MATHEMATICAL DEMOGRAPHY

Vladimir Canudas-Romo
Johns Hopkins Bloomberg School of Public Health, USA.

Kenneth C. Land
Duke University, USA.

Yang Yang
University of Chicago, USA.

Zeng Yi
Duke University, USA; Peking University, China

Keywords: Mathematical demography, formal demography, models, momentum, theory, life table, decomposition, model life tables, stable population, APC models, frailty, heterogeneity, projections, indirect estimation

Contents

Summary

Mathematical demography focuses on population phenomena and their relations with other population phenomena. Its subject is not facts as such, but how to handle them. The methods and tools used in the analysis of population phenomena are continually changing, and as a consequence constantly improving our knowledge of population dynamics. New mathematical developments have occurred in the measures used to describe life tables, decomposition methods, model life tables, models of age-specific rates, and indirect methods of estimation. Changes in mathematical demography also include revisions and updates to the basic definitions of model and theory in the context of demography. This chapter reviews some of the latest developments in the large body of mathematical theory concerned with the growth processes of populations. The topics covered were selected on the basis of elements of mathematical demography that have not been addressed in the rest of the volume. This assessment of the state of the art in mathematical demography complements the rest of the contributions of this special volume on demography. As described in this chapter, mathematical demography remains central to the discipline as a complete body of methods, models, and theories that are used within many different areas of demography.

1. Introduction

Demography is found at the crossroad of the social sciences and biological sciences, and it stands on a firm base of statistics and mathematics foundations. At the heart of demography lies Mathematical Demography. This area of demography accommodates the formalization, through mathematical representations, of relations of population status and changes. Most formulas are straightforward and explain the fundamental relations existing among demographic variables.

The development of many of the equations and theories found in mathematical demography can be traced to such diverse fields of study as actuarial sciences, biodemography, population projections, statistics, and more. The topics covered in a survey of mathematical demography can be numerous and diverse. Therefore, any survey of the field in the confines of a chapter must be selective.

This chapter focuses on describing the elements of mathematical demography that are key foci in this field of study. The chapter is divided into eight major sections, including this introduction. Section 2 discusses the basic elements of mathematical demography: functions, models, and theory. These basic elements are used repeatedly in each of the remaining sections of the chapter, as well as in other chapters. Therefore, it is crucial to set the definitions of these key elements of mathematical demography. Section 3 presents one of the basic models of demography, the life table model, with some of its most important measures. New aspects of the life table which have captured the attention of demographers are examined in that section. Thus the section complements the detailed construction of life tables described in *Demographic Models and Actuarial Science* by Robert Schoen.

The following three sections further explain developments that are related to the life table measures. The fourth, fifth, and sixth sections present decomposition methods, model life tables, and extensions of the stable population theory, respectively. Decomposition methods refer to the procedure of separating demographic phenomena into its constituent parts. Most commonly these methods are used to study change in demographic variables and the contribution of the different components of this change. In the section on model life tables, the available ways of estimating mortality, and complete life tables, for regions with deficient data are reviewed. The section on extensions of stable population theory presents the mathematical relations that connect all major demographic functions in any population: births, deaths and population growth. As shown in this section the set of age-specific growth rates links all of the relationships.

The chapter concludes with sections dedicated to modeling vital events, and indirect methods of estimation. Model schedules, age-period-cohort models, frailty models and considerations in doing population projections are reviewed in the section on modeling of vital events. The section on indirect methods of estimation includes a brief revision of two of the most prominent methods used to approximate child and adult mortality.

2. Models in Demography

This section reviews the theoretical concepts of models as studied since the 1970s by Land and in the 1990s by Casti. Although the section focuses on demographic models and theories, the definitions given here have general use across the sciences. The objective is to position the models studied in mathematical demography within this general framework. A broad discussion on the nature of demographic phenomena and data introduces this section. Sections 2.2 and 2.3 go into depth to definitions of models and formal models respectively.

2.1. The Nature of Demographic Phenomena and Data

Recent interest in bringing together demographers and epidemiologist has also attracted the attention of these scientists to define explicitly their areas of research. In their discussion of the evolution of demography into a cumulative and integrated science, Morgan and Lynch identified four factors intrinsic or internal to the discipline that have facilitated this evolution. They are summarized here because they may be viewed as the basis on which the extraordinary development and application of mathematical models in demography has proceeded.

i) The first is the fact that demographic phenomena are relatively easily measured. Births and deaths, the core events studied in demography, are biologically based and thus are anchored in an unmistakable and universal reality.

ii) A second key feature is that demographic phenomena are inherently quantifiable. This is because births and deaths are categorical (in fact, they are dichotomous) and thus easily counted. Intermediate instances of birth and death are few and rare. The consequence is that repeated measurement and intersubjective agreement among observers would likely be high. This is not to say that demographic measurement is easy for a large population – only that it is relatively straightforward.

iii) Third, the presence of accounting identities has facilitated the successful development of demography as a science. Traditional demography focuses on the description of the composition of human populations by age, sex, and other characteristics and the study of change therein (dynamics). The basic methods of demography are centered on the well-known population accounting or balancing equation. In this equation, the population counts at any time can be derived given a population count earlier in time and the births, deaths, and net migrations that occur between the two times. Using this identity, demographers can perform quality checks on their data and engage in indirect estimation when only fragments of data are available.

iv) The fourth factor is the presence of structural features or relationships among key concepts. Births and deaths can be unambiguously identified and counted in human populations, and they can be used to define populations at risk of one event or another and the corresponding dynamics or rates of occurrence of the events. These, in turn, can be used to build life table/stationary population models, stable population models, and related models to describe and/or explain the corresponding population processes.

Models are the basic elements of mathematical demography and their formal definition is reviewed in Sections 2.2 and 2.3.

2.2. Introduction to Models

Two initial definitions for models follow the concepts studied by Casti. *Models* are tools by which individuals order and organize experiences and observations. This is a broad definition including a vast number of demographic verbal characterizations used to describe population phenomena.

A second definition corresponds to *formal models*, defined as representations of the experiences and observations within the relationships constituting a formal system such as formal logic, mathematics, or statistics. As such, *formal demographic models* are a way of representing aspects of demographic phenomena using formal apparatuses that provide a means for exploring the properties of the demographic phenomena mirrored in the model.

Keyfitz and Land list some of the returns of using models: Models focus research by identifying theoretical and practical issues. Models help in assembling and explaining data. Models permit the design of experiments, simulations, and other research studies out of which causal knowledge can be obtained. Models systematize comparative study across space and time. Models reveal formal analogies between problems that on their surface are quite different. Models help in making predictions. Models provide a "lens" through which patterns can be detected in demographic data that otherwise cannot be perceived. Models help to improve demographic measurement. And, in particular, models provide a locus for defining, developing, and interpreting summary measures of demographic events and phenomena.

2.3. Mathematical Demography Models

Some notation is useful to explicitly describe a formal demographic model. The two basic elements to describe a model are the state space and the corresponding observable:

- *State space* $\Omega = \{\omega_1, \omega_2, ...\}$: all the possible combinations of abstract states of a demographic phenomenon to be modeled.

- *Observable* f of a demographic phenomenon: associating a real number to each ω in the state space Ω, i.e., an observable is a measuring instrument or map $f : \Omega \to \mathbb{R}$, where $\mathbb{R}$ denotes the set of real numbers.

A *formal demographic model* D may now be characterized as an abstract state-space Ω together with a finite set of observables $f_i : \Omega \to \mathbb{R}$, $i = 1, 2, \ldots, n$. Symbolically,

$$D = \{\Omega, f_1, f_2, ..., f_n\}.$$

Several points should be mentioned. Whether or not a demographer can determine the state-space in a particular moment of study depends on the experiences, observations, or measurements (*observables*) at the demographer's disposal. Generally speaking, a full account of the complexities of population phenomena would require an infinite number of observables $f_\alpha : \Omega \to \mathbb{R}$, where the subscript α ranges over a possibly uncountable index set.

However, it is often unnecessary in demography to identify such large sets of observables to build useful demographic models. Normally, demographers focus their attention on *a proper subset* of observables that may or may not capture the full complexity of demographic phenomena.

A formal model, as defined above, is still incomplete if the relationships that link the different observables are not known. A further concept is needed for this:

- *Equations of the state-space* $\Phi_i(.)$: are the mathematical relationships expressing the dependency relations among the observables. Formally, the equations of state can be written as

$$\Phi_i(f_1, f_2, ..., f_n) = 0, \qquad i = 1, 2, \ldots, m \tag{1}$$

There are two forms in which the equations of state that define demographic models are used. The first is to state relationships among observables that produce a *descriptive model or definition* of a demographic index or rate. Many demographic models, from population life tables used to estimate the years of expected life in a population to a formula for calculating the total fertility rate, are used in this descriptive or definitional sense. A second way in which the equations of state are used in demography is to state causal relationships among observables that produce an *explanatory model* in which variations across time or demographic space in certain observables are explained by variations in other observables.

To represent explanatory demographic models, Eq. (1) must be further developed. Suppose that the last m observables $f_{n-m+1},\ldots, f_n$, called *endogenous* (i.e., determined within the system under consideration), are functions of the remaining observables $f_1, f_2,\ldots, f_{n-m}$, which are *exogenous* (i.e., determined outside the system under consideration). In other words, suppose that *m* functional relations are defined, with some finite number r of numerical parameters, $\beta_1,\beta_2,...,\beta_r$, for determining values of the endogenous observables as a function of the exogenous observables. Here the notation

$$\beta \equiv (\beta_1,\beta_2,...,\beta_r)$$

denotes the vector of parameters and the notation

$$x \equiv (f_1, f_2,..., f_{n-m})$$

and

$$y \equiv (f_{n-m+1}, f_{n-m+2},..., f_n)$$

denotes vectors of the exogenous and endogenous observables, henceforth *variables*, respectively. The equations of state become

$$y = \Phi\beta(x) \qquad (2)$$

This last expression, perhaps formulated with stochastic/random components, is in a form that encompasses many explanatory demographic models.

As an example of the application of these formal notations to the definition of a common demographic index, consider the mathematical model that underlies the definition of a common demographic measure of mortality, life expectancy at birth, $e_0(t)$. Period life expectancy at birth is defined as the average number of years that a person born at time *t* would live given the death rates observed in that year.

Formally, let $\mu(a,t)$ denote the force of mortality at age a at time t. The equation of state (1) for the *period life expectancy at birth* for time t is

$$e_0(t) = \int_0^{\omega} e^{-\int_0^x \mu(a,t)da} dx \qquad (3)$$

where ω is the last age attained in the population.

The final concept introduced in this section is the definition of Theory:

- A *theory* is a *family of related models*, and a *model* is a formal manifestation of a particular theory.

In this sense, scientific theories are regarded as more general than scientific models. For instance, the life table is the simplest mathematical demography theory of the age structure of a population, called a stationary population, which is subjected to certain patterns of fertility and mortality interrelated by a set of mathematical functions.

-
-
-

Bibliography

Alho J.M. and Spencer B.D. (2005). *Statistical Demography and Forecasting.* Springer 410 pages. [A unified statistical outlook on both classical methods of demography and recent developments.]

Coale, A.J. and P.Demeny. (1983). *Regional Model Life Tables and Stable Populations* (2d Ed). New York: Academic Press. [The collection of the model life tables which are extensively used, currently with the availability of computers many of the calculations are done in programs.]

Preston, S.H., P. Heuveline, and M. Guillot. (2001). *Demography: Measuring and Modeling Population Processes*, Oxford: Blackwell Publishers. [The preferred textbook for teaching demographic methods.]

Casti, J. L. (1992). *Reality Rules: I Picturing the World in Mathematics—The Fundamentals*. Wiley-Interscience.

Casti, J. L. (1992). *Reality Rules: II Picturing the World in Mathematics—The Frontier*. Wiley-Interscience.[This book (2 volumes) explores the syntax and semantics of the language of mathematical modeling that researchers use today to frame their views of reality.]

Schoen, R. (1988). *Modeling Multigroup Populations*, New York: Plenum [A detailed exposition of multistate life table construction and its applications.]

Schoen, R. (2007). *Dynamic Population Models*. New York: Springer. [A range of demographic topics of current interest to population scientists.]

Siegel J.S. and Swanson D.A. (2004). T*he Methods and Materials of Demography.* Elsevier. [A revised and improved version of a classic book on demographic methods.]

Smith D. and Keyfitz N. (1977). *Mathematical Demography: Selected Papers.* Berlin: Springer. [A substantial collection of the main sections of some of the classic papers in population studies.]

United Nations. (1983). Manual X: Indirect Techniques for Demographic Estimation. New York: United

Nations. [A manual created for evaluation and exploitation of data sources, especially those that are incomplete or deficient. Each of the techniques presented is based on a mathematical model and explained in easy-to-follow examples.]

United Nations (2002). Methods for Estimating Adult Mortality. Department of Economic and Social Affairs, Population Division. ESA/P/WP.175. [This manual represents the state of the art on indirect methods of demographic estimation of adult mortality and a complement to the previous Manual X: Indirect Techniques for Demographic Estimation, 1983.]

Biographical Sketches

Vladimir Canudas-Romo received the degree in Actuarial Sciences from the National Autonomous University of Mexico UNAM, the Master in Population from the Latin-America Faculty of Social Science FLACSO, and the Ph.D. in 2003 from join collaboration between the University of Groningen, the Netherlands, and the Max Planck Institute for Demographic Research, Germany. He has been a faculty member at the Department of Population, Family and Reproductive Health at the Johns Hopkins Bloomberg School of Public Health since 2007. He collaborated with the Human Mortality Database project at the Department of Demography in the University of California, Berkeley, from 2005 to 2007. He was awarded the DeWitt Wallace award from Population Council to work at the Pennsylvania State University for two years from 2003 to 2005. His interests include decomposition methods, life tables and multistate life tables, model life tables, and tempo effects.

Kenneth C. Land is the John Franklin Crowell Professor of Sociology and Demography at Duke University. He is an elected Fellow of the American Statistical Association and was the 1997 recipient of the Paul F. Lazarsfeld Award from the Methodology Section of the American Sociological Association. His research interests are in the development of mathematical and statistical models and methods for substantive applications in demography, criminology, and social indicators/quality-of-life studies.

Yang Yang received the B.A. degree in 1998 from Beijing University, the M.S. degree in Statistics in 2004, and the Ph.D. degree in Sociology and Demography from Duke University in 2005. She has been Assistant Professor of Sociology and Research Associate at the Population Research Center and Center on Aging at NORC at the University of Chicago since 2005. Her research interests include mathematical demography, medical sociology, sociology of aging and the life course, new methods for cohort analysis, and Bayesian statistical methods.

Zeng Yi is a Professor at the Center for Study of Aging and Human Development and Geriatric Division / Dept of Medicine of Medical School, and Institute of Population Research and Dept. of Sociology, Duke University. He is also a Professor at the China Center for Economic Research of Peking University in China, and Distinguished Research Scholar of the Max Planck Institute for Demographic Research (MPIDR) in Germany. He received his doctoral degree from Brussels Free University in May 1986, and conducted post-doctoral study at Princeton University in 1986-87. Up to Aug. 2007, he has had 79 professional articles written in English published in academic journals or as book chapters in the United States and Europe; among them, 51 articles were published in anonymous, peer-reviewed academic journals. He has had 84 professional articles written in Chinese and published in China; among them, 54 articles were published in national top Chinese academic journals. He has published sixteen books, including five research books (as first author), such as “Family Dynamics in China,” published by the University of Wisconsin Press; one textbook on demographic methods (as the sole author); two volumes of demographic software and user’s manuals (as the first author) on family status life table analysis; six edited books (four as the chief editor, and two as the second editor), such as the 2005 and 2008 books published by Springer for which he served as the chief editor. Six of Zeng Yi’s published books were written in English, one was written in both Chinese and English, and the remainders were written in Chinese.

Zeng Yi has been awarded more than ten national and international academic prizes, such as the Dorothy Thomas Prize of the Population Association of America, the Harold D. Lasswell Prize in Policy Science awarded by the international journal *Policy Sciences* and Kluwer Academic Publishers, the second-class prize for advancement of science and technology awarded by the State Sciences and Technology Commission of China, the first-class prize for advancement of science and technology awarded by the State Education Commission, and the highest academic honor of Peking University: "Prize for

Outstanding Contributions in Sciences."

According to the search report, up to Feb. 2, 2007, the internationally most important literature sources SSCI (Social Science Citation Index) and SCI (Science Citation Index), published in the U.S., indicate that Zeng Yi's articles and books have been cited in 696 journal articles by authors other than Zeng Yi. Among them, 416 citations refer to the work of Zeng Yi as the first author; 280 citations refer to the work of Zeng Yi as a co-author. Zeng Yi is one of the authors of "High Impact Papers" worldwide in the period of 1981 -1998, as announced by International Scientific Institute (ISI) in September, 2000.