Modelling Mortality with Actuarial Applications

Actuaries have access to a wealth of individual data in pension and insurance portfolios, but rarely use its full potential. This book will pave the way, from methods using aggregate counts to modern developments in survival analysis.

Based on the fundamental concept of the hazard rate, Part One shows how and why to build statistical models based on data at the level of the individual persons in a pension scheme or life insurance portfolio. Extensive use is made of the R statistics package. Smooth models, including regression and spline models in one and two dimensions, are covered in depth in Part Two. Finally, Part Three uses multiple-state models to extend survival models beyond the simple life/death setting, and includes a brief introduction to the modern counting process approach.

Practising actuaries will find this book indispensable and students will find it helpful when preparing for their professional examinations.

ANGUS S. MACDONALD is Professor of Actuarial Mathematics at Heriot-Watt University, Edinburgh. He is an actuary with much experience of modelling mortality and other life histories, particularly in connection with genetics, and as a member of Continuous Mortality Investigation committees.

STEPHEN J. RICHARDS is an actuary and principal of Longevitas Ltd., Edinburgh, a software and consultancy firm that uses many of the models described in this book with life insurance and pension scheme clients worldwide.

IAIN D. CURRIE is an Honorary Research Fellow at Heriot-Watt University, Edinburgh. As a statistician, he was chiefly responsible for the development of the spline models described in this book, and their application to actuarial problems.

INTERNATIONAL SERIES ON ACTUARIAL SCIENCE

The *International Series on Actuarial Science*, published by Cambridge University Press in conjunction with the Institute and Faculty of Actuaries, contains textbooks for students taking courses in or related to actuarial science, as well as more advanced works designed for continuing professional development or for describing and synthesising research. The series is a vehicle for publishing books that reflect changes and developments in the curriculum, that encourage the introduction of courses on actuarial science in universities, and that show how actuarial science can be used in all areas where there is long-term financial risk.

A complete list of books in the series can be found at www.cambridge.org/statistics. Recent titles include the following:

Claims Reserving in General Insurance
David Hindley

Financial Enterprise Risk Management (2nd Edition)
Paul Sweeting

Insurance Risk and Ruin (2nd Edition)
David C.M. Dickson

Predictive Modeling Applications in Actuarial Science, Volume 2: Case Studies in Insurance
Edited by Edward W. Frees, Richard A. Derrig & Glenn Meyers

Predictive Modeling Applications in Actuarial Science, Volume 1: Predictive Modeling Techniques
Edited by Edward W. Frees, Richard A. Derrig & Glenn Meyers

Computation and Modelling in Insurance and Finance
Erik Bølviken

MODELLING MORTALITY WITH ACTUARIAL APPLICATIONS

ANGUS S. MACDONALD
Heriot-Watt University, Edinburgh

STEPHEN J. RICHARDS
Longevitas Ltd, Edinburgh

IAIN D. CURRIE
Heriot-Watt University, Edinburgh

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107045415
DOI: 10.1017/9781107051386

First published 2018

Printed in the United Kingdom by Clays, St Ives plc

A catalogue record for this publication is available from the British Library.

ISBN 978-1-107-04541-5 Hardback

Contents

Preface

This book brings modern statistical methods to bear on practical problems of mortality and longevity faced by actuaries and analysts in their work for life insurers, reinsurers and pension schemes. It will also be of interest to auditors and regulators of such entities. The following is a list of questions on demographic risks which this book will seek to answer. Practising actuaries will recognise many of them from their daily work.

- Insurance portfolios and pension schemes often contain substantial amounts of individual data on policyholders and beneficiaries. How best can this information be used to manage risk? How do you get the greatest value from your own data?
- Historically, actuarial work modelled mortality rates for grouped data. Does this have drawbacks, and is there a better way of modelling risk? Are there models which recognise that it is individuals who experience insurance events, rather than groups?
- In many markets insurers need to find new risk factors to make their pricing more competitive. How do you know if a risk factor is significant or not? And is a risk factor statistically significant, financially significant, both or neither?
- Even in the very largest portfolio, combinations of some risk factors can be relatively rare. How can you build a model that handles this?
- How do you choose between models?
- Some portfolios are exposed to different modes of exit. An example is a term-insurance portfolio where a policy can lapse, or result in a death claim or a critical-illness claim. How do you build a model when there are competing risks?
- Many modern regulatory frameworks are explicitly statistical in nature, such as the Solvency II regime in the European Union. How do you perform your

analysis in a way that meets the requirements of such regulations? In particular, how do you implement the "events occurring in one year" requirement for risks that are long-term by nature?

- How have mortality rates in your portfolio changed over time? How do you separate time-based trends from changes in the composition of the portfolio?
- After you have built a model, what uncertainty lies over your fitted rates? How do you measure mis-estimation risk in a multi-factor model? And how do you measure mis-estimation risk in terms of the financial consequences for a given portfolio of liabilities?
- The future path of mortality rates is unknown, yet a pension scheme is committed to paying pensions for decades. How do you project mortality rates? How do you acknowledge the uncertainty over the projection?
- How can the analyst or advisor working for a small pension scheme convince a lay audience that a statistical model fits properly?

This book aims to provide answers to these and other questions. While the book is of immediate application to practising actuaries, auditors and regulators, it will also be of interest to university students and provide supplementary reading for actuarial students.

When we refer to "modern" statistical methods, we mean survival modelling. As a sub-discipline of statistics, it developed from the 1950s onwards, mostly with clinical questions in mind, and actuaries paid it little attention. It followed the modern (by then) statistical paradigm, as follows:

- Specify a probabilistic model as a plausible description of the process generating the observations.
- Use that model to deduce the statistical properties of what can be observed.
- Use those statistical properties to test hypotheses, measure goodness-of-fit, choose between alternative models and so on.

Two major themes have emerged in survival modelling in the past 40 years. The first and most obvious is the arrival of cheap computing power and statistics packages. All major statistics packages can now handle the survival models useful in medical research (although not usually those useful to actuaries). In this book we use the R language.

The second development is a thorough examination of the mathematical foundations on which survival modelling rests. This is now also the mathematical basis of the survival models that actuaries use. Few actuaries are aware of it, however. As well as introducing modern statistical methods in a practical setting, which occupies the first two-thirds of this book, we also wish to bring some of the recent work on the foundations of survival models to an actuarial audience. This occupies the last third of the book.

The book is divided into three parts: in Part One, Analysing Portfolio Mortality, we introduce methods of fitting statistical models to mortality data as it is most often presented to actuaries, and assessing the quality of model fit. In Part Two, Regression and Projection Models, we discuss the graduation and forecasting of mortality data in one and two dimensions. In Part Three, Multiple-State Models, we extend the models discussed in Part One to life histories more complex than being alive or dead, and in doing so we introduce some of the modern approach to the foundations of the subject.

Part One consists of nine chapters. Chapter 1 begins by introducing mortality data for individuals and for groups of individuals. Grouped data lead naturally to mortality ratios, defined as a number of deaths divided by a measure of time spent exposed to risk. We give reasons for preferring person-years rather than the number of persons as the denominator of such ratios. Chapter 1 also introduces a case study, a UK pension scheme, which we use as an example to illustrate the fitting of survival models. This chapter discusses the choice of software package for fitting models, and provides a section on the notation that will be followed throughout the book.

Chapter 2 discusses data preparation. We assume that the actuary has data from an insurance or pension portfolio giving details of individual persons. Typically this will contain anomalous entries and duplicates, and we describe how to identify and correct these. Data can then be grouped if an analysis of grouped data is required, or for tests of goodness-of-fit.

In Chapter 3 we introduce the basic probabilistic model that describes the lifetime of an individual person, leading to the key quantity in survival modelling, the hazard rate or force of mortality. Then in Chapter 4 we discuss statistical inference based on the data described by the probabilistic model. We discuss at length the two features that distinguish survival modelling from other branches of statistics, namely left-truncation and right-censoring. Then, in Chapter 5, we focus on parametric models fitted by maximum likelihood, with examples of R code applied to our pension scheme Case Study. We deal with both grouped data and individual data, and show how these analyses are related. By making simplifying assumptions, we obtain the binomial and Poisson models of mortality well known to actuaries. These three chapters form the heart of Part One, and introduce the key model and methodology.

Chapter 6 discusses tests of model goodness-of-fit. We introduce standard statistics such as information criteria that are widely used in statistical practice for model selection, and also the more familiar battery of detailed tests designed to assess the suitability of a fitted model for actuarial use.

One of the main advantages of modelling individual data, rather than grouped data, lies in the possibility of allowing for the effects of covariates by modelling rather than by stratifying the data. Chapter 7 compares stratification and modelling, based on a simple example, and shows how the fitting process in Chapter 5 can be simply extended.

Chapter 8 introduces non-parametric estimates of mortality rates and hazard rates, and their possible use in actuarial practice. These provide a useful, easily visualised presentation of a mortality experience, if individual mortality data are available.

Finally, in Part One, Chapter 9 discusses the role of survival models in risk-based insurance regulation.

Part Two is divided into four chapters. In Chapter 10 we consider regression models of graduation for one-dimensional data. We start with the Gompertz model (Gompertz, 1825) which we fit initially by least squares. This leads to Poisson and binomial models which we describe in a generalised linear model setting. A particular feature of all four chapters is the use of the R language to fit the models; we hope not only that this enables the reader to fit the models but also that the language helps the understanding of the models themselves. Computer code is provided as an online supplement for all four chapters.

In Chapter 11 we discuss smooth models of mortality. We begin with Whittaker's well-known method (Whittaker, 1923) and use this to introduce the general smoothing method of *P*-splines (Eilers and Marx, 1996). We use the R package *MortalitySmooth* (Camarda, 2012) to fit these models.

In Chapter 12 we consider two-dimensional data and model mortality as a function of both age and calendar year. We concentrate on three particular models: the Lee–Carter model (Lee and Carter, 1992), the Cairns–Blake–Dowd model (Cairns et al., 2006) and the smooth two-dimensional model of Currie et al. (2004). We fit the Lee–Carter model with the R package *gnm* (Turner and Firth, 2012), the Cairns–Blake–Dowd model with R's `glm()` function, and the smooth two-dimensional model with the *MortalitySmooth* package.

In the final chapter of Part Two, Chapter 13, we consider the important question of forecasting. We lay particular emphasis on the importance of the reliability of a forecast. We consider both time-series and penalty methods of forecasting; the former are used for forecasts for the Lee–Carter and Cairns–Blake–Dowd models, while the latter are used for the smooth models.

Part Three explores extensions of the probabilistic model used in Part One, which represent life histories more complicated than being alive or dead. It is divided into four chapters.

The framework we use is that of multiple-state models, in which a person occupies one of a number of "states" at any given time and moves between states at random times governed by the probabilistic model. We think these are now quite familiar to actuaries. They are introduced in Chapter 14, mainly in the Markov setting, in which the future is independent of the past, conditional on the present. They are defined in terms of a set of *transition intensities* between states, a natural generalisation of the actuary's force of mortality. The key to their use is a system of differential equations, the *Kolmogorov equations*, which in turn generalises the well-known equation (3.19) met in Chapter 3.

Chapter 15 discusses inference of the transition intensities of a Markov multiple-state model from suitable life history data. Since life histories consist of transitions between pairs of states at random times, what is observable is the number and times of transitions between each pair of states in the model. These are described by *counting processes*. Once these are defined, inference proceeds along practically the same lines as in Part One.

Chapter 16 applies the multiple-state model in a classical setting, that of *competing risks*. This is familiar to actuaries, for example, in representing the decrements observed in a pension scheme. However, it gives rise to some subtle problems of inference, not so easily discerned in a traditional actuarial approach to multiple decrements, which we compare with our probabilistic approach.

Chapter 17 returns to the topic of counting processes. Pioneering work since the 1970s has placed these at the very heart of survival modelling, and no modern book on the subject would be complete without a look at why this is so. It uses the toolkit of modern stochastic processes – filtrations, martingales, compensators, stochastic integrals – that actuaries now use regularly in financial work but not, so far, in mortality modelling. Our treatment is as elementary as we dare to make it, completely devoid of rigour. It will be enough, we hope, to give access to further literature on survival models, much of which is now written in this language.