x}$	x 50,000	60,000	70,000
50,000	0	500,000	2,000,000
60,000	100,000	0	500,000
70,000	200,000	100,000	0

The table can be summarized by the sum $\sum P(x)L(x-\hat{x})$ for each possible forecast $\hat{x}$:

Forecast $\hat{x}$	Expected loss $\sum P(x)L(x-\hat{x})$
50,000	\$750,000
60,000	150,000
70,000	100,000

The right forecast for the particular purpose is 70,000.

12.9 Interval Estimates as Currently Provided

Ex post we know in what fraction of cases the upper and lower projections straddle the subsequent performance. If a collection of projections is in any sense homogeneous, and the method currently in use is similar to the techniques employed in the collection, the fraction of instances in which the subsequent census was straddled constitutes a level of confidence for the present method. A large collection of early projections is available to be compared with subsequent actuality.

Such material may be used to test various hypotheses about the process of projection as it has been practiced in the past. One of these is that, after a steady movement of births and population in a consistent direction (such as occurred from the 1870s to the 1930s), confidence increases, and the range of admitted future possibilities narrows—the projection fan is made smaller. When a sudden change (especially in the birth series) creates uncertainty the fan is made very large. A second hypothesis is that series that are abandoned may ultimately turn out unexpectedly well, perhaps better than those that are continued. Third, the level of births anticipated may be too sensitive to current birth rates; one could test the hypothesis that waves in future population estimates correspond to waves in current births.

12.9.1 Official Agencies Have Backed into Confidence Intervals

What has happened as a result of the insistence on neutral projections rather than forecasts, and on making several so that the user can choose among them, is that the agencies involved have backed into confidence intervals. It seems unlikely that any user does what the theory of projections says she should do—examine the assumptions of the several sets, given more or less completely in fine print in the introduction, use personal judgment as to which is right, and only then look at the corresponding projection and adopt it for use. A user frequently picks the series that seems to give the right answer, and pays only secondary attention to the fine print describing the assumptions. Insofar as she does this, the whole theory of projections is subverted, for she could just as well choose among a set of random numbers.

But users do want the government estimates; they may pick a middle figure from the assortment presented, or a high or low figure according to their need, and suppose that the assumptions made by the official demographers represent the range of reasonable possibilities. This is to say that they are using the published high and low series as defining a confidence interval. Like the producer of the series, they do not formally attach a probability to the range, but even brief experience shows them that there is a finite chance that the performance can fall outside the range given.

In the 1970s, an informal estimate based on 30 cases suggested that the confidence attachable to responsible official estimates was about two-thirds. It seems likely that some such number has implicitly been playing the role in projection that 0.95 plays in tests of statistical significance. Because the instruments are much cruder, no one is interested in 95 percent limits for forecasts; they are apparently too wide for planning purposes. But much more research is needed, especially *ex post* studies that look back on earlier projections and note what fraction turns out to straddle the population performance.

The need for projections steadily increases, and the demand bears no relation to accuracy or inaccuracy as demonstrated *ex post.* The Bureau of the Census will always be required to make projections, independently of how good these projections are in an absolute sense: the least bad forecasts available are what people seek. It can be criticized, not for failing to predict what subsequently occurs, but only for failure to use sound methods and to take account of all relevant data (cf. Lee and Tuljapurkar 1994, Bongaarts and Bulatao 2000).

13

Perturbation Analysis of Matrix Models

The results of a demographic analysis are functions of the vital rates, and through them of biological and environmental variables. For many reasons, the vital rates might have been otherwise, or might be otherwise in the future, and the results of the analysis would be different. This chapter introduces methods for perturbation analysis—asking what would happen to some dependent variable if one or more independent variables were to change. Perturbation analysis is used in several contexts.

- Predicting the results of future changes in the vital rates. These could result from natural changes in the environment, be imposed as a management plan to protect an endangered species or control a pest, or result from family planning or changes in healthcare.
- Accounting for past changes. Given data from several populations that differ because of their environments, one may want to know how much of the resulting change in, say, λ can be attributed to changes in survival, how much to changes in reproduction, and so forth.
- Predicting the action of natural selection. Selection deals with perturbations; it favors changes in the phenotype that increase fitness. Fitness is a demographic concept (the rate at which a genotype propagates itself), so an essential part of the theory of natural selection is understanding how changes in the phenotype affect λ.
- Designing sampling schemes. Errors in estimates of the vital rates produce errors in the estimate of anything calculated from those rates, for example, λ. The most important errors will be those in the rates

to which λ is most sensitive; this information can be used to design sampling procedures that focus on estimating the vital rates where accuracy matters most.

A Note on Examples.

In this chapter we will use examples to show how calculations work and to explore patterns that arise from different kinds of life cycles. Most of the examples are based on the following matrices, which include age-classified, stage-classified, and size-classified life cycles.

- The age-classified matrix for the human population of the United States in 1965, given in Example 7.2.
- The stage-classified model for teasel (*Dipsacus sylvestris*). The life cycle is shown in Figure 11.2 and the matrix is given in Example 11.3. This is an example of a model in which, potentially, transitions are possible between any two stages.
- A size-classified matrix for the desert tortoise (*Gopherus agassizii*), a threatened species living in the southwestern U.S. (Doak et al. 1994). The matrix follows the standard size-classified model (3.8.2), with seven size classes:

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1.30 & 1.98 & 2.57 \\ 0.716 & 0.567 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.149 & 0.567 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.149 & 0.604 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.235 & 0.560 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.225 & 0.678 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.249 & 0.851 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.016 & 0.860 \end{pmatrix}. \tag{13.0.1}$$

  This is an example of a model in which the stages have a natural ordering and individuals may stay in their stage or move to the next stage; no other transitions are allowed.

  The entries in this matrix are expressed in terms of lower-level vital rates. Let $P_i = a_{ii}$ be the probability of surviving and staying in the same size class (*stasis*), and $G_i = a_{i+1,i}$ be the probability of surviving and growing to the next size class (*growth*). Then

$$P_i = \sigma_i (1 - \gamma_i) \tag{13.0.2}$$

$$G_i = \sigma_i \gamma_i, \tag{13.0.3}$$

  where σ_i is the survival probability and γ_i the growth probability of size class i (Caswell 1985).

13.1 Eigenvalue Sensitivity

13.1.1 Perturbations of Matrix Elements

We begin with the sensitivity of an eigenvalue λ to changes in the matrix elements a_{ij}. The functional dependence of λ on the a_{ij} is expressed implicitly in the characteristic equation

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0.$$

When the form of the characteristic equation is known, implicit differentiation can be used to find $\partial\lambda/\partial a_{ij}$ (Hamilton 1966, Demetrius 1969, Emlen 1970, Goodman 1971, Keyfitz 1971a, Mertz 1971a). We will use this approach in Section 13.1.6, but it is limited to cases in which the characteristic equation is sufficiently simple that implicit differentiation yields a simple solution.

A general approach, applicable to matrices of any structure, was introduced by Caswell (1978). The formula dates back at least to Jacobi (1846), and has been rediscovered many times (e.g., Faddeev 1959, Faddeev and Faddeeva 1963, Desoer 1967). For a complete and rigorous discussion, see Kato (1982). Cohen (1978, p. 186) presents an alternative approach to eigenvalue sensitivity.

We begin with the equations defining the eigenvalues and the right and left eigenvectors:

$$\mathbf{A}\mathbf{w}_i = \lambda_i \mathbf{w}_i \tag{13.1.1}$$

$$\mathbf{v}_i^* \mathbf{A} = \lambda_i \mathbf{v}_i^*, \tag{13.1.2}$$

where $\mathbf{v}_i^*$ is the complex conjugate transpose of $\mathbf{v}_i$. For the moment, let us suppress the subscript i; the following formulae apply to any of the eigenvalues and their corresponding right and left eigenvectors. Taking the differential of both sides of (13.1.1) yields

$$\mathbf{A}(d\mathbf{w}) + (d\mathbf{A})\mathbf{w} = \lambda(d\mathbf{w}) + (d\lambda)\mathbf{w}, \tag{13.1.3}$$

where $d\mathbf{A} = (da_{ij})$ is a matrix whose elements are the differentials da_{ij}.

Form the scalar product* of both sides with the left eigenvector $\mathbf{v}$,

$$\langle \mathbf{A}(d\mathbf{w}), \mathbf{v}\rangle + \langle (d\mathbf{A})\mathbf{w}, \mathbf{v}\rangle = \lambda\langle (d\mathbf{w}), \mathbf{v}\rangle + \langle (d\lambda)\mathbf{w}, \mathbf{v}\rangle. \tag{13.1.4}$$

Expanding the scalar products and cancelling terms leaves

$$d\lambda = \frac{\langle (d\mathbf{A})\mathbf{w}, \mathbf{v}\rangle}{\langle \mathbf{w}, \mathbf{v}\rangle} \tag{13.1.5}$$

$$= \frac{\mathbf{v}^* d\mathbf{A}\mathbf{w}}{\mathbf{v}^*\mathbf{w}}. \tag{13.1.6}$$

---

*The scalar product of the vectors $\mathbf{x}$ and $\mathbf{y}$ is $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{y}^*\mathbf{x}$.

Suppose that only one element, a_{ij}, is changed, while all the others are held constant. Then $d\mathbf{A}$ contains only one nonzero entry, da_{ij}, in row i and column j, and (13.1.6) reduces to

$$d\lambda = \frac{\bar{v}_i w_j \, da_{ij}}{\langle \mathbf{w}, \mathbf{v} \rangle}, \tag{13.1.7}$$

where $\bar{v}_i$ is the complex conjugate of v_i. Dividing both sides by da_{ij} and rewriting the differentials as partial derivatives (since all but one of the variables of which λ is a function are being held constant), we get the formula[†]

$$\frac{\partial \lambda}{\partial a_{ij}} = \frac{\bar{v}_i w_j}{\langle \mathbf{w}, \mathbf{v} \rangle}. \tag{13.1.8}$$

That is, the sensitivity of λ to changes in a_{ij} is proportional to the product of the ith element of the reproductive value vector and the jth element of the stable stage distribution. The scalar product term in the denominator is independent of i and j, and can be ignored when considering the relative sensitivities of λ to different elements in the same matrix. Or, the eigenvectors may be scaled so that $\langle \mathbf{w}, \mathbf{v} \rangle = 1$, so that the term can be ignored.

It is easy to calculate a *sensitivity matrix* **S** whose entries give the sensitivities of λ to the corresponding entries of **A**:

$$\begin{aligned} \mathbf{S} &= \left(\frac{\partial \lambda}{\partial a_{ij}} \right) && (13.1.9) \\ &= \frac{\bar{\mathbf{v}} \mathbf{w}^{\mathsf{T}}}{\langle \mathbf{w}, \mathbf{v} \rangle}. && (13.1.10) \end{aligned}$$

The sensitivity is the local slope of λ, considered as a function of a_{ij}.

13.1.2 Sensitivity and Age

The sensitivity of λ to changes in age-specific survival and fertility plays an important role in the theory of senescence (Medawar 1952, Williams 1957, Hamilton 1966, Rose 1984, Wachter and Finch 1997, Carey 2003). Figure 13.1 shows these sensitivities for four populations (a flour beetle, a vole, a whale, and humans). Sensitivities of λ to fertility decline nearly exponentially with age. At early ages, λ is more sensitive to changes in survival than to changes in fertility. At older ages, the pattern is reversed. The sensitivities vary by as much as eight orders of magnitude.

[†] This formula is often written with v_i in place of $\bar{v}_i$. This makes no difference in the analysis of λ_1, since $\mathbf{v}_1$ is real, but could cause errors in computing sensitivities of other eigenvalues, and of eigenvectors.

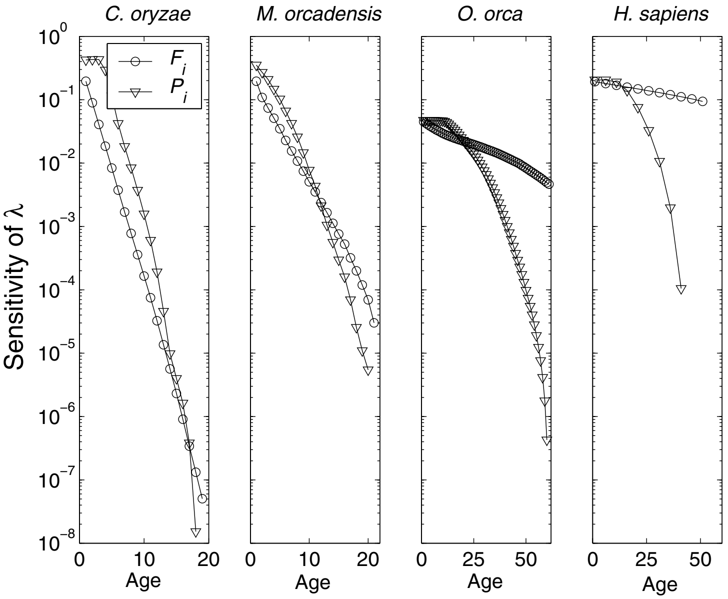


Figure 13.1. The sensitivity of λ to changes in age-specific fertility F_i and survival probability P_i for (from left to right) a laboratory population of the flour beetle *Calandra oryzae* (Birch 1948), a laboratory population of the vole *Microtus orcadensis* (Leslie et al. 1955), the killer whale *Orcinus orca* (Olesiuk et al. 1990), and the human population of the United States in 1965 (Keyfitz and Flieger 1968).

Some of these properties can be derived directly from the sensitivity formula. The eigenvectors **w** and **v** for age-classified populations are

$$w_1 = 1 \tag{13.1.11}$$

$$w_i = P_1 P_2 \cdots P_{i-1} \lambda^{-i+1} \qquad \text{for } i > 1, \tag{13.1.12}$$

and

$$v_1 = 1 \tag{13.1.13}$$

$$v_i = F_i \lambda^{-1} + P_i \lambda^{-1} v_{i+1} \qquad \text{for } i > 1. \tag{13.1.14}$$

We can use **w** and **v** to show how the sensitivities change with age (Demetrius 1969, Caswell 1978, 1982c). From (13.1.8) and (13.1.12) it follows that

$$\frac{\partial\lambda/\partial F_j}{\partial\lambda/\partial F_{j+1}} = \frac{w_j}{w_{j+1}} \tag{13.1.15}$$

$$= \frac{\lambda}{P_j}. \tag{13.1.16}$$

Thus the sensitivity of λ to fertility is a strictly decreasing function of age as long as $\lambda > 1$. If the P_j are constant, then the decrease will be exponential, as is approximately true in Figure 13.1. Other things being equal, the sensitivity of λ to changes in fertility falls off with age more rapidly the greater the value of λ. In a decreasing population, however, the sensitivity will actually increase from age j to $j+1$ if $\lambda < P_j$.

The sensitivities of λ to changes in survival at successive ages satisfy

$$\begin{aligned}
\frac{\partial\lambda/\partial P_j}{\partial\lambda/\partial P_{j+1}} &= \frac{w_j v_{j+1}}{w_{j+1} v_{j+2}} && (13.1.17)\\
&= \frac{\lambda}{P_j}\left(\frac{F_{j+1}\lambda^{-1} + P_{j+1}v_{j+2}}{v_{j+2}}\right) && (13.1.18)\\
&= \lambda\frac{P_{j+1}}{P_j} + \frac{F_{j+1}}{P_j v_{j+2}} && (13.1.19)\\
&\geq \frac{P_{j+1}}{P_j} \quad \text{if } \lambda \geq 1. && (13.1.20)
\end{aligned}$$

Thus the sensitivity of λ to survival decreases monotonically with age provided $\lambda \geq 1$ and $P_{j+1} \geq P_j$. If survival were age-independent, then the sensitivity of λ to changes in survival would decrease monotonically with age provided $\lambda > 1$.

The relative sensitivities of λ to fertility and survival satisfy

$$\begin{aligned}
\frac{\partial\lambda/\partial P_j}{\partial\lambda/\partial F_j} &= \frac{v_{j+1}w_j}{v_1 w_j} && (13.1.21)\\
&= \frac{v_{j+1}}{v_1}. && (13.1.22)
\end{aligned}$$

Thus λ is more sensitive to survival than to fertility if $v_{j+1} > v_1$, which is true at least up to the age of first reproduction. In Figure 13.1, $\partial\lambda/\partial P_j > \partial\lambda/\partial F_j$ for young ages; the inequality is reversed at older ages.

13.1.3 Sensitivities in Stage- and Size-Classified Models

The patterns of sensitivity in size- or stage-classified models can be quite different from those in age-classified models. Because the stable size distribution may exhibit peaks (as is commonly observed in fish and tree size distributions), sensitivities need not be monotonic functions of size or stage.

Example 13.1 Size-specific sensitivity in the desert tortoise

> Figure 13.2 shows the sensitivity of λ to changes in the F_i, P_i, and G_i. None of the sensitivities shows a monotonic trend with size. Population growth rate is very sensitive to changes in the vital rates of size classes 2 and 3, and also size class 7.

Example 13.2 Sensitivity analysis of teasel

> In stage-classified models, there may be no variable, like size, against which sensitivities can be plotted. In this case, it is useful to display

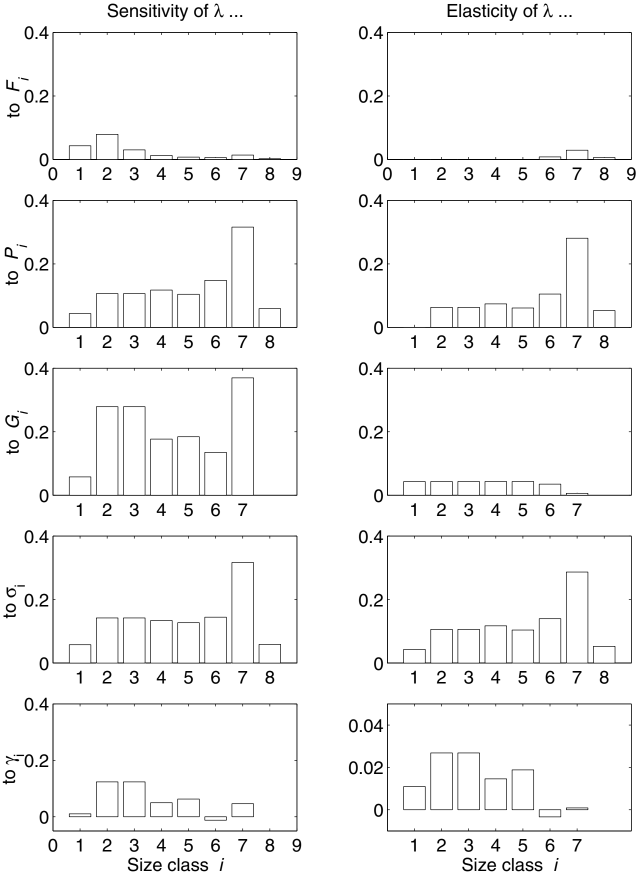


Figure 13.2. Perturbation analysis of population growth rate for the desert tortoise. Left column shows sensitivities, right column elasticities, of λ to F_i, P_i, G_i, σ_i, and γ_i. Calculated from data of Doak et al. (1994), medium-high fertility.

the entire sensitivity matrix $\mathbf{S}$ calculated from (13.1.10):

$$\mathbf{S} = \begin{pmatrix} 0.067 & 0.028 & 0.001 & 0.007 & 0.001 & 0.001 \\ 0.002 & 0.001 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.421 & 0.174 & 0.008 & 0.046 & 0.008 & 0.003 \\ 2.442 & 1.011 & 0.046 & 0.265 & 0.047 & 0.018 \\ 16.432 & 6.801 & 0.313 & 1.786 & 0.315 & 0.119 \\ 47.475 & 19.650 & 0.905 & 5.160 & 0.911 & 0.344 \end{pmatrix}. \quad (13.1.23)$$

The matrix can be displayed graphically as a surface plot or an image plot, or by plotting the entries of vec **S** (Figure 13.3).‡ These plots all show that λ is most sensitive to changes in entries in the lower-left corner of the matrix. These entries correspond to growth directly from seeds to large rosettes and flowering plants. Such growth doesn't happen in this species (i.e., these entries of **A** are all zero), but that does not change the fact that *if* those probabilities were changed, the impact on λ would be considerable.

13.1.4 What About Those Zeros?

The sensitivity (13.1.7) gives the effect on λ of changes in any entry of **A**, including those that may, in a given context, be regarded as fixed at zero or some other value. This is exactly as it should be. The derivative tells what *would* happen to λ *if* a_{ij} were to change, not whether, or in what direction, or how much, a_{ij} actually changes (Caswell et al. 2000). The results of such impossible perturbations may or may not be of interest, but they are *not* zero. It is up to you to decide whether they are useful and whether to display them.

For example, in the projection matrix for teasel, $a_{66} = 0$ because flowering plants neither survive nor produce new flowering plants in one year. But from the sensitivity matrix (13.1.23) we see that

$$\frac{\partial \lambda}{\partial a_{66}} = 0.344.$$

If a_{66} were changed by a small amount Δa_{66}, the resulting change in λ would be approximately $0.344 \Delta a_{66}$. This bit of information would be of no interest to someone concerned with environmental perturbations, but very interesting indeed to someone studying the evolution of annual and biennial life histories.

13.1.5 Total Derivatives and Multiple Perturbations

The sensitivity (13.1.8) is written as the partial derivative of λ with respect to a_{ij}, holding all other parameters constant. If more than one parameter is perturbed simultaneously, the net effect on λ is given by the differential

$$d\lambda = \sum_{k,l} \frac{\partial \lambda}{\partial a_{kl}} da_{kl}. \tag{13.1.24}$$

This result can be used in several ways.

---

‡ The vec operator produces a vector from a matrix by stacking the columns one above the other.

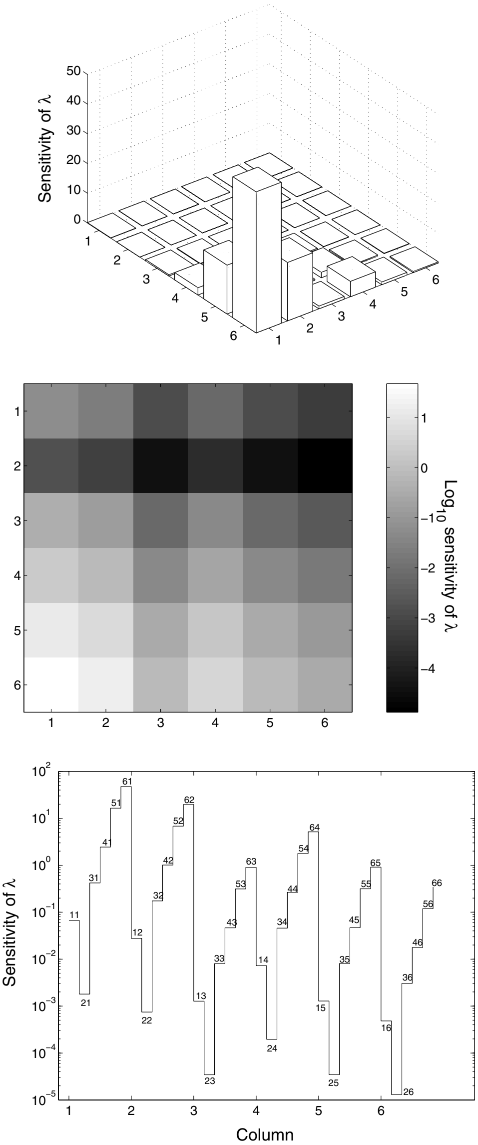


Figure 13.3. Three ways of displaying the sensitivity matrix **S** for teasel (Field A). The upper graph plots sensitivities as a surface, the middle graph as an image, and the lower graph as a plot of vec **S**. In the lower graph, the numbers are coordinates (i, j) of the entry of **S**.

1. Lower-level parameters. Suppose that one or more of the matrix entries are functions $a_{ij}(x)$ of some "lower-level" variable x (e.g., allocation of energy to reproduction, which affects both survival and fertility). Then

$$da_{kl} = \frac{\partial a_{kl}}{\partial x} dx. \tag{13.1.25}$$

Substituting this expression into (13.1.24) yields the chain rule

$$\frac{d\lambda}{dx} = \sum_{k,l} \frac{\partial \lambda}{\partial a_{kl}} \cdot \frac{\partial a_{kl}}{\partial x} \tag{13.1.26}$$

2. Physiological or genetic constraints. One vital rate may be functionally related to another, for many reasons. Increased fertility may result in decreased somatic growth, because of allocation of resources. Increased survival early in life may result in decreased survival later in life due to pleiotropic effects. We suppose that perturbing one vital rate a_{ij} will affect some or all of the other matrix entries, so that

$$da_{kl} = \frac{\partial a_{kl}}{\partial a_{ij}} da_{ij}. \tag{13.1.27}$$

Substituting into (13.1.24), we obtain

$$\frac{d\lambda}{da_{ij}} = \frac{\partial \lambda}{\partial a_{ij}} + \sum_{kl \neq ij} \frac{\partial \lambda}{\partial a_{kl}} \frac{\partial a_{kl}}{\partial a_{ij}}. \tag{13.1.28}$$

3. A special case of constraints arises when individuals face mutually exclusive options. Consider seeds that may either germinate or remain dormant. Clearly, $P[\text{germination}] = 1 - P[\text{dormancy}]$, and a change in one must produce a compensatory change in the other.

Example 13.3 Sensitivity to survival and growth in size-classified models

The parameters G_i and P_i in the standard size-classified model for the desert tortoise are functions of the survival probability (σ_i) and the growth probability (γ_i) of size class i. From (13.1.26) we have

$$\frac{\partial \lambda}{\partial \sigma_i} = \frac{\partial \lambda}{\partial G_i} \frac{\partial G_i}{\partial \sigma_i} + \frac{\partial \lambda}{\partial P_i} \frac{\partial P_i}{\partial \sigma_i} \tag{13.1.29}$$

$$= \frac{\partial \lambda}{\partial G_i} \gamma_i + \frac{\partial \lambda}{\partial P_i} (1 - \gamma_i) \tag{13.1.30}$$

$$= \frac{w_i (v_i + \gamma_i \Delta v_i)}{\langle \mathbf{w}, \mathbf{v} \rangle}, \tag{13.1.31}$$

where $\Delta v_i = v_{i+1} - v_i$ is the change in reproductive value from size class i to $i+1$. Similarly,

$$\frac{\partial \lambda}{\partial \gamma_i} = \frac{\partial \lambda}{\partial G_i}\frac{\partial G_i}{\partial \gamma_i} + \frac{\partial \lambda}{\partial P_i}\frac{\partial P_i}{\partial \gamma_i} \tag{13.1.32}$$

$$= \sigma_i \left(\frac{\partial \lambda}{\partial G_i} - \frac{\partial \lambda}{\partial P_i} \right) \tag{13.1.33}$$

$$= \frac{\sigma_i w_i \Delta v_i}{\langle \mathbf{w}, \mathbf{v} \rangle}. \tag{13.1.34}$$

The sensitivity of λ to growth rate (13.1.34) is negative if $v_{i+1} < v_i$. That is, λ is reduced by increasing the growth rate from $\mathcal{N}_i$ to $\mathcal{N}_{i+1}$ if stage $i+1$ has a lower reproductive value than stage i.

Applying (13.1.31) and (13.1.34) to the size-classified model for the desert tortoise yields the results shown in Figure 13.2. This untangles the survival and growth components of the sensitivities to P_i and G_i. Changes in the survival of stage 7 would have a major impact on λ. The sensitivity of λ to γ_6 is slightly negative because of a slight decline in reproductive value from stage 6 to stage 7.

In the desert tortoise, the fertilities F_i were considered independent of survival and growth. More detailed descriptions of the F_i, however, usually involve both the σ_i and γ_i. In such cases, the sensitivity of λ to the lower-level parameters must also include their effects on fertility.

13.1.6 Sensitivity to Changes in Development Rate

Population growth rate is sensitive to changes in the timing of events in the life cycle (e.g., Lewontin 1965, Mertz 1971, Caswell and Hastings 1980, Caswell 1982c, Hoogendyk and Estabrook 1984, Ebert 1985). In Section 6.3 we saw the effect on the intrinsic rate of increase $r = \log \lambda$ of changes in the mean μ and variance σ^2 of the net maternity function. Increasing μ corresponds to a delay in development, and an analysis using age-classified matrices yields the same conclusion as (6.3.5); i.e., that in an increasing population, delayed reproduction reduces population growth rate.

However, this analysis is based on the transformation of the characteristic equation into a cumulant generating function, and holds only for age-classified models. What can we say about slowing the rate of transition in a general stage-classified model?

To answer this question, we start with a transformation of a life cycle graph (Section 9.1.3). The coefficient on each arrow is multiplied by $\lambda^{-\alpha}$, where α is the number of projection intervals required for the transition. Once transformed, the graph can be simplified by multiplying the coefficients on pathways between any two stages (see MPM Chapter 7). Figure 13.4 shows such a graph, focusing on the transition from stage j to stage

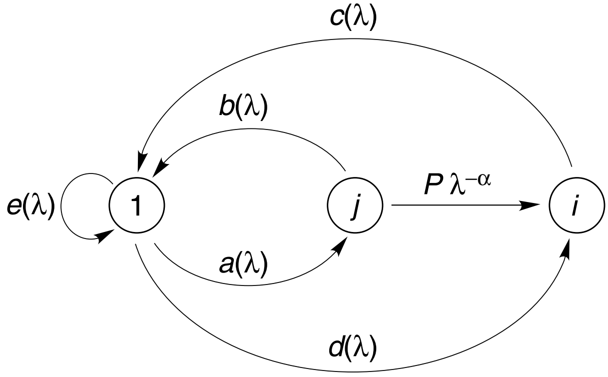


Figure 13.4. A life cycle graph for evaluating the sensitivity of λ to changes in α, the time required for the transition from stage j to stage i.

i, which requires α projection intervals to complete, with a probability P of surviving. The other arcs in this graph represent

- $a(\lambda)$ for all paths between $\mathcal{N}_1$ and $\mathcal{N}_j$ that do not pass through $\mathcal{N}_i$.
- $b(\lambda)$ for all paths between $\mathcal{N}_j$ and $\mathcal{N}_1$ that do not pass through $\mathcal{N}_i$.
- $c(\lambda)$ for all paths between $\mathcal{N}_i$ and $\mathcal{N}_1$
- $d(\lambda)$ for all paths from $\mathcal{N}_1$ to $\mathcal{N}_i$ which do not pass through $\mathcal{N}_j$
- $e(\lambda)$ all loops involving neither $\mathcal{N}_i$ nor $\mathcal{N}_j$.

We assume that there are no disjoint loops in the graph. The following method can be extended to life cycles containing such loops, but the results are more complicated. Since the time (α) required for transition from $\mathcal{N}_j$ to $\mathcal{N}_i$ appears explicitly in the transformed life cycle graph, it is possible to evaluate the sensitivity of λ to changes in development rate between these two stages.

The characteristic equation can be written down from the graph[§] as

$$1 = a(\lambda)c(\lambda)P\lambda^{-\alpha} + \Phi(\lambda) \tag{13.1.35}$$

$$= \mathcal{F}(\lambda), \tag{13.1.36}$$

where $\Phi(\lambda) = a(\lambda)b(\lambda) + c(\lambda)d(\lambda) + e(\lambda)$. Implicit differentiation of λ with respect to α gives

$$\frac{\partial\lambda}{\partial\alpha} = \frac{-\partial\mathcal{F}/\partial\alpha}{\partial\mathcal{F}/\partial\lambda}. \tag{13.1.37}$$

[§] By setting equal to 1 the product of the transmissions around all loops in the graph; see MPM Chapter 7.

The numerator of (13.1.37) depends on whether P is constant or depends on α. We consider these two cases separately.

Case 1: Constant P

In this case,

$$\begin{aligned}\frac{\partial \mathcal{F}}{\partial \alpha} &= -a(\lambda)c(\lambda)P\lambda^{-\alpha}\log(\lambda) \\ &= -\log\lambda\left(1-\Phi(\lambda)\right). \qquad (13.1.38)\end{aligned}$$

We calculate $\partial\mathcal{F}/\partial\lambda$ by rewriting the characteristic equation as

$$1 = \mathcal{F}(\lambda) = \sum_{i=1}^{\infty} k_i\lambda^{-i} \qquad (13.1.39)$$

from which

$$\begin{aligned}\frac{\partial \mathcal{F}}{\partial \lambda} &= -\lambda^{-1}\sum_{i=1}^{\infty} ik_i\lambda^{-i} \\ &= -\lambda^{-1}\bar{A}, \qquad (13.1.40)\end{aligned}$$

where $\bar{A}$ is the mean age of the parents in the stable population (see Section 11.3.5). Thus, from (13.1.37)

$$\frac{\partial\lambda}{\partial\alpha} = \frac{-\lambda\log(\lambda)\left(1-\Phi(\lambda)\right)}{\bar{A}}. \qquad (13.1.41)$$

From this, it follows immediately that $\partial\lambda/\partial\alpha < 0$ whenever $\lambda > 1$; slowing development anywhere in the life cycle of an increasing population reduces λ. The sensitivity of λ to changes in development rate is inversely proportional to generation time.

Case 2: $P = e^{-\mu\alpha}$

The probability of surviving from one stage to another is unlikely to be independent of the time required for the transition. If the survival probability P represents the results of exposure for α time units to a mortality rate μ, then $P = \exp(-\mu\alpha)$, and the numerator of (13.1.37) becomes

$$-\frac{\partial\mathcal{F}}{\partial\alpha} = (\log\lambda + \mu)\left(1-\Phi(\lambda)\right). \qquad (13.1.42)$$

The denominator, $\partial f/\partial\lambda$ is still given by (13.1.40). Thus

$$\frac{\partial\lambda}{\partial\alpha} = \frac{-\lambda(\log\lambda+\mu)\left(1-\Phi(\lambda)\right)}{\bar{A}}. \qquad (13.1.43)$$

Comparing this with (13.1.41), it is apparent that making survival probability dependent on development rate makes it even more difficult for an increase in α to increase λ.

13.1.7 Predictions from Sensitivities

The sensitivities can be used to predict the response of λ to changes, of any size, in any or all of the parameters. Let Δa_{ij} be the perturbation (not necessarily small) of a_{ij} and $\Delta\lambda$ the resulting change in λ. From (13.1.24) we know that the total differential of λ is

$$d\lambda = \sum_{i,j} \frac{\partial \lambda}{\partial a_{ij}} \Delta a_{ij} \tag{13.1.44}$$

and

$$\Delta\lambda \approx d\lambda \tag{13.1.45}$$

with an error that goes to zero faster than Δa_{ij}. Hence the growth rate after the perturbation, $\lambda' = \lambda + \Delta\lambda$ is

$$\lambda' \approx \lambda + \sum_{i,j} \frac{\partial \lambda}{\partial a_{ij}} \Delta a_{ij}. \tag{13.1.46}$$

The approximation (13.1.46) can be made as accurate as desired by making the Δa_{ij} small enough. It would be exact, regardless of the size of Δa_{ij}, if λ were a linear function of the a_{ij}. It is not, but experience suggests that the nonlinearity in λ is not usually severe (e.g., Caswell 2000a, de Kroon et al. 2000, Levin et al. 1996). So the sensitivities can be used to explore scenarios involving even relatively large changes in the vital rates. Of course, if the perturbations are too large, or you need extreme accuracy, any perturbation scenario can always be simulated.

13.1.8 Another Interpretation of Reproductive Value

Our first interpretation of the left eigenvector $\mathbf{v}$ as a measure of reproductive value was in terms of the effects of initial population structure on asymptotic population size (Chapter 9). The sensitivity formula (13.1.8) provides another justification for equating reproductive value with $\mathbf{v}$. Consider a stage j that can contribute individuals to two other stages, 1 and 2. Increases in either a_{1j}or a_{2j} will increase λ, and a reasonable measure of the "value" of stages 1 and 2 is the relative sensitivity of λ to changes in a_{1j} and a_{2j}:

$$\begin{aligned}\frac{\partial\lambda/\partial a_{1j}}{\partial\lambda/\partial a_{2j}} &= \frac{v_1 w_j}{v_2 w_j} \\ &= \frac{v_1}{v_2}. \end{aligned} \tag{13.1.47}$$

That is, the relative sensitivity of λ to changes in a_{1j} and a_{2j} is given by the relative magnitudes of v_1 and v_2. An individual in stage j, confronted with the choice of contributing to stage 1 or stage 2 and desiring to increase

λ, should contribute to the stage with the higher value of v_i—that is, to the more valuable stage.

13.2 Elasticity Analysis

It is often useful to compare perturbations on a proportional scale (what is the effect of a 10 percent decrease in fertility compared to a 10 percent decrease in survival?).¶ The proportional response to a proportional perturbation is known as *elasticity* in economics (e.g., Hicks 1939, p. 205). Caswell et al. (1984) and de Kroon et al. (1986) introduced the calculation of the elasticity of the eigenvalues to changes in the vital rates.

The elasticity of λ with respect to a_{ij} is defined as

$$e_{ij} = \frac{a_{ij}}{\lambda}\frac{\partial \lambda}{\partial a_{ij}} \tag{13.2.1}$$

$$= \frac{\partial \log \lambda}{\partial \log a_{ij}}. \tag{13.2.2}$$

That is, the elasticity e_{ij} is just the slope of $\log \lambda$ plotted against $\log a_{ij}$. Since equal increments on a log scale correspond to equal proportions on an arithmetic scale, the elasticity measures proportional sensitivity. Elasticities can be calculated only for inherently nonnegative quantities like λ.

The elasticities can be conveniently calculated and displayed as an *elasticity matrix*

$$\mathbf{E} = \left(\frac{a_{ij}}{\lambda}\frac{\partial \lambda}{\partial a_{ij}} \right) \tag{13.2.3}$$

$$= \frac{1}{\lambda}\mathbf{S} \circ \mathbf{A}, \tag{13.2.4}$$

where $\mathbf{S}$ is the sensitivity matrix (13.1.10) and $\circ$ denotes the Hadamard (element-by-element) product.

13.2.1 Elasticity and Age

Figure 13.5 shows the elasticity of λ to survival and fertility for four age-classified populations. They are quite different from the sensitivity patterns (cf. Figure 13.1). The elasticity of λ to survival probability consistently declines with age, but the elasticity of λ to fertility first increases and then

---

¶This is of particular concern to biologists who work with organisms in which transition probabilities (which may not exceed 1) and fertilities (which may be much greater than 1) are measured on different scales.

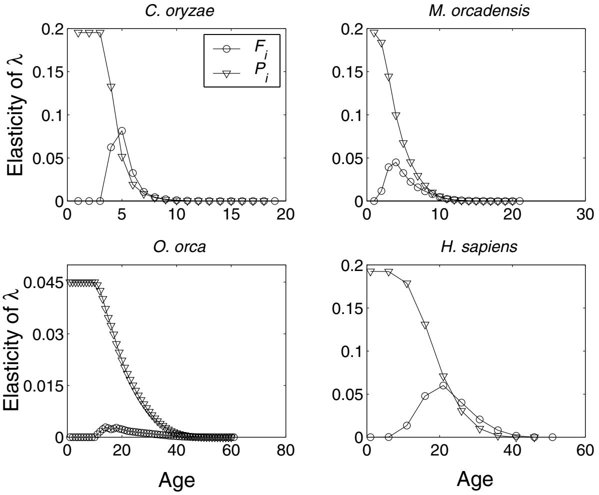


Figure 13.5. Elasticities of λ to changes in age-specific survival (P_i) and fertility (F_i) for the four populations whose sensitivities are shown in Figure 13.1.

declines. These are general properties of age-classified models (cf. Section 13.1.2). The elasticity to fertility at successive ages satisfies

$$\frac{e_{1j}}{e_{1,j+1}} = \left(\frac{F_j}{F_{j+1}}\right)\left(\frac{\lambda}{P_j}\right), \tag{13.2.5}$$

which need not be greater than 1. The elasticity to survival at successive ages satisfies

$$\frac{e_{j+1,j}}{e_{j+2,j+1}} = \frac{P_j}{P_{j+1}} \frac{\partial\lambda/\partial P_j}{\partial\lambda/\partial P_{j+1}} \tag{13.2.6}$$

$$= \lambda + \frac{F_{j+1}}{P_{j+1}v_{j+2}} \tag{13.2.7}$$

[cf. (13.1.19)]. This ratio is always greater than 1 when $\lambda \geq 1$.

Equation (13.2.7) implies that the elasticity of λ to survival is the same for all pre-reproductive age classes.

$$\frac{e_{j+1,j}}{e_{j+2,j+1}} = \left(\frac{P_j}{P_{j+1}}\right) \frac{v_{j+1}w_j}{v_{j+2}w_{j+1}} \tag{13.2.8}$$

and if $F_j = F_{j+1} = 0$, then $v_{j+1} = P_{j+1}\lambda^{-1}v_{j+2}$ and this ratio is equal to 1. See Heppell et al. (2000) for an application of this result to models based on limited demographic data.

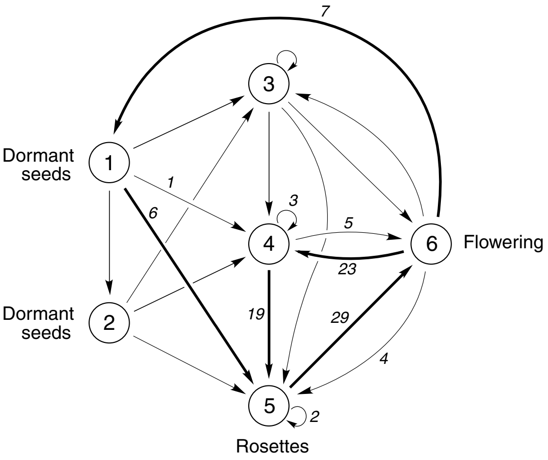


Figure 13.6. The contributions to λ, expressed as percentages, of the pathways in the teasel life cycle (Figure 11.2). Only contributions of more than 1 percent are shown. Heavy arrows indicate those pathways contributing 5 percent or more of λ.

13.2.2 Elasticities as Contributions to λ

The elasticities of λ with respect to the a_{ij} are often interpreted as the "contributions" of each of the a_{ij} to λ. This interpretation relies on the fact that the elasticities of λ with respect to the a_{ij} sum to 1 (De Kroon et al. 1986), a result which follows from Euler's theorem on homogeneous functions (Mesterton-Gibbons 1993).[‖]

An eigenvalue is a homogeneous function of degree one of the a_{ij}, because if $\mathbf{A}\mathbf{w} = \lambda\mathbf{w}$, then $c\mathbf{A}\mathbf{w} = c\lambda\mathbf{w}$; i.e., multiplying all the a_{ij} by c is equivalent to multiplying λ by c. Since λ is homogeneous of degree 1,

$$\sum_{ij} e_{ij} = \sum_{ij} \frac{a_{ij}}{\lambda} \frac{\partial \lambda}{\partial a_{ij}} = 1. \tag{13.2.11}$$

Thus, e_{ij} can be interpreted as the proportional contribution of a_{ij} to λ. This decomposition of λ is the only expression of the form

$$\lambda = \sum_{i,j} a_{ij} b_{ij},$$

[‖] A function $f(x_1, \ldots, x_n)$ is homogeneous of degree k if, for any real constant c,

$$f(cx_1, \ldots, cx_n) = c^k f(x_1, \ldots, x_n). \tag{13.2.9}$$

Euler's theorem states that if $f(x_1, \ldots, x_n)$ is homogeneous of degree k, then

$$x_1 \frac{\partial f}{\partial x_1} + \cdots + x_n \frac{\partial f}{\partial x_n} = kf(x_1, \ldots, x_n) \tag{13.2.10}$$

(e.g., Gillespie 1951).

where the contributions b_{ij} can be written as the product of one term that is a function only of i and another that is a function only of j (Caswell 1986).

The idea of a "contribution" must be interpreted carefully, since λ is not actually composed of independent contributions from each of the a_{ij}. Consider the model for the human population of the United States, in which the elasticity of λ to P_2 is 0.19 (Figure 13.5). We would say that survival of age-class 2 contributes about 20 percent of λ. But if we eliminate this transition, by setting $P_2 = 0$, we eliminate not 20 percent but 100 percent of λ. When $P_2 = 0$, no one survives to reproduce and $\lambda = 0$. Thus the "contributions" of the other vital rates to λ depend on the value of P_2, and vice versa.

Example 13.4 Elasticity in the desert tortoise

> Figure 13.2 shows the elasticities of λ to changes in F_i, P_i, and G_i for the desert tortoise (Doak et al. 1994). The total of the fertility elasticities is only 0.043 (i.e., 4.3 percent). The elasticities to growth (G_i) sum to 25.8 percent and those to stasis (P_i) sum to 69.9 percent. The largest elasticity is e_{77}; i.e., the elasticity of λ to the probability (P_7) of surviving and staying in size class 7.

Example 13.5 Elasticity in the teasel life cycle

> Figure 13.6 shows the largest elasticities in the teasel life cycle (cf. Figure 11.2). If the elasticities are interpreted as contributions, only five arcs, shown by heavy arrows, contribute more than 5 percent to λ. To a good approximation (73 percent of λ), the growth rate of teasel can be described in terms of only three transitions: [flowering plants → medium rosettes → large rosettes → flowering plants]. Adding the pathway [flowering plants → dormant seeds → large rosettes] adds an additional 13 percent of λ.

13.2.3 Elasticities of λ to Lower-Level Parameters

The elasticity of λ to a lower-level variable x is

$$\frac{x}{\lambda}\frac{\partial\lambda}{\partial x} = \frac{x}{\lambda}\sum_{i,j}\frac{\partial\lambda}{\partial a_{ij}}\frac{\partial a_{ij}}{\partial x}. \tag{13.2.12}$$

This gives the proportional change in λ resulting from a proportional change in x, but since there is no reason to expect that λ is a homogeneous function of x, the lower-level elasticities do not in general sum to 1, nor can they be interpreted as contributions to λ.

Example 13.6 Survival and growth in the desert tortoise

The elasticities of λ to changes in survival and growth in the desert tortoise can be obtained directly from (13.1.31) and (13.1.34):

$$\frac{\sigma_i}{\lambda}\frac{\partial\lambda}{\partial\sigma_i} = \frac{\sigma_i(1-\gamma_i)}{\lambda}\frac{\partial\lambda}{\partial P_i} + \frac{\sigma_i\gamma_i}{\lambda}\frac{\partial\lambda}{\partial G_i} \tag{13.2.13}$$

$$= \frac{P_i}{\lambda}\frac{\partial\lambda}{\partial P_i} + \frac{G_i}{\lambda}\frac{\partial\lambda}{\partial G_i} \tag{13.2.14}$$

(i.e., the elasticity of λ to σ_i is the sum of the elasticities to P_i and G_i), and

$$\frac{\gamma_i}{\lambda}\frac{\partial\lambda}{\partial\gamma_i} = \frac{\sigma_i\gamma_i}{\lambda}\left(\frac{\partial\lambda}{\partial G_i} - \frac{\partial\lambda}{\partial P_i}\right). \tag{13.2.15}$$

These results are shown in Figure 13.2.

13.3 Sensitivities of Eigenvectors

Perturbations of $\mathbf{A}$ change the eigenvectors as well as the eigenvalues. We have already seen some results on the sensitivity of the stable age distribution to changes in parameters (Sections 5.7 and 10.1). Here we derive some general sensitivity results (Caswell 1980, following the approach of Faddeev and Faddeeva 1963 and Desoer 1967).

Denote the eigenvalues, right eigenvectors, and left eigenvectors of $\mathbf{A}$ by λ_i, $\mathbf{w}_i$, and $\mathbf{v}_i$, respectively. We assume that the eigenvalues are distinct, and that the eigenvectors have been scaled so that $\langle \mathbf{w}_i, \mathbf{v}_i \rangle = 1$ and $\langle \mathbf{w}_i, \mathbf{v}_j \rangle = 0$ for $i \neq j$.

Suppose that we are interested in $\mathbf{w}_1$ and $\mathbf{v}_1$. As in the derivation of the eigenvalue sensitivities, we begin with

$$\mathbf{A}\mathbf{w}_1 = \lambda_1 \mathbf{w}_1. \tag{13.3.1}$$

Taking the differential of both sides yields

$$(d\mathbf{A})\mathbf{w}_1 + \mathbf{A}(d\mathbf{w}_1) = (d\lambda_1)\mathbf{w}_1 + \lambda_1(d\mathbf{w}_1). \tag{13.3.2}$$

If we consider $d\lambda_1$ as known, (13.3.2) is a linear equation in the differential $d\mathbf{w}_1$:

$$(\mathbf{A} - \lambda_1\mathbf{I})d\mathbf{w}_1 = (d\lambda_1\mathbf{I} - d\mathbf{A})\mathbf{w}_1. \tag{13.3.3}$$

Since $(\mathbf{A} - \lambda_1\mathbf{I})$ is singular, (13.3.3) cannot be solved directly for $d\mathbf{w}_1$. However, we can write any solution as a linear combination of the eigenvectors:

$$d\mathbf{w}_1 = \sum_{m=1}^{s} k_m \mathbf{w}_m \tag{13.3.4}$$

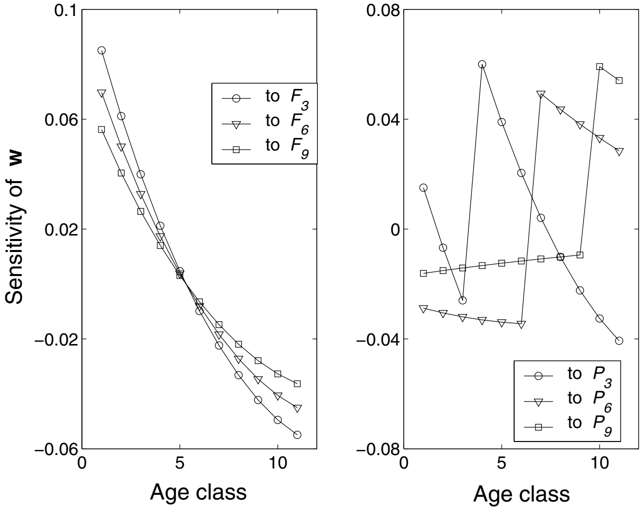


Figure 13.7. The sensitivity of the scaled stable age distribution **w** to changes in fertility (F_3, F_6, and F_9) and survival (P_3, P_6, and P_9) for the population of the United States. The vector **w** is scaled so that $\sum_i w_i = 1$.

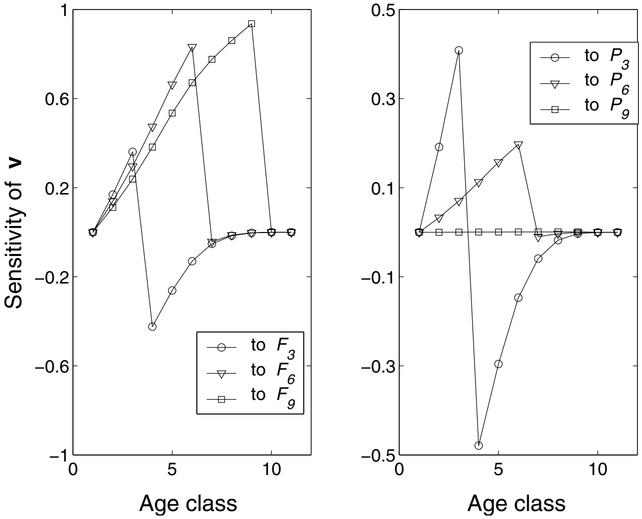


Figure 13.8. The sensitivity of the scaled reproductive value vector **v** to changes in fertility (F_3, F_6, and F_9) and survival (P_3, P_6, and P_9) for the population of the United States. The vector **v** is scaled so that $v_1 = 1$.

for some as yet unknown coefficients k_m. The value of k_1 is irrelevant, since when (13.3.4) is substituted in (13.3.3), $k_1(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{w}_1 = 0$, regardless of the value of k_1. In what follows, we will set $k_1 = 0$; this turns out to be the only reasonable assumption.

Leaving this expression for $d\mathbf{w}_1$ aside for the moment, form the scalar product of both sides of (13.3.2) with $\mathbf{v}_j$, for $j \neq 1$:

$$\langle (d\mathbf{A})\mathbf{w}_1, \mathbf{v}_j \rangle + \langle \mathbf{A} d\mathbf{w}_1, \mathbf{v}_j \rangle = d\lambda_1 \langle \mathbf{w}_1, \mathbf{v}_j \rangle + \lambda_1 \langle d\mathbf{w}_1, \mathbf{v}_j \rangle. \tag{13.3.5}$$

The second term on the left-hand side of (13.3.5) simplifies to $\lambda_j \langle d\mathbf{w}_1, \mathbf{v}_j \rangle$, and the first term on the right-hand side is zero if $j \neq 1$. Simplifying yields

$$\langle d\mathbf{w}_1, \mathbf{v}_j \rangle = \frac{\langle (d\mathbf{A})\mathbf{w}_1, \mathbf{v}_j \rangle}{\lambda_1 - \lambda_j}. \tag{13.3.6}$$

Now, substitute (13.3.4) for $d\mathbf{w}_1$ into (13.3.6) to obtain

$$\sum_{m \neq 1} k_m \langle \mathbf{w}_m, \mathbf{v}_j \rangle = \frac{\langle (d\mathbf{A})\mathbf{w}_1, \mathbf{v}_j \rangle}{\lambda_1 - \lambda_j}. \tag{13.3.7}$$

Since $\langle \mathbf{w}_m, \mathbf{v}_j \rangle = 0$ for $j \neq m$, and equals 1 for $j = m$, (13.3.7) simplifies to an expression for k_j:

$$k_j = \frac{\langle (d\mathbf{A})\mathbf{w}_1, \mathbf{v}_j \rangle}{\lambda_1 - \lambda_j}. \tag{13.3.8}$$

When (13.3.8) is substituted into (13.3.4), we obtain the desired expression for the differential of the right eigenvector $\mathbf{w}_1$:

$$d\mathbf{w}_1 = \sum_{m \neq 1}^{s} \frac{\langle (d\mathbf{A})\mathbf{w}_1, \mathbf{v}_m \rangle}{\lambda_1 - \lambda_m} \mathbf{w}_m. \tag{13.3.9}$$

The corresponding expression for the differential of the left eigenvector $\mathbf{v}_1$ follows from noting that the left eigenvectors of $\mathbf{A}$ are the right eigenvectors of $\mathbf{A}^*$, so that

$$d\mathbf{v}_1 = \sum_{m \neq 1}^{s} \frac{\langle (d\mathbf{A}^*)\mathbf{v}_1, \mathbf{w}_m \rangle}{\bar{\lambda}_1 - \bar{\lambda}_m} \mathbf{v}_m. \tag{13.3.10}$$

The partial derivatives of $\mathbf{w}_1$ and $\mathbf{v}_1$ follow from supposing that only a single entry, say a_{ij}, is perturbed. Using superscripts to distinguish the eigenvectors and subscripts to denote their elements (e.g., $w_j^{(m)}$ is the jth element of $\mathbf{w}_m$), the resulting expressions are

$$\begin{aligned} \frac{\partial \mathbf{w}_1}{\partial a_{ij}} &= \sum_{m \neq 1}^{s} \frac{w_j^{(1)} \bar{v}_i^{(m)}}{\lambda_1 - \lambda_m} \mathbf{w}_m \\ &= w_j^{(1)} \sum_{m \neq 1}^{s} \frac{\bar{v}_i^{(m)}}{\lambda_1 - \lambda_m} \mathbf{w}_m \end{aligned} \tag{13.3.11}$$

$$\frac{\partial \mathbf{v}_1}{\partial a_{ij}} = v_i^{(1)} \sum_{m\neq 1}^{s} \frac{\bar{w}_j^{(m)}}{\bar{\lambda}_1 - \bar{\lambda}_m} \mathbf{v}_m. \tag{13.3.12}$$

13.3.1 Sensitivities of Scaled Eigenvectors

The stable stage distribution is often scaled so that $\sum w_i = 1$ (to represent proportions) or $\sum w_i = 100$ (to represent percentages). The reproductive value vector is usually scaled so that $v_1 = 1$, to measure reproductive value relative to a newborn individual. The sensitivities of these scaled eigenvectors can be calculated using (13.3.11) and (13.3.12).

Let $||\mathbf{w}|| = \sum_i |w_i|$ and suppose the scaled stable stage distribution is given by $\mathbf{w}/||\mathbf{w}||$. Its sensitivity is then given by

$$\frac{\partial}{\partial a_{ij}} \frac{\mathbf{w}}{||\mathbf{w}||} = \frac{\dfrac{\partial \mathbf{w}}{\partial a_{ij}} ||\mathbf{w}|| - \mathbf{w} \displaystyle\sum_m \frac{\partial w_m}{\partial a_{ij}}}{||\mathbf{w}||^2}. \tag{13.3.13}$$

If the eigenvector whose sensitivity is being evaluated is already scaled so that $||\mathbf{w}|| = 1$, this simplifies to

$$\frac{\partial}{\partial a_{ij}} \frac{\mathbf{w}}{||\mathbf{w}||} = \frac{\partial \mathbf{w}}{\partial a_{ij}} - \mathbf{w} \sum_m \frac{\partial w_m}{\partial a_{ij}}, \tag{13.3.14}$$

where $\partial \mathbf{w}/\partial a_{ij}$ is given by (13.3.11).

Similarly, if the scaled reproductive value vector is defined as $\mathbf{v}/v_1$, and if the vector whose sensitivity is being evaluated is already scaled so that $v_1 = 1$, the sensitivity of the scaled reproductive value vector is

$$\frac{\partial}{\partial a_{ij}} \frac{\mathbf{v}}{v_1} = \frac{\partial \mathbf{v}}{\partial a_{ij}} - \mathbf{v} \frac{\partial v_1}{\partial a_{ij}}. \tag{13.3.15}$$

Example 13.7 An age-classified population

> The sensitivities of $\mathbf{w}$ and $\mathbf{v}$ for the U.S. population are shown in Figures 13.7 and 13.8. Increasing fertility shifts the age distribution to younger ages, because it increases λ (fertilities affect the stable age distribution only through λ). Changes in fertility at young ages have a greater effect on λ, and hence on $\mathbf{w}$, than do changes at later ages. An increase in survival at age i increases the representation of age class $i + 1$ in the stable age distribution. Since the age distribution is constrained to sum to 1, this produces a decrease in the representation of some other classes.
>
> Figure 13.8 shows the sensitivity of reproductive value to changes in fertility. Increasing F_i increases the v_i at all age classes up to and including i, and reduces v_i at later ages (because the values are all scaled relative to v_1). Changes in survival probability have qualitatively similar patterns. Thus, increasing survival probability

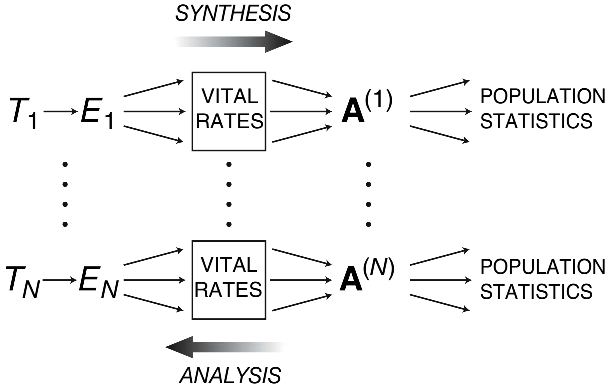


Figure 13.9. The design of a LTRE. Treatments (T_1–T_N) produce environmental conditions (E_1–E_N) that affect all the various vital rates. The vital rates are collected into population projection matrices $\mathbf{A}^{(1)}$–$\mathbf{A}^{(N)}$, from which a variety of population statistics can be calculated. From Caswell (1996a).

or fertility at a given age increases relative reproductive value at earlier ages, and decreases it at later ages.

13.4 Comparative Studies and Life Table Response Experiments

It is often of interest to compare the demography of two or more populations. Biologists often do so by imposing experimental manipulations and measuring the resulting vital rates. Because of this, the methods of this chapter are called *life table response experiments* (LTREs), but even in biological applications the word "experiment" must be understood broadly, including not only designed manipulative experiments but also comparative observations. It is still useful to think in terms of experiments, but to recognize that the treatments may be applied by nature rather than by the investigator.

In a LTRE, a life table (or more generally a set of vital rates) is the response variable in an experimental design or comparative study (Caswell 1989a, 1996a,b, 2000b). Treatments modify the environment and change the vital rates of individuals (Figure 13.9). The effects on the vital rates are usually diverse (affecting survival and reproduction and growth, sometimes in different directions) and stage-specific. Demographic models synthesize these effects into statistics that quantify the treatment effects at the population level. Population growth rate λ is the most frequently used statistic and the one focused on here, but others can be used.

The first true LTRE was Birch's (1953) study of the effects of temperature, moisture, and food on three species of flour beetles. The approach has become widely used in studies of chronic exposure to toxic substances; see Levin et al. (1996) for a recent example and van Straalen and Kammenga (1998) for a review with additional references.

The relation between λ and treatment shows how the treatments affect population growth, but it obscures the cause of those effects. Suppose that λ has been reduced; is it because mortality was increased, or growth impaired, or reproduction limited? Are these causes all equally responsible for the effect on λ, or can parts of that effect be attributed to each of them?

To answer these questions requires a decomposition of the treatment effect on λ into contributions from each stage-specific vital rate. This decomposition pinpoints the vital rates responsible for the population level effect of the treatment. It was introduced by Caswell (1989); several methodological extensions have appeared since (Brault and Caswell 1993, Caswell 1996a,b, 2000b). Applications include Levin et al. (1987, 1996), Levin and Huggett (1990), Walls et al. (1991), Silva et al. (1991), Canales et al. (1994), Brault and Caswell (1993), Caswell and Kaye (2001), Hansen (1997), Horvitz et al. (1997), and Ripley (1998).

LTREs can be classified by their design, in analogy to analysis of variance:

1. Fixed designs: the treatments imposed (by the experimenter or by nature) are of interest in themselves. Examples might include levels of toxicant exposure or food supply.
    (a) One-way designs: comparison of two or more levels of a single treatment factor.
    (b) Factorial designs: two or more levels of each of two or more treatment factors applied in all possible combinations.
2. Random designs: The treatments are a random sample from some distribution of treatment levels. Examples might include quadrats randomly distributed within a region (thereby sampling microhabitat variability), or a sequence of years (randomly sampling climatic conditions). It is often difficult to decide if a factor is fixed or random. One way to decide is to ask if you would use the same levels if you were to repeat the experiment. The answer is probably yes in the case of toxicant levels in a laboratory bioassay (a fixed factor) and no in the case of quadrats randomly located within the forest (a random factor). Random designs come in one-way, factorial, and nested varieties; some of these are only beginning to be explored (Caswell and Dixon in prep.).
3. Regression designs: The treatments represent levels of some quantitative factor (e.g., concentration of pesticide), and the goal is to explore the functional dependence of λ on the factor.

Notation alert. We use superscripts in parentheses to denote treatments, and subscripts to denote matrix elements. Thus $\mathbf{A}^{(i)}$ is the projection matrix obtained under treatment i, $\lambda^{(i)}$ is its dominant eigenvalue, and $a_{kl}^{(i)}$ the (k,l) entry of $\mathbf{A}^{(i)}$. Means are denoted by replacing a superscript by a dot; e.g.,

$$\mathbf{A}^{(\cdot)} = \frac{1}{m} \sum_{j}^{m} \mathbf{A}^{(j)}, \tag{13.4.1}$$

where m is the number of levels of the treatment.

13.5 Fixed Designs

The approach to a fixed design LTRE is to write a linear model for λ as a function of the treatments, and use the sensitivities of λ to obtain the coefficients in the linear model.

13.5.1 One-Way Designs

Consider a one-way design with treatments $T_1, \ldots, T_N$ producing population growth rates $\lambda^{(1)}, \ldots, \lambda^{(N)}$. Choose a *reference matrix* $\mathbf{A}^{(r)}$ as a baseline against which to measure treatment effects. $\mathbf{A}^{(r)}$ might be the mean matrix $\mathbf{A}^{(\cdot)} = \frac{1}{N} \sum_i \mathbf{A}^{(i)}$, or the matrix for a particular level of the treatment, often a "control."

Expanding λ, as a function of the a_{ij}, around $\mathbf{A}^{(r)}$ gives the population growth rate in treatment m as

$$\lambda^{(m)} \approx \lambda^{(r)} + \sum_{i,j} \left(a_{ij}^{(m)} - a_{ij}^{(r)} \right) \left. \frac{\partial \lambda}{\partial a_{ij}} \right|_{\mathbf{A}^{\dagger}} \qquad m = 1, \ldots, N, \tag{13.5.1}$$

where

$$\mathbf{A}^{\dagger} = \left(\mathbf{A}^{(m)} + \mathbf{A}^{(r)} \right) / 2. \tag{13.5.2}$$

The terms in the summation in (13.5.1) are the contributions of the a_{ij} to the effect of treatment m on population growth.

The sensitivities in (13.5.1) are evaluated at a matrix $\mathbf{A}^{\dagger}$ which is "midway" between the two matrices, $\mathbf{A}^{(m)}$ and $\mathbf{A}^{(r)}$, being compared. This is not absolutely essential, and other matrices, such as $\mathbf{A}^{(m)}$ or $\mathbf{A}^{(r)}$, could be used instead. However, using $\mathbf{A}^{\dagger}$ includes some information on the curvature of λ as a function of the a_{ij}. The mean value theorem of calculus guarantees that, for each treatment m, there is a matrix *somewhere* between $\mathbf{A}^{(m)}$ and $\mathbf{A}^{(r)}$ that will make the approximation (13.5.1) exact. Since it is halfway between, $\mathbf{A}^{\dagger}$ has a good chance of being close to this matrix. Using $\mathbf{A}^{\dagger}$ has been shown to give good results, and Logofet and Lesnaya (1997) have shown that it provides, in a sense, the best approximation.

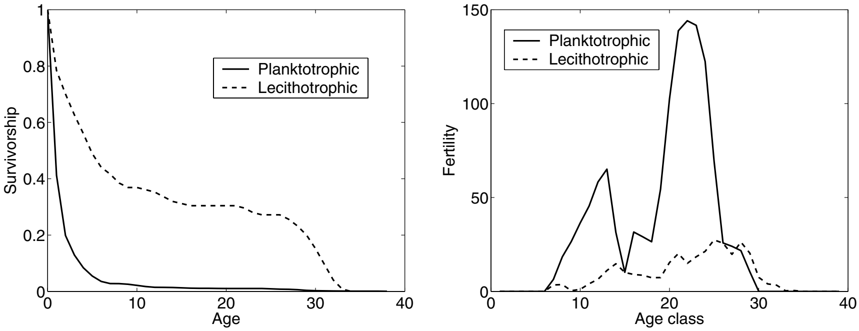


Figure 13.10. Survivorship and fertility for the planktotrophic and lecithotrophic strains of the polychaete *Streblospio benedicti*. From Levin et al. (1987).

These calculations can easily be implemented by defining a matrix of differences

$$\mathbf{D}^{(m)} = \mathbf{A}^{(m)} - \mathbf{A}^{(r)} \qquad m = 1, \ldots, N$$

and then calculating a matrix of contributions as

$$\mathbf{C}^{(m)} = \mathbf{D}^{(m)} \circ \mathbf{S}_{\mathbf{A}^\dagger}, \tag{13.5.3}$$

where $\circ$ is the Hadamard (element-by-element) product.

Example 13.8 Larval development mode in *Streblospio benedicti*

Streblospio benedicti is a marine polycheate worm that is capable of explosive population growth. It reproduces by a planktonic larval stage. Two genetic strains exist; in one the larvae are equipped with yolk and do not feed (lecithotrophic larvae), in the other the larvae are not provisioned with yolk, and must feed in the plankton (planktotrophic larvae); see Levin et al. (1987). Because they invest so much more in each offspring, the lecithotrophic strain has lower fertility, but the offspring have higher survival probability (Figure 13.10). We want to know how these life history differences contribute to differences in λ.

In laboratory experiments, Levin et al. (1987) measured projection matrices $\mathbf{A}^{(l)}$ for lecithotrophs and $\mathbf{A}^{(p)}$ for planktotrophs, with rates of increase $\lambda^{(l)} = 1.319$ and $\lambda^{(p)} = 1.205$.

We choose the lecithotrophic matrix $\mathbf{A}^{(l)}$ as the reference matrix. Figure 13.11 shows the differences in age-specific fertility F_i and survival probability P_i; the planktonic strain has a huge fertility advantage, especially from ages 15–25, and a survival disadvantage early in life. However, the large fertility differences between 20 and 30 weeks of age make almost no contribution to the difference in λ. Indeed, all but a very small proportion of the effect on λ is contributed by fertility and survival effects occurring before 15 weeks of age. Adding the

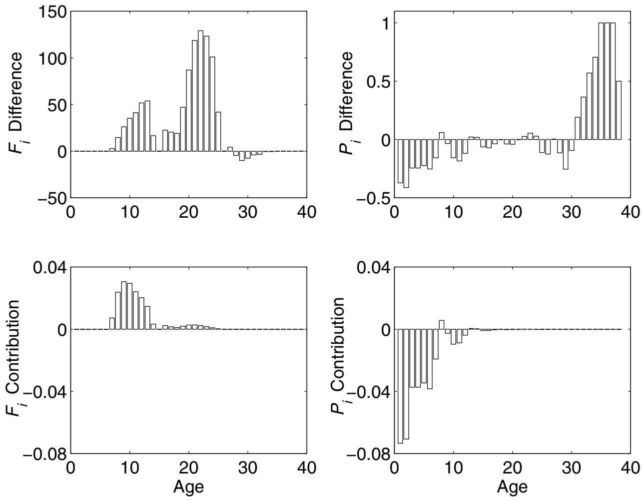


Figure 13.11. (Top row) The differences in age-specific fertility (F_i) and survival (P_i) between the planktotrophic and lecithotrophic strains of *Streblospio benedicti.* (Bottom row) The contributions of those differences to the effect of larval development mode on λ.

fertility contributions (which total 0.1710) and the survival contributions (which total -0.3315) shows that, while the survival differences appear less dramatic than the fertility differences, they actually contribute more to the effect on λ. This analysis quantifies the intuitive notion of a “trade-off” between increased fertility and reduced survival in comparison of larval development modes.

Even though developmental mode has large effects on the vital rates and on λ in this experiment, the first-order approximation in (13.5.1) is very accurate. It predicts

$$\begin{aligned}\lambda^{(p)} &= \lambda^{(l)} + \sum_{i,j} c_{ij} \\ &= 1.1590,\end{aligned}$$

where c_{ij} is the contribution of a_{ij}. This is within 4 percent of the actual value of $\lambda^{(p)}$.

13.6 Random Designs and Variance Decomposition

In a random design, the results are characterized by the variance in λ, and the goal of the analysis is to decompose this variance into contributions from the variances in (and covariances among) the matrix entries (Brault and Caswell 1993, Horvitz et al. 1997).

Let $V(\lambda)$ denote the variance in λ among treatments. To first order, $V(\lambda)$ can be written

$$V(\lambda) \approx \sum_{ij} \sum_{kl} C(ij, kl) s_{ij} s_{kl}, \tag{13.6.1}$$

where $C(ij, kl)$ is the covariance of a_{ij} and a_{kl}, and the sensitivities s_{ij} and s_{kl} are evaluated at the mean matrix. Each term in the summation is the contribution of one vital rate covariance to $V(\lambda)$. Unless there is good reason to believe that the vital rates are independent, the covariance terms in this calculation should not be neglected.

Suppose that there are n stages, and that the mean matrix is $\bar{\mathbf{A}}$. The contributions can be calculated easily by defining a $n^2 \times n^2$ covariance matrix

$$\mathbf{C} = E\left(\text{vec}\,(\mathbf{A})\text{vec}\,(\mathbf{A})^{\mathsf{T}}\right) - \text{vec}\,(\bar{\mathbf{A}})\text{vec}\,(\bar{\mathbf{A}})^{\mathsf{T}} \tag{13.6.2}$$

and then computing a $n^2 \times n^2$ matrix of contributions

$$\mathbf{V} = \mathbf{C} \circ \left(\text{vec}\,(\mathbf{S})\text{vec}\,(\mathbf{S})^{\mathsf{T}}\right). \tag{13.6.3}$$

Contributions to $V(\lambda)$ can also be calculated in terms of lower-level parameters. Suppose that the a_{ij} are defined in terms of stage-specific growth probabilities γ_i, survival probabilities σ_i, and reproductive outputs m_i, for $i = 1, \ldots, n$. Define a parameter vector

$$\mathbf{p} = (\sigma_1, \cdots, \sigma_n, \gamma_1, \cdots, \gamma_{n-1}, m_1, \cdots, m_n).$$

In terms of $\mathbf{p}$, the variance $V(\lambda)$ is

$$V(\lambda) \approx \sum_{ij} \text{Cov}\,(p_i, p_j)\, \frac{\partial \lambda}{\partial p_i} \frac{\partial \lambda}{\partial p_j}, \tag{13.6.4}$$

where the sensitivities are evaluated by applying (13.1.26) to the matrix calculated from the mean of the parameters.

Example 13.9 Interpod variance in λ in killer whales

Killer whales (*Orcinus orca*) live in stable social groups called pods. Brault and Caswell (1993) developed stage-classified models for each of the 18 pods of resident killer whales in the coastal waters of Washington state and British Columbia (Examples 9.1 and 11.1). The model (Figure 3.10) included four stages: yearlings, juvenile females, mature females, and senescent females, with a projection matrix

$$\mathbf{A} = \begin{pmatrix} 0 & F_2 & F_3 & 0 \\ G_1 & P_2 & 0 & 0 \\ 0 & G_2 & P_3 & 0 \\ 0 & 0 & G_3 & P_4 \end{pmatrix}.$$

Pod-specific population growth rates ranged from $\lambda = 0.9949$ to $\lambda = 1.0498$, with a variance $V(\lambda) = 2.90 \times 10^{-4}$.

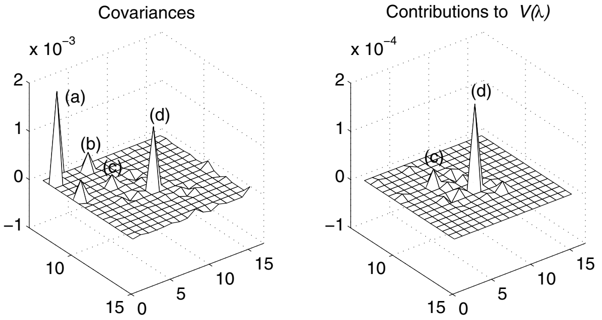


Figure 13.12. (Left) The interpod covariances of the matrix entries a_{ij} for the killer whale. (Right) The contributions of the covariances to $V(\lambda)$. The matrix entries are listed in the order produced by $\operatorname{vec}(\mathbf{A})$. Letters identify the most conspicuous peaks: (a) variance in G_1, (b) covariance of G_1 and G_2, (c) variance in P_2, and (d) variance in F_3. From data of Brault and Caswell (1993).

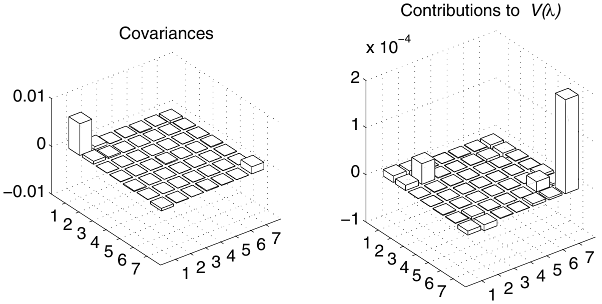


Figure 13.13. The covariances among the lower-level parameters and their contributions to $V(\lambda)$ for killer whales. Parameters 1–4 are σ_1–σ_4, 5–6 are γ_2–γ_3, and 7 is m. Calculated from data of Brault and Caswell (1993).

Figure 13.12 shows the covariance matrix $\mathbf{C}$ and the contribution matrix $\mathbf{V}$ for this population. The largest variance is that of G_1 (yearling survival), but this makes no detectable contribution to $V(\lambda)$. The largest contribution to $V(\lambda)$ is from variance in F_3 (adult fertility). There is a large covariance between G_1 and G_2, but it makes almost no contribution to $V(\lambda)$.

Using the lower-level parameters

$$\mathbf{p} = \left(\begin{array}{ccccccc} \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 & \gamma_2 & \gamma_3 & m \end{array} \right)$$

(γ_1 does not appear because it is always 1) gives a clearer picture of the determinants of $V(\lambda)$. Figure 13.13 shows the covariances among these parameters and their contributions to $V(\lambda)$. The largest contributions come from $V(m)$, $V(\sigma_2)$, and $V(\gamma_3)$, in that order.

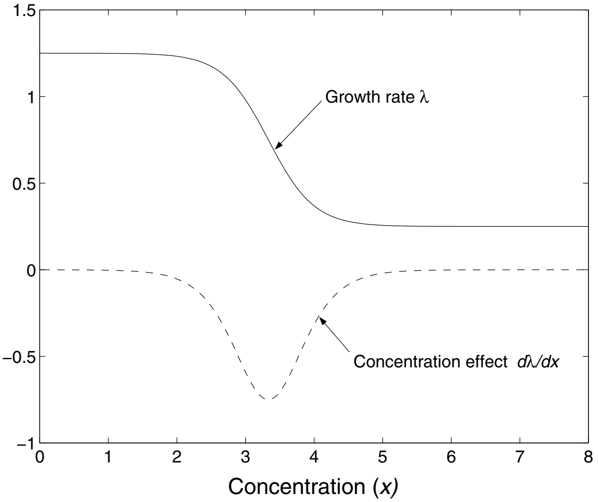


Figure 13.14. A hypothetical example showing λ as a function of concentration (x) and the effect of x, measured as the rate of change of λ with respect to x.

13.7 Regression Designs

The goal of a regression design is to describe the response of λ to the level of some quantitative factor x. In an ordinary linear regression analysis, the "effect" of the independent variable x on the dependent variable y is measured by the slope of the regression line. Accordingly, we define the effect of x on λ as the slope $\partial\lambda/\partial x$. Unless λ is a linear function of x (and there is no reason to expect that it will be) the slope will vary with x. Figure 13.14 shows a hypothetical example, in which λ is measured as a function of the concentration x of some substance. The substance has little effect at low concentrations, large negative effects at intermediate concentrations (from $x = 2$ to $x = 4$), and little or no effect at high concentrations. This intuitive interpretation is captured by using the slope of the function as the measure of effect.

The data for a regression LTRE consist of a set of vital rates $a_{ij}(x)$ that are functions of the treatment variable x. These vital rates generate a set of matrices $\mathbf{A}^{(x)}$ and population growth rates $\lambda(x)$. The effect of x on λ can be decomposed into contributions from each of the vital rates using the chain rule

$$\frac{d\lambda}{dx} = \sum_{i,j} \frac{\partial\lambda}{\partial a_{ij}(x)} \frac{\partial a_{ij}(x)}{\partial x}. \tag{13.7.1}$$

This expression is exact, not an approximation. The derivatives $\partial\lambda/\partial a_{ij}(x)$ are calculated from the matrix $\mathbf{A}^{(x)}$. The vital rate sensitivity $\partial a_{ij}(x)/\partial x$ comes from the functional relationship between the vital rates and x. This relationship can be described in various ways: interpolation of observed values, fitting parametric relationships (linear or nonlinear), or nonparametric smoothing (Caswell 1996a).

The terms in the summation in (13.7.1) are the contributions of each of the vital rates to the treatment effect on λ at a specific value of x. The contribution can be positive or negative, depending on the sign of $\partial a_{ij}(x)/\partial x$. The contribution of $a_{ij}(x)$ will be small if $a_{ij}(x)$ is not very sensitive to x, or if λ is not very sensitive to $a_{ij}(x)$, or both.

These contributions can be summed in various ways, to describe the contributions of different groups of vital rates. Summing over all fertilities, for example, would give an integrated contribution of effects on reproduction:

$$\sum_j \frac{\partial\lambda}{\partial F_j(x)} \frac{\partial F_j(x)}{\partial x}. \tag{13.7.2}$$

Summing the contributions of all the a_{ij} gives an approximation to the derivative of λ with respect to x. Integrating this derivative then gives an estimate of $\lambda(x)$ that can be compared with the observed response to see how well the functions $a_{ij}(x)$ capture the response of the vital rates to x. For an example of this analysis applied to a toxicological experiment, see Caswell (1996a).

13.8 Prospective and Retrospective Analyses

This chapter has avoided discussing the "importance" of the vital rates, because the term admits two meanings. The a_{ij} might be considered "important" because s_{ij} or e_{ij} were large, and hence if a_{ij} were to be perturbed, λ would change a lot. On the other hand, a_{ij} might be considered "important" because the variability in λ is determined more by variability in a_{ij} than by variability in any other rate.

These two definitions are not—nor should they be expected to be—equivalent, and are distinguished as "prospective" and "retrospective" perturbation analysis (Caswell 1997, Horvitz et al. 1997, Caswell 2000a). Prospective analysis looks forward, and asks what will happen if a_{ij} is perturbed. Sensitivity and elasticity analysis provide the answer. Retrospective analysis looks back *at some observed set of matrices* $\mathbf{A}_i$ and asks how much the (co)variance of the a_{ij} contributed to the variance in λ. LTRE analysis, in its various guises, is the appropriate tool.

Prospective analysis is independent of any pattern of variation; indeed, it will happily and correctly give the results of changes in the vital rates that are biologically impossible ("What if this annual plant became a perennial; What if this plant without seed dormancy began to have some; What if pigs could fly ..."). Retrospective analysis, in contrast, is specific to the observed variation in the vital rates. The contribution of the variance in

a_{ij} to $V(\lambda)$ can be low either because λ is insensitive to a_{ij} or because a_{ij} happened not to vary much in the set of observations at hand, or both.

The distinction is particularly important in conservation, where (oversimplifying a bit) managers want to increase λ by imposing some policy, and they want to do so efficiently and economically. All else being equal, the best choice is a policy that targets those vital rates with highest sensitivity or elasticity; the proper tool is prospective sensitivity analysis. For a more detailed discussion, see MPM, Chapter 18.

14

Some Types of Instability

A population can be stable in only one way: by the constancy of its birth and death rates over time, and hence the constancy of its rate of increase. It can be unstable in infinitely many ways: by falling or rising birth rates, by falling or rising death rates, by either birth or death rates rising at some ages and falling at others, by the rise or fall being moderate or rapid, by its being linear, quadratic, or of higher degree. Any and all of these and their combinations could be given the same detailed attention as stability. Needless to say that will not be done here, nor is it likely to be done anywhere else.

The stable model has been studied in detail because it is more informative than any single case of instability. The diversity of kinds of instability distracts attention from the interest in any one kind. To keep down the complexity in the treatment here, either fertility or mortality will be allowed to vary, but not both, and populations will be assumed to be closed to migration as before.

Falling death rates are conspicuous in most of the world today; our first task is to see how their recognition alters the stable model of Chapter 5.

14.1 Absolute Change in Mortality the Same at All Ages

Coale (1963) opened up this subject by analyzing the effects of a steady fall in death rates; he called *quasi-stable* a population in which birth rates remain constant while death rates decline uniformly. The quantitative effect of this on rates of increase from the stable model was calculated by Coale and Demeny (1967). The following is an attempt to see the effect in general terms. In the first example death rates rise or fall by the same absolute amount at all ages. This is hardly realistic, but it will show the approach.

Suppose a rise each year at all ages equal to the constant k (which would be negative to provide for a fall, say of magnitude 0.0001, in which case $k = -0.0001$; the constant k is restricted to values that leave probability positive and less than 1 at all ages). After n years the mortality of age a will not be $\mu(a)$, as it was at the beginning, but $\mu^*(a) = \mu(a) + nk$. If the initial probability of survivorship was $l(x) = \exp[-\int_0^x \mu(a)\,da]$, the survivorship r years later, subject to $\mu^*(a) = \mu(a) + nk$, will be $l^*(x) = \exp\{-\int_0^x [\mu(a) + nk]\,da\} = l(x)e^{-nkx}$. This is true whether both the initial $\mu(a)$ and the subsequent $\mu(a) + nk$ apply to periods or to cohorts.

14.1.1 Inferring the Increase in Births

Apply the cohort case to an observed age distribution to ascertain the rate of increase in births. If $k = 0$, the stable assumption for ascertaining the rate of increase in the population from a census, derived as (5.2.4), gives

$$r_0 = \frac{1}{y-x}\log\left[\frac{B(t-x)}{B(t-y)}\right] = \frac{1}{y-x}\log\left[\frac{c_x/l_x}{c_y/l_y}\right], \tag{14.1.1}$$

where the c_x and c_y are the fractions in finite intervals around ages x and y (Figure 14.1) and are to be identified with observed populations. This is modified to estimate the increase in births between y and x years ago $y > x$, (once we abandon stability, we have to specify the time to which the rate of increase refers). If death rates increase at k per year between cohorts, and the life table for the cohort born y years ago is given by $\mu(a)$ and that for the cohort born x years ago $\mu^*(a) = \mu(a) + k(y - x)$, then $l_x^* = l_x e^{-k(y-x)x}$, and the estimate becomes

$$\begin{aligned} r_1 &= \frac{1}{y-x}\log\left(\frac{c_x/l_x^*}{c_y/l_y}\right) \\ &= \frac{1}{y-x}\log\left(\frac{c_x/\left(l_x e^{-k(y-x)x}\right)}{c_y/l_y}\right) \\ &= r_0 + kx, \end{aligned} \tag{14.1.2}$$

on simplifying and expressing the result in terms of (14.1.1).

For $y > x$ and falling mortality (i.e., $k < 0$), we have $r_1 \leqslant r_0$. In words, under a regime of falling mortality use of the customary formula (14.1.1), which assumes a fixed life table, gives a rate of increase in the earlier births that is too high by $-kx$, the fall in mortality since the younger age group was born. The usual formula (14.1.1) gives too high an r because it

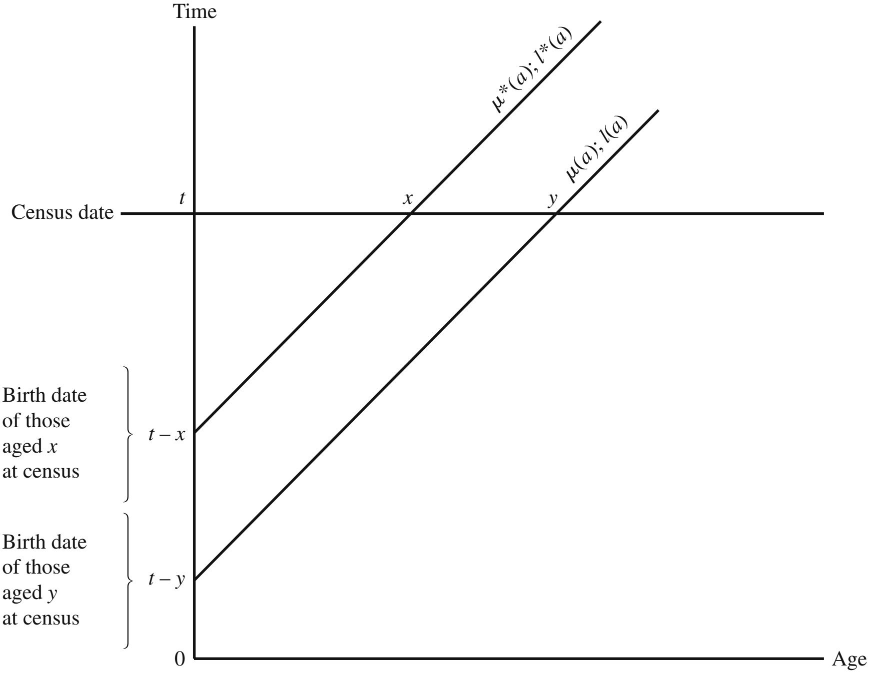


Figure 14.1. Lexis diagram showing census date and comparison of those enumerated at age x and y to ascertain increase between x and y years ago.

disregards the fact that the younger age with the life table l^*_x originates in fewer births than the life table l_x implies.

Let us generalize this to find the average rate of increase in persons aged z between the y cohort and the x cohort, where $x < z < y$. The c_y projected back to age z number $c_y(l_z/l_y)$, and the c_x projected forward to age z number $c_x e^{-k(y-x)(z-x)}(l_z/l_x)$. The average rate of increase r_z is found from

$$\exp[r_z(y-x)] = \frac{c_x e^{-k(y-x)(z-x)}(l_z/l_x)}{c_y(l_z/l_y)}$$
$$= \frac{c_x/l_x}{c_y/l_y} e^{-k(y-x)(z-x)},$$

or, on solving for r_z,

$$r_z = r_0 - k(z-x), \tag{14.1.3}$$

which reduces to (14.1.2) for $z = 0$. For k negative, the increase in any age z greater than x will be more rapid than r_0, the rate inferred on the stable model. [Show that the result applies to any z, whether or not it is between x and y.]

Increase in Person-Years in Cohort. To proceed from how persons aged z were changing over some particular interval to rate of increase in population, a first approach is to compare person-years in the two cohorts that are x and y years of age, respectively, at time t. Person-years in a cohort are births times expectation of life, that is to say, births times the integral of the survivorship function. This is not the usual way to estimate population increase, but is worth starting with because of its mathematical convenience. [Prove that the estimate of change is approximately

$$r_2 \approx r_0 - k(\bar{a} - x), \tag{14.1.4}$$

where $\bar{a}$ is the mean age.]

14.2 Proportional Change in Mortality

Now instead of a fixed change k at all ages suppose a fractional change—that mortality for a given cohort (or period) is $\mu(a)$, while n years later it is $\mu^*(a) = \mu(a)(1 + nk)$. Then that later cohort (or period) will have survivorship $l_x^* = l_x^{1+nk}$. If $k < 0$, then $l_x^* > l_x$.

Rate of Increase of Births. From the numbers c_y enumerated at time t and age y the births $B(t - y)$ must be $B(t - y) = c_y/l_y$, again taking as the standard life table the cohort passing through age y at census time t. The table for the younger cohort aged x at the census must have $\mu^*(a) = \mu(a)(1 + nk)$, so its survivorship to age x will be $l_x^* = l_x^{1+nk} = l_x^{1+(y-x)k}$. Then the number of births $B(t - x)$ from which the c_x are the survivors is

$$B(t - x) = \frac{c_x}{l_x^*} = \frac{c_x}{l_x^{1+(y-x)k}}. \tag{14.2.1}$$

The rate of increase in births between y and x years ago is estimated as

$$\begin{aligned} r_3 &= \frac{1}{y - x} \log \left[\frac{B(t - x)}{B(t - y)} \right] \\ &= \frac{1}{y - x} \log \left[\frac{c_x/l_x^*}{c_y/l_y} \right] \\ &= r_0 + \frac{1}{y - x} \log l_x^{-(y-x)k} \\ &= r_0 - k \log l_x. \end{aligned} \tag{14.2.2}$$

We could have taken our standard $\mu(a)$ to apply to the cohort midway between that aged x and that aged y at the time of the census and obtained approximately the symmetrical expression

$$r_4 \approx r_0 - k \log \sqrt{l_x l_y}. \tag{14.2.3}$$

[Prove this.] Both (14.2.2) and (14.2.3) show that a negative k for declining mortality, along with $\log l_x$ or $\log \sqrt{l_x l_y}$, which are always negative, gives a subtraction from the r calculated from (14.1.1); under the conditions of declining mortality (14.1.1) suffers from an upward bias in estimating the rate of increase of births.

But once again the historic rate of increase of births is not what we need most. We want change in population, and the survivorship will more than offset the correction in (14.2.3). First we review what we know from Section 4.3 of the effect on $\overset{o}{e}_0$ of a proportional change in mortality.

Change of $\overset{o}{e}_0$. The ratio of expectations of life for populations of different periods or cohorts with $l(a)$ changing by fixed fractions will depart substantially from unity, but never by as much as the ratio of age-specific mortality (Section 4.3). If one of two populations has mortality $\mu(a)$ and the other $\mu^*(a) = \mu(a)(1+k)$, the ratio of the expectation of life of the second to the first is

$$\frac{\overset{o}{e}{}_0^*}{\overset{o}{e}_0} = \frac{\displaystyle\int_0^\omega l(a)^{1+k}\,da}{\displaystyle\int_0^\omega l(a)\,da},$$

and, on expanding around $k = 0$ to the first term of a Taylor series, this becomes

$$\frac{\overset{o}{e}{}_0^*}{\overset{o}{e}_0} \approx \frac{\displaystyle\int_0^\omega \Big[l(a) + kl'(a)\Big]\,da}{\displaystyle\int_0^\omega l(a)\,da}, \tag{14.2.4}$$

where $l'(a)$ is the derivative of $l(a)^{1+k}$ with respect to k evaluated at $k = 0$. Now $dl(a)^{1+k}/dk = \log l(a) l(a)^{1+k} = l(a)\log l(a)$ at $k = 0$. Then we have to the linear approximation

$$\begin{aligned}\frac{\overset{o}{e}{}_0^*}{\overset{o}{e}_0} &= 1 + k\frac{\displaystyle\int_0^\omega l(a)\log[l(a)]\,da}{\displaystyle\int_0^\omega l(a)\,da} \\ &= 1 - kH,\end{aligned} \tag{14.2.5}$$

where the quantity H, defined in Section 4.3, is minus the average $\log l(a)$ weighted by $l(a)$. The parameter H of the life table is shown in Table 4.3 as about 0.20 for contemporary male populations and about 0.16 for females, these values being substantially lower than those for a generation earlier. We will apply (14.2.5) with $(y - x)k$ in place of k, since our assumption is that mortality increases by k per year.

Increase in Total Cohort Population. To approach the rate of increase in population rather than in births, we again calculate two cohorts, and compare the number of person-years lived in the cohort born at time $t-y$ with the number for that born at time $t-x$.

For the number of person-years in the cohort we multiply the births by the expectation of life, as was done for r_2 to obtain (14.1.4), and the outcome reduces to

$$r_5 = r_0 - k\log l_x + \frac{1}{y-x}\log\left(\frac{\overset{o}{e}{}_0^*}{\overset{o}{e}{}_0}\right), \tag{14.2.6}$$

on entering the values of (14.2.2). [Prove this.] On applying (14.2.5) with $(y-x)k$ in place of k, we obtain

$$r_5 = r_0 - k\log l_x + \frac{1}{y-x}\log\Big[1-(y-x)kH\Big]. \tag{14.2.7}$$

Approximating with $\log(1-\alpha) \approx -\alpha$, this becomes finally

$$r_5 = r_0 - k(\log l_x + H). \tag{14.2.8}$$

Similarly to the case in which we assumed a constant absolute change in mortality rates, this will apply a positive correction to r_0 when the death rate is falling ($k < 0$), whenever x is a young age.

Increase of Persons of Arbitrary Age. For arbitrary z, whether or not $x < z < y$, we can find the overall rate of increase by comparing the number of individuals in the cohort aged y at time t with those in the cohort aged x at time t. The ratio of the latter to the former is equated to $e^{r_z(y-x)}$, where r_z is the rate of increase sought. Thus we have the equation in r_z, the rate of increase of persons aged z,

$$e^{r_z(y-x)} = \frac{c_x(l_z/l_x)^{1+k(y-x)}}{c_y(l_z/l_y)},$$

and the solution in r_z is

$$r_z = r_0 + k\log\left(\frac{l_z}{l_x}\right), \tag{14.2.9}$$

for the average increase during time $t-y+z$ to $t-x+z$. [Use (14.2.9) to provide an alternative derivation of (14.2.8).]

The correction to r_0 is negative for $z < x$ and positive for $z > x$. The general form (14.2.9) would be a good starting point for deriving the previous expressions of this section as well as others that will occur to the reader. For $x = 0$, $k = -0.01$, and $r_0 = 0.03$ this gives the rate of increase at four ages, with Mexican male mortality of 1966 as the base:

$z =$	$r_z =$
0	0.0300
20	0.0314
40	0.0323
60	0.0347

The steady rise in the inferred rate of increase with age is clearly exhibited.

Demeny (1965) comes a step closer to realism by showing the effect of a decline of mortality (a one-year gain in $\overset{o}{e}_0$ per calendar year) that continues over a limited time (5 years, 10 years, ..., 40 years). Unfortunately this effect cannot be expressed analytically.

14.3 Changing Birth Rates

The effect of changing birth rates on the population rate of increase and on its age distribution is easily determined to a good approximation if the changes are the same for all maternal ages. The following is a generalization of stable theory that permits drawing conclusions as to rates of increase in a population, given its age distribution. In symbols, suppose that the birth rate for women aged a at time t is $m(a,t) = m(a)f(t)$ (Coale, 1963, p. 8; Coale and Zelnik, 1963, p. 83). The argument that follows shows how to infer the rates of birth and natural increase with some relaxation of the restriction on (5.2.4), though not without specializing $f(t)$.

The homogeneous form of the renewal equation (7.5.1) becomes

$$B(t) = \int_{\alpha}^{\beta} B(t-a)l(a)m(a)f(t)\,da.$$

In the birth function $B(t-a)$ under the integral sign, express both time t and age a as departures from A, the mean age of childbearing, so that $B(t-a)$ becomes $B[(t-A)-(a-A)]$, and then expand the birth function as a Taylor series around $t-A$:

$$\begin{aligned} B(t) &= \int_{\alpha}^{\beta} B\left[(t-A)-(a-A)\right] l(a)m(a)f(t)\,da \\ &= \int_{\alpha}^{\beta} \Big[B(t-A) - (a-A)B'(t-A) \\ &\quad + \frac{(a-A)^2}{2!}B''(t-A) - \cdots \Big] l(a)m(a)f(t)\,da. \end{aligned}$$

Dividing by $R_0 = \int_{\alpha}^{\beta} l(a)m(a)\,da$ and carrying through the integration expresses the right-hand side in terms of moments about A, the mean age of childbearing. The first moment is zero, and the second σ^2, the variance of ages of childbearing. Dividing by $B(t-A)$, taking logarithms, and then expanding the logarithm of the series of moments, gives, up to second

$$\log B(t) - \log B(t-A) = \log R_0 + \log f(t) + \frac{B''(t-A)\sigma^2}{2B(t-A)}. \qquad (14.3.1)$$

Since the term involving the second derivative is only about 2 or 3 percent of the value of the main term, it may be neglected in what follows. This amounts to the outrageous assumption that all children are born at the same age of mother. No argument could show that this assumption is reasonable, but numerical tests demonstrate that it makes little numerical difference in this particular application. If the assumption seems objectionable, however, the solution below could be regarded as a first approximation, $B''(t-A)$ calculated and entered in (14.3.1), and a more exact result obtained. (Alternatively, the second derivative could be replaced by a second difference.) Adding both sides over every Ath value and canceling leaves a summation of $\log f(t)$ at intervals of A, the mean age of childbearing:

$$\log B(t) = \log B_0 + \frac{t}{A}\log R_0 + \sum \log f(t),$$

and taking exponentials and writing e^{rt} for $R_0^{t/A}$, approximately true when the unit of time is very nearly a generation, gives

$$\begin{aligned} B(t) = B_0 e^{rt} \exp\{&\log f(0) + \log f(A) + \cdots \\ &+ \log f[(n-1)A]\}. \end{aligned} \tag{14.3.2}$$

After the particular function $f(t) = \exp[k_1 t + k_2 t(t-A)]$ is entered, the summation is evaluated to $t = nA$ as

$$B(t) = B_0 \exp\left[rt + k_1\frac{t(t+A)}{2A} + k_2\frac{t(t^2-A^2)}{3A}\right]. \tag{14.3.3}$$

The intended application being to an age distribution with typical age a, we are interested in births at time $-a$, and these are obtained by putting $t=-a$, in (14.3.3) and then dividing by the current population N:

$$\frac{B(-a)}{N} = b\exp\left[-ra + \frac{k_1 a(a-A)}{2A} - \frac{k_2 a\left(a^2-A^2\right)}{3A}\right], \tag{14.3.4}$$

where b is the current birth rate. Once an age distribution $c(a)$ is given for a population that can be assumed to be closed and whose life table is $l(a)$, $B(-a)/N$ can be estimated by $c(a)/l(a)$. Taking logarithms and fitting to at least four ages produces the four constants b, r, k_1, and k_2. A more detailed derivation of (14.3.4) and the fitting to data are given in Keyfitz et al. (1967).

If $k_1 = k_2 = 0$, (14.3.4) reduces to an equation of Bourgeois-Pichat (1958), shown in Section 5.5. In the more general case where only $k_2 = 0$, we have the result due to Coale and Zelnik (1963, p. 83).

Fitting (14.3.4) to an age distribution, when we are given or may assume a life table, estimates not only the rate of increase and the birth rate but also the change in the latter as indicated by k_1 and k_2. Based on four or more ages, the four constants b, r, k_1, and k_2 are obtained, and these permit a reconstruction of the age distribution by multiplying the right-hand side of (14.3.4) by $l(a)$. The age distribution so obtained can be compared with the observed age distribution. The fit may be compared with

Table 14.1. Estimates of birth rates and other parameters from (9.3.4) for five populations, fitted to ages 5 to 74

Country and year	$1000b$	$1000r$	$1000k_1$	$1000k_2$
Fiji Islands, 1964	38.7	31.9	−6.0	0
France, 1899–1903	19.3	−4.6	−23.5	−0.3
Honduras, 1965	42.9	34.4	−5.6	0
Japan, 1962	16.8	−3.5	−51.6	−0.5
Netherlands, 1901	31.8	14.7	−12.1	−0.2

that resulting when k_2 is put equal to zero, and when $k_1 = k_2 = 0$. Unless it is substantially better, one would avoid the complication of the additional constants; but if the reconstruction of ages is markedly improved with the nonstable method here described, so presumably also are the estimates of b and r.

Table 14.1 gives estimates for certain countries of the four constants contained in (14.3.4). The usefulness of these is suggested by the fact that in all five cases the model reproduces their age distribution at ages 5 to 74 appreciably better than does the stable model with $k_1 = k_2 = 0$. In two instances the k_2 did not appreciably improve the fit; in the other three it did, but the calculated k_2 was small. Note that k_1 always turned out negative, reflecting falling birth rates.

14.4 Announced Period Birth Rate Too High

When successive cohorts bear children at younger and younger ages, each period cross section will tend to catch more births than any cohort. The period births are "too many" in the sense that no one cohort of women has so high an average; the childbearing of successive cohorts overlaps in each period. In the early 1940s there were many marriages and hence first births, and Whelpton (1946) showed from the 1942 registrations that if the pace continued women would average 1.084 first births each. Conversely, when couples are having their children later and later, a given period will catch less than its share of births, that is to say, fewer than pertain to any cohort.

To prove this and similar propositions requires a formal means of translating cohort moments into period moments and vice versa. The problem in its general form has been solved by Ryder (1964). What follows is a self-contained adaptation of his solution.

If the same life table applies at all times, and the probability at time t of a woman of age x to $x+dx$ having a child is $m(x,t)\,dx$, the net reproduction rate R_0 is

$$R_0 = \int_\alpha^\beta l(x)m(x,t)\,dx,$$

where α and β are the youngest and oldest ages, respectively, of childbearing, and $l(x)$ is the chance that a child just born will live at least to age x. The R_0 is a function of time; the fact that fertility varies with time as well as with age is what gives rise to our problem.

The nth period moment at time zero about age A chosen arbitrarily is defined as

$$\frac{R_n(A)}{R_0} = \frac{\displaystyle\int_\alpha^\beta (x-A)^n l(x)m(x,0)\,dx}{\displaystyle\int_\alpha^\beta l(x)m(x,0)\,dx},$$

where the R_0 does not depend on A. The nth cohort moment is similarly

$$\frac{R_n^*(A)}{R_0^*(A)} = \frac{\displaystyle\int_\alpha^\beta (x-A)^n l(x)m(x,x-A)\,dx}{\displaystyle\int_\alpha^\beta l(x)m(x,x-A)\,dx},$$

in which $R_0^*(A)$ does depend on A. The cohort moment contains a time argument $x-A$, contrived so that $m(x,x-A)$ follows the group of women born A years before time t down their life lines. It selects out of the period births as officially published for successive years those appropriate to the particular cohort (Figure 14.2).

The cohort $R_n^*(A)$ is expressible in terms of $R_n(A)$, $R_{n+1}(A), \ldots,$ by means of a Taylor expansion of $m(x,x-A)$ about time $t=0$. For this expansion $m(x,x-A)$ is treated for any fixed age x as a simple function of time, so that for each x,

$$m(x,x-A) = m(x,0) + (x-A)\dot{m}(x,0) + \frac{(x-A)^2}{2!}\ddot{m}(x,0) + \cdots, \quad (14.4.1)$$

the dots representing differential with respect to time. The accuracy of approximation by a given number of terms will depend on the smoothness of whatever changes are taking place in the birth function.

Entering the Taylor expansion (14.4.1) in the expression for $R_n^*(A)$ gives

$$R_n^*(A) \int_\alpha^\beta (x-A)^n l(x)[m(x,0) + (x-A)\dot{m}(x,0) + \cdots]\,dx. \quad (14.4.2)$$

To put this into convenient form we have to convince ourselves that the integral $\int_\alpha^\beta (x-A)^{n+1} l(x)\dot{m}(x,0)\,dx$ is the derivative with respect to time of the period moment, that is, equals $\dot{R}_{n+1}(A)$, and similarly for later terms. The proof consists in representing $\dot{m}(x,0)$ as a difference between two values, $m(x,\Delta)$ and $m(x,0)$, divided by Δ. Before letting Δ tend to

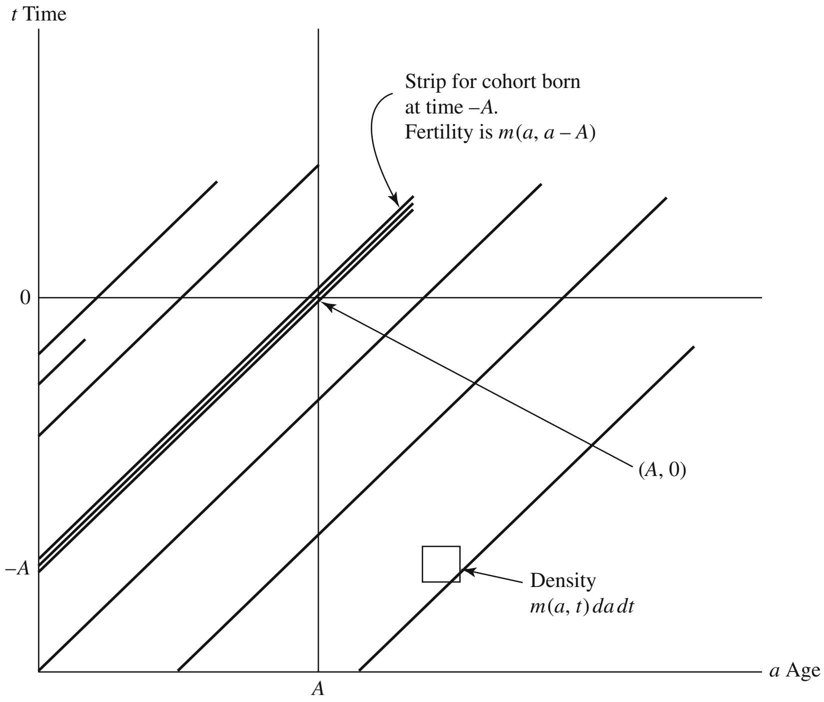


Figure 14.2. Cohort and period fertility.

zero, note that the integral $\int_{\alpha}^{\beta}(x-A)^{n+1}l(x)m(x,\Delta)\,dx$ is $R_{n+1}(A)$ at time Δ, say $R_{n+1}(A,\Delta)$, and similarly when Δ is replaced by zero. The difference of the two integrals divided by Δ is

$$\frac{R_{n+1}(A,\Delta)-R_{n+1}(A,0)}{\Delta},$$

and this becomes $\dot{R}_{n+1}(A,0)=\dot{R}_{n+1}(A)$ as Δ tends to zero. A similar proof would show that later terms of (14.4.2) become $\ddot{R}_{n+2}(A)/2$, and so on. Expressed more briefly, the fact we are using is that the integral of a derivative equals the derivative of an integral under conditions far more general than are required for the demographic application.

On incorporating this fact, expansion 14.4.2 leads to the fundamental result

$$R_n^*(A)=R_n(A)+\dot{R}_{n+1}(A)+\frac{\ddot{R}_{n+2}(A)}{2!}+\cdots, \tag{14.4.3}$$

which estimates the nth cohort moment in terms of the nth and higher period moments (Ryder, 1964). The most interesting application of (14.4.3) will be for $n=0$, and, truncating the series at the second term on the right, we obtain

$$
\begin{aligned}
R_0^*(A) &\approx R_0 + \dot{R}_1(A) \\
&\approx R_0 \left[1 + \frac{\dot{R}_1(A)}{R_0} \right],
\end{aligned}
\tag{14.4.4}
$$

the zeroth period moment not depending on the arbitrary A. In words, we can say that the cohort net reproduction rate is equal to the period net reproduction rate plus the change over unit time in $R_1(A)$.

The theoretical result 14.4.4 can be illustrated arithmetically. To do so, the net maternity function for United States females in 1967 was graduated by a Hadwiger function (Keyfitz, 1968, p. 149); then, leaving R_0 and the variance σ^2 the same in all periods as in 1967, the mean μ was shifted upward by 0.1 year per 5-year period. This produces the results shown in Table 14.2 for 1972 to 2012. It is seen that the period net reproduction rate remains always at 1.205. A change in the mean of 0.1 year per 5 years is a change at the rate of 0.02 per year. The right-hand side of (14.4.4) is therefore $1.205(1 + 0.02) = 1.229$.

If we add the diagonals for the three completed cohorts in Table 14.2, we find $R_0^* = 1.230$, different from the 1.229 of (14.4.4) by rounding error only.

A virtually identical argument expresses the period $R_n(A)$ in terms of the series of cohort moments $R_n^*(A)$, $R_{n+1}^*(A), \ldots$:

$$
R_n(A) = R_n^*(A) - \dot{R}_{n+1}^*(A) + \frac{\ddot{R}_{n+2}^*(A)}{2!} - \cdots, \tag{14.4.5}
$$

the difference from (14.4.3) being that the signs on the right-hand side here come out alternately positive and negative.

Note that (14.4.3) and (14.4.5) serve quite different purposes. The first is useful because it can incorporate the latest period information to suggest how current cohorts are likely to be completed, on which definitive statistics await their members reaching age 50 or so if they are women, older if they are men. Equation (14.4.5), on the other hand, shows how the cohorts, considered as the basic units underlying the process, are translated or distorted in the course of expressing themselves in successive periods.

Among other uses such results provide information on the relative variation through time of cohort and period fertility. If it were true historically that when R_0 rises $R_1(A)$ tends to fall, that is, $\dot{R}_1(A)$ is negative, and when R_0 falls $R_1(A)$ tends to rise, then the variation of $R_0^*(A) = R_0 + \dot{R}_1(A)$ would be smaller than the variation of R_0 by itself. Changes in the period net reproduction rate being in these circumstances offset by contrary changes in the mean age of childbearing, and the cohort R_0^* being subject to less variation than the period R_0, R_0^* is useful in predicting future population, since all prediction depends on finding functions that are relatively constant. (Section 12.4)

Table 14.2. Net maternity function for United States females, 1967, as graduated by Hadwiger function; later periods supposing increase of 0.1 in mean age μ of each successive period, with total R_0 and variance σ^2 fixed at those of 1967

Year	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	Total
1967	0.0095	0.1512	0.3876	0.3656	0.1926	0.0710	0.0208	0.0066	1.205
1972	0.0088	0.1462	0.3846	0.3685	0.1960	0.0727	0.0213	0.0068	1.205
1977	0.0082	0.1413	0.3814	0.3714	0.1995	0.0744	0.0219	0.0071	1.205
1982	0.0076	0.1365	0.3781	0.3742	0.2030	0.0761	0.0224	0.0073	1.205
1987	0.0070	0.1318	0.3746	0.3769	0.2065	0.0778	0.0230	0.0074	1.205
1992	0.0065	0.1271	0.3711	0.3795	0.2100	0.0796	0.0236	0.0076	1.205
1997	0.0060	0.1226	0.3674	0.3821	0.2136	0.0814	0.0242	0.0079	1.205
2002	0.0055	0.1182	0.3636	0.3845	0.2171	0.0833	0.0248	0.0080	1.205
2007	0.0051	0.1138	0.3597	0.3869	0.2207	0.0851	0.0255	0.0082	1.205
2012	0.0047	0.1095	0.3557	0.3891	0.2244	0.0871	0.0261	0.0085	1.205

Note that all totals on rows are 1.205, the period R_0; totals on the three complete cohorts shown are $R_0^* = 1.230$. The purpose of this hypothetical table is to show that the cohort NRR can be constant and different from the period NRR if the timing of fertility undergoes a steady change.

If we knew that every cohort was aiming at exactly three children, (14.4.3) and (14.4.5) would not be needed; we would simply deduct the average number of children already recorded from three, and suppose the remainder to be distributed over time and age in the future in some suitable way. If, on the other hand, cohorts had nothing to do with the matter, we would treat the births to women of given age in successive periods as an ordinary series and extrapolate. The theory of this section is especially useful for the intermediate case in which the R_0^* for cohorts are shifting, but less rapidly than the R_0 for periods.

We could have gone through the argument with gross reproduction rates and obtained the same results, simply by omitting the $l(x)$ throughout. Even more generally, the $m(a, t)$ function and the R's representing the moments of $m(a, t)$ could be interpreted, not as childbearing, but as mortality, marriage, school attendance, income, or some other attribute of individuals. The foregoing relation of periods and cohorts applies to any such characteristic.

14.5 Backward Population Projection

One may need to project a population backward, for instance to estimate age distribution of a period before the first of a series of censuses. Under stability all ages increase in the same ratio; therefore one can calculate the number living at the earlier period by dividing by this ratio. With or without stability one can simply divide the total for an age group by the probability of surviving to obtain the age group one time period earlier. This gives the best possible answer in practice for all ages but the last, to which it is inapplicable. The Leslie matrix is useless for such purposes, being singular for any population that lives beyond its reproductive span; it is of rank $n-1$, where n is the number of age groups recognized.

An easy way to retrieve the last age interval is to suppose that it has been increasing at the intrinsic rate of the population. But in some circumstances one can do better than this by using the generalized inverse, which does not require the assumption of stability (Greville and Keyfitz 1974).

First we express the projection backward along cohort lines in matrix form as

$$\mathbf{X}_0 = \begin{pmatrix} 0 & \dfrac{L_0}{L_5} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \dfrac{L_5}{L_{10}} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \dfrac{L_{10}}{L_{15}} & \cdots & 0 \\ & \vdots & & & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \dfrac{L_{5n-10}}{L_{5n-5}} \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

which premultiplies the vector (say of females) at time t to provide an estimate for time $t-1$ of all age intervals except the last, say 85 and over. This contains the main elements of the required backward projection, and is therefore at least a commonsense inverse of the forward projection $\mathbf{A}$ (Section 3.1).

There are many generalized inverses of a given singular matrix, and the choice of one from among them depends on the use to be made of it. If $\mathbf{A}$ is the given matrix, it is usual to choose a generalized inverse $\mathbf{X}$ that satisfies at least one and preferably both of the two relations (Rao and Mitra 1971)

$$\mathbf{AXA} = \mathbf{A} \tag{14.5.1}$$

and

$$\mathbf{XAX} = \mathbf{X}; \tag{14.5.2}$$

the first says that the backward form $\mathbf{X}$ must be such that projecting forward, then backward, and then forward again is equal to projecting once forward, and an equally obvious interpretation applies to the second.

In fact $\mathbf{X}_0$ is a reasonably satisfactory generalized inverse of $\mathbf{A}$. Like $\mathbf{A}$, it is of order n and rank $n-1$. One can easily verify that it satisfies both (14.5.1) and (14.5.2). A disadvantage of $\mathbf{X}_0$, however, lies in the fact that by its use the population in the oldest age interval at $t-1$ always comes out zero.

This is easily remedied by observing that properties (14.5.1) and (14.5.2) are retained if the zeros in the bottom row of $\mathbf{X}_0$, except the first, are replaced by arbitrary elements. The last-row elements estimate the number in the final age interval at time t as some linear combination of the numbers at time $t+\Delta t$ in all age intervals except the youngest.

Of the many ways of arriving at such a linear estimate the most direct is to use the eigenvalues and eigenvectors of the two matrices. For the classical inverse of a nonsingular matrix $\mathbf{H}$, if

$$\mathbf{Hx} = \lambda\mathbf{x},$$

then

$$\mathbf{H}^{-1}\mathbf{x} = \lambda^{-1}\mathbf{x}.$$

In words, we can say that a nonsingular matrix and its inverse have identical eigenvectors, associated with respective eigenvalues that are reciprocals of each other. It has been shown (Greville 1968) that something of the same kind is true of a singular matrix and its generalized inverse.

The three important eigenvalues of the matrix are the real root, say λ_1, and the conjugate pair of complex eigenvalues closest to the real root in absolute value, denoted as λ_2 and $\bar{\lambda}_2$. The real root is the ultimate ratio of increase in the population that would result if the mortality and natality conditions reflected in the Leslie matrix were perpetuated, while the pair of complex roots is related to the amplitude and period of the oscillations that would precede the attainment of a stable state.

The real cubic polynomial with leading coefficient unity whose three zeros are the reciprocals of these three eigenvalues is $(\lambda_1 > 0)$,

$$q(z) = \left(z-\lambda_1^{-1}\right)\left(z-\lambda_2^{-1}\right)\left(z-\bar{\lambda}_2^{-1}\right) = z^3 + c_2z^2 + c_1z + c_0. \quad (14.5.3)$$

To form a polynomial whose roots are the reciprocals of those of a given polynomial all we need do is reverse the order of the coefficients. [Show that the roots of

$$c_0z^n + c_1z^{n-1} + \cdots + c_n = 0$$

are the reciprocals of the roots of

$$c_nz^n + c_{n-1}z^{n-1} + \cdots + c_0 = 0.$$

Then, if λ_1 is the real root, and $\lambda_2 = a + ib$ and $\bar{\lambda}_2 = a - ib$ are the main complex roots, the characteristic equation with these roots is

$$(z - \lambda_1)(z - \lambda_2)(z - \bar{\lambda}_2) = 0,$$

or

$$\begin{aligned} &z^3 - (\lambda_1 + \lambda_2 + \bar{\lambda}_2)z^2 + (\lambda_1\lambda_2 + \lambda_1\bar{\lambda}_2 + \lambda_2\bar{\lambda}_2)z \\ &- \lambda_1\lambda_2\bar{\lambda}_2 = 0, \end{aligned}$$

or

$$z^3 - (\lambda_1 + 2a)z^2 + (2\lambda_1 a + a^2 + b^2)z - \lambda_1(a^2 + b^2) = 0,$$

and the equation with the reciprocals of these roots is the same with coefficients reversed:

$$\lambda_1(a^2 + b^2)z^3 - (2\lambda_1 a + a^2 + b^2)z^2 + (\lambda_1 + 2a)z - 1 = 0.$$

Hence the coefficients of (14.5.3) are

$$c_2 = -\frac{2\lambda_1 a + a^2 + b^2}{\lambda_1(a^2 + b^2)}, \quad c_1 = \frac{\lambda_1 + 2a}{\lambda_1(a^2 + b^2)}, \quad c_0 = -\frac{1}{\lambda_1(a^2 + b^2)}.$$

Let p_i denote the survival rate from the ith age interval to the $(i + 1)$th (which is the ith subdiagonal element of $\mathbf{A}$). Then, if we take $\mathbf{X}_1$ to be a matrix like $\mathbf{X}_0$ except that the last three elements of the bottom row are

$$-c_0 p_{n-1} p_{n-2}, \quad -c_1 p_{n-1}, \quad -c_2 \tag{14.5.4}$$

(instead of zeros), it is easily verified that the characteristic polynomial of $\mathbf{X}_1$ is $z^{n-3}q(z)$. Thus the eigenvalues of $\mathbf{X}_1$ consist of $n - 3$ zeros and the reciprocals of λ_1, λ_2, and λ_3. It is possible to show that $\mathbf{A}$ and $\mathbf{X}_1$ have in common the eigenvectors associated with these three eigenvalues.

The rule seems to be that for long-term projection backward one cannot improve on the dominant root to estimate the oldest age group. For backward projection over a short interval, however, the first three roots often seem to help. In general, the shorter the interval over which one projects backward the more possible it is to preserve minor roots without finding erratically large and impossibly negative populations.

14.5.1 Application

Let us test these suggestions by backward projection of the older United States female population from 1967: using data from that year only, we will estimate for 1962. For all age intervals but the last this will of course be done by the reciprocal of the survival ratio, a procedure whose properties are straightforward and well known.

From the vital statistics for 1967 the real root is $\lambda_1 = 1.0376$, and the roots following are λ_2, $\bar{\lambda}_2 = 0.3098 \pm 0.7374i$. The polynomial $q(z)$ of

(14.5.3) is

$$q(z) = z^3 - z^2(1.9323) + z(2.4966) - 1.5065.$$

The probability of survival into the last age interval is $p_{n-1} = 0.8024$, and into the second-to-last interval is $p_{n-2} = 0.7030$. Hence the bottom row of the inverse matrix ends up with the three numbers

$$\begin{pmatrix} \multicolumn{4}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ \ldots & 0.8498 & -2.0032 & 1.9323 \end{pmatrix},$$

and they are to premultiply the last three intervals (75 to 79, 80 to 84, 85+) of the 1967 age distribution:

$$\begin{pmatrix} \cdots \\ 2198 \\ 1286 \\ 727 \end{pmatrix}$$

expressed in thousands. The inner product of the two triplets of numbers above constitutes the estimate in this particular way of forming the inverse, and it turns out to be 696 thousands. The observed 1962 figure was 602 thousands. With the procedure of projecting backward by the reciprocal of the real root alone we obtain $727/1.0376 = 701$, which is a slightly larger discrepancy.

For some other populations, the superiority of the three roots method shows more clearly. For example, using data for Belgium, 1960, to estimate women aged 85 and over for 1955, we find that the three roots method gives 31,323 and one root gives 33,543 against an observed 1955 figure of 27,880. Bulgaria, 1965, projects backward to 1960 on three roots at 17,550, and on one root at 23,567, against an observed 15,995. In terms of percentage error, the three cases show the following results:

Data	Three Roots	One Root
United States, 1967	15.6	16.4
Belgium, 1960	12.3	20.3
Bulgaria, 1965	9.7	47.3

The idea of using the generalized inverse occurred to Thomas Greville as a way of formalizing and extending the projection backward along cohort lines, sometimes called *reverse survival*, that is used by demographers. This method contrasts with projecting backward by truncating the Leslie matrix at the last age of reproduction and then using its ordinary inverse. Numerical experiments show no circumstance in which such a method is comparable in accuracy with reverse survival; working back from an observed age distribution gives large negative numbers from the first or second 5-year period onward. If one starts with the artificial age vector obtained by projecting forward on the Leslie matrix, it is possible to project backward on the inverse of the truncated matrix for a few periods, but this serves no purpose, except possibly to study the accuracy attained in double precision by the particular computer in use. All these statements reflect the severely ill-conditioned nature of the truncated matrix.

Attempts to project the population backward for long periods of time run afoul of ergodicity. Any two initial age distributions subjected to the same sequence of vital rates converge toward each other. Thus there are many historical series of age structures that could have led to any present structure, and an attempt to work backwards is doomed to failure. Historical demographers have had to confront this problem in attempts to reconstruct demographic trends from limited data on population size and births and deaths (Lee 1985, 1993).

14.6 The Time to Stability

A baby boom or other irregularity in the time curve of births tends to be echoed in each later generation insofar as the subsequent age-specific rates of birth and death are constant. The mechanism, expressed in words, is that, when the girls born in a baby boom are of reproductive age, mostly about 20 to 30 years later, there will be more mothers in proportion to the population, and consequently again more children. Still in commonsense terms, the narrower and less skewed the range of ages in which women bear children, the more concentrated will be the echo, and the larger the ratio to the original disturbance. With a broad range of ages of childbearing, especially one skewed to older ages, the waves would seem likely to disappear more quickly, again supposing fixed subsequent rates.

Insofar as such fluctuations incur social cost in first overcrowded and then underutilized facilities such as schools, one is interested in the quickness of convergence to stable form of a population that has undergone a perturbation. In recent decades the variance of ages of childbearing in the United States has diminished. Does this mean a slower reversion to stability after a disturbance? Or does skewness help more than variance to speed the convergence to stability? One way of answering these questions is in terms of the main complex roots of the Lotka equation. In the matrix model framework, the corresponding analysis is in terms of the subdominant eigenvalue of the projection matrix (Section 7.3.1).

The literature on time to convergence was initiated by Coale (1968, 1972), and contributions have been made by Sivamurthy (1971), Trussell (1977), Tuljapurkar (1982, 1993), Schoen and Kim (1991) and others.

14.6.1 The Criterion of Convergence

Underlying and facilitating all this work is the fact that the real part of the second largest root in absolute value, r_2 in (7.5.2), largely determines the time to convergence. The magnitude of contribution of the term $e^{r_2 t}$, which can be written as $e^{(x+iy)t}$, depends on e^{xt}. The ratio of the asymptotic effect of the second root to that of the first, as these are projected to t years, is e^{xt}/e^{rt} (this is the inverse of the damping ratio defined in Section 7.3.1; suppose that we want to know when the ratio will be less than ϵ:

$$e^{(x-r)t} < \epsilon.$$

Remembering that $x - r$ is negative, we have

$$t > \frac{\log \epsilon}{x - r}. \tag{14.6.1}$$

Note that this studies convergence in terms of the exponentials in t of (7.5.2) and takes the constants Q into ϵ.

14.6.2 Use of the Characteristic Equation

Start with equation 6.1.2,

$$\int_{\alpha}^{\beta} e^{-ra} l(a) m(a)\, da = 1,$$

divide both sides by R_0, the net reproduction rate, take logarithms, expand the exponential, and so obtain the series of cumulants

$$\mu r - \sigma^2 \frac{r^2}{2!} + \mu_3 \frac{r^3}{3!} - \cdots - \log R_0 = 0, \tag{14.6.2}$$

which is Lotka's equation 161 (1939, p. 69) and is the same as (6.2.1) if $\psi(r)$ is put equal to unity. Entering $r = x + iy$ gives for the real part

$$\begin{aligned}\phi = \mu x &- \frac{\sigma^2(x^2 - y^2)}{2} + \frac{\mu_3(x^3 - 3xy^2)}{6} \\ &- \frac{\kappa_4(x^4 - 6x^2y^2 + y^4)}{24} + \frac{\kappa_5(x^5 - 10x^3y^2 + 5xy^4)}{120} \\ &- \cdots - \log R_0 = 0,\end{aligned} \tag{14.6.3}$$

and for the imaginary part

$$\begin{aligned}\theta = \mu y - \sigma^2 xy &+ \frac{\mu_3(3x^2y - y^3)}{6} - \frac{\kappa_4(x^3y - xy^3)}{6} \\ &+ \frac{\kappa_5(5x^4y - 10x^2y^3 + y^5)}{120} - \cdots = 0.\end{aligned} \tag{14.6.4}$$

The characteristic equation 6.1.2 is expressed as $\phi + i\theta = 0$. Designate derivatives of the real and complex parts with respect to the second cumulants as ϕ_2 and θ_2, respectively, with respect to the third cumulants as ϕ_3 and θ_3, with respect to x as ϕ_x and θ_x, and with respect to y as ϕ_y and θ_y, and similarly for x and y with respect to the cumulants. In this shorthand $\partial x/\partial\sigma^2$, for example, will be written as x_2, and $\partial x/\partial\mu_3$ as x_3. Completely differentiate the real part, $\phi = 0$, by σ^2 and μ_3 in turn:

$$\phi_x x_2 + \phi_y y_2 + \phi_2 = 0 \tag{14.6.5}$$

$$\phi_x x_3 + \phi_y y_3 + \phi_3 = 0, \tag{14.6.6}$$

and the same for the imaginary part, $\theta = 0$:

$$\theta_x x_2 + \theta_y y_2 + \theta_2 = 0 \tag{14.6.7}$$

$$\theta_x x_3 + \theta_y y_3 + \theta_3 = 0. \tag{14.6.8}$$

The first and third of these equations in partial derivatives can be solved for x_2, the derivative of x with respect to variance:

$$x_2 = \frac{\begin{vmatrix} -\phi_2 & \phi_y \\ -\theta_2 & \theta_y \end{vmatrix}}{\begin{vmatrix} \phi_x & \phi_y \\ \theta_x & \theta_y \end{vmatrix}},$$

where $|\mathbf{X}|$ is the determinant of $\mathbf{X}$. The second and fourth equations can be solved for x_3, the derivative with respect to the third moment:

$$x_3 = \frac{\begin{vmatrix} -\phi_3 & \phi_y \\ -\theta_3 & \theta_y \end{vmatrix}}{\begin{vmatrix} \phi_x & \phi_y \\ \theta_x & \theta_y \end{vmatrix}},$$

The denominators are the same for x_2 and x_3, and we want the ratio

$$\frac{x_2}{x_3} = \frac{\begin{vmatrix} -\phi_2 & \phi_y \\ -\theta_2 & \theta_y \end{vmatrix}}{\begin{vmatrix} -\phi_3 & \phi_y \\ -\theta_3 & \theta_y \end{vmatrix}} = \frac{\phi_y\theta_2 - \theta_y\phi_2}{\phi_y\theta_3 - \theta_y\phi_3}.$$

This simple expression provides the ratio of the effect of σ^2 on x to the effect of μ_3 on x, since x_2/x_3 is shorthand for $(\partial x/\partial\sigma^2)/(\partial x/\partial\mu_3)$. It assumes that the second and third moments can vary independently of each other, and that all other moments are fixed, including R_0, the net reproduction rate.

14.6.3 Exact and Approximate Ratios of Partial Derivatives

To evaluate x_2/x_3 we refer back to the expansions of ϕ and θ and calculate the partials:

$$\phi_x = \mu - \sigma^2 x + \frac{\mu_3(x^2 - y^2)}{2} + \cdots$$

$$
\begin{aligned}
\phi_y &= = \sigma^2 y - \mu_3 xy + \cdots \\
\theta_x &= -\sigma^2 y + \mu_3 xy - \cdots \\
\theta_y &= \mu - \sigma^2 x + \frac{\mu_3(x^2 - y^2)}{2} - \cdots \\
\phi_2 &= \frac{y^2 - x^2}{2} \\
\theta_2 &= -xy \\
\phi_3 &= \frac{x^3 - 3xy^2}{6} \\
\theta_3 &= \frac{3x^2 y - y^3}{6}.
\end{aligned}
$$

Entering these in the expression for x_2/x_3, we have

$$
\begin{aligned}
\frac{x_2}{x_3} &= \frac{\phi_y \theta_2 - \theta_y \phi_2}{\phi_y \theta_3 - \theta_y \phi_3} \\
&= \frac{(\sigma^2 y - \mu_3 xy + \cdots)(-xy) - (\mu - \sigma^2 x + \cdots)\left(\frac{y^2 - x^2}{2}\right)}{(\sigma^2 y - \mu_3 xy + \cdots)\left(\frac{3x^2 y - y^3}{6}\right) - (\mu - \sigma^2 x + \cdots)\left(\frac{x^3 - 3xy^2}{6}\right)}, \quad (14.6.9)
\end{aligned}
$$

which is readily evaluated to as many terms as desired.

Choosing the terms that seem most to affect the answer gives the approximation

$$
\frac{x_2}{x_3} \approx \frac{(\sigma^2/\mu)(2x) + 1}{(\sigma^2/\mu)(y^2/3) - x}. \tag{14.6.10}
$$

If σ^2/μ, the ratio of variance to mean in the net maternity function, is 1.3, and $r_2 = x + iy$ is $-0.03 + 0.22i$, then (14.6.10) works out to 18.

14.6.4 Allowance for Different Ranges of Variance and Skewness Among Observed Populations

For a typical population with given age-specific rates of birth and death, we have found that the ratio of the effect of variance to that of skewness, x_2/x_3 above, is about 18. Except for the approximations made after in the interest of ease of calculation, (14.6.9) is an exact result, allowing the several moments to vary independently of one another, and taking account of as many moments as are needed for fit to the net maternity function. But among real populations skewness varies more than variance. Variances are mostly between 25 and 55, whereas skewnesses are between 50 and 200; that is, the range of skewness is 5 times that of variance.

Can we escape this difficulty by somehow finding theoretically comparable measures of dispersion and skewness? A change in the scale on which fertility is measured affects the square root of the variance to the same degree as the cube root of the skewness. Would we have been better off to

Table 14.3. Ratio of Effect of Second Moment to That of Third Moment, All Others Constant: $x_2/x_3 = (\partial x/\partial\sigma^2)/(\partial x/\partial\mu_3)$

Country and year	x_2/x_3 (9.6.9)	x_2/x_3 (9.6.10)	x_2^*/x_3^* (9.6.11)
Algeria, 1965	16.670	16.805	2.673
Austria, 1968	12.636	13.052	2.027
Bulgaria, 1968	15.480	16.273	2.137
Canada, 1968	14.740	15.170	2.385
Colombia, 1965	19.708	19.886	3.276
England and Wales, 1968	16.959	17.330	2.952
France, 1967	17.861	18.322	2.893
Germany, West, 1967	18.834	19.146	3.448
Greece, 1968	18.831	19.094	3.699
Honduras, 1966	16.127	16.090	2.970
Ireland, 1967	38.027	37.953	12.340
Jamaica, 1963	15.219	15.438	2.881
Madagascar, 1966	11.419	11.795	1.553
Mexico, 1966	18.777	18.722	3.906
Nicaragua, 1965	16.034	16.147	2.811
Panama, 1966	15.027	15.362	2.275
Puerto Rico, 1965	8.979	9.638	1.142
Taiwan, 1966	36.809	37.195	5.409
Togo, 1961	11.360	11.660	1.663
Trinidad and Tobago, 1967	13.774	14.057	2.237
United States, 1966	13.713	14.183	2.131
Males	16.172	17.249	1.604
Venezuela, 1965	19.292	19.445	3.316

Female populations only, except for the United States.
Based on data in Keyfitz and Flieger, (1968, 1971).

start from the beginning with the standard deviation and the cube root of the third moment?

It is easy enough to convert our results to what would have been obtained with that procedure. We write $\partial x/\partial\sigma$ as x_2^* and $\partial x/\partial\sqrt[3]{\mu_3}$ as x_3^*. Then

$$x_2^* = \frac{\partial x}{\partial\sigma} = \frac{\partial x}{\partial\sigma^2}\,\frac{\partial\sigma^2}{\partial\sigma} = 2\sigma\frac{\partial x}{\partial\sigma^2}$$

$$x_3^* = \frac{\partial x}{\partial\sqrt[3]{\mu_3}} = \frac{\partial x}{\partial\mu_3}\,\frac{\partial\mu_3}{\partial\sqrt[3]{\mu_3}} = 3\mu_3^{2/3}\frac{\partial x}{\partial\mu_3},$$

so that the new ratio, x_2^*/x_3^*, is

$$\frac{x_2^*}{x_3^*} = \frac{2\sigma}{3\mu_3^{2/3}}\,\frac{x_2}{x_3}. \tag{14.6.11}$$

This modifies the result (14.6.9) by an amount easily evaluated and shown in Table 14.3 for 23 populations, along with (14.6.9) and (14.6.10).

The preceding pages offer a way of finding the relative influence x_2/x_3 of the second and third cumulants (or any other pair) on the real part of the first complex root. It can be used also to find the absolute effect on the root of any cumulant; this would consist of x_2 for the influence of the second moment, x_3 for that of the third, and so on. These methods and results can now be applied to finding the time to convergence.

14.6.5 Time to Convergence

What determines the time to convergence is not x alone but $\log \epsilon/(x-r)$, as in (14.6.1). We need to know, for comparing the second and third moments,

$$\frac{\dfrac{\partial[\log \epsilon/(x-r)]}{\partial \sigma^2}}{\dfrac{\partial[\log \epsilon/(x-r)]}{\partial \mu_3}} = \frac{-\dfrac{\log \epsilon}{(x-r)^2}\dfrac{\partial(x-r)}{\partial \sigma^2}}{-\dfrac{\log \epsilon}{(x-r)^2}\dfrac{\partial(x-r)}{\partial \mu_3}} \tag{14.6.12}$$

$$= \frac{\partial x/\partial \sigma^2 - \partial r/\partial \sigma^2}{\partial x/\partial \mu_3 - \partial r/\partial \mu_3} = \frac{x_2 - r_2}{x_3 - r_3},$$

in an obvious extension of the earlier notation.

The derivatives of r with respect to the moments are obtained from equation 14.6.2 involving r, that may be called $\psi = 0$; by differentiation we have

$$r_2 = \frac{-\psi_2}{-\psi_r} = \frac{r^2/2!}{\mu - r\sigma^2 + \mu_3(r^2/2!) - \cdots}$$

and

$$r_3 = \frac{-\psi_3}{-\psi_r} = \frac{r^3/6}{\mu - r\sigma^2 + \mu_3(r^2/2!) - \cdots};$$

and assembling these with what we found earlier for x_2 and x_3 gives

$$\frac{\dfrac{\partial[\log \epsilon/(x-r)]}{\partial \sigma^2}}{\dfrac{\partial[\log \epsilon/(x-r)]}{\partial \mu_3}} = \frac{x_2 - r_2}{x_3 - r_3} \tag{14.6.13}$$

$$= \frac{\psi_r(\phi_y\theta_2 - \phi_2\theta_y) - \psi_2(\phi_x\theta_y - \phi_y\theta_x)}{\psi_r(\phi_y\theta_3 - \phi_3\theta_y) - \psi_3(\phi_x\theta_y - \phi_y\theta_x)},$$

on substituting the previously obtained x_2, r_2, x_3, and r_3. This result is exact, but unfortunately it includes a score or more terms, even if we stop at squares in r and x and fourth powers in y. To make (14.6.13) tractable for calculation and thinking about, we need to discard smaller terms. The bare minimum appears to be

$$x_2 = \frac{\sigma^2 x/\mu + \frac{1}{2}}{\mu/y^2 + \sigma^4/\mu} \tag{14.6.14}$$

$$x_3 = \frac{\sigma^2 y^2/6\mu - x/2}{\mu/y^2 + \sigma^4/\mu} \tag{14.6.15}$$

$$r_2 = \frac{r^2}{2\mu} \tag{14.6.16}$$

$$r_3 = \frac{r^3}{6\mu}. \tag{14.6.17}$$

For a population with $r = 0.03$, $x = -0.03$, $y = 0.22$, $\mu = 27$, and $\sigma^2 = 35$, the result is

$$\frac{x_2 - r_2}{x_3 - r_3} = \frac{0.000764 - 0.000017}{0.0000422 - 0.0000002} = 17.8$$

against $x_2/x_3 = 18.1$. Apparently even a high rate of increase has little effect on the time to convergence.

14.6.6 Theoretical Versus Empirical Relations

The effect of each moment on the real part of the second root is also obtainable from empirical materials. But this, like all correlations taken from the real world, has a different meaning. It tells what the effect is *when other things vary as they vary.* The formulae of the preceding pages tell what the effect is *when other things are held constant.* Both kinds of information are useful for understanding, but they are not the same (cf. the distinction between prospective and retrospective perturbation analyses in Section 13.8. The difference between theoretical and empirical facts in demography is a main subject of Chapter 20).

Proceeding to the implications of instability in financial, economic, and social matters, we will discuss as a first aspect old-age security.

14.7 Retirement Pensions: Pay-As-You-Go Versus Actuarial Reserves

The cost of pensions to the wage-earner differs according to whether actuarial funding is used. On pay-as-you-go each year's pensions come out of the same year's receipts; the receipts are a tax, whose amount is in principle adjusted to the requirements of the scheme, that is to say, by the number of pensioners and the average amount of payment to them in relation to the number of contributors. Actuarial or reserve plans, on the other hand, are based on calculation of the expected cost to each individual; his premiums are cumulated at interest, and their amount is in principle just sufficient to cover his expected withdrawals from retirement to death. In the actuarial scheme, the expected costs of each individual are covered, which is another way of saying that each cohort pays for itself; in pay-as-you-go each period pays for itself. The two approaches can be compared for cost and equity.

In the stable condition, say with the population growing at rate r, a payment of unity to pensioners will cost

$$\int_{\beta}^{\omega} e^{-ra} l(a)\, da, \tag{14.7.1}$$

per current birth, where β is the age of retirement and ω the oldest age to which anyone lives. Contributions will be equal to

$$p \int_{\alpha}^{\beta} e^{-ra} l(a)\, da, \tag{14.7.2}$$

where α is the age at which uniform contribution begins, and p is the annual uniform payment by each person of working age, again per current birth. Equating (14.7.1) and (14.7.2) and solving for p:

$$p = \frac{\displaystyle\int_{\beta}^{\omega} e^{-ra} l(a)\, da}{\displaystyle\int_{\alpha}^{\beta} e^{-ra} l(a)\, da}, \tag{14.7.3}$$

and this is the premium for a simple case of pay-as-you-go, in accord with the argument of Section 5.7.

The above statements apply whether or not money bears interest. A pay-as-you-go scheme in principle holds no reserves that can accumulate interest, but dispenses what it takes in year by year.

As we saw in Section 2.5, an actuarial scheme is built on reserves that draw interest. If the rate is i per annum, compounded momently, so that the amount of one dollar at the end of t years is e^{it} dollars, then discounting a benefit of 1 dollar per year back to the moment of birth gives $\int_{\beta}^{\omega} e^{-ia} l(a)\, da$. If this amount is to be paid for by a premium of p dollars per year from ages α to β, the present value of the premium to the moment of birth must be $p \int_{\alpha}^{\beta} e^{-ia} l(a)\, da$. Equating the two present values gives for the annual premium p

$$p = \frac{\displaystyle\int_{\beta}^{\omega} e^{-ia} l(a)\, da}{\displaystyle\int_{\alpha}^{\beta} e^{-ia} l(a)\, da}. \tag{14.7.4}$$

Note that the two expressions, 14.7.3 and 14.7.4, for the annual cost of the pension are identical, with the sole difference that one contains the rate of interest i in the place where the other contains the rate of increase in the population. If r is greater than i, the average person will do better in a pay-as-you-go scheme. To prove this we need only note that p is a decreasing function to i. In fact, by taking logarithms of both sides of (14.7.4) and differentiating, we find

$$\frac{1}{p}\frac{dp}{di} = -(m_r - m_w),$$

where m_r is the mean age of the retired persons and m_w is the (necessarily younger) mean age of the working population. In finite terms

$$\frac{\Delta p}{p} \approx -(m_r - m_w)\Delta i; \tag{14.7.5}$$

and, if $m_r - m_w$ is about 30 years, each rise of 1 percent in the rate of interest will lower the premium by 30 percent of itself. This fact is very little affected by variations in the mortality level prevailing under modern conditions. Expression 14.7.5 is formally analogous to (5.7.4), though the two have different content.

When money carries a rate of at least 5 percent and population increases at 0 to 3 percent, why does anyone want a compulsory pay-as-you-go scheme? The answer is that the rate of interest people have in mind is in terms of goods, not money, and in times of inflation the two are not the same. What counts in the comparison of (14.7.3) and (14.7.4) is the increase in the amount of goods that one can have by temporarily foregoing the use of his money. If prospective price inflation is 6 percent per year, and money interest is 5 per cent per year, the goods rate of interest is negative. The decline of the real rate of bank and bond interest virtually to zero helps account for the popularity of pay-as-you-go schemes. A further reason is that the latter are more painless at initiation, when noncontributors can simply be blanketed in, in effect at the cost of a later generation.

Pay-as-you-go is largely proof against inflation, but has demographic troubles. These are suggested in Table 14.4, showing the ratio of working to pensionable ages over a century. In the United States the main pressure will come after the year 2000, with a rise of 50 percent over the years about 2020. The only thing that could prevent this is a large increase of births before the year 2000 that would raise the 21st century labor force, which seems unlikely.

The Social Security scheme can be seen as a way of borrowing from future generations, like the national debt. Besides lacking a contractual character, it differs from the national debt in being five times as large. Martin Feldstein (1976) shows that the scheme reduces private savings: people do not save as much because they are implicitly promised support by the next generation when they are old. But at the same time their smaller savings mean smaller investment than would otherwise occur, so the incomes of the next generation will be less than with private savings for retirement or an actuarial reserve scheme. Our children's having to pay larger benefits to us out of incomes that are smaller than they otherwise

Table 14.4. Persons of working and pensionable ages in the United States, 1950–2050

Year	Age 21–64	Age 65+	Percent 65+/21–64
1950	85,944	12,397	14.42
1960	92,181	16,675	18.09
1970	103,939	20,085	19.32
1980	122,115	24,523	20.08
1990	137,500	28,933	21.04
2000	148,589	30,600	20.59
2025	146,645	45,715	31.17
2050	147,635	45,805	31.03

Source: U.S. *Statistical Abstract*, 1975, p. 6.

would be because of our failure to save may be offset by our having paid for their education; on present trends expenditures on old-age security could substantially exceed those on education. If, however, as many think, our children will be much richer than we are, some borrowing from them is permissible.

But the subject of intergenerational equity is much too big for treatment here. All we need notice is that fluctuations in births are what bring the problem to the fore; under stable population growth most of these issues would not arise (Clark and Spengler 1976).

14.8 The Demography of Educational Organizations Under Changing Age Distributions

The 1950s and 1960s, with all their difficulties, attained a happy balance between graduate teaching and research. At the pinnacle of the academic profession was the faculty member of a major university, whose teaching was largely of graduate students and included or even centered on her own research. Her students were in effect apprentices who would be certified by doctorates, often at the rate of one every year or two, as qualified to take posts in other graduate schools, where they would be a prolongation of the master's research and teaching. One's weight in the university depended on her own discoveries and publications, as well as on former students who would teach her work and cite her publications. These former students would in due course replicate the process and send forth their graduates.

We know now that this process of reproduction cannot always be as rapid as in the 1960s. Even a much slower rate is out of the question in the long run. If the master is 40 years of age, and she turns out a student aged 25, and does the same again when she is 50 and a third time when she is

60, and these immediately go into graduate teaching and follow the same pattern, the population of graduate faculty is approximately trebling each 25 years. Its annual average rate of increase is obtained from the difference equation

$$B(t) = B(t-15) + B(t-25) + B(t-35);$$

to solve this we try $B(t) = e^{rt}$ and obtain the characteristic equation:

$$1 = e^{-15r} + e^{-25r} + e^{-35r}.$$

For an iterative solution we multiply through by e^{25r} and then take logarithms:

$$r^* = \tfrac{1}{25}\log(e^{10r} + 1 + e^{-10r}).$$

Starting with $r = 0$ on the right, we find for the successive values of r^* 0.0439, 0.0465, 0.0468, and 0.0468. A rate of 4.68 percent cannot be maintained over any long period of time. If the general population increases at 1 percent then the fraction who are graduate school teachers would go up over 35 times in a century.

Such pencil-and-paper calculations requiring no data could have been made when the graduate student boom was at its height and would have shown its temporary character. The basic demographic point is that, in the stationary condition that is inevitable sooner or later, a faculty member, in a graduate school or anywhere else, is on the average succeeded by one faculty member.

The system is very sensitive to the rate of increase. If it is expanding at 7 percent per year, and the difference in age between a teacher and his students is 20 years, each teacher is not restricted to one replacement, but will be succeeded by four tenured members on the average. This is so because 7 percent increase is a doubling in 10 years, or two doublings in 20 years. With the same rate of increase, if the time in nontenure is 6 years and in tenure 24 years, then the stable ratio of nontenure to tenure is

$$\int_0^6 e^{-0.07x}\,dx : \int_6^{30} e^{-0.07x}\,dx$$

or

$$1 - e^{(-0.07)(6)} : e^{(-0.07)(6)} - e^{(-0.07)(30)}$$

which is $1:1.56$ as against $1:4$ with the same age limits under stationarity, disregarding mortality while in service.

In public institutions where staffing is based on fixed student–teacher ratios, the slowing down of college entrants results in a sharp deceleration of faculty appointments. And what is true formally in public institutions is true practically in private ones—as their students stabilize or decline, so do tuition and other income, with inevitable consequences for faculty numbers. The greater difficulty of entering college teaching especially is felt by women and minorities, who are just beginning to qualify. To hire them and give them chances for promotion is far harder when the system is contracting than when it was expanding.

14.9 Two Levels of Students and Teachers

Extrapolation of the trend of the 1950s and 1960s could lead only to the conclusion that the entire population would attain the Ph.D. and become college teachers.

A very simple model, recognizing only two levels of education, can avoid such a nonsensical result (Correa 1967). Suppose that the two levels are college and graduate school, and the unit of time is 4 years. Let $S_t^{(1)}$ be the number of students at time t at the lower level, and $S_t^{(2)}$ the number of students at time t at the higher level; let $T_t^{(1)}$ and $T_t^{(2)}$ be the corresponding numbers of teachers at the two levels. If ρ_1 and ρ_2 are the ratios of teachers to students at the two levels, then two equations express the demand for teachers:

$$T_t^{(1)} = \rho_1 S_t^{(1)} \tag{14.9.1}$$

$$T_t^{(2)} = \rho_2 S_t^{(2)}. \tag{14.9.2}$$

If σ_1 is the fraction of the students in level 2 who will become teachers in level 1, and σ_2 the fraction of the students in level 2 who will become teachers in level 2, we have for the supply of teachers in period $t+1$:

$$T_{t+1}^{(1)} = (1-\mu_1)T_t^{(1)} + \sigma_1 S_t^{(2)} \tag{14.9.3}$$

$$T_{t+1}^{(2)} = (1-\mu_2)T_t^{(2)} + \sigma_2 S_t^{(2)}, \tag{14.9.4}$$

where μ_1 and μ_2 represent the loss of teachers through death, retirement, and other causes at the two levels. To complete the model, suppose that a fraction ϕ of the students at the lower level go on to the higher level:

$$S_{t+1}^{(2)} = \phi S_t^{(1)}. \tag{14.9.5}$$

If the unknowns are the students and teachers at the two levels, the system has four unknowns and five equations; it is overdetermined. The unknowns can be increased to 5 by supposing that any one of the seven Greek letter constants is unknown. Not all of the constants are equally likely candidates for this; we would say that σ_1 and σ_2, the fractions of Ph.D. students who become teachers, or else ϕ, the fractions of college students who go on to graduate school, should be the first to be examined. Alternatively, the system could be made solvable by disregarding one of the equations.

The system ought to be considered as dynamic: some of the variables drive the others. If the upper level does the driving, that is, if the number of Ph.D.s determines the number of teaching positions and the number of students at the lower level, so that the operative equations are (14.9.2) and (14.9.4), we have a first-order system, of which the solution is

$$T_t^{(2)} = \left(1 + \frac{\sigma_2}{\rho_2} - \mu_2\right)^t T_0^{(2)}. \tag{14.9.6}$$

The system expands steadily, without oscillations. In this unlikely mode the number of persons who want to attend graduate school and become teachers determines the number of undergraduates. It is worth mentioning here only because it is implicit in some of the educational perspectives of the 1960s.

Taking the more realistic condition that the students at the lower level determine the system, we find that cycles in the number of college students give rise to amplified cycles in the number of new teachers demanded. Any reasonable way of making the system determine will show that the *absolute demand* for Ph.D.s is closely related to the *increase* in the number of college students (Correa, 1967).

The model enables us to follow what happens in the wake of a baby boom. About 17 years after the rise in births comes a rise in the number of college applicants. Every effort is made to satisfy this demand for entrance: teachers are asked to defer retirement; somewhat less qualified persons are given appointments as teachers; newly established colleges take some of the new students and hire as teachers persons who would no earlier have aspired to this occupation. In our model the increase of applicants acts initially on (14.9.1), resulting partly in an immediate expansion of $T_t^{(1)}$, and partly in allowing ρ_1, the fraction of teacher per student to become smaller, which is the same as class sizes becoming larger. Somehow the demand constituted by the rise in applicants for college entrance is *immediately* met; applicants are not ordinarily asked to come back in 8 or 10 years, which could make the system more stable.

The college teacher shortage at this stage is intense, and graduate schools receive applications from many who in earlier times would have been satisfied to leave school with a B.A. The graduate schools, like the applicants, perceive a strong demand for Ph.D.s and accept more than they otherwise would. Even if ϕ of (14.9.5) were fixed, there would be an increase in the number of recruits to graduate school the next period; but in fact ϕ increases in such times, and the graduate schools expand in higher ratios than the increase in the number of undergraduates.

Two periods later, that is to say, 8 years after the baby boom has hit the college level, the first Ph.D.s of the new wave are available. They are seeking jobs in colleges, and particularly in graduate schools, having oriented themselves to the institutions of highest prestige. The supply of graduate school teachers is now given by (14.9.4) with a high $S_t^{(2)}$. The μ, coefficients of death and retirement, are small over the short period here involved. The timing in this crude representation involves the 17 years from the start of the baby boom to college entry of the larger cohorts, plus the 8 years through college and graduate school, a total of 25 years from the first rise in births to the start of an employment crisis for Ph.D.s. Twenty-five years is about the interval from the later 1940s to the early 1970s.

14.10 Mobility in an Unstable Population

The mobility model of Section 5.8 exhibited the demographic factor in promotion, finding it faster for members of populations in rapid growth than for members of stationary populations. A person will get to a middle-level position about 9 years earlier in the fastest growing population than in a stationary one. Mortality also helps (at least for those who survive) but much less; a very high mortality population ($\overset{o}{e}_0 = 35$ years) will advance its surviving members to middle positions only 2 to 3 years sooner than the lowest mortality population known (say $\overset{o}{e}_0 = 75$ years). High subsequent birth rates advance a person's promotion more than do high death rates.

The above results are comparative rather than dynamic—they compare the age of attaining a given level in a fast- and in a slow-growing population, but supposing for each that its rates have been fixed and continuing over a long period. A separate question concerns the slowing down of an actual population: what difference does it make to individuals now alive that the United States was growing at 1.5 percent per year in the late 1950s and is dropping from this rate toward stationarity? Stable population theory, assuming as it does fixed rates, cannot by its very nature answer a question about the effect of change in rates over time.

As an introduction to the problem consider the United States age distribution, as of the mid-1970s, in the stylized version of Figure 14.3. Those approaching retirement at that time were members of the large cohorts born in the 1920s and earlier. The working population then in its forties consisted of the very small cohorts born in the 1930s, and new entrants up to 1975 had been larger, steadily increasing toward the left in Figure 14.3, up to those born about 1960. Following that came a downturn, so that the new entrants of the 1980s were fewer and decreasing.

Those in their forties at the time of this diagram were fortunate in two distinct ways. They were called on to replace the large number of their elders retiring from senior posts, and they had been drawn upward by the large number of new entrants who were younger than themselves. Supervisors and teachers are required in relation to workers and students; the tests of competence for supervisors and teachers are less stringent in times when more are required.

Thus those born in the 1930s, having been pushed upward by the young people coming in behind them, could look forward to further promotion

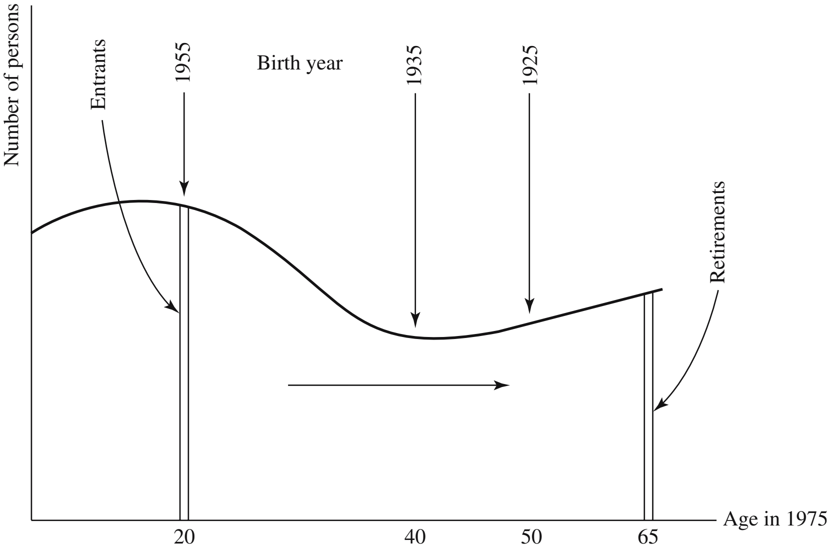


Figure 14.3. Stylized representation of the supply of labor in the United States.

as their seniors retired. However, those born at the peak of the baby boom in the 1950s and early 1960s would never in their later careers be in such great demand as supervisors and teachers, for they were followed by sparse cohorts, nor were they drawn up as strongly to fill posts vacated by retirement, because these were less numerous.

14.11 The Easterlin Effect

The high United States fertility of the 1950s and the subsequent decline have puzzled observers who believe that there are strong economic determinants of age-specific fertility. The phenomenon could not be due to prosperity, since income in the 1950s was not as high as in the 1960s. That children are positively related to income in time series did not seem to help here, until Easterlin (1961, 1968, 1980) noted that the prosperity of couples of childbearing age is what we should look for, not general prosperity. He pointed out that couples of childbearing age in the 1950s, born between the humps of the 1920s and the 1940s, were located in a hollow similar to that of Figure 14.3. Not only was their promotion relatively rapid, but also in any one position, insofar as there is age complementarity in production, they frequently had the advantage of meeting situations in which they were too few to do the necessary work, with resulting appreciation of their services. This was often expressed in material terms and resulted in high wages and good prospects relative to what people of their age would have been paid in a different age configuration; hence these people have enjoyed a sense of security and well-being. Their confidence is well founded, for they will spend their whole working lives in the same advantageous position. They have translated their advantage into childbearing, perhaps projecting their security into the next generation and feeling that their children will be in demand just as they are. Macunovich (1998) has compiled a detailed review of studies on the fertility patterns predicted by Easterlin's hypothesis.

So strong was this effect in the 1950s that it entirely reversed the tendency of the classical model with fixed age-specific rates of Chapter 6. Instead of a dip, echoing that of the 1930s, the 1950s showed births at the highest level in half a century. The 1930s gave a relative advantage to their children by producing few of them, and these later repaid the advantage by having many children. Such a mechanism could produce a very stable result if the rise in birth rates was of just the right amount to compensate for the few parents. In the actual case, however, the rise overcompensated.

The subsequent steady fall of the birth rate in the 1960s might well have been due to the entry into childbearing ages of the large cohorts born in the 1940s. If this was the dominant mechanism operating, we can predict continued low birth rates at least through the 1970s. Not until about 1990 will the parental generation again be small enough to be encouraged to have large families. Instead of the waves of generation length in the free response of the demographic model we find waves two generations in length.

To translate this into quantitative terms we consider females only, and simply suppose that all children are born at the same age of parent, say 25 years or the mean length of generation. If the "normal" female growth ratio is R_0 per generation, births in the tth generation are $B_t = R_0 B_{t-1}$. The conditions imply geometric increase and lead to the solution $B_t = B_0 R_0^t$.

Suppose now that superimposed on this is a component of births in the tth generation that depends on how small is the number of couples aged 25 compared with the number of couples aged 50. When the couples aged 25 are relatively few, that is to say, when B_{t-1} is less than its "normal," suppose an additional $\gamma(R_0 B_{t-2} - B_{t-1})$ births in the t generation, γ positive. These conditions mean that the sizes of cohorts are related by

$$B_t = R_0 B_{t-1} + \gamma \left(R_0 B_{t-2} - B_{t-1}\right), \tag{14.11.1}$$

where R_0 is no longer necessarily the net reproduction rate.

To solve (14.11.1) we try $B_t = x^t$ and find

$$x^2 - (R_0 - \gamma)\, x - \gamma R_0 = 0, \tag{14.11.2}$$

of which the roots are conveniently $x = R_0$ and $-\gamma$. Thus the solution of (14.11.1) is

$$B_t = k_1 R_0^t + k_2(-\gamma)^t, \tag{14.11.3}$$

where k_1 and k_2 depend on the initial conditions. The term in $(-\gamma)^t$ tells us that cycles will be generated two generations in length. If the factor for normal increase R_0 is greater than γ, any waves once started will diminish as a proportion of the geometric increase. If γ is less than unity, they will diminish absolutely.

Lee (1968) applied Easterlin's theory in a much more refined model, one recognizing individual age groups. He also traced out its historical antecedents, starting with a quotation from Yule (1906):

> If ... the supply of labour be above the optimum, the supply of labour being in excess, the birth rate will be depressed, and will stay depressed until the reduction begins to have some effect on the labour market. But this effect will not even commence for fifteen or twenty years, and the labour supply may not be adjusted to the demand for, say, thirty years. The birth rate may now have risen again to normal, but the labour supply will continue to fall owing to the low birth rates formerly prevalent. The birth rate will therefore rise above normal and continue above normal so long as the labour supply is in defect, and so matters will go on, the population swinging about the optimum value with a long period of perhaps fifty to one hundred years, and the birth rate following suit.

The germ of this idea is to be found as far back as Adam Smith, who considered that the supply of labor, like the supply of shoes, was determined by demand. Smith did not go the one further step of recognizing that the period of production of labor is longer than the period of production of shoes, with resulting longer and deeper cycles.

The Easterlin hypothesis is an example of a *nonlinear* demographic model, in which the vital rates are functions of the population itself ("density-dependence" in ecological terminology; see Lee (1987) for a detailed discussion of density effects in human demography). Such models can generate instability in the form of nonstationary solutions: cycles, quasi-periodicity, and chaos. Such analyses are beyond the scope of this book, but general discussions can be found in Cushing (1998) and Chapter 16 of MPM. Nonlinear phenomena, many of them remarkably subtle, have been beautifully documented in careful laboratory experiments on populations of the flour beetles of the genus *Tribolium* (Cushing et al. 2003).

In the case of human populations, a series of analyses followed Lee's (1974) study of the Easterlin effect (Frauenthal and Swick 1983, Wachter and Lee 1989, Wachter 1991, Chu 1998). Cyclic dynamics are clearly possible from the model, but whether parameters estimated from United States demographic data can produce them is not clear.

Some apology is needed for the heterogeneous material contained in the present chapter, which reflects the fact that populations can be unstable in many different ways. To write about stable theory in a coherent fashion is bound to be more straightforward than to attempt to write about everything that is not stable theory. Furthermore, the range of applications here attempted or recounted is especially wide, including estimating the rate of increase of a population when its death rate or its birth rate is falling, backward projection under instability, finding the time to stability after a disturbance, old-age pensions under reserve and nonreserve systems, and the consequences for the educational system, the labor market, and the birth rate of the baby boom and its aftermath.

15

The Demographic Theory of Kinship

This chapter will extract information on kinship numbers from the age-specific rates of birth and death of a population. A fixed set of age-specific rates implies the probability that a girl aged a has a living mother and great-grandmother, as well as her expected number of daughters, sisters, aunts, nieces, and cousins. Certain assumptions are required to draw the implications, some stronger than others. The formulae of this chapter in effect set up a genealogical table, giving not the names of incumbents in the several positions but the expected number of incumbents. Those of Figure 15.1 are based on birth and death rates of the United States in 1965, whose net reproduction rate R_0 was 1.395 and $\overset{o}{e}_0$ was 73.829, all for females. They offer a different kind of knowledge from what would be provided by a kinship census.

Like earlier chapters, this one supposes a population generated by birth and death with overlapping generations. (Generations do not overlap in annual plants, where all the parents have disappeared before the children come to life. This circumstance requires population models that account for processes both within and between years; see MPM Section 13.2.) The considerable longevity of human beings, as well as other large mammals and long-lived birds, after the birth of their offspring produces simultaneously living kin of many kinds—not only parents and children, but also grandparents, nephews, and cousins. Human beings produce most of their offspring in births of discrete individuals, but this is not recognized in the present analysis, which supposes $m(x)\,dx$ of a daughter in each dx of maternal age. For certain kin this introduces serious qualifications, specified in Section 15.3. Development of this field is due to Lotka (1931), Burch

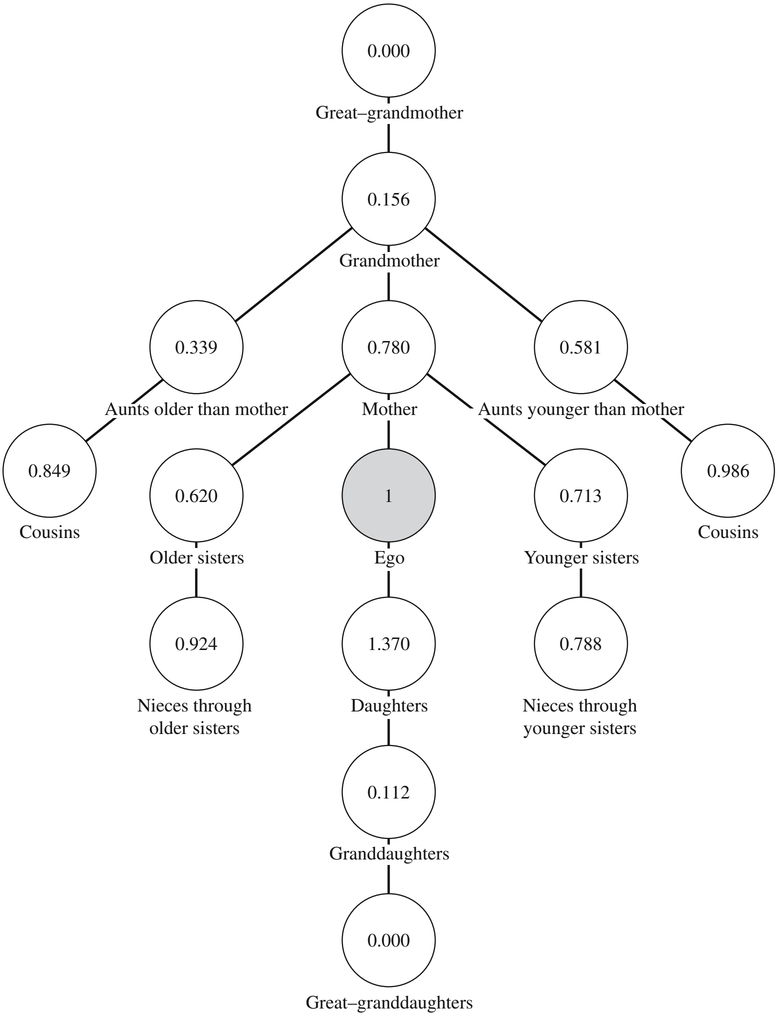


Figure 15.1. Expected number of female kin alive when Ego (hatched circle) is aged 40, based on birth and death rates of the United States, 1965.

(1970), Coale (1965), Goodman, Keyfitz, and Pullum (1974) and Le Bras (1973).

Explicit recognition of the several degrees of living and dead kin varies from one culture to another, and indeed from one family to another. We disregard here cultural, social, and psychological definitions and deal with numerical relations among average numbers of biological kin as they are determined by birth and death rates. To avoid undue complication, all of the following discussion recognizes female kin only.

15.1 Probability of Living Ancestors

Deterministic models concern both population numbers and probabilities. The two perspectives are at least on the surface distinct.

Counting Method. A large population can be seen as developing according to given rules, and in effect we can make counts of the number of individuals having the kin relations of interest. This is an extension of the notion that l_0 is the number of births occurring at one moment and $l(x)$ is the number of those surviving x years later, the cohort conception of the life table $l(x)$ referred to in Section 2.1. (But we still keep $l_0 = 1$.)

Probability Method. We can start by thinking of an individual and work out probabilities and expected values for his various kin. This is an extension of the interpretation of the life table $l(x)$ as the probability that a child just born will survive for x years.

15.1.1 Living Mother by the Counting Method

Our first approach to finding the probability that a girl aged a has a living mother is to see how a population would have developed starting from B girl children born at a moment $a + x$ years ago. At age x of this maternal generation cohort the survivors were $Bl(x)$, and during the interval x to $x + dx$ they could be expected to give birth to $Bl(x)m(x)l(a)\,dx$ daughters who would live to age a. Of the mothers who gave birth at age x a fraction $l(x + a)/l(x)$ would survive over an additional a years; hence the number of living mothers must be $Bl(x)m(x)\,dx[l(x + a)/l(x)]l(a)$. A woman is counted once for each birth that survives.

All this concerns one cohort of the mother generation. But we seek the probability that a girl aged a, standing before us, has a living mother, without any knowledge of which cohort her mother belonged to, or indeed any knowledge other than the regime of mortality and fertility supposed to apply to all generations and at all times. If births as a function of time are $B(t)$, and the present is time t, girls born $x + a$ years ago numbered $B(t - a - x)$, and the number of living mothers (counted once for each daughter) of all cohorts who gave birth to girls now alive and a years of age is the integral

$$\int_{\alpha}^{\beta} B(t - a - x)l(x)m(x)\frac{l(x + a)}{l(x)}l(a)\,dx. \tag{15.1.1}$$

For the same birth function $B(t)$ the number of daughters, that is, girls born a years ago and surviving to the present, is

$$B(t - a)l(a). \tag{15.1.2}$$

Hence the average number of mothers per daughter at time t is the ratio of (15.1.1) to (15.1.2).

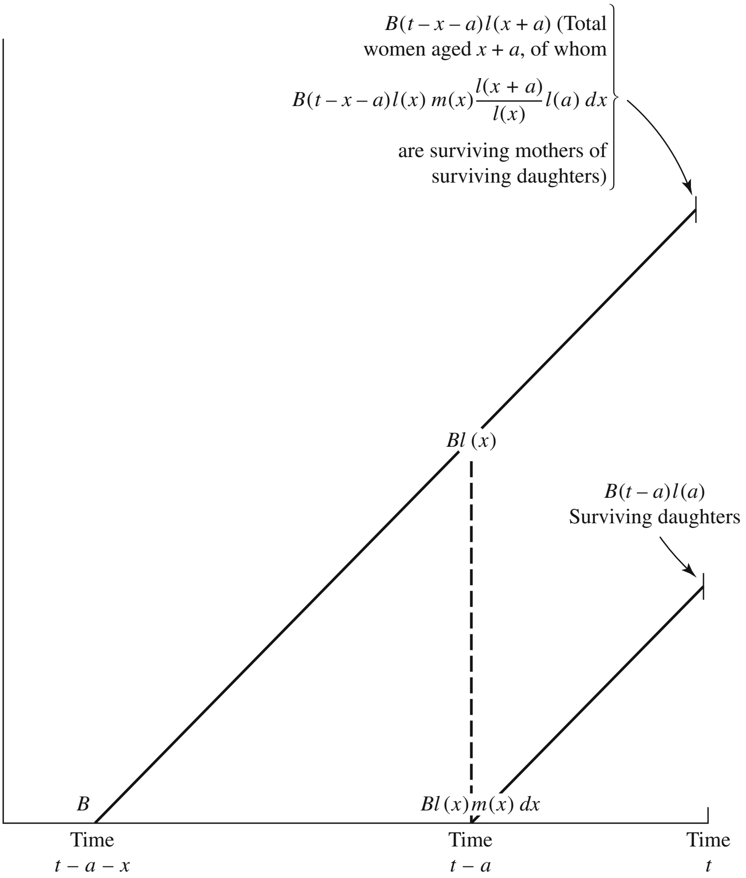


Figure 15.2. Cohort of mothers giving birth at age x, and of daughters born at age x of mother.

Now, if the age-specific rates of birth and death have been in effect for a long period of time, the births will be growing exponentially, say according to the curve $B_0 e^{rt}$, where r is determined by (6.1.2). Entering this for the birth function in each of (15.1.1) and (15.1.2), taking the ratio of the first to the second, and canceling $l(a)B_0 e^{r(t-a)}$ from numerator and denominator gives

$$M_1(a) = \int_\alpha^\beta \frac{l(x+a)}{l(x)} e^{-rx} l(x)m(x)\,dx \tag{15.1.3}$$

for the probability that a girl aged a has a living mother under the given regime of mortality and fertility (Lotka 1931).

All of the problems of this chapter can be solved in this way, by following cohorts through time, then in effect taking a census at a certain moment, and finding the ratio of one census aggregate to another. This does not of course constitute a census of the real population, which would show the result of changing death rates over time, but is at best a simplified abstract argument devised by analogy to counting population by means of a census and births by means of registrations. Made possible by the deterministic assumption of Section 2.1, it requires no explicit considerations of probability, yet it is both unnecessarily complicated and unnecessarily restricted.

15.1.2 Living Mother by Conditional Probability

Alternatively, the life table $l(x)$ column is taken, not as a cohort, but as the probability of living to age x for a child just born, on the regime of mortality assumed, with $l_0 = 1$. The corresponding approach will provide a result for $M_1(a)$ identical with (15.1.3).

In seeking the probability that a girl chosen at random out of a population with birth rates $m(x)$ and survivorship $l(x)$ has a living mother, we first note that the conditional case is easily solved. *If* the mother was a known x years old when she gave birth to the girl, *then* the chance that the mother is alive a years later must be $l(x + a)/l(x)$. This probability, conditional on the mother's age at bearing the girl having been x, is the first part of the solution.

It remains to remove the condition, which is not part of the problem; we do not care about the age x of the mother at childbearing. To eliminate the condition we average over all ages x, giving each x a weight proportional to the number of births occurring at that age of mother under the regime in question. The number of the stable population of ages x to $x + dx$ per current birth is $e^{-rx}l(x)\,dx$, from Euler's argument of Section 5.1, and the births to this fraction are $e^{-rx}l(x)m(x)\,dx$, still taken per one current birth. The last expression is the fraction of births occurring to women aged x to $x + dx$; once again, its total over all x is unity by the (6.1.2) defining r. Hence the unconditional probability that the mother of the girl in question is still alive is obtained by multiplying $e^{-rx}l(x)m(x)\,dx$ by the survivorship $l(x + a)/l(x)$ and totaling over all x.

This is the same expression, derived more compactly, that we obtained as (15.1.3). Because of its compactness the probability method will be preferred in what follows. Note that (15.1.3) is more general than appeared in our interpretation. The derivation did not require the fact that the girl born a years ago is still alive, but consisted in finding the probability of survival of the mother a years after a random birth. Whether the girl born survived does not affect the value of $M_1(a)$, given independent regimes.

15.1.3 Probability of Living Grandmother

Now think of the grandmother of the girl aged a; we once again provisionally take the latter as having been born at age x of her mother. The grandmother has to live $x + a$ years after the birth of her daughter (the mother of our girl aged a) to be alive now. The chance of a woman living $x+a$ years after the birth of her daughter as calculated above is $M_1(x + a)$. This is now to be seen as the (conditional) probability that the grandmother of the girl aged a is alive, given that the girl aged a was born when her mother was x years old. To obtain the unconditional probability we again multiply by $e^{-rx}l(x)m(x)$ and integrate out x:

$$M_2(a) = \int_{\alpha}^{\beta} M_1(x + a)e^{-rx}l(x)m(x)\, dx. \tag{15.1.4}$$

Note that (15.1.4) again does not require the survival of the mother of the girl aged a. It is merely the probability that the grandmother of a randomly selected girl birth of a years ago is now alive.

The procedure supposes not only that the regime is fixed, but also that successive generations are independently subject to the given mortality and fertility. Independence of fertility between generations means, for example, that there is no tendency for daughters born to young mothers themselves to give birth at a young age, and it also disregards social class and other differences in mortality and fertility within heterogeneous populations. This is in addition to requiring birth and death to be independent, both in the same and in different generations; it excludes the possibility that some families have both high birth rates and high death rates. Finally, the chance of a woman having a child at ages x to $x+dx$ is taken as $m(x)\, dx$, whether or not she had a child immediately before this. These several unrealistic features of the model seem to have only a small effect on the probability of living ancestors and descendants, though they are important for the number of sisters, aunts, and nieces.

Once we have $M_2(a)$, the chance that a girl aged a has a living grandmother, we can similarly use it to find the chance of a living great-grandmother. Again suppose that the mother of the girl aged a was x years old at the time of childbearing; the question whether a girl aged a has a living great-grandmother is the probability that the grandmother of the mother, born $x + a$ years ago, is still alive. That the grandmother survives $x + a$ years after her granddaughter was born has probability $M_2(x + a)$, so the unconditional probability that a girl aged a has a living great-grandmother is

$$M_3(a) = \int_{\alpha}^{\beta} M_2(x + a)e^{-rx}l(x)m(x)\, dx. \tag{15.1.5}$$

Once more, this does not require the mother or grandmother of the girl aged a to be alive. The argument may be readily extended to even more remote progenitors (Goodman et al. 1974).

15.1.4 Numerical Examples

These and the other formulae of this chapter have been programmed by Tom Pullum, and Table 15.1 shows his results for three countries—strictly speaking, for three regimes of mortality and fertility—the United States, 1967; Venezuela, 1965; and Madagascar, 1966. The first two resemble each other in mortality and the last two in fertility, as the following standardized rates per thousand population having the United States, 1960, age distribution show:

Country and year	Birth rate	Death rate
United States, 1967	16.71	9.12
Venezuela, 1965	41.82	10.97
Madagascar, 1966	44.48	29.10

Thus Venezuela and Madagascar both have about 2.5 times the fertility of the United States, and Madagascar has about 3 times the mortality of the United States and Venezuela. We will later seek a more precise way of connecting the input mortality and fertility with the output kinship probabilities, but the present comparison is suggestive.

Table 15.1 shows, for example, that the chance that a woman aged 20 has a living mother is about 0.96 for the United States, 0.93 for Venezuela, and 0.71 for Madagascar. The complements of these numbers, 0.04, 0.07, and 0.29, are the probability of orphanhood on the mother's side. It was in an effort to see how serious was the problem of orphanhood that Lotka (1931) first developed (15.1.3). The greater difference between Venezuela and Madagascar than between the United States and Venezuela is to be expected; the chance of having living ancestors depends much more on mortality than on fertility rates. Insofar as fertility affects orphanhood, it is through the age of childbearing rather than through the number of children born, as will appear in Section 15.6.

15.1.5 Stable Results Versus a Kinship Census

These formulae and the numbers of Table 15.1 have been worked out for the specified regimes of mortality and fertility, taken as fixed through time and the same in all generations. They are meant to answer the question: what probability of having a living mother, grandmother, and so on does the given schedule of birth and death rates imply?

The fraction of women aged 20 in the United States having living mothers as ascertained by a survey or census would disagree with the result of calculation by (15.1.3) for several reasons: changing mortality and fertility over the preceding years, presence of immigrants from countries with different regimes, misstatement of age in the survey and in the vital statistics on which our calculations are based, or failure of the various independence

Table 15.1. Probability of living mother, grandmother, great-grandmother, and great-great-grandmother, for a female aged $a = 0$, 20, 40, 60, based on mortality and fertility regimes of the United States, Venezuela, and Madagascar

Ancestor and country	Age a 0	20	40	60
Living mother $M_1(a)$				
United States, 1967	1.000	0.959	0.785	0.298
Venezuela, 1965	1.000	0.932	0.707	0.223
Madagascar, 1966	1.000	0.713	0.386	0.061
Living grandmother $M_2(a)$				
United States, 1967	0.919	0.653	0.165	0.000
Venezuela, 1965	0.867	0.553	0.112	0.000
Madagascar, 1966	0.600	0.256	0.032	0.000
Living great-grandmother $M_3(a)$				
United States, 1967	0.507	0.090	0.000	0.000
Venezuela, 1965	0.397	0.058	0.000	0.000
Madagascar, 1966	0.164	0.017	0.000	0.000
Living great-great-grandmother $M_4(a)$				
United States, 1967	0.049	0.000	0.000	0.000
Venezuela, 1965	0.030	0.000	0.000	0.000
Madagascar, 1966	0.009	0.000	0.000	0.000

assumptions. Probability of living grandmother $M_2(a)$ requires independence of two generations, and $M_3(a)$ of three generations. We know that longevity runs in families, as well as being different for social classes, and experimenting would be required to find the effect of the mortality correlation between generations. All the formulae would become much more complicated if they took account of such departures from the assumptions of fixity and independence of the vital rates. One can only repeat that the kinship implication of fixed and independent rates constitutes a different kind of knowledge from a kinship census.

15.1.6 An Approximation

Insofar as the net maternity function is concentrated close to the mean age of childbearing κ, the quantity $l(\kappa+a)/l(\kappa)$ ought to be an approximation to our $M_1(a)$. For United States, 1967, females with $a = 20$ this would be $l(46.281)/l(26.281)$. The life table shows $l(x+20)/l(x)$ for $x = 25$ as 0.96724, and for $x = 30$ as 0.95122; straight-line interpolation between these values gives $l(46.281)/l(26.281) = 0.963$. This compares with the more precisely calculated $M_1(20) = 0.959$ of Table 15.1; the approximation $l(\kappa + 20)/l(\kappa)$ is slightly high, because the curve of $l(x)$ is at this point concave below.

In the same way we would expect that the chance of a living grandmother would be something like $l(2\kappa + a)/l(\kappa)$. For United States women aged 20 the value is

$$\frac{l(2\kappa + 20)}{l(\kappa)} = \frac{l(72.562)}{l(26.281)} = \frac{0.65431}{0.96734} = 0.676,$$

against the $M_2(20) = 0.653$ in Table 15.1, or about 4 percent high. The greater error for grandmothers than for mothers is due to greater variation in age for the former.

Most of the difference between the crude and the correct estimate is accounted for by the concavity of the survivorship curve, along with the variance of ages of childbearing. To establish this, expand the ratio $l(x + a)/l(x)$ in $M_1(a)$ of (15.1.3) around κ by Taylor's theorem, writing the derivative of $l(x + a)/l(x)$ at $x = \kappa$ as $[l(\kappa + a)/l(\kappa)]'$, and so on:

$$\begin{aligned}\frac{l(x+a)}{l(x)} = \frac{l(\kappa+a)}{l(\kappa)} + (x-\kappa)\left[\frac{l(\kappa+a)}{l(\kappa)}\right]' \\ + \frac{(x-\kappa)^2}{2!}\left[\frac{l(\kappa+a)}{k(\kappa)}\right]'' + \cdots;\end{aligned} \qquad (15.1.6)$$

integrate over x after multiplying by $e^{-rx}l(x)m(x)$, so that the term in $x - \kappa$ vanishes; call the variance of ages of mothers σ^2; and then factor out $l(\kappa + a)/l(\kappa)$ to obtain

$$M_1(a) \approx \frac{l(\kappa+a)}{l(\kappa)}\left[1 + \frac{\dfrac{\sigma^2}{2}\left(\dfrac{l(\kappa+a)}{l(\kappa)}\right)''}{\dfrac{l(\kappa+a)}{l(\kappa)}}\right]. \qquad (15.1.7)$$

The correction in square brackets equals 0.9957 for United States females of 1967, and produces $(0.9631)(0.9957) = 0.9590$ for $M_1(20)$ versus 0.9594 from the printout on which Table 15.1 was based.

For grandmothers the square bracket of (15.1.7) for the correction holds approximately, but with 2κ in place of κ in each numerator. The correction then is 0.963, and multiplying by $l(2\kappa + 20)/l(\kappa)$ gives 0.651, versus the true 0.653.

Presenting $M_1(a)$ in the form of (15.1.7) serves to show what feature of the net maternity function mainly determines the probability of a living grandmother: the mean age of childbearing κ acting through the factor $l(\kappa + a)/l(\kappa)$. The total of the net maternity function is absent, so to the (fairly close) approximation provided by (15.1.7) the level of fertility has little effect on probability of living ancestors. The variance of ages of childbearing has a small effect, whose amount depends mostly on the second derivative (i.e., the curvature) of $l(\kappa + a)/l(\kappa)$. The $l(x)$ curve is for the most part concave downward up to about age 65, so that its second derivative is negative. That the survival of mothers is negatively related to the variance of ages of childbearing wherever $l(\kappa+a)/l(\kappa)$ is concave downward is the conclusion from (15.1.7), but the effect is small.

Further conclusions from (15.1.7) are drawn in Section 15.6.

15.2 Descendants

To illustrate how this chapter is an extension of standard demographic techniques the familiar net reproduction rate R_0 will be put into a form suitable for the counting of descendants.

If a cohort of girl births numbers B, born at time zero, the number of survivors to age x will be $Bl(x)$ on the deterministic model, and in the interval x to $x+dx$ these will bear $Bl(x)m(x)\,dx$ girl babies. The total number of daughters to which the cohort will give birth during its existence will be the integral of this last expression over x, and the average number of daughters will be this integral divided by the girl births B:

$$\begin{aligned} R_0 &= \frac{\displaystyle\int_\alpha^\beta Bl(x)m(x)\,dx}{B} \\ &= \int_\alpha^\beta l(x)m(x)\,dx. \end{aligned} \tag{15.2.1}$$

This, the expected number of girl children to which a girl child will give birth under the regime $l(x)m(x)$, may be regarded as the ratio of one generation to the preceding (cf. Section 11.3.4).

For the number of granddaughters we use (15.2.1) in relation to each of the daughters. Thus, if the average number of girl babies at age x to the mother cohort is expected to be $Bl(x)m(x)\,dx$, and if each of these is expected to have R_0 births, we multiply by R_0 and again integrate up to age β, now to find R_0^2 granddaughters. Similarly the average number of great-granddaughters expected by a girl child will be R_0^3.

For incomplete generations the multiple integrals are formed in the same way, but β is no longer the upper limit. If we want only girl children that will have been born to a female by the time she is age a, where $\beta \geqslant a \geqslant \alpha$, the argument gives an expected

$$\int_\alpha^a l(x)m(x)\,dx \tag{15.2.2}$$

girl children. The expected number of granddaughters by the time the original cohort is aged $a \geqslant 2\alpha$ can be obtained by noting that $l(x)m(x)\,dx$ daughters would be expected to have been born when the woman aged a was $x \geqslant \alpha$ years of age; since each of these has up to age $a-x$ in which to bear children, each would be expected to bear

$$\int_{\alpha}^{a-x} l(y)m(y)\,dy$$

daughters in turn. Thus the total number of granddaughters will be the product of this and (15.2.2) added through the possible x:

$$\int_{\alpha}^{a} l(x)m(x) \int_{\alpha}^{a-x} l(y)m(y)\,dy\,dx. \tag{15.2.3}$$

All the above concern prospective descendants. Now we want to find the average number of girl children that have already been born to women aged a. The same cohort of B births has been followed down from time zero, and by the time it reaches age a there are $Bl(a)$ survivors. The total number of children that have been born up to this time is the integral $\int_{\alpha}^{a} Bl(x)m(x)\,dx$, but we do not want to include all of them in our average—we are concerned only with those born to mothers that lived at least to age a. The fraction of mothers that lived from age x to age a is $l(a)/l(x)$, and this is the fraction that we will take of the daughters born at age x of the mothers. Hence we have, for the total daughters expected to be born to the mothers that survived to age a,

$$\int_{\alpha}^{a} \frac{l(a)}{l(x)} Bl(x)m(x)\,dx = \int_{\alpha}^{a} Bl(a)m(x)\,dx,$$

and on dividing this by $Bl(a)$ mothers living at age a we have the average number of such daughters:

$$B_1(a) = \int_{\alpha}^{a} m(x)\,dx. \tag{15.2.4}$$

This result is obvious when we consider that any woman alive at age $a > \alpha$ was also alive at age $x \leqslant a$, and that her probability of bearing a daughter in the interval x to $x + dx$ was $m(x)\,dx$. Her total female births to age a must therefore be given by (15.2.4), in which the $l(x)$ function does not enter. One minor difficulty is our assumption, unavoidable if existing fertility tables are to be used, that the $m(x)$ function is the same for women who survived to age a as for all women.

How many of the $\int_{\alpha}^{a} m(x)\,dx$ daughters will still be alive by the time the mother cohort is aged a? The fraction of daughters born at age x of their mothers that survive to age a of their mothers, or $a - x$ years, must be $l(a - x)$. Hence the number of daughters still alive of women aged a must on the average be

$$BL_1(a) = \int_{\alpha}^{a} m(x)l(a - x)\,dx. \tag{15.2.5}$$

We can build on these results to find expected granddaughters already born for a woman aged a. Consider a daughter born at age x of the original cohort. By the time the original cohort is aged a, the daughter herself will have averaged $\int_\alpha^{a-x} l(y)m(y)\,dy$ daughters as in (15.2.3). Integrating over all daughters born to the original cohort gives us the double integral

$$B_2(a) = \int_\alpha^a m(x) \int_\alpha^{a-x} l(y)m(y)\,dy\,dx \tag{15.2.6}$$

for the average number of granddaughters so far born to women aged a.

To find the number of such granddaughters who are still alive we must multiply within the inner integral by the chance of survival through the years to the time when the original cohort is aged a (i.e., $a - x - y$ years), that is, by the factor $l(a - x - y)$. Great-granddaughters and further direct descendants raise no new problem.

15.3 Sisters and Aunts

To find the number of older sisters that a girl now aged a is expected to have, we again set the provisional condition that she was born at age x of her mother, when according to (15.2.4) her mother would be expected to have had $\int_\alpha^x m(y)\,dy$ children. The condition on x is removed as before by multiplying by $e^{-rx}l(x)m(x)\,dx$ and then integrating over x. Thus the expected number of older sisters, say S^{old}, still alive or not, is

$$S^{\text{old}} = \int_\alpha^\beta \int_\alpha^x m(y)e^{-rx}l(x)m(x)\,dy\,dx. \tag{15.3.1}$$

Because a mother aged x was necessarily alive at ages y younger than x, we need no allowance for survivorship of the mother. Note that S^{old} does not depend on the age a; a girl can hardly acquire additional older sisters as she ages, and older sisters once born cannot decrease. Some of them, however, are no longer living; to find older sisters now alive we need only include the factor $l(a + x - y)$ in the inner integral of (15.3.1), and this makes the double integral a function of a, the age of the girl with whom the calculation starts.

Younger sisters ever born require an allowance for survivorship beyond the birth of the girl aged a, and their number must depend on a. If the girl aged a was born when her mother was x years old, the chance that the mother lived on to age $x + u$ and then bore a child is $[l(x + u)/l(x)]m(x + u)\,du$; integrating this over the possible values of u, then multiplying by $e^{-rx}l(x)m(x)\,dx$, and again integrating gives

$$S_a^{\text{young}} = \int_\alpha^\beta \int_0^a \left[\frac{l(x+u)}{l(x)}\right] m(x+u)\,du\,e^{-rx}l(x)m(x)\,dx \tag{15.3.2}$$

for the number of younger sisters. Again this result may be interpreted more generally as the number of girls expected to be born to the mother of a random birth in the a years following.

15.3.1 A Paradox: The Average Girl Seems to Have Too Many Sisters

We find that the average number of sisters ever born according to the United States regime of mortality and fertility was 1.251 for a randomly chosen woman aged $a = 60$ (Table 15.2), while the average number of girls in the completed sisterhood or sorority was 1.26, the gross reproduction rate. The two numbers are very close, yet no allowance has been made for the woman aged a herself—one would have thought that the gross reproduction rate would equal $S^{\text{old}} + S_a^{\text{young}} + 1$. Have we inadvertently included the woman as her own sister? An examination of the argument leading to (15.3.1) and (15.3.2) will satisfy the reader that we have not. Nevertheless, how can the sisters of a randomly selected girl number 1.251, so that with her the sisterhood numbers 2.251, whereas the average number of girls obtained as the gross reproduction rate is only 1.26? (Goodman, Keyfitz, and Pullum 1975).

The answer lies in the manner of selection: the number of daughters of a randomly selected mother is decidedly smaller than the number of sisters plus one of a randomly selected girl. Consider the following hypothetical distribution of total number of daughters (i.e., of completed sororities):

Number of Daughters	Proportion of Cases
0	0.79
6	0.21

This would give the same gross reproduction rate of 1.26, but now all girls have five sisters, as needs no calculation to establish. The mean size of sororities when a *girl* is chosen at random is 6, very much larger than the average of 1.26 when a *family* is chosen at random. Whenever there is variation among mothers in childbearing, the estimate of the size of sorority from a sample of daughters will be larger than the estimate from a sample of families.

The difference between these two means can be expressed in terms of the variance. Suppose that of completed sisterhoods fraction f_0 is 0, f_1 is 1, and so on in the following scheme:

Table 15.2. Older and younger sisters ever born of a female aged a, birth and death rates of the United States, 1967

	a			
Sisters	0	20	40	60
Older S^{old}	0.610	0.610	0.610	0.610
Younger S_a^{young}	0.000	0.625	0.641	0.641
Total $S^{\text{old}} + S_a^{\text{young}}$	0.610	1.235	1.251	1.251

Number of daughters	Relative frequency
0	f_0
1	f_1
2	f_2
$\vdots$	$\vdots$

where $f_0 + f_1 + f_2 + \cdots = 1$. Then the gross reproduction rate G is

$$G = 0f_0 + 1f_1 + 2f_2 + \cdots,$$

the mean of the distribution, and the variance of the distribution is σ^2, where

$$\sigma^2 = 1^2 f_1 + 2^2 f_2 + \cdots - G^2.$$

The probability that a randomly chosen girl is a member of a sorority with zero members is 0, that she is a member of a sorority with one member is proportional to f_1, that she is a member of a sorority with two members is proportional to $2f_2$, and so on. This distribution is very different from the preceding one:

Size of sorority	Relative frequency of girls
0	0
1	$1f_1$
2	$2f_2$
3	$3f_3$
$\vdots$	$\vdots$

Now the total frequency is $0 + 1f_1 + 2f_2 + \cdots = G$, and the mean is

$$\begin{aligned} \frac{0 + (1)(1f_1) + (2)(2f_2) + (3)(3f_3) + \cdots}{0 + 1f_1 + 2f_2 + 3f_3 + \cdots} &= \frac{\sigma^2 + G^2}{G} \\ &= \frac{\sigma^2}{G} + G, \end{aligned} \tag{15.3.3}$$

without approximation. The number of sisters of a randomly selected girl equals this minus 1; i.e., $S = (\sigma^2/G) + G - 1$.

If the random selection were of families, the sorority would average G. The fact that the selection is of girls adds the term σ^2/G, the variance of the distribution of girls in families divided by the mean. Random selection of a daughter will always give a larger sorority than random selection of a family, as long as $\sigma^2 > 0$, that is, as long as there is any variation in family size. In the numerical calculation based on the formulae of this chapter the variance of the distribution is nearly equal to its mean, a relation characteristic of the Poisson distribution.

This does not arise from data but from the model. Recall that we made the probability of birth at any moment independent of births at all other moments. No heterogeneity was allowed for among women. That is why we ended with a Poisson distribution. In real populations some women are sterile and others have many children. Such heterogeneity among women negates the independence assumption and tends to make the variance greater than the mean. On the other hand, insofar as a two-child family is popular in birth-controlling populations, the variance is reduced.

15.3.2 Age Incidence of Childbearing Conditional on Birth of One Child

Within any homogeneous group expressions 15.3.1 and 15.3.2 are exact if $m(y)$ in the first and $m(x+u)$ in the second are conditional on the birth of a girl aged a at age x of the mother. Lacking data showing birth rates at the several ages for women who have had a birth at each age x, one is tempted to use overall values of $m(y)$ and $m(x+u)$. Insofar as the chance of another birth is zero in the months after a birth, a further impropriety is thereby added to the one discussed in detail above. The notch in the curve allowing for pregnancy and postpartum sterility would be compensated for in other parts of the range of mother's childbearing ages by the conditional $m(y)$ being higher than the average $m(y)$ for the entire population. Moreover the selection of mothers implicit in the fact that the ones we are concerned with are of proven fertility would probably add further to the conditional $m(y)$ in those ages y where it is nonzero.

The points raised above for sisters apply also to aunts, cousins, nieces, and other kin that are related through sisters.

15.3.3 Aunts

Sisters are aunts when seen from the viewpoint of the daughter of one of them. For the number of aunts that are older sisters of the mother the matter is simple; since S^{old} in (15.3.1) does not depend on the age of the girl, it must be invariant when taken in relation to the daughter of the girl. Thus S^{old} of (15.3.1) is also the expected number of aunts of a girl aged a who are older sisters of the mother. Under a fixed regime of birth and death, the same in both generations, a girl has the same expected number of older sisters and of maternal aunts older than her mother. This of course is not true of surviving sisters or aunts.

Aunts who are younger sisters of the mother, say A_a^{young}, are again first obtained conditionally on the mother having been of age x at the birth of girl now aged a (Figure 15.3). A mother who was then aged x must now be aged $x+a$, if she is alive. The number of younger sisters expected for a woman aged $x+a$ is S_{x+a}^{young}, entering the argument $x+a$ in (15.3.2). As

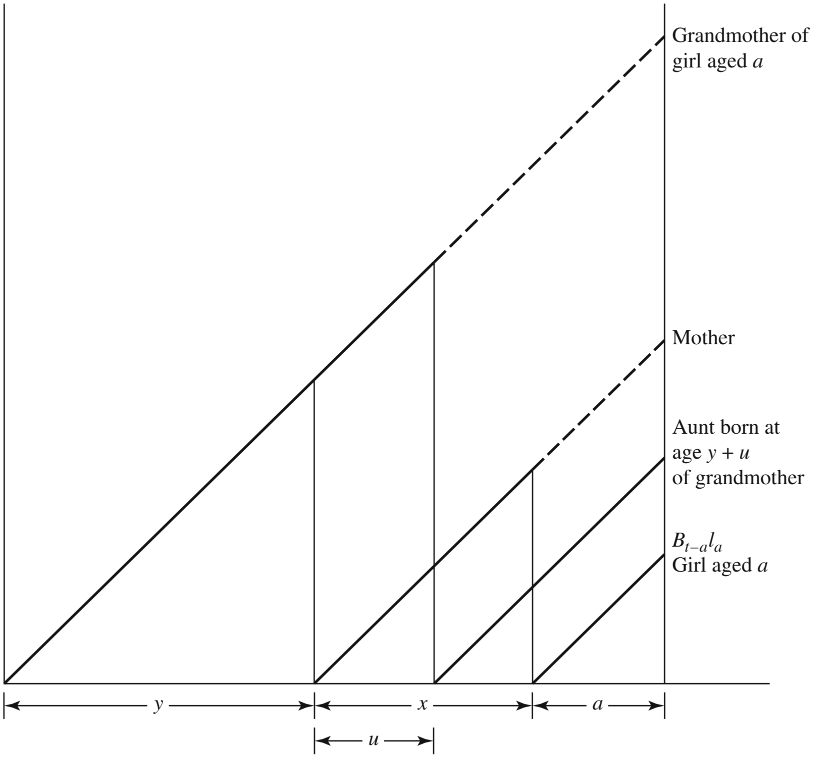


Figure 15.3. Lexis diagram for aunt born after mother.

earlier, we can use this result whether or not the mother is now alive. All that remains is to integrate out the condition that the girl aged a was born when her mother was aged x:

$$A_a^{\text{young}} = \int_\alpha^\beta S_{x+a}^{\text{young}} e^{-rx} l(x)m(x)\,dx,$$

or written out in full so as to accord with Figure 15.3,

$$A_a^{\text{young}} = \int_\alpha^\beta \int_\alpha^\beta \int_0^{x+a} \left(\frac{l(y+u)}{l(y)} m(y+u)\right) \left(e^{-ry} l(y)m(y)\right) \times \left(e^{-rx} l(x)m(x)\right)\,du\,dy\,dx.$$

The application of the same principles to nieces and first cousins (Figure 15.4) is found in Goodman, Keyfitz, and Pullum (1974). [The reader can show how to go on to second- and higher-order cousins, as well as great aunts and other distant relatives.]

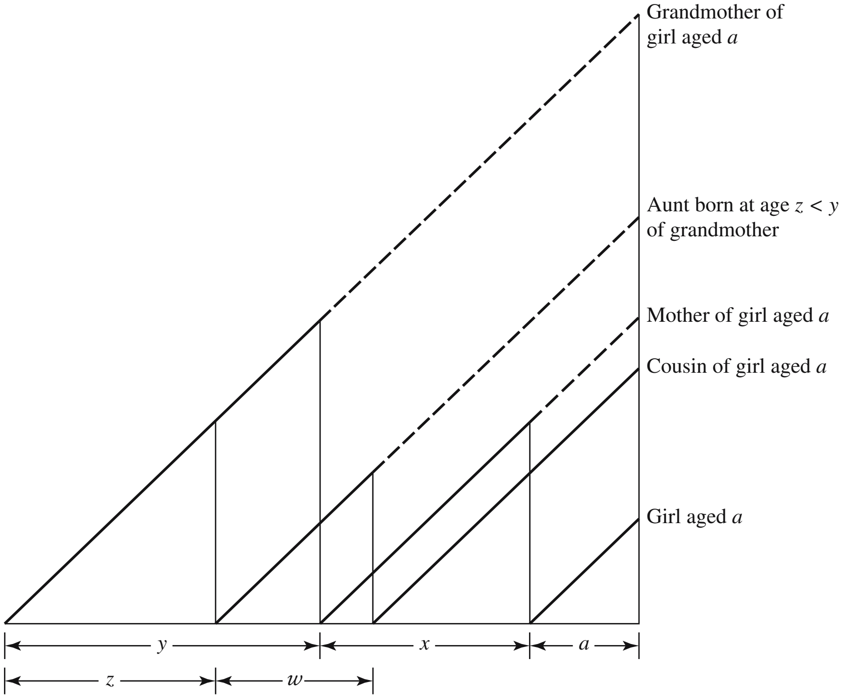


Figure 15.4. Lexis diagram for cousin of a girl aged a through mother's elder sister.

15.4 Mean and Variance of Ages

For each kin whose expected number can be calculated, so can the mean age. Consider the descendants, for example, of daughters already born to a woman aged $a > \alpha$. These number $\int_\alpha^a m(x)\,dx$, and their mean age must be

$$\frac{\displaystyle\int_\alpha^a (a-x)m(x)\,dx}{\displaystyle\int_\alpha^a m(x)\,dx} = a - \bar{x}_a, \tag{15.4.1}$$

where $\bar{x}_a$ is the mean age of the childbearing function $m(x)$ up to age a. The mean age of children still alive to the same woman aged a is

$$\frac{\displaystyle\int_\alpha^a (a-x)m(x)l(a-x)\,dx}{\displaystyle\int_\alpha^a m(x)l(a-x)\,dx}. \tag{15.4.2}$$

The mean age of granddaughters of women aged $a > 2\alpha$ is similarly

$$\frac{\displaystyle\int_{\alpha}^{a} l(x)m(x)\int_{\alpha}^{a-x}(a-x-y)l(y)m(y)\,dy\,dx}{\displaystyle\int_{\alpha}^{a} l(x)m(x)\int_{\alpha}^{a-x} l(y)m(y)\,dy\,dx}. \tag{15.4.3}$$

One can go to the variance of ages of descendants in the successive generations. Thus for the variance of ages of living daughters of a woman aged a we would have

$$\sigma_a^2 = \frac{\displaystyle\int_{\alpha}^{a}(x-\bar{\bar{x}})^2 m(x)l(a-x)\,dx}{\displaystyle\int_{\alpha}^{a} m(x)l(a-x)\,dx}, \tag{15.4.4}$$

where $\bar{\bar{x}}$ is the mean age of women up to age a at the birth of their children, weighted by the survival function $l(a-x)$. Here, as elsewhere, no account is taken of heterogeneity in ages and rates of childbearing, or of the spacing of children imposed by the sterile period of pregnancy and afterward. The simplification has negligible consequences for the expected number of daughters, or for the probability of a living grandmother, but does matter for expected sisters, aunts, and more distant collateral relatives, as well as for variances in all kin.

15.4.1 Ascertainment

Such results illustrate the concept of ascertainment, the way information has been obtained, used in genetics and applied in demography by Mindel C. Sheps. Consider the expected children of a given person. When the person is a child just born, her expected future daughters are $\int_\alpha^\beta l(x)m(x)\,dx$ prospectively; once she has passed age β, they are $\int_\alpha^\beta m(x)\,dx$ retrospectively; the corresponding mean ages are $\int_\alpha^\beta xl(x)m(x)\,dx/R_0$ and $\int_\alpha^\beta xm(x)\,dx/G_0$, the latter always being greater. The expected grandchildren calculated retrospectively for a woman of 85 differ from the prospective number of grandchildren of a child just born. In a cohort, individuals may be ascertained by the occurrence of a "signal" event at some point in their lives—the signal event may be having a second child, or being caught in a survey at time t. (Sheps and Menken 1973, p. 341.) Expressions for extended kin provide further illustrations.

15.5 Changing Rates of Birth and Death

This chapter has been restricted to the stable case, in which we suppose a fixed regime of mortality and fertility to be in force over a long past period and continuing into the present. Yet the theory can be extended to cover certain kinds of change in the regime.

For an example of how changing rates would be accommodated, let us reconsider the probability that a woman aged a has a living mother, the expression $M_1(a)$ of (15.1.3). The conditional probability that the mother is alive, $l(x+a)/l(x)$, must now be determined by the chance of survival appropriate to the changing death rates actually experienced by the cohort aged $x+a$ at time t, that is, born at time $t-x-a$. The ratio $l(x+a)/l(x)$ in the formula would have to be taken from the appropriate cohort life tables, a different table for each value of x. This is certainly possible, though awkward enough that no one is likely to do it.

In addition the distribution of x, the age at childbearing, is affected by the instability; if, for instance, the actual age distribution is younger than the stable one, (15.1.3) has to be modified to allow a greater weight to $l(x+a)/l(x)$ for younger x, thereby increasing the probability that the mother is still alive. Thus the factor $e^{-rx}l(x)m(x)$ in $M_1(a)$ would have to be replaced by numbers proportional to the actual ages of mothers prevailing a years earlier, say $w(x|t-a)$. Result 15.1.3 would thus be replaced by

$$M_1'(a) = \int_\alpha^\beta \frac{l(x+a)}{l(x)} w(x|t-a)\,dx,$$

where $w(x|t-a)$ is the age distribution of women bearing children a years ago, or at time $t-a$.

Analogous considerations permit a straightforward rewriting and reinterpreting of all the formulae of this chapter in a way that dispenses with the stable assumption insofar as it affects earlier age distributions. Goodman, Keyfitz, and Pullum (1974) provide these more general formulae. Interpretation for fixed rates is simple: the given schedules $l(x)$ and $m(x)$ imply certain mean numbers of kin. The corresponding statement for changing rates is unavoidably more complicated.

15.6 Sensitivity Analysis

A main use of the kinship formulae here developed is to ascertain the effect of changes in the demographic variables on kinship. How much does a younger age of marriage and of childbearing reduce the number of orphans? What is the effect of a fall in the birth rate on the number of grandchildren of a person of a given age chosen at random? What does a uniform improvement in mortality at all ages do to the number of living aunts of a girl of given age?

Merely looking at the formulae does not tell much more than we know without them. Intuition suggests that the fraction of girls aged a who have living mothers must depend primarily on death rates (specifically those between the time of childbearing and a years afterward) and secondarily on birth rates—most of all, on whether children are born at young or older ages of mothers. In a smaller way yet it ought to depend on the overall rate of increase, because with given death rates and with birth rates in a given proportional distribution a faster growing population has somewhat younger mothers. But a quick look at (15.1.3) reveals only that $M_1(a)$ is a function of birth and death rates, without clearly suggesting the amount or even the direction of the relation.

15.6.1 Decomposition of $M_1(a)$, the Probability of a Living Mother

Once a computer program for an expression such as (15.1.3) for $M_1(a)$ is available, it is possible to make small variations in any part of the input—add 10 percent to the birth rates at certain ages while leaving the life table intact, for instance—and see the effect on the probability of a living mother. Here we will try to see how such variations operate theoretically, by using the approximation to $M_1(a)$ developed as (15.1.7) or, written slightly differently,

$$M_1(a) \approx \frac{l(\kappa + a)}{l(\kappa)} + \frac{\sigma^2}{2}\left[\frac{l(\kappa + a)}{l(\kappa)}\right]'' . \tag{15.6.1}$$

The main effect of raising the mean age of childbearing is to replace mortality in an interval at the original κ with mortality around $a + \kappa$, as is evident from application of (1.6.5):

$$\frac{l(\kappa + a)}{l(\kappa)} = \exp\left[-\int_{\kappa}^{\kappa+a} \mu(t)\,dt\right] . \tag{15.6.2}$$

If the death rate is nearly constant with age, or if a is small, $M_1(a)$ depends little on the value of κ. The second term on the right-hand side of (15.6.1) is negative through most of the life table and is considerably smaller than the first, unless a is a very old age. The rate of increase of the population enters only through κ, which for a given life table is younger the higher the rate of increase.

If mortality $\mu(x)$ between κ and $\kappa + a$ increases by an amount k at every age, the survivorship $l(\kappa + a)/l(\kappa)$ will diminish in the ratio e^{-ka}, and this is the only effect of a constant mortality addition on $M_1(a)$. Hence the new $\overline{M}_1(a)$ is equal to $e^{-ka}M_1(a)$. The reason that the weighting factor $e^{-rx}l(x)m(x)$ in (15.1.3) remains unaffected is that a uniform change in mortality causes a change in r that offsets the change in $l(x)$ as far as age distribution is concerned, a matter discussed in Section 10.1.

15.6.2 Other Progenitors

If mortality at all ages is increased in the uniform amount k, the probability of a living grandmother will change to

$$\overline{M}_2(a) = \int_\alpha^\beta e^{-k(a+x)} M_1(a+x) e^{-rx} l(x) m(x)\, dx, \tag{15.6.3}$$

which cannot be simplified without approximation. But let x be replaced in $e^{-k(a+x)}$ by κ', the mean age at childbearing for the mothers still alive. Then, taking the exponential outside the integral, we have approximately

$$\overline{M}_2(a) \approx e^{-k(a+\kappa')} M_2(a).$$

For great grandmothers

$$\overline{M}_3(a) \approx e^{-k(a+\kappa'+\kappa'')} M_3(a),$$

where κ'' is the mean age at childbearing for grandmothers still alive. In practice we do not have data on κ' or κ'' and would suppose them to be close to κ; hence the outcome in general is

$$\overline{M}_i(a) \approx e^{-k[a+(i-1)\kappa]} M_i(a). \tag{15.6.4}$$

If k is small, so that e^{-ka} is nearly $1 - ka$, we obtain the following finite approximations for the difference in the several $M_i(a)$ on adding k to the force of mortality:

$$\begin{aligned} \Delta M_1(a) &= -kaM_1(a) \\ \Delta M_2(a) &= -k(a+\kappa)M_2(a) \\ \Delta M_3(a) &= -k(a+2\kappa)M_3(a). \end{aligned}$$

With these approximations, if one of two populations has mortality higher at every age by 0.003, and if κ is 27.5, for women aged $a = 20$ the chance of having a living mother is 0.94 as high as in the other population; of a living grandmother, 0.86; and of a living great-grandmother, 0.78. On the more precise (15.6.4) the last three numbers become 0.942, 0.867, and 0.799, respectively.

Venezuela has somewhat higher mortality than the United States; we note from Table 15.1 that for a Venezuelan girl aged 20 the probability of having a living mother is in the ratio $0.932/0.959 = 0.97$; of a living grandmother, in the ratio 0.85; of a living great-grandmother, in the ratio 0.64, all ratios to the United States. The gradient as one advances to more remote ancestors is steeper than that of (15.6.4) based on fixed differences of $\mu(x)$.

A more complete analysis would decompose the difference between the two countries in, say, the probability of a living grandmother into two components: (1) that due to mortality differences, and (2) that due to differences in the pattern of births. This is readily accomplished arithmetically, once an appropriate computer program is available, simply by permuting the input data, as was done in Table 10.2.

15.6.3 Effect of Birth Pattern on Living Progenitors

The main variation in $M_1(a)$ as far as births are concerned occurs through the mean age of childbearing κ. We found that

$$M_1(a) \approx \frac{l(\kappa+a)}{l(\kappa)};$$

therefore taking logarithms of both sides and differentiating gives

$$\frac{1}{M_1(a)}\frac{dM_1(a)}{d\kappa} = -\left[\mu(\kappa+a) - \mu(\kappa)\right],$$

where $\mu(\kappa)$ is the force of mortality at age κ. In finite terms

$$\frac{\Delta M_1(a)}{M_1(a)} \approx -\left[\mu(\kappa+a) - \mu(\kappa)\right]\Delta\kappa;$$

that is, the proportionate change in the chance of a living mother is minus the difference in death rates over an a-year interval times the absolute change in κ. With the death schedule of Madagascar, 1966 (Keyfitz and Flieger 1971), and its κ of about 27.5, using for $\mu(27.5)$ the approximation ${}_5M_{25} = 0.01740$, and for $\mu(\kappa+a) = \mu(27.5+20) = \mu(47.5)$ the rate ${}_5M_{45} = 0.02189$, we have

$$\frac{\Delta M_1(a)}{M_1(a)} \approx -(0.02189 - 0.01740)\Delta\kappa = -0.0045\Delta\kappa.$$

For each year later of average childbearing the chance of a woman of 20 having a living mother is lower by 0.45 percent.

The change with a is found in the same way to be

$$\frac{\Delta M_1(a)}{M_1(a)} = -\mu(\kappa+a)\Delta a,$$

and for $a = 20$, $\kappa = 27.5$, this is

$$-\mu(47.5)\Delta a = -0.02189\Delta a.$$

For a 5-year interval the proportionate decrease in $M_1(20)$ ought to be 5 times as great or 0.109. In fact Table IIIa of Goodman, Keyfitz, and Pullum (1974) shows

$$\frac{M_1(25) - M_1(15)}{2} = -\frac{0.7817 - 0.6421}{2} = -0.0698;$$

and as a proportion of $M_1(20) = 0.7126$, this is $-0.0698/0.7126 = -0.098$, about as close to -0.109 as we can expect with the crude approximations used.

Table 15.3. Effect of changed birth rate on probability of ancestor being alive, Madagascar females, 1966

	Difference in Probability of Having a Living:		
Age of woman a	Mother $\overline{M}_1(a) - M_1(a)$	Grandmother $\overline{M}_2(a) - M_2(a)$	Great-grandmother $\overline{M}_3(a) - M_3(a)$
	After Lowering the Birth Rate for Women 20–24 by 0.01		
0	0.	−0.00268	−0.00529
20	−0.00025	−0.00423	−0.00109
40	−0.00228	−0.00103	−0.00000
60	−0.00067	−0.00000	−0.00000
	After Lowering the Birth Rate for Women 40–44 by 0.01		
0	0.	+0.00395	+0.00404
20	+0.00109	+0.00345	+0.00057
40	+0.00199	+0.00051	0.00000
60	+0.00036	0.00000	0.00000

To find the effect of a change in fertility at particular ages one can run the program twice, once with the observed regime of mortality and fertility, and once with the specific birth rate for age 20–24 lowered by 0.01. Differences for progenitors are shown in Table 15.3. A drop in fertility at age 40–44 lowers the average age of childbearing and hence raises the chance of a living grandmother. Other items can be similarly interpreted.

15.6.4 Comparison of Effect of Birth and Death Rates

Robert Sembiring (1978) has experimented to determine whether the numbers of particular kin are affected more by birth or by death rates. As an example of the procedure we consider the number of female cousins that a girl aged a would be expected to have through her mother's sister and use the technique of permuting the input data, as in Section 10.1.

To separate the effects of mortality from those of fertility nine calculations were made of the curve of expected cousins by age. Three levels of fertility were used, those of Costa Rica for 1960, the United States for 1959–61, and Sweden for 1958–62, all pertaining to 1960 or thereabout. The gross reproduction rates of the three countries were 3.891, 1.801, and 1.080, respectively. Each of these levels of fertility was paired with each of three levels of mortality taken from the Coale and Demeny West series model tables, with $\overset{o}{e}_0$ values of 70, 55, and 40. The numbers chosen represent approximately the range of mortality and fertility among human populations. The resulting nine curves for the average number of living cousins in the female line (i.e., daughters of maternal aunts) are shown in Figure 15.5.

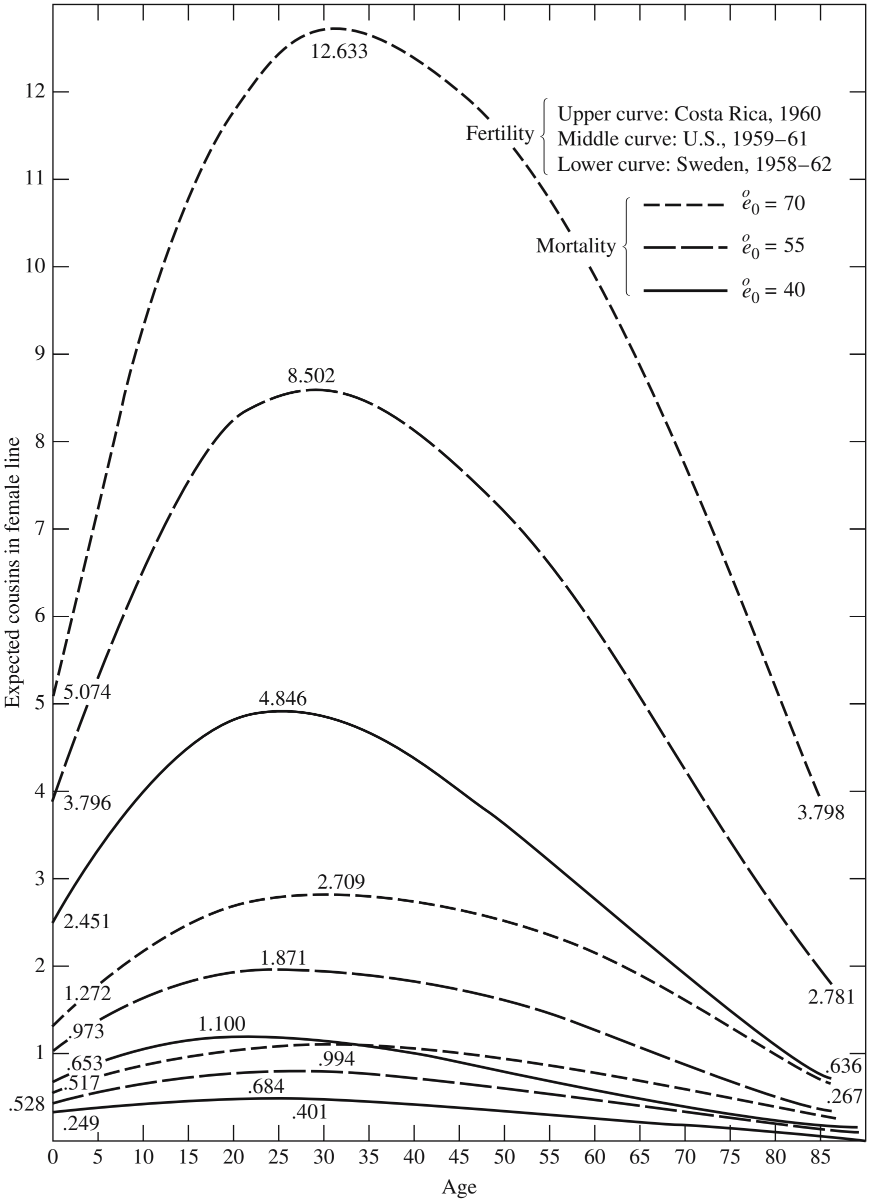


Figure 15.5. Expected number of cousins still alive.

The three curves for high (Costa Rican) fertility are above those for intermediate fertility, and these again are mostly above those for low fertility. Apparently fertility has more effect on the number of living cousins than does mortality. With high fertility the number of cousins reaches a sharp

Table 15.4. Mean number of living first cousins in the female line of a woman aged 20, for artificial populations constructed by fertility of Costa Rica, the United States, and Sweden, about 1960, in all combinations with mortality of Coale–Demeny model West tables having $\overset{o}{e}_0$ values of 40, 55, and 70 years

	Fertility		
Mortality	Costa Rica $G_0 = 3.891$	United States $G_0 = 1.801$	Sweden $G_0 = 1.080$
$\overset{o}{e}_0 = 40$	4.7292	1.0998	0.4009
$\overset{o}{e}_0 = 55$	7.9888	1.8502	0.6789
$\overset{o}{e}_0 = 70$	11.3518	2.6006	0.9657

peak at ages 25 to 45; with lower fertility and lower mortality the curve peaks less sharply. Note that no actual population combines a gross reproduction of 3.891 with an expectation of life of 70 years; therefore the peak of 12.633 female maternal parallel cousins is purely hypothetical. The combination of United States or Swedish fertility with this high expectation of life represent possible real situations.

All of the expressions in this chapter apply to male as well as to female kin, but with one difference. This difference arises out of the fact that we know the mother was alive at the birth of her child, but we know only that the father was alive 9 months before the birth. To apply to males, the formulae would have to be adjusted for the three-quarters of a year of additional mortality, which could be appreciable for certain kin in populations subject to high death rates.

A rough approximation to the total number of first cousins (i.e., of both sexes) would be obtained by multiplying the material parallel female cousins here given by 8; this would be improved by making the corresponding calculation for the male line and for mixed lines. An approximation to the number of cousins implied by other schedules of mortality and fertility would be obtained by two-way quadratic interpolation, using $\overset{o}{e}_0$ and the gross reproduction rate as indices. This is especially feasible for mortality; note that the number of cousins for $\overset{o}{e}_0 = 55$ is almost exactly the mean of the numbers for $\overset{o}{e}_0 = 40$ and $\overset{o}{e}_0 = 70$ (Table 15.4). Interpolation may be useful even with an available program because of the large amount of computer time required for the exact calculation.

15.7 The Inverse Problem: Deriving Rates from Genealogies

The inverse problem is of practical interest to those who must make inferences regarding birth and death rates for areas or times for which registration systems are not in existence, or if in existence are grossly incomplete. For a given genealogy, the problem now is to find the regime of mortality and fertility.

If one-half of women aged 40 have living mothers in a certain population, what is the expectation of life? According to Table 15.1, Venezuela showed a probability of 0.707 and Madagascar of 0.386, and these had $\overset{o}{e}_0$ values of 67.7 and 38.5 years, respectively. By straight-line interpolation, supposing ages at childbearing to be sufficiently similar among all three populations to leave the answer unaffected, we find $\overset{o}{e}_0 = 48.9$ corresponding to our $M_1(40) = 0.5$.

Looked at formally, what were relatively simple integrals to evaluate when the regime of mortality and fertility was given become difficult—in most instances unsolvable—integral equations when the mean number of kin is known but the regime is unknown. If we observed mean numbers of the various living kin, we would have a set of equations, most of them containing multiple integrals, and they would have to be solved as a simultaneous set. Thus all of the expressions for different kin might be equated to observations and solved simultaneously for the unknown rates of birth and death.

Yet for many practical purposes we can avoid most of the difficulties just mentioned by supposing that all life tables can be laid out in a straight line, indexed by the expectation of life at age 0 or age 10, and correspondingly that schedules of childbearing can be arranged according to the gross reproduction rate G_0. All variations in mortality and fertility beyond these two dimensions will be neglected in this simple method.

If the number of cousins to women aged 20 is represented as a height above the $(G_0, \overset{o}{e}_0)$ plane, so that the collection of such information regarding women aged 20 is a quadratic surface over the plane, an observed number of cousins can be represented by a plane parallel to the $(G_0, \overset{o}{e}_0)$ plane and cutting the quadric surface in a second-degree curve, which may now be projected down onto the $(G_0, \overset{o}{e}_0)$ plane. To make the regime entirely determinate we need some other fact about kin. Continuing to confine ourselves to age 20 for the sake of this example (though we need not stay with the same age), we note the fraction of women aged 20 having a living mother. This also can be represented as a quadric surface over the same $(G_0, \overset{o}{e}_0)$ plane, and the given observation as a plane again cutting the quadric surface in a quadratic. The intersection of the two quadratics gives the regime of mortality and fertility. The object is to choose kin that provide curves intersecting as nearly as possible at right angles to each another. Examples are probability of living mother (or grandmother) and number of daughters (or granddaughters) ever born.

In practice any one pair of observations will be unacceptable as neglecting most of the data. Indeed, errors are so pervasive that we will do none too well using all of the information available. A large number of data pairs will each provide a point of intersection, and with given accuracy of enumeration the precision of any point will be greater the closer the lines defining it come to making a right angle with each other. The several estimates obtained from pairs of kin can be weighted by the sine or other suitable function of the angle that the lines make with each other.

15.8 Incest Taboo and Rate of Increase

An incest taboo has the advantage for the group that adheres to it of compelling biological and social mixing, and of stirring individual initiative in the search for a mate. In addition, it avoids the confusion that would result if one's father were also one's uncle, these being very different roles in most societies. It promotes political alliances among families, and it avoids intrafamilial conflict over women. Most such advantages are greater the wider the degree of incest prohibition: a taboo reaching as far as second cousins will compel more mixing than a taboo against brother–sister matings only.

But a price has to be paid for the advantages—the wider the taboo, the more individuals will fail to find mates, especially in sparsely settled populations. From the viewpoint of reproduction the incest taboo is a luxury, and the question is how much of it a group can afford. The ideal approach would be an analysis of trade-offs: find the point at which the advantages of increased mixing are exactly offset by the lesser reproduction. Unfortunately the elements of this equation are incommensurable, and no theory seems to exist that will provide a quantitative measure of the net advantage of mixing.

However, it is possible to deal with one side of the problem: the cost in rate of increase of various degrees of incest taboo. Though determinate, this is mathematically difficult, as are all questions of population increase in which the rate of reproduction depends on the size of the group and the random number of possible mates. Having little hope of finding a closed solution, Hammel (1972, Hammel et al. 1976, 1979) and his coworkers addressed the problem by simulation.

They used mortality rates assembled from the Maghreb and ancient Rome, and fertility rates from the Cocos-Keeling Islands, reduced by 20 percent to be slightly below stationarity without any incest taboo. A group of 65 individuals with a kinship structure of three generations of genealogical depth was derived from the previous evolution of the model. Five runs were then made with no incest prohibition, five with a prohibition of one degree (sibling and parent–child), five with a prohibition of two degrees (up to first cousin), and five with a prohibition of three degrees (effectively up to second cousin). All runs covered 100 years.

The outcome, as anticipated, was a more rapid rate of population decrease the broader the taboo. With no taboo the mean of the five runs was a rate of $r = -0.001$; with sibling and parent–child exclusion the mean was -0.006; with exclusion up to first cousin the mean rate was -0.018; with exclusion up to second cousin the rate was -0.036. The one-degree prohibition apparently costs 0.005, the two-degree an additional 0.012, and the three-degree a further 0.018 lowering of r, on this rough model.

Insofar as the rates used were realistic, a breeding group of the order of 65 individuals could not afford any exclusion, not even siblings. However, as mentioned above, the authors had reduced the Cocos-Keeling fertility by 20 percent. A group that would increase at 1 percent per year with no taboo at all could tolerate the sibling and parent–child taboo, which would reduce it to $0.01 - 0.005 = 0.005$; but it could not afford to go as far as the first cousin taboo, which would bring it down to $0.005 - 0.012 = -0.007$, or a half-life of a century. The calculation suggests that the incest taboo, aside from its other functions, is capable of holding down the rate of increase in small dispersed populations.

15.9 The Bias Imposed by Age Difference on Cross-Cousin Marriage

Among social groups practicing cross-cousin marriage, more instances have been observed of Ego, a male, marrying his mother's brother's daughter (MBD) than his father's sister's daughter (FSD). It is also usual for men to marry women younger than themselves. The question is whether the age bias by itself would lead to the bias toward MBDs. A realistic model would be complicated, but the effect of brides being younger than grooms can be shown with some simplified arithmetic (Hammel, 1972).

Consider a population in which men marry 5 years older than women, and children are born to the couple when the husband is age 25 and the wife age 20. Brothers and sisters are all the same age. Then, if Ego (male) is age E, his father will be age $E + 25$, and his father's sister also $E + 25$. His FSD will be $E + 25 - 20 = E + 5$. On the other hand, his mother will be $E + 20$, his mother's brother also $E + 20$, and his MBD $E + 20 - 25 = E - 5$. Thus, of the two kinds of cross-cousins, the FSD is 5 years older than Ego. If he is seeking a bride 5 years younger, he will find the MBD the right age.

A similar calculation can be made for parallel cousins. Ego's father being $E + 25$ years of age and his father's brother also $E + 25$, his FBD is $E + 25 - 25 = E$, and similarly for the other parallel cousin, designated as MSD. Parallel cousins are the same age as Ego on this simple model.

These purely demographic (some would say, merely logical) considerations mean that the tendency for men to marry kinswomen younger than themselves leads to the MBD marriage. Of the four kinds of cousins only MBDs are the right age to permit men to be older than their brides generation after generation.

Hammel was the first to point out that age preferences for *either* older or younger wives would have equivalent effects, and indeed that any heritable property would work as well as age. The contribution of Hammel and Wachter was to show by simulation that the effect remains considerable even in the face of all the obvious sources of randomness and to study the dependence of the effect on the size of the age gap. Simulation has been useful here and in other instances where analytic solutions are out of reach. Kunstadter et al. (1963) used it to find the fraction of individuals who would have an MBD cousin to marry in a tribe when that was preferred.

The approach in this chapter, via stable population theory, takes a deterministic approach appropriate for large populations. We should, however, point out two other important approaches, both of which take more account of individuals and their properties. First, we might recognize that the vital rates apply as probabilities to discrete individuals. If we suppose that they do so independently, we are led to stochastic branching process models (a simple branching process model appears in Section 16.4; see also Chapter 15 of MPM). These models have been used by Pullum (1982, Pullum and Wolf 1991) to derive entire probability distributions of the numbers of kin of various kinds, but without taking the age-specificity of the vital rates into account.

If we take this approach to its limit, we would want to keep track of each individual, with all of his or her i-state variables (age, marital status, health, employment, etc.) and relationships to other individuals. We would then apply to each individual the probabilities of birth, death, marriage, and any other demographic transitions of interest. Doing so repeatedly would project the population forward in time subject to those rates. Repeating that exercise many times would produce the probability distribution of population trajectories (including all the information on all the individuals) implied by the vital rates. Such models are called i-state configuration models (Caswell and John 1992) or individual-based models (DeAngelis and Gross 1992, Grimm et al. 1999) in the ecological literature, and microsimulation models in the human demographic literature (e.g., Wachter et al. 1997, Wolf 2001). They have been applied to problems of kinship by, e.g., Hammel et al. (1979), Ruggles (1993), Wachter et al. (1997), Wachter (1997), and the chapters in Bongaarts et al. (1987).

The approaches of the 15 chapters through this one may be called *macrodemography*, following a usage going back through sociology, economics, and physics, ultimately to a source in Greek metaphysics. Microsimulation methods are an example of *microdemography*, in which properties of individuals and their random variation are recognized as the source of change in population aggregates. Chapter 16 introduces some aspects of microdemography.

16

Microdemography

Physics accounts for heat by the motion of molecules, medicine accounts for disease by the action of germs, and economics accounts for aggregate prices and production by the activities of individuals seeking to maximize their utility. The fact that nontrivial problems of aggregation arise, and that the microelements often turn out on closer examination to be unrealistic constructions, does not deprive them of explanatory and predictive value.

Microdemography helps us to understand such macrophenomena as birth rates of regions and nations. How much reduction of the birth rate results from couples substituting 99 percent efficient contraception for methods 95 percent efficient? If the number of abortions was equal to the number of births in a country, could we conclude that suppression of abortion would double the number of births? How can the probability of conception be measured? What difference does it make to the increase of a population if parents aim for three children rather than two? These and other questions of microdemography are the subject of the present chapter.

16.1 Births Averted by Contraception

The theory of birth as a Markov renewal process have been developed by Sheps (1964, Sheps and Perrin 1963, Sheps and Menken 1973), and Potter (1970) has shown how this theory can be applied to calculate births averted. Tietze (1962) was a pioneer in this field and Lee and Isbister (1966) made important early suggestions.

In conception and birth models a woman is thought of as going through pregnancy and birth, and then again pregnancy and birth, in periodic fashion, with a longer or shorter cycle. Our main effort here is devoted to finding the length of the cycle under various conditions, and to showing how that tells us the birth rate. First consider a couple who have just married, engage in intercourse without using birth control, and are fecund. Let the probability of conception for a nonpregnant woman in any month be p, and of not conceiving be $q = 1 - p$. Until further notice, all conceptions leading to miscarriage or stillbirth will be disregarded. Thus p is the probability of a conception leading to a live birth. The first question is the expected time to conception.

The probability that the time to conception will be 1 month is p, that it will be more than 1 month is q, that it will be 2 months is qp, that it will be 3 months is q^2p, and so on. Multiplying the probability of exactly 1 month by 1, of 2 months by 2, and so on, gives t, the mean number of months of waiting until conception:

$$t = p + 2qp + 3q^2p + \cdots .$$

To evaluate this we replace each p by $1 - q$ to obtain

$$t = 1 - q + 2q(1 - q) + 3q^2(1 - q) + \cdots ,$$

which permits canceling and leaves only the geometric progression

$$t = 1 + q + q^2 + \cdots ,$$

which converges for q less than 1. To find the sum we multiply by q and note that the right-hand side is the same infinite series for t, except that the 1 is missing. This provides the equation $qt = t - 1$, from which t equals $1/(1 - q)$, a result familiar in high-school algebra. Hence the mean time to conception is $t = 1/(1 - q) = 1/p$. If the chance of conception is zero, the waiting time is infinite and the argument loses its interest. Hence we consider fecund women only, defined as those for whom p is a positive number; the argument does not apply to perfect contraception.

The time to conception is the first of two intervals that make up the conception and birth cycle. The second is the nonfecund period that includes about 270 days of pregnancy plus postpartum sterility. The length of the postpartum anovulatory period depends on lactation and other factors, and we need to be detained by variation and uncertainty regarding its length, but will simply call the entire expected period of pregnancy and postpartum sterility s. This period also includes any time lost by spontaneous or voluntary abortion.

Then the average length of the cycle is

$$w = \frac{1}{p} + s \quad \text{months.} \tag{16.1.1}$$

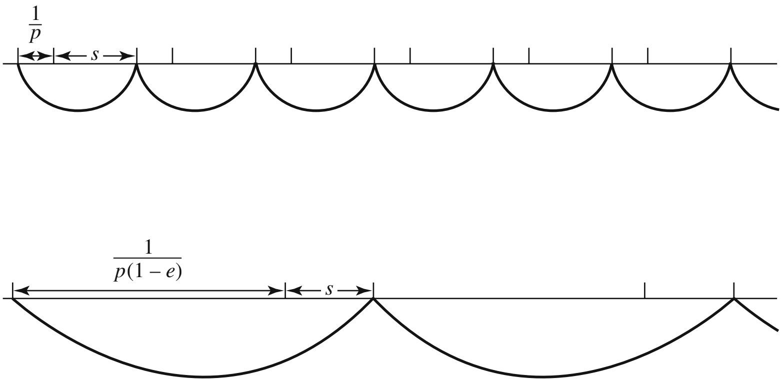


Figure 16.1. Hypothetical average cycles of conception and birth for natural fertility and for contraception of 90 percent efficiency.

Knowing the length of the cycle is the equivalent of knowing the birth rate for the population; if all women produce a child every w months, the average monthly birth rate is $1/w$ and the annual birth rate is $12/w$.

To illustrate (16.1.1) numerically, if $p = 0.2$ and the mean value of s is 17 months, the length of the cycle between births is 22 months, and the birth rate is $1/[(1/0.2) + 17] = 1/22$ per month, or 12/22 per year. The problem of measuring p will be the subject of Section 16.2; in this section it is assumed to be known.

The argument resembles that underlying the stationary population in Chapter 2, except that here no variation in the probability in successive months is allowed. If the death rate is $\mu\, dx$ from age x to $x + dx$, μ being the same for all ages, then the expectation of life at any age, as we saw, is $1/\mu$ years. Month-to-month variation in the probability of conception needs to be recognized, just as variation with age was recognized for deaths, and the life table technique applied to conception is discussed in Section 16.2.

The purpose of contraception is to reduce the probability of conceiving, which has the effect of increasing the time between successive births. We call the efficiency of contraception e, defined by the reduction in the probability of conceiving: if probability for a given month is p without protection, a contraceptive of efficiency e reduces this to $p(1-e)$, say p^*. Conversely, if we know that the probability of conception without protection is p, and with a certain contraceptive is p^*, then solving for e in the equation $p^* = p(1-e)$ gives $e = 1 - p^*/p$. If the probability is $p = 0.2$ without protection and $p^* = 0.02$ with a given contraceptive, the efficiency is $1 - 0.02/0.2 = 0.90$ or 90 percent.

To find w^*, the length of the cycle with contraception, we need not begin the argument over again, but only multiply p by $1 - e$ in the expression for w to obtain the new lower probability of conception, and enter this product in (16.1.1):

$$w^* \frac{1}{p(1-e)} + s. \tag{16.1.2}$$

Contraception is considered to have no effect on the sterile period s.

Imagine two groups of women, the first group not practicing contraception and having length of cycle $w = (1/p) + s$, and the other practicing contraception of efficiency e, with length of cycle w^* as in (16.1.2). Then w^*/w of the cycles of the first group fit into each of the cycles of the second group (Figure 16.1), and this is the ratio of the birth rates for the two groups. In other words, if the birth rate for the unprotected group is $1/w$ and for the contracepting group is $1/w^*$, the first has a birth rate $(1/w)(1/w^*) = w^*/w$ times that of the second.

The model is crude in omitting many factors, including end effects that could dominate if birth rates are very low. But it suffices to show that 90 percent efficient contraception does not lower the birth by 90 percent. For if, as before, the probability of conception with unprotected intercourse is 0.2 and the sterile period of pregnancy and afterward is 17 months, the waiting time will average $(1/0.2) + 17 = 22$ months. With 90 percent efficient contraception this increases to $1/[(0.2)(0.1)] + 17 = 67$, or just 3 times as long. Ninety percent efficient contraception reduces the birth rate by only two-thirds, rather than the nine-tenths that would apply if there were no sterile period. To put the matter intuitively, contraception can serve no purpose during a time when a woman is sterile anyway.

The technique of comparing two groups of women, one using no contraception and the other using a method of given efficiency, disregards the facts that women differ from one another in fecundity, that each woman does not continue to use a particular method indefinitely, and that the natural fecundity of each woman declines, ultimately to zero around age 50. The model hardly simulates what happens when contraception is used for a relatively short period of time like a year or two by a heterogeneous group of women; it offers a way of thinking about the matter that retains some of the essential features, and it helps to avoid some gross fallacies.

Abstention. One absolutely certain method of contraception is complete abstention from intercourse, but this is too drastic for general use, and partial abstention is the utmost that can be aspired to in practice. Suppose that a couple decide to abstain on every second occasion, without regard to the time of month. Even this represents a considerable degree of restraint; what is its effect on births? Insofar as the restraint reduces the probability of conception by one-half, it is equivalent to a contraceptive of efficiency $e = 0.5$. With parameters $p = 0.2$ and $s = 17$ months, the ratio of monthly birth rate with 50 percent abstention to that with no abstention must be

$$\begin{aligned}\frac{1/w^*}{1/w} = \frac{w}{w^*} &= \frac{(1/p) + s}{[1/p(1-e)] + s} \\ &= \frac{(1/0.2) + 17}{[1/(0.2)(1-0.5)] + 17} = \frac{22}{27} = 0.81.\end{aligned} \tag{16.1.3}$$

The 50 percent restraint would reduce the birth rate by only about 19 percent. This would be improved if something better than random timing of intercourse could be arranged.

16.1.1 Births Averted—The Causal Inference

The original inspiration for such models as these was the attempt to relate activity in the dissemination of the means of contraception to decline in the birth rate. A typical question concerns the effect on births of inserting intrauterine devices (IUDs) in 1000 women. First suppose that the IUDs are of 97 percent efficiency, that they will remain in place indefinitely, and the women in question have previously been unprotected. With our same p and s again, the ratio of the new to the old birth rate would be

$$\begin{aligned}\frac{1/w^*}{1/w} = \frac{w}{w^*} &= \frac{(1/0.2) + 17}{[1/(0.2)(1-0.97)] + 17} \\ &= \frac{22}{184} = 0.12,\end{aligned}$$

which tells us that 88 percent of births would be averted from then onward. In absolute numbers, without the IUDs the 1000 women would have $1000(12/22) = 545$ births per year; with the IUDs they would have $1000(12/184) = 65$ per year, so that 480 births would be averted.

This calculation may well be correct. Natural fertility is indeed to be found in some places. But far more common, even in very high fertility populations with birth rates of 40 per thousand or above, are primitive but not wholly ineffective means of contraception. Even when the population as a whole shows a very high birth rate, it can contain some couples practicing relatively efficient contraception.

The crucial assumption of the above calculation that 88 percent of births are averted is that the IUDs would take the place of unprotected intercourse. The question, “How many births would be averted by the insertion of 100 IUDs?” implies a causal analysis and cannot be answered without first answering a subsidiary question: “What would the women be doing in the absence of the IUDs?” If there were a selection of the women who turned up for insertion of IUDs, by which most were in any case practicing relatively effective birth control, the answer in regard to births averted would be very different from the 88 percent above.

To go to an extreme, suppose that the couples were using periodic abstention in the hope of avoiding the fertile period, or condoms, or a combination, and somehow attaining 90 percent efficiency. This means that in the absence of the IUDs the probability of conception would be 0.1 of the chance with unprotected intercourse, and hence the interval between pregnancies, again supposing $p = 0.2$ and $s = 17$ months, would be $1/[(0.2)(1-0.90)]+17 = 67$ months.

If the same conditions applied in the future, the birth rate would be 1/67 per month without the IUDs, and $1/\{[1/(0.2)(1-0.97)]+17\} = 1/184$ per month with them. Now births would be $67/184 = 0.36$ of what they would have been, so 64 percent would be averted. Births per year among the 1000 couples would be $12{,}000/67 = 179$ without the IUDs, and 65 with them. On these assumptions the credit to the activity of fitting 1000 IUDs would be 114 births per year. If, however, the alternative that the couples would use in the absence of IUDs was 95 percent efficient, avoidance of only $103 - 65 = 38$ births could be credited to the IUDs.

To suppose that no contraceptive would be used in the absence of the IUD gives 480 births averted per 1000 insertions; if the alternative to the IUD is 90 percent efficient contraception, 114 births are averted; if the alternative is 95 percent efficient contraception, the births averted are 38. The variation in the apparent effectiveness of the family planning effort with seemingly small variation in the assumptions is distressing, but any attempt to get around it must face the basic fact that no birth can be averted that was not going to occur.

An example of the way the arithmetic operates is given by comparing two cases: (1) a population using contraception of 50 percent efficiency, and (2) another population of which half use contraception of perfect efficiency and the other half use no contraception. The birth rate per month for the first case is $1/[(\frac{1}{2}p) + s]$; for the second, $\frac{1}{2}\{1/[(1/p) + s]\} + \frac{1}{2}(0)$. With $p = 0.2$ and $s = 17$, the first is 1/27, the second, 1/44. Thus the birth rate in the first case would be 44/27 times the birth rate in the second, that is, 63 percent higher.

16.1.2 Marginal Effect

One of the applications of this approach is to determine the result of a slight improvement in efficiency—a better IUD, a superior condom, slightly more careful use of the condom. Suppose that efficiency is raised from e to $e+\delta$. Then average births averted per year are

$$\begin{array}{l}\text{Births averted}\\ \text{per year per woman}\end{array} = \frac{12}{[1/p(1-e)]+s} - \frac{12}{[1/p(1-e-\delta)]+s}.$$

If $f(e) = 12/\{[1/p(1-e)] + s\}$, the additional births averted are approximately equal to $f'(e)\delta$; hence we need the derivative $f'(e)$, which

Table 16.1. Birth rates per month with various degrees of contraceptive efficiency, based on $p = 0.2$ without protection, and $s = 17$

e	$\dfrac{1}{\frac{1}{p(1-e)} + s}$
0.00	0.04545
0.01	0.04535
0.05	0.04492
0.50	0.03704
0.75	0.02703
0.90	0.01493
0.94	0.00997
0.95	0.00855
0.96	0.00704
0.97	0.00544
0.98	0.00375
0.99	0.00193
1.00	0.00000

is

$$f'(e) = -\left(\frac{12}{\{[1/p(1-e)] + s\}^2}\right)\left(\frac{1}{p(1-e)^2}\right).$$

Thus the additional births averted by a year of increase δ in efficiency, as a fraction of the number that would otherwise have occurred, are

$$-\frac{f'(e)\delta}{f(e)} = \frac{\delta}{(1-e)[1+sp(1-e)]}. \tag{16.1.4}$$

The values of this expression for four values of e and $p = 0.2,\ s = 17$ are as follows:

e	Further fraction averted by improvement δ in efficiency
0	0.23δ
0.9	7.46δ
0.95	17.09δ
0.97	30.25δ

Evidently to go from no protection to a contraceptive of 1 percent efficiency averts only $0.23\delta = (0.23)(0.01) = 0.23$ percent of births. On the other hand, to go from a contraceptive of 97 percent efficiency to one of 98 percent efficiency averts over 30 percent of births. The extra 1 percent of efficiency is 130 times as effective in lowering pre-existing births at the 97 percent level as at the 0 percent level. [Express the relative efficiency in terms of possible births.]

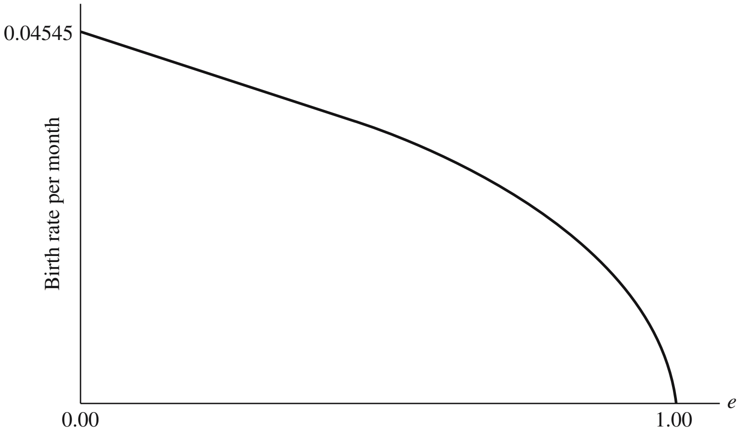


Figure 16.2. Curve of monthly birth rates at various efficiencies of contraception, $p = 0.2$ and $s = 17$ months.

Table 16.1 provides both a verification of these numbers and an illustration of their meaning. To go from 0 to 0.01 efficiency (a difference $\delta = 0.01$) decreases the monthly birth rate by $0.04545-0.04535 = 0.00010$ on 0.04545, or 0.0022, which is 0.0022/0.01 or 0.22, and multiplying by δ this is the same except for rounding as the 0.23δ above. To go from 0.96 to 0.98 efficiency is to lower the birth rate from 0.00704 to 0.00375, a difference of 0.00329, or 0.00164 per 0.01 of increase in efficiency. As a fraction of births at 0.97 efficiency (Table 16.1) this is $0.00164/0.00544 = 0.301$, or 30.1δ, since $\delta = 0.01$, in agreement except for slight curvature and rounding with the 30.25δ given above.

Figure 16.2 shows monthly birth rates as a function of efficiency of contraception for the simple model here used. The numbers in the table following (16.1.4) give the relative slope at four points in the curve.

16.1.3 Dropping the Contraceptive

We have been considering an IUD, a supply of pills, or other "segment," as Tietze and Potter call it, of contraception, and have calculated the effect per year, as though everyone would use the supply according to instructions and without interruption. But we know that different individuals continue to be careful for different lengths of time. A rough way of taking this into account is to suppose each woman has a probability d of dropping the contraceptive in each month. During one segment the probability of conception in a given month is p' (supposed the same for all women and for all times), and the probability in any month of dropping the contraceptive for the women who do not become pregnant while using it is d.

All of the above refers to a particular month, and we suppose that the women are followed through time until they either become pregnant or drop the contraceptive. One of these happening in the first month has probability $p' + d$; its not happening in that month but happening in the next month has probability $(1 - p' - d)(p' + d)$; its happening in the third month for the first time has probability $(1 - p' - d)^2(p' + d)$; and so on. If we write q for $1 - p' - d$, the sequence is

$$(1 - q),\ q(1 - q),\ q^2(1 - q),\ \ldots .$$

The mean number of months of exposure is

$$(1 - q) + 2q(1 - q) + 3q^2(1 - q) + \cdots = 1 + q + q^2 + \cdots = \frac{1}{1 - q}$$

$$= \frac{1}{1 - (1 - p' - d)} = \frac{1}{p' + d},$$

before the woman passes out of the group either through becoming pregnant or dropping the contraceptive.

Among the women who leave the group of nonpregnant contraceptive users during a particular month, the proportion $d/(p' + d)$ do so through dropping the contraceptive, and $p'/(p' + d)$ do so through becoming pregnant. If, in each case, this is the fraction in every single month, it is also the fraction in all months together. Thus $d/(p' + d)$ of the original group of women will sooner or later drop the contraceptive and we suppose that they revert to natural fertility; their chance of conceiving in any month becomes p. The model follows all women to pregnancy, either while using the contraceptive or subsequently. With the same argument, now applied to p rather than to $p' + d$, their mean time to pregnancy after dropping the contraceptive will be $1/p$.

Thus all women average $1/(p' + d)$ months until they drop the contraceptive or become pregnant; $d/(p' + d)$ of them drop the contraceptive, and these take another $1/p$ months, on the average, to become pregnant. Then the expected time to pregnancy for all of the women is

$$t = \frac{1}{p' + d} + \frac{d}{p' + d}\left(\frac{1}{p}\right). \tag{16.1.5}$$

The model for births averted once again consists of comparing two groups of women, one group initially using the contraceptive and the other never using it, and following through successive segments. Those initially using the contraceptive will take an average of t months to become pregnant; those not using it, an average of $1/p$ months. Suppose again that the nonfecund time of childbearing plus the postpartum anovulatory period is s months for all women, and that we count only pregnancies leading to live births (i.e., disregarding miscarriages and stillbirths). Then those practicing contraception will average a live birth every $t + s$ months, that is to say, their monthly birth rate will be $1/(t + s)$. The group not using the contraceptive will average a birth every $(1/p) + s$ months, and the corresponding rate is $1/[(1/p) + s]$. The problem is now solved: the reduction in monthly birth rates due to the contraceptive is the difference

$$\frac{1}{(1/p)+s}-\frac{1}{t+s},$$

and as a fraction of the birth rate without the contraceptive this reduces to

$$1-\frac{(1/p)+s}{t+s}. \tag{16.1.6}$$

If an IUD is 95 percent efficient and is inserted in each of a group of women of natural fertility with $p = 0.2$ chance of a pregnancy leading to a live birth each month, so that for protected women the chance of an accidental pregnancy is 0.01 each month, and if the chance of a woman's dropping the contraceptive and reverting to natural fertility in any given month is $d = 0.03$, and if the nonfecund period of pregnancy and its aftermath is $s = 17$ months, then $t = 28.75$ months. Expression 16.1.6 says that users of the method avert 0.52 of the births that would otherwise occur.

But we should take account of the fact that women do not ordinarily go directly from natural fertility to modern contraception. The recruits to an IUD program have ordinarily been restricting their births in one way or another. Suppose that they have been practicing rhythm with care and attaining 90 percent efficiency, so that the chance of childbearing if they did not have the IUD is $p = 0.02$, and it is to this practice that they revert if they drop the IUD. Then $t = 62.5$; and, entering $p = 0.02$ in place of $p = 0.2$ in (16.1.6), we find the fraction of births averted to be 0.16 rather than 0.52 as in the preceding paragraph.

The above argument, due to Potter (1970), illustrates once more the general point of causal inference that nothing can be said about the effect of the IUD without specifying what the couples concerned would be doing without it. That such specification is important appears from our numbers: with natural fertility as the alternative to the IUD 52 percent of births are averted, whereas with 90 percent efficient contraception in the background only 16 percent are averted.

The above argument emerges in the simplest possible form when we try to see the effect of abortion on the birth rate.

16.1.4 Why 1000 Abortions Do Not Prevent 1000 Births in a Population

That the logic of individuals becomes grossly misleading when applied to populations is implicit in much of the work in this book. The contrast between individuals and populations is especially sharp in regard to births averted by abortion. If we think of a woman aborting a pregnancy that would have led to a live birth, then one abortion has indeed prevented one birth. But 1000 abortions in a population generally prevent far fewer than

1000 births. To find how many they do prevent we must reckon in terms of each woman's time—how long she takes to have a birth, how long she is tied up in having an abortion. Once again only conceptions potentially leading to live births will be considered, that is to say, spontaneous abortions will be disregarded (Potter 1972).

A woman who has just conceived may decide to have an abortion in the second month and be sterile for 1 further month, a total time from conception of 3 months. Suppose that she is then fecund again, with the same probability 0.2 of conceiving in each month without contraceptive protection. To arrive at this point from the last previous fecund condition has taken her the 3 infertile months before and after the abortion, plus the preceding expected 5 months to pregnancy, 8 months in all. The 8 months represent the time out from childbearing due to one abortion. Only if this length of time were sufficient to have a child would one abortion prevent one birth in the population. If the cycle for having a child is 22 months on the average, as in this illustration, the abortion has prevented only 8/22 of a birth. On these assumptions nearly three abortions are required to avert one birth.

More generally, we can see how many births are prevented by abortions taking place at such time after the onset of pregnancy that the woman is infecund for an expected a months. The length of the cycle involving one abortion averages $(1/p)+a$, and the length of the cycle involving one birth averages $(1/p)+s$; the number of the former that will fit into the latter is

$$\frac{(1/p)+s}{(1/p)+a}. \tag{16.1.7}$$

This being the number of abortion cycles required to fill the time that will be taken by one birth cycle, it is also the number of abortions that will prevent one birth. The model is deterministic; it compares two women going through repeated cycles, one involving births and the other involving abortions, without allowing for variation in the length of cycle or in fecundity among women.

16.1.5 Abortion as a Backup to Contraception

Expression 16.1.7 refers to a population that does not use contraception. In populations that do practice birth control, the fractional effect of abortion is much greater, and (16.1.7) can be readily modified to show how much.

To apply the argument to our new problem we write $p(1-e)$ in place of p as the probability of conceiving in a particular month and again go through the whole of the preceding argument. The mean length of time to pregnancy for fertile couples becomes $1/p(1-e)$, and the number of abortions that will prevent one birth is now

$$\frac{[1/p(1-e)]+s}{[1/p(1-e)]+a}. \tag{16.1.8}$$

Entering $p = 0.2$, $e = 0.95$, $s = 17$ months, and $a = 3$ months gives

$$\frac{(1/0.01)+17}{(1/0.01)+3} = \frac{117}{103} = 1.14$$

abortions to prevent one birth.

This is a very different outcome from the no-contraception case. With unprotected intercourse nearly three abortions are required to prevent one birth. With 95 percent efficient contraception only about one and one-seventh abortions are needed to prevent one birth. If the efficiency of contraception were higher than 0.95, an abortion would have even more impact.

Although this section has used the length of the conception and birth cycle to obtain birth rate in the context of contraceptive effects, it is relevant in many other contexts. The interval between births is an important parameter in such species as whales (e.g., Barlow and Clapham 1997, Caswell et al. 1999), elephants (Wu and Botkin 1980, Moss 2001), and the great albatrosses (e.g., Croxall et al. 1990). This interval may change in response to environmental factors, and the result can be used to project their impact on the population (Caswell et al. 1999).

The absorbing Markov chain approach of Chapter 11 can be used to calculate the interbirth interval in complex stage-classified life cycles if reproducing females are identified as a stage (Fujiwara and Caswell 2001, Fujiwara et al. 2004). The approach is to make reproduction an absorbing state and calculate the mean time to absorption in a chain conditional on reaching that state before death (see Section 11.1.2.2). It could be applied directly to matrix versions of multistate models of fertility, such as those of Wood et al. (1994) or Yashin et al. (1998).

16.2 Measurement of Fertility and Fecundity

According to the usual English language definitions, fecundity is the uninhibited biological capacity of women to bear children; fertility is the number of children borne under existing social conditions. Fertility is fecundity modified by contraception and other kinds of intervention. (Ecologists often switch the definitions.)

Fertility is directly measurable—the birth of a child is a publicly recognized event, and its recording is merely a matter of organization and attention, especially by those, parents and doctors, who are in a position to observe the event as it occurs and so have the necessary facts about it.

But the underlying fecundity depends on decidedly private circumstances not generally known even to the couples concerned. These include viability of sperm and ovum as this affects the length of the fertile period, and other factors hidden from direct observation. The problem of measuring these and their influences on the birth rate will be the subject of this section.

16.2.1 Probability of Conception by Days of the Month

A rough way of calculating probability of conception, making use only of the frequency of intercourse and the length of the fertile period, is implied in the work of Glass and Grebenik (1954). Suppose that a couple have intercourse n times in the menstrual cycle, which includes a fertile period of f days out of a total of 25 nonmenstruating days per month. Then if coitus is unplanned in relation to the fertile period, the probability that it will occur at least once during the f fertile days is the complement of the probability that it will not occur at all during those days: the chance that any particular coitus will take place during the nonfertile period is $1-f/25$. If different occurrences of intercourse are independently random, with no spacing, the chance that all n will take place during the nonfertile period is this quantity to the nth power: $(1-f/25)^n$. The chance p that at least one coitus will take place during the fertile period must be the complement of this last:

$$p = 1 - \left(1 - \frac{f}{25}\right)^n ,$$

and if ovum and sperm are healthy and behave as expected, this is the chance of conception during the month in question.

For a given n the probability is increased insofar as there is a degree of regularity in intercourse, for instance, if it occurs only once in each 24-hour period (Jain 1969). Divide the nonmenstruating part of the month into 25 separate days, each a 24-hour interval, and suppose that in f of these the woman is fertile; then the chance of avoiding the first of the f days with coitus on n (separate) random days is $1-n/25$. The chance of avoiding all f days with n acts of coitus spread over different days is $(1-n/25)^f$, and hence the chance of conception is

$$p = 1 - \left(1 - \frac{n}{25}\right)^f .$$

We need to allow not only for frequency of intercourse but also for the moment when intercourse occurs in relation to the moment of ovulation, using the best knowledge or guesses regarding the probability of conception for intercourse on the day of ovulation, 1 day earlier, 2 days earlier, or 1 day later. Lachenbruch (1967) set up a model that incorporates these probabilities and simulated it by computer to obtain numerical results. For couples using rhythm as a means of contraception and having their intercourse in two "humps," one before and one after ovulation, he found probabilities of conception of 0.07 to 0.20, and commented that the time he allowed for intercourse would be too short for most couples. "Bracketed rhythm," in which the couple have intercourse on the last "safe" day before ovulation and the first "safe" day after, leads to a fairly high value of the probability of conception—0.20 with the assumptions made. A feature of bracketed intercourse is that the total frequency of intercourse has almost no effect on conception.

To estimate probabilities on conception on the several days of the cycle, Barrett and Marshall (1969) followed 241 fertile British couples, mostly 20 to 40 years of age and not using birth control. They obtained from each couple each month a calendar showing dates of intercourse, and a temperature chart from whose rise at midmonth the time of ovulation could be read. Then, if p_i is the probability that intercourse on the ith day will lead to conception, the chance $1-p$ that conception did not occur during a particular month must be the product

$$1 - p = \prod_i (1 - p_i)^{x_i}$$

taken over the days of the month, where x_i is 1 if intercourse took place on the ith day and 0 if it did not.

One would like to find values of p_i such that the above product for $1-p$ comes as close as possible to 1 for the months when conception did not occur, and to 0 for the months when it did. Barrett and Marshall took the logarithm of the likelihood and maximized for the whole sample of cycles. Their estimates for the 5 days before ovulation and the 1 day after it were as follows:

$p_{-4.5}$	0.13
$p_{-3.5}$	0.20
$p_{-2.5}$	0.17
$p_{-1.5}$	0.30
$p_{-0.5}$	0.14
$p_{0.5}$	0.07

Thus the highest probability of conception, 0.30, was for the day 24 to 48 hours before ovulation. Outside of the above 6-day range the probabilities were not significantly different from zero.

From these numbers it follows that daily intercourse gives a probability of conception

$$p = 1 - (1 - 0.13)(1 - 0.20) \cdots = 0.68.$$

The probability is 0.43 for intercourse every second day, 0.31 for every third day, and 0.24 for every fourth day (numbers rounded after calculation, and without making allowance for the effect of frequency of intercourse on the production of sperm).

16.2.2 Mean Fecundity from Surveys

The probability p of conceiving in a given month for a group of fertile women is typically sought in order to compare it with p', the corresponding ratio for another group of women. When p applies to women not practicing contraception, it is an estimate of fecundity or natural fertility; the amount by which p', for a group of women practicing contraception, is lower measures the efficiency of that form of contraception. The most obvious way of obtaining estimates is by observing waiting times until pregnancy, a subject to which we proceed.

Homogeneous Populations. The measurement of fertility, either natural or with contraceptive protection, depends on data for a group of women all of whom are having intercourse and are nonpregnant; suppose all to be subject to the same probability p of conceiving in each month, where p is greater than zero. Suppose that N fertile women, just married, are surveyed month by month until they become pregnant; as each becomes pregnant, she drops out of observation. Let the number who become pregnant in the first month be N_1, the number who become pregnant in the second month N_2, and so on. Then the probability of conception in any month, p, is estimated from the first month's data as the ratio N_1/N. This leaves $N - N_1$ women starting the second month in a fecund condition, and the estimate of p from that month's data is the ratio $N_2/(N-N_1)$, and similarly for later months. A series of estimates of p is thus provided

$$\begin{aligned}
\hat{p}_1 &= \frac{N_1}{N} \\
\hat{p}_2 &= \frac{N_2}{N - N_1} \\
\hat{p}_3 &= \frac{N_3}{N - N_1 - N_2} \\
&\vdots \\
\hat{p}_m &= \frac{N_m}{N - N_1 - N_2 - \cdots - N_{m-1}},
\end{aligned}$$

where observation stops after the mth month.

To make use of all the information we must average the several values of $\hat{p}_1, \hat{p}_2, \ldots$. If the true probability p is the same for all women and for all months, the right average is one that weights the several months by their sample sizes: we need to weight our $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_m$ by the number of exposed women on which each is based. Since $\hat{p}_1$ is based on N women, $\hat{p}_2$

on $N - N_1$, and so on, the estimate is

$$\frac{N\hat{p}_1 + (N - N_1)\hat{p}_2 + \cdots + (N - N_1 - N_2 - \cdots - N_{m-1})\hat{p}_m}{N + (N - N_1) + \cdots + (N - N_1 - N_2 - \cdots - N_{m-1})},$$

or, entering the estimates $\hat{p}_1, \hat{p}_2, \ldots$ from above,

$$\frac{N_1 + N_2 + \cdots + N_m}{N + (N - N_1) + \cdots + (N - N_1 - N_2 - \cdots - N_{m-1})}, \tag{16.2.1}$$

supposing that all women are followed to the mth month.

This widely used index, due to Pearl (1939, p. 296), will be referred to as $\hat{p}_p$. It contains the total number of conceptions during the m months of observation in its numerator, and its denominator is the number of woman-months of exposure, if the month of conception is counted into the exposure. The index is intuitively appealing quite apart from the statistical argument above. Multiplied by 1200, it gives pregnancies per 100 woman-years of intercourse, and this rate is often calculated and published. Not only does $\hat{p}_p$ seem intuitively reasonable, but also, if all women were equally susceptible, it would be the correct measure, and we would need to go no further in the search for a measure of fecundity.

A Heterogeneous Population with Fecundity Constant for Each Woman. However, we know that some women are more fecund than others, and we seek from the survey a suitable average of their several p values. The women who are most fecund will tend to become pregnant first, so the $\hat{p}_1, \hat{p}_2, \ldots$ for the several months are estimating different quantities. The estimate for the first month $\hat{p}_1 = N_1/N$, refers to unselected women and is an unbiased estimate of the mean p. Since those who become pregnant drop out of observation, no later month refers to unselected women. The ith month, for any $i > 1$, omits some women selected for their fecundity, and so the estimate derived from it, $N_i/(N - N_1 - N_2 - \cdots - N_{i-1})$, must be an underestimate of the fecundity of the original N women.

The p values differed from month to month in the sample size on which they were estimated in the model underlying p_p, and they differed in no other way. With such homogeneous material the correct way to weight a number of estimates of the same parameter is by the quantity of information contained in each estimate, that is, by the size of sample available in each month. With heterogeneity among women the pregnancy ratios are genuinely different in the different months, and to weight by the quantity of information, that is, the sample size, would be incorrect. To avoid considering two different problems at once we will now suppose that the sample is large, so that random variation can be disregarded. What is wanted is a population average, in which each woman counts once, and hence we must weight the women of a given fecundity class according to the number of women in that class in the population. (See Chapter 19 for a more general discussion of heterogeneity.)

Suppose (Sheps, in Sheps and Ridley 1966) that the N representative women with whom we start out include $Nf(p)\,dp$ women of susceptibility or fecundity between p and $p+dp$, where dp is small. Among women of fecundity p we will expect the fraction p to become pregnant in the first month, $1-p$ to go on to the second month, and, of these, $(1-p)p$ to become pregnant then, and so on. In other words, we will expect

$$Npf(p)\,dp$$

women to become pregnant in the first month,

$$N(1-p)pf(p)\,dp$$

in the second month, . . . , and

$$N(1-p)^{m-1}pf(p)\,dp$$

in the mth month—all this for a given susceptibility p.

To find the corresponding numbers for all women we add through the several susceptibility groups, which in the limit is the same as integrating with respect to p. The total expected conceptions for the ith month will be

$$N\int_0^1 (1-p)^{i-1}pf(p)\,dp. \tag{16.2.2}$$

Expression 16.2.2 shows that for $0 < p < 1$ the number of conceptions steadily decreases with time. Most of the decrease is due to women dropping out through pregnancy, and there is no provision for bringing them back in this way of measuring fertility.

The number of conceptions cannot but decline to zero in the course of time as the entire group drops out through pregnancy.

But aside from this gross fact, which applies to a homogeneous group as well, there is a selection factor that can be important if the variation in p, expressed by $f(p)$, is substantial. The pregnancy rate during the ith month is (16.2.2) divided by the number of women still under observation

$$p_i = \frac{\displaystyle\int_0^1 (1-p)^{i-1}pf(p)\,dp}{\displaystyle\int_0^1 (1-p)^{i-1}f(p)\,dp}, \tag{16.2.3}$$

which can be regarded as the mean of p weighted by $(1-p)^{i-1}f(p)$. Because $1-p$ is less than unity, the weighting function $(1-p)^{i-1}f(p)$ shifts downward (i.e., to the left on the usual form of chart) as i increases, so p_i must decline steadily with i. This is a selection arising from the removal by pregnancy of the more fertile women.

The two processes are shown in Table 16.2, with a simple example in which half the women have $p = 0.2$ and half have $p = 0.3$, and q is written for $1-p$ in (16.2.2) and (16.2.3). It shows in column 1 the chance of

Table 16.2. Example of removal of more fertile women in successive months, using $f(p) = 0.5$ for $p = 0.2$ and $f(p) = 0.5$ for $p = 0.3$; that is, half the women are of fecundability $p = 0.2$ and half are of fecundability $p = 0.3$

Month	Probability of conceiving in ith month for all women entering $\int_0^1 pq^{i-1}f(p)\,dp$	Fraction of women remaining at beginning of month $\int_0^1 q^{i-1}f(p)\,dp$	Probability of conceiving in ith month for women who have not conceived by $(i-1)$th month $\dfrac{\int_0^1 pq^{i-1}f(p)\,dp}{\int_0^1 q^{i-1}f(p)\,dp}$
1	0.25	1.000	0.25
2	0.185	0.750	0.247
3	0.138	0.565	0.243
4	0.103	0.428	0.240
5	0.077	0.325	0.237
6	0.058	0.248	0.234
7	0.044	0.190	0.231
8	0.033	0.146	0.228
9	0.025	0.113	0.226
10	0.019	0.087	0.223
15	0.0054	0.025	0.213
20	0.0016	0.0078	0.207
30	0.00016	0.00079	0.2020
50	0.0000018	0.0000089	0.20014

becoming pregnant in the ith month as forecast for a woman when she comes under observation in month 1, and also (column 3) the chance of a pregnancy in the ith month for a woman who has gone through the first $i-1$ months without becoming pregnant. The rapid decline in the first column is no surprise, while the third column exhibits the more subtle selection effect, which is appreciable even with the small variation of our example. The mean fecundability started at $p = 0.25$, the mean of 0.2 and 0.3, and by the tenth month had fallen to 0.223. The drop is more than halfway to 0.2, to which it would ultimately tend as i became large. The greater the variation, the more rapid is the decline in average fertility among the women who remain.

To express (16.2.3) in terms of variation among women, expand $(1-p)^{i-1}p$ of the numerator and $(1-p)^{i-1}$ of the denominator in a Taylor series around $\bar{p}$, the mean $\int_0^1 pf(p)\,dp$. The constant terms are $(1-\bar{p})^{i-1}\bar{p}$ in the numerator and $(1-\bar{p})^{i-1}$ in the denominator, and the linear terms vanish, with the result

$$p_i \approx \bar{p} - \frac{(i-1)\sigma^2}{1-\bar{p}}, \tag{16.2.4}$$

on using the fact that $1/(1+\alpha) \approx 1-\alpha$ for α small.

In application to the distribution of Table 16.2, consisting of two spikes, $\bar{p} = 0.25$. The general rule is that the standard deviation of two numbers is half the interval between them; if the numbers are a and b, their variance is

$$\sigma^2 = \frac{a^2 + b^2}{2} - \left(\frac{a+b}{2}\right)^2 = \left(\frac{a-b}{2}\right)^2 .$$

In this case $a = 0.2$ and $b = 0.3$, so $\sigma^2 = 0.0025$. Then from (16.2.4) we have for p_{10}

$$\begin{aligned} p_{10} &\approx \bar{p} - \frac{(i-1)\sigma^2}{1-\bar{p}} \\ &= 0.25 - \frac{(9)(0.0025)}{0.75} = 0.220, \end{aligned}$$

as opposed to the 0.223 shown in Table 16.2. The difference is due to neglect of higher moments and other approximations, the effect of which becomes more serious as i increases.

16.2.3 The Pearl Index Is the Harmonic Mean of the Distribution

The Pearl index $\hat{p}_p$ of (16.2.1) is shown to estimate the harmonic mean of the several p. For the expected value of N_1 is the initial number N times p, of N_2 is Nqp, of N_3 is Nq^2p, and so on. The denominator of (16.2.1) can be taken as the numbers of women nonpregnant at the beginning, after 1 month after 2 months, and so on. Entering these for any fixed p and then integrating over the range of p [overlooking that $E(X/Y) \neq EX/EY$]:

$$p_p = \frac{\displaystyle\int_0^1 pf(p)\,dp + \int_0^1 (1-p)pf(p)\,dp + \cdots}{\displaystyle\int_0^1 f(p)\,dp + \int_0^1 (1-p)f(p)\,dp + \cdots},$$

where we suppose that the survey continues until the last woman becomes pregnant, or, more reasonable, that nonpregnant women left at the end of the survey are excluded from all of its records. In the numerator again we change p to $1-q$ and assemble the integrals to obtain

$$\begin{aligned} p_p &= \frac{\int_0^1 [(1-q) + q(1-q) + q^2(1-q) + \cdots] f(p)\,dp}{\int_0^1 (1 + q + q^2 + \cdots) f(p)\,dp} \\ &= \frac{\int_0^1 f(p)\,dp}{\int_0^1 (1 + q + q^2 + \cdots) f(p)\,dp} \end{aligned}$$

$$= \frac{1}{\int_0^1 [1/(1-q)] f(p)\,dp} = \frac{1}{\int_0^1 (1/p) f(p)\,dp}. \tag{16.2.5}$$

The quantity p_p, of which $\hat{p}_p$ given by (16.2.1) is an estimate, is thus shown to be the harmonic mean of the p's, a statement that is meaningful once all infecund women have been removed from the records.

16.2.4 The Gini Fertility Measure

Gini's (1924) way of providing a measure of fertility that is unbiased in the face of variation among women is effectively to confine the index to the pregnancies and exposure of the first month, bound to be unselected for susceptibility if the sample is unselected. His index is

$$\hat{p}_g = \frac{N_1 - N_{m+1}}{N_1 + N_2 + \cdots + N_m}, \tag{16.2.6}$$

where the denominator totals the women of proven fertility, and the numerator has the relatively unimportant subtraction, if m is large, of those who become pregnant in the $(m+1)$th month.

If one wants to see formally what is happening in $\hat{p}_g$ when those not pregnant by the mth month cannot be neglected, the algebra is only slightly more involved:

$$\hat{p}_g = \frac{N_1 - N_{m+1}}{N_1 + N_2 + \cdots + N_m}$$

is an estimate of

$$\frac{N\int_0^1 pf(p)\,dp - N\int_0^1 pq^m f(p)\,dp}{N\int_0^1 (p + pq + pq^2 + \cdots + pq^{m-1}) f(p)\,dp} = \frac{\int_0^1 p(1-q^m) f(p)\,dp}{\int_0^1 (1-q^m) f(p)\,dp}. \tag{16.2.7}$$

This is the arithmetic mean value of p, not quite in the original distribution $f(p)$, but in the slightly different distribution proportional to $(1-q^m)f(p)\,dp$

Comparison of Pearl and Gini Estimates. If p does not vary among fertile women, or does not vary greatly, we can be indifferent as to whether their arithmetic or harmonic mean is estimated. Thus in the homogeneous case $\hat{p}_p$ is the best estimate of the common p in that it has the smallest sampling error. If, however, p does vary considerably, the harmonic mean is no substitute for the arithmetic, being always below it, and to use $\hat{p}_p$ is to minimize sampling error at the cost of substantial bias. Take the arithmetic example of Table 16.2, in which a group of women is equally divided among those with probability of conceiving 0.2 and those with probability 0.3; then the arithmetic mean of their fecundity is 0.25 and the harmonic mean is 0.24, a small difference. For a group equally divided among those with probabilities of conceiving of $\frac{1}{6}$, $\frac{1}{4}$, and $\frac{1}{2}$, the arithmetic mean is

$$\frac{\frac{1}{6}+\frac{1}{4}+\frac{1}{2}}{3}=\frac{0.917}{3}=0.306.$$

The harmonic mean is

$$\frac{3}{1/\frac{1}{6}+1/\frac{1}{4}+1/\frac{1}{2}}=\frac{3}{6+4+2}=0.25,$$

or about 18 percent low.

Thus mean fecundity is understated by $\hat{p}_p$ whenever variation among women is considerable. The Pearl index, given as pregnancies per 100 woman-years of exposure, if unadjusted for selection will underestimate the average probability of conception per month for all women in the unselected group. In commonsense terms this is so because $\hat{p}_p$ puts weight on later months when the less fecund women are disproportionately represented. This defect of $\hat{p}_p$ is unimportant if couples are followed over a short period. It is not a defect at all if we are interested in the mean fecundity over a period of time of a group of initially fecund women, in which those who become pregnant drop out and are not replaced. Moreover, the mean waiting time to pregnancy in a heterogeneous group of women will be approximated by $1/\hat{p}_p$, so that, if waiting time is the subject of interest, the harmonic mean implicit in $\hat{p}_p$ is the one wanted.

16.2.5 Excursus on Averages

In the present section we have a type of selection to produce p_p that lowers the average below the unselected p_g. The opposite occurred in Section 15.3, where it appeared that choosing a woman at random gave the expected number of her daughters as the arithmetic mean $\sum_i if_i$, where f_i is the fraction of women who have i daughters; choosing a girl at random and asking how many daughters her mother had gave a larger average. It is worth contrasting the four main kinds of average that enter demographic work.

If families are chosen with probability proportional to number of daughters, the expected value is found by weighting by i and dividing by the total number of daughters ($\sum f_i = 1$ but $\sum if_i \neq 1$):

$$\frac{\sum(i)(if_i)}{\sum if_i}=\frac{\sum i^2 f_i}{\sum if_i}.$$

This has the old name of contraharmonic mean (according to the *Oxford English Dictionary*). If it is designated as C and the arithmetic mean as

A, the difference is (Section 15.3)

$$C - A = \frac{\sum i^2 f_i}{\sum i f_i} - \sum i f_i$$
$$= \frac{\sum i^2 f_i - (\sum i f_i)^2}{\sum i f_i} = \frac{\sigma^2}{A},$$

where σ^2 is the variance of the distribution. Thus C is always greater than A, except in the trivial case where the numbers being averaged are the same.

A similar aspect of selection, but over time, appears in Feller (1971, p. 12), who explains why we have to wait so long at a bus stop. Suppose, for example, that the buses arrive independently at random according to a Poisson process with constant α, so that the mean time between buses is $1/\alpha$. When a would-be rider shows up at the stop, he ought on the average to be midway between two buses and have to wait only $\frac{1}{2}(1/\alpha)$. That is one way of calculating. The second is to say that, since the Poisson process has no memory, and one moment of arrival has the same waiting prospect as another, the time to the next bus must on the average be $1/\alpha$. Unfortunately the second answer is the correct one. A selection occurs by which one more often arrives in a long interbus period than in a short one.

A different aspect of waiting times is evident when women are subject to different chances of having a child. What is the mean interchild period in a group of women if the probability of pregnancy is p_i for the ith woman, say in a given month? The expected period for the ith woman is $1/p_i$, and the mean time to conception $(1/n)\sum(1/p_i)$ months. This is the reciprocal of the harmonic mean of the probabilities of conception; it is not $n/\sum p_i$, the reciprocal of the arithmetic mean. The Pearl index p_p is another example of the same harmonic mean.

The geometric mean turns up in applications of the life table. The chance of a person surviving from age 30 to age 40, say, is the geometric average chance of surviving over the 10 years taken to the tenth power. It has to be a geometric average because the chance of survival is

$$\frac{l_{40}}{l_{30}} = \left[\sqrt[10]{\left(\frac{l_{40}}{l_{39}}\right)\left(\frac{l_{39}}{l_{38}}\right)\cdots\left(\frac{l_{31}}{l_{30}}\right)} \right]^{10}.$$

Alternatively, if the chance of dying in the ith of n years is μ_i, with an arithmetic average $\bar{\mu}$, the chance of surviving over the period is $(e^{-\bar{\mu}})^n$, and this is the geometric mean G of the chances of surviving the individual years.

Thus we have instances of the arithmetic (A), geometric (G), harmonic (H), and contraharmonic (C) means. For integral variables i with weights

f_i such that $\sum f_i = 1$, these may be written, respectively, as

A	G	H	C
$\sum i f_i$	$\prod i^{f_i}$	$\dfrac{1}{\sum(f_i/i)}$	$\dfrac{\sum i^2 f_i}{\sum i f_i}$

For N unweighted values X_i, integral or not, these means are

$$\frac{\sum X_i}{N} \qquad \sqrt[N]{\prod X_i} \qquad \frac{N}{\sum(1/X_i)} \qquad \frac{\sum X_i^2}{\sum X_i}$$

Differences are not trifling. For a population with two values, $X_1 = 1$ and $X_2 = 4$, we have for the four averages

$$2.5, \qquad 2, \qquad 1.6, \qquad 3.4.$$

[Show that the contraharmonic mean is as much greater than the arithmetic as the arithmetic is greater than the harmonic; that is, $C - A = A - H$. Show also that $G = \sqrt{AH}$. Under what conditions do they hold?]

16.2.6 Graduation Uses Information Efficiently

Given that in practice samples used for fertility studies are small, one would like to avoid the sampling instability of the Gini index, as well as what is, from one point of view, the bias of the Pearl index. One would also like to know something of the variation in fecundity among women that is not revealed by $\hat{p}_p$ or by $\hat{p}_g$. Variance could be inferred from the difference by solving (16.2.4) for σ^2, but this would have substantial error. Fortunately a graduation due to Potter and Parker (1964) provides variance, along with low sampling error and absence of bias in the mean.

16.2.7 Mean and Variance Simultaneously Estimated by Graduation

Potter and Parker (1964) fitted a beta distribution proportional to

$$p^{a-1}(1-p)^{b-1}$$

whose mean fecundity $\bar{p}$ and variance σ^2 are as follows:

$$\begin{aligned} \bar{p} &= \int_0^1 p f(p)\,dp = \frac{a}{a+b} \\ \sigma^2 &= \int_0^1 (p-\bar{p})^2 f(p)\,dp = \frac{ab}{(a+b+1)(a+b)^2}. \end{aligned} \tag{16.2.8}$$

The a and b are not directly known, but we do know the mean and variance of the waiting times of those becoming pregnant in terms of the same constants a and b:

$$\bar{w} = \frac{a+b-1}{a-1}, \qquad a > 1; \qquad \sigma_w^2 = \frac{(ab)(a+b-1)}{(a-1)^2(a-2)}, \qquad a > 2.$$

All that is needed is to solve for a and b in terms of the known $\bar{w}$ and σ_w^2. After extended but straightforward algebra the estimates

$$a = \frac{2\sigma_w^2}{\sigma_w^2 - \bar{w}^2 + \bar{w}}, \qquad b = (\bar{w}-1)(a-1) \tag{16.2.9}$$

are reached. Substituting for a and b in (16.2.8) would give the mean and variance of the probabilities of conceiving (not directly observable) in terms of the mean and variance of waiting times as observed. In the example used by Potter and Parker, $\bar{w} = 5.47$ months and $\sigma_w^2 = 89.98$, and from (16.2.9) these provide $a = 2.746$ and $b = 7.806$; therefore mean fecundity $\bar{p}$ is 0.260 and the standard deviation of p among women is $\sigma = 0.129$. (A more detailed secondary account is found in Keyfitz 1968, p. 386.)

Improvements on the moments fitting sketched above have been published by Majumdar and Sheps (1970). They develop maximum likelihood estimators that make more effective use of the data, these data still consisting of waiting times for the whole group of women followed.

There is a certain arbitrariness in picking a family of probability distributions to represent fecundity on the basis of mathematical convenience rather than empirical evidence, and then drawing conclusions without at least trying alternative distributions. This problem arises generally in studies of heterogeneity (Chapter 19). Recent discussions can be found in Wood and Weinstein (1990), Wood (1994), and Weinberg and Dunson (2000). Understanding patterns of heterogeneous fecundity among women and how they vary among different populations is important in studies of the effects of heterogeneous fertility on variance in completed family size, or the relative effects of male and female fecundity within marriage.

The variation among couples is especially great for a contraceptive method that is ineffective for some users. The ineffective users are selected out by pregnancy, leaving a residue of effective users whose probability of conception is very low. For this reason and others, surveys of groups of women practicing contraception are often tabulated by periods—first year, second year, and so on, so that the decline in the conception rate can be traced. Tabulating in periods has the advantage that cases subsequently lost to follow-up can be included as long as they are observed. The life table method described below rescues the incomplete records.

16.2.8 Life Table Methods for Fertility

The theory so far has dealt with two cases: homogeneous, where all women have the same probability of conception p, and heterogeneous, where their several p values are distributed according to an (unknown) probability distribution. In both we supposed that any given woman has an unchanging probability of conception. If the sample is to be followed for a long period, however, we need to allow also for change in individual women, either because of a decline in fecundity with age or because of increased motivation and skill in using contraception. The life table method in a fashion embraces both factors. It also allows for the selection effect of pregnancy.

To make the table we first calculate the probability of conception month by month. Like any life table, that for conception is based on two kinds of data: number of events, and numbers exposed to risk. In the present case the events are pregnancies, and the exposed are the women under observation, for each month. If P_i women are under observation through the ith month, and A_i (standing for “accident”) conceptions occur among them, the conception rate for the group in that month is $p_i = A_i/P_i$, and the probability of not conceiving is $1 - p_i$ (Potter 1967).

We can multiply together the $1 - p_i$ for successive months, and obtain a column analogous to the life table l_i that represents the chance of a child just born surviving to exact age i. The technique is identical to that for mortality, discussed at length in Chapter 2.

The life table model in which allowance is made for the single decrement of pregnancy can be extended to provide for other risks, including the death of the person, divorce of the couple, discontinuance of contraception, and other contingencies. Among these the possibility that the couple will drop the contraceptive is of the greatest interest for our analysis. Tietze (1962) tells us that pregnancy rates for various IUDs during the 2 years after insertion were considerably lower than discontinuance rates. This is a standard problem in competing risks, of the kind dealt with in Section 2.6, and any of the methods there used would serve in this case too. But the refinements useful for mortality are not necessary for conception, where small samples and biased data are general.

See Weinberg and Dunson (2000) for a discussion of some recent developments along these same lines, and Chapter 19 for discussion of individual heterogeneity in general. Life table methods can be generalized to methods based on hazard functions; these are reviewed in the context of fertility by Wood et al. (1992) and Wood (1994).

16.2.9 Relation of Micro to Population Replacement

We saw that the replacement of a population, the ratio of girl children in one generation to girl children in the preceding generation, is given by

$$R_0 = \int_\alpha^\beta l(a)m(a)\,da.$$

We can factor $m(a)$ into $v(a)f(a)$, where $v(a)$ is the fraction married at age a, and $f(a)$ is the marital fertility rate. We can also go one step further and say of married fertility that

$$f(a) = \frac{12}{1/\{p(a)[1-e(a)]\}+s},$$

where $p(a)$ is natural monthly probability of conceiving at age (a), and $e(a)$ is the efficiency of contraception, both in a particular population. Then we have for the net reproduction rate

$$R_0 = \int_{\alpha}^{\beta} l(a)v(a)\frac{12}{1/\left\{p(a)\left[1-e(a)\right]\right\}+s}\,da, \qquad (16.2.10)$$

but this remains purely formal without some source of information on $p(a)$ and $e(a)$.

16.2.10 How Surer Contraception Reduces the Interval Between Births

At one time "child-spacing" was a euphemism for contraception, and much was said about the benefits to the mother's health if she spaced out her children. This may have been good public relations at a time when authorities frowned on contraception intended simply to reduce family size. While the supposed healthful effects of spacing are still referred to in some countries, yet in fact with the spread of safe and certain contraception women often *reduce* the interval between such births as they decide to have. For the mother who must give up a job in order to look after the family, two or three children of about the same ages cost less in lost earnings than would the same number of children spread over her reproductive life.

Whatever the motivation to compress childraising into a small time interval, the couple could not afford to yield to it when contraception was uncertain. Compromise was necessary; the number of planned children would be held below what was desired in order to allow for accidents. Even if the chance of conception in any one month is as low as 0.01, and the couple have 240 fertile months ahead of them, their expected prospective children are 2.4, whether they want them or not. If they want 2 or 3 children, they cannot afford to have any deliberately. Even if they can achieve a probability of conceiving in any month as low as 0.003, the chance of an accidental pregnancy during 240 months is over one half:

$$1-(0.997)^{240} = 0.514.$$

Thus only perfect contraception, or its equivalent in the form of easy legal abortion, permits thoroughgoing family planning, and when such planning is aimed at saving the time of the mother it will space the children as closely as possible.

The idea of studying fertility in terms of the time required for the different processes involved in conception, pregnancy, lactation, and so on has proven fruitful. Recent surveys have gone further and further in the direction of integrating the underlying biological processes with their demographic consequences (Campbell and Wood 1994, Wood 1994, Wachter and Bulatao 2003), including comparisons with some of our near but nonhuman relatives (Altmann and Alberts 2003).

16.3 Why Three-Child Families Constitute a Population Explosion, Whereas Two-Child Families Would Lead to the Extinction of Mankind

Let us see what three children mean for the growth of a population subject to United States mortality. There are four steps in the calculation, which we will carry out for females; the first three steps establish by how many fertile women a woman past childbearing will be replaced in the next generation.

1. If fertile women surviving through childbearing average three children in all, they will average 0.488×3 or 1.464 girls, the proportion of United States, 1967, births that were girls being 0.488.
2. Not all of the girls will live to reproduce in turn. Again according to data for the United States in 1967, the average fraction who survive to childbearing, that is, the ratio of the net to the gross reproduction rate, is 0.9665. By multiplying 0.9665 by 1.464 we go from gross to net reproduction and find $0.9665 \times 1.464 = 1.415 = R_0$.
3. An unknown fraction of girls are physiologically incapable of having children or do not wish to have them; let us allow 10 percent to cover these. This brings us to $1.415 \times 0.9 = 1.273$ as the number of fertile women by which a fertile woman is replaced on the average.
4. The period over which this replacement ratio applies is one generation, or 26.14 years for 1967 in the United States, about the same as the mean age of childbearing. We need the 26.14th root of the replacement ratio 1.274 to obtain the annual ratio of increase:

   $$1.274^{1/26.14} = 1.00931,$$

   or, as an annual rate compounded momently, $\log 1.274/26.14 = 0.00926$, either way an increase of 0.0093 or 0.93 percent.

Thus, with the assumptions here made, including that the ages of childbearing are the same as those of the United States in 1967, and that 10 percent of women do not have any children, a three-child average for fertile couples implies an increase of 0.93 percent per year. This works out to a doubling in 75 years, and a multiplication by 16 in less than 300 years. With three-child families from now on the United States would have a population of about 3 billion within three centuries.

On the other hand, an average of two children per fertile married woman would not suffice to maintain the population. Calculation similar to that above shows that the population would ultimately change in the ratio of $2 \times 0.488 \times 0.9665 \times 0.9 = 0.849$ per generation, or the 26.14th root of this, 0.9938 per year, which is a decline of 0.62 percent per year. The half-life would be 111 years, and with two-child families to fertile women the United States population would fall to 30 million in a little over three centuries.

Interpolation from the above will give the average that would hold the population stationary. More directly, call the number x, and solve the equation

$$x \times 0.488 \times 0.9665 \times 0.9 = 1.$$

The result is 2.36 children per fertile married women, In summary, with present United States death rates an average of three children leads to a population of more than 3 billion in three centuries; what we need is an average of just 2.36 children.

An average of 2.1 children is often quoted as the bare replacement level. What is meant is 2.1 children averaged over all women, married and fertile or not. This is obtained by some such equation for the net reproduction rate as $x \times 0.488 \times 0.9665 = 1$, or $x = 2.12$, still with United States, 1967, data. Our 2.36 is the average number of children for couples who will have children. Its greatest weakness is the 10 percent allowance for celibacy and sterility.

These numbers show that the average family size over a period of time must be finely adjusted; even small departures continued for long lead to intolerable increase or decrease. Our present average for fertile married women is about two children. This number is less than we want indefinitely, but we have time to make the adjustment; intolerable changes come only over the course of generations and centuries.

Countries with high fertility, on the other hand, are not in a position to wait. Mexico in the late 1960s was an example. Its fertile women averaged six children each, implying a growth rate of 3.5 percent per year, a doubling in 20 years, and a multiplication by 16 in 80 years. But as of 2004, Mexico has a fertility of only 2.49 children per woman, and a growth rate of only 1.14 percent per year.

Patterns of fertility have evolved since this chapter was originally written, but the importance of the replacement level remains. It is now estimated that, as of early 2004, more than half of the world population now lives in countries or regions where fertility is below replacement level (Cohen 2003, Wilson 2004). The global average is above replacement because the half of the world with above-replacement fertility averages 3.6 children, while that with below-replacement fertility averages 1.6 children (Wilson 2004).

When both mortality and fertility are very high, as they must have been through most of unrecorded history, say at 40.00 and 40.02 per thousand, respectively (Section 1.2), an equilibrium mechanism analyzed by Frisch (1972) could have been operative. Undernourished women reach menarche later and menopause sooner than well-nourished women. They are also subject to more gaps in fecundability during the course of their reproductive lives. Hence periods of severe food scarcity result in lower births, by a purely biological causation, and when death rates are high this means population decrease. Although the mechanism does not produce stability in the face of short-period fluctuations, it can adapt population numbers to long-term declines or increases in food supplies. Evidently no such mechanism can produce an optimum population under modern conditions, where the gap between birth and death rates in many countries has become very wide.

16.4 A Family-Building Strategy to Avoid Extinction

Common sense suggests that the more children one has the less likely is the extinction of one's line of descent. This is only partly true, however, for extinction depends especially on the *variation* in the number of children in later generations, and in particular on the chance of having zero children. In avoiding the extinction of family lines a high average birth rate does not necessarily provide an advantage over a low birth rate.

Observing that many of the great men of the past had no living descendants, some writers inferred the deterioration of the race. Galton responded that before drawing such a conclusion one must know what fraction of people in general have no descendants. He initiated branching process theory (Galton and Watson 1874), which subsequently became a rich field of mathematical research with many and diverse applications. The analysis shows that in an increasing population each member either has zero descendants or has many—the chance that any of us will have exactly one descendant after ten generations is remote. For the mathematics to explore this aspect and others see Harris (1963). For the generalization to processes that distinguish multiple types of individuals (by age or stage), see Pollard (1973) and Chapter 15 of MPM. Fortunately it is possible to derive the probability of extinction with virtually no theory at all. The exposition below, like Galton's statement of the problem, is given in terms of the male line.

Call the probability of having no sons p_0, of having one son p_1, of having two sons p_2, and so on, these probabilities applying independently to all males in all generations, and to be interpreted as sons living to maturity. Designate the chance of extinction of a male line starting with one person as x. Then the chance of extinction of two separately and independently developing male lines must be x^2, of three lines x^3, and so on. If a man has

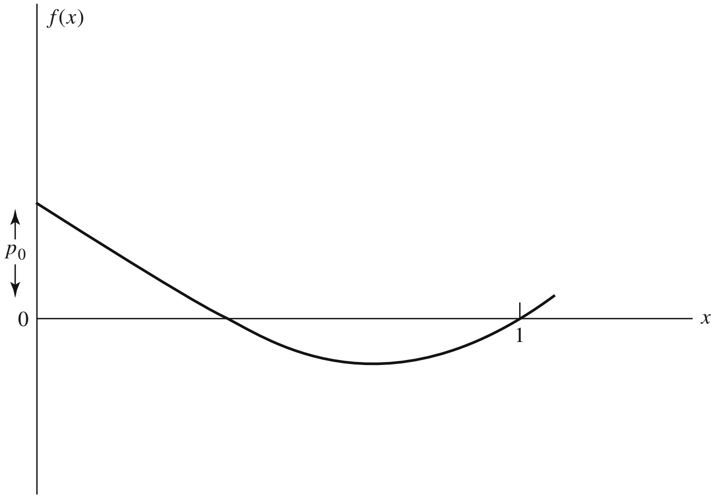


Figure 16.3. General form of the curve $f(x) = p_0 + p_1x + p_2x^2 + \cdots - x$, when the derivative $f'(x) = p_1 + 2p_2x + \cdots - 1$ is positive at $x = 1$, so that mean children are more than 1, showing $f''(x)$ always positive.

no sons his (conditional) chance of extinction is one, if he has one son his chance of extinction is x, if he has two sons his chance of extinction is x^2, and so on. Thus his whole (unconditional) chance of extinction through his sons is $p_0 + p_1x + p_2x^2 + \cdots$, and by definition this must be equal to his chance of extinction x. Hence the equation for x is

$$x = p_0 + p_1x + p_2x^2 + \cdots. \tag{16.4.1}$$

If $f(x) = p_0 + p_1x + p_2x^2 + \cdots - x$, we have $f(0) = p_0$, a positive quantity, and $f(1) = 0$; $f''(x)$ is always positive so $f(x)$ is concave upward. The question whether $f(x) = 0$ has a root between 0 and 1 is answered by whether it approaches 1 on the left from below, the condition for which is $f'(1)$ greater than zero or

$$p_1 + 2p_2x + 3p_3x^2 + \cdots - 1 > 0,$$

and $f'(x)$ is greater than zero at the point $x = 1$ if the mean number of sons is greater than one. Thus, if the mean is greater than one, the curve is sloping upward toward zero at $x = 1$, and there will be at least one root in the interior of the interval between zero and one. In this case the curve will have the general shape of Table 16.3 and Figure 16.3 and the chance of extinction is greater than zero and less than one.

Why do irrelevant roots, such as that at $x = 1$, appear so often in applied work? The reason can be seen only if we make a distinction between the substantive problem and its mathematical formulation. In demography, as in other fields, the substantive problem facing us is identified with a mathematical formulation that is somehow broader, and after solving the

Table 16.3. Values of $f(x) = \frac{1}{4} + \frac{x}{4} + \frac{x^2}{4} + \frac{x^3}{4} - x$, $f'(x) = \frac{1}{4} + \frac{x}{2} + \frac{3x^2}{4} - 1$, and $f''(x) = \frac{1}{2} + \frac{3x}{2}$

x	$f(x)$	$f'(x)$	$f''(x)$
0	0.25	−0.750	0.50
0.1	0.178	−0.692	0.65
0.2	0.112	−0.620	0.80
0.3	0.054	−0.532	0.95
0.4	0.006	−0.430	1.10
0.5	−0.031	−0.312	1.25
0.6	−0.056	−0.180	1.40
0.7	−0.067	−0.032	1.55
0.8	−0.062	0.130	1.70
0.9	−0.040	0.308	1.85
1.0	0.000	0.500	2.00
1.1	0.060	0.708	2.15
1.2	0.142	0.930	2.30

mathematical formulation we have to hand-pick among its answers the one that corresponds to our narrower substantive concern. In a sense we ourselves unknowingly put extra roots into the mathematical representation, and we have no easy mathematical way of eliminating them.

Let us try to get some feeling for what it is in the pattern of childbearing that most affects the chance of extinction of a family line. First the case (Table 16.3) in which the chance of having no sons surviving to maturity is 0.25, of having one son is 0.25, of having two sons is 0.25, of having three sons is 0.25. (This would be an easy population to simulate with a pair of coins.) The equation for x, the probability of extinction, is

$$x = 0.25 + 0.25x + 0.25x^2 + 0.25x^3, \tag{16.4.2}$$

which is

$$x^3 + x^2 - 3x + 1 = 0.$$

The equation is satisfied by $x = 1$, a root of no demographic interest, so we divide by $x - 1$ to obtain the quadratic

$$x^2 + 2x - 1 = 0$$

containing the other roots. This is solved by completing the square as $x = -1 + \sqrt{2} = 0.414$, the negative root being extraneous.

The above hypothetical population has a net reproduction rate of 1.5, which is somewhere between the values for presently less developed and more developed countries. Mexico in 1959–61 had much higher fertility, its net reproduction rate being over 2.5, and yet the chance of extinction (Keyfitz 1968, p. 409) at 0.4066 was just about as high as that of our hypothetical population. (The extinction calculation was based on 1960 census data following the female line for Mexico.)

A population could have even higher fertility than Mexico and yet also show a greater probability of extinction. Suppose that the chance of having no sons surviving to maturity is 0.5 and the chance of having six sons is 0.5. This would give a net reproduction rate of 3.0—higher than the values for recorded populations. But the probability of extinction of a male line would be the zero of $x = \frac{1}{2} + x^6/2$, or $x^6 - 2x + 1 = 0$. A convenient iterative form for solution is obtained by rearranging the equation as $x = (x^6 + 1)/2$. Starting with $x = 0$ on the right-hand side, the iterates are 0.5, 0.508, and 0.509, the root being 0.50866.

At the opposite extreme, a population in which everyone married and each couple had exactly one son (and one daughter) surviving to maturity would be stationary, but the probability of extinction of either line would be zero.

Couples in the United States and Europe seem (*a*) to devote much effort to overcoming sterility, (*b*) to aim at having two children with a uniformity unknown in the fertility schedules of the past, and (*c*) to want to have at least one boy and one girl child. Whether or not they have such an objective in mind, couples act as though each is doing its utmost to maintain both its male and female lines and wants the smallest possible number of children consistent with a high chance of infinite lines of descent.

16.5 Sex Preference and the Birth Rate

Insofar as parents wish to have at least one boy, or at least one boy and one girl, and keep having children until they attain their wish, the birth rate is higher than it would otherwise be. A considerable literature (Goodman 1961a, Repetto 1972, Hatzold 1974, Sheps 1963, McDonald 1973) analyzes the magnitude of this effect, using observed births to infer preferences, and assesses by how much the ability to control the sex of children would lower the birth rate.

If the probability of a boy on a particular birth is p, and of a girl is $q = 1 - p$, and we think of couples whose sole aim in family building is to have at least one boy, the proportion of such couples who will stop after exactly one child is p, who will stop after exactly two children is qp, who will stop after exactly three is q^2p, and so on. For them the mean number of children is

$$p + 2qp + 3q^2p + \cdots,$$

which as we saw in Section 16.1 is

$$1 - q + 2q(1 - q) + 3q^2(1 - q) + \cdots,$$

or, on canceling

$$1 + q + q^2 + q^3 + \cdots = \frac{1}{1 - q} = \frac{1}{p}.$$

If for example, the chance of a boy is 0.5, couples who continue to have children until a boy arrives will average two children. (Note the formal analogy to the number of months to pregnancy in Section 16.1; births have taken the place of months and a boy takes the place of conception.)

For these couples the fraction of boys is 1 divided by the mean number of children; that is, $1/(1/p) = p$. Just as a gambler cannot influence his winnings by choosing the time of leaving a fair game, so parents cannot influence the proportion of boys by a stopping rule. This argument applies to any target number of children, so long as the couples considered are homogeneous, in that all have the same chance p of a boy on any birth.

If the couples are not homogeneous, a selective effect can occur in the population. Suppose that p is distributed according to $f(p)$, where $\int_0^1 f(p)\,dp = 1$, so there are $Nf(p)\,dp$ couples between p and $p + dp$. [We need to suppose that $f(0) = 0$; that is, no couples incapable of having boys are included.] Then for the $Nf(p)\,dp$ couples with a given p the average number of children is $1/p$, as appeared above, so the total number of children contributed by those between p and $p + dp$ is $(N/p)f(p)\,dp$; adding through the distribution of p gives the grand total of children, $N\int_0^1 (1/p)f(p)\,dp$. Since the total number of boys is N, one to each couple in this problem, the fraction of boys in the population is

$$\frac{N}{N\displaystyle\int_0^1 (1/p)f(p)\,dp} = H,$$

say, which is the harmonic mean of the p's (Goodman, 1961a).

Since the harmonic mean is less than the arithmetic mean, we have proved two things about the effect of heterogeneity on the outcome of this stopping rule: the fraction of boys is H, which is smaller than the fraction of boys if no stopping rule were used, and the average number of children per couple is $\int_0^1 (1/p)f(p)\,dp = 1/H$, and hence is greater than the $1/p$ that would occur with homogeneity. Though an individual couple of fixed p cannot affect the proportion of boys among its offspring by any stopping rule, a population can: each couple continuing until it attains a boy reduces the fraction of boys in the population, as though parents were testing themselves to see whether they were boy-producers or girl-producers, and only in the latter case having further children.

As an indication of the numerical effect of this rule, suppose that one-third of parents have $p = 0.25$, one-third $p = 0.50$, and one-third $p = 0.75$. Then the harmonic mean is

$$\frac{1}{(1/0.25)\frac{1}{3} + (1/0.50)\frac{1}{3} + (1/0.75)\frac{1}{3}} = \frac{9}{22} = 0.4091,$$

as against the arithmetic mean of 0.50.

An Approximation to the Harmonic Mean. For small variations in the p values pertaining to the several couples in the population we can find a general expression for the amount by which parents following the rule of stopping with a boy will decrease the proportion of boys in the population. We first need an expansion of $1/p$ (where $0 < p \leqslant 1$) around the reciprocal of the mean value $\bar{p}$. By Taylor's theorem

$$\phi(p) = \phi(\bar{p}) + (p - \bar{p})\phi'(\bar{p}) + \frac{(p - \bar{p})^2}{2!}\phi''(\bar{p}),$$

where we suppose $p - \bar{p}$ to be small enough that $(p - \bar{p})^3$ and higher powers are negligible. Then, if the function $\phi(p)$ is $1/p, \phi'(\bar{p})$ is $-1/\bar{p}^2$ and $\phi''(p)$ is $2/\bar{p}^3$. Hence

$$\frac{1}{p} = \frac{1}{\bar{p}} + (p - \bar{p})\left(-\frac{1}{\bar{p}^2}\right) + \frac{(p - \bar{p})^2}{2!}\frac{2}{\bar{p}^3} + \cdots,$$

and, entering the right-hand side in place of $1/p$ in the expression for the harmonic mean, we have approximately

$$\frac{1}{\displaystyle\int_0^1 \frac{f(p)}{p}\,dp} \approx \bar{p} - \frac{\sigma^2}{\bar{p}}. \tag{16.5.1}$$

This follows from the definition of the mean, $\bar{p} = \int_0^1 pf(p)\,dp$, and of the variance, $\sigma^2 = \int_0^1 (p - \bar{p})^2 f(p)\,dp$, the approximation in (16.5.1) requiring the variation of p to be small.

To test the approximation of the numerical example above, in which parents are equally divided among those having 0.25, 0.50, and 0.75 probability of producing a boy on each birth, we find that σ^2 is 0.04167, and hence that the estimate of the harmonic mean $\bar{p} - \sigma^2/\bar{p}$ is $\frac{1}{2} - (0.04167/\frac{1}{2}) = 0.4167$, only 2 percent higher than the 0.4091 above.

What in fact are the sex preferences of parents? We ought to be able to find out by observing, for example, in what proportion of cases parents with given constitutions of family go on to further children. The proportions of first children, of second children, and so on that are boys are out of the parents' control and so can tell us nothing about parental preferences; on the other hand, the proportions of the last child and the second-to-last child are determinable by parents, even without any ability to determine a given birth. If couples with one girl go on to a further child in a higher fraction of cases than those with one boy, this suggests a wish for boys; if those with one boy and one girl go on to a further child less often than those with two boys or two girls, this shows a wish for at least one child of each sex. How subtly do parents play the game? If they conclude, after having a girl, that they tend to be girl producers, at least some of those with boy preference will stop at that point.

Some data that classify children according to birth order and sex are available, but the complexities of their analysis will not be undertaken here. Instead we go on to anticipate the time when parents will have a measure of control over the sex of their offspring. (Some assert that techniques for this are available now, but their effectiveness is controversial.)

16.6 Family-Building Strategy with Parental Control over Sex of Children

If parents want at least one son and one daughter and the chance of either is one-half on a given birth, they have no choice but to proceed at random, and they will average three children. Suppose that the chance of having a boy when they are trying for a boy is b and of having a girl when trying for a girl is g, both b and g being appreciably larger than one-half. A couple want at least B boys and at least G girls; what should they do to attain these while exceeding $B+G$ total children by as small a number as possible? A direct attack on this problem would be complicated; we will see that it can be made simple by proceeding one step at a time (McDonald 1973).

The problem can be represented on paper as a lattice of points (Figure 16.4), of which the one on the upper right is labeled zero. The several points represent the number of children needed to attain the target. Thus at the points on the top, reading from right to left, the couple wants zero, one, two, ... boys to complete its (self-determined) total. At the points along the right, reading from top to bottom, the couple wants zero, one, two, ... girls to complete the number sought. At the interior point A the couple want one boy and one girl, and so on. The problem is then to calculate at each such point the expected number of further children in total to reach the desired constitution of the family. One visualizes families climbing toward the target, one child at a time, and we can take advantage of the additivity over the several steps of the expected number of children acquired.

Consider the couple that are lacking exactly one girl (point C). Suppose that they try for a girl and stop as soon as they have one. They stand a probability g of having a girl first, in which case they add just one child, a chance $(1-g)g$ of having first a boy and then a girl (i.e., two children), a chance $(1-g)^2g$ of having three children, and so on; hence their expected number of children is

$$g + 2(1-g)g + 3(1-g)^2g + \cdots,$$

or, using the by now familiar device of putting $f = 1 - g$ and then consolidating terms,

$$\begin{aligned} 1 - f + 2f(1-f) + 3f^2(1-f) + \cdots &= 1 + f + f^2 + \cdots \\ &= \frac{1}{1-f} = \frac{1}{1-(1-g)} = \frac{1}{g}. \end{aligned}$$

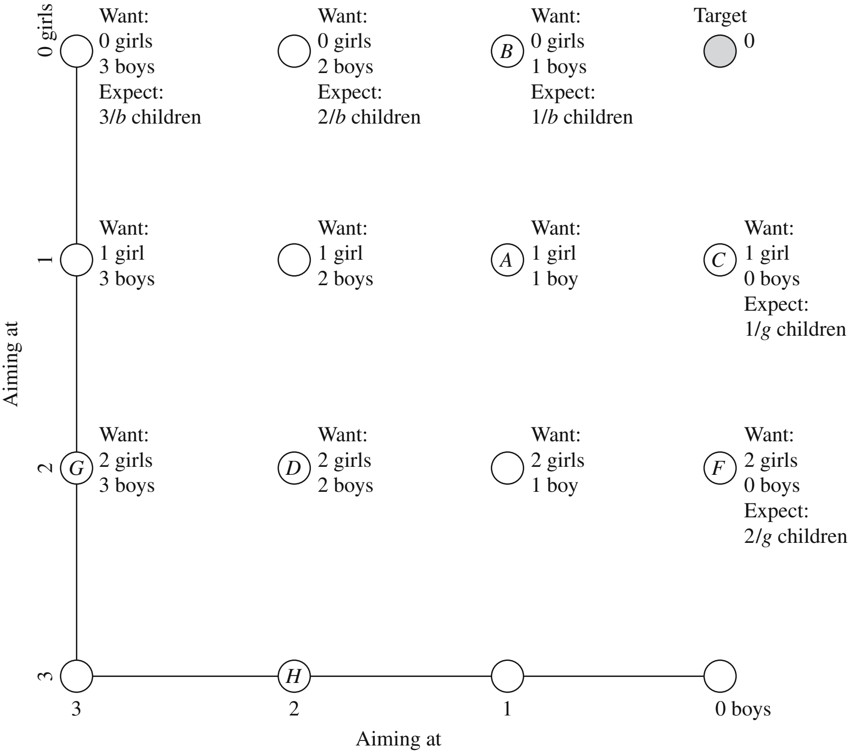


Figure 16.4. Distance to target for couples aiming at specified numbers of boys and girls.

The expected number of children is $1/g$, and we have labeled the point just below the target accordingly (Figure 16.4). With $g = 0.7$ the expected number of children is $1/0.7 = 1.43$. With $g = 0.5$, that is, no sex control, the expected value is $1/0.5 = 2$.

Suppose that the couple wants two girls. Now they would be starting at the second point below zero. If they try for a girl, they stand a chance g of getting one, which will move them up one point and give them an expected number of children $1 + (1/g)$, so we can write $g[1 + (1/g)]$ in respect to this possibility. If they fail to have a girl, they have one more child and still need two girls, so the expectation for this is $(1 - g)(1 + e)$, where e is the expected number of children when one is trying for two girls. Then, adding both possibilities, we have

$$e = g\left(1 + \frac{1}{g}\right) + (1 - g)(1 + e),$$

or, on solving the simple equation for e,

$$e = \frac{2}{g}.$$

We could have anticipated this by noting that to move up one girl takes on the average $1/g$ children; hence to move up two girls takes $2/g$ children. To move up three girls takes an average of $3/g$ children, and similarly for higher numbers. The cells for boys only, across the top of the diagram, are the same but with b replacing g.

Hence the expectations shown on the right-hand points and across the top of the diagram; having attached probabilities to these boundary points, we now turn to the interior. Consider the interior point marked A in Figure 16.5, standing for a couple wishing one girl and one boy, either altogether or in addition to the children they already have.

Suppose they try for a girl first. Then the chance of moving up in the diagram is g, and if they succeed they reach the point just above, from which the expectation is $1/b$. If they fail they reach the point at the right, from which the expectation is $1/g$. Hence the total expected number of children from point A is

$$A_g = g\left(1 + \frac{1}{b}\right) + (1-g)\left(1 + \frac{1}{g}\right),$$

where 1 has been added within the parentheses to allow for the girl or boy born in the move. On cancellation this reduces to

$$\frac{g}{b} + \frac{1}{g},$$

and with $b = 0.8$ and $g = 0.7$ equals $0.7/0.8 + 1/0.7 = 2.30$.

Using the opposite strategy of trying for a boy first (Figure 16.5) would give

$$A_b = b\left(1 + \frac{1}{g}\right) + (1-b)\left(1 + \frac{1}{b}\right) = \frac{b}{g} + \frac{1}{b},$$

or, with our numbers, $0.8/0.7 + 1/0.8 = 2.39$. Evidently the right strategy is to try for a girl first. This suggests a general rule: parents should leave the more controllable contingency to the last if they seek equal numbers of boys and girls.

We want always to choose the strategy that gives the minimum number of total children, for example, that corresponds to

$$\text{Min}\left(\frac{g}{b} + \frac{1}{g}, \frac{b}{g} + \frac{1}{b}\right).$$

If $b > g$, that is, if we are surer to get a boy if we aim at a boy than a girl if we aim at a girl, it is easily seen by algebra that $b/g + 1/b > g/b + 1/g$. This is so because, if $b > g > \frac{1}{2}$, then $b - \frac{1}{2} > g - \frac{1}{2}$ and $(b - \frac{1}{2})^2 > (g - \frac{1}{2})^2$, or $b^2 - b > g^2 - g$; therefore $b^2 + g > g^2 + b$, and, on dividing both sides by $bg, b/g + 1/b > g/b + 1/g$. If $b > g$ we should try for a girl first to minimize total children.

The ability of parents to influence the sex of their children will reduce the average number of offspring for parents targeting on a minimum number of

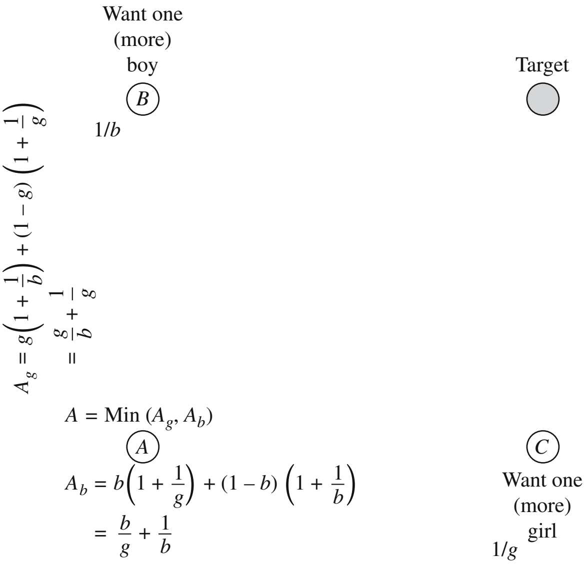


Figure 16.5. Enlargement of part of Figure 16.4, showing expected children for couple wanting one boy and one girl (position A). Trying for a boy first gives an expected $A_b = b(1+1/g)+(1-b)(1+1/b) = b/g+1/b$ children. Trying for a girl first gives $A_g(1+1/b)+(1-g)(1+1/g) = g/b+1/g$ children, where b is the chance of having a boy if the couple try for a boy, and g is the chance of having a girl if they try for a girl.

boys and of girls. Figure 16.6 shows expected numbers from various points with $b = 0.8,\ g = 0.7$, and in parentheses with $b = 0.5,\ g = 0.5$.

Let us see how sensitive the result is to the degree of reliability with which sex is determined. Table 16.4 shows the cases of parents aiming at one boy, and at one boy and one girl. Making $b = g = 0.5, 0.6, \ldots,$ we see the average number of children they will have. If $b = g = 0.8$ rather than $b = g = 0.5$, the expected total drops from 2.0 to 1.25 for couples aiming at one boy, and from 3.0 to 2.25 for couples aiming at one boy and one girl. In both cases the reliability of 0.8 in sex determination eliminates three-quarters of the unwanted children.

A number of surveys have asked newly married couples how many boy and how many girl children they want. One such survey in Hull, England, showed 45.7 percent wanting one boy and one girl, 15.4 two boys and one girl, and 12.6 two boys and two girls; the average was 2.55 children. Under present circumstances, that is to say with $b = g = 0.5$ approximately, we find by applying the numbers in parentheses in Figure 16.6 to these percentages that the average attained would be 3.74 children, or 1.19 more than wanted (Table 16.5).

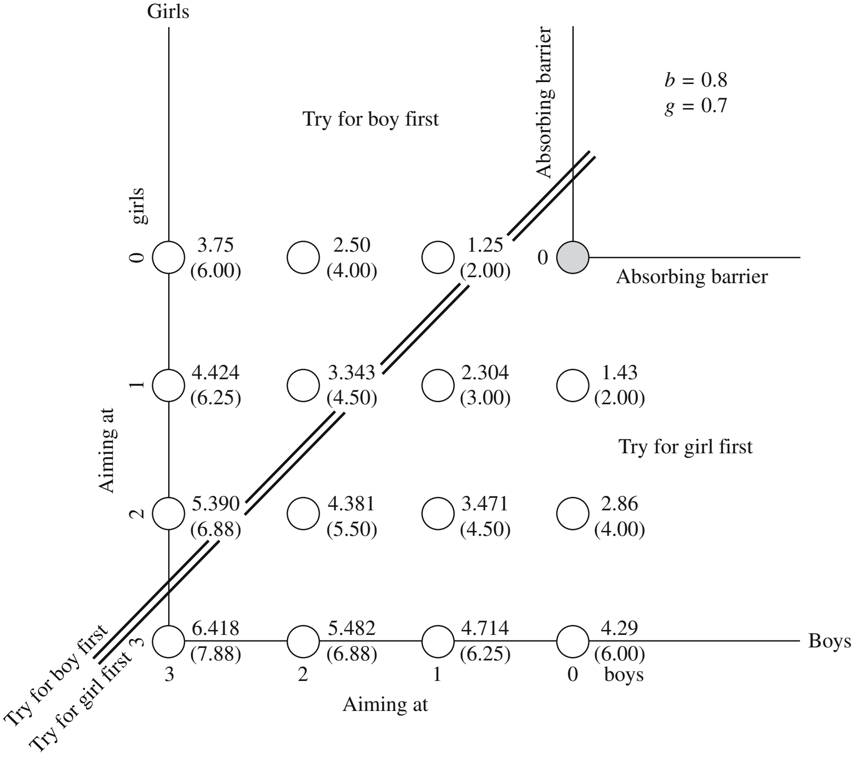


Figure 16.6. Expected number of children for various family objectives from zero boys and zero girls up to three boys and three girls, when $b = 0.8$, $g = 0.7$, and (in parentheses) when $b = g = 0.5$.

Table 16.4. Expected number of children born to couples who stop after one boy, and who stop after one boy and one girl

	Expected total stopping after	
	One boy	One boy and one girl
$b = g$	$\frac{1}{b}$	$\frac{g}{b} + \frac{1}{g} = \frac{b}{g} + \frac{1}{b} = 1 + \frac{1}{b}$
0.5	2.0	3.0
0.6	1.67	2.67
0.7	1.43	2.43
0.8	1.25	2.25
0.9	1.11	2.11
1.0	1.00	2.00

With $b = 0.8, g = 0.7$, numbers chosen because they are about halfway to complete sex control represented by $b = g = 1.0$, the average would be 2.88 children, or only 0.33 more than wanted.

A technology that goes halfway to sex control would eliminate nearly three-quarters of the unwanted children born because of the sex preferences of parents. Full sex control would reduce the birth rate for these parents by 32 percent; half sex control would reduce their birth rate by 23 percent.

Five kinds of qualification are required. The least important is that children are born one at a time in the model, which has no easy way of taking account of multiple births. Moreover, not all parents have preferences regarding the sex composition of their families, and the model deals only with parents whose preferences are strong enough to be overriding. In fact, they must be willing and able to have an infinite number of children if necessary. The model also supposes that the parental decision regarding composition is made at the outset rather than in stages, or child by child. It supposes homogeneity—that all parents have the same chances b and g.

The advent of partial or complete control of the sex of children would have the effect of changing the sex of the new generation, in some cultures in favor of boys; in those cultures the advent of a high sex ratio in the new generation would reduce the number of marriages; when the children came to childbearing age marriages would be limited by the less numerous sex. Unwed males would decrease the overall birth rate. But well before that came about the culture would shift in favor of the initially less favored sex. Within a decade or so of the birth of a disproportionate number of boys couples would come to value girls more highly. Perhaps a series of waves would ensue, not unlike those familiar in a market economy arising from the interval of time between capital investments and the returns therefrom (Westoff and Rindfuss 1974). For an insightful analysis of some of the demographic consequences of sex selection in recent China, see Tuljapurkar et al. (1995).

Table 16.5. Application of average numbers of expected children to sex preferences of parents in survey in Hull, England

		Number of children		
Parents' preferences	Percent	Wanted	Average born with $b = g = \frac{1}{2}$	Average born with $b = 0.8$, $g = 0.7$
One boy, one girl	45.7	2	3.0	2.304
Two boys, one girl	15.4	3	4.5	3.343
Two boys, two girls	12.6	4	5.5	4.381
Average of survey		2.55	3.74	2.88
Excess with $b = g = \frac{1}{2}$			$3.74 - 2.55 = 1.19$	
Excess with $b = 0.8, g = 0.7$			$2.88 - 2.55 = 0.33$	

16.7 Mean Family Size from Order-of-Birth Distribution

Birth certificates almost invariably require that mothers report whether a particular birth is their first, second, third, and so on. The resultant order-of-birth tabulations are available for many countries and regions, and one would like to extract their implication for the mean number of children born to mothers. The problem was considered by Burks (1933) and others.

The early writers assumed a stationary condition, in which the distribution of births by order in a calendar year or other period of observation is the same as in the cohorts of mothers that are passing through the period in question; the several cohorts are taken as of equal size. Suppose that the observed number of first births is N_1, of second births is N_2, and so on, and that these numbers apply in all periods and for all cohorts. This is to say that N_1 is the total of women who have any children at all, N_2 the number who have two or more children, N_3 the number who have three or more children, for all periods and all cohorts. If we call ϕ_1 the number who stop with one child, ϕ_2 the number who stop with two children, and so on, the N's are the cumulative sums of the ϕ's:

$$\begin{aligned} N_1 &= \phi_1 + \phi_2 + \phi_3 + \cdots \\ N_2 &= \phantom{\phi_1 + {}} \phi_2 + \phi_3 + \cdots \\ N_3 &= \phantom{\phi_1 + \phi_2 + {}} \phi_3 + \cdots \\ &\vdots \end{aligned}$$

and the ϕ's are the differences of the N's:

$$\begin{aligned} \phi_1 &= N_1 - N_2 \\ \phi_2 &= N_2 - N_3 \\ \phi_3 &= N_3 - N_4 \\ &\vdots \end{aligned}$$

The mean number of children in a family that has children is

$$\begin{aligned} \frac{\phi_1 + 2\phi_2 + 3\phi_3 + \cdots}{\phi_1 + \phi_2 + \phi_3 + \cdots} &= \frac{N_1 + N_2 + N_3 + \cdots}{N_1} \\ &= \frac{N}{N_1} = \frac{\text{Total births}}{\text{First births}}. \end{aligned} \tag{16.7.1}$$

This highly simplified formula applied to United States data for 1933 requires only the fact that births in that year were 74 per thousand native white women 15 to 44 years old, and first births were 24, making a mean family size of 3.1 children. For 1956 the corresponding figures were 115, 33, and a mean family size of 3.5 children; for 1973 the mean was 2.40.

16.8 Parity Progression and Population Increase

How parents decide the size of their families has been the subject of much investigation since effective birth control has made that size subject to deliberate decision. Some may have a concept at the beginning that they retain throughout—they are fixed on three children, say. The majority, however, seem to make up their minds as they go along—they have one child and then, depending on how things look, they decide whether to have another. The following argument suggested by Norman Ryder formalizes parental decision in terms of the probabilities of successively proceeding to each further child.

Disregard time and consider the married couples that have children of successive orders. Suppose that a certain fraction of women have a first child; of these a certain fraction have a second child; of those that have two children a certain fraction go on to a third; and so on. We might take as the cohort a group of couples that married at the same time, or a group of women born at the same time. We could include a symbol for the fraction r_0 of the cohort that survives, and another for the fraction that marries, but simplicity is served by just supposing that r_1 of the girls born at a given moment will grow up and have a first child, of these the fraction r_2 will go on to a second child, and so on.

Then the probability that a girl child will have at least one child is r_1, that she will have two children is $r_1 r_2$, and so on. Suppose that the ratio of boys to girls among the births of any order is s, so the fraction $1/(1+s)$ of the births are girls. The net reproduction rate of the cohort of which the girl in question is a member is

$$R_0 = \frac{1}{1+s}(r_1 + r_1 r_2 + r_1 r_2 r_3 + \cdots). \tag{16.8.1}$$

Due allowance is implicitly made for mortality, nonmarriage, illegitimacy, and voluntary or involuntary sterility within marriage.

We can enable the notion of parity progression to serve us better by reducing the parameters in (16.8.1) from one for each child to just two in all. Ryder found that $r_1 \approx r_2 = h$, say, and that $r_3 \approx r_4 \approx r_5 \approx \cdots = k$, say. Then (16.8.1) becomes

$$\begin{aligned} R_0 &= \frac{1}{1+s}(h + h^2 + h^2 k + h^2 k^2 + h^2 k^3 + \cdots) \\ &= \frac{1}{1+s}\left(h + \frac{h^2}{1-k}\right). \end{aligned}$$

Table 16.6 shows values of R_0 for various combinations of h and k; the combinations on the lower right give an increasing population.

With this result we can do experiments of various kinds. We can, in particular, see how R_0 would be modified by lowering k, which is the way in which fertility change seems largely to have taken place in the United

Table 16.6. Net reproduction rate corresponding to mothers' parity progression ratio, h for first and second child, and k for subsequent children, according to (16.8.2)

h	k						
	0.40	0.45	0.50	0.55	0.60	0.65	0.70
0.60	0.59	0.61	0.64	0.68	0.73	0.79	0.88
0.65	0.66	0.69	0.73	0.78	0.83	0.91	1.00
0.70	0.74	0.78	0.82	0.87	0.94	1.02	1.14
0.75	0.82	0.86	0.91	0.98	1.05	1.15	1.28
0.80	0.91	0.96	1.01	1.08	1.17	1.28	1.43
0.85	1.00	1.06	1.12	1.20	1.30	1.42	1.59

States. Now, since

$$R_0 = \frac{1}{1+s}\left(h + \frac{h^2}{1-k}\right), \tag{16.8.2}$$

we have

$$\frac{dR_0}{dk} = \frac{1}{1+s}\left(\frac{h}{1-k}\right)^2 .$$

Around $s = 1.05, h = 0.80, k = 0.65$, we have $dR_0/dk = 2.55$, so that each increase of 0.01 in k produces an increase of 0.0255 in R_0, and similarly for decreases.

To see the effects of successive childbearing decisions on population growth we need also to take account of timing. The age at which a woman has her first child, and the successive interbirth intervals, will evidently make a difference in the rate at which the population grows. Timing is the one element lacking in the present model. For the effect of its omission, consider $h = 0.80$ and $k = 0.65$, so that $R_0 = 2.63/2.05 = 1.28$. If the mean age of childbearing (strictly, the length of generation) is 25 years, the intrinsic rate is 0.0099; if it is 30 years, the intrinsic rate is 0.0082.

16.9 For a Given Probability of Survivors, Lower Mortality Lowers the Rate of Increase

When mortality is high, a man who wants to have a son who will see him through his old age requires many children. This point has often been made before, but we still need clarification of the relation between mortality and the rate of population increase among people who want a certain assurance of *surviving* children. With number of births given, the rate of increase r goes up as mortality μ_x goes down; we will see that the relation between r and μ_x is reversed if the birth rate is determined by the wish to have surviving sons.

Table 16.7. Rate of increase r of populations with values of l_{30} from 0.5 to 0.9 and probability p of at least one living son of 0.8 and 0.9; tabulation of (16.9.4)

l_{30}	$p = 0.8$	$p = 0.9$
0.5	0.0050	0.0169
0.6	0.0017	0.0137
0.7	−0.0022	0.0097
0.8	−0.0074	0.0045
0.9	−0.0155	−0.0035

Call the probability of at least one surviving son p; this must equal one minus the probability that all sons will die. Consider survivorship to father's age 60; then the probability that not all sons will die is

$$1 - (1 - l_{60-a_1})(1 - l_{60-a_2}) \cdots (1 - l_{60-a_n}) = p \tag{16.9.1}$$

if the first son is born at age a_1 of the father, the second at age $a_2, \ldots,$ and the nth at age a_n.

Now the rate of increase of the population for this particular family, still on the one-sex model but for males, is the real root in r of the equation

$$\sum_{i=1}^{n} e^{-ra_i} l_{a_i} = 1. \tag{16.9.2}$$

This discrete form of Lotka's characteristic equation can be solved for r to tell us what r would be in a population if the mortality and childbearing patterns of a particular family were general.

Our problem is to find r as a function of the life table l_a from (16.9.2), given that the number of children n will be determined by (16.9.1). Even to define the solution of this would be awkward, so we approximate by supposing all the children to be born at the same age of the father: $a_1 = a_2 = \cdots = 30$, say. Then the first equation, (16.9.1), is

$$1 - (1 - l_{30})^n = p,$$

or

$$n = \frac{\log(1-p)}{\log(1-l_{30})}. \tag{16.9.3}$$

The second equation, (16.9.2), is

$$ne^{-30r} l_{30} = 1;$$

and substituting n from (16.9.3) and taking logarithms gives

$$r = \frac{1}{30} \log \left[l_{30} \frac{\log(1-p)}{\log(1-l_{30})} \right]. \tag{16.9.4}$$

The values of r are shown in Table 16.7, whose most important message is that each rise in l_{30} of 0.1 causes a decline in r of 0.003 to 0.008. The point seems to be that the improvement in mortality permits a decline in the number of children sufficiently great that the births drop more than enough to offset the fall in mortality. Heer (1966) has applied simulation to this problem, and his more complex model produces an inverted U-curve in r.

Note that n, being the number of children, ought to be an integer, and if p is to be assured it ought to be the next higher integer to the expression on the right-hand side of (16.9.3). For some combinations of high p and low l_{30} the n implied in (16.9.3) will be an impossibly large number of children. This does not apply, however, to the range shown in Table 16.7, for which 1 to 4 male children suffice.

Note that as l_{30} moves toward unity a discontinuity exists, for r goes steadily downward with $1 - l_{30}$ for any fixed value of p; but we know that $l_{30} = 1$ and $r = 0$ will provide with certainty the one child living at age 60 of the father. Evidently (16.9.4) is not to be taken seriously as mortality falls very low, or when other than small positive values of $n \geqslant 1$ are implied.

17

The Multi-State Model

The life table was in use for some centuries before anyone noticed that the transition from life to death is an instance of much more general transitions—from well to sick, from school to work, from single to married, from working to unemployed, from living in Province A to living in Province B, from fecund to pregnant. For each of these, and many other transitions, a decrement table can be made, and the usefulness of doing so has been recognized in field after field. To study schooling, one routinely makes life tables of survival in school as against dropping out; to study the effectiveness of a contraceptive, one makes a life table of those remaining fecund versus those falling into pregnancy.

The technique could well be applied more extensively. For unemployment, we ought to know not only the total number at a given moment and the fraction they are of the labor force, but also the probability that an unemployed individual will find a job next month, the month after, or the month after that. With ten percent unemployment in the country, it makes a great deal of difference whether everyone in the labor force is out of a job for 5 weeks a year, or one tenth of the labor force is out for 50 weeks. The life table form would tell a person of certain characteristics who has just lost his job the chance of being out of work for 1 month, 2 months, etc.

All these would be single decrement life tables, calculable by exactly the same technique as has been used for the representation of mortality these last 300 years. There is no need to adapt the life table method to them; the method is immediately applicable. This chapter goes beyond them into the realm of multiple contingencies. The person now out of work has a certain chance in the next month of finding a job, of moving to another province, of going back to school, etc.

The matter can become complex: What is the chance that a person who is now unemployed will be back at work within the year? This would include the probability that he finds a job next month, then loses it four months later, finds another in the seventh month, and keeps this to the end of the year, plus millions of similar combinations. To show the formulae that would take account of all the combinations, but without the user having to specify any of them, is the purpose of this chapter. The mathematics is essentially due to Kolmogorov, put into convenient form by Andrei Rogers (1975, 1995). Schoen (1975) and Schoen and Land (1979) independently recognized the essential principle and expressed it without the use of matrices, necessarily in longer formulae. All the methods used for the ordinary life table turn out to apply, with no modification other than the replacement of scalars by vectors and matrices, and specification of the order in which multiplication is carried out. Formulae include expressions for converting age-specific rates into probabilities of survival to the next age, successive multiplication of these from the beginning of life to obtain probabilities that a person born into a certain state will be in (another or the same) state at successive ages; integration of the survivors to obtain the stationary population; cumulative adding of the stationary population from the end of the table backwards to obtain at each age the expectation of future time in each state.

Most of demography, like the present exposition, starts with transitions of people from one state at a certain moment to another state 1 year or 5 years later: single to married, married with 1 child to married with 2 children, at school to in the labor force, living to dead. These transitions have to be calculated from raw statistical data of various forms. Sometimes individual movements are registered: Robert Jones died at age 51 on April 23, 1984; Mary Henderson, age 28, gave birth to a baby boy on July 17, 1984; Henry Johnson retired at the end of September 1984. The individual movements are aggregated into groups and published as official statistics: There were 1372 deaths of males aged 50–54 in 1984; 987 girl babies were born to women aged 25–29 in 1984. Some of the data are not events but a count of the individuals in a region: 116,572 males aged 50–54 were living in a certain area on July 1, 1984. These are stocks, in contrast to the flow data describing events.

The life table is a transition model in which observed death rates, within an age interval, are the basis of probabilities of dying and then of the stationary population, the expectation of life, and other parameters of interest. Migration analysis, on the other hand, often starts from a census question asking respondents where they were living 5 years earlier. A kind of transition probability is directly given by the aggregation of the resulting answers. With some qualifications one can thus, in a sense, observe transitions and infer moves from them; the opposite applies in mortality statistics, where it is moves that are observed and transitions inferred from them. That the Jones family lived in Denver in 1980 and in Omaha in 1985 is a *transition*; the family *moved* from Denver to Chicago in June 1981 and from Chicago to Omaha in November 1984.

That there were 1,160,000 divorces in the United States in 1984 tells a very small part of the story of marriage dissolution; anyone studying divorce wants to know the probability of divorce, say within 5 years of marriage, for couples in various categories. Only in that way can proper comparisons be made over time and between social groups. The original counts of numbers of divorces year by year do not even tell whether the propensity to divorce is increasing, let alone by how much.

A couple is provided with some means of birth control—perhaps the wife is fitted with an IUD. What is the chance that the IUD will still be in place 1 month, 2 months, 3 months later? And how does this compare with the steadfastness of another couple in using a stock of pills with which they are provided? What is the chance that a recruit to a particular job will still be holding on and doing the work 1 year, 2 years, 20 years later?

All the hazards implied above—divorce, failing to retain an IUD, losing a job—can be represented as hurdles at heights suitably determining the risks of failure. When we know the runners who fail on the ith hurdle as a fraction of the number who arrive at it, for all values of i, then the cumulative product of the probabilities of not failing tells us what fraction of the number that started the race will still be in the running after the xth hurdle. This cumulative fraction remaining in the race after the xth hurdle among those that started is the l_x column of the ordinary single-decrement life table. The average number of hurdles cleared before the runner misses is the expectation or $\overset{o}{e}_0$ column of the same life table.

The works of Fix and Neyman (1951), Mertens (1965), Sverdrup (1965), Chiang (1960a,b, 1961, 1968), Oechsli (1971), Schoen and Nelson (1974), Schoen (1975, 1988), Hoem (1975), Hoem and Fong (1976), Schoen and Land (1979), and especially of Rogers (1975, 1984, 1995), Willekens (1978), Rogers and Ledent (1976), and Land and Rogers (1982) provide further expositions and examples of multivariate demography. Al Mamun (2003) reviews more recent applications, and gives an introduction to the theory in the context of cardiovascular disease and its risk factors. What follows is an elementary outline of the theory, stressing the analogy to single decrement.

17.1 Single Decrement and Increment-Decrement

Early in the history of demography two questions were asked: What is the probability of surving to age x, and what is the average age at death—the expectation of life? Smith and Keyfitz (1977) provide excerpts from original papers. The answer, in current notation, is that if the chance of dying between age a and $a + da$ for those aged a is $\mu(a)\,da, \mu(a)$ being the force or intensity of mortality, then the probability $l(x)$ of survival to age x is obtainable by solving the differential equation

$$dl(x)/dx = -\mu(x)l(x), \tag{17.1.1}$$

which gives $l(x) = \exp(-\int_0^x \mu(a)\,da)$, and the expectation of life is then calculated as

$$\overset{o}{e}(x) = \int_x^\omega l(a)\,da/l(x), \tag{17.1.2}$$

where ω is the oldest age of life. Chapter 2 describes one of the numerous ways of calculating $l(x)$ and hence $\overset{o}{e}(x)$ from empirical data.

When births are taken into account as well as deaths, the question arises: By how many girls will a girl baby be replaced? The answer is given in terms of the chance $l(x)$ that she will live to age x and then have a girl baby between age x and $x+dx, m(x)l(x)\,dx$, where $m(x)$ is the age-specific fertility rate. Integrating this over the range of reproductive life gives the required $R_0 = \int_\alpha^\beta m(x)l(x)\,dx$, which is the ratio of the size of one generation to the preceding at the specified rates of birth and death. It gives the implication for population growth of the prevailing schedule of mortality and fertility, in disregard of peculiarities of the existing age distribution.

These are just about the most complicated demographic problems that can be presented and solved in one dimension. Everything else requires two or more dimensions. To study mortality by itself, or even mortality and fertility, recognizing age only, is to take a very small part of the demographic process out of its natural context. What follows will generalize the preceding formulae to an arbitrary number of dimensions. The result is a theory that links continuous time or age processes to the discrete-time population projection matrix approach of Chapters 3 and 7.

17.1.1 Matrix of Inputs

To generalize the original life table theory, we deal not only with the movement from life to death represented by the scalar rate $\mu(x)$ but with the matrix $\boldsymbol{\mu}(x)$, standing for the instantaneous rates of movement between states.

Construction of the $\boldsymbol{\mu}(x)$ matrix from actual data, at least to a suitable approximation, is straightforward. The off-diagonal elements of $\boldsymbol{\mu}(x)$ are each the corresponding observed rate of movement in a small time interval with sign reversed. Thus $-\mu_{ij}(x)\,dx$ is minus the chance that a person in state j transfers to state i during the short period of time and age dx. Each diagonal element of $\boldsymbol{\mu}(x)$ contains the rate $\mu_{\delta i}$ of dying, with positive sign, along with the total of the off-diagonal elements of the column, $\sum_{i \neq j} \mu_{ij}$, also with positive sign. The reason for this is that the column total has to be conservative—that is, to add zero with respect to movements among units.

Table 17.1. Matrix $\boldsymbol{\mu}(x)$ of moves for the case of three states.

$$\boldsymbol{\mu}(x) = \begin{pmatrix} \mu_{\delta 1}(x) + \sum_{i\neq 1}^{n} \mu_{i1}(x) & -\mu_{12}(x) & -\mu_{13}(x) \\ -\mu_{21}(x) & \mu_{\delta 2}(x) + \sum_{i\neq 2}^{n} \mu_{i2}(x) & -\mu_{23}(x) \\ -\mu_{31}(x) & -\mu_{32}(x) & \mu_{\delta 3}(x) + \sum_{i\neq 3}^{n} \mu_{i3}(x) \end{pmatrix}$$

The quantities from the jth state added into the ith state must also be subtracted from the jth, so an increment to $\mu_{ij}, i \neq j$, has to be subtracted from μ_{jj}. The net total of the column is just the death rate.

The notation is indicated in Table 17.1 for the case of three states, giving the matrix $\boldsymbol{\mu}(x)$ in some detail. The right-hand subscript is state of origin, the left-hand subscript state of destination. Thus $\mu_{23}(x)$ is the movement from state 3 to state 2 for persons aged x during a given interval. All other matrices use the same subscripting, essentially that of Rogers (1975). The matrix $\mathbf{M}(x)$ will be the finite approximation to $\boldsymbol{\mu}(x)$. It is obtained as raw data by dividing the transitions from the jth state to the ith state by the exposure, i.e., the mid-period population times the length of the interval.

Throughout we shall make the one major assumption that is common to all life tables and to increment-decrement tables and without which demographic processes cannot be conveniently portrayed. The probability of an individual making any transition will be taken to depend only on the state in which he is located at the start of the transition period—the Markov condition. The chance of a man of 55 dying before the age of 60 in the ordinary life table depends only on the fact that he is 55 and belongs to a certain defined population; it does not depend on his health as a baby or whether he smokes or is a nonsmoker or whether his father died young or old. If we want to take into account anything beside his age at the beginning of the interval, we have to do so by dividing up the population—say into smokers and nonsmokers—and then allowing the same Markov condition to apply within each of these groups.

Researchers in some fields find this assumption more restrictive than those in other fields. There is not much complaint about it in the ordinary life table. On the other hand, for mobility studies the history of the individual does seem to be important—for example, the longer a person has been in a given region, the less likely he is to move away in the next time interval. Such considerations introduce the history of the person in a way that precludes the treatment of these pages. We shall always assume that the entire history of the person is summed up in the state in which he is found at the beginning of each interval.

17.2 The Kolmogorov Equation

Identical with equation (17.1.1), except that the elements are now matrices and vectors, is the basic

$$d\mathbf{l}(x)/dx = -\boldsymbol{\mu}(x)\mathbf{l}(x), \tag{17.2.1}$$

which is due originally in this application to Kolmogorov (Krishnamoorthy 1978, Willekens 1978). Here $\mathbf{l}(x)$ is a column vector in which the ith element is the number of the population surviving and in the ith category at age x. In general, where people are going in and out of the several categories, we cannot say that the elements of $\mathbf{l}(x)$ represent probabilities, yet probabilities are what we seek. Now suppose that in the small interval of time and age dx no one will be affected by more than one event. We would like to pass from $\boldsymbol{\mu}(x)$ and the vector $\mathbf{l}(x)$ to a matrix $\mathbf{L}(x)$ whose typical element is $l_{ij}(x)$, the chance that a person born in the jth state will be in the ith state by age x.

The theory for doing this is available from standard works on linear differential equations (Coddington and Levinson 1955, Gantmacher 1959, vol. 2, p. 113). If there are n states, and so the matrix $\boldsymbol{\mu}(x)$ is $n \times n$, and if the n eigenvalues of that matrix are distinct, then there will be n linearly independent vectors $\mathbf{l}_j(x)$, $j = 1, \ldots, n$ that satisfy equation (17.2.1). When this is so, the matrix made by setting those vectors side by side will obviously also satisfy the equation, and it can be shown to be the complete solution. Call $\mathbf{L}(x)$ the matrix made up of the several $\mathbf{l}_j(x)$ We shall see how to obtain the elements of $\mathbf{L}(x)$ so as to ensure that the ijth element is the probability that a person born in the jth category finds himself in the ith category by age x. The procedure is due to Rogers (1975).

17.2.1 The Multiplicative Property

One mathematical property of the $\mathbf{L}(x)$ will be important for the demographic application: its multiplicativity. It may be shown (though not here) that if the interval from zero to y is broken into two subintervals at any point, say $x < y$, then (Gantmacher, 1959, vol. 2, p. 127)

$$\mathbf{L}(y) = \mathbf{L}(y|x)\mathbf{L}(x), \tag{17.2.2}$$

where the ijth element of $\mathbf{L}(y|x)$ will in our interpretation mean the probability of being in the ith state at age y given that the person was in the jth state at age x. Since the interval from x to y may also be split into subintervals with the same property, we can divide the whole of any range into sufficiently small intervals (usually 1 or 5 years) that within each interval $\boldsymbol{\mu}(x)$ may be approximated by a matrix whose elements are constants independent of age. This will be the key to the numerical solution of (17.2.1).

What we cannot do is calculate directly the exponential of minus the integral of $\boldsymbol{\mu}(x)$, in analogy to what is possible for the one-region life table solution of equation 17.1.1. The exponential of an integral has meaning only when the matrix commutes. The relation of exponentials $e^{A+B} = e^A \times e^B$ requires commutativity, as *a fortiori* does $e^{\int A(x)\, dx}$. Only diagonal matrices and other uninteresting special cases are commutative. We must break down the interval from zero to x into subintervals short enough, say h in length, that the $\mu_{ij}(x)$ may be taken as constant within each of them. If in the interval $x, x+h, \mu_{ij}(x)$ is constant, say m_{ij} for all i and j, and $\mathbf{M}_x$ is the array of the m_{ij}, then from property 17.2.2 we can write

$$\boldsymbol{L}(x+h) = e^{-h\mathbf{M}_x}\boldsymbol{L}(x). \tag{17.2.3}$$

With an arbitrary radix $\mathbf{L}(0)$, equation (17.2.3) permits the construction of $\mathbf{L}(x)$ step by step at intervals of h all the way to the end of life. Alternatively, expanding the exponential in (17.2.3) to its first two terms gives

$$\mathbf{L}(x+h) = (\mathbf{I} - h\mathbf{M}_x)\,\mathbf{L}(x). \tag{17.2.4}$$

This approximation can be improved by first premultiplying (17.2.3) on both sides by $\exp(h\mathbf{M}_x/2)$ and then expanding to obtain the more symmetric

$$\left(\mathbf{I} + \frac{h\mathbf{M}_x}{2}\right)\mathbf{L}(x+h) = \left(\mathbf{I} - \frac{h\mathbf{M}_x}{2}\right)\mathbf{L}(x),$$

or on multiplying by $(\mathbf{I} + h\mathbf{M}_x/2)^{-1}$ on the left,

$$\mathbf{L}(x+h) = \left(\mathbf{I} + \frac{h\mathbf{M}_x}{2}\right)^{-1}\left(\mathbf{I} - \frac{h\mathbf{M}_x}{2}\right)\mathbf{L}(x). \tag{17.2.5}$$

Thus (17.2.3) is an approximation to (17.2.2) for $y - x = h$, and (17.2.4) and (17.2.5) are approximations to (17.2.3). The approximation (17.2.5) is close enough for many kinds of data with intervals of 1 year or even 5 years. It can always be improved by graduating the original data down to tenths of a year or smaller, and this was essentially what Oechsli (1971, 1975) did, using spline functions.

It is possible to escape from the restriction, implicit in (17.2.3) and (17.2.5), that the rates be constant over the interval h. Willekens (1978) and Krishnamoorthy (1978) have done this by using the Volterra theory of integration. As a consequence of (17.2.1),

$$\mathbf{L}(x+h) = \mathbf{L}(x) - \int_x^{x+h} \boldsymbol{\mu}(t)\mathbf{L}(t)\,dt$$

$$= \left[\mathbf{I} - \int_x^{x+h} \boldsymbol{\mu}(t)\mathbf{L}(t)\mathbf{L}^{-1}(x)\,dt\right]\mathbf{L}(x).$$

With observed or appropriately constructed curves for $\boldsymbol{\mu}(t)$ and $\mathbf{L}(t)$ within the interval h, the square bracket can be evaluated.

A third approach is due to Schoen and Land (1979). They obtain *flow* equations, the multi-dimensional analogue to $l_{x+h} = l_x - {}_hd_x$, representing relations within the life table. Alongside these are *orientation* equations, analogous to ${}_hM_x = {}_hd_x/{}_hL_x$, where ${}_hM_x$ is the observed rate. Finally the set is completed with numerical integration equations analogous to ${}_hL_x = (h/2)(l_x + l_{x+h})$. As in the single-region case, the solution can be given explicitly with a straight-line integration formula. With more elaborate integration formulae iteration is required.

The initial $\mathbf{L}(0)$ is arbitrary as far as the differential equation (17.2.1) is concerned; we shall define it as the identity matrix $\mathbf{I}$. In instances where a population model is to be constructed rather than a set of probabilities, so that radices other than $\mathbf{I}$ are required, those will be entered by multiplication: $\mathbf{L}(x)\mathbf{Q}$, where $\mathbf{Q}$ is a diagonal matrix containing the starting numbers or births in the several categories recognized.

17.2.2 Probabilities over Long Intervals

The most obvious question to ask is: what is the probability that a person in the jth state at age x will find himself in the ith state at age y, where the difference $y - x$ need not be small? Without matrix methods the problem is difficult and has even been thought unsolvable. It has to take account not only of movement out of the jth state but also of movement into the ith state of persons not in the jth state at age x. It may be solved by the multiplicative property referred to above as applicable wherever the interval (x, y) can be broken down into subintervals of width h, within each of which $\mathbf{L}(x+h)\mathbf{L}^{-1}(x)$ can be calculated by any of the methods cited earlier. If $l(y|x)$ is the desired set of probabilities, we know that $\mathbf{L}(y|x)\mathbf{L}(x) = \mathbf{L}(y)$, so multiplying on the right by $\mathbf{L}^{-1}(x)$ we get

$$\mathbf{L}(y|x) = \mathbf{L}(y)\mathbf{L}^{-1}(x), \tag{17.2.6}$$

where the probability of going from the jth state at age x to the ith state at age $y > x$ is the jth element of the ith row of $\mathbf{L}(y|x)$.

It can be argued that the differential equation (17.2.1) is a background of mathematical theory used only to provide a context for our symbols far more general than will be called for by demographic applications. Why could we not be satisfied to build up the $\mathbf{L}(x)$ matrix from the $\mathbf{M}_x$ matrix, step by step, in 5-year age groups starting from the unit matrix $\mathbf{L}(0) = \mathbf{I}$, using (17.2.5) at each step? We could even go through the arithmetic and obtain every probability required without ever introducing matrix notation, but the process of tracing individual combinations would be tedious. Moreover, the general theory assures us of the multiplicativity of the $\mathbf{L}(x)$ matrix, in the sense of (17.2.2), and from this all else follows.

17.3 Expected Time in the Several States

Beyond probabilities we would like to know the expected time lived between age x and $x+h$ in the several states, where in the first instance h is small. Let ${}_h\mathbf{C}_x$ be a matrix whose (i,j) entry is the time of residence in the ith state for those initially in the jth state. A straight-line approximation gives

$${}_h\mathbf{C}_x = (h/2)\left(\mathbf{L}_x + \mathbf{L}_{x+h}\right),$$

and a cubic gives

$${}_h\mathbf{C}_x = (13h/24)\left(\mathbf{L}_x + \mathbf{L}_{x+h}\right) - (h/24)\left(\mathbf{L}_{x-h} + \mathbf{L}_{x+2h}\right). \tag{17.3.1}$$

Adding ${}_h\mathbf{C}_x$ gives the person-years over any interval of age large or small. Cumulating ${}_h\mathbf{C}_x$ back from the end of the table gives the expected years in the ith state from age x to the end of life measured prospectively from birth in the jth state:

$$\mathbf{T}(x) = \int_x^{\omega-h} \mathbf{C}(a) = \int_x^{\omega} \mathbf{L}(a)\,da.$$

For an individual just born in the jth state, the probability of being in the ith state by age x is the ijth element of $\mathbf{L}(x)$. And if the expected number of years beyond age x in the kth state for those who survive to the ith state by age x is the ikth element of $\overset{o}{\mathbf{e}}(x)$ we must have

$$\mathbf{T}(x) = \overset{o}{\mathbf{e}}(x)\mathbf{L}(x), \tag{17.3.2}$$

where the right-hand side gives for the jth state at birth the number of years that can be expected if one reaches age x (and is then in the ith state) times the probability of reaching the ith state by age x.

Consider, for example, those in the second state at birth and let us find their expectation beyond age x in the first state. The second column of $\mathbf{L}(x)$ gives the chance that the person born in the second state is in the first, the second, and so forth, state at age x: If residing in the first state at age x, he has an expected $\overset{o}{e}_{11}(x)$ in the first state; if residing in the second, he has an expected $\overset{o}{e}_{12}(x)$ in the first; and so on. In short his total expectation in the first state, given that he was born in the second, is prospectively from age zero

$$\overset{o}{e}_{11}(x)l_{12}(x) + \overset{o}{e}_{12}(x)l_{22}(x) + \overset{o}{e}_{13}(x)l_{32}(x) + \cdots.$$

For the whole collection of states, we have

$$\overset{o}{\mathbf{e}}(x)\mathbf{L}(x) = \begin{pmatrix} \overset{o}{e}_{11}(x) & \overset{o}{e}_{12}(x) & \overset{o}{e}_{13}(x) & \cdots \\ \overset{o}{e}_{21}(x) & \overset{o}{e}_{22}(x) & \overset{o}{e}_{23}(x) & \cdots \\ \overset{o}{e}_{31}(x) & \overset{o}{e}_{32}(x) & \overset{o}{e}_{33}(x) & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} l_{11}(x) & l_{12}(x) & l_{13}(x) & \cdots \\ l_{21}(x) & l_{22}(x) & l_{23}(x) & \cdots \\ l_{31}(x) & l_{32}(x) & l_{33}(x) & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Note that here as in other expressions indexes are read right to left in order to use column vectors and the conventional subscripting of matrix elements.

Multiplying equation (17.3.2) by $\mathbf{L}^{-1}(x)$ on the right, we have for the expectation in the ith state for a person in the jth state at age x the ijth element of

$$\overset{o}{\mathbf{e}}(x) = \mathbf{T}(x)\mathbf{L}^{-1}(x). \tag{17.3.3}$$

An example is Table 17.2, calculated by Frans Willekens from 1967–72 United States data covering both sexes combined. It shows a total expectation of life of 71.08 years for those born in the Northeast. Of this time they will spend 13.16 years in the South. On the other hand, those born in the South will spend only 7.73 years in the Northeast, all on the average and provided that the rates of the given period, 1967–72, continue to apply.

Note that, on the definitions provided, the ijth element of $\mathbf{T}(x)$ is average time after age x spent in the ith state by those born in the jth state. For $\mathbf{T}(0)$ birth and initial residence are the same; but for any later $\mathbf{T}(x)$ they are different and give rise to two different expectations. Multiplying on the right by $\mathbf{L}^{-1}(x)$, as was done in equation (17.3.3), provides the expectations of stay in the ith state for each jth state of residence at age x.

To find the expected stay in the ith state for each state of birth requires a different denominator. The chance that the person born in state j is in state i at age x is $l_{ij}(x)$; the total $\sum_i l_{ij}(x)$ of the jth column of this through all i gives the chance that the person born in state j is still alive at age x, irrespective of where he lives at that time. If the diagonal matrix of these totals is called $\bar{\mathbf{L}}(x)$, then we have for the expectations $\bar{\overset{o}{\mathbf{e}}}(x) = \mathbf{T}(x)\bar{\mathbf{L}}^{-1}(x)$. Region of residence at age x has been duly summed out.

The multi-group life table often imposes distinctions not required in the ordinary life table.

Table 17.2. Life expectancies at birth by region, both sexes together, United States: 1967–72

	Place of birth			
Place of residence	Northeast	North Central	South	West
Northeast	41.73	5.84	7.73	6.57
North Central	8.19	39.89	11.95	11.64
South	13.16	14.69	39.52	15.24
West	8.01	10.66	11.30	37.70
Total	71.08	71.08	70.50	71.15

Source: Calculated by Frans Willekens.

Table 17.3. Fertility expectancies by region, both sexes together, United States: 1967–72

	Place of birth			
Place of residence	Northeast	North Central	South	West
Northeast	0.74	0.09	0.13	0.10
North Central	0.13	0.74	0.21	0.20
South	0.20	0.23	0.70	0.24
West	0.12	0.16	0.18	0.65
Total	1.19	1.22	1.22	1.19

Source: Calculated by Frans Willekens.

17.3.1 Fertility Expectations

Rogers (1975, p. 106) and Willekens and Rogers (1977) go on to discuss fertility expectations based on the same data, now relating births to total population age by age. The result is in effect an average of the male and female net reproduction rates. Whether for females alone or for both sexes we can write the age-specific birth rate in the ith region over a short time interval dx as $\boldsymbol{\nu}_i(x)$, construct the diagonal matrix $\boldsymbol{\nu}(x)$, and postmultiply by $\mathbf{L}(x)$:

$$\boldsymbol{\nu}(x)\mathbf{L}(x) = \begin{pmatrix} \nu_1(x) & 0 & \cdots & \cdots \\ 0 & \nu_2(x) & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} l_{11}(x) & l_{12}(x) & \cdots & \cdots \\ l_{21}(x) & l_{22}(x) & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

to obtain

$$\begin{pmatrix} \nu_1(x)l_{11}(x) & \nu_1(x)l_{12}(x) & \cdots & \cdots \\ \nu_2(x)l_{21}(x) & \nu_2(x)l_{22}(x) & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix},$$

whose ijth element gives the probability that a person (or woman) born in the jth region gives birth to a child (or girl child) in the ith region. When this is integrated over all x we have the multi-regional net reproduction rate $\mathbf{R}_0 = \int_\alpha^\beta \boldsymbol{\nu}(x)\mathbf{l}(x)\,dx$.

That $\mathbf{R}_0$ is shown in Table 17.3 for the United States (1967–72). The child born in the Northeast could expect to have 1.19 babies on the average. Of these only 0.74 would be born in the Northeast; 0.13 would be born in the North Central region, and so forth. The table shows the implications of the data about 1970 for the birthplaces of successive generations, averaging the sexes. Homogeneity is assumed here as throughout this chapter.

Table 17.4 shows e_{ij} for five states: single, once married, widowed, divorced, and in a second or later marriage. Thus the total expected future lifetime of Canadian males at the mortality levels of 1981 is 52.430 years for those single at age 20 and 53.587 for those already married. The single can expect 11.687 years single, $28.896 + 7.580 = 36.476$ years married.

17.4 Projection

A common demographic activity is population projection. If a given set of probabilities applies, and we know the vector representing the several categories of population at a certain point of time, the expected numbers at a later point of time are calculable. Suppose the age interval as well as the time interval to be h years as before; then for a population that happened to be concentrated at ages $0, h, 2h, \ldots$, the matrix $\mathbf{L}(x+h)\mathbf{L}^{-1}(x)$ would be appropriate for the projection in all cases where $\mathbf{L}(x)$ is nonsingular. But to approximate the observed population by a series of spikes seems inferior to approximation by a series of histograms.

We need a ratio corresponding to $S_x = {}_hL_{x+h}/{}_hL_x$ of the single-region case. This is obtained by applying the multiplicative property (17.2.2) to show that $\mathbf{L}_{x+h} = \mathbf{S}_x\mathbf{L}_x$, and multiplying on the right by $\mathbf{L}_x^{-1}$ to obtain $\mathbf{S}_x = \mathbf{L}_{x+h}\mathbf{L}_x^{-1}$, again assuming the inverse of $\mathbf{L}_x$ exists. (It does not exist for a marriage or labor force table prior to age 15 or so, and some other device is needed.) Now suppose the transitions pertaining to each given age,perhaps transitions among regions, are assembled into a block. Such blocks for the several ages x may be assembled into a matrix $\mathbf{S}$ with zero blocks everywhere except in the subdiagonal. The arbitrary (that is, observed) $\mathbf{n}(t)$ is projected to time $t+h$ by

$$\mathbf{n}(t+h) = \mathbf{S}\mathbf{n}(t). \tag{17.4.1}$$

This projects the part of the population already alive, but it gives no attention to births. To allow for them we need a matrix $\mathbf{F}$, whose nonzero elements are in its first row of submatrices, so the complete projection with time-invariant coefficients is (Rogers 1975, Feeney 1970)

$$\mathbf{n}(t+h) = (\mathbf{S}+\mathbf{F})\,\mathbf{n}(t). \tag{17.4.2}$$

Table 17.4. Multi-state life table: expected number of years in each of five states for Canadian males aged 20; vital statistics of 1981

	Single	Once married	Widowed	Divorced	Later marriage
Single	11.687	0	0	0	0
Once married	28.896	36.788	0	0	0
Widowed	1.904	2.182	10.669	2.164	2.182
Divorced	2.363	3.063	2.495	6.765	3.063
Later marriage	7.580	11.554	38.197	44.407	48.342
Total	52.430	53.587	51.361	53.337	53.587

Source: Author's calculation.

The procedure described by (17.4.2) permits a certain kind of experimentation with the elements of $\mathbf{S}$ and $\mathbf{F}$ and serves to ascertain the effects on future population of marriage ages, migration rates from more to less fertile regions, and so on. For forecasting purposes, some gains can be obtained by deliberately allowing the coefficients to change according to what one thinks may be the trends over future time. One must be cautious here; in more than one case varying the coefficients has produced a less accurate forecast than fixed coefficients would have done.

17.5 Transition Versus Instantaneous Probability of Moving

Most data on geographical mobility come in the form of transitions over a finite period—we know that the person was in state j at one moment of time and in state i 1 year or 5 years later. The census, for instance, asks people where they were 5 years ago without inquiring about intermediate moves. We shall speak of transitions in terms of a $\mathbf{P}$ matrix, whose p_{ij} is the probability that a person is in the jth state this year and in the ith state 1 year or 5 years later, in contrast to the $\boldsymbol{\mu}$ matrix, where the chance of moving from the jth state to the ith in infinitesimal time dx is $\mu_{ij}(x)\,dx$.

Although our theory started out in terms of movements in time separated by an infinitesimal interval dx, it then dropped the $\boldsymbol{\mu}$ and took no account of what happens within the finite interval beyond the probability that a person in the jth state at the beginning is in the ith state at the end. This is satisfactory if the time interval is short enough that two moves do not occur within it—that a person does not move and die in the same specified h years, for example. But what if the data cover a wide enough interval that several moves are possible?

Fortunately, the bias does not apply to most of the quantities calculated from the model. For projecting population according to the spike representation of the $\mathbf{L}(x)$, the observed transitions are obviously the right ones to use; one wants to omit intermediate transitions (people going from j to i and back again, all within one time interval) in the future as they have been omitted in the past. Slightly less obviously, the same applies to projection by histograms (equation 17.4.2). On the other hand, the expected number of moves calculated from the model will be understated if multiple moves occurring within the unit time interval of the data are neglected.

The Markov assumption of our model assigns the same probabilities to everyone in a given state at the beginning of each interval; no past history is allowed to influence the chance of transition beyond what is implied by the state that the person is then occupying. If we suppose this also applies in all subintervals, then the Markov assumption allows us to capture and add in those jumps that occur within intervals.

Table 17.5. Transition probability p, corresponding instantaneous rate μ, and probability of 0, 1, 2, and 3 moves in two-state model with $P_{ij} = P_{ji} = p$

Transition probability (p)	Corresponding instantaneous rate (μ)	Probability of			
		0 Moves	1 Move	2 Moves	3 Moves
0.001	0.001	0.999	0.001	0.000	0.000
0.050	0.053	0.949	0.050	0.001	0.000
0.100	0.112	0.894	0.100	0.006	0.000
0.150	0.178	0.837	0.149	0.013	0.001
0.200	0.255	0.775	0.198	0.025	0.002
0.250	0.347	0.707	0.245	0.042	0.005
0.300	0.458	0.632	0.290	0.066	0.010
0.350	0.602	0.548	0.330	0.099	0.020

Source: Calculated by Frans Willekens.

Consider transitions of one period between two states, in which $p_{ij} = p_{ji} = p$, so that the fraction of persons initially in the jth state that are found in the ith state at the end of the unit period is p. Suppose also that the instantaneous probability of moving (not known) in the time dt is $\mu\, dt$; then we can write out in terms of μ what fraction of the individuals initially in the jth state would be found in the ith state at the end of the period. It is those who made an *odd* number of moves, 1 or 3 or 5 or 7, this being as far as we need to go. If the instantaneous rate is μ, then by integration the probability of one transition in unit time is $\mu e^{-\mu}$, the probability of three transitions is $(\mu^3/3!)e^{-\mu}$, and so on, and the sum of these can be equated to p:

$$\begin{aligned} p &= \mu e^{-\mu} + \left(\mu^3/3!\right) e^{-\mu} + \cdots \\ &= \left(e^{-\mu}/2\right)\left(e^{\mu} - e^{-\mu}\right) = \left(1 - e^{-2\mu}\right)/2. \end{aligned} \tag{17.5.1}$$

Solving for μ gives

$$1 - 2p = e^{-2\mu}$$

$$\mu = -\tfrac{1}{2}\log(1 - 2p).$$

(We are grateful to Jan Hoem for this form of the solution.)

Table 17.5 gives values of p and the corresponding values of μ. It says, for example, that if the transition probability from the jth to the ith state is 0.2, and the transition probability from the ith to the jth is the same as this, and there are no other states, then an instantaneous rate of 0.255 will produce the 0.200 chance of transition. There will be a probability $e^{-\mu} = 0.775$ of no moves, $\mu e^{-\mu} = 0.198$ of one move, $(\mu^2/2!)e^{-\mu} = 0.025$ of two moves, and so on.

To generalize, if p_{ij} is the probability of a transition from state j to state i in unit time, we would like to find the instantaneous probability μ_{ij} (fixed over the interval) that would provide gross movement but still show a net of p_{ij}. We have discussed only the simple case where $p_{ij} = p_{ji} = p$ is the probability of transition in either direction between the two states.

The generalization that follows from the Kolmogorov equation (Kitsul 1978) is the matrix relation $\mathbf{P} = \exp(-\boldsymbol{\mu})$, or written out:

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & \cdots \\ p_{21} & p_{22} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} = \exp\left[-\begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \cdots \\ \mu_{21} & \mu_{22} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}\right], \qquad (17.5.2)$$

where the diagonals μ_{ii} are such as to make the columns sum to zero. If the matrices have distinct eigenvalues, (17.5.2) may be solved numerically in either direction.

Let us prove that (17.5.1) is a special case of (17.5.2). Take the matrices as 2×2 and make the μ's equal to μ or $-\mu$, so that (17.5.2) becomes

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \exp\left[-\begin{pmatrix} \mu & -\mu \\ -\mu & \mu \end{pmatrix}\right].$$

To evaluate the right-hand side, we need eigenvalues, eigenvectors, and spectral components. For eigenvalues,

$$\begin{vmatrix} \mu - \lambda & -\mu \\ -\mu & \mu - \lambda \end{vmatrix} = 0,$$

or

$$\mu^2 - 2\lambda\mu + \lambda^2 - \mu^2 = 0.$$

Therefore $\lambda = 0$ or 2μ. Left eigenvectors are $\begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathsf{T}}$ and $\begin{pmatrix} 1 & 1 \end{pmatrix}^{\mathsf{T}}$, up to constant factors. Right eigenvectors are $\begin{pmatrix} 1 & -1 \end{pmatrix}^{\mathsf{T}}$ and $\begin{pmatrix} 1 & -1 \end{pmatrix}^{\mathsf{T}}$. The spectral components are

$$\mathbf{Z}_1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad \mathbf{Z}_2 = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Hence

$$\exp(-\boldsymbol{\mu}) = \left(\frac{e^0}{2}\right)\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \left(\frac{e^{-2\mu}}{2}\right)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Multiply by $e^{-\mu}e^{\mu}$ to obtain as the $\mathbf{P}$ matrix

$$\left(\frac{e^{-\mu}}{2}\right)\left\{e^{\mu}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + e^{-\mu}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right\}.$$

The lower left element of this is

$$e^{-\mu}\left[\frac{e^{\mu} - e^{-\mu}}{2}\right] = e^{-\mu}\left(\mu + \frac{\mu^3}{3!} + \frac{\mu^5}{5!} + \cdots\right),$$

which is identical to (17.5.1).

Equation 17.5.2 allows us to infer the instantaneous probabilities of movement that are equivalent to a given set of transitions and hence to know the multiple transitions expected in any finite period. This is based on the Poisson process in which the probability of moving from the jth to the ith state is equal for all moments within the unit time period. In the case of death, a movement is the same as a transition and no reversal is possible; for other changes, transition in either direction has to be admitted. Under the stated assumptions, equation 17.5.2 read as an equation in $\mathbf{P}$ tells the transitions that correspond to a given set of movement probabilities; read as an equation in $\boldsymbol{\mu}$, it tells the movement probabilities that account for the given transitions.

In a series of papers Singer and Spilerman (1976) have shown how to find the continuous-time Markov chain of which an observed set of transitions can be considered the manifestation. Not every series of transitions is embeddable in a stationary Markov chain (that is, one with fixed parameters). They represent the solution of $\mathbf{P} = e^{-\boldsymbol{\mu}}$ formally as $\boldsymbol{\mu} = -\log \mathbf{P}$ and proceed to create a suitable definition of $\log \mathbf{P}$ in terms of the decomposition of $\mathbf{P}$ in spectral components. Because the logarithm has multiple branches, the result is not unique, and identifying the answer that corresponds to the problem in hand requires some ingenuity. The reader is referred to Singer and Spilerman for a highly sophisticated treatment of the relation between $\mathbf{P}$ and $\boldsymbol{\mu}$.

17.6 Stable Population

When we seek probabilities and expected values for individuals, the radix is taken as $\mathbf{L}(0) = \mathbf{I}$. But for stationary populations the radix must be the number of births into each of the categories. The point does not arise in tables of the single and married population, nor of the working population or the school population, because for each of these classifications everyone starts out in the same state—single, not in the labor force, and not at school. But the classification by territory is different: Portrayal of stationary conditions in the several regions requires that we enter as the radix a diagonal matrix $\mathbf{Q}$ showing the number of equivalent births in each region; here we use not $\mathbf{L}(x)$ but $\mathbf{L}(x)\mathbf{Q}$.

Like the usual single-region life table, our multi-group table is not only suited to provide probabilities and expectations but is also a population model for such groups as regions, occupations, marital statuses, or years of schooling. It gives the age and group distribution to which the age-group-specific rates would lead if they were in operation long enough that the peculiarities of the original distribution were forgotten.

But for this purpose the stationary model can readily be improved on by incorporating population increase. In the stable model the people now aged x were born x years ago when the population was smaller than it is now in the ratio e^{-rx}, so the number that would be counted as of age x would be proportional to the survivors from one birth times e^{-rx}. On a radix of one birth in each region the stable multi-group population would show the age and region distribution $e^{-rx}l(x)$; on a basis of q_i births in the ith region it would show

$$e^{-rx}l(x)\mathbf{Q}, \tag{17.6.1}$$

where $\mathbf{Q}$ is the diagonal matrix containing the q_i. The jth element of the ith row of $e^{-rx}l(x)\mathbf{Q}$ is $e^{-rx}l_{ij}(x)q_j$, which is the number of persons of age x out of the q_j born in the jth region that will be found at stability in the ith region. The $\mathbf{Q}$ is the multi-region analogue of the stable equivalent (Keyfitz 1968; see Chapter 8).

Such a model mimics observed populations with a closeness that depends on how nearly constant is the regime of mortality and fertility to which they have been subject. The model tells the implication of present rates—what the outcome will be if they continue unchanged for two or three generations. It tells, for example, where the Northeast would stand with respect to the rest of the United States if the age-specific rates of 1967–72 continued to prevail.

Formulae of this chapter have been applied to census and vital records that profess to cover the whole population. Another way of saying this is that the errors of such data are not of a random nature, and so the guarantees of accuracy that probability sampling can offer are not available for them.

Demographers increasingly gather their own data by sample survey methods. The results of such enquiries are usually tabulated in extensive cross-classifications, and bring the authors face to face with difficult questions of statistical significance. The pioneer on searching through contingency tables to find what conclusions can properly be drawn from them is Leo A. Goodman, whose log-linear methods are now widely used; for presentations relevant to our applications, see Bishop et al. (1975) and Fingleton (1984), for applications to the problem of choosing demographic state variables see Caswell (1988a, MPM Chapter 3).

18

Family Demography

For most purposes of population study there is no need to consider any unit intermediate between the individual and the larger group consisting of all the individuals included within some area—state, province, county, or nation. Recognizing only two units, population and individual, permits the construction of models that are readily expounded and understood. Demographers, following in the footsteps of Lotka and other predecessors, have worked hard to simplify this much-too-complex material.

The choice of unit for demographic (as for any other) analysis depends on the problem to be solved. For forecasts of total population, of the future labor force, or of the pension burden, it has seemed sufficient to work with the individual, and often without characterizing each individual beyond age and sex. One supposes that individuals give rise to other individuals over the course of time in a renewal process, irrespective of marriage or co-residence; individuals are discrete from birth; they live their separate lives, reproduce, and die.

For some purposes the recognition of an intermediate unit is unavoidable. Individual demography can tell us little about how the population fits into the housing stock; it can tell nothing about the kin networks among which mutual aid and protection take place. It falls short of explaining fertility change, insofar as the couple rather than the individual is the decision-making unit. Children are not born to couples independently at random, but couples take account of the number of children already born to them and of other aspects of their family situation at any given moment. When family constitution was stable it attracted less attention from scholars and lay people. Enormous changes in family and residential arrangements since World War II have aroused the current interest in family demography.

18.1 Definitions

This chapter concerns three kinds of unit intermediate between the individual and the population: the kin group, the residential family, and the household. The kin group are parents and children, siblings, nephews, nieces and cousins, irrespective of where they are living. The family, for census purposes, is not the totality of relatives but only those that live together in a given household or dwelling. This is the convenient unit of census enumeration—since the members know one another, the census information on all can be obtained from whoever answers the door when the enumerator calls. The members of the kin groups recognized by the census taker are in the first place the nuclear family—husband, wife, and children—and then the resident extended family that includes parents and in-laws of the husband or wife. The household is the persons living in a given residential unit, however that may be defined—by separate entrance from the street (i.e., without the need to pass through the premises of some other household), or full complement of cooking, washing, and other facilities. It is not easy to apply such definitions across cultures; those cultures in which privacy is important are likely to have more stringent physical requirements for a living space to be considered a separate household than those in which less privacy is desired or can be afforded.

Thus the unit for the analysis with which this chapter deals can be:

1. The kin group, irrespective of its living arrangement, and however dispersed. In demographic (though not in genealogical) work the data are confined to living members. As among kin groups one can recognize: (i) the nuclear family, consisting of father, mother, and children; (ii) the stem family, including ancestors in direct line and children of all generations; (iii) the extended family that includes all generations and all collaterals. Individuals belong in general to more than one family; even on the narrowest definition, individuals are usually members of a family of orientation, into which they are born, and a family of procreation, where they in their turn become parents.

2. The residential group, people living in a household, which is to say, sharing kitchen, bathroom, and other facilities, whether or not they are related.

3. Those members of a family that live in a given household. This cross between the kin group and the residential group is the natural object of census inquiries, and it is widely enumerated and tabulated. But it omits much; the common case where a couple have parents living in an

adjacent apartment is unrecognized; for many purposes propinquity is important and makes for what is in some respects a joint family, not captured by censuses.

18.1.1 Classifications

The problems of family demography do not end with the definition of the unit, but go on with the classification of types of that unit, whatever it may be. For the residential family, i.e., kin members within a household, the classifications used by censuses are many. The problem is to recognize the main forms of co-residence without multiplying excessively the number of types.

One summary solution is to classify by generations. Such a summary has been used by the Chinese census of 1982 and elsewhere. Comparing the Chinese material with that for Canada 1981, we find that one-person families were 20.3 percent of the whole in Canada, against 7.9 percent in China; one couple with no others, 22 percent in Canada against 4.8 percent in China. In both countries the biggest group was two generations (typically a couple and their children with no others) numbering 47.4 percent of the whole in Canada, 64.8 percent in China. Three or more generations, without unrelated members were 2.3 percent in Canada, 17.2 percent in China which accords with the common stereotype.

Trends over time are substantial and unidirectional. The United States proportion of one-person households was 13.3 percent in 1960, 17.5 percent in 1970. Three or more generations, with or without unrelated members, dropped from 4.7 percent to 3.5 percent in the United States in the course of the 1960s. Over longer periods the differences are dramatic; one-person households were 4 percent of the whole in 1790 and 23 percent in 1980.

One can speculate on the effect of China's new marriage law; will requiring children to look after their elderly parents prevent China from following the evolution of the West toward separate living for old people?

Another equally simple classification is given in the Canadian census of 1981. Five categories in all are recognized, and these are tabulated by age of wife and other characteristics. Thus we have

Husband-wife, no children	851,000
Husband-wife with children	3,267,000
Husband-wife, empty nest	968,000
Male lone parent	83,000
Female lone parent	464,000

out of a total of 5,632,000 families renting or owning their dwellings. By using the distribution of these by age of wife it is possible to see in cross-section the various stages in the family life cycle (*Canadian Statistical Review*, September 1984, p. viii).

Once the unit intermediate between the individual and the population is defined and enumerated in a census, so that classified counts are available of the number of households and the number of persons in individual households, tabulations can be made of the distribution of units by type (husband and wife and children, widow with child, etc.) according to census characteristics of individuals (occupation, earnings of senior member) and of the dwellings that contain them (number of rooms, rent paid).

Beyond these static qualities at the time of the census are family history materials, portraying longitudinally how individuals circulate through families over time—arrival of children, their departure 20 or so years later, the deaths of successive family members. Analysis of typical family histories and the way a family is transformed from one type to another, i.e., following the family through its event history, is the object of extensive modelling and simulation that will be discussed below. Comparative study between countries and cultures, as well as between historical epochs, is an important part of family history.

18.1.2 Theory and Statistical Compilation

The Italian statistician Barberi complained about the shortage of statistical data on the family, attributing it to the lack of theoretical study of the family as such (*in quanto tale*) as against its individual components (Barberi 1972). A statistical agency could not usefully collect data on an entity whose essential features had not been brought out by theoretical analysis.

On the other hand, the United Nations (1963) says the opposite—that the lack of theory is due to the lack of statistical information: "The paucity of demographic studies of families and households is due largely to the lack of pertinent census and survey data."

This is the classical problem of the direction of causation that dogs all social science. Both contentions are right. Theory and data influence each other reciprocally, and the absence of one is a handicap to the other. But an underlying factor operates to hamper both—as John Bongaarts (1983) tells us, the complexity of the subject makes it difficult both to gather data and to develop rational understanding.

Thus at the meta-level of the discussion of family demography, i.e., before we get into the subject proper, there lies a problem: does A cause B (Barberi), does B cause A (United Nations), or are both A and B caused by C (Bongaarts)?

18.2 Kinship

To see the effect of the flows of mortality, fertility, and divorce on kinship we need not census-type data but a model in which cross-sectional numbers of various kin are expressed in terms of those flows. A number of such models are given in Chapter 15 above; others in Goodman, Keyfitz, and Pullum (1974).

All of these formulae assume independence in the births. It requires a simulation experiment to find how the results must be modified to recognize nonindependence, in particular to impose the condition that after a birth at least 9 months (in practice usually much longer) must elapse before there can be another birth. Recently Le Bras (1984) has carried out such experiments and found that the average number of kin do not come out very differently with the recognition of dependence between successive births. Experimenting to ascertain distributions remains to be done; the microsimulation (or individual-based model) approach discussed in Section 15.9 is available for this.

One can start from the viewpoint of the parents as above and see how many children they would have alive (for example, after age 65) at given rates of mortality and fertility, or alternatively one can start with the children, and see how many would come into pension still having parents who are on pension. We saw (Section 15.1) that the probability that a woman aged α has a living mother is given by first considering the probability conditional on the child having been born at age x of the mother, in which case the chance that the mother is alive is $l(x+\alpha)/l(x)$, and then taking out the condition by averaging this quantity over all the ages of childbearing. The expression for the number of persons who have a living father is identical, provided we make the definitional modification that ages are measured from the time of conception (when we know that the father was alive) rather than from birth, say ages x^* and α^*, where $x^* \approx x + \frac{3}{4}$, etc. Hence the probability that a person aged α has a living father and mother is

$$M_1(\alpha)M_1^*(\alpha^*), \tag{18.2.1}$$

where we suppose that father's and mother's mortality are independent.

We want to translate these probabilities for individuals into a number for the population. Suppose that the age distribution of the female population is $p(\alpha)\,d\alpha$ and of the male population is $p^*(\alpha^*)$, where $\int_0^\omega p(\alpha)\,d\alpha$ is the total female population, etc.

Then the number of individuals, say over 65, who have a living mother and father is

$$\int_{65}^{\omega}\int_{65}^{\omega} M_1(\alpha)M_1^*(\alpha^*)\,\Big[p(\alpha)+p^*(\alpha^*)\Big]\;d\alpha\,d\alpha^*. \tag{18.2.2}$$

As with other formulae, the use of these is not to estimate the number of persons in the population with living mother and father (which can be done much better by a census or survey) but to find how that number varies with the mortality schedule, when all other variables are constant.

18.2.1 Inference from Kin Counts

One of the uses of expressions for the number of kin is what may be called backward inference, going not from the flow inputs—birth rates, etc.—to the number of kin, but from the numbers of kin as counted to the birth and other rates. Goldman (1978) showed how kin expressions for the number of sisters could be used to calculate the rate of increase of a population. She equates the observed ratio of younger to older sisters to the theoretical rates.

Suppose that we have a survey, such as an anthropologist might make, in which women are asked how many older and how many younger sisters were ever born to their mother, and the ratio of younger to older designated Z. The ratio of expressions such as those of Section 15.3 above can be equated to the observed Z and the unknown r, the rate of increase, calculated. An exact solution to the equation may be found by one iterative process or another, or else one may resort to approximations.

McDaniel and Hammel (1984) extended the idea to instances where one enquires not on the number ever born, but only whether the respondent is the first- or the last-born of her sorority or sibset. This is even easier for the respondent to recall. The ratio S of those who are youngest to those who are oldest provides an estimate of the rate of increase r whose sampling error is slightly greater than r obtained from Goldman's Z but is less subject to reporting error. It is hard to think of a survey question less demanding than whether the respondent is the eldest child, the youngest child, or somewhere between.

18.2.2 Widowhood

A couple are married at age x of the groom and age y of the bride; after t years the probability that both husband and wife are still alive is $l^*(x+t) \times l(y+t)/l^*(x)l(y)$, where $l^*(x)$ is the probability of surviving to age x from birth for a male, and $l(y)$ to age y for a female; we will throughout distinguish the male life table with an asterisk. The expected future lifetime together of the couple is the integral of this over t.

We can generalize by making x and y the ages of the couple at any time subsequent to marriage without needing to change the expression. Thus when either member has attained age 65 the expected future number of years together is given by the formula

$$\overset{o}{e}_{xy} = \int_0^\infty \frac{l^*(x+t)}{l^*(x)} \frac{l(y+t)}{l(y)}\, dt, \tag{18.2.3}$$

where now x and y are the ages of husband and wife at the given later point in their lives.

The probability that the wife will die first is

$$\int_0^\infty \frac{l^*(x+t)}{l^*(x)} \frac{l(y+t)\mu(y+t)}{l(y)}\, dt \tag{18.2.4}$$

and that the husband will die first is the same, but with $\mu^*(x+t)$ replacing $\mu(y+t)$. The sum of these is readily shown to be equal to unity in the same way that $\int_0^\infty l(y+t)\mu(y)\, dy/l(y)$ is unity. Continuing we can find the number of years of widowhood as the expected number of years a woman will live, $\overset{o}{e}_y$, less the number of years in the marriage, $\overset{o}{e}_{xy}$, that is $\overset{o}{e}_y - \overset{o}{e}_{xy}$, and of widowerhood similarly $\overset{o}{e}{}^*_x - \overset{o}{e}{}^*_{xy}$, assuming no remarriage.

Applying such formulae to the United States from 1950 to 1980 Goldman (1983) explains a large part of the observed increase in the number of widows as compared with widowers: with current life tables and ages at marriage the theoretical probability of a wife outliving her husband is about 70 percent. To bring the probability down to 50 percent would require that brides be about 7 years older than grooms. The 70 percent probability of the wife outliving her husband translates into an expected three to one ratio of widows to widowers in the population—partly due to the widow living longer after the dissolution of the marriage by death than does the widower, partly to higher remarriage rates for males.

But many marriages are dissolved otherwise than by death. It is not wholly realistic to neglect divorce, separation, and annulment. If we know the rate at which these occur we can assimilate it into the preceding formulae by adding it to the death rates of the partners. Call $\delta(t)$ the rate of dissolution otherwise than by death at t years after the marriage (or more generally t years after the couple are aged x and y, respectively). The probability of the marriage holding out to time t and then breaking up in the interval $(t, t+dt)$ is

$$\exp\left[-\int_0^t \delta(\tau)\, d\tau\right] \delta(t)\, dt \tag{18.2.5}$$

if we abstract from mortality. Including mortality requires entering this last expression in the preceding integrals. Thus the probability of the marriage dissolving at time t either by death of the male partner or divorce is

$$\frac{l^*(x+t)}{l(x)} \frac{l(y+t)}{l(y)} \exp\left[-\int_0^t \delta(\tau)\, d\tau\right] \Big[\mu^*(x+t) + \delta(t)\Big]\ dt. \tag{18.2.6}$$

18.2.3 Theoretical Number of Families in the Population

For the number of families and their ratio to the population, as expressed by the headship rate, a different approach is available. Suppose that we define any woman above a certain age who does not have a living mother to be the head of a household. The problem of finding the number of households in the country is then reduced to finding the number of women whose mother has died.

To derive this we started (Chapter 15) with the (conditional) probability $l(y+\alpha)/l(y)$ that a woman now aged α who was born at age y of her mother has a living mother. To remove the condition we averaged over all ages of childbearing, i.e., weighted by $e^{-ry}l(y)m(y)$, to find

$$M_1(\alpha) = \int_a^b \frac{l(y+\alpha)}{l(y)} e^{-ry} l(y)m(y)\, dy. \tag{18.2.7}$$

In terms of $M_1(\alpha)$ the fraction of women aged α who are heads of families must be

$$1 - M_1(\alpha) \tag{18.2.8}$$

and the total number of families is this weighted by the age distribution $p(\alpha)\, d\alpha$:

$$\int_a^\infty \left[1 - M_1(\alpha)\right] p(\alpha)\, d\alpha, \tag{18.2.9}$$

where a is some suitable minimum age.

The formulae are exact, but the model on which they are based is not as realistic as might be wished. Note that considerable departure from realism in these and other formulae need not vitiate conclusions drawn from them by sensitivity analysis. They are primarily used to find how much difference it makes to family constitution of mortality declines by 100δ percent, or fertility rises by 100ε percent.

18.2.4 Decomposing Widowhood

Widowhood and widowerhood are not mere consequences of improved longevity alone, but depend on the profiles of age-specific death rates of men and women. That a population has high or low mortality does not as such imply anything about widowhood.

If the peaking of the male curve is earlier than the peaking of the female curve then this by itself will result in many widows and relatively few widowers. But another factor is dispersion around the peak. Wide dispersion will produce many widows and widowers, even abstracting from the sex differential, i.e., if the peaks of mortality for men and women coincide. With little dispersion about the peaks, and peaks coinciding, then there will be few widows and widowers in the population. The model built on formula 18.2.4 will permit more precise statements than these, and they can be verified on the increment–decrement model of Chapter 17.

Calculations of the probability of widowhood that suppose a uniform age of marriage for men and another for women, i.e., that omit variability in marriage ages, will underestimate the amount of widowhood and widowerhood. The underestimate disappears if the age of marriage of men and women is the same and men and women all die at the same age, in which case there would be no widows or widowers. Allowing variation in the age of marriage for men and for women would make some widows and some widowers in the model. Thus one of the components of the amount of widowhood is variability in the age of marriage (a) for men and (b) for women. If the probability of dying varies among individual males of a given age, that will add what may be called a frailty component to widowhood.

The general supposition, then, is that the number of survivors at any given moment of marriages broken by death does not depend much on whether overall mortality is high or low, but on the difference between male and female mortality, and the breadth or narrowness of spread of both the male and female curves. This suggests a decomposition in which widows and widowers may be attributed to:

(1) Unequal average ages at marriage of men and women.

(2) Variations around that mean age for each of men and women.

(3) Unequal average mortality for men and women.

(4) Variation around that mortality for men and women, on the usual life table supposition that everyone in the population has the same chance of dying at any given age, but ages differ.

(5) Variation in frailty—i.e., probability of dying—of individuals of each sex about the average.

Each of these represents a certain kind of heterogeneity. For example, (2) is heterogeneity in ages at which individuals marry, and (5) is heterogeneity around the mean mortality of the person's sex at given age, unmeasurable for an individual but very real nonetheless. The full decomposition may be carried out by microsimulation.

18.3 The Life Cycle

The family, whether of kin or of residence, changes over the course of time. If one starts with a couple that have just married, then follows through as they have their children and as the children grow, one sees the residential family expanding over time to a certain maximum, perhaps 5 to 20 years after marriage in contemporary America, after which the children leave home, and the family shrinks back to the original couple, and sooner or later to only one of them. To compare two populations in regard to residential family size in effect averages much disparate material; what one should be comparing is the family sizes of the two populations at given stages as they go through this evolution.

Paul Glick, a pioneer in this field, made the first serious effort to trace the family life cycle statistically. His initial work analyzed cross-sectional census data; he compared family sizes at different ages of head. Subsequent work compared real cohorts; Table 18.1 gives an example.

The family life cycle is important in the study of migration. A high proportion of moves are associated with the beginning of working life, with marriage, and with retirement (Rogers 1984). A population in which there is a high proportion of young people will, other things being equal, have more movers than one in which there are many middle-aged. Those in which there are many retired persons who have made their move from the place of work to the place to which they retire will again have few movers. Movement is considerable for couples that have not had, or have just had, a child, and less for those with school-age children (Burch 1984, p. 182).

Life cycle theory has been developed in economics to allow for the fact that persons have some time-related options in regard to expenditures; within some limits they can advance or delay their purchases, and one can suppose they time their expenditures so as to maximize total satisfaction, which means to equalize marginal satisfaction at the various junctures of the life cycle. Such equalization must be subject to a personal discount rate. The distribution of expenses over time cannot be understood except in terms of the formation of a family and later its dissolution. When marrying and setting up a home people often dissave, buying on installment or borrowing in some other way until they have the "standard package"; only after that do their savings turn positive. The life cycle can be studied by demographic models with many states, either directly from a life cycle graph as in Chapters 3 and 7, or beginning with the multi-state life table as presented in Chapter 17.

18.3.1 Shape of Family Tree

As birth and death rates decline, the living kin group becomes elongated; in each generation there are fewer members, but more generations coexist (Glick 1977, 1979). The old person has fewer children and each of these has fewer children, but the chain extends further down—even to great grand-

Table 18.1. Comparison of live cycles of two cohorts, showing median ages of mother*

	First marriage about	
Median age of mother at	1905	1975
First marriage	21.4	21.2
Birth of first child	23.0	22.7
Birth of last child	32.9	29.6
Marriage of last child	55.4	52.3
Death of one spouse	57.0	65.2

**Source*: Paul Glick (1977).

children. To how many generations the chain extends depends on age at marriage and childbearing as well as on longevity.

The concept of a family cycle was early presented by Glick (1955, 1977). Riley et al. (1983) discuss the number of contemporary generations in an extended family with present longevity as against earlier high mortality. Probability of orphanhood, number of siblings, probability of three, or even four, successive generations alive at the same time, are all of importance and trends in them in relation to the trends in mortality and fertility need study. The elongation of the chain can be traced using expressions such as those of Chapter 15 above, developed in Goodman, Keyfitz, and Pullum (1974). The problem will be to remove some of the restrictive assumptions, or where this is impossible to estimate their effect on the results.

18.3.2 Headship

Brass (1983) has developed a useful way of tracing the effects of birth, death, marriage, and divorce. In the real world these are indeed proximate determinants; in the analysis they are inputs to the representation of family formation and dissolution. Brass's approximations show that at the low levels of mortality now attained it is age at marriage and fertility that count most for resident family size, with divorce becoming important in some circumstances.

Brass's way of doing family demography is to select a "marker" and to suppose a certain rate at which others attach themselves to this marker. One can imagine a girl child as a marker, first being born, sooner or later getting married—i.e., having a spouse attached—then having a child, then another child, then becoming divorced. Alternatively, one could start the process with a male marker and proceed similarly to the construction and ultimately the dissolution of the family unit.

That theory helps relate the resident family to various demographic factors (Burch 1979, Burch et al. 1983, Bongaarts 1983). In one application of this theory Brass ascertained the direct influence of fertility: a 10 percent increase in fertility results in a 4 to 6 percent increase in the size of the resident family. Using females as the markers for determining the advent of new resident families, it turns out that an increase of 1 year in the average age at which females leave home raises the family size by about $2\frac{1}{2}$ to 5 percent for a fixed level of fertility.

18.4 Household Size Distribution

The enormous differences in the size distribution of households are shown by the successive censuses (Table 18.2, provided by the Population Ref-

Table 18.2. Household size: 1790–1980
(Percentage distribution of households by number of persons)

Number of persons	1790	1900	1930	1940	1950	1960	1970	1980
Total	100	100	100	100	100	100	100	100
One	4	5	8	7	11	13	17	23
Two	8	15	23	25	29	28	29	31
Three	11	18	21	22	22	19	17	17
Four	14	17	18	18	18	18	16	15
Five	14	14	12	12	10	11	10	8
Six or more	49	31	18	16	10	11	11	6
Average persons per household	5.79	4.76	4.11	3.67	3.37	3.33	3.14	2.75

Sources: Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1970*, Part 1 (Washington, DC: U.S. Government Printing Office, 1975) Series A291 and A335–349, p. 42, and “Provisional Estimates of Social, Economic and Housing Characteristics,” *1980 Census of Population and Housing, Supplementary Report* (Washington, DC: U.S. Government Printing Office, 1982) Table P-1, p. 3.

erence Bureau). In 1790 just 4 percent of all households consisted of one person; in 1980, 23 percent were one-person households. In 1790, 49 percent of households consisted of six or more persons; in 1980, only 6 percent had six or more persons. During that time the average number of persons went down from 5.79 to 2.75. The trend to separate living was not uniform over the period, but seems to have been accelerating, with far more change from 1960 to 1980 than over the entire first century after the founding of this country.

18.4.1 Separate Living

The liberation to which our age testifies so amply has roots that go back to the Enlightenment, in ideas of individual worth, as Lesthaege points out (1983, cited in Burch 1984), and built into this is the liberation from the need to live with others. Pampel (1983) attempts to find to what degree the living alone that we observe can be explained by compositional variables, including age, income, etc. If the increase is fully accounted for by the fact that there are more old people, that incomes are higher, etc., then we do not need to search further. However, he finds that these compositional variables go only part of the way to explaining the observed increase in living alone.

Various writers have found that factors associated with modernization are associated with living alone. Correlations are positive between living alone and income or education and negative between living alone and fertility. Some of this could be due to intermediate variables—if fertility is lower for the better off, then they are more likely to live alone because they have fewer kin. (The effect is opposite to that in many nonwestern societies, where the joint family represents an upper-class way of living: the family property is what holds a large number of individuals in one household. The propertyless more often separate into nuclear families.)

The American positive association of living alone with income seems to imply that the better off are more isolated. But this is hardly in accord with many kinds of evidence that the better off are more likely to participate in social activities of all kinds, that they have more extensive kin and friendship networks. One can believe, however, that relatives are a larger fraction of the total social contacts of the poor than of the well off.

The components of the rise in separate living since World War II are multiple. Young people leave the parental home earlier, setting up separate households without either marrying or going to college. Retired people more and more live by themselves instead of with their married children. Divorce has increased, and both members of the former couple do not always remarry.

Underlying this is another level of causation, of which the literature stresses three elements.

(a) People have always wanted the privacy and independence of separate living, but only now have incomes risen enough to support it. This applies especially to the old, who have been aided by the growth of private pensions and social security.

(b) The simple absence of kin. With smaller birth rates there are fewer relatives in every generation, so a widow is less likely to have a child with whom she can live. What may be thought of as a narrowing of the genealogical tree must be a part of any explanation.

(c) Changed preferences. There has been a major change in the culture, of which one manifestation is the greater value set on privacy. This explanation has no value without an independent measure of it. Without such a measure the expression "wish for privacy" is just another name for separate living, and we are no further ahead.

Burch (1984) does better than simply telling us that there has been a cultural change; he argues that there has been a change in age and sex roles. With women's liberation and increasing freedom for children has come a situation where the members of the household are more or less autonomous, and all tend to resemble one another in their skills and activities. There is thus less room for a division of labor among them. That means less differentiation and so less of the solidarity that comes from the division of labor among differentiated individuals in the social as in the biological world, and consequently less to be gained from living together.

On the other side, "The young, the old, women, the unmarried, servants, boarders—all have been accorded more nearly equal 'rights' to various household goods previously reserved to the patriarch or breadwinner." This in effect makes the household more "crowded" in the sense that there is more competition for scarce, space-related goods. It is as though the democratization of the family has deprived members of the super- and subordinate niches that each formerly occupied. Thus on the one side adult numbers compete with one another for the scarce good of space, and on the other have little dependence on one another, so they might as well live separately.

18.5 Economic, Political, and Biological Theory

The economic theory of the family has been developed by Gary Becker (1991). Women have a comparative advantage in work in the home because it is they who have to bear and raise the children in any case, and while they are at home doing this they may as well also do the housework. Women need marriage to protect themselves against being abandoned with children whom they would not have the means to support. Within the family a degree of altruism exists: each person's utility function depends positively on the utility of others, and the family as a whole can be thought of as having a collective utility function.

Very different is a political theory of the family by which there is indeed a division of labor, but it is determined by the power and solidarity of labor, and not in altruism, but in their having no choice. In exchange for protection for life they were sheltered from (or kept out of) the world. Traditional societies could impose an acceptance of this breadwinner–homemaker family by suitable indoctrination of girls from earliest ages. Kingsley Davis has developed this realistic perspective in unpublished work.

A game theory model due to Luce and Raiffa (1957) offers a very persuasive explanation of past and present changes. In what they call the Battle of the Sexes there are two players, A and B. A, the husband, favors activity I and B, the wife, activity II. That alone would cause them to separate, except that both prefer activities in common. For A the utility of I is a, and this is greater than the utility b of II, which in turn is greater than c, the utility of the couple breaking up. Similarly, with primes, for B, except that for B the utility of I is b' and of II is a', with $a' > b'$. That is to say, A wants to do I and B wants to do II. For the marriage to be stable c, the utility for A of breaking up the marriage must be less than a or b, and similarly with primes for B. With this condition the solutions are both doing I or both doing II. The model admits two ways in which these can be arrived at: by altruism or by imposition. The outcome depends on which player is more anxious to avoid separation; until women worked separation could have been disastrous for the wife, indifferent for the husband. The observer is hard put to distinguish between the effect of altruism and the effect of power; the Luce–Raiffa model is convincing because it accepts both possibilities.

The utilities in the Luce and Raiffa payoff matrix are

$$\begin{array}{cc|c|c|l} & & \multicolumn{2}{c}{\text{(A) Husband's choice}} & \\ & & \text{I} & \text{II} & \\ \hline \text{(B) Wife's} & \text{I} & a, b' & c, c' & a > b > c, \\ \cline{2-4} \text{choice} & \text{II} & c, c' & b, a' & a' > b' > c'. \\ \cline{2-4} \end{array} \tag{18.5.1}$$

If $a' > b' > c'$, and $c' > c$ then the man has power. If c and c' are more nearly equal, then the sexes are more equal and the woman has a chance of attaining her goal. The model applies to a matter as trivial as going out to dinner versus staying home, as well as to the most solemn decisions that married couples make.

The implications of a utilitarian or contractual relation have been worked out in a tradition of thought about the family that runs through Durkheim, Schumpeter (1950, p. 157), and more recently William Goode. Thus Schumpeter in a characteristically farsighted phrase speaks of "the heavy personal sacrifices that family ties and especially parenthood entail under modern conditions," and Goode, "If larger part of one's life benefits are to be derived from job holding, in a social setting where emotional relations can be fleeting and superficial without incurring social disapproval, then it seems likely that future investments in the family ... may be lessened" (p. 79, in *Toqueville Review*). The stability of the family was all along based on the division of labor within it, but this could only be maintained with a concentration of authority, almost invariably in the male. Such a view, with its pessimistic conclusions, is supported by statistics of divorce, of later marriage, of nonmarriage, of increasing illegitimacy.

More upbeat is the argument that democracy within the family of course changes its nature, and of course no one can deny that divorce has increased, but the fact is that most people still marry, and the usual purpose of divorce is to escape one marriage in order to engage in another. It is almost as though the very concern about the quality of marriage and family life is what leads to divorce and trying again. No empirical evidence is likely to disprove this.

18.6 Family Policy

One test of theories of the family is whether policies based on such theories work. There has been a good deal of such testing in the United States by policy analysis and supplementation in recent years, especially in the field of welfare, and the results are not encouraging (Glazer 1984): "A program meant to reduce the distress of widows and unmarried mothers and their children, and designed to maintain them at a minimal but decent level, seemed to be accompanied by a rising number of such women. Welfare assistance was an incentive for mothers to push fathers out of the home."

The hope had been that mothers and children would have husbands and fathers, and that with the support of unemployment insurance, old-age pensions, and full-employment policies the family would be held together, with the husband-father the principal wage earner and Aid to Families with Dependent Children (AFDC) a transitory necessity only. But in fact AFDC became a program for mothers of illegitimate children, and the number of these increased rapidly, with half of the number being black.

If that incentive system was wrong, then could not ingenuity devise a better one? Just as the income tax does not deprive people of the incentive to work, so support should not prevent them from working, provided they were able to keep some fraction—say half—of their earnings. This application of the income tax principle did operate as theory said it should, but it too had a perverse effect: in repeated trials it raised the divorce rate. Evidently the grant made wives independent and a larger number of couples took advantage of this to separate or divorce. The Negative Income Tax could maintain the incentive to work, but only at the cost of increasing divorce (Hannan et al. 1977).

An example of the perverse way in which policy measures can operate is shown by the Swedish experience. In the postwar period Sweden experienced a labor shortage, and its response was various measures to encourage women to enter the labor force. The measures were successful. As among women with children, 44.6 percent were in the labor force in 1965 and 69.0 percent in 1975. But the longer-run effect was delay of childbearing and a lower birth rate. As the resultant small cohorts reach maturity the entrants into the labor force will diminish.

Thus measures to increase the labor force that were successful in the short run can have the effect of diminishing the labor force in the longer run.

Boudon (1977) gives other examples of the perversity of social life in the face of measures to influence it. More specifically in relation to our field, Henripin (1977) shows how difficult it is to alter demographic trends. Yet in certain less-developed countries genuine examples of the success of the policy are apparently to be found.

19

Heterogeneity and Selection in Population Analysis

Heterogeneity in the underlying population places difficulties in the way of interpretation of all statistical data based on averages. No two persons are equally likely to die in the next year; no two marriages are equally likely to be broken up by divorce; no two businesses are equally likely to fail; no two automobiles are equally likely to break down. That on the average a given make of automobile will travel 50,000 miles without major repairs offers little assurance for any particular automobile. Averages can be applied to individual cases only at great risk. This gross aspect of heterogeneity is *not* the subject of the present chapter.

People vary in respect of age, and mortality comparisons between populations have attempted to take account of age at least since the time of John Graunt. More recently, matrix and multi-state methods have made it straightforward to incorporate measurable sources of heterogeneity other than, or in addition to, age. Nothing more need be said here beyond the customary exhortation to break down aggregates into homogeneous subgroups for purposes of analysis.

But what about differences among individuals not ordinarily tabulated in mortality statistics? Mr. A has a heart murmur, or is a heavy drinker, and these affect his chances of survival. Beyond such differences, that could be distinguished in the data collection but are usually neglected, are differences that could not possibly be described. One cannot provide examples of the indescribable, but one can be certain that two individuals, even though alike in every possible statistical categorization, do not have identical probabilities of dying in the next year.

People thought of heterogeneity as at most affecting the precision of a mean; now we know that under many conditions it introduces a clear bias in the mean as customarily estimated. That bias arises because of a selection effect: in the random process under consideration those persons likely to die first disappear from the exposed group, leaving the group on average more robust. This selective effect acts always in the same direction—in the case of mortality to *underestimate* the rates to which the average individual will be subject.

19.0.1 Historical Note

The selection within a heterogeneous population that is our main concern has turned up in various areas of statistics and demography. An early recognition of the effect is that of Gini (1924) in a two-page paper. In the 1920s, Raymond Pearl followed a group of women from marriage to first pregnancy, counting the number of pregnancies over a considerable number of months, and dividing by the total woman-months of exposure. Gini showed that this gives too low a ratio to represent the pregnancy rate for an unselected group, because the less fecund women are steadily increasing their weight among the observations. Only the first month's observations provide a meaningful ratio, and one must throw away all the data for later months as having no reference to an unselected population. Potter and Parker (1964) tried by means of graduation to rescue these rejected data, but they were compelled to use strong assumptions on the distribution of fecundability among women (see also Wood and Weinstein 1990, Wood 1994)

In respect to migration it was observed (Blumen, Kogan, and McCarthy 1955) that to project a population forward through time as though everyone were subject to the rates of a given period exaggerates the subsequent movement. It is an improvement to suppose that the population consists of two kinds of people: those who never move, and those who have the same chance of moving in each period.

The Gini problem came to life again in the work of Sheps and Menken (1973). In respect of mortality it was taken up by Vaupel, Manton, and Stallard (1979), Manton, Stallard, and Vaupel (1981), and Vaupel and Yashin (1985). Shepard and Zeckhauser (1980) showed the issue to be important for the interpretation of medical trials. Reliability engineers have studied the question under the heading of mixed distributions (Mann, Schafer, and Singpurwalla 1974). Automobiles show higher failure rates in the first weeks of use than later. This is not likely to be due to individual items of equipment becoming better as they are used, the way an organism can gain strength; most of it must be due to defective machines being repaired or eliminated.

19.1 Conditioning and the Interpretation of Statistical Data

19.1.1 Simpson's Paradox

Some numbers will illustrate how unrecognized heterogeneity can make a comparison uncertain, in a situation more general than those discussed above in that it does not depend on selection over time (Cohen 1986, Simpson 1951). This pervasive aspect of heterogeneity deserves more attention than it has received. It can be summarized in the assertion that every comparison depends on the variables on which it is conditioned.

There are two regions, A and B, each with a 1,000,000 labor force. In A there are 100,000 unemployed, in B 90,000. That is, we have

A			B		
Labor force	Unemployed	Percent unemployed	Labor force	Unemployed	Percent unemployed
1,000,000	100,000	10	1,000,000	90,000	9

Apparently, A's unemployment rate at 10 percent is greater than B's at 9 percent. Abstracting from sampling error as we do throughout, nothing would seem more incontestable.

Yet, the obvious conclusion can be wrong. Suppose each of A and B to break down into two groups, young and old, with the following proportions:

	A			B		
	Labor force	Unemployed	Percent unemployed	Labor force	Unemployed	Percent unemployed
Young	500,000	75,000	15	300,000	48,000	16
Old	500,000	25,000	5	700,000	42,000	6

For the young, B has the higher unemployment, and likewise for the old. We have to reverse the conclusion based on the aggregate data; now B's unemployment is higher. Might a further breakdown make another reversal? There is no way of knowing, and the effect of underlying heterogeneity can only induce modesty in the presentation of the comparison. We have to visualize a penumbra of uncertainty based on the ever-present possibility that some unsuspected source of heterogeneity will show up, or exists but will never show up. It would be useful to develop some theory showing the bounds of that penumbra.

> An actual occurrence of this paradox was observed (Cohen and Nagel 1934, p. 449) in a comparison of tuberculosis deaths in New York City and Richmond, Virginia, during the year

> 1910. Although the overall tuberculosis mortality rate was lower in New York, the opposite was observed when the data were separated into two racial categories; in both the white and non-white categories, Richmond had a lower mortality rate. (Wagner 1982.)

A more recent real instance appears in the Canadian census of 1976:

	Mean number of children 1971–76	
	French	English
Canada	1.85	1.95
Quebec	1.80	1.64
Other provinces	2.14	1.97

Source: Réjean Lachapelle (1980), *La situation demolinguistique au Canada: Evolution passee et prospective.* L'Institut de Recherches Politiques.

Average children for the period between 1971 and 1976 were 1.85 for the population of French mother tongue and 1.95 for the population of English mother tongue. Should we conclude, then, that English fertility is higher than French? Not at all: in a breakdown of rates between Quebec and the rest of Canada, Quebec showed 1.80 for the French and 1.64 for the English. The same relation held for other provinces, with the French higher at 2.14 as compared with 1.97 for the English. The initial conclusion is vulnerable to challenge on the basis of heterogeneity.

The paradox can be grasped intuitively as an aspect of weighting. French speakers in Canada are mostly in Quebec; English speakers are mostly in the other provinces. Thus the Canada figure for the French is mostly a reflection of Quebec fertility; for the English it mostly reflects the other provinces. For the statistician there is no paradox here; he understands that every comparison depends on the variables on which it is conditioned.

Yet others are likely to feel uncomfortable. The man in the street might well say, "I am asking the simple question and looking for a yes or no answer: after all these years of legendary high fertility for the French has English fertility now come to be higher?" To say that it is literally higher but that we do not know whether to attribute this to people's being English or to their living in a certain part of the country sounds like an evasion. Yet we have to insist that the simple question is more difficult than it looks.

Extending somewhat an argument due to Colin Blyth (1972), suppose that we are trying out a new medical treatment in two hospitals and comparing it with the standard treatment. Suppose also that one of the hospitals is very famous and attracts the more difficult cases; the other gets easy cases most of whom would recover however they were handled. If the new treatment is tried mostly in the hospital with the more difficult cases, and the standard one in the hospital with the easy cases, then the new treatment will show a higher failure rate. What spoils any such comparison is the correlation between treatments and hospitals; the only way to escape the difficulty and find out which treatment is really better is to assign cases to the treatments at random in both hospitals. That is a recourse we do not have in the French–English fertility question. Randomization enables us to avoid the error that can be called misnaming: attributing to treatment what is really due to hospital.

Note that the difficulty is not removed when the whole experiment is done in a single hospital. If the assignment is not made at random there is always the possibility that some other variable—unobserved and perhaps unobservable—is influencing the comparison of the way the two hospitals were doing.

In Colin Blyth's (1972) notation that expresses the principle in its most general form, where it is known as Simpson's paradox, it is possible to have

$$P(A|B) < P(A|B') \qquad (19.1.1)$$

and yet both

$$P(A|BC) > P(A|B'C) \qquad (19.1.2)$$

and

$$P(A|BC') > P(A|B'C'). \qquad (19.1.3)$$

Here $B \neq B'$ and $C \neq C'$. In words, the probability of A in the Group B is less than that in B', even though within each of two subgroups, say C and C', the probability of A is greater for B than for B'.

Some deep and very general questions of interpretation of statistics are raised by Simpson's paradox. The effect of heterogeneity in the unobservables applies in all comparisons, and so affects nearly any interpretation of statistical data. The apparent solution, adding additional dimensions of cross-classification, creates its own problems, because as the number of categories increases, the data become more and more thinly spread over the possible combinations. Cohen (1986) refers to this as a demographic uncertainty principle.

19.2 Heterogeneity and Selection

If we take an unbiased sample of the unemployed and follow the individuals in it month by month to establish the duration of unemployment, we come face to face with the selective effect of heterogeneity. For those who are more capable or more active will find jobs sooner and drop out of observation, leaving in the observations those less likely to find jobs, and ultimately only those most difficult to place. Statistics based on this group will exaggerate the duration of unemployment, as well as the rate of unemployment.

If everyone had the same chance of finding a job the selective effect could not exist; the bias is the consequence of variability in that chance, i.e., of heterogeneity. This effect cannot occur instantaneously; it acts when a group is followed over time. If the heterogeneity is between observable categories it can be removed; the unobservable heterogeneity is unremovable, though indirect ways of estimating its effect have been suggested. Matters such as unemployment where the event can occur to an individual more than once are more tractable than mortality, and indeed one could in principle work out the "within person" variability of the chance of falling unemployed. Yet even in this case it is highly uncertain that heterogeneity as measured for one epoch is the same as heterogeneity for another.

Similar statements can be made about divorce. In most populations the probability of divorce is low in the first years of marriage, then rises to a peak ("the seven-year itch"), then falls off. Does that mean that the chance of a particular marriage breaking up rises to a peak, then declines? Not necessarily: if, for example, there are two kinds of couples—one with a low and constant probability of divorce, and one with a steadily rising probability, then the observations would be accounted for by an argument similar to that applying to death (Vaupel and Yashin 1985, Hoem 1990).

The subject matter fields affected are numerous: mortality, unemployment, divorce, risk of conception, migration, mechanical failure, in short any field where individuals drop out of the observed category when the contingency in question materializes.

19.3 Application to Mortality

In the usual deterministic life table model, for a person of given age the probability of surviving a year, say 0.99, is based on the collection of observed deaths and the corresponding exposed population. We may carelessly argue that different probabilities for individuals, being unmeasurable, have no meaning; after all the person will be alive or dead in the succeeding period—there is no middle possibility. But that argument falls to the ground when we think of one person in the hospital with a diagnosis of incurable cancer, and another of the same age, going about his business in evident good health. The unrealistic assumption of homogeneity has been thought to be innocent, in that it would not affect the overall conclusions drawn from the numbers. Differences between individuals, therefore, being both unmeasurable and inconsequential for averages, can be disregarded. This is the viewpoint that recent research has shown to be unacceptable. A person of average health or frailty* has something of the order of 1 year

---

*The "frailty" of an individual is supposed to measure the susceptibility of the individual to risks of death, beyond what is determined by age or other measured co-less expectation of life than is given by the life tables for the population of which he is a member.

If an insurance company were to set its rates from population statistics, and then insure a random subset of the population, it would break even (abstracting from expenses, etc.). But if its underwriters were to accept as risks only persons of average robustness, then it would lose.

A medical improvement can save many lives—i.e., it can prevent the death of persons who without it would have died—and yet *raise* the overall death rate in the population. It does this if it changes sufficiently the mix of frail and robust individuals in the population.

A mix of populations, in both of which mortality is rising steadily and indefinitely, can show a rise to a peak in the death rate, followed by a decline, then a further rise. That is to say there would be a period when mortality at the individual level is actually rising, while the overall statistics, the "observations" at the population level, show a fall.

This issue has become inescapable in studies of mortality among very old individuals. Studies of humans, as well as of laboratory populations of fruit flies, nematodes, and yeast, show mortality curves that decline, or at least decelerate, at the oldest ages (Vaupel et al. 1998). These data have been obtained by following huge cohorts (millions or even billions of individuals) to increase sample sizes among the rare extremely old individuals. Evolutionary theory predicts that the selective pressure against mortality should decline with age, so that traits with beneficial effects early in life and detrimental effects late in life should be favored (e.g., Hamilton 1966). This should lead to the accumulation of genes that increase mortality late in life, and to mortality rates that increase with age. The observations of declining mortality with age could mean that the evolutionary theory has left something out (e.g., Lee 2003), or that the predicted increase is masked by heterogeneity. The problem becomes only more acute (and interesting) as more is learned about the biological determinants of longevity (e.g., Wachter and Finch 1997).

19.3.1 A Mixture of Populations Having Different Rates of Increase

In one sense the mortality application is a special case of changing proportions of subpopulations that applies to overall population increase or decrease.

A population of initial size Q growing at rate r numbers Qe^{rt} at time t, r being taken as fixed and the population as homogeneous. Now suppose heterogeneity—a number of subpopulations, of which the ith is initially Q_i

---

variates. It could be determined by genetics, physiology, behavior, or the current or past environment.

growing at rate r_i, so that at time t the total number is $P(t) = \sum_i Q_i e^{r_i t}$. We showed (Section 1.5) that the total never stabilizes, that its rate of increase forever increases, and that the composition constantly changes.

The argument of Section 1.5 leads to the conclusion that the change in the mean rate of change is $\sigma^2(t)$, the variance among the rates r_i, each weighted according to its current subpopulation $Q_i e^{r_i t}$. That proves that with all r_i fixed the rate of increase steadily increases, unless the r_i are equal.

For two subpopulations that start out equal with unequal rates of increase the fact that the sum of the separate projections will always be greater than the combined population projection may be shown as a consequence of the arithmetic–geometric inequality. The sum of the separate projections from initial values of unity is $e^{r_1 t} + e^{r_2 t}$ and the single projection is $2e^{(r_1+r_2)t/2}$. The former is greater if

$$e^{r_1 t} + e^{r_2 t} > 2e^{(r_1+r_2)t/2},$$

i.e., if

$$\left(\sqrt{e^{r_1 t}} - \sqrt{e^{r_2 t}}\right)^2 > 0,$$

which is always the case if $r_1 \neq r_2$. The central proposition is that when one projects a heterogeneous population in disregard of the heterogeneity, which is to say using the average rate of increase for the whole, one *underestimates* the subsequent population. To project the population of the United States, for example, with the parameters of the country as a whole necessarily gives a lower answer than projecting each state with its own parameters, and then taking the total for the United States.

19.3.2 Two Classes of Frailty

An especially simple application of the foregoing is where a population with fixed mortality rates for individuals shows a spurious time-dependence. If at the outset half of the population suffers mortality rate μ_1, and half μ_2, and there are no births, then after x years there will be $e^{-\mu_1 x}$ individuals of the first type, and $e^{-\mu_2 x}$ individuals of the second type. The mortality rate will then be:

$$\frac{e^{-\mu_1 x}\mu_1 + e^{-\mu_2 x}\mu_2}{e^{-\mu_1 x} + e^{-\mu_2 x}}$$

and this is bound to be less than the initial rate $(\mu_1 + \mu_2)/2$. To show that we note by multiplying up that

$$\frac{e^{-\mu_1 x}\mu_1 + e^{-\mu_2 x}\mu_2}{e^{-\mu_1 x} + e^{-\mu_2 x}} < \frac{\mu_1 + \mu_2}{2}$$

if and only if

$$(\mu_1 - \mu_2)(e^{-\mu_1 t} - e^{-\mu_2 t}) < 0$$

and this last is bound to be true unless $\mu_1 = \mu_2$. We can prove it if $\mu_1 \neq \mu_2$ by trying first $\mu_1 > \mu_2$, then $\mu_1 < \mu_2$. Thus the mixed population in question with fixed death rates in each of its two subpopulations will show a spurious fall in its death rate over time.

19.3.3 Numerical Effect on Mortality

Since in its very nature unobserved heterogeneity cannot be measured statistically, it is not clear how we are to put to practical use the unquestioned fact that individuals vary in their probability of dying (or of becoming sick, divorcing, etc.). A person being either alive or dead, neither before nor after he dies is there any way of ascertaining his individual probability. All the facts on numbers exposed and dying contribute no information on individual variability and we are forever unable to make a life table for a single individual, except in the trivial *ex post* sense that his probability was unity of living to any age up to his death, and from then on was zero. Selection enters whenever we make a table for a group.

Yet we can consider the individual as the limiting case of a group. Suppose we move from the unselected population to those who are active and at work, to those who jog every day, etc. By projecting the sequence we obtain at least the order of magnitude of the probability for the individual least likely to die. And in the other direction, we can move from the unselected to the disabled, sick, in hospital, etc., to find the individual most likely to die.

Without empirical materials we can at least try out a hypothetical degree of heterogeneity. Suppose the population in three homogeneous groups, one with standard mortality, one with mortality of 20 percent of the standard, and the third with 180 percent of the standard, all taken in relation to a current life table. This may seem like a wide range, but it corresponds to a range of expectations of 65.3 to 81.8 years; that is not unbelievable.

Consider then that there is within a population a group whose life expectation is 81.8 years, another whose expectation is 72.10 years, a third whose expectation is 65.3 years, and suppose that these three groups are initially numerically equal. Under these conditions, the expectation as observed is 73.09, against 72.10 for the true expectation, i.e., the expectation of the average individual. The life table as calculated in disregard of heterogeneity exaggerates expectation by 0.99 years.

Thus a moderate conclusion from what we know now is that the observed expectation for ordinary populations calculated by the usual life table methods (Chapter 2) is high by about a year in application to the individual of average frailty at birth. We expect the effect to be much greater for mobility, divorce, or morbidity, since in these the individual is not restricted to the outcomes of 0 or 1 as for living or dead, but can move, or divorce, or become sick 0, 1, 2, ... times.

19.4 Modelling Heterogeneity

While measured heterogeneity can be incorporated into the life cycle graph and made part of a demographic model, unmeasured heterogeneity (sometimes called unobserved or unmeasurable) confronts the analyst with including the effects of something that he or she hasn't seen. Not surprisingly, this can be difficult. This section summarizes some of the approaches used and some of the problems that arise. For more details and extensive references, see Trussel and Rodríguez (1990), Vaupel (1990), Yashin et al. (1994), Yashin and Iachine (1997), Vaupel and Yashin (1999), and Link et al. (2002). We will speak in terms of mortality, but the principles apply to any kind of transition from one state to another.

19.4.1 Continuous Distribution of Frailty

Think of a standard individual, whose frailty at age zero is the mean of that of the population at age zero, and in the first round disregard any changes of frailty over the lifetime of a given person (Vaupel and Yashin, 1985). Call the frailty of that standard individual unity, and the frailty of any other person z, so that with $\mu(x)\,dx$ the conditional probability of dying between age x and $x+dx$ for the standard individual, the chance for the individual of frailty z is $z\mu(x)$. We express this assumption as $\mu(x|z) = z\mu(x)$. To suppose the frailty fixed through life for any individual simplifies the mathematics without committing us to any stand on the relative effects of heredity and environment in the action of mortality.

If for individuals the mortality is $z\mu(x)$ then the observed rate at birth $\bar{\mu}(0)$ depends on the distribution of frailty, say $f_0(z)$:

$$\bar{\mu}(0) = \frac{\int f_0(z) z\mu(0)\,dz}{\int f_0(z)\,dz} = \bar{z}(0)\mu(0). \qquad (19.4.1)$$

Following Vaupel and Yashin (1985) we will henceforth use a bar for the observed mortality, and the unbarred μ for the standard or initially average individual.

For a later age x the distribution of frailty is no longer $f_0(z)$ but $f_0(z)l(x|z)$ when $l(x|z)$ is the probability that an individual of frailty z survives to age x. We have for the observed mortality at age x

$$\bar{\mu}(x) = \frac{\int f_0(z) l(x|z) z\mu(x)\,dz}{\int f_0(z) l(x|z)\,dz} = \bar{z}(x)\mu(x). \qquad (19.4.2)$$

In words: the observed death rate is the mean frailty multiplied by the death rate for the standard individual. This simple but fundamental result factors $\bar{\mu}(x)$ into a part involving z and a part involving μ.

Calculating the derivative of $\bar{z}(x)$ will show that mean frailty steadily declines through life under all circumstances where mortality is positive.

In the first place we know that

$$\begin{aligned} l(x|z) &= \exp\left[-\int_0^x \mu(\alpha|z)\,d\alpha\right] \\ &= \exp\left[-\int_0^x z\mu(\alpha)\,d\alpha\right] \\ &= \exp\left[-\left(\int_0^x \mu(\alpha)\,d\alpha\right)(z)\right] = \left[l(x)\right]^z . \end{aligned}$$

Entering this in an expression for the mean frailty gives

$$\bar{z}(x) = \frac{\int f_0(z)\left[l(x)\right]^z z\,dz}{\int f_0(z)\left[l(x)\right]^z\,dz}. \tag{19.4.3}$$

To find how $\bar{z}(x)$ changes over successive ages we take the logarithm of $\bar{z}(x)$ and differentiate to obtain

$$\frac{1}{\bar{z}(x)}\frac{d\bar{z}(x)}{dx} = \frac{\overline{\left[z(x)\right]^2}\mu(x)}{\bar{z}(x)} + \bar{z}(x)\mu(x),$$

where we write $\overline{(z(x))^2}$ for the mean square of $z(x)$. The algebra is a further variant on that employing the arithmetic–geometry inequality. We finally have

$$\frac{d\bar{z}(x)}{dx} = -\sigma^2(z|x)\mu(x), \tag{19.4.4}$$

which is necessarily a negative quantity. We have thus shown that with positive mortality and positive variance of z mean frailty can only decline. Expression 19.4.4 describes the selection effect as it diminishes frailty through the life of a cohort. Together with (19.4.1) it shows that the observed $\bar{\mu}(x)$ is increasingly biased (relative to the true frailty $\mu(x)$ at the individual level) as the cohort ages.

Note that the results here obtained are applicable starting from any age, and not only from age zero. The selection that starts to operate in a cohort just born operates equally in a cohort starting at an arbitrary age. Even for an advanced age, say 70, whatever distribution of frailties is to be found will gradually change with differential survival through ages 75, 80, etc. Age zero is a perfectly arbitrary starting point for this as for other purposes; one could develop a demography for age 15, age 70, or any other age, disregarding the individuals below that age, and consider a birth to be the entry into that starting age. One could go the other way and measure from conception as was done in Chapter 18.

The results in (19.4.1) to (19.4.4) can be extended if we parameterize the distribution $f_0(z)$. A convenient choice is the gamma distribution:

$$f_0(z) = \frac{k^k z^{k-1} e^{-kz}}{\Gamma(k)},$$

where k is the reciprocal of the variance σ^2. Without assuming anything about the value of k, Vaupel (1979) shows that with this $f_0(z)$ mean frailty at age x is

$$\bar{z}(x) = \frac{1}{1 + \sigma^2 \left(\int_0^x \mu(\alpha)\, d\alpha \right)}. \tag{19.4.5}$$

This last tells us that the course of frailty with age is a simple function of death rates below each age.

As an example of the use of (19.4.5), if mortality is age-invariant, so that $\mu(x) = \mu$, then the course of observed mortality, calculated from (19.4.2), will be hyperbolic

$$\bar{\mu}(x) = \frac{\mu}{1 + \sigma^2 \mu x}.$$

This expression permits calculating the effect numerically knowing only the variance of the z, assuming a gamma distribution of frailty.

Alternatively, if the baseline mortality follows the Gompertz–Makeham model

$$\mu(x) = ae^{bx} + c \tag{19.4.6}$$

with a, b, and c constants, then the true mortality of each individual increases exponentially with age. But the observed mortality will be

$$\bar{\mu}(x) = \frac{ae^{bx}}{1 + \sigma^2 \frac{a}{b}\left(e^{bx} - 1\right)} + c, \tag{19.4.7}$$

which is a logistic increase with age, reaching an asymptote of $b/a\sigma^2$ (Yashin et al. 1994). The effect of selection completely obscures the pattern of mortality increase at the individual level.

19.4.1.1 Parameter Estimation

Suppose that data are available on a cohort of individuals, followed until the last one dies. We want to estimate, via maximum likelihood, a set of parameters defining $f_0(z)$ and $\mu(x)$. The likelihood of a set of parameter values is proportional to the probability of observing the data given those values. The fate of an individual depends on a frailty z specific to that individual; in a continuous-frailty model, each individual has a unique value of z. Thus the number of parameters to be estimated appears to exceed the number of individuals in the sample; this is not promising.

A solution is to focus on the distribution $f_0(z)$ rather than on the values of z realized by the individuals. This treats the individual frailties as random variables rather than as fixed but unknown constants. This view is familiar in the Bayesian framework, where *all* parameters are viewed as random variables. If we let Y denote the data and $\boldsymbol{\theta}$ a vector of parameters, then a Bayesian analyst writes

$$f(\boldsymbol{\theta}|Y) = \frac{P(Y|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int P(Y|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{19.4.8}$$

where $f(\boldsymbol{\theta}|Y)$ is the distribution of $\boldsymbol{\theta}$ after observing the data Y (called by Bayesians the *posterior distribution*), $P(Y|\boldsymbol{\theta})$ is the probability of the data given $\boldsymbol{\theta}$ (the *likelihood*), and $p(\boldsymbol{\theta})$ is the unconditional distribution of $\boldsymbol{\theta}$ before the data Y are obtained (the *prior distribution*). That is, the prior distribution of θ is updated, after data are obtained, via the likelihood function. The posterior distribution shows which values of the parameters are probable and which are not. The mean or the mode of the posterior distribution can be used as a point estimate of the parameter if desired.

Bayesian statistics inspires philosophical arguments about the influence of the prior distribution on the conclusions. Depending on your point of view, this influence is either good, representing the growth of scientific knowledge by updating prior beliefs in the light of new data, or bad, corrupting the scientific process by permitting a subjective prior distribution to influence the conclusions. We can leave the philosophers to their own devices, however, by taking what is sometimes called an *objective Bayesian* approach (e.g., Link et al. 2002). If the prior distribution in (19.4.8) is flat (or very nearly so), it provides no (or nearly no) information about $\boldsymbol{\theta}$ prior to obtaining the data. The posterior distribution is then proportional (or nearly so) to the likelihood function. The mode of the posterior distribution is then the maximum likelihood estimate of $\boldsymbol{\theta}$. Thus Bayesian methods can be used to arrive at maximum likelihood estimates uninfluenced by prior beliefs. The advantage is that this includes models where some parameters, such as individual frailty, are treated as random variables characterized by their distributions.

The integral in the denominator of (19.4.8) can be extremely difficult to evaluate, but the so-called Markov chain Monte Carlo (MCMC) algorithm solves this problem, and powerful software for this purpose is now available (Spiegelhalter et al. 1999; see Link et al. 2002 for a description in the context of population biology).

MCMC methods are attracting attention from both human demographers (e.g., Bolstad and Manda 2001) and population ecologists. As an example, Cam et al. (2002; see also Link et al. 2002) studied survival and fertility in the black-legged kittiwake (*Rissa tridactyla*, a gull of temperate and arctic waters in the Atlantic and Pacific oceans). Over a period of 22 years, some 845 individuals were followed from first reproduction until death or the end of the study. For each bird, the time it entered the study and the time of its death are known (unless it survived to the end of the study), and for each year that it was in the study it is known whether or not it reproduced. Cam et al. (2002) modelled survival and fertility as

$$\operatorname{logit}\big(\phi(i,x,t)\big) = a(t) + b(x) + z_i^{(\phi)} \tag{19.4.9}$$

$$\operatorname{logit}\big(\beta(i,x,t)\big) = c(t) + d(x) + z_i^{(\beta)}, \tag{19.4.10}$$

where x denotes age and t denotes time, and i refers to a particular individual; the frailties $z_i^{(\phi)}$ and $z_i^{(\beta)}$ are random effects on survival and fertility specific to individual i.

Assuming that the frailties $z_i^{(\phi)}$ and $z_i^{(\beta)}$ were fixed, individual-specific effects would require estimating $845 \times 2 = 1690$ parameters to account for inter-individual heterogeneity. However, assuming that the individual frailties were drawn from a bivariate normal distribution with a mean of zero reduces the number of parameters to 3 (two variances and a correlation). Flat priors were used for these 3 parameters, as well as the effects of age and time, and MCMC methods used to compute the posterior distributions of all the parameters. A variety of models for the time and age effects were explored; the best-fitting models all included significant individual heterogeneity. At the individual level (i.e., the age effect $b(x)$), survival declined linearly with age. Because of selection, however, survival at the population level was nearly age-invariant. The estimated correlation between the individual effects on survival and reproduction was unambiguously positive, indicating that, at the level of individual heterogeneity, “frailty” affects both survival and reproduction in the same way.

19.4.2 *Finite Mixture Models*

Often there is little theoretical basis for choosing a parametric distribution of frailty. It is tempting to choose a function on the basis of convenience (e.g., the gamma distribution, which has support on the non-negative numbers), but in at least some circumstances different choices of the distribution lead to different estimates of the functional relationship between survival and other (measured) variables (Heckman and Singer 1982). A nonparametric approach based on a mixture of a finite number of frailty classes may be preferable; it is certainly appropriate when there are really classes within the population (those with and without some medical condition, for example). In any case, many of the effects of heterogeneity become apparent by considering only two groups (Vaupel and Yashin 1985).

Suppose there are N types, with frailties z_i, $i = 1, \ldots, N$, and that $p_0(z_i) = P(z = z_i)$. Then we replace (19.4.1) with

$$\bar{\mu}(0) = \sum_i p_0(z_i) z_i \mu(0) = \bar{z}(0)\mu(0). \tag{19.4.11}$$

The theory proceeds as in the continuous case, but with summations replacing integrations (Vaupel and Yashin 1999). Maximum likelihood can be used to estimate the parameters in the mortality schedule $\mu(x)$ and the probabilities $p_0(z_i)$. In cases where the follow-up of individuals is less than perfect, so that an individual may disappear either because of death or because of imperfect sampling, the heterogeneity can affect not only mortality but the likelihood of detection (e.g., Pledger 2000, Pledger et al. 2003). The choice of the number of groups to include (N) can be based on prior knowledge or on a comparison of models with different numbers of classes. Hoem (1990), however, reports on a failed attempt to use this approach to study heterogeneity in the breakup of conjugal unions (married or otherwise) in Sweden; he found the number of parameters too large and the likelihood too difficult to maximize. Whether Bayesian MCMC methods would help is an open question.

19.4.3 Age-Dependent Frailty

The models discussed so far assume that each individual has a fixed frailty, assigned at birth and kept until death. An alternative conceptualization assumes that all individuals are born with the same frailty, and that frailty then changes with age (e.g., Yashin et al. 1994). Suppose there are discrete frailty states $0, 1, \ldots, n$, representing increasing levels of frailty, and that frailty affects mortality as

$$\mu(z, x) = \mu_0(x) + z\mu. \tag{19.4.12}$$

Here, μ is the rate at which mortality increases with increased frailty. Frailty changes over time, either increasing according to

$$P(z(t+dt) = z+1) = (\gamma_0 + z\gamma)\, dt \tag{19.4.13}$$

$$P(z(t+dt) = z) = 1 - (\gamma_0 + z\gamma)\, dt. \tag{19.4.14}$$

Here, γ is the rate of increase in frailty.

This model leads to a system of differential equations that can be solved for the probability distribution of individuals among frailty classes at any age. From this, as $n \to \infty$, the observed mortality is

$$\bar{\mu}(x) = \mu_0 + \frac{\mu\gamma_0 \left(1 - e^{(\gamma+\mu)x}\right)}{\mu + \gamma e^{-(\gamma+\mu)x}}, \tag{19.4.15}$$

which produces a logistic pattern of observed mortality (Yashin et al. 1994).

19.4.4 Parameter Identification and Model Discrimination

Not surprisingly, any attempt to estimate the effect of invisible variables leads to problems with identification and model discrimination.

First, as emphasized by Trussell and Rodríguez (1990), there are two ways to write the mortality rate in the presence of heterogeneity. One is to specify a distribution of frailty and incorporate it in the formula (19.4.2). The other is to carry out the integration over levels of frailty in (19.4.2) and obtain a resulting mortality curve in which the distribution of frailty is no longer apparent. There is no way to tell, from a mortality schedule alone, whether it is the result of heterogeneity. Thus observed mortality data could follow the logistic age trajectory in (19.4.7) for two reasons. Each individual in the population might experience the same logistic trajectory of mortality, or each individual might experience the exponentially increasing Gompertz–Makeham mortality trajectory (19.4.6) modified by its unique, gamma-distributed, frailty. No observations on mortality alone can distinguish the two hypotheses, because they lead to exactly the same mortality trajectory.

The emphasis here must be on "mortality alone," because it is certainly possible that additional information could help distinguish the models. One can reduce heterogeneity (e.g., by conducting experiments on genetically homogeneous cohorts) and see if it changes the form of the mortality function. Or one can use information from relatives, whose frailty may be expected to exhibit some correlation, and see how much heterogeneity is reduced (e.g., Yashin and Iachine 1997).

It is also impossible, under at least some hypotheses, to distinguish fixed from age-dependent frailty. Yashin et al. (1994) showed that the parameters of the age-dependent frailty model (19.4.15) can be put in one-to-one correspondence with the parameters of the gamma–Makeham model (19.4.7). Thus observations of mortality alone cannot distinguish between a frailty model where each individual has its own fixed level of frailty assigned at birth and one where all individuals are born with the same frailty, which increases stochastically through their lives.

19.5 Experimentation

As an example of the way heterogeneity can distort even a careful medical trial, consider a population that is heterogeneous, containing two equal subpopulations, both initially stable, equal in number and increasing at the same rate r, but with different mortality (and correspondingly different birth) rates. A procedure is applied that lowers mortality by amount ε at all ages for one of the subpopulations. Under what circumstances will the use of this procedure *raise* the overall crude death rate of the population?

If the initial common rate of increase is r, the two birth rates b^* and b, then the initial crude death rate is

$$(\mathrm{cdr})_0 = ((b^* - r) + (b - r))/2.$$

Suppose the procedure lowers mortality by ε uniformly at all ages for the starred subpopulation, increasing its r to $r + \varepsilon$, and having no effect on its crude birth rate b^*. The combined population that results at time t will have a weight of $e^{\varepsilon t}$ for the starred subpopulation and unity for the other, so its crude death rate will be

$$(\mathrm{cdr})_t = (e^{\varepsilon t}(b^* - r - \varepsilon) + (b - r))/(e^{\varepsilon t} + 1).$$

Then $(\mathrm{cdr})_t$ will be greater than $(\mathrm{cdr})_0$ if and only if

$$\frac{e^{\varepsilon t}(b^* - r - \varepsilon) + (b - r)}{e^{\varepsilon t} + 1} > \frac{b^* - r + (b - r)}{2}.$$

Ultimately as t becomes large, the condition becomes $\varepsilon < (b^* - b)/2$ which is the same as

$$\varepsilon < \frac{(b^* - r) - (b - r)}{2}. \tag{19.5.1}$$

This shows that if the decrease in mortality rates affects the subpopulation of higher mortality only, and if the improvement is less than half the difference between the preceding death rates of the two subpopulations, then the effect of the improvement will be an *increase* in the overall mortality rate.

20

Epilogue: How Do We Know the Facts of Demography?

Demographers know that a slowly increasing population has a higher proportion of old people than one that is increasing rapidly, and that differences in birth rates have a larger influence on the age distribution than do differences in death rates. They often claim that a poor country whose population is growing rapidly will increase its per capita income faster if it lowers its birth rate rather than maintaining it at a high level.

How do demographers know these things? Many readers will be surprised to learn that in a science thought of as empirical and often criticized for its lack of theory the most important relations cannot be established by direct observation, which tends to provide enigmatic and inconsistent reports. Confrontation of data with theory is essential for correct interpretation of such relationships, even though on a particular issue it more often generates an agenda for further investigation than it yields useful knowledge. Much of this book is devoted to examining the ways in which demographers distill knowledge from observation and from theory. The present summing up shows a relatively heavy weight of evidence for theory, illustrated briefly with an application to economic demography. We thank Paul Demeny for many improvements and clarifications in this account.

Let no one think that the questions of demography, and the issues of method for finding answers to them, are remote or purely abstract. The resolution of major policy issues of our time depends on the answers. How much of its development effort should a poor country put into birth control if it deems its rate of population growth excessive? Some would put nothing, in the expectation that rapid increase of income will by itself bring population under control. Once people have automobiles, once their countryside is

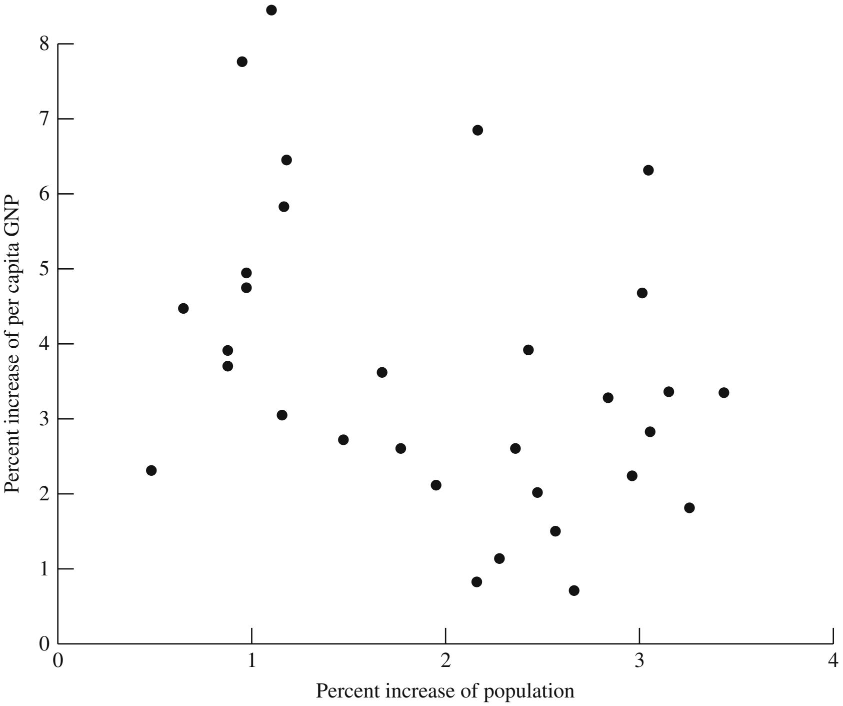


Figure 20.1. Average annual increase of per capita gross national product and of population for countries with over 20 million population, 1960–72.

paved over with roads, once enough air-conditioned houses are built, they will lower their fertility. But is this not a circuitous way of getting people to use pills and IUDs? Surely direct persuasion aimed at lowering fertility would help to reach desired developmental goals more quickly.

Any answer to such questions must take into account the degree to which a low rate of population increase promotes development and conversely. This is no simple matter. Figure 20.1 shows the relation between rates of population growth and increase of income per capita. Even the most imaginative viewer would hardly perceive the negative relationship that the dominant theory (later to be summarized) requires. In the pages ahead, the irregularity of empirical data as they appear in charts and tables will be repeatedly contrasted with the clear-cut mathematical relationships of the theory.

The theoretical approach can be described as "holding unmentioned variables constant"; the empirical, for example in the form of a regression between measured variables, as "allowing unmentioned variables to vary as they vary in actuality." This distinction is familiar, more so in the natural than the social sciences, in the context of experiments. An experiment is, and is intended to be, an artificial situation in which some factors are manipulated while others are held constant. The power of experiments is precisely their ability to draw conclusions about the manipulated factors, in the context of the factors held constant (or randomized, which has much the same effect). Theoretical models of the sort considered in this book can be thought of as mathematical experiments (Caswell 1988b). The principles of good experimental design are many, but they do not include trying to duplicate nature, or to manipulate all the relevant factors at once. Still, the results of an experiment are always conditional on the choice of which variables to manipulate and which to hold constant, and the ability of an experiment to predict the behavior of the real world depends on how those unmentioned variables actually vary.

The contrast between the theoretical and empirical approaches becomes clear when we look at an example in which we think we know the true nature of the relationship between two variables.

20.1 Growing Populations Have Smaller Proportions of Old People

In the late 1960s the population of Mexico grew at 3.5 percent per year; its proportion of ages 65 and over was about 3 percent. The United States was growing at less than 1 percent per year; its proportion 65 and over was about 10 percent. The relation can be expressed as a linear equation. For 1966 (Keyfitz and Flieger 1971, pp. 344, 354) the four numbers were as follows:

Variable	Mexico	United States
Rate of natural increase (percent)	3.44	0.89
Percent aged 65 and over	3.31	9.42

Call the annual percentage rate of increase $100r$, and $100P_{65+}$ the percentage over age 65. Then the straight line from the 1966 information for the United States and Mexico is

$$P_{65+} = 0.116 - 2.40r,$$

which tells us that for each 1 percent by which the rate of increase is higher there is a decrease of 2.4 percent in the proportion aged 65 and over. With zero increase the percentage over 65 would be 11.6; with 3 percent increase it would be $11.6 - 7.2 = 4.4$ percent.

We should be able to obtain a more reliable result with a larger and more homogeneous group of countries, so let us try those of Latin America shown in Table 20.1. The regression is $P_{65+} = 0.0846 - 1.63r$. Apparently the more homogeneous group gives a less steep slope than do the United States and Mexico. Now each 1 percent increase in r is associated with a

Table 20.1. Proportion aged 65 and over and rate of natural increase, 18 Latin American countries

Country	Percent aged 65 and over $100P_{65+}$	Percent rate of natural increase $100r$
Argentina, 1964	6.05	1.40
Brazil, 1950	2.45	2.80
Chile, 1967	4.47	1.89
Colombia, 1964	3.00	2.85
Costa Rica, 1966	3.18	3.44
Dominican Republic, 1966	3.57	2.85
Ecuador, 1965	3.16	3.25
El Salvador, 1961	3.18	3.81
Guatemala, 1964	2.77	2.89
Honduras, 1966	1.76	3.55
Martinique, 1963	4.96	2.50
Mexico, 1966	3.31	3.44
Nicaragua, 1965	2.90	3.57
Panama, 1966	3.57	3.29
Peru, 1963	3.42	2.83
Puerto Rico, 1965	5.77	2.36
Uruguay, 1963	7.81	1.03
Venezuela, 1965	2.99	3.65

drop of 1.6 in P_{65+}—only two-thirds as much. A scatter diagram (Figure 20.2) shows that we could have chosen two countries that would provide almost any given slope; if we wanted to show a positive rather than a negative relation, we could have taken Mexico and Brazil, or else Panama and Guatemala. Moreover, much of the overall negative correlation is due to three countries of the southern cap—Argentina, Uruguay, and Chile—that are culturally distinct from those farther north, along with Puerto Rico and Martinique. To exaggerate a little, it looks as though countries fall into two groups, those with low r and high P_{65+}, and those with high r and low P_{65+}. In short, much of the pertinent information was contained in the comparison of the United States and Mexico with which we started.

What about taking one country and following changes through time in the two variables? Sweden provides information over nearly 200 years, and also yields a very different regression from any obtained cross-sectionally.

The comparisons and regressions summarizing them are highly inconsistent in reporting how much difference in the proportion over 65 is to be associated with differences in the rate of increase. A large research project could be undertaken to see why they fail to agree; it might reveal that the changing mortality over 200 years in Sweden is confounded by the changing birth rate, or that the more homogeneous the group, the lower the correlation and the lower the slope of regression. It happens that in this instance no one will undertake such research because a simple theory is available

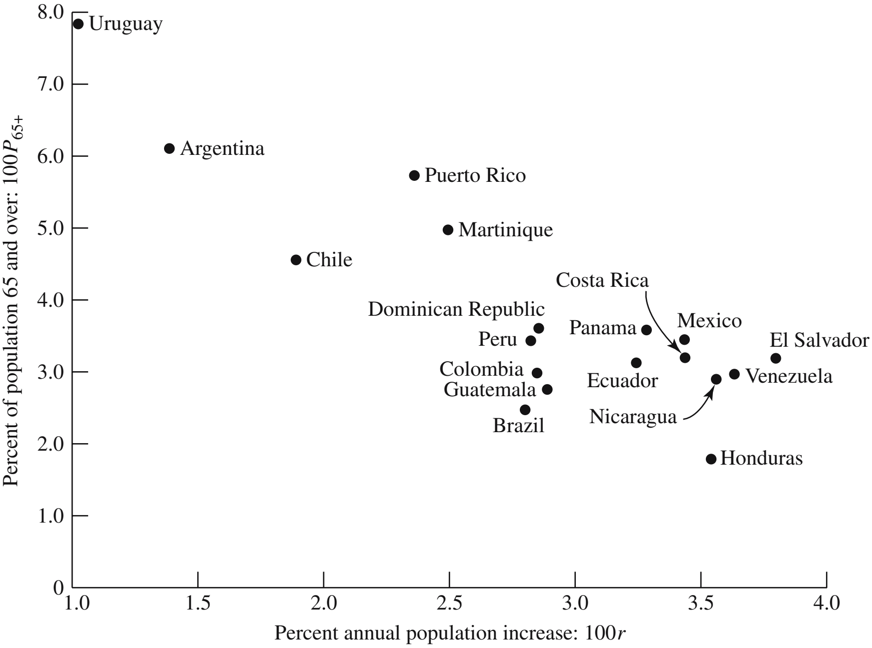


Figure 20.2. Relation of proportion of the population over age 65 to the rate of population increase: 18 Latin American countries.

that provides a better insight into the nature of the relationship between growth rate and age distribution. Let us use this theory to back off and take a fresh approach to the question.

20.1.1 Older Population as a Function of Rate of Increase When All Else Is Constant

The approach is that of Chapter 5. Suppose that the fraction of births surviving to age x is given by a fixed survival function $l(x)$, and the annual rate of increase in births is r, so that compared with x years ago the number of births is now e^{rx} greater. Then for each present birth there were e^{-rx} births x years ago, and of these a fraction $l(x)$ have survived, the survivors being now aged x. Thus the number of present persons of age x must be $Be^{-rx}l(x)$, where B is the number of current births. This applies for all ages and suffices to specify the age distribution, as we saw in Chapter 5.

The expression for persons aged x will tell the relation between a fraction of the age distribution and the rate of increase r:

$$P_{65+} = \frac{\displaystyle\int_{65}^{\omega} e^{-rx} l(x)\, dx}{\displaystyle\int_{0}^{\omega} e^{-rx} l(x)\, dx}. \tag{20.1.1}$$

Table 20.2. Fraction over 65, P_{65+}, tabulated as a function of rate of increase r, and first and second approximations

r (1)	P_{65+} (2)	$0.1379e^{-40r}$ (3)	$0.1379e^{-40r-135r^2}$ (4)
0.000	0.1379	0.1379	0.1379
0.005	0.1140	0.1129	0.1125
0.010	0.0932	0.0924	0.0912
0.015	0.0753	0.0757	0.0734
0.020	0.0602	0.0620	0.0587
0.025	0.0477	0.0507	0.0466
0.030	0.0374	0.0415	0.0368
0.035	0.0291	0.0340	0.0288
0.040	0.0225	0.0278	0.0224

Based on mortality of Coale and Demeny (1966, p. 20) Model West Females, $\overset{o}{e}_0 = 65$

If $l(x)$ is fixed, (20.1.1) establishes P_{65+} as a function of r and of nothing else. The equation is not very instructive in this form, for we cannot easily see whether P_{65+} increases or decreases with r, let alone by how much.

Taking the logarithm and expanding the exponential provides a series of cumulants that are a good approximation to the full expression 20.1.1. Up to the first two terms we have, by the method of Chapter 5,

$$P_{65+} = \frac{\int_{65}^{\omega} l(x)\,dx}{\int_{0}^{\omega} l(x)\,dx} \exp\left[-(m_2 - m_1)r - (\sigma_1^2 - \sigma_2^2)\frac{r^2}{2}\right], \tag{20.1.2}$$

where m_2 is the mean and σ_2^2 the variance of those 65 and over, and m_1 the mean and σ_1^2 the variance of all, in the stationary population. With $m_2 - m_1 = 40$, and disregarding the square term, we have column 3 of Table 20.2; including the square term with $\sigma_1^2 = 400$ and $\sigma_2^2 = 130$, so that $(\sigma_1^2 - \sigma_2^2)/2 = 135$, gives column 4, which is a good fit, and it is based on little more data than a demographer carries in his head.

Such a theoretical relation largely escapes defects of the data. Another advantage of the theoretical approach is that we know exactly its assumptions. In this instance, our model specifies that the comparison be among populations closed to migration, with the same life table but different rates of increase, and that each of them have had births increasing exponentially during the lifetimes of persons now alive or, alternatively, have had fixed age-specific birth and death rates over a long past period. Consequently, this model does not tell anything about the change through time from one such condition to another; the trajectory from rapid increase to stationarity for a given population requires a more difficult kind of mathematics. That the theory here, like the comparative statics of economics, permits the comparison of stable conditions only is both a strength and a weakness.

Instead of supposing fixed rates in a closed population, the empirical regression takes into account migration, in whatever proportion it has been occurring in the populations whose data are included. Insofar as mortality has been falling, the influence of that fall is also incorporated. Thus it is a better description of the state of affairs covered by the data; it is a worse statement of the intrinsic relationship between the stated variables. If underlying conditions are the same in the future, the regression will predict more accurately; if they change substantially, the theory will be more dependable. If an underlying interference is caused by some known and measurable variable, the empirical regression can partial it out, and in this degree approach closer to theory.

In another aspect the regression inevitably depends on a data base, and that base is determined by what data are available. One can hardly apply sampling notions to it, since whether one takes countries or some other unit, the number of existing populations that are truly independent is small and predetermined for the observer, beyond any possibility of his arranging a randomized design. Moreover, data on many countries are lacking. Even if each entity describable as a nation could be thought of as providing independent evidence, and if all had good data, the collection of nations is not easy to conceptualize as a homogeneous universe.

This simple example shows how uncertain our knowledge would be if analytical tools like the stable model were not available. One can imagine methodological controversies, and schools of opinion, some perhaps taking the view that the relationships were really different for different races or different continents. One who has been through the theory would no sooner say that the underlying relationship between increase and age composition is different for different continents than he would say that the laws of thermodynamics so differ.

20.1.2 Are Births or Deaths Decisive?

The same stable model can help one to decide whether the age distribution of a population depends more on its births or on its deaths.

Venezuela in 1965 had a greater proportion of children plus old people than did Sweden in 1803–07. To compare a contemporary nonindustrialized country with one in the early nineteenth century reveals an aspect of the difference in the process of getting development launched then and now. A high dependency ratio (children under age 15 plus adults over 65 as a proportion of the number of working ages 15 to 65) is a disadvantage for development: Venezuela's dependency ratio in 1965 of 1.021 was 73 percent greater than Sweden's in 1803–07 of 0.589. One would like to know to what extent this was due to Venezuela's lower death rate and to what extent to its higher birth rate. No such decomposition is even conceivable, however, on the observed rates—they show what they show.

The stable model of Chapter 5, in which the number of persons aged x is proportional to $l(x)e^{-rx}$, allows one to synthesize dependency ratios from various combinations of birth and death rates:

Venezuelan births and Venezuelan deaths	1.021
Swedish births and Venezuelan deaths	0.703
Swedish births and Swedish deaths	0.589

as shown in Table 10.2. The effect of the birth rate when the death rate is constant is $1.021 - 0.703 = 0.318$; the effect of the death rate when the birth rate is constant is $0.703 - 0.589 = 0.114$; of the total difference of 0.432, the part due to births is about 74 percent, and that due to deaths about 26 percent. The LTRE approach in Section 13.4 extends this kind of calculation to individual age- or stage-specific vital rates.

We could alternatively have used as the intermediate term in the decomposition the dependency ratio with Swedish deaths and Venezuelan births, which is 0.856. The death effect would have been $1.021 - 0.856 = 0.165$, and the birth effect $0.856 - 0.589 = 0.267$. Now 62 percent of the difference is due to births, still the larger part. We can say that between 62 and 74 percent is due to births, the interval between these numbers being an interaction that cannot be allocated.

Any other feature of age can be similarly analyzed. Sweden's percentage under age 15 was 31.3, and Venezuela's 47.7; the combination of Swedish births and Venezuelan deaths would produce 34.2 percent. Hence, of the difference of 16.4 percentage points obtained as $47.7 - 31.3$ only 2.9 ($= 34.2 - 31.3$) is due to deaths and 13.5—over 4 times as much—to births.

20.2 Promotion in Organizations

Everyone knows that in a fast-growing organization promotion is likely to be faster than in one growing slowly. Neither elaborate empirical data nor a model is required to demonstrate that bare fact. What one would like to know is the quantitative relation: in a fast-growing organization does one advance to a middle position a few months sooner than in a slow-growing one, or several years sooner?

We can imagine collecting a great volume of data to settle this point. The survey organizer would have to give attention to the universe of organizations from which he proposed to sample—perhaps he would settle on all commercial, transport, and manufacturing firms in the United States. He would have to define the boundaries of each organization, whether it included all establishments constituting a firm or whether each establishment was to be considered a separate organization. A lower limit would have to be set on the size of organization considered, say 100 employees. One would want to distinguish family-run enterprises, since the conditions for promotion in these would certainly be different. If a one-time survey were to be made, the information on promotion would have to be obtained retrospectively, with the errors of recollection that this entails. On the other hand, a succession of surveys that statistically followed careers of individuals would take time and be expensive. Many decisions would have to be made to establish the universe and to conduct the sampling operation within it.

Moreover, when the results were in we would notice that in some organizations there were many resignations, so that promotion was rapid for personnel that remained; indeed, this effect might be strong enough to hide the effect of growth. We would have to classify organizations into homogeneous groups according to their turnover, or else obtain an index of turnover for each and use regression analysis to partial it out. This is only one of many disturbing elements that could be expected to make the results, so painstakingly obtained, uncertain in interpretation in relation to the question to which an answer is being sought.

A simpler approach that avoids the errors to which a survey is subject is to compare the number of employees ahead of a representative individual—let us call him Ego—in a fast-growing and in a slow-growing organization, as if promotion depended only on age. Superimposed on individual ability, assiduity, influence, luck, and all the other elements that determine promotion in the real world is the pure effect of organizational growth on individual careers, and that is what we want to ascertain. That effect can be studied by use of a suitable model as we saw in Section 5.8.

First suppose a given schedule of survival—knowing that the deaths of his contemporaries help Ego's promotion, we do not want differences in mortality to cloud our analysis. Then suppose an age distribution that is a function only of this survival function and rate of increase, so that the stable model is applicable. Finally, take as the arbitrary benchmark for measurement the age at which individuals arrive at a level where one-half of their fellow-workers are below them and one-half above, say a junior supervisory position.

After that, the simple mathematics of Section 5.8 shows that, for given rates of death or resignation, the age x at which Ego reaches such a position is shortened by two-thirds of a product of three factors:

1. The time from age x to retirement, discounted at the rate of population increase;
2. The difference between the mean age of the group senior to the point of promotion considered, and the mean age of those junior to it. This difference cannot be far from half of the length of working life;
3. Increase of the rate of increase of the population.

With an entry age of 20 and a retirement age of 65, comparison of two populations whose increase differs by Δr percent gives for the difference in ages

$$\Delta x = -\left(\frac{2}{3}\right)(15)(22.5)\Delta r = -225r.$$

Thus the time of promotion is delayed by 2.25 years for each 1 percent by which population growth is lower. This demographic factor is overlaid on all individual differences of ability, influence, and luck. Although the model is based on pure seniority, some such effect will apply if any element of seniority is present. Only if length of service in the organization is wholly disregarded in promotion will the model be irrelevant.

20.3 No Model, No Understanding

A large volume of data is on hand regarding breast cancer. Despite stepped-up efforts to deal with it, expensive operations and other forms of treatment, and widespread publicity urging women to examine themselves and to see their doctors at once if there is any suspicion, the increase in deaths from breast cancer is considerable in North America and Western Europe, just where the most intensive therapeutic effort is being made. In the 1980s, breast cancer was the leading cause of death for women aged 35 to 54 and was second only to heart disease for older ages. Some of the increase may have been due to more awareness and hence more frequent diagnosis than in the past, and to better diagnosis in America and Europe than in Asia and Latin America, but apparently this was not the whole cause. Women who bear children early seem to have a lower risk of breast cancer, but no one thinks that having children—early or late—can prevent the disease or account for the differences. Breast cancer is less common in warm climates and among poor populations, but that warm climate or poverty is a preventive seems unlikely.

Such statistical differentials will continue to be unsolved puzzles until someone comes along with a model that explains the differences. In the meantime, all that can be done is to continue gathering data to discriminate among proposed models.

The ratio of male to female births is a similar case, in that there is no obvious model, and no clear-cut result has so far emerged from differentials and correlations. We know that births to young mothers have a higher sex ratio (males to females) than births to older ones, that first births to mothers have a higher sex ratio than later births, and that children of young fathers have a higher sex ratio than children of older fathers. But among age of mother, parity of mother, and age of father, which is the effective cause? The high intercorrelations among the possible causes make it difficult to distinguish. Mechanisms have been suggested involving the relative activity and viability of sperm producing male and female babies but until some such mechanism is shown to be the operative one, our knowledge is tentative and uncertain.

20.4 Too Many Models

India and some other countries have raised the legal age of marriage partly with the aim of lowering the birth rate. Implicit in the thinking of legislators and others is a theory in which marital fertility age for age is relatively fixed, and a legal minimum age effectively eliminates the part of the fertility curve below the age specified. Given the curve, the amount of effect is easily calculated. Sample survey data for India show that, of 18.14 million births in 1961, some 3.24 million or 18 percent were to mothers under age 20 years (Keyfitz and Flieger 1968, p. 659).

This seems a potentially powerful argument for restriction of marriage supposing it feasible to raise the age as high as 20 for women. But before one reaches a firm conclusion, it ought to be noted that on an opposite model raising age at marriage would be wholly ineffective.

Suppose that married couples are not reproducing to the maximum, but that they want a certain number of children, and will have later what the law forbids them to have sooner. After all, the birth rate of India (under 40 per thousand) is well below the physiologically possible maximum. Under these circumstances, the only gain from a legal minimum age for marriage would be a delay of a few years that would lengthen the time separating successive generations and hence lower the average annual increase, but by a small amount. Illegitimacy is also a problem; it is low in India, but one of the reasons parents want their daughters married early is to avoid their engaging in premarital sex. If the parents' fears are not altogether imaginary, there could be some increase in the number of children born outside of marriage.

Yet these arguments are in the end unconvincing; one has the impression that couples that lose time before they are 20 may make up some part of the lost ground but not all, and that extramarital fertility would remain low. To determine the net drop in overall fertility as a result of the restriction, behavioral data would be required. These alone can discriminate between the competing models and predict the quantitative effects of an induced change in age at marriage.

20.5 Development and Population Increase

This brief chapter is not the place to consider intricate issues of population economics, which is an entire academic discipline having scores of specialists, a literature running to many hundreds of articles and books, and its own lines of cleavage and of controversy. It is worth saying here only enough to show that both theoretical and empirical methods are applied in this field, and that, notwithstanding their extensive and skilled use, much remains to be done in disentangling the lines of causation. The literature speaks of "development" as the socioeconomic transformation into the modern condition, and of "income" as sufficiently correlated with development to be used as a proxy.

20.5.1 The Effect of Development on Population

First, in regard to the effect of development on population, a quick look at cases suggests a familiar negative relationship, with which theory conforms. Development seems sooner or later to have brought a reduction in population growth in all the instances where it has occurred. All of the rich countries have low birth rates today, and the very richest are not replacing themselves. For example, West Germany had fewer births than deaths in 1973, and in 1974 it had fewer births plus net immigrants than deaths, so that its population actually declined. But the countries of Eastern Europe are much less rich, and they also have low birth rates, while in Britain the birth rate first started to fall almost a century after development was under way. Thus the correlation is not perfect, but still history seems to be saying that, with more or less lag, industrialization has led to reduced family size.

In theory this may be due to women finding jobs and sources of prestige outside the home, so that they do not need to rely on childbearing for their standing, and to children being more expensive and, on the other hand, less directly useful to their parents as income increases, both effects being related to the decline of the family as a productive unit with the growth of industry. As contraception is made easier and more available, relatively weak motivation suffices to cut the birth rate. What we ought to believe in this matter, summed up in the concept of demographic transition, is relatively unambiguous because the dominant theories and the most conspicuous anecdotal evidence all point the same way.

Yet even here, the more closely and systematically scholars have looked at the data, the less clear they have found the effect of development on family size. Taking income as a proxy for development, Adelman (1963) made "an analysis of fertility and mortality patterns as they are affected by economic and social forces." Her materials, based mostly on national statistics for 1953, showed a decidedly *positive* relation between income and fertility. Friedlander and Silver (1967) partialed out more variables and found that for developed nations the relation is negative. Heer (1966) calculated correlations for 41 countries that suggested that the direct effect of economic development is to increase fertility, and the indirect effects (through education, and so on) are to reduce it. But it makes a difference when the data for the 24 less-developed countries are separated from those for the 17 more developed and more than one point of time is introduced so that changes rather than levels are correlated. Ekanem (1972) used two points of time, the 1950 decade and the 1960 decade, but the effect of his greater care seems to be a less clear-cut result than Heer's. Again, Janowitz (1971) followed five European countries and found that variables shifted enough through time so that the longitudinal relations, more likely to indicate causation, are decidedly different from the cross-sectional regressions.

It would be unkind to say that these efforts constitute raw empiricism. They are oriented by an economic theory: that increased affluence causes people to buy more of most things, the exceptions being labeled inferior goods. Since no one considers children inferior goods, many argue that children and income "really" are positively related, but the relation is concealed by the intervention of other factors. The better off have access to contraceptives of which the poor are ignorant; the better off have higher quality (i.e., more expensive) children and so can afford fewer of them (Becker 1960, Leibenstein 1974).

20.5.2 The Effect of Population Growth on Development

The writers cited above were trying to find the impact of development on fertility, a situation where, despite some complications and contradictions causation seems clearer than in the inverse problem: in which direction and to what extent does rapid population growth affect development. Among all the questions that demographers seek to answer, this last is the one that is truly important for policy.

In the classic theory rapid growth means many children—40 percent or more of the population under age 15. The children have to be fed, clothed, and educated, and however the cost is divided between parents and the state, it requires resources that compete with industrial and other investments. In addition, children grow up and want jobs, and hence growth requires that a larger labor force be provided with capital goods. Thus a fast-growing population is doubly handicapped.

So much for the static aspect of the demographic–economic relationship. In regard to dynamics, when fertility falls from an initially high level, the dependency ratio begins to shift immediately in an economically favorable direction. Thus investment can be greater than it was before. Lagging 15 or 20 years behind is a longer-run dynamic effect: a slackening of the growth of the population in the labor force ages. When relatively fewer children grow up to enter these ages, there is less competition for productive jobs and each entrant may have more capital to work with than if the birth rate had not been cut (Coale and Hoover 1958).

All this is based on the view that development is capital-limited rather than resource-limited. But if it is resource-limited, population is an even more serious consideration, although now the absolute level of population is of concern rather than the birth rate; the more people, the less resources at the disposal of each, on a theory dating back to John Stuart Mill and ultimately to Malthus. In the most general statement, certain ratios of labor to the other factors of production—land and capital—are more favorable than others, and most developing countries are moving away from the optimum with present population sizes and birth rates.

How would matters stand if no theory had ever been presented? Let us try to wipe theory out of our minds, and look at the data with complete naiveté. Among developing countries, Pakistan is increasing at over 3 percent and India at less than 2.5 percent, yet Pakistan seems to be making more economic progress. Iran's rate of population increase is much greater than Nepal's, and so is its economic advance. Brazil and Venezuela are not increasing in population less rapidly than their economically stagnant neighbors; indeed, Argentina and Chile, with very low birth rates, may be becoming poorer absolutely. Mexico is advancing economically with an annual population increase of 3.5 percent per year, one of the highest in the world and higher than that in Paraguay or Bolivia, where economic dynamism is absent. On the other hand, sub-Saharan Africa has high rates of population increase and low income growth. Figure 20.1 depicts the broad array of relationships between population growth and increase in income for large countries in the contemporary world. As noted at the start of this chapter, the relation that theory predicts is not at all evident.

It makes a difference if we compare birth rates rather than natural increase, and for the theory, births less infant deaths might be the best indicator of the economic impact. But whatever measure is used, the inverse correlation with economic dynamism simply does not appear.

Of course individual countries can be analyzed, and by making allowance for such nonpopulation aspects as leadership, political conditions, the educational system, religion, and the dissolving of patrimonial social relations as expressed in landholding and other ways, along with resource endowment, we need not be at a loss to account for the observed national differences. This explanation a posteriori can be made to sustain the theory, but hardly answers a disturbing question: to what extent would naive examination of population and income data for the poor countries of the world have revealed any clear effect of population on development? Would the effect have been as blurred as the effect of population increase on age distribution?

It is just this incapacity of the raw data to speak for themselves that permits some to argue that population and its growth do not harm development and should be allowed to take care of themselves. One might expect the facts to silence anyone who could utter such opinions, but as presented either anecodotally as above or in simple correlations they do not. To make the facts speak clearly to this issue is not easy.*

---

*Since these words were written, the problem has become no less difficult, but even more important, as demographers and ecologists explore the inter-relationships of the environment, development, and population. Keyfitz (1993, 1994, 1996) discusses the issues and their history.

20.6 How Nature Covers Her Tracks

The reason for bringing these very difficult matters into the present exposition is the hope that their investigation can be aided by going back to some simpler issues, like the relation between age distribution and the rate of increase of a population. In that case most would agree that theory gives the right answer: the rate of increase determines the proportion of old persons (as well as middle-aged and young) in the population. Where the relationship is obscured by migration or by changing birth and death rates, as it commonly is, these are seen as mere disturbances. Such "noise" could drown out the relationship in the observed data without weakening our conviction that the relationship is "really" as stable theory says it is. Up to this point stable theory has the immutability of the laws of logic: if over a sufficient period of time death rates are the same in two populations, the one with the higher birth rate will have the lower proportion at older ages. Belief in this is unshaken by the fact that El Salvador is higher than Honduras both in rate of population increase and in percentage over age 65. A supporter of the theory would convincingly argue that the official data must be wrong (perhaps registration of births is differentially incomplete), or there has been age-selective migration, or some other reason underlies the discrepancy between expected and observed relationships.

Although stable theory can never be disproved, it could be deprived of all interest if in the real world certain things that it assumes constant were in fact steadily changing. If death rates were always falling at a certain pace, the proportion of old people would everywhere be different from that given by stable theory, and a different theory would be required for interpreting reality. Any steady change that was universal would make us want to replace stable theory with its fixed rates by some other, inevitably more complicated, theory that would have equal force of logic but be more applicable. In fact, change is not so uniform under different real conditions, but is rather erratic, a means by which nature covers up her mechanisms, rendering their interpretation less amenable to a universal theory.

But change, whether steady or erratic, is not the means by which the mechanisms of nature are most effectively concealed. More deceiving is the clinging together of variables. Suppose all countries of natural growth were countries of emigration, so that they lost their young people to countries of slower growth; then the conclusion derived from the application of stable theory would be downright misleading. We would want some other theory, perhaps one on which populations tend to spread out evenly in relation to resources. Such a view is held regarding internal migration, where free movement occurs and people go to distant places unless they are attracted to intervening opportunities (Stouffer 1940).

20.6.1 The Oblique Use of Data to Challenge Theory

How then can the classical theory that rapid population growth checks development be challenged? The matter is important because a theory that is immune to challenge has little value for science.

One way is by declaring that there is a trend toward development everywhere in the world, as well as a trend toward smaller families, and that the latter makes no difference to the former. Suppose that the trend to development occurs everywhere sooner or later and nothing can either stop it or hasten it. On this comfortable view of development as immanent in human history no detailed causal theory would be possible, and no policy measures would be sought or needed. Such a view is not entirely absent from contemporary discussion, although by its very nature little evidence can be summoned for or against it.

Another direction of attack is to adduce evidence that enterprising personalities are more often born into large families and to show quantitatively that this greater enterprise is sufficient to overcome the capital and land shortage caused by the large family's propensity to consume. Or else that couples with more children have a greater incentive to save and so increase investment funds. Or else that having many children indeed increases consumption but fathers of large families work correspondingly harder to offset this. All of these are statements on the individual level that there is a sticking together of the variables concerned with development—population growth, motivation to work, and motivation to save. Nothing in logic proves that such sticking together does not occur, but it is the obligation of anyone who challenges the theory to adduce evidence.

On the national level, the countries that are developing may be the ones in which the authorities are development minded and persuade their people to make sacrifices that more than offset the disadvantages of population increase. Again, evidence bearing on this specific point would be required.

To take an example that, alas, may not be entirely unrealistic, if dictatorial technocratic regimes are effective in producing development, and if these are lukewarm about population control, the population effect might be dominated by the dictator effect. But one would give up the classical theory only if there were shown to be some necessary relationship between technocratic dictators and development, on the one hand, and dictators and large families, on the other. Otherwise one would still have to insist that the dictator was paying a price for population growth, and the price could be avoided.

Why, then, does the failure of a correlation-type approach to show that development follows a slowing of population growth present no challenge to the theory? Certainly a purely cross-sectional analysis can at best be suggestive of propositions concerning longitudinal changes, and prove nothing concerning them. Aside from this, the major difficulty is that many other factors affect the correlations. In principle, the disturbing factor of "motivation to work" or "making sacrifices" could be partialed out or held constant while the relation of population to development was examined. Yet, even if one or two disturbing factors could be identified and measured, many others would remain. And to partial out a number of variables simultaneously raises logical difficulties if any of them is correlated with the independent variable of interest.

Nonetheless, empirical data have to be applied to verify theory, and such application is the heart of demography, as of any other science, but data must be brought to bear in an oblique fashion. If we are going to detect nature's hidden mechanisms, we need a subtlety that approaches hers. An attack by directly correlating the variables of immediate interest is less promising than a search to determine what other subsidiary variables stick to them. This applies equally to the analysis of age distribution, the effects of age at marriage on the birth rate, and population and development.

20.7 The Psychology of Research

A footnote concerning the mental conditions in which research occurs may help to illuminate the way we get to know the facts of demography. Faced with a variety of data, the investigator listlessly surveys them, in the hope of somehow tying them together. He or she is swamped by the multiplicity of observations and tries to fit them into a scheme, if only to economize his own limited memory. He becomes more animated when he sees that some general connections do subsist in the data, and that a model, however crude, helps him to keep their relationships in mind. The model is much more than a mnemonic device, however; it is a machine with causal linkages. Insofar as it reflects the real world, it suggests how levers can be moved to alter direction in accord with policy requirements. The question is always how closely this constructed machine resembles the one operated by nature. As the investigator concentrates on its degree of realism, he more and more persuades himself that his model is a theory of how the world operates.

But now he is frustrated—he has just turned up an incontrovertible observation that is wholly inconsistent with his theory. Such an observation is a stubborn fact, an exception to the theory that cannot be avoided or disregarded. A struggle ensues as the investigator attempts to force the theory to embrace the exception. As his efforts prove vain, he questions the theory, and looks back again at the raw data whose complexity he thought he had put behind him. The intensity of the struggle that follows is one of the hallmarks of scientific activity and distinguishes it from mindless collecting of data, on the one hand, and complacent theorizing on the other.

The problem and its possible solutions have now taken possession of the investigator. In this phase of research his unconscious is enmeshed and is working on the question day and night. Sleep is difficult or impossible; eating and the daily round of life are petty diversions. He is irritable and distracted. Whatever he does, the contradiction he has uncovered comes into his mind, and stands between him and any normal kind of life.

During the struggle the investigator is like a person with high fever. Then with luck he comes on the answer, or his unconscious does. He finds a model that fits, perhaps nearly perfectly, perhaps only tolerably, but well enough to provide a handle on the varied data. His tension relaxes, and he proceeds with the normal and dull work of establishing the details of the fit and presenting his results. He must indeed revert to a calmer state before he can hope to communicate his finding to an audience that is perfectly normal. An immediate test of his result will be whether it makes sense to his contemporaries; an ultimate test is whether it can predict outcomes involving data not taken into account in the establishment of the model.

Only with exceptional good fortune will one cycle of feverish concern produce a final theory and permit immediate relaxation. More often a long series of false starts and disappointments will precede the resolution. Sometimes the problem turns out to be unsolvable in the existing state of knowledge or is beyond the capacity of the investigator, and then he has the unhappy task of winding himself down without the desired denouement.

None of the psychological accompaniment of scientific production is special to demography, but this field may show it in heightened form, at least in comparison with other social sciences. The abundant data of demography cause an inappropriate theory or an erroneous prediction to stand out more clearly than a corresponding failure in interpreting history or in the general analysis of society. When the possibility of a sharp rejection by hard data is lacking, the game of research loses its seriousness—it is like playing solitaire with rules that are adjustable to the cards that appear.

Mathematical demography (or, as it is sometimes called, formal demography) has become too big a field to be covered in one book, or even to be effectively summarized. Many important themes are omitted altogether from the foregoing, and many others are treated much too summarily. More extensive theory is available than has been included, and (a much graver omission) only a minute fraction of the data bearing on the issues raised has been brought under examination. What points are made in the book had best be considered as examples of the ideas and methods of the subject. A full and systematic account remains to be written.

Formal demography provides the core of the discipline, the framework within which detailed studies of micro-level individual phenomena can be interpreted (Lee 2001, Goldstein 2001). This is equally true in human and nonhuman demography, in both of which the link from the individual to the population is critical, and is provided by formal demographic models (Tuljapurkar and Caswell 1997, Lee 2002, MPM). The vital rates that parameterize all the models in this book are, in the end, properties of individuals. But the population consequences of those rates become apparent only when they are linked together in a model that describes stocks and flows of those individuals. The ability to do that is the power of applied mathematical demography.

Some of the directions that demography will take in the near future seem clear. As more and more types of individual data become available, models that can classify individuals freely by criteria in addition to age will become ever more important. In this book we have approached this topic with methods, rooted in the work of Leslie (1945), that lead to population projection matrices. Other approaches are also the subject of current research, in both human and non-human demography (e.g., Tuljapurkar and Caswell 1997, Thieme 2003). Heterogeneity (Chapter 19) will continue to attract attention. Part of it can be dealt with by incorporating it into *i*-state variables, but there is always some left unresolved. The effects of the latter, on parameter estimation and both short- and long-term dynamics, are still important problems.

Of the models in Figure 3.2, perhaps none has seen as much important theoretical development in recent years as those including environmental stochasticity (Tuljapurkar 1990, MPM Chapter 14). Incorporating stochasticity into population projections, so that the uncertainty of the projection can be treated as part of the results, rather than as obscuring them, seems likely to continue to attract attention (e.g., Lee and Tuljapurkar 1994, Bongaarts and Bulatao 2000).

Nonlinear effects, in which the vital rates depend on the current state of the population, lead to exciting mathematical behavior (Cushing 1998, MPM Chapter 16), but have always been difficult to apply because such models require following the influence of the current population through often complicated causal pathways back to the vital rates. In human populations, where these pathways include social and economic systems that are themselves extraordinarily complicated (Lee 1987, Keyfitz 1996), the difficulties will be even greater, and will require both new data and new models. But the link between demography and the environment is of critical importance, from the global scale (Cohen 1995, Bongaarts and Bulatao 2000) on down to the local scale.

The links between human demography and biology seem likely to become closer and intellectually richer. Studies of mortality, lifespan, and fertility are already benefiting from this link, as are studies of epidemiological demography, a topic we have not considered here.

The tools of mathematical demography are essential for the study of populations—of humans and of other species alike—and we hope that this book will contribute to that study.

Bibliography

Adelman, I. 1963. An econometric analysis of population growth. American Economic Review 53:314–339.

Agnew, R.A. 1972. Crude confidence interval estimates for future U.S. population levels. Letter in The American Statistician, February, pp. 47–48.

Al Mamun, A. 2003. Life History of Cardiovascular Disease and its Risk Factors. Rozenberg Publishers, Amsterdam, The Netherlands.

Alonso, W. 1972. Problems, purposes, and implicit policies for a national strategy of urbanization. In S.M. Mazie, editor, Population, Distribution, and Policy, pp. 635–647. U.S. Commission on Population Growth and the American Future. U.S. Government Printing Office, Washington DC.

Alonso, W. 1978. A theory of movements. pp. 197–211 in N.M. Hansen, editor. Human Settlement Systems: International Perspectives on Structure, Change and Public policy. Ballinger, Cambridge, Massachusetts, USA.

Altmann, J. and S.C. Alberts. 2003. Intraspecific variability in fertility and offspring survival in a nonhuman primate: Behavioral control of ecological and social sources. Pages 140–169 in K.W. Wachter and R.A. Bulatao, editors. Offspring: Human Fertility Behavior in Biodemographic Perspective. National Academies Press, Washington, DC, USA.

Anderson, H. 1971. The sample period in economic prediction, The American Statistician 25:46–50.

Anderson, R.M. and R.M. May. 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford, United Kingdom.

Armstrong D.P. and I.G. McLean. 1995. New Zealand translocations: Theory and practice. Pacific Conservation Biology 2:39–54.

Arriaga, E.A. 1968. New Life Tables for Latin American Populations in the Nineteenth and Twentieth Centuries. Population Monograph Series No. 3. Berkeley, California, Institute of International Studies.

Arthur, W.B., and J.W. Vaupel. 1984. Some general relationships in population dynamics. Population Index 50:214–226.

Barberi, B. 1972. Elementi di Statistica Economica. Torino: Boringhieri.

Barlow, J. and P.J. Clapham. 1997. A new birth-interval approach to estimating demographic parameters of humpback whales. Ecology 78:535–546.

Barlow, R.E., and F. Proschan. 1964. Mathematical Theory of Reliability. Wiley, New York, New York, USA.

Barlow, R.E. and J.L.M. Saboia. 1973. The use of reliability notions to make predictions and establish bounds for population growth. In Reliability and Biometry: Statistical Analysis of Lifelength, edited by F. Proschan and R.J. Serfling. SIAM, Philadelphia, Pennsylvania, USA.

Barrett, J.D., and J. Marshall. 1969. The risk of conception on different days of the menstrual cycle. Population Studies 23:455–461.

Bartholomew, D.J. 1982. Stochastic Models for Social Processes. 3rd ed. Wiley, New York, New York, USA.

Bartlett, M.S. 1960. Stochastic Population Models in Ecology and Epidemiology. Methuen, London, UK.

Bartlett, M.S. 1962. An Introduction to Stochastic Processes, with Special Reference to Methods and Applications. Cambridge University Press, Cambridge, UK.

Beckenbach, E., and R. Bellman. 1961. An Introduction to Inequalities. Random House, New York, New York, USA.

Becker, G.S. 1960. An economic analysis of fertility. In Demographic and Economic Change in Developed Countries, edited by A.J. Coale. Princeton University Press, Princeton, New Jersey, USA.

Becker, G.S. 1991. A Treatise on the Family. 2nd ed. Harvard University Press, Cambridge, Massachusetts, USA.

Berman, A. and R.J. Plemmons. 1994. Nonnegative Matrices in the Mathematical Sciences. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, U.S.A.

Bernardelli, H. 1941. Population waves. Journal of Burma Research Society 31:1–18.

Berry, B.J.L. 1973. Growth Centers in the American Urban System, Vol. 1. Ballinger, Cambridge, Massachusetts, USA.

Bickel, P.J., E.A. Hammel, and J.W. O'Connell. 1977. Sex bias in graduate admissions: Data from Berkeley. In Statistics and Public Policy. Addison-Wesley, Reading, Massachusetts, USA. (Reprinted from Science 187:398–404.)

Birch, L.C. 1948. The intrinsic rate of natural increase of an insect population. Journal of Animal Ecology 17:15–26.

Birch, L.C. 1953. Experimental background to the study of the distribution and abundance of insects. Ecology 34:698–711.

Bishop, Y.M.M., S.E. Fienberg, and P.W. Holland. 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, Massachusetts, USA.

Blumen, I., M. Kogan, and P.J. McCarthy. 1955. The Industrial Mobility of Labor as a Probability Process. Vol. VI of Cornell Studies of Industrial and Labor Relations. Cornell University Press, Ithaca, New York, New York, USA.

Blyth, C. 1972. On Simpson's paradox and the sure-thing principle. Journal of the American Statistical Association 67:364–366.

Boldrini, M. 1956. Demografia. Revised ed. Dott. A. Giuffre, Milan, Italy.

Boldrini, M. 1928. Lezioni di Demografia. Tipografia Mariani, Milan, Italy.

Boling, R.H., Jr. 1973. Toward state-space models for biological populations. Journal of Theoretical Biology 40:485–506.

Bolstad, W.M. and S.O.M. Manda. 2001. Investigation child mortality in Malawi using family and community random effects: A Bayesian Analysis. Journal of the American Statistical Association 96: 12–19.

Bongaarts, J. 1983. The formal demography of families and households: An overview. IUSSI Newsletter 17:27–42.

Bongaarts, J. and R.A. Bulatao. 2000. Beyond Six Billion: Forecasting the World's Population. National Academy Press, Washington, DC, USA.

Bongaarts, J., T. Burch, and K.W. Wachter. 1987. Family Demography: Methods and Their Applications. Clarendon Press, Oxford, UK.

Bortkiewicz, L.V. 1911. Die Sterbeziffer und der Frauenüberschuss in der stationären und in der progressiven Bevölkerung. Bulletin de l'Institut International de Statistique 19:63–183.

Boucher, D.H. 1997. General patterns of age-by-stage distributions. Journal of Ecology 85:235–240.

Boudon, R. 1977. Effets Pervers et Ordre Soci. Presses Universitaires de France, Paris, France.

Bourgeois-Pichat, J. 1958. Utilisation de la notion de population stable pour mesurer la mortalité et al fécondité des populations des pays sous-développés [Using the concept of a stable population for measuring mortality and fertility in underdeveloped countries]. Bulletin de l'Institut International de Statistique (Stockholm) 36:94–121.

Bowley, A.L. 1924. Births and population of Great Britain. Journal of the Royal Economic Society 34:188–192.

Brass, W. 1953. The derivation of fertility and reproduction rates from restricted data on reproductive histories. Population Studies 7:137–166.

Brass, W. 1963. The construction of life tables from child survivorship ratios. In International Population Conference (IUSSP) New York, 1961: Proceeding, Vol. 1, pp. 294–301. International Union for the Study of Population, London, UK.

Brass, W. 1971a. A critique of methods for estimating population growth in countries with limited data. Bulletin of the International Statistical Institute 44:397–412.

Brass, W. 1971b. On the scale of mortality. In Biological Aspects of Demography, edited by William Brass, pp. 69–110. Taylor & Francis, London, UK.

Brass, W. 1974. Perspectives in population prediction, illustrated by the statistics of England and Wales. Journal of the Royal Statistical Society, Series A: General 137:532–583.

Brass, W. 1975. Methods for Estimating Fertility and Mortality from Limited and Defective Data. An Occasional Publication. International Program of Laboratories for Population Statistics, University of North Carolina, Chapel Hill, North Carolina, USA.

Brass W. 1983. The formal demography of the family: An overview of the proximate determinants. Occasional paper 31. Office of Population Census and Surveys, London, UK.

Brass, W., and A.J. Coale. 1968. Methods of analysis and estimation. In The Demography of Tropical Africa, edited by William Brass et al., pp. 88–142. Princeton University Press, Princeton, New Jersey, USA.

Brass, W., et al., Editors, 1968. The Demography of Tropical Africa. Princeton University Press, Princeton, New Jersey, USA.

Brault, S. and H. Caswell. 1993. Pod-specific demography of killer whales (*Orcinus orca*). Ecology 74:1444–1454.

Bravo de la Parra, R., P. Auger and E. Sanchez. 1995. Aggregation methods in discrete models. Journal of Biological Systems 3:603–612.

Bravo de la Parra, R., P. Auger and E. Sanchez. 1997. Time scales in density dependent discrete models. Journal of Biological Systems 5:111–129.

Burch, T.K. 1970. Some demographic determinants of average household size: An analytic approach. Demography 7:61–70.

Burch, T.K. 1979. Household and family demography: A bibliographic essay. Population Index 43:173–195.

Burch, T.K. 1984. Changing Age-sex Roles and Household Crowding: A Theoretical Note. Population Studies Centre, University of Western Ontario, London, Ontario, Canada.

Burch, T.K., et al. 1983. Measures of Household Composition and Headship Based on Aggregate Routine Census Data. Population Studies Centre, University of Western Ontario, London, Ontario, Canada.

Burks, B.S. 1933. Statistical method for estimating the distribution of sizes of completed fraternities in a population represented by a random sampling of individuals. Journal of the American Statistical Association 28:388–394.

Burnham, K.P. and D.R. Anderson. 2002. Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach. 2nd. ed. Springer-Verlag, New York, New York, USA.

Cam, E., W.A. Link, E.G. Cooch, J-Y. Monnat, and E. Danchin. 2002. Individual covariation in life-history traits: Seeing the trees despite the forest. American Naturalist 159:96–105.

Campbell, K.L. and J.W. Wood. 1994. Human reproductive ecology: Interactions of environment, fertility, and behavior. Annals of the New York Academy of Sciences 709.

Canales, J., M.C. Trevisan, J.F. Silva, and H. Caswell. 1994. A demographic study of an annual grass *Andropogon brevifolius* (Schwarz) in burnt and unburnt savanna. Acta Oecologica 15:261–273.

Cannan, E. 1895. The probability of cessation of the growth of population in England and Wales during the next century. Economic Journal 5:505–515.

Carey, J.R. 1983. Practical application of the stable age distribution: Analysis of a tetranychid mite (*Acari: Tetranychidae*) population outbreak. Environmental Entomology 12:10–18.

Carey, J.R. 1993. Applied Demography for Biologists with Special Emphasis on Insects. Oxford University Press, Oxford, UK.

Carey, J.R. 2003. Longevity: The Biology and Demography of Life Span. Princeton University Press, Princeton, New Jersey, USA.

Carey, J.R. and S. Tuljapurkar. 2003. Life span: Evolutionary, ecological, and demographic perspectives. Population and Development Review, Supplement to Volume 23.

Caswell, H. 1978. A general formula for the sensitivity of population growth rate to changes in life history parameters. Theoretical Population Biology 14:215–230.

Caswell, H. 1980. On the equivalence of maximizing reproductive value and maximizing fitness. Ecology 61:19–24.

Caswell, H. 1982a. Stable population structure and reproductive value for populations with complex life cycles. Ecology 63:1223–1231.

Caswell, H. 1982b. Optimal life histories and the maximization of reproductive value: A general theorem for complex life cycles. Ecology 63:1218–1222.

Caswell, H. 1982c. Life history theory and the equilibrium status of populations. American Naturalist 120:317–339.

Caswell, H. 1985. The evolutionary demography of clonal reproduction. Pages 187–224 in J. B. C. Jackson, L.W. Buss, and R.E. Cook, editors. Population Biology and Evolution of Clonal Organisms. Yale University Press, New Haven, Connecticut, USA.

Caswell, H. 1986. Life cycle models for plants. Lectures on Mathematics in the Life Sciences 18:171–233.

Caswell, H. 1988a. Approaching size and age in matrix population models. Pages 85–105 in B. Ebenman and L. Persson, editors. Size-Structured Populations. Springer-Verlag, Berlin, Germany.

Caswell, H. 1988b. Theory and models in ecology: A different perspective. Ecological Modelling 43:33–44.

Caswell, H. 1989. The analysis of life table response experiments. I. Decomposition of effects on population growth rate. Ecological Modelling 46:221–237.

Caswell, H. 1996a. Demography meets ecotoxicology: Untangling the population level effects of toxic substances. Pages 255–292 in M.C. Newman and C.H. Jagoe , editors. Ecotoxicology: a Hierarchical Treatment. Lewis Publishers, Boca Raton, Florida.

Caswell, H. 1996b. Analysis of life table response experiments. II. Alternative parameterizations for size- and stage-structured models. Ecological Modelling 88:73–82.

Caswell, H. 1997. Methods of matrix population analysis. Pages 19–58 in Tuljapurkar, S. and H. Caswell, editors. Structured Population Models in Marine, Terrestrial and Freshwater Systems. Chapman and Hall, New York, New York, USA.

Caswell, H. 2000a. Prospective and retrospective perturbation analyses and their use in conservation biology. Ecology 81:619–627.

Caswell, H. 2000b. Life table response experiments in ecotoxicology. in J. Kammenga and R. Laskowski, editors. Demography in Ecotoxicology. Wiley, London, UK.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. 2nd ed. Sinauer, Sunderland, Massachusetts, USA.

Caswell, H. and A. Hastings. 1980. Fecundity, developmental time, and population growth rate: An analytical solution. Theoretical Population Biology 17:71–79.

Caswell, H. and A.M. John. 1992. From the individual to the population in demographic models. Pages 36–61 in D.L. DeAngelis and L.J. Gross, editors. Individual Based Models and Approaches in Ecology. Chapman and Hall, New York, New York, USA.

Caswell, H. and T. Kaye. 2001. Stochastic demography and conservation of *Lomatium bradshawii* in a dynamic fire regime. Advances in Ecological Research 32:1–51.

Caswell, H. and P.A. Werner. 1978. Transient behavior and life history analysis of teasel (*Dipsacus sylvestris* Huds.). Ecology 59:53–66.

Caswell, H., H.E. Koenig, J.A. Resh, and Q.E. Ross. 1972. An introduction to systems science for ecologists. Pages 3–78 in B. C. Patten, editor. Systems Analysis and Simulation in Ecology, Volume 2. Academic Press, New York, New York, USA.

Caswell, H., R.J. Naiman and R. Morin. 1984. Evaluating the consequences of reproduction in complex salmonid life cycles. Aquaculture 43:123–134.

Caswell, H., M. Fujiwara, and S. Brault. 1999. Declining survival probability threatens the North Atlantic right whale. Proceedings of the National Academy of Sciences, USA. 96:3308–3313.

Caswell, H., S. Brault, J.-D. Lebreton, M. Neubert, R. Sibly, T. Takada, and S. Tuljapurkar. 2000. No inconsistencies in sensitivity analysis. Trends in Ecology and Evolution 15:204.

Caughley, G. 1977. Analysis of Vertebrate Populations. Wiley, New York, New York, USA.

Chen, W.K. 1976. Applied graph theory: Graphs and electrical networks. 2nd ed. North-Holland Publishing, New York, New York, USA.

Chiang, C.K. 1960a. A stochastic study of the life table and its applications: I. Probability distributions of the biometric functions. Biometrics 16:618–635.

Chiang, C.L. 1960b. A stochastic study of the life table and its applications: II. Sample variance of the observed expectation of life and other biometric functions. Human Biology 32:221–238.

Chiang, C.L. 1961. On the probability of death from specific causes in the presence of competing risks. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 4:169–180.

Chiang, C.L. 1968. Stochastic Processes in Biostatistics. Wiley, New York, New York, USA.

Cho, L.-J. 1972. On estimating annual birth rates from census data on children. In Proceedings of the Social Statistics Section, 1971, pp. 86–96. American Statistical Association, Washington, DC, USA.

Chu, C.Y.C. 1998. Population Dynamics: A New Economic Approach. Oxford University Press, Oxford, UK.

Clark, R.L., and Spengler, J. J. 1976. Changing demography and dependency costs. Unpublished manuscript.

Coale, A.J. 1956. The effects of changes in mortality and fertility on age composition. Milbank Memorial Fund Quarterly 34:79–114.

Coale, A.J. 1957. A new method for calculating Lotka's r—the intrinsic rate of growth in a stable population. Population Studies 11:92–94.

Coale, A.J. 1963. Estimate of various demographic measures through the quasi-stable age distribution. In Emerging Techniques in Population Research (39th Annual Conference of the Milbank Memorial Fund, 1962), pp. 175–193. Milbank Memorial Fund.

Coale, A.J. 1965. Appendix: Estimates of average size of household. In A. J. Coale et al., Aspects of the Analysis of Family Structure, pp. 64–69. Princeton University, Office of Population Research, Princeton, New Jersey, USA.

Coale, A.J. 1968. Convergence of a human population to a stable form. Journal of the American Statistical Association, 63:395–435.

Coale, A.J. 1971. Age patterns of marriage, Population Studies 25:193–214.

Coale, A.J. 1972. The Growth and Structure of Human Populations: A Mathematical Investigation. Princeton University Press, Princeton, New Jersey, USA.

Coale, A.J. 1973a. The demographic transition reconsidered. International Population Conference, Liege. International Union for the Scientific Study of Population.

Coale, A.J. 1973b. Age composition in the absence of mortality and in other odd circumstances. Demography 10:537–542.

Coale, A.J., and P. Demeny. 1966. Regional Model Life Tables and Stable Populations. Princeton University Press, Princeton, New Jersey, USA.

Coale, A.J., and P. Demeny. 1967. Method of Estimating Basic Demographic Measures from Incomplete Data. Manuals on Methods of Estimating Population, Manual 4. ST/SOA/Series A/42. United Nations.

Coale, A.J., and E.M. Hoover. 1958. Population Growth and Economic Development in Low-Income Countries: A Case Study of India's Prospects. Princeton University Press, Princeton, New Jersey, USA.

Coale, A.J., and D.R. McNeil. 1972. Distribution by age of the frequency of first marriage in a female cohort. Journal of the American Statistical Association 67:743–749.

Coale, A.J., and T.J. Trussell. 1974. Model fertility schedules: Variations in the age structure of childbearing in human populations. Population Index 40:185–258.

Coale, A.J., and M. Zelnik. 1963. New Estimates of Fertility and Population in the United States. Princeton University Press, Princeton, New Jersey, USA.

Cochran, M.E. and S. Ellner. 1992. Simple methods for calculating age-based life history parameters for stage-structured populations. Ecological Monographs 62:345–364.

Coddington, E.A., and N. Levinson. 1955. Theory of Ordinary Differential Equations. McGraw-Hill, New York, New York, USA.

Cohen, J.E. 1973. Economic evaluation of small changes in the life table. Paper presented at Conference of International Union for the Scientific Study of Population, Liege.

Cohen, J.E. 1978. Derivatives of the spectral radius as a function of non-negative matrix elements. Mathematical Proceedings of the Cambridge Philosophical Society 83:183–190.

Cohen, J.E. 1979a. Ergodic theorems in demography. Bulletin of the American Mathematical Society 1:275–295.

Cohen, J.E. 1979b. The cumulative distance from an observed to a stable age structure. SIAM Journal of Applied Mathematics 36(1):169–175.

Cohen, J.E. 1986. An uncertainty principle in demography and the unisex issue. American Statistician 40:32–39.

Cohen, J.E. 1995. How Many People Can the Earth Support? Norton, New York, New York, USA.

Cohen, J.E. 2003. Human population: The next half century. Science 302:1172–1175.

Cohen, M.R., and E. Nagel. 1934. An Introduction to Logic and Scientific Method. Harcourt-Brace, New York, New York, USA.

Cole, L. 1954. The population consequences of life history phenomena. Quarterly Review of Biology 19:103–137.

Coleman, J.S. 1973. Loss of power. American Sociological Review 38:1–17.

Cook, R.C., Editor. 1962. How many people have ever lived on earth? Population Bulletin 18:1–19.

Cook, R.E. 1985. Growth and development in clonal plant populations. Pages 259–296 in J.B.C. Jackson, L.W. Buss, and R.E. Cook, editors. Population Biology and Evolution of Clonal Organisms. Yale University Press, New Haven, Connecticut, USA.

Coombs, C.H., L.C. Coombs, and G.H. McClelland. 1975. Preference scales for number and sex of children. Population Studies 29:273–298.

Correa, H. 1967. A survey of mathematical models in educational planning. In Mathematical Models in Educational Planning. Technical Report. Organization for Economic Co-operation and Development, Paris, France.

Cramer, J. 1972. Personal communication.

Crowcroft, P. 1991. Elton's Ecologists. University of Chicago Press, Chicago, Illinois, USA.

Croxall, J.P., P. Rothery, S.P.C. Pickering and P.A. Prince. 1990. Reproductive performance, recruitment and survival of wandering albatrosses *Diomedia exulans* at Bird Island, South Georgia. Journal of Animal Ecology 59:775–796.

Cull, P. and A. Vogt. 1973. Mathematical analysis of the asymptotic behavior of the Leslie population matrix model. Bulletin of Mathematical Biology 35:645–661.

Cull, P. and A. Vogt. 1974. The periodic limit for the Leslie model. Mathematical Biosciences 21:39–54.

Cull, P. and A. Vogt. 1976. The period of total population. Bulletin of Mathetical Biology 38:317–319.

Cummings, S.R., D.M. Black, and S.M. Rubin. 1989. Lifetime risks of hip, Colle's, or vertebral fracture and coronary heart disease among white postmenopausal women. Archives of Internal Medicine 149:2445–2448.

Cushing, J.M. 1988. Nonlinear matrix models and population dynamics. Natural Resource Modeling 2:539–580.

Cushing, J.M. 1997. Nonlinear matrix equations and population dynamics. Pages 205–243 in S. Tuljapurkar and H. Caswell, editors. Structured-Population Models in Marine, Terrestrial, and Freshwater Systems. Chapman and Hall, New York, New York.

Cushing, J.M. 1998. An introduction to structured population dynamics. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.

Cushing, J.M. and Z. Yicang. 1994. The net reproductive value and stability in matrix population models. Natural Resources Modeling 8:297–333.

Cushing, J.M., R.F. Costantino, B. Dennis, R.A. Desharnais, and S.M. Henson. 2003. Chaos in Ecology. Academic Press, New York, New York, USA.

Das Gupta, P. 1976. Age-parity-nuptiality-specific stable population model that recognizes births to single women. Journal of the American Statistical Association 71:308–314.

David, P.A. 1974. Fortune, risk, and the microeconomics of migration. In Nations and Households in Economic Growth, edited by P. A. David and M. W. Reder, pp. 21–88. Academic Press, New York, New York, USA.

DeAngelis, D.L. and L.J. Gross. 1992. Individual based models and approaches in ecology. Chapman and Hall, New York, New York.

Deevey, E. S., Jr. 1950. (1974) The probability of death. Scientific American. Reprinted in Readings from Scientific American, edited by E.O. Wilson, pp. 106–108. W. H. Freeman, San Francisco, California, USA.

de Kroon, H., A. Plaisier, J. van Groenendael, and H. Caswell. 1986. Elasticity: The relative contribution of demographic parameters to population growth rate. Ecology 67:1427–1431.

de Kroon, H., J. van Groenendael, and J. Ehrlén. 2000. Elasticities: A review of methods and model limitations. Ecology 81:607–618.

Demeny, P. 1965. Estimation of vital rates for populations in process of destabilization. Demography 2:516–530.

Demetrius, L. 1969. The sensitivity of population growth rate to perturbations in the life cycle components. Mathematical Biosciences 4:129–136.

Demetrius, L. 1971. Primitivity conditions for growth matrices. Mathematical Biosciences 12:53–58.

Demetrius, L. 1974. Demographic parameters and natural selection. Proceedings of the National Academy of Sciences USA 12:4645–4647.

Demetrius, L. 1983. Statistical mechanics and population biology. Journal of Statistical Physics 30:709–753.

de Roos, A.M. 1997. A gentle introduction to physiologically structured population models. Pages 119–204 in S. Tuljapurkar and H. Caswell, editors. Structured-Population Models in Marine, Terrestrial, and Freshwater Systems. Chapman and Hall, New York, New York, USA.

de Roos, A.M., O. Diekmann, and J.A.J. Metz. 1992. Studying the dynamics of structured population models: A versatile technique and its application to *Daphnia*. American Naturalist 139:123–147.

Desoer, C.A. 1967. Perturbations of eigenvalues and eigenvectors of a network. Pages 8–11 in Fifth Annual Allerton Conference on Circuit and System Theory. University of Illinois, Urbana, Illinois, USA.

de Vries, J.J., P. Nijkamp, and P. Rietveld. 2000. Alonso's general theory of movement: Advances in spatial interaction modeling. Discussion Paper TI 2000-062/3 of the Tinbergen Institute, Amsterdam, the Netherlands.

Diekmann, O. and J.A.P. Heesterbeek. 2000. Mathematical epidemiology of infectious diseases: Model building, analysis, and interpretation. Wiley, New York, New York, USA.

Diekmann, O., J.A.P. Heesterbeek, and J.A.J. Metz. 1990. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology 28:365-382.

Doak, D., P. Kareiva, and B. Klepetka. 1994. Modeling population viability for the desert tortoise in the western Mojave Desert. Ecological Applications 4:446–460.

Dobbernack, W., and G. Tietz. 1940. Die Entwicklung von Personengesamtheiten vom Standpunkt der Sozialversicherungstechnik, Twelfth International Congress of Actuaries 4:233–253.

Dorn, H. F. 1950. Pitfalls in population forecasts and projections. Journal of the American Statistical Association 45:311–334.

Durkheim, E. 1973. De la Division du Travail Social. PUF, Paris, France.

Easterlin, R.A. 1961. The baby boom in perspective. American Economic Review 51:869–911.

Easterlin, R.A. 1966. On the relation of economic factors to recent and projected fertility changes. Demography 3:131–153.

Easterlin, R.A. 1968. The current fertility decline and projected fertility changes. Chapter 5 in Population, Labor Force, and Long Swings in Economic Growth: The American Experience. Columbia University Press, New York, New York, USA.

Easterlin, R.A. 1980. Birth and Fortune. Basic Books, New York, New York, USA.

Easterling, M.R., S.P. Ellner, and P. Dixon. 2000. Size-specific sensitivity: Applying a new structured population model. Ecology 81:694–708.

Ebert, T.A. 1985. Sensitivity of fitness to macroparameter changes: An analysis of survivorship and individual growth in sea urchin life histories. Oecologia 65:461–467.

Efron, B. and R.J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York, New York, USA.

Ehrlich, P.R., and A.H. Ehrlich. 1972. Population, Resources, Environment: Issues in Human Ecology. 2nd ed. W. H. Freeman, San Francisco, California, USA.

Ekanem, I. 1972. A further note on the relation between economic development and fertility. Demography 9:383–398.

Emlen, J.M. 1970. Age specificity and ecological theory. Ecology 51:588–601.

Espenshade, T.J. 1975. The stable decomposition of the rate of natural increase. Theoretical Population Biology 8:97–115.

Euler, L. 1760. Recherches générales sur la mortalité et la multiplication [General researches on mortality and multiplication]. Mémoires de l'Académie Royale des Sciences et Belles Lettres 16:144–164. Translated by N. Keyfitz and B. Keyfitz. Theoretical Population Biology 1:307–314, 1970.

Faddeev, D.K. 1959. The conditionality of matrices ("Ob obuslovlennosti matrits"). Matematicheskii institut Steklov, Trudy No. 53:387–391.

Faddeev, D.K., and V.N. Faddeeva. 1963. Computational methods of linear algebra. W. H. Freeman, San Francisco, California, USA.

Feeney, G. 1970. Stable age by region distributions. Demography 7:341–348.

Feeney, G. 1973. Two models for multiregional population dynamics. Environment and Planning 5:31–43.

Feeny, P. 1976. Plant apparency and chemical defense. Recent Advances in Phytochemistry 10:1–40.

Feichtinger, G. 1971. Stochastische Modelle demographischer Prozesse. In Lecture Notes in Operations Research and Mathematical Systems, Vol. 44, edited by M. Beckmann and H. P. Künzi). Springer, Berlin, Germany.

Feichtinger, G. 1972. Markoff-Modelle zur sozialen Mobilität, Jahrbuch für Sozialwissenschaft.

Feichtinger, G. 1973. Markovian models for some demographic processes. Statistiche Hefte 14:310–334.

Feldstein, M. 1976. The social security fund and national capital accumulation. Paper presented at the Federal Reserve Bank of Boston conference on "Funding Pensions: The Issues and Implications for Financial Markets."

Feller, W. 1941. On the integral equation of renewal theory. Annals of Mathematical Statistics 12:243–267.

Feller, W. 1957. An Introduction to Probability Theory and Its Applications, Vol. 1. 2nd ed. Wiley, New York, New York, USA.

Feller, W. 1971. An Introduction to Probability Theory and Its Applications, Vol. 2. Wiley, New York, New York, USA.

Fingleton, B. 1984. Models of category counts. Cambridge University Press, Cambridge.

Fisher, R.A. 1930 (republished 1958). The Genetical Theory of Natural Selection. Dover Publications, New York, New York, USA.

Fix, E., and J. Neyman. 1951. A simple stochastic model of recovery, relapse, death and loss of patients. Human Biology 23:205–241.

Flieger, W. 1967. A re-examination of the demographic transition in the light of newly collected data. Ph.D. Dissertation, Department of Sociology, University of Chicago.

Forsyth, D.M. and R.P. Duncan. 2001. Propagule size and the relative success of exotic ungulate and bird introductions to New Zealand. American Naturalist 157:583–595.

Frauenthal, J. and K. Swick. 1983. Limit cycle oscillations of the human population. Demography 20:285–298.

Fredrickson, A. G. 1971. A mathematical theory of age-structure in sexual populations. Mathematical Biosciences 10:117–143.

Freeman, H. 1962. Finite Differences for Actuarial Students. Cambridge University Press, Cambridge, UK.

Frejka, T. 1973. The Future of Population Growth: Alternative Paths to Equilibrium. John Wiley, New York, New York, USA.

Friedlander, S., and M. Silver. 1967. A quantitative study of the determinants of fertility behavior. Demography 4:30–70.

Frisch, R.E. 1972. Weight at menarche: Similarity for well-nourished and undernourished girls at differing ages, and evidence for historical constancy. Pediatrics 50:445–450.

Frisch, R.E. 1984. Body fat, puberty and fertility. Biological Reviews 59:161–188.

Frobenius, G. 1912. Uber Matrizen aus nicht negativen Elementen. Sitzungsberichte Preussischer Akademie, 456–477.

Fuchs, W. 1951. Über die Zahl der Menschen die bisher gelebt haben. Zietschrift für die gesamte Staatswissenschaft 107:440–450.

Fujiwara, M. and H. Caswell. 2001. Demography of the endangered North Atlantic right whale. Nature 414:537-541.

Fujiwara, M., H. Caswell, and C.M. Hunter. 2004. A stage-classified demographic analysis of the North Atlantic right whale. (in prep.)

Furstenberg, H. and H. Kesten. 1960. Products of random matrices. Annals of Mathematical Statistics. 31:457–469.

Galton, F., and H.W. Watson. 1874. On the probability of extinction of families. Journal of the Anthropological Institute 6:138–144.

Gantmacher, F.R. 1959. Matrix Theory. Chelsea Publishing Company, New York, New York, USA.

Gilbert, J.P., and E.A. Hammel. 1966. Computer simulation and analysis of problems in kinship and social structure. American Anthropologist 68:71–93.

Gillespie, R.P. 1951. Partial Differentiation. Oliver and Boyd, Edinburgh, Scotland.

Gini, C. 1924. Premières recherches sur la fécondabilité de la femme [Preliminary research on female fertility]. In Proceedings of the International Mathematics Congress, Vol. 2, pp. 889–892.

Glass, D.V., and E. Grebenik. 1954. The Trend and Pattern of Fertility in Great Britain: A Report on the Family Census of 1946. Vol. 6 of Papers of the Royal Commission on Population. H.M. Stationery Office, London, UK.

Glazer, N. 1984. Reforming the American welfare family:1969–1981. Toqueville Review 6(1):149–168.

Glick, P.C. 1955. The life cycle of the family. Marriage and Family Living 17:3–9.

Glick, P.C. 1977. Updating the life cycle of the family. Journal of Marriage and the Family 39, 1:5–13.

Glick, P.C. 1979. The Future of the American Family. Current Population Reports, Special Studies, P-23, No. 78. United States Bureau of the Census, Washington, DC.

Goldman, N. 1978. Estimating the intrinsic rate of increase from the average numbers of younger and older sisters. Demography 15:499–508.

Goldman, N. 1983. Sex differences in life cycle measures of widowhood. Demography 20:177–195.

Goldstein, J. 2001. Formal demography. Visions of the future: DBSB long-range planning. National Institute of Child Health and Human Development, Demographic and Behavioral Sciences Branch, Washington DC. `http://www.nichd.nih.gov/cpr/dbs/sp/golstein1.htm`

Goldstein, S. 1976. Facets of redistribution: Research challenges and opportunities. Presidential address presented at the Annual Meeting of the Population Association of America, Montreal, Canada.

Golini, A. 1967. L'Influenza delle variazioni di mortalità sulla struttura per sesso della popolazione. Società Italiana di Statistic Atti della XXV Riunione Scientifica, pp. 865–898.

Golubitsky, M., E.B. Keeler, and M. Rothschild. 1975. Convergence of the age structure: Applications of the projective metric. Theoretical Population Biology 7:84–93.

Goode, W.J. 1984. Individual investment in family relationships over the coming decade. Toqueville Review 6(1):51–83.

Goodman, L.A. 1953. Population growth of the sexes. Biometrics 9:212–225.

Goodman, L.A. 1961a. Some possible effects of birth control on the human sex ratio. Annals of Human Genetics 25:75–81.

Goodman, L.A. 1961b. Statistical methods for the mover-stayer model. Journal of the American Statistical Association 56:841–868.

Goodman, L.A. 1968. An elementary approach to the population projection-matrix, to the population reproductive value, and to related topics in the mathematical theory of population growth. Demography 5:382–409.

Goodman, L.A. 1969. The analysis of population growth when the birth and death rates depend upon several factors. Biometrics 25:659–681.

Goodman, L.A. 1971. On the sensitivity of the intrinsic growth rate to changes in the age-specific birth and death rates. Theoretical Population Biology 2:339–354.

Goodman, L.A., N. Keyfitz, and T.W. Pullum. 1974. Family formation and the frequency of various kinship relationships. Theoretical Population Biology 5:1–27.

Goodman, L.A., N. Keyfitz, and T.W. Pullum. 1975. Addendum to "Family formation and the frequency of various kinship relationships." Theoretical Population Biology 8:376–381.

Greville, T.N.E. 1943. Some methods of constructing abridged life tables. Record of the American Institute of Actuaries 32, Part 1 29–43.

Greville, T.N.E. 1968. Some new generalized inverses with spectral properties. In Theory and Application of Generalized Inverses of Matrices. Texas Technological College Mathematics Series No. 4, pp. 26–46. Texas Technological Press, Lubbock, Texas, USA.

Greville, T.N.E., and N. Keyfitz. 1974. Backward population projection by a generalized inverse, Theoretical Population Biology 6:135–142.

Grimm, V., T. D. Wyszomirski, J. Aikman, and J. Uchman. 1999. Individual-based modelling and ecological theory: Synthesis of a workshop. Ecological Modelling 115: 275-282

Gurney, W.S.C. and R.M. Nisbet. 1998. Ecological Dynamics. Oxford University Press, Oxford, UK.

Gurney, W.S.C., E. McCauley, R.M. Nisbet, and W.W. Murdoch. 1990. The physiological ecology of *Daphnia*: A dynamic model of growth and reproduction. Ecology 71:716–732.

Gurtin, M.E. and R.C. MacCamy. 1974. Non-linear age-dependent population dynamics. Archive for Rational Mechanics and Analysis 54:281–300.

Hadwiger, H. 1940. Eine analytische Reproduktionsfunktion für biologische Gesamtheiten [An analytic reproduction function for biological groups]. Skandinavisk Aktuarietidskrift (Stockholm) 23:101–113.

Hajnal, J. 1955. The prospects for population forecasts. Journal of the American Statistical Association 50:309–322.

Hallam, T.G., R.R. Lassiter, J. Li, and L.A. Suarez. 1990. Modelling individuals employing an integrated energy response: application to *Daphnia*. Ecology 71:938–954.

Hamilton, W.D. 1966. The moulding of senescence by natural selection. Journal of Theoretical Biology 12:12–45.

Hammel, E.A. 1972. The matrilateral implications of structural cross-cousin marriage. In Proceedings of the Advanced Seminar in Anthropology and Demography edited by E. Zubrow. School of American Research, Santa Fe, California, USA.

Hammel, E.A., et al. 1976. The SOCSIM Demographic-Sociological Microsimulation Program. Institute ofInternational Studies. University of California, Berkeley, California, USA.

Hammel, E.A., C.K. McDaniel, and K.W. Wachter. 1979. Demographic consequences of incest tabus: A microsimulation analysis. Science 205:972–977.

Hannan, M. T., N.B. Tuma, and L.P. Groenevald. 1977. Income and marital events: Evidence from an income maintenance experiment. American Journal of Sociology 82:6, 1186–1211.

Hansen, F.T. 1997. The effects of chronic exposure to 4-n-nonylphenol on life-history traits and population dynamics of hermaphrodites in the marine polychaete *Capitella* sp. 1. Thesis, University of Roskilde, Denmark.

Harris, T.E. 1963. The Theory of Branching Processes. Englewood Cliffs, N. J.: Prentice-Hall.

Hastings, A. 1992. Age dependent dispersal is not a simple process: Density dependence, stability, and chaos. Theoretical Population Biology 41:388–400.

Hatzold, O. 1974. "Possibilities of a pre-conceptional sex-determination and their effects on sex ratio at birth." Program summary of the research project. University of Munich, Germany. (Mimeographed)

Heckman, J.J. and B. Singer. 1982. Population heterogeneity in demographic models. Pages 567–599 in K.C. Land and A. Rogers (editors). Multidimensional mathematical demography. Academic Press, New York, New York, USA.

Heer, D.M. 1966. Economic development and fertility. Demography 3:423–444.

Henripin, J. 1982. Discrétion et témérité à propos des politiques de population Montréal: Université de Montréal.

Henry, L. 1960. Mesure indirecte de la mortalité des adultes Population 15:457–465.

Henry, L. 1969. Schémas de nuptialité: déséquilibre des sexes et célibat, Population 24:457–486.

Henry, L. 1972. On the Measurement of Human Fertility: Selected Writings of Louis Henry. Translated and edited by M.C. Sheps and E. Lapierre-Adamcyk. Elsevier, Amsterdam, The Netherlands.

Heppell, S.S., L.B. Crowder, and D.T. Crouse. 1996. Models to evaluate headstarting as a management tool for long-lived turtles. Ecological Applications 6:556–565.

Heppell, S.S., H. Caswell, and L.B. Crowder. 2000. Life histories and elasticity patterns: Perturbation analysis for species with minimal demographic data. Ecology 81:654–665.

Herglotz, G. 1908. Über die Integralgleichungen der Elektronentheorie [Concerning the integral equations of the electron theory]. Mathematische Annalen 65:87–106.

Hernes, G. 1972. The process of entry into first marriage. American Sociological Review 27:173–182.

Hertz, P. 1908. Die Bewegung eines Elektrons unter dem Einflusse einer stets gleich gerichteten Kraft [The movement of an electron under the influence of a force of constant unchanging direction]. Mathematische Annalen 65:1–86.

Hicks, J.R. 1939. Value and Capital. Oxford University Press, Oxford, UK.

Hirschman, C., and J. Matras. 1971. A new look at the marriage market and nuptiality rates, 1915–1958. Demography 8:549–569.

Hoem, J. 1975. The construction of increment-decrement life tables: A comment on articles by R. Schoen and V. Nelson. Demography 12:661.

Hoem, J., and Fong, M. 1976. A Markov Chain Model of Working Life Tables: A New Method for the Construction of Tables of Working Life. Working Paper No. 2. Laboratory of Actuarial Mathematics, University of Copenhagen, Copenhagen, Denmark.

Hoem, J. M. 1971. On the interpretation of the maternity function as a probability density. Theoretical Population Biology 2:319–327.

Hoem, J.M. 1990. Limitations of a heterogeneity technique: Selectivity issues in conjugal union disruption at parity zero in contemporary Sweden. Pages 133–151 in J. Adams, D.A. Lam, A.I. Hermalin, and P.E. Smouse (editors) Convergent Issues in Genetics and Demography. Oxford University Press, New York, New York, USA.

Hoogendyk, C.G. and G.F. Estabrook. 1984. The consequences of earlier reproduction in declining populations. Mathematical Biosciences 71:217–235.

Horn, R.A. and C.R. Johnson. 1985. Matrix Analysis. Cambridge University Press, Cambridge, UK.

Horst, T.J. 1977. Use of the Leslie matrix for assessing environmental impact with an example for a fish population. Transactions of the American Fisheries Society 106:253–257.

Horvitz, C.C. and D.W. Schemske. 1995. Spatiotemporal variation in demographic transitions of a tropical understory herb: projection matrix analysis. Ecological Monographs 65:155–192.

Horvitz, C.C., D.W. Schemske, and H. Caswell. 1997. The "importance" of life history stages to population growth: prospective and retrospective analyses. Pages 247–272 in S. Tuljapurkar and H. Caswell, editors. Structured population models in marine, Terrestrial and Freshwater Systems. Chapman and Hall, New York, New York, USA.

Hubbell, S.P. and P.A. Werner. 1979. On measuring the intrinsic rate of increase of populations with heterogeneous life histories. American Naturalist 113:277–293.

Hughes, T.P. and J.B.C. Jackson. 1985. Population dynamics and life histories of foliaceous corals. Ecological Monographs 55:141–166.

Hunter, C.M. and H. Caswell. 2004. The use of the vec-permutation matrix in spatial matrix population models. Ecological Modelling (in press).

Iosifescu, M. 1980. Finite Markov Processes and Their Applications. Wiley, New York, USA.

Jackson, J.B.C., L.W. Buss, and R.E. Cook. 1985. Population Biology and Evolution of Clonal Organisms. Yale University Press, New Haven, Connecticut, USA.

Jacobi, C.J.G. 1846. Uber ein leichtes Verfagren die in der Theorie der Sacularstorungen Vorkommenden Gleichungen numerisch aufsulosen. J. Reine Angew. Math. 30:51–95.

Jain, A.K. 1969. Fecundability and its relation to age in a sample of Taiwanese women. Population Studies 23:69–85.

Janowitz, B.S. 1971. An empirical study of the effects of socioeconomic development on fertility rates. Demography 8:319–330.

Janzen, D.H. 1976. Why bamboos wait so long to flower. Annual Review of Ecology and Systematics 7:347–391.

Jordan, C.W. 1952. Life Contingencies. Society of Actuaries, Chicago, Illinois, USA.

Karmel, P.H. 1948a. The relations between male and female nuptiality in a stable population. Population Studies 1:352–387.

Karmel, P.H. 1948b. An analysis of the sources and magnitudes of inconsistencies between male and female net reproduction rates in actual populations. Population Studies 2:240–273.

Karmel, P.H. 1950. A note on P. K. Whelpton's calculation of parity-adjusted reproduction rates. Journal of the American Statistical Association 45:119–124.

Kato, T. 1982. A Short Introduction to Perturbation Theory for Linear Operators. Springer-Verlag, New York, New York, USA.

Kemeny, J.G. and J.L. Snell. 1976. Finite Markov Chains. Springer-Verlag, New York, New York, USA.

Kemeny, J.G. J.L. Snell, and G.L. Thompson. 1974. Introduction to Finite Mathematics. Prentice-Hall, New York, New York, USA.

Kendall, D. 1949. Stochastic processes and population growth. Journal of the Royal Statistical Society, Series B 11:230–264.

Kendall, M.G. and A. Stuart. 1958. The Advanced Theory of Statistics. 3 vols. Hafner, New York, New York, USA.

Kermack, W.O. and A.G. McKendrick. 1927. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Statistical Society of London, Series A, 115:700-721.

Keyfitz, N. 1964. The population projection as a matrix operator. Demography 1:56–73.

Keyfitz, N. 1967. Reconciliation of population models: Matrix, integral equation and partial fraction. Journal of the Royal Statistical Society 130:61–83.

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

Keyfitz, N. 1969. Age distribution and the stable equivalent. Demography 6:261–269.

Keyfitz, N. 1971a. Linkages of intrinsic to age-specific rates. Journal of the American Statistical Association 66:275–281.

Keyfitz, N. 1971b. On the momentum of population growth. Demography 8:71–80.

Keyfitz, N. 1972a. On future population. Journal of the American Statistical Association 67:347–363.

Keyfitz, N. 1972b. Population waves. In Population Dynamics, edited by T. N. E. Greville, pp. 1–38. Academic Press, New York, New York, USA.

Keyfitz, N. 1973. Individual mobility in a stationary population. Population Studies 27:335–352.

Keyfitz, N. 1977. Applied Mathematical Demography. Wiley, New York, New York, USA.

Keyfitz, N. 1980. Multistate demography and its data: A comment. Environment and Planning A 12:615–622.

Keyfitz, N. 1981. The limits of population forecasting. Population and Development Review 7:579–593.

Keyfitz, N. 1982. Can knowledge improve forecasts? Population and Development Review 8:729–751.

Keyfitz, N. 1993. Are there ecological limits to population? Proceedings of the National Academy of Sciences USA 90:6895–6899.

Keyftiz, N. 1994. Influence of human reproduction on environment: What features of biology and economics cause them to see population differently? pp. 331–346 in K.L. Campbell and J.W. Wood, editors. Human Reproductive Ecology: Interactions of Environment, Fertility, and Behavior. Annals of the New York Academy of Sciences 709.

Keyfitz, N. 1996. Population growth, development and the environment. Population Studies 50:335–359.

Keyfitz, N. and W. Flieger. 1968. World Population: An Analysis of Vital Data. The University of Chicago Press, Chicago, Illinois, USA.

Keyfitz, N. and W. Flieger. 1971. Population: Facts and Methods of Demography. W.H. Freeman, San Francisco, California, USA.

Keyfitz, N., D. Nagnur, and D. Sharma. 1967. On the interpretation of age distributions. Journal of the American Statistical Association 62:862–874.

Kitagawa, E.M. 1964. Standardized comparisons in population research. Demography 1:296–315.

Kitsul, P. 1978. On a probabilistic approach to transition migration processes. Unpublished manuscript.

Kooijman, S.A.L.M. 1993. Dynamic Energy Budgets in Biological Systems. Cambridge University Press, Cambridge, UK.

Koppel, H. 1987. Lifetime likelihood of victimization. Bureau of Justice Statistics Technical Report, U.S. Department of Justice.

Krakowski., M. 1972. Lacunology, entropy, and keeping up with the Joneses. In The Mathematics of Large-Scale Simulation, edited by Paul Brock, pp. 55–59. Simulation Councils Proceedings Series, Vol. 2, No. 1.

Krishnamoorthy, S. 1978. The demography of the life cycle. Unpublished doctoral dissertation, School of Public Health, Harvard University.

Kunstadter, P., et al. 1963. Demographic variability and preferential marriage patterns. American Journal of Physical Anthropology, New Series 21:511–519.

Lachapelle, R. 1980. La situation démolinguistique au Canada: Evolution passée et prospective. Montreal: Institute de Recherches Politiques.

Lachenbruch, P.A. 1967. Frequency and timing of intercourse: Its relation to the probability of conception. Population Studies 21:23–31.

Lachenbruch, P.A., M.C. Sheps, and A.M. Sorant. 1973. Applications of POPREP, a modification of POPSIM. In Computer Simulation in Human Population Studies, edited by B. Dyke and J.W. MacCluer, pp. 305–328. Academic Press, New York, New York, USA.

Land, K.C. and A. Rogers. 1982. Multidimensional Mathematical Demography. Academic Press, New York, New York, USA.

Law, R. and M.T. Edley. 1990. Transient dynamics of populations with age- and size-dependent vital rates. Ecology 71:1863–1870.

Lebreton, J.-D. 1996. Demographic models for subdivided populations: The renewal equation approach. Theoretical Population Biology 49:291–313.

Lebreton, J.-D. and G. Gonzalez-Davila. 1993. An introduction to models of subdivided populations. Journal of Biological Systems 1:389–423.

Le Bras, H. 1973. Parents, grandparents, diaeresis bisaïeux. Population 28:9–37 (English translation K. Wachter, editor, 1978. Statistical Studies of Historical Social Structure, Academic Press, New York, New York, USA.)

Le Bras, H. 1984. Simulation des liens de famille. Unpublished manuscript.

Le Bras, H. and J.-C. Chesnais. 1976. Cycle de l'habitat et âge des habitants. Population 31:269–298.

Lee, B.M., and J. Isbister. 1966. The impact of birth control programs on fertility. pp. 737–758 in Family Planning and Population Programs, edited by B. Berelson. University of Chicago Press, Chicago, Illinois, USA.

Lee, R.D. 1968. Population cycles. Unpublished manuscript. Department of Economics, Harvard University.

Lee, R. . 1974. The formal dynamics of controlled populations and the echo, the boom and the bust. Demography 11:563–585.

Lee, R.D. 1985. Inverse projection and back projection: Comparative results and sensitivity tests for England. Population Studies 39:233–248.

Lee, R.D. 1987. Population dynamics of humans and other animals. Demography 24:443-465.

Lee, R.D. 1993. Inverse projection and demographic fluctuations: A critical assessment of new methods. pp. 7–28 in D. Reher and R. Schofield, editors, Old and New Methods in Historical Demography. Oxford University Press, Oxford, United Kingdom.

Lee, R.D. 2001. Formal demography. Visions of the future: DBSB long-range planning. National Institute of Child Health and Human Development, Demographic and Behavioral Sciences Branch, Washington DC. `http://www.nichd.nih.gov/cpr/dbs/sp/lee.htm`

Lee, R.D. 2003. Rethinking the evolutionary theory of aging: Transfers, not births, shape senescence in social species. Proceedings of the National Academy of Sciences 100:9637-9642.

Lee, R.D. and S. Tuljapurkar 1994. Stochastic population forecasts for the United States: Beyond high, medium, and low. Journal of the American Statistical Association 89:1175–1189.

Lefkovitch, L.P. 1963a. Estimating the Malthusian parameter from census data. Nature 204:810

Lefkovitch, L.P. 1965. The study of population growth in organisms grouped by stages. Biometrics 21:1–18

Lefkovitch, L.P. 1971. Some comments on the invariants of population growth. Pages 337–360 in G. P. Patil, E. C. Pielou, and W. E. Waters, editors. Statistical Ecology, volume 2. Pennsylvania State University Press, University Park, Pennsylvania, USA.

Leibenstein, H. 1974. An interpretation of the economic theory of fertility: Promising path or blind alley? Journal of Economic Literature 12:467–479.

Leslie, P. H. 1945. On the use of matrices in certain population mathematics. Biometrika 33:183–212.

Leslie, P.H. 1948. Some further notes on the use of matrices in population mathematics. Biometrika 35:213–245.

Leslie, P.H., J.S. Tener, M. Vizoso, and H. Chitty. 1955. The longevity and fertility of the Orkney vole, *Microtus orcadensis*, as observed in the laboratory. Proceedings of the Zoological Society of London 125:115–125.

Levin, L.A. and D. V. Huggett. 1990. Implications of alternative reproductive modes for seasonality and demography in an estuarine polychaete. Ecology 71:2191-2208.

Levin, L.A., H. Caswell, K.D. DePatra, and E.L. Creed. 1987. Demographic consequences of larval development mode: planktotrophy vs. lecithotrophy in *Streblospio benedicti*. Ecology 68:1877–1886.

Levin, L.A., H. Caswell, T. Bridges, C. DiBacco, D. Cabrera, and G. Plaia. 1996. Demographic response of estuarine polychaetes to pollutants: Life table response experiments. Ecological Applications 6:1295–1313.

Lewis, E.G. 1942. On the generation and growth of a population. Sankhya, Indian Journal of Statistics 6:93–96.

Lewis, E.R. 1972. Delay-line models of population growth. Ecology 53:797–807.

Lewis, E.R. 1976. Applications of discrete and continuous network theory to linear population models. Ecology 57:33–47.

Lewis, E.R. 1977. Network Models in Population Biology. Springer-Verlag, New York, New York, USA.

Lewontin, R.C. 1965. Selection for colonizing ability. Pages 77–94 in H. G. Baker and G. L. Stebbins, editors. The Genetics of Colonizing Species. Academic Press, New York, New York, USA.

Lexis, W. 1875. Einleitung in die Theorie der Bevölkerungs-Statistik [Introduction to the Theory of Population Statistics]. Strasbourg: Trubner.

Li, N. and S. Tuljapurkar. 1997. Population momentum. Working Paper 0070, Morrison Institute for Population and Resource Studies, Stanford University.

Liaw, K.-L. 1980. Multistate dynamics: The convergence of an age-by-region population system. Environment and Planning A 12:589–613.

Link, W.A., E. Cam, J.D. Nichols, and E.G. Cooch. 2002. Of BUGS and birds: Markov Chain Monte Carlo for hierarchical modeling in wildlife research. Journal of Wildlife Management 66:277–291.

Liu, P.T., L.P. Chow, and H. Abbey. 1972. A study of IUD retention by curve fitting. Demography 9:1–11.

Livi, L. 1974. Trattato di Demografia. 2 volumes. Cedam, Padova, Italy.

Logofet, D.O. and E.V. Lesnaya. 1997. Why are the middle points the most ‘sensitive’ in the sensitivity experiments? Ecological Modelling 104:303–306.

Longstaff, B.C. 1984. An extension of the Leslie matrix model to include a variable immature period. Australian Journal of Ecology 9:289–293.

Lopez, A. 1961. Problems in Stable Population Theory. Princeton University Press, Princeton, New Jersey, USA.

Lotka, A.J. 1922. The stability of the normal age distribution. Proceedings of the National Academy of Sciences 5:339–345.

Lotka, A.J. 1931. Orphanhood in relation to demographic factors. Metron (Rome) 9:37–109.

Lotka, A.J. 1939. Théorie analytique des associations biologiques. Part II. Analyse démographique avec application particulière à l'espèce humaine. Actualités Scientifiques et Industrielles, No. 780. Hermann et Cie, Paris, France. (Published in translation as: Analytical Theory of Biological Populations, translated by D.P. Smith and H. Rossert. Plenum Press, New York, 1998)

Lotka, A.J. 1945. Population analysis as a chapter in the mathematical theory of evolution. Pages 355–395 in W.E. LeGros Clark and P.B. Medawar, editors. Essays on growth and form. Oxford University Press, Oxford, UK.

Lotka, A. J. 1948. Application of recurrent series in renewal theory. Annals of Mathematical Statistics 19:190–206.

Lowry, I.S. 1964. A Model of Metropolis. Rand Corporation, Santa Monica, California, USA.

Luce, R.D., and H. Raiffa. 1957. Games and Decisions: Introduction and Critical Survey. Wiley, New York, New York, USA.

Macdonald, G. 1965. The dynamics of helminth infections, with special reference to schistosomes. Transactions of the Royal Society of Tropical Medicine and Hygiene 59:489–506.

Macunovich, D.J. 1998. Fertility and the Easterlin hypothesis: an assessment of the literature. Journal of Population Economics 11:53–111.

Majumdar, H., and M.C. Sheps. 1970. Estimators of a Type I geometric distribution from observations on conception times. Demography 7:349–360.

Mann, N.R., R.E. Schafer, and N.D. Singpurwalla. 1974. Methods for Statistical Analysis of Reliability and Life Data. Wiley, New York, New York, USA.

Manton, K.G., E. Stallard, and J.W. Vaupel. 1981. Methods for comparing the mortality experience of heterogeneous populations. Demography 18:389–410.

March, J.G. 1975. Comments at Conference on the Demography of Education, Harvard University, April.

Marks, E.S., W. Seltzer, and K.J. Krotki. 1974. Population Growth Estimation: A Handbook of Vital Statistics Measurement. Population Council, New York, New York, USA.

Martin, M.D.P. 1967. Une application des fonctions de Gompertz à l'étude de la fécondité d'une cohorte. Population 22:1085–1096.

Masnick, G.S., and J.A. McFalls, Jr. 1976. The twentieth century American fertility swing: A challenge to theories of fertility. Paper presented at the Annual Meeting of the Population Association of America.

May, R.M. 1973a. Stability and Complexity in Model Ecosystems. Princeton University Press, Princeton, New Jersey, USA.

May, R.M. 1973b. Time-delay versus stability in population models with two and three trophic levels. Ecology 54:316–325.

McDaniel, C.K., and E.A. Hammel. 1984. A kin-based measure of r and an evaluation of its effectiveness. Demography 21:41–51.

McDonald, J. 1973. Sex predetermination: Demographic effects. Mathematical Biosciences 17:137–146.

McFarland, D.D. 1969. On the theory of stable populations: A new and elementary proof of the theorems under weaker assumptions. Demography 6:301–322.

McFarland, D.D. 1970. Effects of group size on the availability of marriage partners. Demography 7:411–415.

McFarland, D.D. 1975a. Some mathematical models of marriage formation and their relevance to fertility models. Social Forces 54:66–83.

McFarland, D.D. 1975b. Models of marriage formation and fertility. Social Forces 54:66–83.

McKendrick, A.G. 1926. Applications of mathematics to medical problems. Proceedings of the Edinburgh Mathematical Society 40:98–130.

McGinnis, R. 1968. A stochastic model of migration. American Sociological Review 33:712–722.

McNeil, D.R. 1973. Pearl-Reed Type Stochastic Models for Population Growth. Technical Report 35, Series 2. Princeton University, Office of Population Research, Princeton, New Jersey, USA.

McNicoll, G. 1972. Optimal policies in economic–demographic growth models. Ph.D. dissertation, Graduate Division, University of California, Berkeley.

Medawar, P.B. 1952. An Unsolved Problem in Biology. H. K. Lewis, London, UK.

Menken, J. 1975. Biometric models of fertility. Social Forces 54:52–65.

Mertens, W. 1965. Methodological aspects of construction of nuptiality tables. Demography 3:317–348.

Mertz, D.B. 1971. Life history phenomena in increasing and decreasing populations. Pages 361–399 in G. P. Patil, E.C. Pielou, and W.E. Waters, editors. Statistical Ecology. II. Sampling and Modeling Biological Populations and Population Dynamics. Pennsyvania State University Press, University Park, Pennsylvania, USA.

Mesterton-Gibbons, M. 1993. Why demographic elasticities sum to one: a postscript to de Kroon et al. Ecology 74:2467-2468.

Metz, J.A.J. 1977. State space models for animal behavior. Annals of System Research 6:65–109.

Metz, J.A.J. and A.M. de Roos. 1992. The role of physiologically structured population models within a general individual-based modeling perspective. Pages 88–111 in D.L. DeAngelis and L.J. Gross, editors.

Individual-Based Models and Approaches in Ecology. Chapman and Hall, New York, New York, USA.

Metz, J.A.J. and O. Diekmann. 1986. The Dynamics of Physiologically Structured Populations. Springer-Verlag, Berlin, Germany.

Mode, C.J. 1975. Perspectives in stochastic models of human reproduction: A review and analysis. Theoretical Population Biology 8:247–291.

Morrison, P.A. 1971. Chronic movers and the future redistribution of population. Demography 8:171–184.

Moss, C.J. 2001. The demography of an African elephant (*Loxodonta africana*) population in Amboseli, Kenya. Journal of Zoology, London 255:145–156.

Mosteller, F., et al. 1949. The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-election Polls and Forecasts. Social Science Research Council, New York, New York, USA.

MPM. Refers to Caswell, H. 2001. Matrix population models: Construction, analysis, and interpretation. 2nd ed. Sinauer, Sunderland, Massachusetts, USA.

Muhsam, H.V. 1956. The utilization of alternative population forecasts in planning. Bulletin of the Research Council of Israel 5:133–146.

Murphy, E.M., and Nagnur, D.N. 1972. A Gompertz fit that fits: applications to Canadian fertility patterns. Demography 9:35–50.

Neubert, M.G. and H. Caswell. 2000. Demography and dispersal: invasion speeds of stage-structured populations. Ecology 81:1613–1628.

Nisbet, R.M. 1997. Delay-differential equations for structured populations. Pages 89–118 in S. Tuljapurkar and H. Caswell, editors. Structured-Population Models in Marine, Terrestrial, and Freshwater Systems. Chapman and Hall, New York, New York, USA.

Nisbet, R.M. and W.S.C. Gurney. 1982. Modelling Fluctuating Populations. Wiley, New York, New York, USA.

Nortman, D. and J. Bongaarts. 1975. Contraceptive practice required to meet a prescribed crude birth rate target: A proposed macro-model (TABRAP) and hypothetical illustrations. Demography 12:471–489.

Notestein, F.W. 1945. Population—The long view. In Food for the World, edited by T.W. Schultz. University of Chicago Press, Chicago, Illinois, USA.

Oechsli, F. 1975. A population model based on a life table that includes marriage and parity. Theoretical Population Biology 7:229–245.

Oechsli, F.W. 1971. The parity and nuptiality problem in demography. Unpublished Ph.D. dissertation, Graduate Division, University of California, Berkeley.

Olesiuk, P.F., M.A. Bigg, and G.M. Ellis. 1990. Life history and population dynamics of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State. Report of the International Whaling Commission, Special Issue 12:209–243.

Orcutt, G.H., et al. 1961. Microanalysis of Socioeconomic Systems: A Simulation Study. Harper, New York, New York, USA.

Orcutt, G.H., et al. 1976. Population Exploration Through Microsimulation. Washington, D. C.: Urban Institute.

Oster, G. 1976. Internal variables in population dynamics. Lectures on Mathematics in the Life Sciences 8:37–68.

Oster, G. and Y. Takahashi. 1974. Models for age-specific interactions in a periodic environment. Ecological Monographs 44:483–501.

Pampel, F.C. 1983. Changes in the propensity to live alone: Evidence from consecutive cross-sectional surveys, 1960–1976. Demography 20:433–447.

Parlett, B. 1970. Ergodic properties of populations, I: The one sex model. Theoretical Population Biology 1:191–207.

Pearl, R. 1939. The Natural History of Population. Oxford University Press, New York, New York, USA.

Pearl, R. and L.J. Reed. 1920. On the rate of growth of the population of the United States since 1790 and its mathematical representation. Proceedings of the National Academy of Science 6:275–288.

Perron, O. 1907. Zur Theorie der Matrizen. Mathematische Annalen 64:248–263.

Pinero, D., M. Martinez-Ramos, and J. Sarukhan. 1984. A population model of *Astrocaryum mexicanum* and a sensitivity analysis of its finite rate of increase. Journal of Ecology 72:977–991.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. Biometrics 56:434–442.

Pledger, S., K.H. Pollock, and J.L. Norris. 2003. Open capture-recapture models with heterogeneity. I. Cormack-Jolly-Seber Model. Biometrics 59:786–794.

Pollard, J.H. 1966. On the use of the direct matrix product in analysing certain stochasic population models. Biometrika 53:397–415.

Pollard, J.H. 1973. Mathematical Models for the Growth of Human Populations. Cambridge University Press, Cambridge, UK.

Potter, R.G. 1967. The multiple decrement life table as an approach to the measurement of use effectiveness and demographic effectiveness of contraception. In Contributed Papers, Sydney Conference, pp. 871–883. International Union for the Scientific Study of Population, Sydney, Australia.

Potter, R.G. 1970. Births averted by contraception: An approach through renewal theory. Theoretical Population Biology 1:251–272.

Potter, R.G. 1972. Births averted by induced abortion: An application Of renewal theory. Theoretical Population Biology 3:69–86.

Potter, R.G., and M.P. Parker. 1964. Predicting the time required to conceive. Population Studies 18:99–116.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. 1986. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, Cambridge, UK.

Pressat, R. 1961. L'Analyse Démographique: Méthodes, Résultats, Applications [Demographic Analysis: Methods, Findings, Applications]. Presses Universitaires de France, for L'Institut National d'Etudes Démographiques, Paris, France.

Preston, S.H. 1969. Mortality differentials by social class and smoking habit. Social Biology 16:280–289.

Preston, S.H. 1970. The age-incidence of death from smoking. Journal of the American Statistical Association 65:1125–1130.

Preston, S.H. 1974. Effect of mortality change on stable population parameters. Demography 11:119–130.

Preston, S.H. 1975a. Seminar on Infant Mortality in Relation to the Level of Fertility. Committee for the International Coordination of National Research in Demography (CICRED), Paris, France.

Preston, S.H. 1975b. Estimating proportion of marriages that end in divorce. Sociological Methods and Research 3:435–459.

Preston, S.H. and A.J. Coale. 1982. Age structure, growth, attrition, and accession: A new synthesis. Population Index 48:217–259.

Preston, S.H., N. Keyfitz, and R. Schoen. 1972. Causes of Death: Life Tables for National Populations. Studies in Population Series. Seminar Press, New York, New York, USA.

Price, D.O. and M.M. Sikes. 1975. Rural-Urban Migration Research in the United States: Annotated Bibliography and Synthesis. U.S. Government Printing Office, Washington, DC.

Pullum, T.W. 1982. The eventual frequencies of kin in a stable population. Demography 19:549–565.

Pullum, T.W. and D.A. Wolf. 1991. Correlations between frequencies of kin. Demography 28:391–409.

Rago, P.J. and C.P. Goodyear. 1987. Recruitment mechanisms of striped bass and Atlantic salmon: Comparative liabilities of alternative life histories. American Fisheries Society Symposium 1:402–416.

Ralston, A. 1965. A First Course in Numerical Analysis. New McGraw-Hill, New York, New York, USA.

Rao, C.R. and S.K. Mitra. 1971. Generalized Inverse of Matrices and Its Applications. John Wiley, New York, New York, USA.

Reed, L.J. and M. Merrell. 1939. A short method for constructing an abridged life table. American Journal of Hygiene 30:33–62.

Repetto, R. 1972. Son preference and fertility behavior in developing countries. Studies in Family Planning 3:70–76.

Rhodes, E.C. 1940. Population mathematics, I, II, and III. Journal of the Royal Statistical Society (London) 103:61–89, 218–245, 362–387.

Ridker, R.G., Editor. 1976. Population and Development: The Search for Selective Intervention. Johns Hopkins University Press, Baltimore, Maryland, USA.

Ridley, J.C., et al. 1975. Technical Manual—Reproductive Simulation Model: REPSIM-B. Kennedy Institute Center for Population Research, Georgetown University, Washington, DC, USA. .

Riley, M., et al. 1983. The family in an aging society Journal of Family Issues 4:439–454.

Ripley, B.J. 1998. Life history traits and population processes in marine bivalve molluscs. Ph.D. Dissertation, Woods Hole Oceanographic Institution and Massachusetts Institute of Technology.

Rogers, A. 1966. The multiregional matrix growth operator and the stable interregional age structure. Demography 3:537–544.

Rogers, A. 1968. Matrix Analysis of Interregional Population Growth and Distribution. University of California Press, Berkeley, California, USA.

Rogers, A. 1975. Introduction to Multiregional Mathematical Demography. Wiley, New York, New York, USA.

Rogers, A. 1980. Introduction to multistate mathematical demography. Environment and Planning A 12:489–498.

Rogers, A. 1984. Migration, Urbanization, and Spatial Population Dynamics. Westview Press, Boulder, Colorado, USA.

Rogers, A. 1985. Regional Population Projection Models. Sage Publications, Beverly Hills, California, USA.

Rogers, A. 1995. Multiregional Demography: Principles, Methods and Extensions. Wiley, New York, New York, USA.

Rogers, A. 1995b. Population forecasting: do simple models outperform complex models? Mathematical Population Studies 5:187–202.

Rogers, A. and J. Ledent. 1976. Increment-decrement life tables: A comment. Demography 13:287–290.

Rogers, A. and F. Willekens. 1975. Spatial population dynamics. RR-75-24. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rogers, A. and F. Willekens. 1976. Spatial zero population growth. RM-76-25. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rose, M.R. 1984. The evolution of animal senescence. Canadian Journal of Zoology 62:1661–1667.

Rose, M.R. 1991. Evolutionary Biology of Aging. Oxford University Press, Oxford, UK.

Rosenblatt, D. 1957. On the graphs and asymptotic forms of finite Boolean relation matrices and stochastic matrices. Naval Research Logistics Quarterly 4:151–167.

Ruggles, S. 1993. Confessions of a microsimulator—Problems in modeling the demography of kinship. Historical Methods 26:161–169.

Ryder, N.B. 1963. The translation model of demographic change. In Emerging Techniques in Population Research (39th Annual Conference of the Milbank Memorial Fund, 1962), pp. 65–81. Milbank Memorial Fund, New York, New York, USA.

Ryder, N.B. 1964. The process of demographic translation. Demography 1:74–82.

Ryder, N.B. 1974. The family in developed countries. in The Human Population, pp. 81–88. A Scientific American Book. W. H. Freeman, San Francisco, California, USA.

Ryder, N.B. and C. Westoff. 1967. The trend of expected parity in the United States. Population Index 33:153–168.

Samuelson, P.A. 1976. Resolving a historical confusion in population analysis. Human Biology 48:559–580.

Sanz, L. and R. Bravo de la Parra. 1999. Variables aggregation in a time discrete linear model. Mathematical Biosciences 157:111–146.

Sanz, L. and R. Bravo de la Parra. 2000. Time scales in stochastic multiregional models. Nonlinear Analysis: Real World Applications 1:89–122.

Sauer, J.R. and N. A. Slade. 1985. Mass-based demography of a hispid cotton rat (*Sigmodon hispidus*) population. Journal of Mammology 66:316–328.

Sauer, J.R. and N. A. Slade. 1986. Size-dependent population dynamics of *Microtus ochrogaster*. American Naturalist 127:902–908.

Sauer, J.R. and N. A. Slade. 1987. Uinta ground Squirrel demography: Is body mass a better categorical variable than age? Ecology 68:642–650.

Savage, I.R. 1973. Sociology Wants Mathematics. FSU Statistics Report M-247. Florida State University, Tallahassee, Florida, USA.

Scarborough, J.B. 1958. Numerical Mathematical Analysis. 4th ed. Johns Hopkins Press, Baltimore, Maryland, USA.

Schoen, R. 1970. The geometric mean of the age-specific death rates as a summary index of mortality. Demography 7:317–324.

Schoen, R. 1975. Constructing increment-decrement life tables. Demography 12:313–324.

Schoen, R. 1988. Modeling Multigroup Populations. Plenum Press, New York, New York, USA.

Schoen R. and Y.J. Kim. 1991. Movement toward stability as a fundamental principle of population dynamics. Demography 28:455–466.

Schoen, R. and K.C. Land. 1979. A general algorithm for estimating a Markov-generated increment-decrement life table with application to marital status patterns. Journal of the American Statistical Association 74:761–776.

Schoen, R. and V. Nelson. 1974. Marriage, divorce, and mortality: A life table analysis. Demography 11:267–290.

Schultz, T.P. 1973. A preliminary survey of economic analysis of fertility. American Economic Review 63:71–78.

Schultz, T.W., Editor. 1974. Economics of the Family: Marriage, Children, and Human Capital. University of Chicago Press, Chicago, Illinois, USA.

Schumpeter, J. 1950. Capitalism, Socialism and Democracy. Harper and Row, New York, New York, USA.

Sembiring, R. 1978. Demographic Treatment of Kinship. Ph.D. Dissertation, University of California, Berkeley.

Seneta, E. 1981. Non-Negative Matrices and Markov Chains. 2nd ed. Springer-Verlag, New York, New York, USA.

Sharpe, F.R. and A.J. Lotka. 1911. A problem in age-distribution. Philosophical Magazine 21:435–438.

Shepard, D.S. 1976. Prediction and incentives in health care policy. Ph.D. dissertation, Kennedy School of Government, Harvard University.

Shepard, D. and R. Zeckhauser. 1975. The assessment of programs to prolong life, recognizing their interaction and risk factors. Discussion Paper 32D. Kennedy School of Government, Harvard University.

Shepard, D.S. and R.J. Zeckhauser. 1980. The choice of health policies with heterogeneous populations. Paper presented at NBER Conference on Economic Aspects of Health, Palo Alto, CA. Xerox.

Sheps, M.C. 1959. An examination of some methods of comparing several rates or proportions. Biometrics 15:87–97.

Sheps, M.C. 1963. Effects on family size and sex ratio of preferences regarding the sex of children. Population Studies 17:66–72.

Sheps, M.C. 1964. On the time required for conception. Population Studies 18:85–97.

Sheps, M.C. and J.A. Menken. 1973. Mathematical Models of Conception and Birth. University of Chicago Press, Chicago, Illinois.

Sheps, M.C. and E.B. Perrin. 1963. Changes in birth rates as a function of contraceptive effectiveness: Some applications of a stochastic model. American Journal of Public Health 53:1031–1046.

Sheps, M.C. and J.C. Ridley, Editors. 1966. Public Health and Population Change: Current Research Issues. University of Pittsburgh Press, Pittsburgh, Pennsylvania, USA.

Shryock, H.S. and J.S. Siegel. 1971. The Methods and Materials of Demography. 2 vols. U.S. Bureau of the Census. Government Printing Office, Washington, DC.

Siegel, J.S. and D.S. Akers. 1969. Some aspects of the use of birth expectations data from sample surveys for population projections. Demography 6:101–115.

Silva, J.F., J. Raventos, H. Caswell, and M.C. Trevisan. 1991. Population responses to fire in a tropical savanna grass *Andropogon semiberbis*: A matrix model approach. Journal of Ecology 79:345–356.

Simpson, E.H. 1951. The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society, Series B13:238–241.

Singer, B. and S. Spilerman. 1976. The representation of social processes by Markov models. American Journal of Sociology 82:1–54.

Sinko, J.W. and W. Streifer. 1967. A new model for age-size structure of a population. Ecology 48:910–918.

Sinko, J.W. and W. Streifer. 1969. Applying models incorporating age-size structure of a population to *Daphnia*. Ecology 50:608–615.

Sivamurthy, M. 1971. Convergence of age–sex distributions and population change in the presence of migration. Doctoral thesis, Institute of Advanced Studies, Australian National University, Canberra.

Smith, D. 1974. Generating functions for partial sex control problems. Demography 11:683–689.

Smith, D. and N. Keyfitz, Editors. 1977. Mathematical Demography: Selected Readings Springer-Verlag, New York, New York, USA.

Spiegelhalter, D.J., A. Thomas, and N.G. Best 1999. WinBUGS Version 1.2 User Manual, MRC Biostatistics Unit, Cambridge, United Kingdom.

Spiegelman, M. 1968. Introduction to Demography. Revised ed. Harvard University Press, Cambridge, Massachusetts, USA.

Spurgeon, E.F. 1932. Life Contingencies. Cambridge University Press, Cambridge, UK.

Stearns, S.C. and J.C. Koella. 1986. The evolution of phenotypic plasticity in life-history traits: Predictions of reaction norms for age and size at maturity. Evolution 40:893–913.

Stone, L.O. 1971. On the correlation between metropolitan area in- and out-migration by occupation. Journal of the American Statistical Association 66:693–701.

Stone, R. 1972. A Markovian education model and other examples linking social behavior to the economy. Journal of the Royal Statistical Society, Series A 135:511–543.

Stone, R. 1975. Towards a System of Social and Demographic Statistics. United Nations Statistical Office, Studies in Methods, Series F, No. 18. Sales No. E. 74, XVII. 8. United Nations, New York, New York, USA.

Stouffer, S.A. 1940. Intervening opportunities: A theory relating mobility and distance. American Sociological Review 5:845–867.

Süssmilch, J.P. 1788 (1941). Die göttliche Ordnung in den Veränderungen des menschilichen Geschlechts, aus der Geburt, dem Tode und der Fortpflanzung desselben erwiesen. 3 volumes. Verlag der Buchhandlung der Realschule, Berlin, Germany.

Sverdrup, E. 1965. Estimates and test procedures in connection with stochastic models for deaths, recoveries and transfers between different states of health. Skandinavisk Aktuarietidskrift 40:184–21.

Svirezhev, Y.M. and D.O. Logofet. 1983. Stability of Biological Communitites. Mir Publishers, Moscow, Russia.

Sykes, Z.M. 1969. Some stochastic versions of the matrix model for population dynamics. Journal of the American Statistical Association 64:111–130.

Tabah, L. 1968. Representations matricielles de perspectives de population active [Matrix representations of projections of active population]. Population 23:437–476.

Taylor, F. 1979. Convergence to the stable age distribution in populations of insects. American Naturalist 113:511–530.

Thieme, H.R. 2003. Mathematics in Population Biology. Princeton University Press, Princeton, New Jersey, USA.

Thompson, J.N. and M.E. Moody. 1985. Assessing probability of interaction in size-structured populations:*Depressaria* attack on *Lomatium*. Ecology 66:1597–1607.

Thompson, W.R. 1931. On the reproduction of organisms with overlapping generations. Bulletin of Entomological Research 22:147–172.

Thonstad, T. 1967. A mathematical model of the Norwegian educational system. pp. 125–158 in Mathematical Models in Educational Planning. Technical Report. Organization for Economic Co-operation and Development, Paris, France.

Tibshirani, R. 1992. Bootstrap hypothesis testing. Biometrics 48:969–970.

Tietze, C. 1962. Pregnancy rates and birth rates. Population Studies 16:31–37.

Todaro, M.P. 1969. A model of labor migration and urban unemployment in less developed countries. American Economic Review 59:138–148.

Trussell, J.T. 1977. Determinants of roots of Lotka's equation. Mathematical Biosciences 36:213-227.

Tuljapurkar, S.. 1982. Why use population entropy? It determines the rate on convergence. Journal of Mathematical Biology 13:325–337.

Tuljapurkar, S.. 1985. Population dynamics in variable environments. VI. Cyclical environments. Theoretical Population Biology 28:1–17.

Tuljapurkar, S. 1990. Population Dynamics in Variable Environments. Springer-Verlag, New York, New York.

Tuljapurkar, S. 1993. Entropy and convergence in dynamics and demography. Journal of Mathematical Biology 31:253–271.

Tuljapurkar, S. and H. Caswell. 1997. Structured Population Models in Marine, Terrestrial, and Freshwater Systems. Chapman and Hall, New York, New York, USA.

Tuljapurkar, S. and R. Lee. 1997. Demographic uncertainty and the stable equivalent population. Mathematical and Computer Modelling 26:39–56.

Tuljapurkar, S.D. and S.H. Orzack. 1980. Population dynamics in variable environments I. Long-run growth rates and extinction. Theoretical Population Biology 18:314–342.

Tuljapurkar, S., N. Li, and M.W. Feldman. 1995. High sex ratios in China's future. Science 267: 874-876.

Trussell, J. and G. Rodríguez. 1990. Heterogeneity in demographic research. Pages 111–132 in J. Adams, D.A. Lam, A.I. Hermalin, and P.E. Smouse. Convergent Issues in Genetics and Demography. Oxford University Press, Oxford, UK.

United Nations. 1955. Age and Sex Patterns of Mortality: Model Life Tables for Underdeveloped Countries. United Nations, New York, New York, USA.

United Nations. 1963. Determinants and Consequences of Population Trends. United Nations, New York, New York, USA.

Usher, M.B. 1966. A matrix approach to the management of renewable resources, with special reference to selection forests. Journal of Applied Ecology 3:355–367.

Usher, M.B. 1976. Extensions to models, used in renewable resource management, which incorporate an arbitrary structure. Journal of Environmental Management 4:123–140.

van Sickle, J. 1977a. Analysis of a distributed-parameter population model based on physiological age. Journal of Theoretical Biology 64:571–586.

van Sickle, J. 1977b. Mortality rates from size distributions. Oecologia 27:311–318.

van Straalen, N.M. and J.E. Kammenga. 1998. Assessment of ecotoxicity at the population level using demographic parameters. Pages 621–644 in G. Schüürmann and B. Markert, editors. Ecotoxicology. Wiley, New York, New York, USA.

Vaupel, J.W. 1990. Kindred lifetimes: Frailty models in population genetics. Pages 155–170 in J. Adams, D.A. Lam, A.I. Hermalin, and

P.E. Smouse. Convergent Issues in Genetics and Demography. Oxford University Press, Oxford, UK.

Vaupel, J.W., and A.I. Yashin. 1983. The Deviant Dynamics of Death in Heterogeneous Populations. Research Report RR-83-1. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Vaupel J.W. and A.I. Yashin. 1985. Heterogeneity's ruses: Some surprising effects of selection on population dynamics. American Statistician. 39:176-85.

Vaupel, J.W. and A.I. Yashin. 1999. Unobserved population heterogeneity. pp. 345–358 in Démographie: Analyse et synthèse. Volume 1. Dipartimento di Science Demografiche, Università degli Studi di Roma "La Sapienza"; Institut national d'études démographiques, Rome, Italy.

Vaupel, J.W., K. Manton, and E. Stallard. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16:439–454.

Vaupel, J.W. et al. 1998. Biodemographic trajectories of longevity. Science 280:855–860.

Veltman, C.J., S. Nee, and M.J. Crawley. 1996. Correlates of introduction success in exotic New Zealand birds. American Naturalist 147:542–557.

Verhulst, P.F. 1838. Notice sur la loi que la population suit dans son accroissement. Correspondence Mathématique et Physique Publiée par A. Quételet (Brussels) 10:113–121.

Vincent, P. 1945. Potential d'accroissement d'une population stable. Journal de la Société de Statistique de Paris 86:16–29.

von Foerster, H. 1959. Some remarks on changing populations. Pages 382–407 in F. Stohlman, Jr., editor. The Kinetics of Cellular Proliferation. Grune and Stratton, New York, New York, USA.

von Foerster, H., P.M. Mora, and L.W. Amiot. 1960. Doomsday: Friday, 13 November, A.D. 2026. Science 132:1291–1295.

Wachter, K.W. 1991. Elusive cycles: Are there dynamically possible Lee-Easterlin models for U.S. births? Population Studies 45:109–135.

Wachter, K.W. 1997. Kinship resources for the elderly. Philosophical Transactions of the Royal Society of London B 352:1811–1817.

Wachter, K.W. and R.A. Bulatao. 2003. Offspring: Human Fertility Behavior in Biodemographic Perspective. National Academies Press, Washington, DC, USA.

Wachter, K.W. and C.E. Finch. 1997. Between Zeus and the Salmon: The Biodemography of Longevity. National Academy Press, Washington, D.C., USA.

Wachter, K.W. and R.D. Lee. 1989. U.S. births and limit cycle models. Demography 26:99–115.

Wachter, K.W., D. Blackwell, and E.A. Hammel. 1997. Testing the validity of kinship microsimulation. Journal of Mathematical and Computer Modeling 26:89–104.

Wagner, C.H. 1982. Simpson's paradox in real life. The American Statistician 36:46–48.

Walls, M., H. Caswell, and M. Ketola. 1991. Demographic costs of *Chaoborus*-induced defenses in *Daphnia pulex*: A sensitivity analysis. Oecologia 87:43–50.

Warner, R.R. and T.P. Hughes. 1988. The population biology of reef fishes. Proceedings of the Sixth International Coral Reef Symposium 1:149–155.

Waugh, W.A.O. 1971. Career prospects in stochastic social models with time-varying rates. Fourth Conference on the Mathematics of Population. East-West Population Institute, Honolulu, Hawaii.

Weck, F.A. 1947. The mortality rate and its derivation from actual experience. Record of the American Institute of Actuaries 36:23–54.

Weinberg, C.R. and D.B. Dunson. 2000. Some issues in assessing human fertility. Journal of the American Statistical Association 95:300–303.

Werner, P.A. and H. Caswell. 1977. Population growth rates and age versus stage-distribution models for teasel (*Dipsacus sylvestris* Huds.). Ecology 58:1103–1111.

Westoff, C.F. and R.R. Rindfuss. 1974. Sex preselection in the United States: Some implications. Science 184:633–636.

Westoff, C.F., R.G. Potter, Jr., and P.C. Sagi. 1963. The Third Child: A Study in the Prediction of Fertility. Princeton University, Office of Population Research, Princeton, New Jersey, USA.

Whelpton, P.K. 1936. An empirical method of calculating future population. Journal of the American Statistical Association 31:457–473.

Whelpton, P.K. 1946. Reproduction rates adjusted for age, parity, fecundity, and marriage. Journal of the American Statistical Association 41:501–516.

Wicksell, S.D. 1931. Nuptiality, fertility, and reproductivity. Skandinavisk Aktuarietidskrift (Stockholm), pp. 125–157.

Widjojo N. 1970. Population Trends in Indonesia. Ithaca, N.Y.: Cornell University Press.

Wilkinson, J.H. 1965. The Algebraic Eigenvalue Problem. Clarendon Press, Oxford, UK.

Wilkinson, J.H. and C. Renisch. 1971. Linear Algebra. Springer-Verlag, New York, New York, USA.

Willekens, F. 1978. The Demography of Labor Force Participation. Research memorandum RM-78-17. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Willekens, F. and A. Rogers. 1977. More Computer Programs for Spatial Demographic Analysis. Research memorandum RM-77-30. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Williams, G.C. 1957. Pleiotropy, natural selection, and the evolution of senescence. Evolution 11:398–411.

Williams, G.C. 1966. Natural selection, the costs of reproduction, and a refinement of Lack's principle. American Naturalist 100:687–690.

Wilson, C. 2004. Fertility below replacement level. Science 304:207–209.

Wilson, E.B. and R.R. Puffer. 1933. Least squares and population growth. Proceedings of the American Academy of Arts and Sciences 68:285–382.

Wilson, E. O. 1975. Sociobiology: The New Synthesis. Cambridge, Mass.: Harvard University Press.

Winkler, W. 1959. Wieviele Menschen haben bisher auf der Erde gelebt? International Population Conference, Vienna, 1959, pp. 73–76. International Union for the Scientific Study of Population.

Winsor, C.P. 1932. A comparison of certain symmetrical growth curves. Journal of the Washington Academy of Sciences 22:73–84.

Wolf, D.A. 2001. The role of microsimulation in longitudinal data analysis. Canadian Studies in Population 28:313–339.

Wolf, S.L., H.X. Barnhart, et al. 1996. Reducing frailty and falls in older persons: An investigation of tai chi and computerized balance training. Journal of the American Geriatric Society 44:489.

Wolfenden, H. H. 1925 (rev. ed. 1954). Population Statistics and Their Compilation. University of Chicago Press, Chicago, Illinois, USA.

Wood, J.W. 1994. Dynamics of Human Reproduction: Biology, Biometry, Demography. de Gruyter, New York, New York, USA.

Wood, J.W. and M. Weinstein. 1990. Heterogeneity in fecundability: The effect of fetal loss. Pages 171–188 in J. Adams, D.A. Lam, A.I. Hermalin, and P.E. Smouse (editors) Convergent Issues in Genetics and Demography. Oxford University Press, New York, New York, USA.

Wood, J.W., D.J. Holman, K.M. Weiss, A.V. Buchanan, and B. Lefor. 1992. Hazards models for human population biology. Yearbook of Physical Anthropology 35:43–87.

Wood, J.W., D.J. Holman, A.I. Yashin, R.S. Peterson, M. Weinstein, and M.-C. Chang. 1994. A multistate model of fecundability and sterility. Demography 31:403–426.

Wu, L.S.Y. and D.B. Botkin. 1980. Of elephants and men—a discrete, stochastic model for long-lived species with complex life histories. American Naturalist 116:831–849.

Wunsch, G. 1966. Courbes de Gompertz et perspectives de fécondité. Recherches Economiques (Louvain) 32:457–468.

Yashin, A.I. and I.A. Iachine. 1997. How frailty models can be used for evaluating longevity limits: Taking advantage of an interdisciplinary approach. Demography 34:31–48.

Yashin, A.I., J.W. Vaupel, and I.A. Iachine. 1994. A duality in aging: The equivalence of mortality models based on radically different concepts. Mechanisms of Ageing and Development 74:1–14.

Yashin, A.I., I.A. Iachine, K.F. Andreev, and U. Larsen. 1998. Multistate models of postpartum infecundity, fecundability and sterility by age and parity: Methodological issues. Mathematical Population Studies 7:51–78.

Yule, G. U. 1906. Changes in the marriage and birth rates in England and Wales during the past half century. Journal of the Royal Statistical Society 68:18–132.

Zelinsky, W. 1971. The hypothesis of the mobility transition. Geographical Review 46:219–249.

Index