Computational Intelligence:

Mortality models for the actuary

Willem Jan Willemse

Computational Intelligence:

Mortality models for the actuary

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.F. Wakker,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op dinsdag 6 maart 2001 om 13.30 uur

door

Willem Jan WILLEMSE
informatica ingenieur
geboren te Leidschendam

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. H. Koppelaar

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter

Prof. dr. H. Koppelaar, Technische Universiteit Delft, promotor

Prof. dr. J. M. Aarts, Technische Universiteit Delft

Prof. dr. R. M. Cooke, Technische Universiteit Delft

Prof. dr. Ing.habil W. Gerhardt-Hackl, Technische Universiteit Delft

Prof. dr. S. Haberman MA PhD FIA FSS, City University London

Prof. dr. H. E. Lamur, Universiteit van Amsterdam

Dr. ir. J. C. A. van der Lubbe, Technische Universiteit Delft

Dit onderzoek is gefinancierd door Practis B.V. en mede gesubsidieerd door het Ministerie van Economische Zaken.

Published and distributed by: DUP science

DUP Science is an imprint of
Delft University Press
P.O. Box 98
2600 MG Delft
The Netherlands
Telephone: +31 15 2785121
Telefax: +31 15 2781661
E-mail: DUP@Library.TUDelft.NL

ISBN 90-407-2160-2

Keywords: actuarial science, computational intelligence, mortality forecasts

Copyright © 2001 by W.J. Willemse

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilised in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the publisher: Delft University Press.

Printed in The Netherlands

"There 's nothing serious in mortality
All is but toys",
Macbeth: Act II, Scene III

Contents

Preface **xi**

1 Introduction **1**
1.1 Subject of this thesis . . . 3
1.2 Method of this thesis . . . 4
1.3 Outline of this thesis . . . 5
1.4 Some preliminary definitions . . . 6

2 Models **9**
2.1 The objective of modelling . . . 9
2.1.1 The model theories of Zadeh and Maciejowski . . . 10
2.1.2 Outline of this chapter . . . 12
2.2 The first mortality models . . . 13
2.2.1 The first models based on observation . . . 14
2.2.2 The first explanatory models . . . 15
2.2.3 Mortality laws . . . 20
2.2.4 Survival distributions . . . 22
2.3 Competing risks models . . . 24
2.3.1 General description . . . 25
2.3.2 Assumptions of competing risks models . . . 31
2.3.3 Relaxing the assumptions . . . 33
2.3.4 Discussion . . . 34
2.4 Heterogeneous population models . . . 37
2.4.1 General model . . . 38
2.4.2 The frailty model . . . 40

2.4.3 Other definitions of conditional mortality intensity 42
2.5 A general model of mortality 47
2.5.1 Gompertz' law of mortality reformulated 47
2.5.2 The vitality model 59
2.5.3 Generalising the model 63
2.5.4 Comparison with other models 64

3 Measuring **67**
3.1 Purpose of measuring 67
3.1.1 Measurement limitations 68
3.1.2 Outline of this chapter 69
3.2 Mortality probabilities 70
3.2.1 Graduation of measurement data 70
3.2.2 The concept of probability 71
3.3 Metrics for mortality 74
3.3.1 The mortality surface 74
3.3.2 Mortality quotients 76
3.4 Interpolation methods 80
3.4.1 Calendar year-structured data 81
3.4.2 Number of dead and size of a population 82
3.4.3 The arcsin-law 86
3.5 Identifying competing risks models 90
3.5.1 Independence of risks 91
3.5.2 Measuring the parts 93
3.5.3 Applications of competing risks models 95
3.6 Measuring population heterogeneity 100
3.6.1 Relative mortality experience 101
3.6.2 Multiplicative hazards 102
3.6.3 The choice of the vitality distribution 103
3.7 Measuring Gompertz' populations 105
3.7.1 Method of moments 105
3.8 Actuarial model identification 107
3.8.1 Zadeh's embeddability applied 108
3.8.2 Maciejowski's modelling approach 110

4 Applications **117**
4.1 Mortality forecasts . . . 117
4.1.1 Time and age . . . 118
4.1.2 Life expectancies . . . 119
4.1.3 Heterogeneous populations . . . 121
4.1.4 Forecast of the CBS . . . 124
4.2 Computational intelligence . . . 126
4.2.1 Life insurance mathematics . . . 127
4.2.2 Life insurance germs . . . 133
4.2.3 Representing insurance knowledge . . . 137
4.2.4 Ordering of actuarial risks . . . 145

5 Results **151**

6 Conclusions and discussion **155**

Appendix: the Gamma function and Meijer's G-function **157**

References **161**

Samenvatting **168**

Summary **171**

Curriculum Vitae **173**

Preface

In the age of Herodotus, the oracle of Apollo in Delphi was one of Greek society's most authoritative institutions. People journeyed from all over the world seeking advice from the oracle. Often the questions involved domestic affairs, such as whether or not to take out a loan, or whether the harvest would be good that year. Sometimes delegates came to Delphi to ask whether their state should go to war against a tyrant. Pythia (the priestess of Apollo) answered on behalf of the god. She did so in a trance and priests translated her (incoherent) cries into more or less comprehensible, often rhyming sentences. Because so many different persons and states came there with questions, the oracle had an excellent notion of what was going on in the region, and could tailor the advice accordingly.

Nowadays, the role of Pythia with priests as predictors can be found in several disciplines. A meteorologist predicts the weather, a market analyst anticipates stock prices, and an actuary determines the level of the premiums for our insurances and pensions. These predictors no longer derive their authority from a special relationship with the gods. They employ complicated computer models and use a different language. Words like 'prophesise' and 'predict' are not readily used. One prefers to speak of calculating instead of predicting.

This thesis deals with the predictions an actuary makes. There are two people in particular who introduced me to the field of the actuarial sciences and made it possible for me to write this thesis. The first is my Socratic Ph.D. supervisor Henk Koppelaar, master in reasoning by analogy and midwife (or rather, midman) pur sang. The second is Nico Top, a master of pragmatics and director of Practis B.V., the company where I conducted this doctoral research project. I wish to thank both of them for the confidence they displayed in me and the freedom offered in which I was able to work. I would like to thank professor Cooke and professor Haberman for many improvements and corrections. I also thank the co-sponsor of this research, the Ministry of Economic Affairs, and in particular Wout Sorgdrager of Senter, the agency which handles the grant procedures. Finally, I would like to thank my colleagues at Practis for

a stimulative environment, and above all my family and friends for their love and support.

Chapter 1

Introduction

This thesis is the report of a study undertaken in order to improve knowledge about and therefore insight into actuarial 'tables'. The word 'table' is expressly placed in quotes because a table forms a representation of data about which so little is known that they cannot be adequately (e.g. not accurately enough) captured in a simple and succinct mathematical formulation. In everyday language, such a mathematical formulation is referred to as a law or model.

Unfortunately, the term 'representation' also brings with it its contradictory use. In the natural, linguistic and social sciences, the word 'model' is used for the (succinct) representation of data, ideas and theories. By contrast, logics and mathematics have formal propositions or formulas in which situations either from reality or from other logics or mathematics serve as a representation (or indeed 'realisation' or 'model'). In mathematics, that for which a construction (concept, proposition or theory) stands, what it represents, is referred to as a 'model' of that construction. In mathematics and logics, a model of a construction can only be an object, or a set. In thesis, we stick to the natural science tradition: a model represents data and theory.

To determine the amount of premiums for life insurances and predict the development of reserves, knowledge is required of the future development of mortality. Such knowledge emerges from collecting data and then abstracting of details. Through abstraction, simple and succinct descriptions can be generated which model behaviour from measurements and predict measurements, which have not yet been performed. With increasing abstraction come progressively simpler descriptions. Knowledge increases in each case if new models are devised as earlier ones are repudiated.

For example, the gravitational theory of Descartes was refuted when it appeared that planets do not in fact move in circles around their sun. Newton not only explained all the measurements

which Descartes' theory could explain, but also the non-circular orbits of the planets. Descartes' theory was thereby refuted. Newton's theory then replaced that of Descartes. Later, Newton's theory was itself refuted and replaced by Einstein's general relativity theory, for it appeared impossible with Newton's theory to precisely describe the orbit of the planet Mercury. Einstein correctly predicted Mercury's orbit. In 1982, the solar physicist Hill shook general relativity theory itself by challenging its oldest confirmation. Hill showed that one could successfully calculate the orbit of Mercury with Newton's theory. To do this, he used measurement data, which were not available in Einstein's time.

A similar sort of development can be seen in mortality research. To predict the probability of dying at a certain age, one uses tables, which 'model' the behaviour of previous measurements. The first actuarial tables were representations of the behaviour of the limited measurement data and were used without actually modelling the behaviour of the measurement data. As better measurement data became available, models developed which abstracted irrelevant details out of the measurements. Thus, there has been a perceptible interaction between the model and the measurement data. Other factors also influence models, such as the availability of powerful computers, which allow researchers to use increasingly complex models.

This work reports on a quest for the interaction between mortality model and mortality data. This is done using a description of the development of modelling in the actuarial sciences. We will also look at the theory of how perceptions arise. It was Popper's [Pop59] hypothesis that there is a distinction between theoretical and empirical statements. But theoretical statements are impossible without a few very elementary experiences and observations (experiences, perceptions, sensations) too are in one way or another always contaminated with expectations and models, or in other words: with theory. Therefore, in turn, observations are by no means independent from theory.

The main goal of mathematical modelling is to predict (consequences of contemporary) observations. For example, observation led to Makeham's law of mortality, with which someone's life expectancy can be predicted. In our contemporary perspective, the term 'predict' strikes us as odd. After all, we are convinced that we have done such a good job of modelling the regularities of survival probabilities that we speak of 'calculate'. We no longer need to read tables, because we have mortality laws. However, it is of course mythology to think that a phenomenon would obey a law.

A law makes it easier to identify a phenomenon, while at the same time it poses a question about why something repeats itself. The questions posed give more insight and lead to new laws being proposed. Knowledge is not so much a knowing ('ontology') as it is the crystallisation of a

line of reasoning. As a field of knowledge evolves and progressively better laws are formulated, facts and events can be predicted. In short, insight is achieved.

The criteria for insight are that the laws involved be

- conceptually accepted. As soon as a model is conceptually accepted, it can be elaborated and made more complex in an attempt to increase its adequacy.
- mathematically and succinctly formulated, in such a way that verbal concepts which are processed in it are interpretable.
- adequately tested, i.e. in accordance with observations. The condition for this is that the formal form be calculable, and not merely an unusable and purely qualitative model.
- and allow prediction of facts which were not observed earlier.

If, after time has passed, a scientific theory has emerged it can then be cast into a formal form. A model is expressed in a formal language because this makes it clearer and simpler than the corresponding verbal formulation. The concepts are thereby sharpened and simplification of the model is facilitated. In this process, simplification is not a goal in itself, but a means for increasing insight.

The highest form of insight is attained if we are able to convert the formal description in which a model is expressed into an algorithm. A model can then be interpreted as an algorithm, which in turn can be seen as a computer program. By using different programming languages, the programs will have different properties with respect to (for example) length of the description (the complexity), speed and accuracy. If the program has a limited length, one model will often be preferred on that basis to another model with a more complex program.

The field which deals with the acceleration of software with the aid of human intelligence is called 'computational complexity'. A related field, 'computational intelligence', concentrates on the automatic acceleration of software with the aid of the machine itself.

1.1 Subject of this thesis

As understood by Szczepaniak, computational intelligence is "a set of machine capabilities which allow the machine to use acquired knowledge and to solve new tasks efficiently in the presence of new conditions of their performing" [Szc99, page VII]. Generally, computational intelligence integrates a hardware construction with computational procedures and uses human (expert)

primitive and self acquired knowledge. Examples of computational intelligence can be found in fuzzy, genetic and neural computing. This thesis is about the application of computational intelligence in life insurance mathematics.

Life insurance mathematics is a field with a rich history and applications of it are widely used these days in the administrative systems of insurance companies and pension funds. In contrast to (for example) risk theory, life insurance mathematics is not heavily influenced by other disciplines. Because of this, the field has developed its own tools, such as the life table as a simple representation of measured mortality data.

Although many models derive their grounds of existence from their successful application, it must be possible to derive knowledge from existing mortality models and use it to improve actuarial models. This view is endorsed by Greenwood [Gre28]. In 1928, he found that 'to an actuary, a life table is not a subject for curious speculation but a working tool, sufficient to explain why, in actuarial circles, interest in biological 'laws' of mortality is lukewarm'. According to Greenwood, in principle it was indeed possible to develop a scientific theory, which would allow researchers to understand the relation between age and mortality.

This study is directed towards improving actuarial models and calculations by using computational intelligence. This is achieved by introducing a general framework from which existing models can be abstracted. Tools are designed in order to improve existing models so that actuarial tables are cleaned up and predictions improved. Given the complexity of the phenomenon of mortality and the large number of factors which have an impact on mortality, this thesis primarily concentrates on the question of how existing mortality models can be abstracted from in such a way that they can be generalised and applied in actuarial models. This is done without emphasis on empirical testing, since this often is what the existing models rely on for their justification.

1.2 Method of this thesis

The research method employed in this study is 'rational reconstruction'. There is a difference of opinion about how such rational reconstruction should be carried out. There are two main currents of rational reconstruction. Stegmüller [Ste79] has labelled them the 'Carnap school' and the 'Suppes school'.

According to Carnap, a theory must be reconstructed in a formal language. It was Carnap's conviction that only formal languages were appropriate tools for attaining the desired degree of precision, and so he postponed consideration of already-existing empirical theories until an

universal formal artificial language had been constructed. For this reason, the Carnap school is also called the school of the formal language approach.

Suppes' approach was different. Suppes and his colleagues used informal logic and set theory only for the axiomisation. With this approach, Suppes and his colleagues were able to axiomatize physical theories without reverting to formal languages, thus making their work more readable than attempts with the aid of formal languages from the Carnap school.

With Carnap's approach, mathematical reconstruction (in the most formal sense of the word) of theories would long remain merely a 'logical' possibility, one which was not practically feasible. Devotees of the Carnap school are forced to offer small, simplified examples instead of existing theories. The work is too voluminous for one person to approach existing theories in this way. Stegmüller thus also argues that Carnap's approach is 'humanly impossible'.

This work follows the approach of Suppes. The first step is the reconstruction and clarification of the internal mathematical structure of models. A model is autonomous in the sense that neither the significance nor the (possible) truth depends on something outside of the model. This means that the reconstruction is independent of the model's relation to the entities outside of it.

Models are based upon quantitative concepts or functions. The values of some of these functions are obtained by calculations of functions, which appear in other, more elementary, models. Ultimately we reach a moment when a quantitative concept is associated with a qualitative concept. This is the moment when the theory touches reality: experiments and observations refer to qualitative relations, which underlie the quantitative structures. The connection between theory and actual observations is an essential component.

The result is a theory, which consists of two layers. The upper part which uses the quantitative concept, and a lower part in which axiomatically defined operations and relations are described.

1.3 Outline of this thesis

After the introduction come three chapters, all focusing on the word 'to see'. The 'Models' chapter concentrates on the question 'What does one see with?'. The organ, with which we see (in the wide meaning of the word, not just 'looking') is located in our head. Intelligence allows us to reconstruct reality in our head and to correspond our actions to this reconstruction. In science, such a reconstruction of (a part of) reality corresponds to a model. The chapter tries to clarify what mortality models are. The first step in the rational reconstruction of theories

and models is axiomatization, the purpose of which is to expose (or indeed to discover) the underlying formal structure of a model. The work of Cox [Cox72] on survival data analysis can be seen as an attempt to provide a general mortality model. The work of Vaupel [VY83] and Manton [MS84] on heterogeneous populations will play a part in the general mortality model introduced in this chapter. The chapter comprises the first layer of Suppes' theory.

The second question is 'How does one see?'. We call this operation 'Measuring' and it forms the subject of the third chapter. Without measuring, a model has no meaning. That measuring without theory is impossible and is no trivial action shall be made clear. The theory about measuring forms the second layer of Suppes' theory. In this same chapter, the third question 'What is seen?' is answered. Data is then processed using the concepts described in the first chapter. Only then can we judge whether a model actually satisfies the desired criteria. That these are three essentially different questions is one of the things clarified by this thesis. By two examples of actuarial model identification, computational intelligence is applied.

After that, the chapter 'Applications' shows how applications of computational intelligence lead improvements of actuarial calculations. It describes how mortality models as laid down in the previous chapters can be used for actuarial calculations and for ordering of actuarial risks [Hee91]. The work concludes with two chapters: one in which the results are summarised, and another with the conclusions and recommendations of this thesis.

1.4 Some preliminary definitions

Terminology for mortality models differs from one knowledge domain to another. This section is intended to provide insight into the terminology used in this thesis. The study of mortality models and data is also known as survival data analysis. Survival time can broadly be defined as the time to the occurrence of a given event, for example the development of a disease or death. Survival data can include survival time, survival or the development of a disease [Lee92].

The survival function (or survivorship function) $S(x)$ is defined as the probability that an individual survives an age $x \geq 0$, that is

$$S(x) = \Pr[\text{survival to age } x] \tag{1.1}$$

If $S(x)$ is defined for only (some) integer ages of x, one can speak of a life table or an abridged

life table. Survival function $S(x)$ is a non-increasing function of age x with the properties

$$S(x) = 1 \text{ for } x = 0$$

and

$$S(x) = 0 \text{ for } x = \infty$$

that is, the probability of surviving at least age zero is one and that of surviving an infinite age is zero. This function is also known as the cumulative survival rate and the graph of $S(x)$ is known as a survival curve. The cumulative distribution function $F(x)$ is defined by

$$F(x) = 1 - S(x)$$

Because the survival time can be interpreted as a probability distribution, a corresponding probability density function can be defined as the limit of the probability that an individual dies in the short interval x to $x + \Delta$ per unit width Δ, or simply the probability of dying in a small interval per unit time. The density function, also called an unconditional failure rate, can be expressed by

$$f(x) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr[\text{die in } [x, x + \Delta)]$$

The probability of dying during a very small time interval, assuming that the individual has survived to the beginning of that interval is the mortality intensity $\mu(x)$. This conditional failure rate is the limit of the probability that an individual dies in a short interval, x to $x + \Delta$, given that the individual has survived to age x, or

$$\mu(x) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr[\text{die in } [x, x + \Delta) \mid \text{survival to age } x] \tag{1.2}$$

The mortality intensity can also be defined with the distribution function $F(x)$ and the probability density $f(x)$

$$\mu(x) = \frac{f(x)}{1 - F(x)} = -\frac{1}{S(x)} \frac{d}{dx} S(x)$$

The mortality intensity is also known as the hazard function and the conditional mortality rate. In reliability theory, the instantaneous failure rate or the age-specific failure rate can be found for the mortality intensity. It plays an important role in this thesis and in mortality

models in general. The cumulative hazard function is defined by

$$H(x) = \int_0^x \mu(t)\, dt = -\ln S(x) \tag{1.3}$$

The cumulative hazard rate is another often used concept in survival data analysis.

Chapter 2

Models

The quest for suitable models can be seen as the most important goal of science. The reasons for this change as we move from the 'hard' sciences (physics, chemistry) to the 'softer' sciences (economics, sociology). In the former, our objective is to imitate, copy, simulate, etc., and in the latter, we rely on prediction and speculation.

Whereas physics can create its own objects when formulating explanatory theories, the softer sciences are committed to the observation of their research subjects due to a lack of explanatory theory. For example, quarks and gravitons are theoretical objects of which we have no direct sensual experience (their existence is not proven). It is much easier on the other hand to determine whether a person is alive or dead.

2.1 The objective of modelling

The objective of modelling is defined by the application of the model. The general goal of actuarial modelling is to (re)construct mortality data at a given time, in order to use the reconstruction for actuarial calculations such as determining life insurance premiums. Reconstructing involves the transformation of the observed data into mortality probabilities and the extrapolation of these probabilities to future expectations (prediction). Reconstructions are also known as representations. Modelling can therefore be seen as the process that leads to a representation. Though the representation may not be physically linked to the entity modelled, a priori knowledge of the measurement data and of the language used to formulate the representation is always required.

Another important objective of modelling is that of deriving knowledge. Predicting future

mortality based on observations requires knowledge of specific external events, which cause jumps in the observed data. Classifying events of a temporary or permanent character is an important element in extrapolating to the future. Modelling mortality enables us to identify trends and derive causal relations.

Reconstructing mortality is not the only objective; if it were, keeping tables of measurements would be sufficient and there would be no need to reduce the dimensional space of the model. Models that describe mortality by a direct mapping of the observed data will seldom be used to gather knowledge or to make forecasts. Therefore, an important aspect of formulating mortality models is that of reducing the complexity of working with them, i.e. computational complexity.

Furthermore, modelling results in the omitting of irrelevant details since modelling leads to an abstraction of the observed reality and a compression of the data into less complex elements. Since modelling reduces complexity, it gives a summary of reality. Besides omitting details, it also accentuates certain properties of reality. In so doing, models reveal and conceal at the same time.

Models are never derived from data alone. Besides omitting details, modelling also requires (essential) assumptions. These assumptions do not follow from the original observations, and so cannot be proven. They are needed to reduce complexity to a required level.

In the next section, two general approaches of modelling are described. Throughout this thesis, we will refer to these modelling strategies.

2.1.1 The model theories of Zadeh and Maciejowski

This chapter deals with mortality models. Without predefining the objective of mortality models, because that would close unexplored lines of research to us, we will first look at two model theories: Zadeh's model theory [Zad56] and Maciejowski's model theory [Mac78].

Zadeh's theory requires that the preconceptions at the root of modelling must first be laid down in a model type. The type or class of models used to identify the data is the preconception of the modelling work. The wrong class of models would result in an inability to imbed the data. The gravitational laws presuppose the circular motion of the planets in line with Descartes' theory. When discrepancies in measurement arose as the model was bedded in, the preconceptions could be refuted and replaced and better models could be made.

Once the preconceptions have been determined, the process of modelling can start. Zadeh stated modelling to be the identification of a model from model class M. Take black box B whose input and output relations are not known beforehand, an input domain, and a class of

black boxes known as M, of which we know, due to prior knowledge of black box B, that B is an element of M. Then, modelling is identifying the optimum element of M that is equivalent to B so that all outcomes of black box B are identical to that element of M. The definition results in a first demarcation.

Maciejowski's model theory is about the search for mathematical models, or, in other words, mental models that can be formally described as computer programs and that can adequately reconstruct and explain the behaviour of a corresponding system. Modelling is making available an algorithm to construct a model for a system, based on a quality criterion. It is assumed that the system to be modelled is definable as an ordered pair of input and output observations (in which some or all of the input or output is unknown). A system model is an algorithm that assigns an output to all input observation subsets (of the system under study), where the term algorithm can be interpreted as a computer program. The quality criterion is one of weighing complexity against accuracy.

The description is broad enough to contain tables, for example, but these are of no use in finding out how the system reacts to a new situation, and that is usually the main objective of modelling. However, this type of model, the trivial model, is a point of reference for the success of the modelling process. The quality criterion is the information gained by the algorithm expressed in programming language. The information gained is defined as 'the difference between the length of the trivial model program and the length of the program for the model under consideration, where the trivial model of the system is the one that calculates output observation by reading a table'. The model that gains the most information is the best.

If the gain in information is zero the model uses all the output observations, and that is precisely what the trivial model is: a table. If the information gain is close to maximum (i.e. close to the length of the trivial model) then the model is simple enough to be constructed from a small part of the data set, and the model explains the rest of the set. The information gain can also be negative. This shows that the model has more random parameters than are justified by the quantity of information in the data set.

The higher the gain in information is the greater the confidence in a model will be. All knowledge of a system resides in the set of observations. This implies that the confidence that someone may have in a certain system model can be restricted by the size of the trivial model for the system. This corresponds with the intuitive observation that a limited number of system observations lead to a lack of confidence in this type of system model. Furthermore, it implies that small models are good because they offer a higher information gain and engender greater confidence. Good models will therefore tend to re-use the same (short) algorithms for as many

calculations as possible. This is why simplicity is desirable in scientific hypotheses. This ties in with the general conviction that the more frequently an observation is done, the more likely it will occur in the future.

Zadeh and Maciejowski compared

The advantage of Maciejowski's procedure for selecting a model is that it can be applied to complex, non-linear, dynamic models, and that the structures of the models compared need not be the same. This is achieved by joining the model structure and parameter estimation. The strategy requires fully specified models (with parameters filled in). In Zadeh's model theory, the best model always belongs to a model class of the same structure chosen beforehand.

In Maciejowski's approach, models are assessed using a quality criterion depending on the programming language in which the models are written, and on the way in which the observations are coded. However, this disadvantage appears in every model strategy. The approach always depends on the a priori information used to define the programming language and on the coding of data. Furthermore, it is uncertain whether the length of the program gives sufficient indication of quality. Different programming languages differ also in efficiency.

The need for a priori knowledge in non-linear systems has made it impossible to find the best and 'true' model for a system. All we can do is compare models that use this a priori knowledge. If the real model is not in the set of candidate models, it will not be found.

In the end, there can never be a general, effective method of finding the best model for a system. Applying a standard identification technique mechanically does not guarantee a usable model. It is not possible to find a model without specific expertise in the system discipline. The boundaries of system modelling depend on the progress made in that particular area.

2.1.2 Outline of this chapter

This chapter discusses a number of model classes. We will start with a brief historical summary of the first mortality models (2.2). These aimed to describe whole population mortality (without consideration of the cause of death and other variables) and resulted in a life table, or age-structured mortality probabilities.

An important element of the early models is the explanation they were given. These explanations enable us to justify and assess the models, but often they do not provide an adequate description of how the models work. We should take a critical view of informal explanations and should remember that a phenomenon can always be explained in several ways. We can

gain new insights by comparing all the explanations.

After this, we will discuss a class of models with several (final) states (for example dependent on the cause of death): competing risks models (2.3). In competing risks models, various causes of death are used to aggregate a survival function. The general competing risks model is described, followed by model assumptions to simplify applications. Then, a new class of models based on heterogeneous populations is described (2.4) and the parameter frailty is identified to distinguish higher risk sub-populations from a total population.

In section 2.5 a general, explanatory mortality model is introduced. The use of extreme value theory lead to a reformulation of an important mortality law (2.5.1) and in section 2.5.2, it is applied in the general mortality model. The earlier described models will then be compared with this general model (2.5.4).

The model classes are essentially deterministic but can have stochastic or deterministic interpretations. This differentiation only appears when the notion of probability is translated to a usable interpretation, i.e. if it is measured. The next chapter is concerned with measurement. Stochastic or deterministic models do not imply essential improvements, or, in other words, the (Suppes) structure remains unchanged.

2.2 The first mortality models

By looking into the past, we can gain insight in the way that people used to think about mortality models. Why did people start looking for mortality models? How did models come about? To answer these questions, it is necessary to look at the models which the study of mortality has put forward in order to forecast mortality. Although they were not seen explicitly as models in the way they are today, we will treat them as such.

Mortality modelling started in the seventeenth century. The modelling process resulted in a life table that represented the size of a population per age, based on a population aged zero. The first tables were composed from observations by interested parties. The better the recording process, the more accurately the death could be defined. However, the recorded deaths contained more irregularities because of the specific mortality pattern of the population during the period of observation. This resulted in a demand for a method of graduating life tables and (later) of modelling them by analytical functions. The study of population dynamics also called for simple ways to model mortality.

2.2.1 The first models based on observation

Only once the concept of 'probability' had entered the human repertoire and information had been gathered on a larger group of people, it was possible to measure mortality probabilities. Modern mortality research into the Egyptian, Greek or Roman civilisations is only possible in specific cases, such as the study of mummies or with the help of excavations. However, these few cases cannot be used to make generalisations for larger populations. We possess genealogical information concerning the German nobility in the twelfth century, but these too involve such a small group that generalisation is impossible [Lan90].

The mortality data in the monasteries and parish records became available late in the sixteenth century. Although much information about the causes of death such as sickness, starvation and war can be deduced, they were not accurate records of death.

Graunt in 1662 [Gra1662] carried out the first large study of mortality probabilities. Although there have been studies before this, by Bacon for example, Graunt was the first to make detailed observations of mortality in London. The sources of his observations were the so-called 'Bills of Mortality'. The frequent plagues, a major cause of death in those times, created the need to keep these reports. To stop the plagues having too big an effect, Graunt tended to conduct his observations during periods when there were relatively few plagues about. Graunt was the first to demonstrate that there were regularities to be found in life's events, such as birth, marriage, death and high mortality in the first years of life. He was critical of his sources and careful with the new hypotheses that he developed.

By a coincidence, Graunt's work became possible through a schism between the Roman and English churches. One effect of the schism was that it was no longer the sole preserve of the church to register births and deaths. Another reason was the emergence of life insurances.

In those days, obtaining information was no easy matter because the population was often reluctant to co-operate. Graunt's mortality table therefore contained a strong element of guesswork. Using Graunt's table, Christiaan Huygens and his brother Lodewijk calculated the life expectancy of a new-born at 18.2 years on average.

In 1671, after some speculation, de Witt assumed that from a group of 128 individuals with the age of four, 28, 14.7 and 4.7 would survive to the 54th, 64th and 74th year respectively. De Witt also assumed that no one would live beyond the age of 80. He used the following life table

$$\begin{aligned} S(4) &= 128 \\ S(54) &= 28 \end{aligned}$$

$$\begin{aligned} S(64) &= 14.7 \\ S(74) &= 4.7 \\ S(80) &= 0 \end{aligned}$$

To calculate the ages in between, he used a first-order linear function as an approximation. De Witt's table appeared more realistic that Graunt's.

The astronomer Halley understood that a table could be constructed using the date of birth and the date of death [Hal1693]. The observation should be made in a population in which migration is negligible. He chose the city of Breslau (Wrøclaw). This work gave rise to the first age-structured table based on statistical data. Halley also noted high mortality probabilities during the first years of life.

Another contribution was the life table of Deparcieux of 1746. With his observations, he was able to demonstrate that the life expectancies of adults had risen in the latter half of the century. Furthermore, in 1766 Bernoulli applied continuous analysis and introduced $\mu(x)$, which later became known as the mortality intensity. The reciprocal of $\mu(x)$ was in 1772 introduced by the self-taught man Lambert who called it the force of vitality. Lambert was also the first to graduate a life table graphically and tried to find mathematical formulas for $S(x)$ with some semblance of a crude theoretical basis [Daw80].

2.2.2 The first explanatory models

During the years the observation-based mortality models were increasingly further formalised, and laws began to emerge. These laws led to the first explanatory models, also known as mortality laws that could be used to calculate mortality probabilities. This made it easier to calculate the value of life insurance policies. To describe these models the description put forward by Smith and Keyfitz in [SK77] is used.

In 1725, de Moivre described a time-based model based on the assumption of a straight curve $S(x)$ for t successive years, as the following equation shows

$$S(x) - S(x+t) = a \cdot t$$

where $a \geq 0$ is a real-valued constant. By dividing the total domain into pieces and defining a constant for each, it was possible to construct an - at that time - adequately accurate table.

De Graaf proposed another model in 1729. He supposed that the vitality of a person decreased with an increasing rate as the age x increased. He used this theory to express the

following survival probability

$$1 - \left(\frac{x}{92}\right)^n$$

and the complement gives the probability of death

$$\left(\frac{x}{92}\right)^n$$

for $n > 0$.

De Graaf assumed that the maximum age of a human was 92. The positive parameter n represents the speed with which vitality diminishes. With $n = 5$ a reasonable approximation of de Witt's life table is obtained.

Since population growth increases with the size of the population, other models were examined, such as exponential and logistic models. Population growth models of Malthus are an example of exponential models. These models use an increasing growth function rather than a linear increase. The exponential function uses the following definition $N(0)\, e^{at}$, where $N(0)$ is the population at moment 0, $t \geq 0$ the time and a a real-valued parameter, so-called intrinsic growth coefficient. However, no population can continue to grow exponentially (i.e. have a positive growth coefficient). With a negative growth coefficient, a population will gradually disappear.

To make the population model more realistic, Verhulst formulated in 1838 a logistic differential equation with the population growth $\frac{d}{dt}N(t)$ proportional with the product of the size $N(t)$ and a fertility coefficient $(1 - N(t)/K)$ in which K is the carrying capacity, the population size beyond which $N(t)$ cannot grow. This results in the following differential equation

$$\frac{d}{dt}N(t) = aN(t)\left(1 - \frac{N(t)}{K}\right) \tag{2.1}$$

The solution to this differential equation, with $N(0)$ as the initial population, is the following equation

$$N(t) = \frac{K}{1 + \left(\frac{K}{N(0)} - 1\right)e^{-at}}$$

It appeared that this model could be applied easily to mortality data.

Although exponential functions were later formalised in differential equations for population dynamics like (2.1), the first to make use of them was Lambert. In 1772, he gave a formula to approximate a life table he obtained from the deaths of the London Bills of Mortality, 1728-1757.

The model structure consisted of a de Graaf model with $n = 2$ corrected by two exponential functions

$$S(x) = \left(1 - \frac{x}{96}\right)^2 - a\left(e^{-b_1 x} - e^{-b_2 x}\right)$$

with $a, b_1, b_2 \geq 0$. Lambert did not have a scientific basis for his formula but his basic concept is that of the gradual 'exhaustions of a man's power to avoid death' [Daw80]. A similar explanation is used fifty years later by Gompertz.

In 1825, Gompertz noted in [Gom1825] that several life tables could be modelled by a geometrical series. He noticed this because the differences between the life table logarithms were approximately constant.

Gompertz assumed two causes of death

> *'It is possible that death may be the consequence of two generally co-existing causes; the one, chance, without previous disposition to death or deterioration; the other, a deterioration, or an increased inability to withstand destruction.'* [Gom1825, p. 7]

The first, a probability, depending on age, and the second, which he described as an increasing inability to withstand destruction. If we suppose, for example, that certain illnesses are equally infectious at every age, then their effect on mortality will be constant. If, furthermore, there were no other illnesses, then the number of survivors would decrease by a geometric ratio. This is de Graaf's model. However, since man's resistance falls it follows that the number of survivors of a given age would fall with an increasingly higher number than a geometric ratio. Gompertz applied this idea explicitly to the population. In his article he states

> *'If the average exhaustion's of a man's power to avoid death were such that at the end of equal infinitely small intervals of time, he lost equal portions of his remaining power to oppose destruction which he had at the commencement of those intervals, then at the age x his power to avoid death, or the intensity of his mortality might be denoted by $a \cdot q^x$...'* [Gom1825, p. 8]

Gompertz described man's resistance as $a \cdot c^x$ (nowadays we write c for the real-valued parameter q), in which a and c were positive real-valued parameters to be estimated and x was the age. He then defined $S(x) = k \cdot g^{a \cdot c^x}$, with $k > 0$ and $0 < g < 1$, for the number of survivors at age x, where k and g are real-valued parameters. Because g and a are both constants, g^a could be replaced by g, resulting in

$$S(x) = k \cdot g^{c^x} \tag{2.2}$$

The relationship with Verhulst's differential equation can be found by substituting an exponential decreasing growth coefficient in the equation and assuming that there is no limit to which $N(t)$ cannot grow, resulting in

$$\frac{d}{dt}N(t) = a(0) \cdot e^{-kt} \cdot N(t)$$

where $a(0)$ is an initial growth coefficient and $k > 0$ the decreasing growth coefficient.

A good thirty years later, in 1860, Makeham made a small adjustment to Gompertz' model. In fact, he actually took Gompertz' work a little further. Besides the exponential rise in mortality intensity, Makeham recognised different causes of death that could be represented by a constant mortality intensity. He extended the model with an extra term resulting in the following model

$$S(x) = k \cdot s^x \cdot g^{c^x} \tag{2.3}$$

in which the positive real-valued parameters k, s, g, and c were estimated in order to fit the observed data. This function is the basis for current methods of modelling mortality data. In [Mak1867] Makeham added another term to his model, so that a second, lesser-known model was created

$$S(x) = k \cdot s^x \cdot w^{x^2} \cdot g^{c^x} \tag{2.4}$$

The mortality laws of Gompertz in 1825 and of Makeham in 1860 can be seen as early attempts to find a simple model of mortality intensity $\mu(x)$ defined by definition (1.2). Both models assume that the mortality intensity rises exponentially. Gompertz' model fits accurately for 30 to 90 years. Before and after this interval, the model does not fit precisely. In those days, the older ages had a minor influence on the value of life insurance and life insurances were not often used by younger age groups. Therefore, the imperfect fit was not seen as a disadvantage.

To accommodate to all the ages such as younger age groups, it became necessary to provide a more accurate approximation of mortality data. Complex expressions were found, such as Thiele's model of 1872 that looked at mortality intensity $\mu(x)$ as follows

$$\mu(x) = a_1 e^{-b_1 x} + a_2 e^{-\frac{1}{2}b_2(x-c)^2} + a_3 e^{b_3 x} \tag{2.5}$$

where $a_1, a_2, a_3, b_1, b_2, b_3$ and c are positive real-valued parameters. The last term in the model is an exponential function for old ages, the first term being a decrement function for child mortality and a quadratic function for ages in-between.

Perks (as described in [SK77]) introduced a new class of models for mortality intensity

$$\mu(x) = \frac{a + bc^x}{kc^{-x} + 1 + dc^x} \tag{2.6}$$

where a, b, c, d and k are positive real-valued parameters. Like Gompertz, he had a rational explanation for the functional form of the model based on the physical concept of entropy change.

Besides mortality intensities, other mortality quantities were being modelled, such as the number of deaths $\mu(x) S(x)$ by Phillips, Clarke and Beard (see for example Benjamin and Pollard in [BP86]). In 1953, Ogborn introduced the following rational function

$$\mu(x) = \frac{a + x^2}{b_0 + b_1 x + b_2 x^2 + b_3 x^3}$$

Later, researchers focussed on how to predict mortality probabilities of specific groups such as children, young males and females and older persons. A mortality model that represents three distinct components of mortality, not unlike that of Thiele's model, was in 1980 introduced by Heligman and Pollard in [HP80] and is defined by

$$\frac{q(x)}{p(x)} = A^{(x+B)^C} + De^{-E(\ln x - \ln F)^2} + GH^x \tag{2.7}$$

where $q(x)$ is the probability of dying within one year for a person aged x, the probability of surviving $p(x) = 1 - q(x)$ and $A, B, ..., H$ are real-valued parameters. By defining the law in this way, the $q(x)$ only takes values between zero and one only. The first part represents a rapidly declining exponential function that reflects the fall in mortality during early childhood. The third term incorporates a Gompertz' model and the remaining term reflects accident mortality for males and accident and maternal mortality for the female population.

The mortality models put forward over the decades show a rising complexity. This relates to the detailed picture that we nowadays have of all aspects of mortality. Linked to this, there is an increasing amount of data available, and there are complex techniques for managing the growth in data. However, it should be born in mind that the increasing complexity of the models often comes down to the addition of parameters. It can often be contested whether a fundamentally new mortality model was introduced since many models are based on earlier elements. We can see this, for example, in Gompertz' model, which is used repeatedly in later models, albeit in modified form.

The summary above shows how the explanation of the model plays an important role. All discoverers of the above 'mortality laws' have given rational, yet often vague, explanations of the law's functional form, for example reduced vitality (de Graaf in 1729 and Gompertz in 1825) and physical entropy (Perks in 1932). An explanation is important to the conceptual acceptance of the model. At first sight, there is a direct relation between the explanation and the analytical formalisation of the model. The formalisation is then presented as an explication of the intuitive idea.

As we saw with Gompertz' explanation, he proposed that an individual's death process is the same as that for the rest of the population. Other explanations of mortality models are biomedical too. The unobservable risk of dying is explained on an individual basis. On the other hand, the model is not applied individually, and it is always geared towards modelling the mortality of a population.

The use of explanations was mainly intended as a means of justifying a model and making it easy to assess. The explanation was not primarily designed to explain how the model worked. This is supported by the fact that the parameters of the mortality models usually have merely a vague meaning. Though adding parameters to the model resulted in a more accurate model for mortality, it did not prove that the explanation was better. Researchers repeatedly make the mistake of confusing the explanation of a model with the way it works.

There is a danger of loosing oneself in a unilateral explanation of a model. The fact that a given explanation of a phenomenon can be neatly formulated in mathematical terms, and that we can even model it, is no reason for giving the explanation greater status than the hypothesis: it is one of the possible explanations.

2.2.3 Mortality laws

> *'Most of us retain, consciously or unconsciously, a feeling that, underlying all the roughness in our data referable to errors of observation and an ever-changing environment, there may be an inherent mathematical system of law and order, which if it could but be discovered would give such insight into the meaning of the unadjusted figures that a considerable advance would be made in the practical application of our science.'* (Perks as cited in [OC97])

The mortality models of the past have a long history of methods and model classes. The fascination for a general law of mortality has resulted in a number of elegant models that, though not general enough to describe 'actual mortality', did exert an enormous influence on

later developments.

Inspired by the progress made in physics, many scientists have also sought a law of human mortality. The first part of the title of Gompertz' article [Gom1825] ('On the nature of the function expressive of the Law of Human Mortality') indicates that he had much the same view of mortality laws as did many physicists. However, the sciences of biology and economics are largely bound by their research subjects.

Nowadays, there is no discussion over whether or not a law of mortality exists. Researchers have been looking for proof that biological processes are governed by physical principles, independent of quantitative formulations. The reason why it seems impossible to find the proof is that the phenomenon is complex. Whether there is biological limit to human life expectancy, what the effect is of genetic material and life style on (causes of) death and whether the (exponential) mortality probability increases with age are questions that cannot be answered with certainty, and probably will never be answered. Studies of fruit flies show that the mortality probabilities decrease with old age [OC97] (which would immediately refute many laws of mortality), but we do not know whether this is true of people. The only certainty seems to be that all people are mortal, as Socrates knew.

That does not mean that the mortality laws are not inherently deterministic. If they were not they would not be laws. A law must always have a variable that stands for 'for (almost) all' or 'for most'. The current and frequently heard stochastic interpretation of the mortality laws does not detract from the fact that they were all originally put forward as mortality laws.

The way in which we look at the past is determined by the events that disclose themselves to us. No model can stand on its own. The way that we recognise a model is not determined by the model itself, but by the way we use it.

There is a difference between the explanation and the use of mortality models. Almost every model can be explained at the individual level. This means that the explanations are biological or medical and that they describe the process of decline over the lives of separate individuals. On the other hand, every mortality model is used to describe the mortality of a population. So, the model is explained for an individual and used for populations. An individual reaches a certain age and then his life ends. We cannot predict what that age will be without looking at the internal fluctuations of that particular individual, as well as the external influences, two components with unknown micro laws and they are necessarily modelled by using probabilities. If we set the explanation aside we find that the mortality laws are macroscopic, average outcomes of the underlying micro laws. A population as a whole can behave according to a mortality law, but the collective properties of a population do not apply to individuals within that population.

An individual never dies in accordance with Gompertz' law of mortality. The mortality law only has meaning when used for a population.

When viewing the mortality laws as macroscopic and not considering the explanations of the past we again ask ourselves what the definition is of a mortality law. Are there even deeper reasons as to why a mortality law can be used with success, and with which the meaning of the parameters can be clarified? To do this we have to extract knowledge from the mortality laws. First, we will look at some often used survival distributions.

2.2.4 Survival distributions

The mortality laws described above show an isolated development of mortality models where new models improved old ones. In survival analysis, one is interested in probability distributions for reliability of machines and electronic systems. However, they can also be used for human populations. Below, three often-used distributions in survival analysis will be described.

The exponential distribution

The exponential distribution is often referred as a purely random failure pattern. One of its properties, 'lack-of-memory', requires that the age of an individual does not affect future survival. This property of course limits its application for mortality models but the distribution will facilitate description of other distributions. The exponential distribution with parameter $\lambda > 0$ has a probability density function defined by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases} \tag{2.8}$$

The cumulative distribution equals

$$F(x) = 1 - e^{-\lambda x}$$

and the survival function is

$$S(x) = e^{-\lambda x}$$

The constant mortality intensity $\mu(x)$ is equal to λ.

The Weibull distribution

A generalisation of the exponential distribution is the Weibull distribution. However, it does not assume a constant mortality intensity and has therefore a broader application. The probability density function of the Weibull distribution is defined by

$$f(x) = \lambda^k k x^{k-1} e^{-(\lambda x)^k} \tag{2.9}$$

for $x \geq 0$ and $\lambda, k > 0$. Parameter k determines the shape of the distribution and λ determines the scaling.

The cumulative distribution is

$$F(x) = 1 - e^{-(\lambda x)^k}$$

and the survival function is

$$S(x) = e^{-(\lambda x)^k}$$

The mortality intensity is defined by

$$\mu(x) = \lambda^k k x^{k-1}$$

The Gamma distribution

Suppose that death takes place in n stages or as soon as n subfailures have happened. Total failure or death occurs at the end of the n-th stage when the n-th subfailure happens. The survival time spent in each stage is assumed to be independently exponential distributed with parameter λ. Then the total distribution is called an Erlang distribution. A natural generalisation of this distribution is to replace parameter n restricted to integers by parameter k taking any real positive value. Then we obtain the Gamma distribution with density defined by

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)} \tag{2.10}$$

for $k > 0$. The cumulative distribution has the following form

$$F(x) = \frac{1}{\Gamma(k)} \int_0^{\lambda x} t^{k-1} e^{-t} dt$$

The integral is the (lower) incomplete Gamma function which can be derived from the

(above) incomplete Gamma function, which defined by

$$\Gamma(k, \lambda x) = \int_{\lambda x}^{\infty} t^{k-1} e^{-t} dt$$

(see the Appendix). Therefore, we also have $F(x) = 1 - \Gamma(k, \lambda x) / \Gamma(k)$.

The mortality intensity is the ratio of $f(x)$ to $1 - F(x)$. If k is an integer n, it equals

$$\mu(x) = \frac{\lambda^n x^{n-1}}{\Gamma(n) \sum_{k=0}^{n-1} (\lambda x)^k / \Gamma(k+1)} \tag{2.11}$$

Note: if $\mu(x)$ is defined by (2.11), then $\mu(x) \geq \lambda^n x^{n-1} e^{-\lambda x} / \Gamma(n)$, for all $x \geq 0$.

When applied to human populations, the mortality laws described in the previous sections are often more accurate than the three survival distributions described here. Therefore, in actuarial science the mortality laws are appreciated above survival distributions despite the lack of a sound theoretical basis. However, there exist relations between some mortality laws and survival distributions. For example, it is possible to express Gompertz' law of mortality as a survival distribution. Before doing this, we will discuss two mortality models that are essential to an understanding of the general model we are about to introduce.

2.3 Competing risks models

In previous sections, mortality was modelled without differentiation between causes of death. These models are often referred to as models of total mortality. In this section, we discuss deaths classified into disjoint groups, for example, based on medical conditions reported in death certificates.

Cause-specific models are known under different names in various disciplines. Within actuarial science, the termination from a given status (often the cause of death) is called decrement and models based on more than one decrement are called multiple decrement models. Within biostatistics and reliability theory, these models are referred to as competing risks models. The term risk is due to Chiang [Chi68] who termed the cause of death in hindsight more properly a risk in foresight. In this view, an individual is exposed to several competing risks of death while only one risk ultimately causes death. For example, death and disability are competing risks for purposes of calculations of premiums for disability insurances. One application of these models is that they allow eliminating a cause or risk of death from a population. Therefore, the term cause elimination models is often used.

First, we present a general framework of the multiple decrement or competing risks model (see for example Bowers et at. in [BGH86] and Preston et al. in [PKS72]). In this section, it is not important to differentiate between a cause and a risk. To simplify the model, often assumptions are made which are explained in the following section. We conclude with a discussion.

2.3.1 General description

A traditional age-structured life table aggregates data through various causes of death. This may result in a table with irregularities due to separate causes. A method of constructing less irregular tables uses competing risks models and is described in [Nam87]. We deal with this below.

In a population, unique risks R_k, $k = 1, ..., n$, ($n \geq 1$ causes of decrement) are different. Examples of risks are cardiovascular disease, cancer and accident but also withdrawal or termination of premium payment can be seen as risks. Everyone eventually leaves the population because of one of the n risks. First, we consider the population with only one acting risk R_k. The k-th decrement function $S_k(x)$ for the population with acting risk R_k, $k = 1, ..., n$, is defined for $x \geq 0$ by

$$S_k(x) = \Pr[\text{survival to age } x \text{ where } R_k \text{ is the only acting risk}]$$

This k-th decrement function, which is also known as the k-th subdistribution function or marginal survival function, satisfies

$$S_k(0) = 1$$

and

$$\lim_{x \to \infty} S_k(x) = 0$$

for $k = 1, ..., n$.

The multiple decrement function or joint survival function is defined as the probability to survive ages $x_1, ..., x_n$ in each sub-population and may be characterised by the function

$$S^{(n)}(x_1, x_2, ..., x_n) = \Pr\left[\bigcap_{k=1}^{n} \text{survival to age } x_k \text{ where } R_k \text{ is the acting risk}\right] \tag{2.12}$$

Note that this function depends on all n ages x_k, $k = 1, ..., n$. This function is also called the survivor function. Sometimes, the superscript (n) is deleted (see for example Gail in

[Ga75]). Given the properties of the k-th decrement functions, we define $S^{(n)}(0,...,0) = 1$ and $\lim_{x\to\infty} S^{(n)}(x,...,x) = 0$. Generally, the joint survival function cannot be expressed in terms of $S_k(x)$. Additional assumptions are needed, which will be explained later.

This formulation of a joint survival function is very general and applicable whether the survival times are statistically independent or not. The total survival function $S(x)$ is defined as the probability that death will not occur before age x given that all risks of death $R_1,...,R_n$ are acting, or

$$S(x) = S^{(n)}(\underbrace{x,...,x}_{n \text{ times}}) = \Pr[\text{survival to age } x \text{ where all risks are acting}]$$

To break down the total survival function into its risk components, we introduce the k-th crude survival function. It is defined as the probability that an individual will survive up to age x and then eventually die from risk R_k. We define the following probability as the k-th crude survival function for $k = 1,...,n$

$$P_k(x) = \Pr[\text{survival to age } x \text{ and die in } [x,\infty) \text{ of risk } R_k]$$

for $x \geq 0$. The $P_k(x)$ is the probability that both the individual will die of risk R_k and that he survives age x. It is not, in general, equal to the marginal survival function $S_k(x)$, which just gives the probability that someone survives age x with only acting risk R_k. The probability $P_k(x)$ is also known as the (fixed) subsurvival function. Because one will eventually die from one of the acting risks, the following always holds

$$S(x) = \sum_{1\leq k\leq n} P_k(x) \tag{2.13}$$

The following theorem, by Tsiatis in [Tsi74], indicates that any given multiple decrement function $S^{(n)}$ determines uniquely the crude survival functions $P_k(x)$.

Theorem 1 *The derivative $P_k'(x)$ of the k-th crude survival function is equal to the partial derivative of (2.12) with respect to x_k evaluated at $x_1 = x_2 = ... = x_n = x$, or*

$$P_k'(x) = \frac{\delta}{\delta x_k} S^{(n)}(x_1,...,x_n)\Big|_{x_1=x_2=...=x_n=x}$$

Proof. By differentiating $P_k(x)$ we mean

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\left(P_k(x)-P_k(x+\Delta)\right)=\lim_{\Delta\to 0}\Pr\left[\text{survival to age } x \text{ and die in } [x,x+\Delta) \text{ of risk } R_k\right]$$

Obviously, for a certain Δ^* such that $0<\Delta<\Delta^*$, this has lower limit

$$\begin{aligned}&\lim_{\Delta\to 0}\frac{1}{\Delta}\left(P_k(x)-P_k(x+\Delta)\right)\\ \geq\ &\lim_{\Delta\to 0}\frac{1}{\Delta}\left(S^{(n)}(x+\Delta^*,...,x,...,x+\Delta^*)-S^{(n)}(x+\Delta^*,...,x+\Delta,...,x+\Delta^*)\right)\end{aligned}$$

and upper bound

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\left(P_k(x)-P_k(x+\Delta)\right)\leq\lim_{\Delta\to 0}\frac{1}{\Delta}\left(S^{(n)}(x,x...,x)-S^{(n)}(x+\Delta,x...,x)\right)$$

By passing the to limit as $\Delta\to 0$, one obtains

$$\frac{\delta}{\delta x_k}S^{(n)}(x_1,...,x_n)\Big|_{x_1=x_2=...=x_n=x}\leq P_k'(x)\leq\frac{\delta}{\delta x_k}S^{(n)}(x_1,...,x_n)\Big|_{\substack{x_k=x\\ x_j=x+\Delta^*}},\ \text{for } j\neq k$$

Limiting to $\Delta^*\to 0$ yields the result.

■

With the total survival function $S(x)$, all kind of interesting information can be deduced. Generally, the probability of dying in the age interval $[0,x)$ equals $1-S(x)$, or

$$\int_0^x S(t)\,\mu(t)\,dt=-\int_0^x S'(t)\,dt=-S(t)\Big|_0^x=1-S(x)\tag{2.14}$$

where the total mortality intensity is defined by

$$\mu(x)=\lim_{\Delta x\to 0}\frac{1}{\Delta x}\Pr\left[\text{die in } [x,x+\Delta x)\mid\text{survival of all risks to age } x\right]$$

For multiple decrement functions, similar equations can be found. First, we define cause-specific mortality intensity $\mu_k(x)$ as the mortality intensity for risk R_k where all risks R_i, $i=1,...,n$, are acting, or

$$\mu_k(x)=\lim_{\Delta x\to 0}\frac{1}{\Delta x}\Pr\left[\text{die in } [x,x+\Delta x) \text{ of risk } R_k\mid\text{survival of all risks to age } x\right]$$

We note that

$$\mu_k(x) = -\frac{P_k'(x)}{S(x)} \tag{2.15}$$

Because the risks are supposed to be mutually exclusive, the mortality intensity corresponding to the total survival function equals the sum of the cause-specific forces of mortality, or component forces $\mu_k(x)$. The total force of mortality $\mu(x)$ is equal to

$$\mu(x) = \sum_{1 \le k \le n} \mu_k(x) \tag{2.16}$$

Sometimes a superscript (τ) is used to denote 'all causes'.

By analogy of (2.14), we find the probability for a new-born dying from risk R_k in the age interval $[0, x)$ by

$$\int_0^x S(t)\,\mu_k(t)\,dt = -\int_0^x P'(t)\,dt = \pi_k - P_k(x) \tag{2.17}$$

where $\pi_k = P_k(0)$ and $P_k(x)$ the probability to eventually die from risk R_k in $[x, \infty)$. We define $Q_k(x)$ as the probability of decrement before age x from risk R_k, so

$$Q_k(x) = \pi_k - P_k(x) \tag{2.18}$$

Although $P_k(x)$ is decreasing in x and has $P_k(0)$ as its highest value, there is no guarantee that $P_k(x)$ will behave like the usual survival function. In the common case, $P_k(0)$ will be unequal to one. Because we have $\lim_{x\to\infty} P_k(x) = 0$, the probability associated with risk R_k can be found by evaluating (2.18) for $x \to \infty$ so that

$$\pi_k = \lim_{x\to\infty} Q_k(x) \tag{2.19}$$

Everyone dies from one of the n causes of death, so using (2.16) and (2.19) gives

$$\sum_{1 \le k \le n} \pi_k = 1$$

The total survival and death probability is the accumulation of all n causes of death

$$\begin{aligned} Q(x) &= \sum_{1 \le k \le n} Q_k(x) = 1 - S(x) \\ S(x) &= \sum_{1 \le k \le n} P_k(x) \end{aligned}$$

For each risk R_k recognised in the competing risks model, it is possible to define a single risk model that depends only on that particular risk R_k. We define the associated single decrement model function as

$$S_k^*(x) = e^{-\int_0^x \mu_k(t)dt} \tag{2.20}$$

Then $S_k^*(0) = 1$ and generally $S_k(x) \neq S_k^*(x)$. Probabilities associated with this function are called net probabilities of decrement or absolute rate of decrement. By definition, we have

$$S^{(n)}(x_1, x_2, ..., x_n) = \prod_{1 \leq k \leq n} S_k^*(x_k)$$

It should be noted that (2.20) is not necessarily a survival function because it is possible that for one risk the integral

$$\lim_{x \to \infty} \int_0^x \mu_k(t)\, dt$$

is finite. Because one seldom has the opportunity to observe a system where only one decrement operates, this function is often used for analysis.

The mortality intensity of the population with only one acting risk R_k cannot be derived from $\mu_k(x)$ because this mortality intensity is based on a population where all risks are acting. Therefore, we define marginal mortality intensity $\mu_k^*(x)$ as the mortality intensity of a population with only one acting risk R_k

$$\mu_k^*(x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \Pr\left[\text{die in } [x, x + \Delta x) \text{ of risk } R_k \mid \text{survival of } R_k \text{ to age } x\right]$$

Then we have

$$S_k(x) = e^{-\int_0^x \mu_k^*(t)dt}$$

As noted in the introduction, between the models described in this section several terminology differences exist via application in actuarial science and in biostatistics and reliability

theory. These differences are summarised in the following table.

	Multiple decrement models	**Competing risks theory**
$S^{(n)}(x_1, x_2, .., x_n)$	multiple decrement function	joint survival distribution
$P_k(x)$	k-th crude survival function	subsurvival function
$S_k(x)$	k-th decrement function	marginal survival distribution
$S_k^*(x)$	single decrement function due to cause k	conditional survival function
$\mu(x)$	force of decrement	hazard function
$\mu_k(x)$	force of decrement due to cause k	sub-hazard function
$\mu_k^*(x)$	marginal force of decrement due to cause k	marginal hazard function

In the following example, several decrement functions will be derived from one joint survival function.

Example 2 *Consider the following bivariate exponential joint survival function (described in [Cro94])*

$$S^{(2)}(x_1, x_2) = e^{-\lambda_1 x_1 - \lambda_2 x_2 - \nu x_1 x_2}$$

in which $\lambda_1, \lambda_2 > 0$ *and* $0 \leq \upsilon \leq \lambda_1 \lambda_2$*. Then the total survival function is*

$$S(x) = e^{-(\lambda_1+\lambda_2)x - \nu x^2}$$

and forces of decrement can be found by applying (2.15), which results in $\mu_1(x) = \lambda_1 + \nu x$ *and* $\mu_k(x) = \lambda_2 + \nu x$*. The single decrement functions equal*

$$S_1^*(x) = e^{-\lambda_1 x - \frac{1}{2}\nu x^2} \text{ and } S_2^*(x) = e^{-\lambda_2 x - \frac{1}{2}\nu x^2}$$

The decrement functions can be found by applying (2.17) which results in an expression with the standard normal distribution function

$$P_1(x) = \frac{1}{2}S(x) + \frac{1}{2}(\lambda_1 - \lambda_2)\sqrt{\frac{\pi}{\nu}}e^{\frac{\lambda_1+\lambda_2}{4\nu}}\Phi\left(-\sqrt{2\nu}\left(x + \frac{\lambda_1+\lambda_2}{2\nu}\right)\right)$$

and

$$P_2(x) = \frac{1}{2}S(x) + \frac{1}{2}(\lambda_2 - \lambda_1)\sqrt{\frac{\pi}{\nu}}e^{\frac{\lambda_1+\lambda_2}{4\nu}}\Phi\left(-\sqrt{2\nu}\left(x + \frac{\lambda_1+\lambda_2}{2\nu}\right)\right)$$

where $\Phi(x)$ *is the standard normal distribution.*

In order to simplify competing risks models, often additional assumptions are needed.

2.3.2 Assumptions of competing risks models

A frequently used measure of the impact of a given risk is the cause-elimination life expectancy gain, the change in life expectancy that would occur if all deaths due to that given cause were prevented. These measures are based on the basic assumption (of competing risks analysis) that if risk R_k is eliminated then this has no effect on $S^{(n)}$ itself but only nullifies the corresponding argument of $S^{(n)}$. Thus, given a population with three acting risks ($n = 3$), the multiple decrement function for a population exposed to risk R_1 and R_3 only is given by $S^{(3)}(x, 0, x)$ and a population exposed to risk R_2 only is assumed to be reigned by 2-th decrement function $S_2(x) = S^{(3)}(0, x, 0)$ and the joint survival function can be identified by

$$S^{(n)}(x_1, x_2, ..., x_n) = \prod_{1 \leq k \leq n} S_k(x_k)$$

However, this basic assumption is always suspect, since measures which eliminate R_k may alter other risks.

Another assumption is the Makeham's assumption. One might expect the risk R_k to differ in two individuals, one who has survived exposure to all risks until x and the other who had survived only exposure to R_k until x. In general their mortality intensities will not be equal, $\mu_k(x) \neq \mu_k^*(x)$, and $\mu(x) \neq \sum_{1 \leq i \leq n} \mu_k^*(x)$. However, because it greatly simplifies subsequent derivations an often used assumption is

$$\mu_k(x) = \mu_k^*(x)$$

This is known as the Makeham assumption because his law is based on it. It says that the two hazard functions for risk k, the one in the presence of the other risks and the other in their hypothetical absence, are equal. If the Makeham assumption holds then

$$S^{(n)}(x, ..., x) = \prod_{1 \leq k \leq n} e^{-\int_0^x \mu_k(t)dt} \overset{Makeham}{=} \prod_{1 \leq k \leq n} e^{-\int_0^x \mu_k^*(t)dt} = \prod_{1 \leq k \leq n} S_k(x) \tag{2.21}$$

so, the k-decrement functions determine the multiple decrement function uniquely. Under the

Makeham assumption one can derive the survival function from $\mu_k^*(x)$ by

$$\mu_k^*(x) \overset{Makeham}{=} -\frac{P_k'(x)}{S(x)}$$

and this set of crude survival functions $P_k(x)$ determines $S_k(x)$ uniquely because by definition it yields that

$$S_k(x) = e^{-\int_0^x \mu_k^*(t)dt}$$

which by (2.21) determines $S^{(n)}(x, ..., x)$.

An additional assumption which relates the marginal force of decrement $\mu_k^*(x)$ to the total force of decrement $\mu(x)$ is from Chiang and described in [Ga75]. He assumed that marginal force of decrement is equal to a weighted intensity $\mu(x)$

$$\mu_k^*(x) = w_k \mu(x)$$

where w_k are constant weights satisfying $w_k > 0$ and $\sum_{1 \le k \le n} w_k = 1$. With the Makeham assumption we have

$$w_k = \pi_k$$

and

$$P_k(x) = e^{-w_i \int_0^x \mu(x)dt} = (S(x))^{w_k}$$

Weight w_k can be identified as a weight for frailty: the higher the weight is, the higher the mortality intensity of risk k is and its influence on the total mortality intensity.

A stronger assumption is the assumption of statistically independence of the risks. This implies the Makeham assumption but the reverse does generally not hold. Only under the assumption of independence, $S_k(x)$ determines the multiple decrement function $S^{(n)}(x_1, ..., x_n)$ by

$$S^{(n)}(x_1, ..., x_n) = \prod_{1 \le k \le n} S_k(x_k)$$

The assumption of statistical independence greatly simplifies the model but in practical cases, one can not prove that there is independence between the risks. In the next chapter, problems with the measurement of competing risks models will be discussed.

2.3.3 Relaxing the assumptions

Without independence assumptions, bounds are derived, as described by Peterson in [Pet75]. A multiple decrement function with fixed crude survival functions $P_k(x_k)$, $k = 1, ..., n$ is bounded above and below as follows

$$\sum_{1\leq k\leq n} P_k(x)\Big|_{x=\max(x_1,...,x_n)} \leq S^{(n)}(x_1, ..., x_n) \leq \sum_{1\leq k\leq n} P_k(x_k)$$

and the k-th survival probability is bounded by

$$\sum_{1\leq i\leq n} P_i(x) \leq S_k(x) \leq P_k(x) + (1 - \pi_k) \tag{2.22}$$

For example, for case $n = 2$, we obtain the upper bound of the joint survival function

$$S^{(2)}(x_1, x_2) \leq P_1(x_1) + P_2(x_2)$$

By observing that for $x = \max(x_1, x_2)$ the upper bound

$$S^{(2)}(x, x) \leq S^{(2)}(x_1, x_2)$$

gives with equation (2.13) the following lower bound

$$P_1(x) + P_2(x) \leq S^{(2)}(x_1, x_2)$$

The bounds for $S_k(t)$ follow from this immediately by setting the arguments of $S^{(n)}(x_1, ..., x_n)$ to zero except for x_k.

The Peterson bounds are sharp under continuity of the subsurvival functions P_k, in the sense that through every point between the two corresponding lower and upper bound functions there passes the marginal distribution function of the corresponding lifetime variable, for some feasible multiple decrement function. However, these pointwise sharp bounds are not functionally sharp, that is, not every distribution function between the Peterson bounds is the marginal distribution $S_k(x)$ of some feasible joint distribution: a feasible marginal leaves a nondecreasing gap with its corresponding Peterson bound, because rewriting (2.22) leads to

$$0 \leq S_1(x) - S(x) \leq Q_2(x)$$

and

$$0 \leq S_2(x) - S(x) \leq Q_1(x)$$

where $Q_k(x)$ are nondecreasing functions defined by $Q_k(x) = \pi_k - P_k(x)$ in (2.18).

Based on this result, Bedford and Meilijson's main result in [BM96] is a complete characterisation of the possible marginal distribution functions of these variables with given subsurvival functions, without any assumptions on the underlying joint distribution. However, with a set of feasible marginals still infinitely many joint distributions are possible.

The Peterson's bounds can be regarded as the most extreme models of dependency corresponding to fixed crude survival functions. The bounds only coincide if $x_1 = x_2$. The bounds show how large the error can be if independence is wrongly assumed. To discover where between the bounds in reality are, additional information is needed.

Example 3 *Consider the following subsurvival functions which are derived from a rewritten Gompertz' law of mortality*

$$P_1(x) = \frac{9}{10} e^{e^{-7.5}} e^{-e^{\frac{x-75}{10}}}$$

and

$$P_2(x) = \frac{1}{10} e^{e^{-5}} e^{-e^{\frac{x-50}{10}}}$$

then $\pi_1 = 9/10$ *and* $\pi_2 = 1/10$. *The first sub-population has a life expectancy of approximately* 75 *and the second of approximately* 50 *(see section 2.5.1). Then the following bounds exist*

$$P_1(x) + P_2(x) \leq S_1(x) \leq P_1(x) + \pi_2$$

This is illustrated in the following figure. It is easy to see that the gap between the lower and upper bound is nondecreasing.

The assumption that the decrement functions S_k are continuous can be dropped (see Miller in [Mil77]). The assumption that death does not result from simultaneous risks can also be dropped (see Langberg, Proschan and Quinzi in [LPQ81]). Because typical applications occur in reliability theory, this will not be discussed here.

2.3.4 Discussion

Multiple decrement survival functions arose from the desire to zoom in on detailed causes of mortality. By modelling observable causes of death, it was attempted to find a better reflection

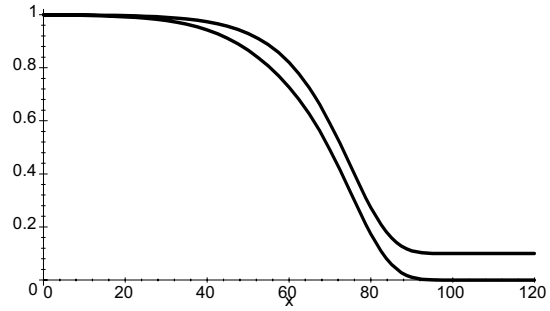


Figure 2-1: The Peterson bounds for $S_1(x)$ in case of two Gompertz sub-populations

of reality. One effect of differentiating between the causes was the cause elimination model, whereby specific causes of death were removed from the survival function.

The original observation made when developing multiple decrement survival functions is that cause-specific survival functions can be simple while the total survival function can have a complex form. The separate survival functions need not always be akin to each other. In general, the mortality intensity is not the same as the mortality intensity per cause of death derived from the survival function. The problem shifts to one of defining the survival functions of each cause of death.

The main disadvantage of multiple decrement survival functions is that these functions are based on the assumption that every individual dies of one unique risk or cause of death. By defining a survival probability for each cause of death, we actually assume that individuals will die of something that can be defined in advance. This is not a given fact, and there is not always an observable biological basis for this assumption. The dependencies between the causes of death need to be included in the model. We have still not managed to answer questions such as what happens if a certain cause of death disappears.

It should be noted that there are some differences between the terms risk and cause of death. The classification system of causes of death may not permit each cause of death to be defined as a risk. Secondly, one may not wish to deal with detailed categories of causes when defining risks in a given analysis. Some aspects of disease dependence can be resolved through appropriate risk definitions.

In the usual formulation of the cause-elimination survival model given above, one may assume that the risks are independent. This could be problematic for analysis of underlying-cause mortality data because risks that are truly dependent may be mistakenly treated as independent. When dependent risks are taken to be independent, serious errors can be made in estimating the survival functions.

An improvement of this form the lethal-defect pattern-of-failure models described by Manton and Stallard in [MS84]. The patterns-of-failure are defined by combinations of one or more diseases. If a pair of diseases is suspected to be dependent, then the patterns with that disease pair can be used to define a single group that represents the mortality risk of the disease pair as an independent physiological process. The pattern-of-failure model, in effect, views the question of disease dependency as an issue of disease definition. Rather than dealing with mortality risks as separate diseases, this model deals with mortality risks as 'lethal defects' defined by combinations of diseases. From this perspective, the analysis of single diseases involves a specialisation of the lethal defect definition to patterns with the disease occurring only as the underlying cause. The lethal defects are identified with independent physiological processes so that the pattern-of-failure model is, by definition, a model of independent mortality risks.

Prentice and Kalbfleisch argued in [PK78] that consideration of multiple decrement function $S^{(n)}$ is unjustified because it only is a theoretical function which can never be observed directly. Prentice considered the interpretation of the life expectancy of a population with acting risk R_k. This would be observed if the other risks where removed. The quantities like the probability to die of a risk when only that risk is operating are called 'non-observable' on the basis of the convention that only functions of the cause-specific forces of mortality corresponding to the original observations when all risks are acting are 'observable'. Removal of some risks, even if possible, might alter the circumstances in which the remaining risks operate so that the new risk cannot be identified with the old one. Secondly, the interpretation of the life expectancy as a sort of potential variate which only acquires physical existence when death arises from risk R_k was criticised on the grounds that removal of some risks might change the marginal distributions of the remaining S_k from their former ones.

Concluding, the criticism of Prentice and Kalbfleisch boils down to the overall lack of physical meaning of the unobserved life expectancies. Their aim is to set up models for only observable phenomena. Because the components of multiple decrement models are non-observable, one should ignore them and build models based on mortality intensities for which it is assumed that they simplify statistical modelling and inference.

Although the lack of identifiability is a general problem of competing risks models and heterogeneous population models (to be described in the next sections), it can be weakened by using regression techniques (see for example Crowder in [Cro94]). One aim of modelling in this thesis is to simplify existing models which allows applications of these regression techniques. In this sense, it is true that the multiple decrement function is possibly a non-realistic and too complicated form. Instead of this function, mortality intensities are easier to handle. However, to build models based on mortality intensities obviously has the same disadvantages as competing risks models, that is that removal of certain risks will also influence the remaining mortality intensities. For example, one might well be prepared to believe that there are real ageing and disease processes simultaneously active in a human body. The processes probably interact and in a different individual where perhaps some are absent and other ones present, the processes common to both individuals might well run differently. In this sense, the lack of identifiability is more oppressive in population dynamics than it is in reliability theory.

The identification problem is a result of the lack of biophysical meaning of concepts used in a model. As we have seen, concepts used in actuarial models (based on mortality laws) often do not refer to biophysics. Competing risks models also lack justified meaning of the concepts used. What risks or causes do we have to choose in order to use the independence assumption and does this not limit our applications too much? The models discussed here are all theoretical and stand alone; no assumptions about possible measurements are used. Therefore, neither the significance nor the (possible) truth depends on the real world, i.e. something outside of the model. Creators base these models on quantitative concepts, which still have to be related to measurements. The quantities cannot be estimated without additional assumptions. For example, in case of competing risks models one may assume that only the minimum of these failure times is observed. These considerations are introduced when the quantities are to be measured and therefore subject of next chapter.

2.4 Heterogeneous population models

All the models discussed above are based on homogeneous sub-populations. The models presuppose that there is no difference between individuals of a population. All individuals of the same age (and gender and cause of death) in a population have the same mortality probability. This means that assumptions about a single individual can easily be generalised to populations.

The first mortality models, the cause elimination model and the multiple decrement survival functions are all examples of models with discrete states. The division into sub-populations was

based on age, gender and cause of death. By differentiating between the states, it was possible to give a mortality probability for each state. The main consideration in this type of model is the temporary change in the conditional density of the individuals per age, gender and so on, given the situation at the beginning of such a group. In a traditional life table, the initial value is $S(0)$ and a definition is made of the probabilities of transition to other states, such as the next year of life. The individuals that reach or leave a state at a given time are viewed as identical. Individuals cannot be distinguished and that is why it is not possible to recognise individual differences.

This section describes mortality models based on so-called heterogeneous populations. Heterogeneous populations consist of a number of sub-populations assumed to be homogeneous. It is assumed that the individuals in the homogeneous populations have the same properties (with respect to the probability of death). The difference with homogeneous models is that in heterogeneous models the probability of specific types of illness or death are related to the value of one or more physiological variables that are relevant to health, such as blood pressure, or cholesterol levels. These models are therefore explicitly individual, since they use data at this level.

This raises the question as to what the relationship is between the observed mortality probabilities of a population over a given period and the mortality probabilities of (homogeneous) sub-populations. It is possible that mortality probabilities of these smaller sub-populations are easily described to enable prediction [VY83].

2.4.1 General model

Involving the effects of individual differences in models is known as the problem of population heterogeneity. Here, our main interest is in the differences between the mortality probabilities of sub-populations. If sub-populations have different mortality probabilities, it is possible that there is some sort of selection between them. The population with the best physiological attributes will, after time, survive the others and so form the majority of the total population. This happens, for example, when the mortality probabilities for the total population are measured. Since men have a higher mortality probability, the main mortality probabilities measured over time will be those of the female sub-population.

Below, we endow death with a real valued risk factor z, contrary to the enumeration variable k defined previously. We start with a single group at age x with individual mortality intensities

$\mu_z(x)$. The conditional mortality intensity shows the probability of death with a risk factor z

$$\mu_z(x) = \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr[\text{an individual with risk factor } z \text{ dies during interval } [x, x+\Delta)]$$

The conditional mortality intensity is $\mu_z(x) = -\frac{\delta}{\delta x} S_z(x) / S_z(x)$. This leads to

$$\mu_z(x) = -\frac{\delta}{\delta x} \ln S_z(x)$$

The conditional mortality hazard function $H_z(x)$ is defined by

$$H_z(x) = \int_0^x \mu_z(t)\, dt$$

and the conditional survival function is

$$S_z(x) = e^{-H_z(x)}$$

For risk factor z a risk density $f_x(z)$ is defined to reflect the density of z under the groups living at the age of x. The heterogeneous mortality intensity can be found by integrating the conditional mortality intensity multiplied by the frailty density

$$\overline{\mu}(x) = \int_{-\infty}^{\infty} \mu_z(x) \cdot f_x(z)\, dz$$

where z ranges over the interval $(-\infty, \infty)$. When looking at possible choices of $\mu_z(x)$, this interval can be restricted to, for example, positive values. Similarly, the heterogeneous hazard function is defined by

$$\overline{H}(x) = \int_0^x \overline{\mu}(t)\, dt$$

The heterogeneous survival function is

$$\overline{S}(x) = e^{-\int_0^x \overline{\mu}(t) dt}$$

In a general heterogeneous model the mortality intensity is conditioned by a risk factor z. The heterogeneous survival function is defined by weighting the mortality intensity with a density $f_x(z)$. It is essential to determine how the conditional mortality intensity is defined. We will now look at choices of $f_x(z)$ and $\mu_z(x)$.

2.4.2 The frailty model

This section describes the application of a general model of heterogeneous populations. It is based on the model that Vaupel presents in [VY83]. This model describes heterogeneity with the help of a non-observable notion of frailty. A base-line mortality intensity $\mu(x)$ is used (it is assumed that $\mu(x)$ is defined properly). From this mortality intensity, the conditional mortality intensities can be derived. The conditional mortality intensity is

$$\mu_z(x) = z \cdot \mu(x) \tag{2.23}$$

If the mortality intensity is an exponential function then (2.23) results in age correction. The expected mortality intensity of the entire population then is

$$\overline{\mu}(x) = \int_0^\infty \mu_z(x) \cdot f_x(z)\, dz \tag{2.24}$$

with $f_x(z)$ as the conditional probability density for the relative risk at age x and z is confined to positive values (otherwise a negative intensity would be possible). If we take (2.23), then the average of z is

$$\overline{z}(x) = \int_0^\infty z \cdot f_x(z)\, dz$$

Then (2.24) can be simplified to

$$\overline{\mu}(x) = \overline{z}(x) \cdot \mu(x) \tag{2.25}$$

Therefore, z can be seen as a measure of the 'frailty' of a population. When $\overline{z}(x)$ does not depend on x, a proportional hazard model is obtained by

$$\overline{H}(x) = \overline{z} \cdot H(x) \tag{2.26}$$

where $H(x)$ is the base-line hazard function defined by definition (1.3).

The proportional hazard model was introduced by Cox in 1972 [Cox72]. In this model, the average frailty $\overline{z}$ is estimated by explanatory variables and regression techniques. Measuring the average frailty is the subject of the following chapter; here we are only interested in the model class M of which (2.26) is an element.

The relationship between the mortality intensity of the entire population $\overline{\mu}(x)$ and the base-line intensity $\mu(x)$ depends on the distribution used for the relative risk z. A number of researchers have noted that the Gamma distribution is the most suitable, but that others can

be used too, such as the inverse Gaussian distribution in [Hou84].

In the following, we assume an initial and unchanging frailty at birth. To find the conditioned probability density of $f_x(z)$ an initial probability density $f_0(z)$ is used. The probability density, given age x, is

$$f_x(z) = \frac{f_0(z) S_z(x)}{\overline{S}(x)} \tag{2.27}$$

where the heterogeneous survival function is $\overline{S}(x) = \int_0^\infty f_0(y) S_y(x) dy = \psi_{f_0}(-H(x))$, with ψ_{f_0} as the moment generating function of $f_0(z)$. Throughout this thesis, the symbol ψ will refer to a moment generating function.

In [VMS79] Vaupel, Manton and Stallard assume an unchanging frailty and that this frailty can be represented by a Gamma density (2.10) at age 0, thus

$$f_0(z) = \frac{\lambda^k z^{k-1} e^{-\lambda z}}{\Gamma(k)}$$

If $\lambda = k/\overline{z}(0)$, then at birth $\overline{\mu}(0) = \overline{z}(0)\mu(0)$. The Gamma density depends on the average frailty $\overline{z}(0)$ and shape parameter k. If the average frailty at birth is set to one, i.e. $\overline{z}(0) = 1$, it can be demonstrated that the average frailty at age x is equivalent to

$$\overline{z}(x) = \frac{k}{k + H(x)} \tag{2.28}$$

because, with (2.27), we find

$$f_x(z) = f_0(z) \frac{e^{-zH(x)}}{\psi_{f_0}(-H(x))} = \frac{(k + H(x))^k z^{k-1} e^{-(k+H(x))z}}{\Gamma(k)}$$

which has expected value $\overline{z}(x)$.

The heterogeneous mortality intensity can be obtained by substituting (2.28) into (2.25)

$$\overline{\mu}(x) = \frac{k}{k + H(x)} \mu(x)$$

The heterogeneous survival function is

$$\overline{S}(x) = \left(1 + \frac{1}{k} H(x)\right)^{-k}$$

with $H(x) = -\ln S(x)$. It therefore follows by substitution that

$$\overline{\mu}(x) = \mu(x) \cdot \left(\overline{S}(x)\right)^{1/k}$$

If the Gamma distribution is used and if the mortality intensity corresponds with Gompertz' law of mortality, this leads to paradoxical conclusions. It is possible to demonstrate that the independent sub-populations aged 'quicker' than the entire population. In other words, the curve of the survival function for the separate sub-populations would drop more rapidly than the one for the overall population. A population can have a high mortality probability even if the mortality probabilities for all its sub-populations are lower for all ages. It is also the case that the mortality probability for the entire population can rise, even if the probabilities in all sub-populations fall [VY83].

2.4.3 Other definitions of conditional mortality intensity

The above definition of conditional mortality intensity, where the base-line intensity is multiplied by the risk factor, is not the only option. Below, we will take a closer look at another choice of $\mu_z(x)$. Instead of multiplying z by the mortality intensity, the risk factor z can also be added to the mortality intensity and thereby increasing the probability to die at age x. Then we have

$$\mu_z(x) = \mu(x) + z \tag{2.29}$$

The following heterogeneous mortality intensity can be obtained by substituting (2.29) into (2.24)

$$\overline{\mu}(x) = \mu(x) + \overline{z} \tag{2.30}$$

In this case, the average frailty does not depend on age x and the heterogeneous hazard function is defined by

$$\overline{H}(x) = H(x) + \overline{z}x$$

If $\mu(x)$ is an exponential function, as is the case in Gompertz' law of mortality (2.2), then (2.30) corresponds with Makeham's mortality law (2.3), which is defined by $S(x) = s^x g^{c^x}$. If we rewrite the parameters according to $\overline{z} = -\ln s$, and g and c as in Gompertz' law of mortality, this gives us

$$S(x) = e^{-\overline{z}x} g^{c^x} \tag{2.31}$$

According to (2.31), the mortality intensity is

$$\overline{\mu}(x) = -c^x \ln c \ln g + \overline{z} \tag{2.32}$$

In the theory of heterogeneous populations Makeham's mortality law (2.3) can be derived from Gompertz' law of mortality. The average correction $\overline{z}$ serves as a value with which the mortality intensity is increased. This increases our insight into Makeham's mortality law. Makeham's law supposes that the $f_x(z)$ density is not important as long as the average $\overline{z}$ is independent of age x.

The constant $\overline{z}$ added to the base-line mortality intensity can also be represented by the multiplication factor described previously. This results in

$$\overline{z}'(x)\,\mu(x) = \overline{\mu}(x) = \mu(x) + \overline{z}$$

Substitution of (2.32) gives

$$\overline{z}'(x) = 1 - \frac{\overline{z}}{c^x \ln c \ln g}$$

The mortality intensity increases with age x and the $\overline{z}'(x)$ is changed. Figure 2-2 gives an example of $\overline{z}'(x)$ for a base-line population with a life expectancy of 75, standard deviation of 15, and average vitality of $\overline{z}'(x) = 0.0001$. The influence is particularly high in the first years of life.

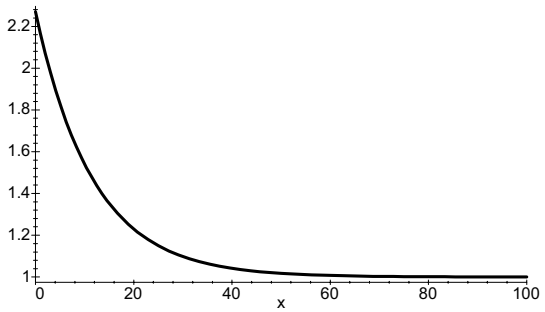


Figure 2-2: The multiplication factor as a result of the addition of the vitality with $\mu = 75$, $\sigma = 15$ and $\overline{z} = 0.0001$.

Because of the positive $\overline{z}$, the mortality intensity is multiplied by a high value for low ages. Since the intensity rises the effect of the constant is decreasing. Therefore, the $\overline{z}'_x$ converges to the value 1 for high ages. The improvement brought by Makeham's mortality law is mostly in the first few years, and this is precisely the area where Gompertz' law does not correspond well to measurement data.

The (non-observable) risk factor can also be represented by an increase in life expectancy. The higher the risk factor the older a sub-population is assumed to be. The age correction reduces to

$$\mu_z(x) = \mu(x - z)$$

for $z \in (-\infty, \infty)$.

In actuarial science, age correction is probably the most common way to specify the mortality pattern of a group. If using an exponential mortality intensity, the age correction does not differ from Vaupel's heterogeneous population mortality model, because

$$\mu_z(x) = \mu(x - z) = \mu(-z)\,\mu(x)$$

Then, the heterogeneous mortality intensity is defined by

$$\overline{\mu}(x) = \overline{z}(x)\,\mu(x)$$

where $\mu(x)$ is Gompertz' mortality intensity and the average $\overline{z}(x)$ is equivalent to the moment generating function of the probability density $f_x(z)$ because

$$\overline{z}(x) = \int_{-\infty}^{\infty} e^{-z \ln c} \cdot f_x(z)\, dz = \psi_{f_x}(-\ln c)$$

where ψ_{f_x} is the moment generating function of $f_x(z)$.

If Gompertz' mortality law is used, then

$$S_z(x) = S(x - z) = g^{c^{x-z}}$$

This results in

$$S_z(x) = (S(x))^{\frac{G(z)}{\ln c}}$$

with $G(z) = \ln c e^{-z \ln c}$, an exponential distribution with parameter $1/\ln c$. Also, $\mu(-z) = G(z)/\ln c$.

If the probability density for the risk factor is exponential with parameter β

$$f_0(x) = \frac{1}{\beta} e^{-\frac{1}{\beta}x}$$

then a substitution of $x = G(z)$ gives

$$f_0(G(z)) = \frac{1}{\beta} e^{-\frac{1}{\beta}(\ln c) e^{-z \ln c}}$$

The moment generating function for a Gamma distribution is identified by

$$\psi(t) = \frac{1}{(1-\beta t)^{\alpha+1}} = \int_0^\infty e^{ty} e^{-\frac{y}{\beta}} \frac{y\alpha}{\beta^{\alpha+1}\Gamma(\alpha+1)} dy$$

A general property of the Gamma distribution is

$$\psi_{f_0}(tG(z)) = \int_0^\infty e^{tG(z)} f_0(G(z)) \, dG(z)$$

If we assume that $\beta = 1/\ln c$ then there is a simple solution to this integral because this yields $f_0(y) = -G'(y)/\ln c$. Making use of this gives

$$f_x(z) = \frac{\left(\ln^2 c\right) f_0(z) S_z(x)}{-\psi_{f_0}\left(\frac{\ln S(x)}{\ln c} + \ln c\right)} = f_0(z) \frac{S_z(x) \ln S(x)}{S(x) - 1}$$

Substitution with heterogeneous mortality intensity

$$\overline{\mu}(x) = \frac{\mu(x) \ln S(x)}{S(x) - 1} \int_0^\infty f_0(2z) \, e^{G(z)\frac{\ln S(x)}{\ln c}} dz$$

gives

$$\overline{\mu}(x) = \ln c - \frac{S(x)\mu(x)}{1 - S(x)}$$

This gives an expression of $\overline{S}(x)$

$$\overline{S}(x) = c^{-x} \frac{1 - S(x)}{1 - S(0)}$$

Above, the age correction was the similar for every age. In reality, this correction will probably have less of an effect. The disadvantage with the risk factor is therefore smaller.

Given a probability density $f_x(z)$ we define the heterogeneous intensity by

$$\overline{\mu}(x) = \int_{-\infty}^{\infty} \mu_z(x) \cdot f_x(z)\, dz$$

We can also model the heterogeneous population by using a different correction for every age. For the heterogeneous mortality intensity, we may use a general case, i.e.

$$\mu_z(x) = \mu(x - w(x)\, z)$$

where $w(x)$ is a continuous age-dependent function.

If we substitute $\overline{\mu}(x) = \mu(x - g(x))$ and solve $g(x)$ this gives

$$g(x) = -b \ln(\psi_{f_x}(-w(x) \ln c))$$

Another definition for the conditional mortality intensity is

$$\mu_z(x) = \mu(x)^z \tag{2.33}$$

For $z > 1$, the standard deviation decreases and for $0 < z < 1$ the standard deviation increases. For $z = -1$, the reciprocal of $\mu(x)$ is obtained. A first sight, negative values for z are not very useful because this results in very low mortality for high ages. However, they could be useful for modelling childhood mortality as will be described later.

In theory, this definition can also be used in Vaupel's model. One case in which it cannot is described in [MSV86], and it is as follows

$$\mu_z(x) = z\mu(zx)$$

Several ways to link the risk distribution $f_x(z)$ to the mortality intensity have been described. If the frailty was added to an exponential intensity, it is shown that this resulted in a well-known law of mortality. More laws can be explained likewise. Furthermore, when multiplying the frailty to the mortality intensity and the frailty was assumed independent of the age, a proportional hazard model appeared. We have now reached a point where all the different models can be forged together to form a general model.

2.5 A general model of mortality

If we look back at history we see that the first life tables were created from observed data. Good models corresponded well to what was observed in reality. In view of the limited information that was available on death, it was not possible to expect too much accuracy. The models consisted of tables because so little was known that no laws had yet been discovered.

In the nineteenth century the argument changed. Hidden forces in mortality were revealed and all kinds of laws were discovered. The reason why these models were used lies in the explanation and significance that they gave to mortality. Improvements were made through successive models with deeper and deeper explanations.

When data that is not simply age-structured became available, there arose a desire to include this data too. The models did not appear to be good enough and researchers started looking for a better correspondence to reality. Detailed data became available and researchers started to differentiate more variables, such as cause of death and living conditions.

In what follows, a model will be formulated with a high level of knowledge and for detailed data. To do this we need to link together concepts from reality and concepts within the model. In a heterogeneous population, there is an explicit link between the vitality of a population and its mortality intensity. It is possible to relate physiological variables to this vitality. In this way, Gompertz' law of mortality is expressed as a probability distribution below. Using this new definition of Gompertz' law of mortality a general (heterogeneous) mortality model is developed below.

2.5.1 Gompertz' law of mortality reformulated

Many mortality models present death as a shock resulting in a jump in the observed mortality quotients. Shock models are not suitable for the analysis of chronic degenerative illnesses and death, in, for example, older populations. Manton [MS84] suggests to use continuous (process oriented) models. A mortality model should represent the death of an individual as a continuous process. Janssen explains in [JS95] that shock models are essentially static, i.e. not process oriented. Dynamic models, on the other hand, model the state of health of an individual. In Janssen's model 'death' takes place when a stochastic process reaches a critical value for the first time.

Here Gompertz' law of mortality is being cast in a different mould, making it possible to apply the law easily to homogeneous sub-populations or individuals. This leads to a dynamic, process-oriented model in which the state of health of a sub-population or individual can be

modelled explicitly. Physiological variables can be tied in directly to the mortality law.

The meaning of parameters in Gompertz' law is explained below. By eliciting this knowledge, we can look at the law in a knowledge-oriented way. For this explanation recent research from extreme value theory is used [RT97].

Gompertz' law of mortality

For more than one hundred and fifty years Gompertz' law of mortality is the simplest and most effective representation of mortality probabilities. Derived from it we have Makeham's mortality law (2.3), which is still used as a model today. Though there are other, more accurate models, none of them combines as much accuracy with so few variables (in fact just two).

Gompertz' law of mortality is known most of all for its low computational complexity. However, not much is known about its analytical formulation. Gompertz gave a biomedical explanation based on the increasing inability to withstand destruction, but, as we said earlier, this does little to explain how the law works.

Definition In [Gom1825] Gompertz proposed the following law

$$S(x) = k \cdot g^{c^x} \tag{2.34}$$

with g, c and k being real-valued parameters to be estimated. The $S(x)$ indicates the number of survivors aged x.

Gompertz' law of mortality is mainly used to describe the dying process in populations. Contrary to Gompertz' notable explanation, his law is explained in terms of populations. We will start with a population change $\frac{d}{dx}N(x) = a \cdot N(x)$ with an age-dependent changing $a > 0$.

The law can be seen as a growth model by the relationship of $N(x) = N(0) - S(x)$ where $N(x)$ equals the number of survivors up to the age of x and $N(0)$ the starting population. Change in the size of population $N(x)$ is described by

$$\frac{d}{dx}N(x) = a \cdot N(x) \tag{2.35}$$

with a as the growth coefficient. If the growth coefficient itself changes with age, which we illustrate by the relationship of $\frac{d}{dx}a(x) = -\lambda \cdot a(x)$ with λ, the decrement coefficient of the growth coefficient, then $a(x) = a(0) \cdot e^{-\lambda x}$, with $a(0)$ as its initial value. By substituting the

decremental growth coefficient with (2.35) Gompertz' law of mortality is obtained

$$\frac{d}{dx}N(x) = a(0) \cdot e^{-\lambda x} \cdot N(x)$$

This can be rewritten to

$$N(x) = N(0) \cdot e^{-\frac{a(0)}{\lambda}e^{-\lambda x}}$$

This is the same law as (2.34) but with different parameters $c = e^{-\lambda}$, $g = e^{-a(0)/\lambda}$ and $k = N(0)$. Gompertz' law of mortality therefore is based on the assumption of an exponential decremental growth coefficient.

Properties Gompertz' law of mortality has some interesting properties. If the law is represented as a curve, it takes on an S shape, which is not symmetrical, contrary to the frequently used logistic curve. Gompertz' law of mortality assumes that the period of increasing mortality is longer than the period of decreasing mortality. The number of deaths rises slowly and then falls rapidly. Gompertz' law of mortality ensures that the mortality intensity $\mu(x)$, defined by (1.2), increases exponentially.

The double exponential function makes the law a nice representation of the so-called step function. For the (basic) function

$$F(x) = e^{-e^x} \tag{2.36}$$

the following holds

$$\lim_{x \to -\infty} F(x) = 1 \text{ en } \lim_{x \to \infty} F(x) = 0$$

Around 0, the function decreases from 1 to 0, while the speed of the decrement becomes higher. The function $1 - F(x)$ is also known as the converse Gumbel distribution (described in [RT97]) and is frequently used to find extreme values of certain processes.

Diagram (2-3) gives the graph of $F(x)$ and its derivative $f(x) = -e^x e^{-e^x}$. From this, it is obvious that the curve is not symmetrical.

The derivative corresponds with a pulse function. Pulse functions are used in signal theory. For example, the Gaussian function (standard normal distribution)

$$e^{-ax^2},\ a > 0 \tag{2.37}$$

As opposed to the asymmetrical $\frac{d}{dx}F(x)$, (2.37) is symmetrical.

If the curve of Gompertz' law of mortality is compared with the curve given by the survivors

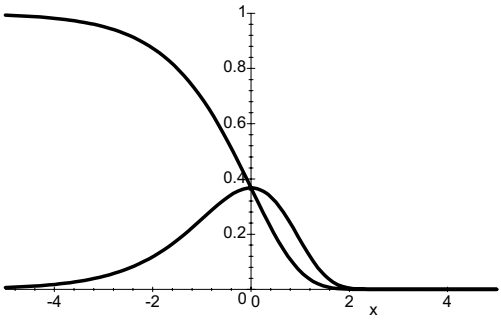


Figure 2-3: Gompertz' curve and its derivative.

of an observed group of people, it seems that there are many similarities. The older the population gets, the more deaths will be observed. After every death, the population will actually drop, and so at a given time there will be a turning point at which the number of deaths will fall. This fall continues until everyone in the population dies.

One difference between Gompertz' law of mortality and the observed population is that the mortality law never reaches the exact value of zero. Every population dies out, i.e. the size of the population reaches exactly the value of zero. However, this is not of practical importance because the values that the law produces at very old ages approximates zero very well.

One well-known property of Gompertz' law of mortality is that it meets the so-called 'Law of Uniform Seniority'. This implies that actuarial sizes dependent on n ages can be reduced to sizes dependent on n equivalent ages, so

$$k \cdot g^{c^{x_1}+c^{x_2}+\ldots+c^{x_n}} = k \cdot g^{nc^z}$$

with

$$z = \frac{\ln\left(\frac{1}{n}\sum_{1\leq i\leq n} c^{x_i}\right)}{\ln c}$$

Makeham's mortality law also has this property.

To fit Gompertz' law of mortality to data, g and c must be estimated (the parameter k is removed for simplicity). The parameters indirectly describe the age at which the highest

number of individuals will die, and the age distribution of the deaths. Traditionally, the least squared method is used to fit data and iteration is used to find which values of the law fit the data best, but if knowledge is available then modelling the parameters in a coherent way leads to an alternative method to estimate the parameters.

An alternative formulation

As was stated above, the basic function (2.36) is related to the Gumbel distribution known from extreme value analysis. In extreme value theory, one is interested in the statistical analysis of extreme data. For example, in [AH93] Aarssen and de Haan study the maximal life span using extreme value theory. The actual distribution of a maximum is usually replaced by an extreme value distribution. The Gumbel distribution is one the standard extreme value distributions which is identical to Gompertz' law of mortality, as is shown below.

In [Kun98] Kunimura examines the moments of the Gompertz distribution in terms of the original parameters g and c. Here, this is simplified by introducing a location and a scale parameter in order to relate the basic function (2.36) with Gompertz' law of mortality.

Gompertz' law of mortality can be described as a translation of (2.36). As we have shown above, the graph of the derivative of Gompertz' law of mortality gives a pulse. After time, this pulse reaches a turning point, and then falls. Differentiating the pulse function and equating it to 0 identifies this turning point.

It is highly probable that a similar turning point in the population data exists. Suppose the age a indicates the age at which the highest number of deaths is observed. We write

$$F(x-a) = e^{-e^{x-a}}$$

The second step is to identify the variance. For populations of a large size the pulse will have a higher deviation. By introducing an extra dilation parameter b, the variance is accommodated

$$F\left(\frac{x-a}{b}\right) = e^{-e^{\frac{x-a}{b}}} \tag{2.38}$$

Turning point a is the moment at which most people die and does not change with variation in b. Parameter b reflects the variance.

However, the turning point a does not correspond to the true life expectancy and b is not exactly equivalent to the standard deviation. By making a small correction using two constants, equivalence with life expectancy and standard deviation is achieved. The parameters are derived

below. By writing (2.36) as probability distribution, the converse Gumbel distribution appears

$$W(x) = 1 - e^{-e^x} \tag{2.39}$$

We can calculate explicitly the average life expectancy and variance using the moment generating function of (2.39).

Proposition 4 *The moment generating function of the Gumbel distribution is equivalent to*

$$\psi_W(t) = \Gamma(t+1) \tag{2.40}$$

Proof. By substitution of $x = \ln y$ in $\psi_W(t) = \int_{-\infty}^{\infty} e^{tx} w(x)\, dx$, it can be found that

$$\psi_W(t) = \lim_{M\to\infty} \left(\int_1^M y^t e^{-y} dy - \int_1^M y^{-2-t} e^{-\frac{1}{y}} dy \right)$$

The second integral gives by substitution of $y = \frac{1}{z}$, $dy = -z^{-2}dz$, it yields

$$\int_1^M y^{-2-t} e^{-\frac{1}{y}} dy = -\int_{1/M}^1 z^t e^{-z} dz$$

This results in (2.40) with the Gamma function defined by $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ (see also the Appendix).

■

By using (2.40), the moments can be derived.

Corollary 5 *The expected value of the Gumbel distribution is defined by*

$$\int_{-\infty}^{\infty} x dW(x) = -\gamma \tag{2.41}$$

where γ is equivalent to Euler-Mascheroni constant.

Proof. By taking the first derivative of (2.40), we find

$$\int_{-\infty}^{\infty} x dW(x) = \frac{d}{dt}\psi_W(t)\Big|_{t=0} = \int_0^\infty \ln(x)\, e^{-x} dx = -\gamma$$

■

The Euler - Mascheroni constant is defined by

$$\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln n\right) \approx 0.5772156649$$

The constant is seen as important alongside other constants such as π and e (see also the Appendix).

Corollary 6 *The variance of the Gumbel distribution is defined by*

$$Var = \frac{1}{6}\pi^2$$

Proof. By taking the second derivative of (2.40) and using (2.41), we find

$$Var = \frac{d^2}{dt^2}\psi_W(t)\Big|_{t=0} - \gamma^2 = \int_0^\infty \ln(x)^2 e^{-x}dx - \gamma^2 = \frac{1}{6}\pi^2$$

■

The Gompertz distribution is defined using a translated distribution (2.39).

Definition 7 *The Gompertz distribution $G(x \mid \mu, \sigma)$ with average $\mu \geq 0$ and standard deviation $\sigma > 0$ is defined by*

$$G(x \mid \mu, \sigma) = W\left(\frac{x-a}{b}\right) \tag{2.42}$$

where $W(x)$ is the Gumbel distribution defined in (2.39) and the constants a and b satisfy

$$a = \mu + \sigma\gamma\frac{\sqrt{6}}{\pi} \text{ and } b = \frac{\sqrt{6}}{\pi}\sigma \tag{2.43}$$

($\sigma = \frac{\pi}{\sqrt{6}}b$ and $\mu = a - b\gamma$).

The distribution is named after Gompertz because $G(x \mid \mu, \sigma)$ is equivalent to Gompertz' law of mortality. The distribution defined in (2.42) satisfies

$$G(x \mid \mu, \sigma) = 1 - g^{c^x}$$

where

$$g = e^{-e^{-\frac{a}{b}}} \text{ and } c = e^{\frac{1}{b}} \tag{2.44}$$

The moment generated function for distribution is $G(x \mid \mu, \sigma)$ equals

$$\psi_G(t) = e^{at}\Gamma(bt+1) \tag{2.45}$$

Because the mortality intensity is represented by the same symbol as the mean, i.e. life expectancy at age 0, it could easily be confused. To avoid confusion, the mean μ is only used within the Gompertz distribution $G(x \mid \mu, \sigma)$ and the mortality intensity is defined with use of parameters a and b.

For the sake of completeness it should be mentioned that the density $g(x)$ is nonzero where the age is negative, i.e. $x < 0$, and only if this range is included equations like (2.44) hold exactly. Consequently, at age zero the distribution $G(0) > 0$. To deal with this flaw, traditionally the distribution is scaled with parameter k at age zero in order to let $G(0) = 1$. This parameter k equals

$$k = \left(1 - W\left(-\frac{a}{b}\right)\right)^{-1} = e^{e^{-\frac{a}{b}}} \tag{2.46}$$

However, this rescaling leads to minor differences in the life expectancy. Another possibility is to determine the life expectancy such that it equals the life expectancy of only $x \geq 0$.

To do this, i.e. to calculate the life expectancy at an arbitrary age $x \geq 0$, a general result about the Gompertz distribution is needed. This is developed in the 'Applications' chapter of this thesis where formulas for the life expectancy at an arbitrary age are derived.

If we calculate the mortality intensity of (2.42), we find

$$\mu(x) = \frac{1}{b}e^{\frac{x-a}{b}} \tag{2.47}$$

The hazard function is

$$H(x) = \int_0^x \mu(t)\,dt = b\,\mu(t)|_0^x = e^{\frac{x-a}{b}} - e^{-\frac{a}{b}}$$

and the survival function of (2.47) is

$$S(x) = ke^{-e^{\frac{x-a}{b}}}$$

where k is defined in (2.46).

Modelling mortality with the Gompertz distribution

In this section, some examples are given that show how the Gompertz distribution can be used for modelling heterogeneous populations. In the first example, we model a population with a Gompertz distribution. In the second, we will examine how a sub-population with a higher risk (accident mortality) can be modelled. In the last example, childhood mortality will be modelled. At pointed out earlier, in this chapter we are mainly interested in the model structure and not with how the parameters are to be measured.

Example 8 *The Gompertz distribution described above is used to model a recent not graduated life table (for men). The life expectancy and standard deviation (from [Ag97]) are calculated and substituted in the Gompertz distribution. The resulting Gompertz distribution (Figure 2-4), with parameters* $\mu = 74.8$ *and* $\sigma = 13.4$ *and mortality intensity*

$$\mu(x) = \frac{1}{10.5} e^{\frac{x-80.2}{10.5}}$$

, can be compared to this data (child mortality was ignored).

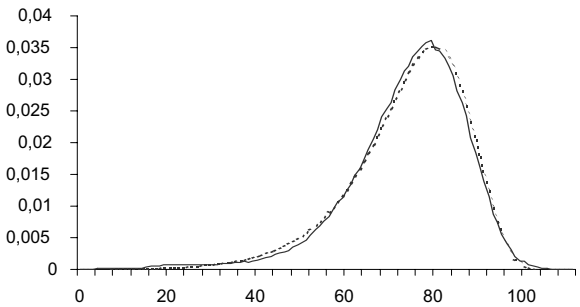


Figure 2-4: The probability for a newborn to die at age x with measured data (solid) compared to Gompertz distribution (dotted)

There are several advantages of the reformulation above.

- The life expectancy and variance are more elementary parameters and meaningful than g and c.

- In life insurance mathematics, application of a distribution is easier than application with help of life tables based on Gompertz because more analytical results can now be obtained.

- It is also easy to generalise these parameters to more general, for example heterogeneous, models with several sub-populations with different life expectancies, as follows.

In the following two examples, the heterogeneous populations is divided into two homogeneous sub-populations. Initial frailty is assumed with probability p to belong to the first population and $1-p$ to belong to the second. By using (2.27), the heterogeneous mortality intensity equals

$$\overline{\mu}(x) = \frac{p\mu_1(x) S_1(x) + (1-p)\mu_2(x) S_2(x)}{pS_1(x) + (1-p) S_2(x)} \tag{2.48}$$

where $S_1(x)$ and $S_2(x)$ are the population corresponding to $\mu_1(x)$ and $\mu_2(x)$.

To model accident mortality for the male population and accident plus maternal mortality for the female population often an additional term is added to the mortality model, see for example the model of Heligman and Pollard in (2.7). The 'accident hump' is found in all populations and appears as a distinct hump in the mortality curve, generally between ages 10 and 40 [HP80]. In the following example it is shown how a sub-population with a high risk can be modelled within a heterogeneous population.

Example 9 *Consider a heterogeneous population with two Gompertz sub-populations. The first population is equal to the population in the previous example and has mortality intensity $\mu_1(x)$. The population with the high risk has parameters a_2 and b_2. Observe that this population can be obtained by transforming the first, with a correction of the life expectancy by $a_1 - a_2$ and multiply the standard deviation by b_2/b_1, or*

$$\mu_2(x) = \mu_1(x + a_1 - a_2)^{b_1/b_2}$$

Initial frailty is assumed with probability p to belong to the first population and $1-p$ to belong to the second. Therefore, the heterogeneous mortality intensity equals (2.48).

When we assume that 5% of the initial total population belongs to the population with the higher risk and that the difference in life expectancy equals 60 years (resulting in a life expectancy of 20 years), and when we further assume that the standard deviation is half of the normal deviation, we find

$$\mu_2(x) = \mu_1(x + 60)^2$$

The 'accident hump' in the mortality rates $q(x)$ can be observed clearly (Figure 2-5) around age 20.

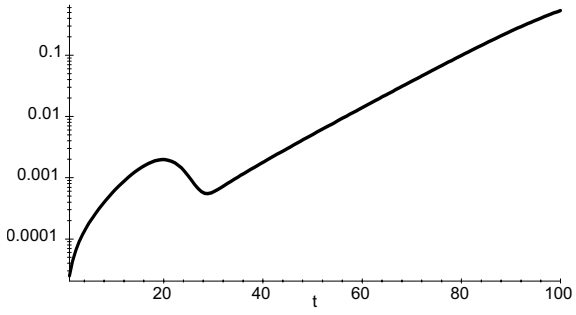


Figure 2-5: The heterogeneous mortality probability $q(x)$ with the accident hump

The example above shows that the second component of Heligman and Pollard's model can be modelled with the theory of heterogeneous populations by using a standard population and a high-risk population. Another example relating child mortality is shown below.

Example 10 *In order to model childhood mortality, we take a heterogeneous population with two sub-populations. The second sub-population represents childhood mortality*

$$\mu_2(x) = \frac{1}{d} e^{\frac{c-x}{d}}$$

(this can also be seen as the reciprocal of a base-line $\mu(x)$ and thus a transformation of the mortality intensity). The first represents the remaining population with parameters defined in the previous example.

For the initial vitality distribution at age 0 *we use a Bernoulli variable with probability p of being in the first population and $1-p$ of being in the second population, so the heterogeneous mortality intensity defined in (2.48) may be used. The first intensity converges to a certain value that represents the total childhood mortality with respect to the total population.*

In order to determine probability p, we define the following constraint. Suppose that $q_c \in [0,1)$ *is the ratio of the population dying during childhood mortality then (for an age* $w > 0$*)*

$$e^{-(1-p)H_1(w)} = 1 - q_c$$

Evaluating this, we find

$$p = 1 + \frac{\ln(1-q_c)}{e^{c/d} - e^{(c-w)/d}} \tag{2.49}$$

If we assume that childhood mortality is placed around age 0 *and for the deviation of the childhood mortality we assume* $b_2 = 1$*. The probability to survive the first two years is according to [Ag97] approximately* 0.992*. Then, for probability p to be in the first population, we find*

$$p = 1 + \frac{\ln(1-0.008)}{1 - e^{\frac{0-2}{1}}} = .99071$$

The resulting distribution (Figure 2-6) shows a fast decreasing survival function in the first year of life.

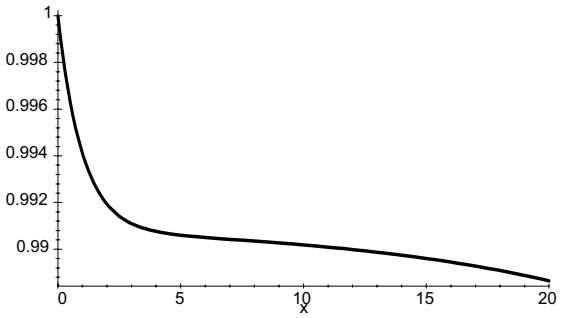


Figure 2-6: The first years of the heterogeneous survival distribution with childhood mortality

Another way to model childhood mortality is to use the same strategy as in accident mortality, that is by using a Gompertz sub-population with a very high risk at the first years of life, for example by using $\mu_2(x) = e^x$.

As the examples above show, the Gompertz distribution is a fine way to represent base-line populations. Specific demographic events, like accident and child mortality, can be modelled by transforming the base-line mortality intensity. It is therefore possible to use the Gompertz distribution for a general model, which is introduced in the next section.

2.5.2 The vitality model

After Gompertz' publication in 1825 [Gom1825], the new mortality laws appeared more accurate than before. This accuracy was mainly increased by adding (new) parameters to the models. However, these parameters did not always refer to an observable quantity. The explanations related the verbal concepts to the mathematical structures but these verbal concepts bear no direct relation with reality.

We give a new overview based on the explicit choice of one model class in the sense of Zadeh (section 2.1.1 of this thesis). A number of the mortality laws we have covered belong to class M. In the theory of heterogeneous populations we have seen how a quantity, that is not directly observable, v (vitality) can be an aid in describing the heterogeneity of a population. Vitality can be seen as the degree of susceptibility to illness and death, but this susceptibility cannot be directly observed.

In this section, we link the heterogeneous model with Gompertz' law of mortality. This gives us a new model that is based on a non-observable and summarising population 'vitality'. As stated earlier the mortality intensity of a heterogeneous population can be defined by

$$\overline{\mu}(x) = \int_{-\infty}^{\infty} \mu_v(x) \cdot f_x(v)\, dv$$

Pearl and Minor [OC97] argued in 1935 against a universal mortality law. They concluded that survival functions for animals never overlapped perfectly, even when life expectancy was adjusted. The most important problem is to bring internal as well as external causes to bear on a model [OC97]. However, it cannot be denied that mortality intensity is exponential in a large domain and the law does apply to many species (including humans) for the majority of their lives.

The model covers a base-line population for which Gompertz' law of mortality is used. The reason for this is that an exponentially increasing mortality intensity is ideally suited to representing a base-line population. It had to be reformulated to clarify its application. Heterogeneity can be modelled by choosing an appropriate vitality distribution. However, it is

possible to use other mortality laws for the mortality intensity.

For the model in this section, we use the base-line intensity derived from Gompertz' law of mortality

$$\mu(x) = \frac{1}{b} e^{\frac{x-a}{b}}$$

However, the mortality law cannot be applied to all ages $x \geq 0$. In very old age and just after birth, the law does not appear to coincide with observations. To battle this, an important question in the study of heterogeneous populations is that of splitting a population into homogeneous sub-populations. The advantage of heterogeneous populations is that they are not restricted to discrete states, and that one or more physiological variables relevant to vitality can be modelled, such as blood pressure and cholesterol levels. Not only that, but other variables, such as income, can be related with the notion of vitality.

The population can also be divided into families with specific, genetically defined mortality patterns. The average age of death for each family is then decisive for the vitality of the family, compared to the total population. In [YI97] an analysis is made of how genetic material and environment affect longevity, using collected data on twins.

Below three methods of presenting the vitality of a sub-population are described. To use Carnap's terminology [Car50]: the 'explicandum' of vitality can be transformed to three 'explicata', i.e. vitality by reducing the intensity, vitality through an increased life expectancy and vitality through a lowering of the standard deviation.

The three methods result in a positive or negative correction of the mortality intensity, which will give the population an advantage or disadvantage compared to the base-line population. The correction can be applied at three levels. For the sake of clarity, different symbols for the three 'explicata' are used.

- Vitality by reducing the mortality intensity

$$\mu_{v_i}(x) = \mu(x) - v_i, \text{ for } x \geq 0,\ v_i \in (-\infty, \mu(0)] \tag{2.50}$$

- Vitality by increasing the life expectancy

$$\mu_{v_a}(x) = \mu(x - v_a), \text{ for } x \geq 0,\ v_a \in (-\infty, \infty) \tag{2.51}$$

- Vitality by lowering the standard deviation

$$\mu_{v_s}(x) = \mu(x)^{v_s}, \text{ for } x \geq 0,\ v_s \in (-\infty, \infty) \tag{2.52}$$

  Notice that when $v_s < 0$ this results in a high mortality for the lower ages and a low mortality for the higher ages.

In the third possibility, vitality is presented as a lower standard deviation with regard to the base-line population. Although life expectancy has not changed by this representation of vitality, and it appears to offer no advantage, this vitality is relevant. The increasing angularity of the survival function, or the drop in standard deviation, is a process that can be observed for some time in many populations.

We add vitality (2.50) to the theory of heterogeneous populations. This creates a new class that embeds existing mortality laws. These laws are seen as attempts to introduce heterogeneity to the population. Starting from (2.50), we get for heterogeneous mortality intensity

$$\overline{\mu}(x) = \mu(x) - \overline{v_i}(x)$$

where $\overline{v_i}(x)$ is the average of the conditional probability $f_x(v_i)$.

By using different approximations of the average $\overline{v_i}(x)$, we newly derive the following mortality laws.

- Makeham's first mortality law (2.3), $S(x) = k \cdot s^x \cdot g^{c^x}$. The mortality intensity of this law is

$$\mu(x) = -c^x \ln c \ln g - \ln s$$

  which, after rewriting the variables, amounts to an approximation of $\overline{v_i}(x)$ by using a constant

$$\overline{v_i}(x) = c_0$$

  with $c_0 = \ln s$.

- Makeham's second mortality law (2.4), $S(x) = k \cdot s^x \cdot w^{x^2} \cdot g^{c^x}$, can be rewritten as

$$\overline{v_i}(x) = c_0 + c_1 x$$

  with $c_0 = \ln s$ and $c_1 = -2 \ln w$.

By taking higher order polynomial terms we can construct more laws with finer approximations of $\overline{v}_i(x)$, as is also discussed by Benjamin and Pollard in [BP86]. In fact, the two mortality laws are part of the same larger class based on average probability distribution v_i.

The second form (2.51) corresponds with Vaupel's model [VY83] because, given the condition of an exponential mortality intensity, we can write (2.51) as $\mu_{v_a}(x) = e^{-v_a/b} \cdot \mu(x)$. With some constraints, Cox' model proportional model [Cox72] can also be derived from the second form. Other mortality laws can be derived from the model. There are further variants, such as $\mu_v(x) = v\mu(vx)$ from [MSV86].

The third form (2.52) is used as an element in other existing laws. As is described earlier, it is possible to model childhood mortality by using a negative v_s. Laws introduced by Thiele, Perks and Heligman and Pollard use an exponential function with negative constants. Only Thiele's model can partially be expressed by the third form.

- (Partially) Thiele's mortality law (2.5), $\mu(x) = a_1 e^{-b_1 x} + a_3 e^{b_3 x}$. An age-independent Bernoulli variable is used for the vitality function

$$\mu(x) = a_1 e^{-b_1 x} + a_3 e^{b_3 x} = \frac{p}{d} e^{\frac{-x}{d}} + \frac{1-p}{b} e^{\frac{x-a}{b}}$$

  with p defined by (2.49). It is not possible to express the middle component of Thiele's model with one of the three forms. However, as the previous example about modelling accident mortality showed, this component can be replaced by adding a high-risk sub-population, which has the same effect as the middle component in Thiele's model.

It is probable that the influence of vitality on mortality will decrease with age. If we look at the three different forms of vitality, there is a difference in their effect on mortality. The effect of correcting the mortality intensity in line with (2.50) will ultimately diminish as age increases. Correcting the age in (2.51) will result in the constant correction for every age. In (2.52), it is only the variance that changes and the change is the constant for each age.

Therefore, only the first form (correcting the mortality intensity) results in age-dependent vitality. This corresponds with the observation that the advantage a sub-population has over a base-line population diminishes over the years. It explains the success enjoyed by the early mortality laws. The early laws of mortality used this form of vitality.

2.5.3 Generalising the model

The general model uses the non-observable and summarising variable of 'vitality'. The reason for this is that we would very quickly become bogged down in causal relations between the external variables if we used them.

Example 11 *In general, people with a low income, low level of education and low social status have a worse state of health and a higher probability of premature death than people who find themselves higher up on the social ladder. However, it appears that socio-economic differences and lifestyles are not sufficient to explain differences in heart attacks and related deaths. Part of the risk of a heart infarction is related to family history and so points to a genetic influence. There is a link between genetic risk and lifestyle, but it is not clear, what that relation is and whether it is self-enforcing.*

When endowing a population with a number of variables the model described above can be used. Generalisation to more variables can be achieved by defining the influence that these variables have on the summarising factor of 'vitality'. This factor affects the mortality intensity in one of the ways described above. However, if we want to describe developments over time this strategy will not do.

Another possibility is that of using a variable to condition the variance of vitality. The variance of vitality is always conditioned by the age x. The variance of v_i can also be conditioned by other variables, such as moment t. A population becomes more vital over the years as the result of medical and socio-economic progress. By relating the moment with the vitality of the population, it is easy to make a prognosis of mortality.

A final generalisation of the new model is possible by use of Markov chains [Wol94]. By distinguishing several states based on associated transition probabilities, individuals with different properties are modelled. In life insurance mathematics Markov chains are used to calculate disability insurance [Ast90] and to model consequences of aids. Models based on Markov chains can be applied well if the populations in the states are very homogeneous. If there is a lack of homogeneity in a sub-population, because the duration of an individual's stay in a given state is relevant, for example, and if certain properties apply continuously, such as smoking and drinking habits, it is better to work from heterogeneous populations. Since a person is not always fully incapacitated (but often has a percentage disability), the heterogeneous model can be used for disability.

2.5.4 Comparison with other models

In this section, the model structure of the general model is compared to other model structures described earlier: competing risks models and multiplicative hazards.

Competing risks models

As with the theory of heterogeneous populations, the competing risks models that we described earlier in section 2.3 use a conditioned mortality intensity $\mu_k(x)$. The difference is that the mortality intensity is conditioned by a natural variable, in our case the risk of death or competing risk R_k. Heterogeneity is represented by a discrete distribution of homogeneous sub-populations. Nothing is said about the relation between cause of death and mortality intensity. It is probable that the mortality intensity per cause of death will differ greatly and defy explanation. The problem reduces to one of defining a survival function for each risk or cause of death. However, that brings us nowhere since the survival function per cause of death is nothing more than the survival function of a sub-population.

Data-oriented competing risks models go astray on the fact that it is not possible to correlate a cause of death with a specific mortality pattern. The independence of the causes of death assumed in competing-risk models is not always accounted for.

The relation between the variable that conditions the mortality intensity and the ensuing mortality intensity is made explicit in the description of heterogeneous populations. The variable's effect on the population is described exactly. By linking physiological variables with vitality, we can develop dynamic models. Not only that, but it appears possible to view the early mortality laws as an attempt to bring heterogeneity to the populations, by introducing an additive model. This leads to a general model that uses vitality as the binding element.

Multiplicative hazards

The general proportional hazard model with lifetime dependent explanatory variables, described in section 2.4, does not prescribe a model class for the base-line hazard. The base-line hazard can be determined (semi)parametrically or nonparametrically. However, if a parametrically base-line hazard (for example based on Gompertz' or Makeham's law of mortality) is used, the proportional hazard model is an element of the general class introduced above. Instead of a distribution of frailty, proportional hazard models assume an averaged frailty, which can be estimated with explanatory variables and regression techniques.

Furthermore, if the Gompertz' law of mortality is used for the base-line hazard, the multiplicative mortality intensity essentially changes the life expectancy. If $\mu_v(x) = v\mu(x)$ is used, the model reduces to a proportional hazard model because

$$\int_0^x \mu_v(t)\,dt = vH_0(x)$$

so when the base-line hazard is exponential, $H_0 = e^{(x-a)/b}$, then the vitality factor v results in an increase of life expectancy of $b \ln v$.

Cox' proportional hazard model [Cox72] with a Gompertz base-line hazard can therefore be seen as a heterogeneous population with unchanging vitality with sub-populations with different life expectancies and a constant standard deviation. In Cox' model the vitality v is equal to a vitality function $\psi(\mathbf{v})$ with set of explanatory factors $\mathbf{v}$.

Disadvantages

One disadvantage of the general model introduced in section 2.5.2, is that the vitality distribution, which is an individual quantity, is not directly identifiable. It is not possible to divide the variation into that within and between individuals, if there is one observation per individual. It can be assumed that the vitality has the same variation for several individuals. Assuming proportional hazards for several groups is another solution.

A drawback of the model is that there is no biological basis for the three possible forms of vitality. The notion of vitality will not affect every age in the same way. If the effect of vitality becomes less and less, we will need to add an extra term to the last two forms, in order to decrease vitality in proportion to the age. It is still not clear how diminishing vitality will affect the mortality probability.

Chapter 3

Measuring

The most elementary form of measurement is the recognition of correspondences and of differences between objects and events which makes naming and classification possible. Beyond this, the ordinal, ratio and interval scale of measurement come into play [Ste62].

Measuring describes how one sees. The attribution occurs by means of rules. The quality of these rules naturally determines to a large degree the quality of the measurement.

3.1 Purpose of measuring

The purpose of actuarial measuring is to connect the concepts described in models - such as mortality probabilities and vitality - with the observed (properties of) objects and events. Previous mortality models, which are described in the chapter on models, embed quantitative concepts. Only when the concepts are explicated, the model can be tested and used for predictions.

Measuring entails imperfections. Even actuarial counting can, via the classification which lies at the basis of the counting, produce distortions. Every observation is indirect and during the observation, crucial assumptions are used which distort the data which is to be measured. Measuring itself is the result of a theory (although often at a low level).

Without measuring, every model remains incomprehensible and its usefulness stays undemonstrated. The semantic potency of a model (to use that term) is ineluctably linked to the result of measurements.

The measurement of mortality probabilities is subject to limitations, which cannot be easily overcome. On the one hand, there are problems which limit the theory formulation, and there are more practical problems about how numbers must be abstracted into probabilities. These

limitations are examined below.

3.1.1 Measurement limitations

There are a number of limitations relating to the collection of actuarial data. Experiments are impossible. No single repetition of a random sample makes use of the same elements or cases. The equivalence of cases is an assumption of the logical reasoning process and in physics most likely a plausible one. The assumption is less easy to maintain in relation to mortality, because in each random sample different individuals are observed. These individuals need not be comparable with the individuals from an earlier random sample. In other words, there could be physiological, biological, psychological or sociological changes, which mean that other outcomes are generated at the moment that a random sample is repeated.

A simple but unavoidable limitation is that performing a random sample takes time, which can scarcely be reduced. The number of dead is measured during a certain period. A short period results directly in a small data set, and that reduces the certainty about the data.

Another obvious limitation has to do with ethics. One cannot explore the impact of smoking by obliging half of a population to smoke and comparing their mortality behaviour with that of non-smokers. If a random sample with voluntary smokers and non-smokers is taken, then it is not known whether one directly observes the effects of smoking or the effects of some concealed cause which lies at the basis of smoking. Smoking might be one component of a complex of factors which all can have a separate effect on mortality. Influencing is virtually impossible because this causes changes in comparison with the natural situation, which in the final analysis is to be modelled.

The population whose mortality probabilities must be determined can be composed in a variety of ways. The national population and the population in a region or city are examples of populations whose mortality probabilities can be measured. Demographers and sociologists are primarily interested in such regional mortality patterns. Actuaries are interested in mortality probabilities of a group or subgroup of policyholders in one or several insurance companies, or in the mortality probabilities of people who belong to a certain risk group.

To measure the mortality probabilities of a population, the population will have to be observed during a period. A naive approximation of a ‘probability’ is the fraction of the number of deaths divided by the size of the population. Such fractions are called the ‘observed mortality probabilities’ or ‘mortality quotients’. The first step is to determine the number of deaths, and the second step is to determine the average population.

In the first step, the number of deaths of a particular age must be determined. Often the number of deaths in a population is determined by censuses. These censuses are performed at fixed times, for example at the end of the year. If 1100 dead of 25 years of age are registered at the end of a year, what does this value signify? Not all 1100 dead will have been exactly 25 years old, and yet it is assumed that all individuals in an age group are equal. We will call this the calendar age problem. It is impossible to derive the age precisely from the year of death and the birth year alone.

In the second step, the average population is determined. The mortality quotient depends on how the average population is defined. If a group is very large, the number of individuals who belong to a population can only be determined via estimations. The estimations take place at fixed times. When people die or emigrate, the population size declines. With births and immigration, the size increases. These fluctuations in the size of a population can only be registered indirectly and after the fact.

If, from the group to be measured, data of a number of subgroups is available, they can be used for determining the mortality probability. However, a problem arises if the age distributions of the subgroups differ from one another.

The measurement problems cannot be resolved by making the measurements more precise. With more data, the complexity of the calculation in order to use this data properly also increases.

3.1.2 Outline of this chapter

The measurement of a mortality probability does not simply come down to counting the number of dead. Measuring requires necessary assumptions relating to the objects to be measured. In essence, mortality probabilities and mortality quotients are seemingly legitimised concepts or artefacts. The mortality quotient, which earlier was called the rough mortality probability, is by no means 'rough'. With the methods for determining mortality quotients, a sort of primitive probability theory was already applied. It is necessary to use sub-models, which consist of micro-laws that describe relations and operations between counted numbers and other quantities such as mortality quotients and mortality probabilities.

First, we will examine the mortality probability and the concept of probability (3.2). The next section gives a theoretical explanation of mortality quotients (3.3). Then we describe interpolation methods for determining mortality probabilities (3.4). Solutions are described which were proposed in the past in order to measure mortality probabilities. These methods

are sub-models in their own right. Many methods make use of interpolation techniques based on certain assumptions about values between two points. In section 3.4.3, interpolation methods are improved by application of a statistical law.

Then, we will examine the identification of competing risks models (3.5) and the measurement of population heterogeneity (3.6) and we will show how these methods can be improved using the mortality model earlier defined.

After the theory of measurement, we are in a position to confront the data (3.8) by presenting two ways of model discovery. With the aid of a test to be defined, one can look at whether a model is appropriate to use for mortality data.

3.2 Mortality probabilities

Mortality probabilities are fundamental demographic quantities. Before looking at how mortality probabilities will be determined, we will explore exactly what a mortality probability means. The measurement of probabilities requires a statement of what the concept of probability entails.

Traditionally, the gap between measurement data and mortality probabilities is bridged by graduation of measurement data. We will first look at this graduation procedure.

The assumptions, which lie at the basis of graduation, cannot always be maintained. Before going into how mortality probabilities are determined, we will first look at two perspectives on the concept of probability. The section concludes with a discussion of determinism and probability.

3.2.1 Graduation of measurement data

Graduation of measurement data is an important part of the measuring process. Graduation is seen as an attempt to represent a phenomenon by systematic revision of a number of observations of that phenomenon [Hic87]. A much-used graduation method is the approximation of measurement data with the aid of a (continuous) mathematical function, for example Makeham's law of mortality.

Graduation of mortality probabilities is permissible because it is known from experience that as a population gets larger, the corresponding mortality probabilities become smoother. Then, it is assumed that, for an infinite number of individuals, the observed mortality probabilities are just as smooth as continuous functions. Graduation thus produces 'more accurate' mortality

probabilities. Because one measures from a part of a much larger imaginary population, one may expect that there are deviations from the total group. These fluctuations can arise due to seasonal influences, disasters or the popularity of life-endangering activities such as bungee jumping. Because it is practically impossible to observe a large population, the errors are eliminated by graduation of the data.

It may happen that not all irregularities are arbitrary. Sometimes these intrinsic irregularities can be determined and can be explained from the observed population. Upon repetition of the observation in another period, these intrinsic irregularities will reoccur. An example of this is the rapid increase of the mortality probability among men around age twenty.

The assumption that mortality probabilities from the past, purified of irregularities, correspond with those of the future is only realistic if no changes occur within a population. Graduation is only a criterion when it is assumed that a population does not change and there are no determinable intrinsic irregularities to be found.

The assumptions, which lie at the basis of graduation, cannot always be demonstrated. From observations it appears that the mortality probability has been dropping for a considerable time now, and the expectation is that this decline will continue for a while yet. This could be happening because of the degree to which medical science influences cure and disease prevention, or because of permanent and global social-economic displacements. Secondly, it appears that there are determinable irregularities at a global level, which arise in every period.

Graduation of measurement data with a single simple formula leads to an excessively crude approximation. Inevitably, this raises the question of what is the significance of the concept of probability.

3.2.2 The concept of probability

Carnap [Car50] distinguishes between two concepts of probability. These show up in actuarial science as the stochastic concept and the deterministic concept. Although the stochastic concept is increasingly being taught, the deterministic concept is still widely used in practice. The traditional practical methods find their origin in the life table of Graunt [Gra1662]. The life table refers directly to observable reality. Deterministic models imply that the data has a perfect, precisely determined structure. The stochastic concept, introduced from other disciplines, opposes this and interprets the life table as a probability distribution. The stochastic concept leaves room for the concept of probability within the framework of the model. In the following, these concepts are compared. This section is based on [Bun67a], [Bun67b] and [Keu94].

The stochastic concept, referred to as $\mathcal{P}_1$, reflects the degree of correspondence of a hypothesis h with respect to a proof b, or an observation. This is a logical semantic concept. Use of this concept does not lead to conclusions which are based on observable facts but to logical analysis. An example of this is the probabilistic statement that the probability that an individual will die in the coming year is equal to p. Although formally the statement applies for an arbitrary individual in a group, the statement only has meaning if the probability holds true for all individuals in the group.

The second concept, referred to as $\mathcal{P}_2$, is based on the (long-term) relative frequency of a property of events or things with respect to one another. Use of this concept leads to factual and empirical conclusions. An example of this is the frequentistic statement that the relative frequency or percentage of the number of deceased individuals from the total population of the past year is equal to f.

To call the probabilistic statement a translation of the frequentistic statement is misleading, because the statements are not equivalent on three levels.

- A probabilistic statement refers both to a described class and to an undefined individual. A frequentistic statement refers solely to a group. From the concept $\mathcal{P}_2$, determining the mortality probability of a single individual is senseless.

- The fact that the numeric variable p can be equated to f still does not mean that p also signifies the same as f. The variable p is a theoretical concept and f is an empirical concept. The value of p is fixed, while various empirical outcomes can be found for f, as estimations of p. A frequentistic statement does not need to contain any theoretical statements, while a probabilistic statement must be at least semi-theoretical, even when it relates to empirically-found values.

- The value of p may well be equated to f, but this still does not mean that the values are also equal. For statistical generalisation, it was a collective property, while for the probabilistic concept it is a potential property of each individual in the group. The introduction of an empirically found value in a probabilistic statement must be justified by a methodological rule, which says that the numeric value of the probability is approximated by the corresponding (long-term) frequency.

Often only one of the two concepts is used as explicandum, without knowledge of the importance or the existence of the other concept. This has led to an important difference of opinion.

Von Mises a representative of the second concept, gives the following example in [Mis39]

> *'According to certain insurance tables, the probability that a men forty years old will die within the next year is* 0.011*. Where are the 'equally likely cases' in this example? Which are the favourable ones? Are there* 1000 *different possibilities, eleven of which are 'favourable' to the occurrence of death, or are there* 3000 *possibilities and thirty-three 'favourable' ones? It would be useless to search the textbooks for an answer for no discussion on how to define equally likely cases in questions of this kind is given.'*

Von Mises did not see how a mortality probability could say anything about a single individual. A probability is estimated using observations from the real world. Therefore, the probability can only be applied to a population of individuals. According to Von Mises, the first concept was highly subjective.

Jeffreys, a defender of the first concept, in turn did not understand the standpoint of Von Mises [Keu94]. He did not see how an objective definition of probability could be given in terms of actual or possible observations. Von Mises should accept the existence of an a priori limit. Both defend their own perspective on the concept of probability and vigorously attack that of the other.

Carnap sees both concepts as attempts to explain the concept of probability. However, the question is whether both concepts refer to the same thing. Carnap asserted that the two meant different things by 'probability'.

If $\mathcal{P}_1$ and $\mathcal{P}_2$ are seen as quantitative concepts, then p and f are functions with two arguments of which the value lies between 0 and 1. $\mathcal{P}_1$ has the hypothesis and proof (although the latter is often omitted). The arguments of f are properties. $\mathcal{P}_1$ is relative to the proof. It is wrong to think that p was dependent on our knowledge and thus was not proven. Thus, p too has a relative character. Thereby the term 'event' is rather confusing because often one refers to a 'sort' of event.

Jeffreys sees probability as the degree of confirmation of a hypothesis and observation. Given an observation and given a number of hypotheses that are capable of explaining the observation, what is the relative expectation of the predictions of the hypotheses? This is the same thing as the determination of the quality of models. In the final analysis, Jeffreys is dealing with a property of a model when he talks about a probability, and he will defend the view that models must be as simple as possible. Popper [Pop59], by contrast, defends the view that simple models are less probable. Popper used an a priori property of observations. The criterion of Jeffreys is

not practical. Empirical data usually requires complex models. In practice, a simpler model is always used.

What a probability is exactly, according to $\mathcal{P}_1$ or $\mathcal{P}_2$, is in fact a pseudo-question, which still seems to predominate as a 'problem'. Whether a mortality quotient is equal to a mortality probability depends on the definition one gives to a 'probability'. One can adopt various positions with respect to this definition. In any case, one cannot use both concepts in the same model by measuring mortality probabilities in a frequentistic manner, and then using them in a stochastic model.

3.3 Metrics for mortality

Measuring mortality probabilities calls for a metrics. Next, we will look at the measurement of mortality quotients.

3.3.1 The mortality surface

To represent and analyse mortality data, a metric representation of time is convenient. In the actuarial world one uses the Lexis diagram [TPW94], named after the German economist Lexis (1837 -1914). This section describes the diagram and its properties.

Lexis' diagram is two-dimensional with the birth moment at the horizontal axis, while the vertical axis represents age. The surface that results makes the Lexis diagram a geometric representation of time. For an event (a death case), the following data is established.

- date of death
- date of birth
- age

In the ideal-typical case, time would be continuously observed. In that case, the moment of every event can be measured exactly, and a life line can be drawn for each individual.

The figure shows four life lines a, b, c and d. Consider line c and let P be an event of someone with age x at moment t. The individual to whom line c refers was born at moment $t - x$. The age, period, and cohort variables are linearly dependent by

$$t_0 = t - x$$

where x is the age, t the period and t_0 the cohort.

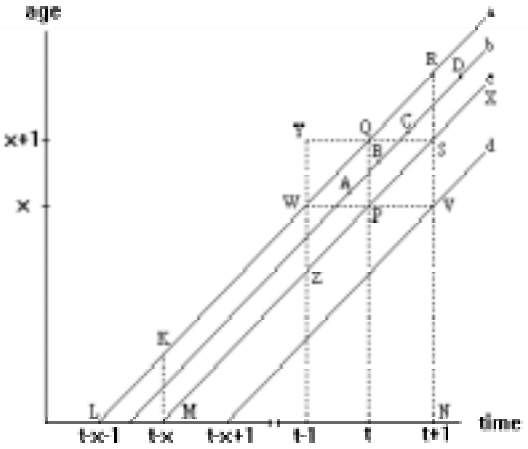


Figure 3-1: The Lexis diagram

Observation intervals are also represented in a Lexis diagram. Consider for example the age interval $[x, x+1)$, the observation time interval $[t, t+1)$ and the cohort interval $[t-x-1, t-x)$. The cohort refers to the persons who were born between year $t-x-1$ and $t-x$. Events are usually represented with three variables.

In the Lexis diagram, death cases are represented by a point. In order to measure the number of death cases, four different methods can be used.

- cohort-age. The date of birth and age are used. Thus, the parallelogram $WQSP$ corresponds with the persons who were born between $t-x-1$ and $t-x$ and are between the x and $x+1$ year. The measurement then extends over two years.

- period-cohort. The calendar year and date of birth are used. Between period t and $t+1$ one looks at how many persons, born between $t-x-1$ and $t-x$, have died. This corresponds with the parallelogram $PQRS$.

- period-age. The calendar year and the age are used. In the period $[t, t+1)$ one looks at

how many persons with age between x and $x+1$ have died. This corresponds with area $PQSV$.

- age-period-cohort. Here three pieces of data are used. In period $[t, t+1)$ one looks at the persons of age between x and $x+1$ who were born between $t-x-1$ and $t-x$. This corresponds to the triangle PQS. This data is double grouped.

In general, period-age observations are used. This has a drawback. As can be seen from the diagram, two dates of birth, namely the period $t-x-1$ and $t-x+1$, are considered in this method. Thus, it is impossible to derive the date of birth.

3.3.2 Mortality quotients

The measurement of a data set is possible if each element of the set can be measured, but also cross-sections, combinations and differences of elements from the set can be measured. A 'class' is defined as a set of several sets. The powerset $\mathcal{P}(X)$ is an example of such a class which consists of all possible sets of elements from X. Such a data set with operations is called a σ-algebra. Given an arbitrary set X and a non-empty class $S \subset \mathcal{P}(X)$, this class is a σ-algebra if

- $A \in S \Rightarrow A' \in S$
- $A_i \in S, \bigcup_{i\geq 1} A_i \in S$

This definition requires some clarification. The two conditions form a minimal set of requirements such that cross-sections, combinations and differences of arbitrary elements of S are again an element of S. Regardless of the operations which one applies to elements of S, one remains in the class S.

The measure

A measure adds a certain value to each data set E_i from a class S. Given a set (or space) X and a σ-algebra of sets $S \subset \mathcal{P}(X)$. A measure is defined as a nonnegative scalar function $\mu : S \to [0, \infty]$, with every element $E_i \in S$ satisfying

$$\mu\left(\bigcup_{i\geq 1} E_i\right) = \sum_{i\geq 1} \mu(E_i) \stackrel{\text{def}}{=} \lim_{N\to\infty} \sum_{1\leq i\leq N} \mu(E_i)$$

whereby the E_i's do not overlap with one another. Furthermore it holds that the measure of an empty set is zero, $\mu(\emptyset) = 0$.

One example of a measure is the probability. A measure $p : S \to [0, \infty]$ is a probability if for every separate event it holds that $0 \leq p(A) \leq 1$ and $p(S) = 1$. This measure links a probability with each element from the σ-algebra.

The mortality quotient is determined in three steps. First of all the measurement space is defined. Then, the mortality quotient of a single group is defined, and then this concept is generalised to different groups.

The measurement space

A measurement space (X, S, d) consists of a set (or space) X, a σ-algebra $S \subset \mathcal{P}(X)$ of measurable sets and a measure $d : S \to [0, \infty]$. A measurement space can be defined for measuring the number of dead. The set or space X is the set of all possible points in time, or the time domain. The observation periods form the set K

$$K = \bigcup_{0 \leq i \leq n-1} K_i$$

where $K_i = [t_i, t_{i+1})$.

The mortality quotient of a population

A minimum σ-algebra $S(K)$ always exists for every set K. Further, we possess a mortality measure d, which adds to each interval from $S(K)$ the number of observed death cases. The result is the measurement space Q consisting of $\left(X_{[t_0,t_n)}, S(K), d\right)$. The measurement space Q can be used for further analysis and ultimate determination of the mortality quotient.

For this, the measurement space of the number of dead is described. For the mortality quotient, the number of dead must be divided by the size of a population. The population size changes through birth, death, emigration and immigration.

The size of a population can be determined in a variety of ways. We proceed on the assumption that the size of a population is determined via censuses. Other methods exist that approximate the growth and decline by analytical functions with which at each arbitrary moment the size of a population can be determined. However, these are indirect methods. Determining the size of a population by a census is the most obvious method.

We assume a set of measurement moments $T = \{t_0, ..., t_n\}$. The size of a population which is

exposed to an observation period is a function $E : T \to N$. For the time being, we will proceed on the assumption that the mortality quotient of a single population must be measured, for example a population of 25 year old smokers. The question of how the size of the average population E^c (central exposed to risk) is determined is answered later.

The actuarial literature offers various definitions for determining the mortality quotient. The most precise would be to divide the age (measured until the moment of death) directly by the size of a population at that moment. With the definition of a measure in mind, one could imagine that in this way just as many small intervals would be generated as registered death cases. Each death case occurs at the beginning of each small interval, and each small interval always ends at the occurrence of the following death case. By dividing by the size of the population at that moment, the exact mortality quotient is obtained. Formalisation of this definition leads to the mortality quotient $m\left(t_i; t_{i+1}\right)$ during the period $[t_i, t_{i+1})$

$$m\left(t_i; t_{i+1}\right) = \sum_{0 \leq k \leq n-1} \frac{d\left(t_{i+\frac{k}{n}}; t_{i+\frac{k+1}{n}}\right)}{E^c\left(t_{i+\frac{k}{n}}\right)} \tag{3.1}$$

where $m\left(t_i; t_{i+1}\right)$ is also known as the central death rate or average mortality for a certain group.

However, the size of the population E^c is often not known at the moment of a death case. In practice, the size of a population is only known at the measurement moments. Provisionally, we will proceed on the assumption that we know the average population E^c which for a certain period represents the central or average size of a population by census.

In order to determine the quality of the calculated mortality quotients, the following consistency criterion can be used. Upon application of mortality quotients to a population, the same number of death cases would have to be calculated. Given a number n of observation moments on which the size of a population is known, then the following is true for the mortality observation measured in period $[t_i, t_{i+1})$

$$d\left(t_i; t_{i+1}\right) = \sum_{0 \leq k \leq n-1} E^c\left(t_{i+\frac{k}{n}}\right) m\left(t_{i+\frac{k}{n}}; t_{i+\frac{k+1}{n}}\right)$$

Notice that the function E^c does not represent the size of a population at moment $t_{i+\frac{k}{n}}$ but the average population in period $\left[t_{i+\frac{k}{n}}, t_{i+\frac{k+1}{n}}\right)$, as we saw in the definition (3.1) of the mortality quotient.

The result (3.1) is a mortality quotient m, which can serve as the basis for further measu-

rements. The advantage is that nothing is yet defined about the population that is measured. Another advantage is that the measurement points do not have to be at identical intervals. Approximation algorithms converge quickly if the measurement points are defined according to the zero points of Chebyshev polynomials.

The mortality quotient of groups

The next step is to distinguish groups. An arbitrary population can be divided into groups. Prevalent practice is to create mortality quotients by age and by sex. However, other distributions are also possible, such as smokers and non-smokers. We distinguish the set Θ of intervals $\theta_1, ..., \theta_n$. The intervals do not overlap. We assume that age is between 0 and a fixed N, then the age groups can be represented by $\theta_i = [i, i+1)$ with $i = 0, ..., N$.

The mortality quotient of group θ_1 is represented by $m_{\theta_1}(X)$. The standardised mortality quotient is

$$m^*_{\theta_1}(X) = \frac{m_{\theta_1}(X)}{m_\Theta(X)}$$

with $0 < m_{\theta_1}(X) \leq m_\Theta(X) \leq 1$.

For the construction of a normal life table, we distinguish one group for each entire age, and in total, we distinguish just as many groups as the maximum age. The mortality quotient is measured for each group.

Two problems arise in choosing the size of the intervals. As the groups become smaller by reducing the width of the interval, the number of dead in the group is also reduced. Secondly, the data at high ages is scarce, because few people die at a very high age. The result is that with smaller groups there will be larger differences between mortality probabilities in successive groups. With very many groups, the mortality probability of most groups will be zero. With the aid of credibility theory [Kli93], one attempts to supplement these virtually empty groups with information from other groups. Via weighted averages, a probability can nevertheless be determined.

Normally speaking, the period over which one measures is equal to $[x, x+1)$ so

$$m(x; x+1) = \sum_{0 \leq k \leq n-1} \frac{d\left(x + \frac{k}{n}; x + \frac{k+1}{n}\right)}{E^c\left(x + \frac{k}{n}\right)} \tag{3.2}$$

If $n = 1$, the commonly used central death rate at age x is obtained from (3.2)

$$m\left(x; x+1\right) = \frac{d\left(x; x+1\right)}{E^c\left(x\right)}$$

The central death rate or average mortality $m\left(x; x+1\right)$ cannot be used for an ordinary life table because it uses an average population. The quotient that is related to initial populations is called the rate of mortality, which can be seen as a frequency based probability, and is defined with the initial population E as follows

$$q\left(x; x+1\right) = \frac{d\left(x; x+1\right)}{E\left(x\right)}$$

Often, one year probabilities like $q\left(x; x+1\right)$ and $d\left(x; x+1\right)$ are abbreviated to $q\left(x\right)$ and $d\left(x\right)$, yielding $q\left(x\right) = d\left(x\right)/E\left(x\right)$.

The quotients measured above are rates or ratios and are as such related to frequency distributions as discussed earlier. In the following, we examine interpolation methods for determining the size of the initial population $E\left(x\right)$ and average population $E^c\left(x\right)$.

3.4 Interpolation methods

Because censuses are often performed at fixed times in the year, the situation is measured only at that moment. Not everyone is born at the same moment, and thus a census at a specific moment does not correspond with the average population of that year. Thus, the average size of a population must be estimated between the census moments and from that the number of dead per age must also be estimated. Interpolation methods are needed to determine the average population during an observation period.

First, we describe how one can deal with calendar year-structured data. Then we look at methods for determining the number of dead and the average population. We thereby distinguish between uncensored and censored interpolation methods. With the latter methods, one knows at what moment the individuals join a population, with the former one does not. The described methods can be found in [BP86], [DLS49], [Pre72] and [Nam87]. Lastly, a new method is presented for calculating the average population for both uncensored and censored data.

3.4.1 Calendar year-structured data

For many studies, the acquired data must be grouped before further analyses are possible. The purpose of grouping is to simplify the analysis in such a way that it is possible to derive mortality probabilities. Measurements generally take place at fixed moments in a year (generally at the end of the year). Here interpolation methods are described which interpolate calendar year-structured data. One of these interpolation methods is used by the Dutch Central Bureau of Statistics (CBS) to determine ages from calendar year-structured data.

The following rule is a simple method for determining the age x

$$\theta_x : x = \text{calendar year of death } - \text{ year of birth}$$

The age cannot be precisely determined from this, because those who fall in group Θ_{35} can be between 34 and 36 years old. If the grouping is done by age, the above rule creates an inaccuracy. Therefore, rules are developed which approximate the age as well as possible, i.e. which minimise the error.

An improved solution is based upon the assumption that someone is born and dies in the exact middle of the year. For instance, θ_{35} contains persons with age between $34\frac{1}{2}$ and $35\frac{1}{2}$ by the following improved rule

$$\theta_x : x = \overset{\text{july the first}}{\text{calendar year of death}} - \overset{\text{july the first}}{\text{year of birth}} \tag{3.3}$$

The method which the CBS uses to determine the age proceeds from (3.3). It is assumed that children are born evenly spread over a calendar year. At the end of a calendar year they are then one half year old on average. The life tables the CBS presents do not contain entire ages but half ages, thus explicitly the group with age $34\frac{1}{2}$ and $35\frac{1}{2}$. The transformation to entire ages is omitted [Cbs97a].

Along with groups of one year, smaller groups can of course also be used. Probably the day of death is the most exact information which can be had, and that thus defines the practical upper limit of the number of groups to be determined (leaving aside the question of how reasonable it is to distinguish between so many groups).

Another method is based on probabilities. Proceeding from a one-time observation at the end of the year of death, what is the probability that someone has the exact age? With observed age x the real age will lie between $(x-1, x+1)$. Assuming that the births are uniformly spread out during the year, the distribution of the age at the end of a calendar year is given by the

probability density with the following function

$$f(t) = \begin{cases} 1 - |t| & \text{for } -1 < t < 1 \\ 0 & \text{otherwise} \end{cases}$$

The average population with age x can then be derived from this. If $N(x)$ represents the number of individuals aged x at the end of a calendar year, then the average population E^c of age $x + t$, with $0 \leq t < 1$, can be found by linear interpolation of $N(x)$

$$E^c(x + t) = (1 - t) N(x - 1) + tN(x)$$

3.4.2 Number of dead and size of a population

The previous section described how censuses at the end of calendar years can be transformed to age-structured data. In this section, we will describe how mortality quotients can be calculated from this data.

If interim censuses have been held, or if data of individuals is separately obtained, data is available for calculating mortality probabilities with respect to the observation period. The mortality quotients will then be more accurate, but another method must be used in order to deal with this extra data.

The result of the analysis is a sub-division into groups. The mortality probability can be determined per group.

Uncensored data

In some cases it is not known on what date someone enters or leaves the observed group. It is only known what the number of deaths is and how many persons are part of the observed group at a particular moment. For determining the mortality quotient the actuarial literature offers various definitions. A number of definitions will be described here.

The average size $E^c(x)$ of a population of individuals aged x which is observed during observation period X is a function of $E : X \to N$. Our purpose is to determine the mortality quotient during age period $[x, x + 1)$. Based on the formula for the mortality quotient in (3.2), three methods of determining it will be explained.

A first method defines the mortality quotient by dividing the number of death cases by the

size of a population at the beginning of the observation period, or

$$m(x) = \frac{d(x)}{E(x)}$$

In this case, the average size is the size at the beginning of the observation period. The drawback of this is that a population can change during the period. The death cases, which are observed just before the end of the observation period, belong to a different population than the death cases at the beginning. In order to take this into account, the average of the size of a population at the beginning and at the end of the observation period can be used

$$m(x) = \frac{d(x)}{\frac{1}{2}(E(x) + E(x+1))}$$

In order to avoid inaccuracies, one must divide by the size a population which lies as close as possible to the population size in the observation period. The following method attempts to reduce this difference by counting the number of death cases during a short time $t \leq 1$ after a census and extrapolating to the year.

$$m(x) = \frac{1}{t}\frac{d(x; x+t)}{E(x)}$$

The drawback of this is that the mortality quotient becomes susceptible to possible events at the beginning of the observation period.

Censored data

In the ideal case, individuals can be observed separately during a period. The actual age can be derived at any moment from the date of birth and the time that someone forms part of the group can be observed. Not all individuals will belong to the group during the entire period and the observed period can differ from person to person. The impact, which someone has on the mortality quotients, then also depends on the observed duration. In calculating the mortality quotients, account can be taken of this individual observation period, whereby it can be precisely determined. If a national population is observed as a group, the number of individuals can vary during the observation period by emigration.

Traditional methods (such as described in [Nam87] and [BP86]) distinguish the following variables: $\widehat{l}$, n and w. With $\widehat{l}(x)$ the starters are indicated, i.e. the persons who participate from the beginning of the observation period and attain exact age x during the observation period.

Normally, the period is one year and the number at the end of the period is $\widehat{l}(x+1)$. Persons who within the observation period have been added to the group are indicated with $n(x+r)$. Interim terminators (through emigration or by redemption of an insurance) are indicated with $w(x+r)$. Upon leaving, the reason for departure can be registered so that the collected data can be used for multiple decrement survival functions.

Based on these data, the mortality quotient in the age interval $[x, x+t)$ can be obtained. Initially (at moment x) $\widehat{l}(x)$ persons belong to the group. During the period $[x, x+t)$ new persons are included in the observed group and persons leave the group. For the number of those entering with age x we note the entrance intensity $n(x)$ and for the terminators at age x we note the departure intensity $w(x)$. Thus, there applies

$$\widehat{l}(x) + \int_0^1 [n(x+t) - w(x+t)]\, dt - d(x) = \widehat{l}(x+1)$$

For simplicity's sake, we write $c(x) = w(x) - n(x)$ for the total decline of a population except mortality, so

$$\widehat{l}(x) - \int_0^1 c(x+t)\, dt - d(x) = \widehat{l}(x+1)$$

We define $C(x; x+s) = \int_0^s c(x+t)\, dt$. The exposed population at age $x+s$ is $\widehat{l}(x) - C(x; x+s)$ and the number of death equals

$$\int_0^1 \left(\widehat{l}(x) - C(x; x+s)\right) p(x; x+s)\, \mu(x+s)\, ds = d(x) \tag{3.4}$$

where $p(x; x+s)$ is the probability for an person aged x to be alive at age $x+s$ and $\mu(x+s)$ the mortality intensity at time $x+s$. Simplifying (3.4) with

$$q(x+s; x+1) = \int_s^1 p(x; x+t)\, \mu(x+t)\, dt$$

leads to

$$\widehat{l}(x)\, q(x) - \int_0^1 C(x; x+s)\, p(x; x+s)\, \mu(x+s)\, ds = d(x)$$

Then, by using partial integration, the equation

$$\int_0^1 C(x; x+s)\, p(x; x+s)\, \mu(x+s)\, ds = \int_0^1 c(x+s) \int_s^1 p(x; x+t)\, \mu(x+t)\, dt ds$$

the following equation is obtained

$$\widehat{l}(x)\, q(x) - \int_0^1 c(x+s)\, q(x+s; x+1)\, ds = d(x) \tag{3.5}$$

In order to be able to derive $q(x)$ from (3.5), the values of $q(x+s; x+1)$ must be approximated. The behaviour of mortality probabilities between two entire ages is called a mortality pattern. Prediction of this pattern for each interval is essential if the resulting approximation must correspond with the observed data. There exist various interpolation methods, which make assumptions about the pattern of mortality between two points in time. In general, it is assumed that the function must be simple. We will examine three of them and subsequently combine them.

Uniform distribution In the early eighteenth century, Abraham de Moivre proposed a first-order linear function as a representation of the life table. His hypothesis, that every year there is an identical number of deaths, was found unrealistic. Yet, his hypothesis is still widely used today in order to determine the number of dead in an interval. By assuming that the dead during a year x are uniformly distributed, the mortality probability $q(x; x+t)$ mounts during the year. This is represented by

$$q(x; x+t) = t \cdot q(x)$$

for $0 \leq t < 1$.

The Balducci hypothesis Another method for determining the pattern of mortality within an interval was proposed in 1920 by Gaetano Balducci. Although not easy to describe, it appears that this method, in contrast to uniform distribution, is in many cases a reasonable indicator for the life table.

Balducci assumed that the function $f(t) = q(x+t; x+1)$ is first-order linear over the interval $0 \leq t < 1$. From this it follows for the mortality probability

$$q(x+t; x+1) = (1-t) \cdot q(x)$$

for $0 \leq t < 1$. The result is a hyperbola between two points in the life table. A drawback is that the mortality-intensity drops locally between two points while it rises over the whole. However, this drawback does not offset the favourable impact, which the hypothesis has for determining the mortality probabilities from observed data, as shall become evident later.

Constant mortality-intensity A third method, which lies between the first two, assumes that the mortality-intensity is constant. Under the assumption that

$$\mu(x+t) = \mu(x)$$

for $0 \leq t < 1$. This results in

$$q(x; x+t) = 1 - e^{-\mu(x)t}$$

and

$$\widehat{l}(x+t) = \widehat{l}(x) \cdot e^{-\mu(x)t}$$

for $0 \leq t < 1$. The actuarial literature has a number of other definitions as well.

If we substitute the three different methods described above into (3.5), it appears that the simplest formula is generated by using the Balducci method. This produces

$$q(x) = \frac{d(x)}{E(x)} = \frac{d(x)}{\widehat{l}(x) - \int_0^1 (1-s)\, c(x+s)\, ds} \tag{3.6}$$

The complexity of computation does not increase if there are several moments at which persons exit or are admitted into the observed group.

An often used method, the so-called actuarial method, approximates (3.6) with

$$\int_0^1 (1-s)\, c(x+s)\, ds \approx \frac{1}{2} C(x) \tag{3.7}$$

so that

$$q(x) = \frac{d(x)}{\widehat{l}(x) - \frac{1}{2} C(x)} \tag{3.8}$$

The actuarial method is based on the assumption that the events are uniformly distributed over the course of the year. A method is described below which does not assume that. The average population is derived by averaging over all possible sequences of events.

3.4.3 The arcsin-law

In this section, a method is described for calculating the average population for censored data. The method is based on [Jab69] and [Shi84].

We assume an initial population $\widehat{l}(x)$ at the beginning of the observation. The population changes through immigration, birth, emigration and death. These events will not be concen-

trated but spread over the year. The total decrease at moment t (after starting age x) we represent with $c(x) = w(x) + d(x) - n(x)$. The sum S gives the total increase or decrease of a population

$$S(x+k) = \sum_{1 \leq t \leq k} c(x+t)$$

The size of a population at moment $x+k$ is then $\widehat{l}(x) - S(x+k)$.

It is often unknown at precisely what moment the population changes take place. The total numbers are known but not the moment in the year. The average depends on the sequence of the positive and negative events. If the positive population changes occur first and then the negative population changes, this yields a different average then if things occur the other way around. Other sequences yield other population averages.

We may permute the sequence in which the increase $c(x+t)$ takes place. By then looking at all possible permutations, it is possible to say something about the size of a population. We use the permutation $\pi : i \to i\pi$ for $i = 1, ..., n$. The sequence of the population changes is thereby $c_{1\pi}, c_{2\pi}, ..., c_{n\pi}$. Then there applies

$$S_\pi(x+k) = \sum_{1 \leq t \leq k} c_{i\pi}(x+t)$$

Now $S_\pi(x) = 0$ still applies. The set of all possible permutations is Π. We seek to derive the properties of $S_\pi(x+k)$. If it can be stated that

$$\begin{aligned} P(\pi) &= \text{number of positive } S_\pi(x+k) \\ M(\pi) &= \text{moment of the first maximum of } S_\pi(x+k) \end{aligned}$$

then it can be demonstrated that the number of times that t positive $S_\pi(x+k)$ appear in all possible permutations is equal to the number of times that at moment t the first maximum occurs, or

$$|\{M = t\}| = |\{P = t\}|$$

Analogously, the same can be demonstrated for the number of negatives and the moment of the first minimum. We define the number as follows

$$|\{M = t\}| = |\{P = t\}| = A_n(t)$$

Summing this for all possible values of t, then we obtain all possible permutations of Π,

thus

$$\sum_{0\leq t\leq n} A_n(t) = |\Pi| = n!$$

and $A_n(t) = A_n(n-t)$ applies. Also

$$A_n(t) = \binom{2t}{t}\binom{2n-2t}{n-t}\frac{n!}{2^{2n}}, \text{ for } t = 0, ..., n$$

We then define a probability $a_n(t) = A_n(t)/n!$. With the limit the continuous case is obtained

$$\lim_{n\to\infty} \sum_{a\leq \frac{t}{n}\leq b} a_n(t) = \int_a^b \rho(x)\, dx$$

where

$$\rho(x) = \begin{cases} \frac{1}{\pi}\frac{1}{\sqrt{x(1-x)}} & 0 < x < 1 \\ 0 & x = 0, x = 1 \end{cases}$$

It is known that

$$\int_0^x \rho(y)\, dy = \frac{2}{\pi}\arcsin\sqrt{x} \tag{3.9}$$

and

$$\int_a^b \rho(y)\, dy = \frac{2}{\pi}\left(\arcsin\sqrt{b} - \arcsin\sqrt{a}\right)$$

The course of $\rho(x)$ is represented in Figure 3-2.

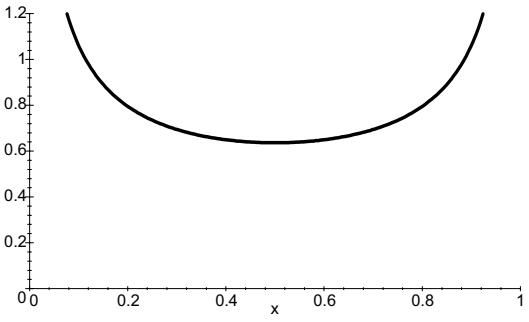


Figure 3-2: The $\rho(x)$ curve

Relatively few of the paths have an approximately equal number of positive and negative changes. The majority has predominantly positive or negative values. This has an impact on the average population. In Figure 3-3, the arcsin-law is compared with a uniform distribution.

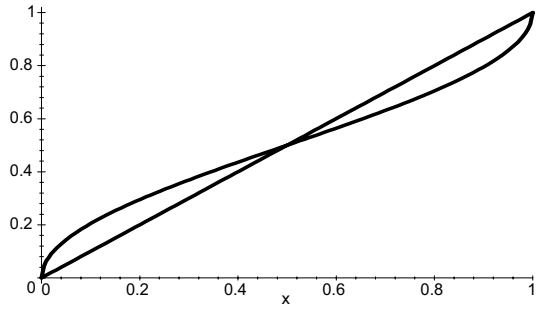


Figure 3-3: The arcsin-law compared with the uniform distribution

From the above method we obtain for the population size at moment $x+s$ the following

$$\widehat{l}\,(x+s) = \widehat{l}\,(x) - \frac{2}{\pi} C\,(x) \arcsin\sqrt{s} \tag{3.10}$$

This is a new determination for the calculation of the mortality probability. It follows from the fact that it is incorrect to expect the peak of the population size to occur in the middle of the observation period. The peak is more likely to be found either at the beginning or at the end of the interval. The average of (3.10) is equal to $\widehat{l}\,(x) - \frac{1}{2}C\,(x)$. Substitution of the Balducci hypothesis (3.10) in (3.4) yields

$$\widehat{l}\,(x)\, q\,(x) - \frac{2}{\pi} C\,(x) \int_0^1 \arcsin\left(\sqrt{s}\right) p\,(x; x+s)\, \mu\,(x+s)\, ds = d\,(x)$$

From this and result (3.9) it follows that

$$\widehat{l}\,(x)\, q\,(x) - \frac{1}{\pi} C\,(x) \int_0^1 \frac{q\,(x+s; x+1)}{\sqrt{s\,(1-s)}} ds = d\,(x)$$

Substitution of $q(x+s;x+1) = (1-s)\,q(x)$ yields the result (3.8) of the actuarial method

$$q(x) = \frac{d(x)}{\widehat{l}(x) - \frac{1}{2}C(x)}$$

The arcsin-law shows that, without making assumptions about $C(x)$ like (3.7), it is possible to derive result (3.8) by averaging over all possible events. Furthermore, the law makes it possible to use interim assumptions about the size of a population within a year.

3.5 Identifying competing risks models

Having defined how mortality probabilities of (sub-)populations can be measured, the next step is identifying competing risks models. The importance of grouping for calculating mortality quotients was already mentioned. A population of insured people is never uniform with regard to death or other risks. An age grouping is usually used, but there is also e.g. the distinction between smokers and non-smokers. Grouping makes sense. Grouping is an a priori classification with regard to dying. This is necessary in premium calculations (for which this thesis offers instruments). An a posteriori method, i.e. by using causes of death, distinguishes various unique risks or causes of death $R_1, ..., R_n$.

We will proceed with the assumption that random variable X is the potential survival time where all risks are acting and X_k is the potential survival time where R_k is the only acting risk, $k = 1, ..., n$. Then we have the marginal survival distribution

$$S_k(x) = \Pr[X_k \geq x]$$

Because the risks are unique, the joint survival distribution is defined by

$$S^{(n)}(x_1, x_2, ..., x_n) = \Pr\left[\bigcap_{k=1}^{n} (X_k \geq x_k)\right]$$

Then, the total survival function is

$$S(x) = S^{(n)}(\underbrace{x, ..., x}_{n \text{ times}}) = \Pr[X \geq x]$$

and the subsurvival function equals

$$P_k(x) = \Pr\left[(X_k \geq x) \bigcap_{i \neq k} (X_i \geq X_k)\right]$$

The potential survival times are contrasted with the actual survival time when the individual is exposed to all the competing risks. If we assume that one only observes the minimum of survival times, or

$$X = \min(X_1, X_2, ..., X_n) \tag{3.11}$$

Then, only X and K, where $X = X_K$, are observed. Random variable X is the survival time (or time of failure in reliability theory) and K the cause of failure. It is also assumed that no ties appear, that is the probability of $X_j = X_k$ equals zero for all $j \neq k$. With (3.11), the subsurvival function becomes identifiable because then

$$P_k(x) = \Pr[X \geq x, K = k]$$

which specifies the joint distribution of K and X. This is called the crude survival function and is not equal to the marginal survival function $S_k(x)$.

Obviously, the total survival function $S(x)$ is also directly identifiable. However, the joint survival distribution $S^{(n)}(x_1, x_2, ..., x_n)$ and the marginal distribution $S_k(x)$ cannot be measured from X and K alone. Additional assumptions are necessary, which is explained in the following section.

3.5.1 Independence of risks

If only X and K are observed then we do not know the joint survival distribution $S^n(x_1, .., x_n)$ and we do not know whether potential survival times are mutually independent. As shown above, the crude survival functions $P_k(x)$ can be estimated and we shall assume that they are known. The crucial question is how much information about the joint distribution of potential survival times does the set of crude functions provide.

Suppose that the set of $P_k(x)$, $k = 1, ..., n$, is given for a model with dependent risks. Then, as Tsiatis showed in [Tsi74], there exists a unique model with independent risks yielding

identical $P_k(x)$, which is defined by

$$P'_k(x) = -\mu_k(x) \prod_{k=1}^{n} S_k(x), \text{ where } S_k(x) = e^{-\int_0^x \mu_k(t)dt}$$

and the $\mu_k(x)$ are derived from the given $P_k(x)$. It is assumed that the functions are absolutely continuous. Therefore, to each dependent risks model there corresponds a unique model with independent risks.

This indicates that the agreement of the computed crude survival function $P_k(x)$ with their empirical counterparts cannot be considered as any kind of confirmation of the hypothesis of independence of the survival times. A unique independent risk model corresponds to every dependent risk model. One can never know, from observation of $P_k(x)$ only whether the risks are independent or not. Serious errors can be made in estimating the (potential) survival functions in the competing risks problem if the risks are assumed independent when in fact they are not. There is no way of knowing from the data in the competing risks problem whether or not an error is being made, since the data contain no information on whether or not the risks are independent or not.

Concluding, given any joint survival function with arbitrary dependence between the variates, there exists a different joint survival function in which the variates are independent and which reproduces the sub-densities precisely. There exists a system of net survival probabilities $S'_k(x)$, which combined with independent hypothesis implies the crude survival functions $P'_k(x)$ that coincide with the given $P_k(x)$.

Several studies are based on empirical counterparts of this function. Recent studies are based on assumptions specifying the functional form of the joint distribution function with a few variables or on the hypothesis that the potential survival times $X_1, ..., X_m$ are mutually independent. Without this hypothesis, the system is unidentifiable, that is the set of crude survival functions $P_k(x)$ is consistent with an infinity of joint distributions of potential survival times. Thus, a fully realistic treatment of the problem of competing risks depends on the validated detailed hypotheses about the joint distribution of the X's.

If the Makeham assumption (see section 2.3) does not hold then one cannot find an independent-risks model, which will reproduce both $P_k(x)$ and $S_k(x)$. However, this does not rule out the existence of a dependent-risks model which will do the job. Suppose one could observe R_k in the absence of the other risks as well as in their presence. Then, in principle, unlimited observation would identify the joint survival distribution $S^{(n)}(x_1, ..., x_n)$ as well as $P_k(x)$. It would however

be less likely to be feasible in biological applications.

In [Cro94], Crowder shows that the extra knowledge of the marginal survival functions in addition to the survival function is still not enough to identify the joint function. To each independent risks models there corresponds a whole class of dependent risks models which have the same subsurvival functions.

Suppose that failing complete identification for $S^{(n)}(x_1, ..., x_n)$, we ask about the parametric form of it. In some applications it has been argued that the failure time must have distributions of certain parametric form, for example Gompertz' model for human populations. Unfortunately, the theorem shows that even if the $P_k(x)$ are given and parametric forms of $S_k(x)$ are specified, $S^{(n)}(x_1, ..., x_n)$ is still not identified.

If independence is not assumed, the competing risks model is not identifiable. However, this does not trouble the practising statistician for too long because he seeks for an approximately fitting model, not necessarily for a 'true' model. In the case of independence of risks, the identifiability problem disappears but one cannot know solely from the observed data alone that independence is in fact the case. What is being estimated under the assumption of independent risks is the proxy model, which is at least well defined. Therefore, in the following we will assume independence.

3.5.2 Measuring the parts

Because it is assumed that only one cause of death can arise, the random variables $X_1, ..., X_n$ cannot be observed simultaneously. The random variable which is first realised is the observed cause. Thus, we observe the minimum of the random variables $X_1, ..., X_n$ as stated in (3.11).

The probability that someone at age x has not died can be written as

$$S(x) = \Pr[X \geq x] = \Pr\left[(X_1 \geq x) \bigcap ... \bigcap (X_n \geq x)\right]$$

We can determine the probability density per cause of death in the following way: if $\mu_k(x)$ is the mortality probability for cause of death k, then the probability distribution per age and per cause of death can be written as follows

$$Q_k^*(x) = \Pr\left[X < x \bigcap K = k\right] = \int_0^x S(t)\,\mu_k(t)\,dt$$

where K is an indicator variable such that it takes the value k if the cause of death is C_k and zero otherwise. The different probabilities π of the causes of death C_k can be determined by

summing up over all ages x

$$\pi_k = \lim_{x\to\infty} \int_0^x S(t)\,\mu_k(t)\,dt$$

A life and mortality probability per cause of death can be defined as follows

$$\begin{aligned} P_k^*(x) &= \Pr\left[X \geq x \bigcap K = k\right] \\ Q_k^*(x) &= \Pr\left[X < x \bigcap K = k\right] \end{aligned}$$

The total mortality, i.e. the probability that someone dies at age x regardless of the cause of death, is

$$Q(x) = \sum_k Q_k^*(x)$$

The total life probability, i.e. the probability that someone lives at age x regardless of the ultimate cause of death, is

$$P(x) = \sum_k P_k^*(x)$$

It is customary to approximate all defined random variables with discrete data. Assume that a $d(x)$ is the number of individuals aged x who die. If $S(x)$ is the fraction of survivors with age x then we estimate

$$\widehat{S}(x) = \prod_{0\leq t<x} \left(1 - \frac{d(t)}{\widehat{l}(t)}\right)$$

where $\widehat{l}(x)$ the number of observed individuals with age x. The cause-specific survival function can be determined as follows

$$\widehat{S}_k(x) = \prod_{0\leq t<x} \frac{d_k(t)}{\widehat{l}(t)} \widehat{S}(t)$$

with $d_k(x)$ the number of those who died due to cause of death k.

In order to determine the cause-specific survival function, the following formula can be used

$$\pi_k = \frac{S_k(0)}{\sum_k S_k(0)}$$

If there is also data available about the number of individuals who depart from the observed group, then the mortality probability can be determined by applying the arcsin-law, yielding (3.8) for cause R_k

$$q_k(x) = \frac{d_k(x)}{\widehat{l}(x) - \frac{1}{2}C(x)}$$

An alternative method is described in [MAL77]. The number of individuals that leaves the observed group can be eliminated by representing this group as a separate cause of death.

3.5.3 Applications of competing risks models

Independent right censoring

An application of competing risks models is censored data. In an censored data problem, one of the random variables R_k represents the time at which an individual becomes 'unobservable' for a reason other than death, while the remaining variables represent various causes of death. We identify one of the risks as the censoring variable. In the following, we will assume right censored data. The lifetime and risk or cause of death are only observed at the end of a life. These variables are called marks of the endpoint. These variables are correlated to lifetime and observed data and therefore are biased.

Two risks are assumed where the first is the risk of death and the other risk is the censoring variable. Let $(X_{i1}, X_{i2}), i = 1, ..., n$ be n independent identically distributed pairs of random variables. The minimum of X_{i1} and X_{i2} is observed and also the indicator of whether the minimum is X_{i1} or X_{i2} is observed for each pair (X_{i1}, X_{i2}), so

$$X_i = \min(X_{i1}, X_{i2})$$

and risk R_i with

$$R_i = \begin{cases} 1 & \text{if } X_i = X_{i1} \\ 2 & \text{if } X_i = X_{i2} \end{cases}$$

Although the censored data problem is symmetric in X_1 and X_2, it is common to refer to X_1 as the random variable of interest and to refer to X_2 as the censoring variable. Consistent with this convention we say that observation X is uncensored if $R_i = 1$ and censored if $R_i = 2$. The pair (X_i, R_i) is sometimes said to be a subsurvival pair.

Kaplan and Meier suggested an estimator for $S_1(x)$ that is a limit of life table estimators as the time interval increase in number and go to zero in length (see [Pet77]). Their estimator is

$$\widehat{S}_1(x) = \begin{cases} \prod_{i:X_{(i)} \leq x} \left(\frac{n-i}{n-i+1}\right)^{I_{(R_i=1)}} & \text{for } x \leq X_{(n)} \\ 0 & \text{if } R_{(n)} = 1, \text{ for } x > X_{(n)} \\ \text{undefined} & \text{if } R_{(n)} = 2, \text{ for } x > X_{(n)} \end{cases}$$

where the pairs $\left(X_{(i)}, R_{(i)}\right)$, $i = 1, ..., n$, are the n pairs of observations ordered on the $X_{(i)}$, i.e.

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$$

This estimator is a step function with jumps at those observations of X_i for which $R_i = 1$. If $R_i = 1$ for all i (if there is no censoring) the K-M estimator reduces to a step function with jumps of height $1/n$ at each of the X_i, which is the usual empirical distribution function. Although some of the X_i may be identical, the ordering is not unique. The K-M estimator however, is.

Peterson [Pet77] expresses the Kaplan-Meier estimator in terms of empirical subsurvival functions. The estimators of the subsurvival function $P_k(x)$, $k = 1, 2$, defined by

$$\widehat{P}_k(x) = \frac{1}{n} \sum_{1 \leq i \leq n} I_{[X_i > x, R_i = k]}$$

are maximum likelihood estimators.

Proof.

The likelihood is

$$\prod_{i=1}^{n} S_1^*(X_i)^{I_{[R_i=1]}} S_2^*(X_i)^{I_{[R_i=2]}} = (\pi_1^{n_1} \pi_2^{n_2}) \prod_{\substack{i=1 \\ R_i=1}}^{n} S_1^*(X_i) \prod_{\substack{i=1 \\ R_i=2}}^{n} S_2^*(X_i)$$

where n_k, $k = 1, 2$ is the number of observations i for which $R_i = k$ and the conditional survival functions $S_k^*(x)$ is defined by

$$S_k^*(x) = \frac{P_k(x)}{\pi_k}$$

The parts can be maximised separately. The first part is maximised when

$$\widehat{\pi}_k = \frac{n_k}{n}$$

The second part is maximised when

$$\widehat{S}_k^*(x) = \frac{1}{n_k} \sum_{1 \leq i \leq n} I_{[X_i \geq x, R_i = k]}$$

and thus the maximum likelihood estimator is

$$\widehat{P}_k(x) = \widehat{\pi}_k \widehat{S}_k^*(x) = \frac{1}{n} \sum_{1 \le i \le n} I_{[X_i \ge x, R_i = k]}$$

for $k = 1, 2$, which completes the proof.

■

The estimators are unbiased. With the subsurvival functions, the marginal survival function can be found. Because of the invariance principle of maximum likelihood estimators, this will provide us with the maximum likelihood estimators of the marginal survival function. It can be shown that estimation of the subsurvival function in the censored data problem consists of estimating the subsurvival functions in the obvious way and then applying to these estimators the function that relates subsurvival functions to the marginal survival function.

We assumed that no simultaneous causes could occur. They can however occur in typical applications in reliability theory where a piece of equipment is subject to random shocks, which knock out one or more components at a time. Applied to biomedical problems, the assumption of the simultaneous causes does not limit the applications much because the death certificate always represents one cause of death. It is possible to drop the assumption of simultaneous causes, the assumption of no simultaneous discontinuities and the assumption that the joint distribution is continuous, see Landberg, Proschan and Quinzi in [LPQ81].

Age-dependent right censoring

Cooke [Coo93] examines two classes for right censoring variables. The first class is random right censoring. For the second class, it is assumed that whether a variable is censored is independent of its age, but given that it is censored, the censoring time may depend on age. Individuals who feel healthy after a treatment might be more likely to change their insurance, thus causing a censored problem.

- Class 1: random right censoring X_1 independent of X_2
- Class 2: age-dependent censoring $X_2 = X_1 - \xi$

Under what circumstances can right censored life data distinguish between these two model classes?

Any continuous pair defines a unique independent model of class 1 (see Tsiatis in [Tsi74]). Therefore, a unique random censoring model will always describe the process. To get this type of result for class 2, we must restrict to continuous strictly monotonic subsurvival pairs.

For class 2, the probability that X_1 is being censored does not depend on the age. Therefore, a random variable δ may be assumed, for which $\delta \perp X$. The probability of being censored is π_2, so $\Pr[\delta = 1] = \pi_2$ and $\Pr[\delta = 0] = \pi_1$. Note that $\pi_2 = P_2(0)$, that is the initial probability to 'die' of risk R_2. If $\delta = 1$, then the age-dependent censoring variable ξ is chosen such that

$$X_2 = S_2^{*-1}(S_1^*(X_1))$$

and therefore

$$\Pr[X_2 \geq x] = S_2^*(x)$$

If $\delta = 0$, then X_1 is not censored.

Theorem 12 *Let $P_1(x)$ and $P_2(x)$ be a pair of continuous strictly monotonic subsurvival functions, then the following are equivalent*

(i) There exists random variables ξ and X_1 with $\operatorname{sgn}(\xi)$ independent from X_1 such that

$$P_1(x) = \Pr\left[(X_1 \geq x) \bigcap (\xi \leq 0)\right]$$
$$P_2(x) = \Pr\left[(X_1 - \xi \geq x) \bigcap (\xi \geq 0)\right]$$

(ii) For the conditional survival functions, the following yields for $x \geq 0$

$$S_1^*(x) \geq S_2^*(x) \tag{3.12}$$

Proof.

We have

$$\begin{aligned} P_1(x) &= \Pr\left[(X_1 \geq x) \bigcap (\xi < 0)\right] \\ &= \Pr[X_1 \geq x \mid \xi < 0] \Pr[\xi < 0] \\ &= \Pr[X_1 \geq x \mid \xi \geq 0] P_1(0) \end{aligned}$$

and

$$\begin{aligned} P_2(x) &= \Pr\left[(X_1 - \xi \geq x) \bigcap (\xi \geq 0)\right] \\ &= \Pr\left[X_1 - \xi \geq x \mid \xi \geq 0\right] P_2(0) \end{aligned}$$

Concluding, $P_1(x)/\pi_1 \geq P_2(x)/\pi_2$, and thus, for all $x \geq 0$, we have

$$S_1^*(x) \geq S_2^*(x)$$

The converse, that (ii) implies (i), can also be proven. We take a random variable $\delta \perp X$ with $\Pr[\delta = 1] = 1 - \Pr[\delta = 0] = \pi_2$, and put

$$\xi = \delta\left(X_1 - S_2^{*-1}\left(S_1^*(X_1)\right)\right) - (1 - \delta)$$

then, $\delta = 1$ is equal $\xi \geq 0$ and $\delta = 0$ is equal to $\xi = -1$, so $X \perp \operatorname{sgn}(\xi)$.

$$\begin{aligned} \Pr\left[X_1 - \xi \geq x \bigcap \xi \geq 0\right] &= \Pr\left[X_1 - \xi \geq x \mid \xi \geq 0\right] P_2(0) \\ &= \Pr\left[S_1^*(X_1) \geq S_2^*(x)\right] P_2(0) \\ &= \Pr\left[X_1 \geq S_1^{*-1}\left(S_2^*(x)\right)\right] P_2(0) \\ &= S_1^*\left(S_1^{*-1}\left(S_2^*(x)\right)\right) P_2(0) \\ &= P_2(x) \end{aligned}$$

and

$$\begin{aligned} \Pr\left[X_1 \geq x \bigcap \xi < 0\right] &= S_1^*(x) \Pr[\xi < 0] \\ &= S_1^*(x) P_1(0) \\ &= P_1(x) \end{aligned}$$

which proves the theorem.

■

The more observations we have of right censored process X the more closer the empirical subsurvival pair will be to $P_1(x)$ and $P_2(x)$. The subsurvival pair expresses the maximum of information of the empirical process.

An age-dependent censoring model will describe the process only if condition 3.12 is satisfied. Because the distribution of ξ is not given for $\xi < 0$, uniqueness is not possible. The theorem described above says that an age-dependent censoring model will describe the process if (3.12) holds or, in words, that a necessary and sufficient condition is that the conditional subsurvival function of the life variable dominates the conditional subsurvival function of the censoring variable.

The process of observation may be independent of the insured's age. Policyholders might be more likely to stop their insurance policy after being observed and thus causing a censored observation, whereas unobserved persons tend to hold their insurance policy. The probability of censoring after observation depends on the age.

Example 13 *We assume two Gompertz' distributions, one for a population $P_1(x)$ and one distribution for a population after being observed. Assumed is that they are consistent with the random signs censoring model. Suppose that the subsurvival functions are defined by*

$$P_1(x) = \frac{\alpha}{\alpha+\beta} e^{e^{\frac{-a}{b}}} e^{-e^{\frac{x-a}{b}}}$$
$$P_2(x) = \frac{\beta}{\alpha+\beta} e^{e^{\frac{-c}{b}}} e^{-e^{\frac{x-c}{b}}}$$

with $S(x) = P_1(x) + P_2(x)$. Then, $\pi_1 = \alpha/(\alpha+\beta)$ and the probability of being observed is $\pi_2 = \beta/(\alpha+\beta)$. Not every set of observations is consistent with the random signs model. Condition (3.12) must hold, or

$$a \geq c$$

Only if the life expectancy based on the first distribution is equal or higher than the life expectancy of the censored population then the subsurvival functions $P_1(x)$ and $P_2(x)$ are consistent with a random signs censoring model.

3.6 Measuring population heterogeneity

In the 'Models' chapter, a method is described where a distribution of vitality v among subpopulations was represented with a probability density $f_v(x)$. It was assumed that mortality was the only factor in changes of the heterogeneous population distribution. In order to develop a model that can estimate the degree of unobserved heterogeneity, two types of functions are necessary. Firstly, a function for the age-dependent mortality-intensity, and secondly a

probability density of the distribution of the specific mortality risk of individuals.

Several ways to measure population heterogeneity can be distinguished. First, a method to measure relative mortality experience based on Haberman's work [Hab88] will be described. After that, Cox's regression model [Cox72] will be explained and incorporated in the general mortality model from the 'Models' chapter. The last section is about the choice of vitality distribution when measuring population heterogeneity and is based on Manton's work in [MSV86] and [MSV81].

3.6.1 Relative mortality experience

Measuring the specific mortality risk of individuals consists of comparing the relative mortality experience with a standard mortality experience. This can be done by comparing the number of deaths and the number of expected deaths given a standard experience.

Haberman [Hab88] describes and discusses commonly used methods of comparing an observed mortality experience with that of a standard. The period over which one measures the number of observed deaths $d'(x)$ and the number of deaths of the standard population $d(x)$ is equal to $[x, x+1)$. The methods are all based on the Interval Mortality Ratio defined by

$$k(x) = \frac{d(x)}{d'(x)}$$

Because this equals $k(x) = q(x)/q'(x)$ and often the simplification $q(x) \approx \mu(x)$ is made, one can also write $\mu'(x) = k(x)^{-1}\mu(x)$, where $\mu'(x)$ is the 'observed' mortality intensity.

The progression of $k(x)$ depends on the underlying impairments of the observed population. With increasing x the progression of $k(x)$ can be downward (for example, breast cancer) or upward (for example, diabetes and overweight) or constant (for example, cerebrovascular disease). With the Interval Mortality Ratio several indices can be described.

De average of $k(x)$ over an n-year period is called the Average Interval Mortality Ratio and defined by

$$AK(n) = \frac{1}{n}\sum_{t=0}^{n-1} k(t)$$

The Annual Exposure Cumulative Mortality Ratio is the weighted average of $k(x)$ with the number of observed death and is defined by

$$AE(n) = \frac{\sum_{t=0}^{n-1} d'(t)\,k(t)}{\sum_{t=0}^{n-1} d'(t)}$$

Finally, the Ratio of Arithmetic Average Mortality Rates is the weighted average of $k(x)$ with the observed mortality probability

$$RAAD(n) = \frac{\sum_{t=0}^{n-1} q'(t) k(t)}{\sum_{t=0}^{n-1} q'(t)}$$

All indices try to summarise the set of $k(x)$ for the purpose of comparison where the last index is less sensitive for age intervals with low mortality probabilities. When the underlying pattern of $k(x)$ is complex and the observed mortality rates are changing rapidly then an average is not a good representation of the $k(x)$. Only when $k(x)$ is constant then these indices are useful.

3.6.2 Multiplicative hazards

The methods above can be generalised to deal with covariables like blood pressure and weight. In this section, we follow the terminology of Zadeh's embeddability (3.8.1). The multiplicative hazard was introduced by Cox in [Cox72] and here defined by

$$H(x; v) = \varphi_{\boldsymbol{\beta}}(\mathbf{v}) H_0(x)$$

with $H_0(x)$ the base-line hazard, $\varphi_{\boldsymbol{\beta}}(\mathbf{v})$ a vitality function from the class of equations φ defined by

$$\varphi = \{v = \varphi_{\boldsymbol{\beta}}(\mathbf{v}) \mid \mathbf{v} \in \mathbb{R}^n, v \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^r\}$$

where $\mathbf{v}$ is a vector of n covariates and $\boldsymbol{\beta}$ a vector of r regression parameters.

The most popular functional form of $\varphi_{\boldsymbol{\beta}}(\mathbf{v})$ is the log linear form

$$\varphi_{\boldsymbol{\beta}}(\mathbf{v}) = e^{-\boldsymbol{\beta}^T \mathbf{v}} \tag{3.13}$$

The reason why this form is used is its easily understood interpretation. There is also empirical evidence to support the assumption of proportionality of hazards in distinct treatment groups. Therefore, (3.13) appears to be a stable functional form. In addition, the log linear form has some nice properties with respect to censoring and statistical inference [CO84].

If the base-line hazard is according to Gompertz' law of mortality then it equals

$$H_0(x) = e^{\frac{x-a}{b}}$$

This leads to the following multiplicative hazard

$$H(x;\mathbf{v}) = e^{\frac{x-a(\mathbf{v})}{b}}$$

with $a(\mathbf{v}) = a + b\boldsymbol{\beta}^T\mathbf{v}$.

The survival distribution based on this hazard then equals

$$S(x) = 1 - G\left(x \mid \mu + b\boldsymbol{\beta}^T\mathbf{v}, \sigma\right) \tag{3.14}$$

If $\varphi_{\boldsymbol{\beta}}(\mathbf{v})$ is of a log linear form, it increases the life expectancy of the base-line population. In addition, the log linear form leads to a constant variance of a population.

After identifying the explanatory variables, the regression parameters $\boldsymbol{\beta}$ can be estimated by using a log likelihood [CO84]. Furthermore, using the mortality model in the 'Models' chapter, a Generalised Linear Interactive Modelling (GLIM) approach, as described in [Ren88], simplifies current methods.

3.6.3 The choice of the vitality distribution

Manton et al. [MSV86] study the sensitivity of the choice of functions for the estimates of heterogeneity of a population. In the article, the total mortality is compared to the mortality due to lung cancer. The cause-specific mortality-intensity, for n causes of death, is defined as follows

$$\overline{\mu}(x) = f_1\mu_1(x) + ... + f_n\mu_n(x)$$

with $\mu_k(x)$ as the mortality-intensity for the k-th cause of death and f_k as the weighting factor for the k-th cause of death.

In their article, results are presented based on different assumptions of the distribution of the mortality risk (Gamma, Inverse Gaussian and Bernoulli) and the mortality-intensity (Makeham and Weibull). The estimates of the structure parameters appeared to be less sensitive for the choice of the distribution of the mortality risk than for the choice of the mortality-intensity.

In [MSV81], the degree of heterogeneity of different populations is compared. Using measurements, the degree of heterogeneity of populations is estimated. By coupling the degree of heterogeneity to the mortality risk within a population, the specific heterogeneity within a population can be filtered out. This allows one to compare mortality risks between populations more efficiently.

A high deviation of the mortality pattern indicates a high heterogeneity of a population.

The heterogeneity of a population is measured as follows. It is assumed that the distribution of the mortality risk is a Gamma distribution. The variance of the risk distribution is then $var(z) = \overline{z}^2/k$, whereby $\overline{z}$ is the average frailty, and k the shape parameter of the Gamma distribution. The value of k then indicates the degree of homogeneity; the lower the value of k, the higher the homogeneity of a population.

The article proceeds on the assumption that the average risk influences the mortality-intensity as follows

$$\overline{\mu}(x_0) = \overline{z} \cdot \mu(x_0)$$

If it is assumed that at birth the basic mortality-intensity is equal to the cohort-intensity, then

$$\overline{z} = 1,\ x_0 = 0$$

The initial distribution only depends on parameter k. This reduction to one parameter is achieved because a Gamma variable may be rescaled to produce another Gamma variable without changing the shape parameter.

If frailty is represented with a Gamma distribution, then mortality-intensity of the heterogeneous population is equal to

$$\overline{\mu}(x) = \mu(x) \cdot \overline{S}(x)^{1/k}$$

(as was explained in section 2.4.2).

In order to estimate the mortality-intensity, the following must be done. It is assumed that $\overline{S}(x)$ and $d(x)$ are measured. A Gompertz mortality-intensity is taken as an underlying mortality model

$$\mu(x,t) = \alpha_{t_0} e^{\beta x}$$

with $t_0 = t - x$ the year is indicated in which a cohort was born. Consequently, the variable α_{t_0} depends on the cohort.

Proceeding from the earlier-defined representation of Gompertz' law of mortality, we here use

$$\ln \mu(x,t) = \frac{x-a}{b} - c(t-x)$$

with $c(t-x)$ indicating the difference of a cohort which is born at time $t-x$ and the basic cohort.

This model is then extended with an age-dependent proportional factor

$$\ln \mu(x,t) = \frac{x-a}{b} - c(t-x) + b(x)(1885-(t-x)) + d(t-x)\,x$$

with $b(x)$ an age-dependent function (1885 is the year in which the first data was available), and $c(t-x)$ and $d(t-x)$ cohort-dependent functions. From the research of Manton [MSV86] it appears that the last addition yielded no significant improvement.

The models described in this section show how explanatory variables can used for describing heterogeneous populations and make mortality forecasts possible by assuming a time conditioned mortality intensity. By using the earlier-defined representation of Gompertz' law of mortality these models are better understood because it shows the relation between the life expectancy on one hand and multiplicative hazards and mortality intensities on the other hand.

3.7 Measuring Gompertz' populations

In this section, we explain how populations can be measured when a Gompertz' distribution is assumed.

Obviously, the life expectancy μ has an easily understood interpretation. Cohort- and generation-effects and the influence of diseases are generally expressed in terms of changes in the life expectancy of an individual. Furthermore, changing the life expectancy is equal to a generally accepted actuarial method of age-addition, i.e. increasing the age of an insured if the insured belongs to a high-risk population. Increasing the age of an insured decreases the life expectancy accordingly. The standard deviation σ influences the angularity of the survival distribution but not the life expectancy. The higher the standard deviation, the less angular the survival distribution is. In some countries, one can observe an increasing angularity over time, which can be modelled by changing parameter σ. In the following, a method of estimating parameters μ and σ is given. Regression of parameters g and c is less obvious.

3.7.1 Method of moments

The method of moments generates possibly good estimators and in case of the Gompertz distribution, it results in easy estimates of parameters μ and σ. Generally, the method of moments attempts to equate the first r population moments to the first r sample moments (each non-central) and take the resulting solution in $\theta_1, ..., \theta_r$ (if one exists) as an estimator of $\theta_1, ..., \theta_r$.

Let $X_1, ..., X_n$ be independent and identically distributed random variables each with distribution function $G(x \mid \mu, \sigma)$. If the indicated moments exist, then the method of moments estimators of $\theta_1, ..., \theta_r$ are the solutions of the equations

$$E\left[X_1^k\right] = m_k'$$

for $k = 1, ..., r$, where m_k' is the k-th noncentral sample moments defined by

$$m_k' = \frac{X_1^k + ... + X_n^k}{n}$$

The basic method of moments sometimes yields estimates that can be found more easily in comparison with the maximum likelihood estimation. For the Gompertz distribution, it yields a solution for the life expectancy and variance. Because we have

$$\frac{d^n}{dt^n}\psi_X(t) = e^{at}\sum_{0\leq k\leq n} a^{n-k}\binom{n}{k}\frac{d^k}{dt^k}\Gamma(bt+1)$$

For the first two moments $n = 1, 2$ we find

$$\mu = \frac{d}{dt}\psi_X(t)\Big|_{t=0}$$

and

$$\mu^2 + \sigma^2 = \frac{d^2}{dt^2}\psi_X(t)\Big|_{t=0}$$

Substitution leads to the following estimates for the life expectancy

$$\widehat{\mu} = m_1' = \overline{X}$$

where $\overline{X} = \frac{1}{n}\sum_{1\leq k\leq n} X_k$ and for the variance

$$\widehat{\sigma}^2 = m_2' - \widehat{\mu}^2 = \frac{1}{n}\sum_{1\leq k\leq n}\left(X_k - \overline{X}\right)^2$$

In this way, it is possible to obtain life tables for very small populations.

Example 14 *Consider the following observed survival times* $x_1 = 85$, $x_2 = 62$, $x_3 = 81$,

$x_4 = 78$, $x_5 = 48$, $x_6 = 96$. *Then we have*

$$\widehat{\mu} = 75 \text{ and } \widehat{\sigma} = 15.727$$

Applying (2.43), we obtain

$$a = 82.078 \text{ and } b = 12.262$$

The survival function of the Gompertz distribution $G(x \mid 75, 15.727)$ *and the density are given in Figure 3-4 and Figure 3-5 respectively.*

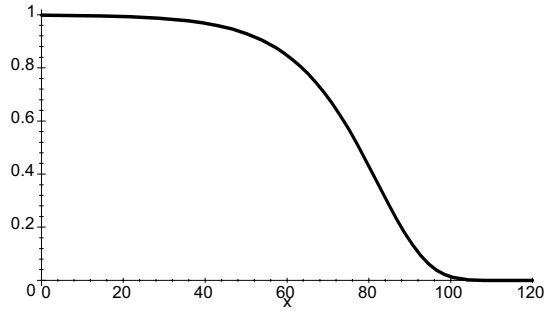


Figure 3-4: The survival function of the Gompertz distribution with $\mu = 75$ and $\sigma = 15.727$

It is difficult to fit the Gompertz curve in this data set by traditional fitting methods.

3.8 Actuarial model identification

Because mortality models are generally expressed with the aid of quantitative formulations, the foundation for every realistic prediction is formed by data. In this chapter, it was already mentioned that a kind of theory is necessary of every measurement. The result of measurement is, albeit only partially, a product of the observer himself. The previous sections described how the mortality probabilities could be determined.

A fundamental aspect of modelling is the confrontation of the model with the data. This confrontation, which is generally referred to as 'fitting' the model to the data, is only correct

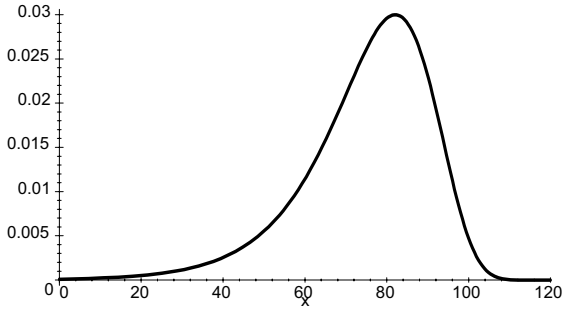


Figure 3-5: The density of the Gompertz distribution with $\mu = 15$ and $\sigma = 15.727$

if the data meets the requirements placed on (the data of) the model. The fitting of the requirements of the data on the model is also called 'embeddability'. Computationally, it is possible to fit any population data set to Gompertz' law of mortality but this does not mean the law meets the requirements of the data.

The two modelling theories described in the first chapter are applied for discovering models. The first applies Zadeh's embeddability by defining a computational criterion to determine the embeddability of models. After that, model discovery by Maciejowski's programmes will be explained.

3.8.1 Zadeh's embeddability applied

The problem of embeddability is often seen as balancing between complexity and accuracy. On the one hand, the model must do a good job of approximating the data, while on the other hand the model must, to ensure any predictive ability, abstract from data. For accuracy, a large number of tests exist, such as the Kolmogorov-Smirnov test and the Chi-square test [Kli93]. Combinations of tests are often used.

Statistical approximation of curve fitting is done in two steps. First of all the functional form M is determined in which yet undefined parameters are used. This step is described in the 'Models' chapter. The second step is making the best estimation of values of the parameters. This is, according to Zadeh [Zad56], the identification of the model from a model class M. For identification, a measure of accuracy must be defined. Here, the standard solution is the least

squares method.

The choice of the model class M is much more complicated because a complexity criterion has to be used. There is no generally accepted measure for the desired complexity. Well-known methods are the minimisation of the number of parameters in a class and / or information criteria. For these complexity criteria, however, there are no fundamental justifications.

The criterion for complexity and the criterion for accuracy are in conflict. Assume that we want to determine a linear regression equation for a certain observable outcome Y in terms of predictive and independent observable input variables $X_1, ..., X_n$. Furthermore assume that $Z_1, ..., Z_r$, which are functions with one or more X's, represent a set of state variables from which the equation structure, i.e. the model, can be chosen and that this set contains each arbitrary function. Then we have two opposed criteria for the choice of the resulting model.

In order to make the model fit accurately the variables $X_1, ..., X_n$, the model must contain as many Z's as needed, i.e. that reliable approximated values are be determined by maximising the use of state variables (the complexity of the model structure).

The trade off between accuracy and complexity thus in fact comes down to the balancing between accuracy and generality. As soon as the number of the Z's goes down, the balance will shift and accuracy will become more important than generality. Computers can deal with hundreds of parameters without difficulty. Thus, it appears that complexity is connected with generality. Researchers prefer beautiful models, i.e. with few parameters.

Stability as criterion

To test the embeddability of a model class in relation to the data collected for it or vice versa, Turney [Tur90] finds the curve with best balance between stability of consistent estimation and accuracy. He presents a measure for stability of equations, which will be discussed here. In this section, stability is regarded as a test for embeddability.

A class of equations F is assumed, to which the following applies

$$F = \{\mathbf{y} = f_{\mathbf{a}}(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m, \mathbf{a} \in \mathbb{R}^r\}$$

with variable vector $\mathbf{x}$ and $\mathbf{y}$ and parameter vector $\mathbf{a}$ where D consists of points in a space

$$D = \{\langle x, y\rangle \mid x \in R^n, y \in R^m\}$$

We seek a fitting procedure for a function f from a class F on data D. For $n = m = 1$ we

can use for this the sum of the squares

$$S = \sum_{1 \leq i \leq k} (y_i - f(x_i))^2$$

This gives the values

$$D_f = \{\langle x_1, f_c(x_1)\rangle, ..., \langle x_k, f_c(x_k)\rangle\}$$

Again, we seek the $f_c(x)$ for which $S(D, D_f)$ is minimal. Take each point $\langle x, y\rangle$ in D whereby to y is added a small perturbation, of which the distribution

$$D_p = \{\langle x_1, y_1'\rangle, ..., \langle x_k, y_k'\rangle\}$$

The data yields

$$D_{fp} = \{\langle x_1, f_d(x_1)\rangle, ..., \langle x_k, f_d(x_k)\rangle\}$$

We seek $f_d(x)$ so that $S(D_p, D_{fp})$ is minimal. The embeddability is a measure for the sensitivity of class F for perturbations

$$I(F, D) = E[S(D_f, D_{fp})]$$

The measure is semantic instead of syntactic and quantitative instead of qualitative. Embeddability is a property of models but lies above all in the pragmatic use of models.

For the choice of the model class, the degree of embeddability is important. If with a number of experiments under the same conditions the same model from the model class is always identified, then the embeddability is high and we prefer the model class over other, less stable, model classes.

A different is approach of modelling is taken by the Actuarial Association in the Netherlands. Their approach has much in common with Maciejowski's programmes and will be explained in the next section.

3.8.2 Maciejowski's modelling approach

In Maciejowski's model theory [Mac78], no model class is selected in advance; rather, modelling is providing an algorithm to construct a model of a system. This way of modelling is followed by the 'Mortality tables' committee of the Dutch association of actuaries (Actuarieel Genootschap) for the construction of life tables. These life tables are published every five years by the AG and

are used by insurance companies and pension funds. The life tables are based on the mortality quotients measured by the CBS. How the modelling process developed over time is described below.

Historical overview

Since the 1950's, insurance companies and pension funds have been using the life tables of the Actuarial Association. For this period, the national mortality probabilities were graduated by actuaries themselves, which led to a proliferation of different rounding methods. In 1950, Van Rooijen and Yntema were appointed by the AG to make a rounded life table which could be used by everyone. The method that was used consisted of the approximation of the observed mortality probabilities from the period 1951-1955 with Makeham's first law of mortality. For the life table of 1956-1960, the assignment was given to Van Rooijen. Both times the work was in fact performed by Van Oosterwijk Bruyn.

As of 1961, a 'Mortality tables' committee was established which produced a rounded table of average observations from the previous period every five years. Along with the mortality table other actuarial quantities were also calculated and published. Thus emerged the 1966-1970, 1971-1975 and 1976-1980 tables. As of 1980, the designation of the data provided by the CBS was changed. Analogous to the designation of the CBS, the name of the following table would be 1980/81-1984/85. This was abbreviated to 1980-1985 and later the 1985-1990 mortality table. As of 1990, one speaks of 'life tables'. The most recent life table is 1990-1995.

The following three starting points were and are still used. Firstly, a table must graduate the observed life data because it contains too many irregularities. Secondly, a life table must represent a monotonic rising mortality probability. The last starting point is that the life table must satisfy the Law of Uniform Seniority. This law makes it easier to do actuarial calculations for more than one life.

In 1966 higher requirements were imposed on the numerical accuracy of the life tables. The mortality probabilities at low ages were not accurate enough. In order to satisfy the changed conditions, the method of approximation was changed every five years. Improvements for approximation of low ages were based on life expectancy and, later, on approximations with the aid of polynomials.

The solutions for better approximation resulted in life tables, which were approximated with different techniques per interval. This led to a problem at the points where the technique was changed. The outcome at the boundary points differed as a result of the technique, and this

resulted in small jumps. Minimising these jumps is currently important in graduation of the life table.

The current AG model

The 1990-1995 life table contains several similarities and modifications compared to preceding publications. How the life tables (for men and women) were created is described below.

In the most recent CBS publications, there is mention of mortality quotients and life tables instead of 'observed mortality probabilities' and 'mortality tables'. The mortality table committee has also adopted this change of terminology (although this committee left its own name unchanged).

In 1995, the CBS provided the observed mortality quotients up to the age of 99 years only. A part of the population still has not died at that age. It was therefore decided to let the rounded life table run to a higher age by extrapolation. Later it appeared the CBS was indeed capable of providing this missing data, yet it is not included in the current table.

The starting points and their evolution

The model that is used in the AG's most recent life table, as well as in the preceding tables, is Makeham's first law of mortality. With this law, the life table can be modelled with only a limited number of parameters (low complexity of the model) and the table satisfies the Law of Uniform Seniority. For more than a century now, this law has been an important argument in favour of using Makeham's law of mortality. The law presently appears less important because, with the method used, the law has only limited validity. Perhaps one can say that the Law of Uniform Seniority is a nice extra property.

The technique used for estimating the parameters is essentially a Generalised (Non-)Linear Models as for example described by Renshaw in [Ren91]. These models are often used in actuarial graduation practice. Gompertz' law can be generalised to a GLM because, after a transformation, a linear form appears

$$\ln\left(-\ln\left(S\left(x\right)\right)\right) = \frac{x-a}{b}$$

Makeham's law is non-linear so an iterative procedure has to be used in order to find the most appropriate parameter values.

In order to estimate the parameters of Makeham's law of mortality, a criterion must be used to determine which approximations are better than others are. The criterion that is used is

the standard deviation: the best life table has the smallest standard deviation compared to the observed life data. This criterion was first used in the previous publication [Ag92]. The average error was used for this.

Two other criteria in the estimation of the parameters are small jumps at the interval boundaries and as large as possible a number of sign changes of the difference of the approximated curve and the observed life data. These are primarily of practical importance.

If Makeham's law of mortality is to be used as model for the entire interval and the parameters have the same value over the entire domain, then the life data is not approximated accurately enough. Therefore the domain is split in two and the parameters are defined in such a way that within each sub-interval the standard deviation is minimal.

Due to this technique, however, jumps arise at the interval boundaries. To do something about this, it is added as an extra constraint that the jumps at the interval boundaries must be minimal. The jumps are the result of the way the model is used.

Although the criteria are used as a test, some questions remain to be answered. It is still the question how the criteria must ultimately be used. Reducing the size of the jumps comes at the price of the accuracy of the approximation within the interval. Exactly what weighing was done between the two is not described.

Because different criteria are used, the possibility exists that criteria are in conflict with one another. A first conflict is for example that, however the parameters are defined, Makeham's law of mortality yields a rather high standard deviation in the first 28 years (for men). This interval is therefore not approximated with Makeham. The interval is split into two parts. Up to 12 years, one approximates with a constant mortality probability and after that, from 13 to 28 years, with a third-degree polynomial. The interval boundaries are determined using an 'iterative process'. The details of this process are not described. The introduction to the last published life tables [Ag97] says the following about this

> *'The age at which one shifts from the rounding off according to Makeham's law of mortality to rounding off with the aid of a polynomial in* $\ln p_x$ *is chosen in such a way that an optimal correspondence to the observations was obtained with as little discontinuity as possible in the course of the mortality quotients at the boundary age. From test calculations it appears that the transition for men was most favourable between 28 and 29 years, and for women between 23 and 24 years.'* [Ag97, page VI]

The deviations after around 70 years also appear to be systematic. Given the importance of the figures at high ages (for pension calculations), it has been attempted to approximate this

part with a higher accuracy (regarding to the standard deviation). This is done by using two Makeham formulas, one adapted to the figures from before and one adapted to the figures after 73 years, which are then stuck together, so to speak. The reason why the committee chose this specific year of life is not made clear.

This correction results in a kink (which conflicts with the criterion that the jumps must be small) and means that the Law of Uniform Seniority only applies for the interval (and consequently, appears of little importance). A simple improvement is to remove the kinks manually.

Gradually, the habit of using Makeham's law of mortality less strictly as a model was developed. Makeham's law of mortality was primarily used to describe the mortality probabilities of a population with only one analytical function. Later the committee was confronted with the problem that Makeham's law of mortality no longer satisfied contemporary requirements. The initial response to this was to make minor modifications in the method of approximation, but in reality, by allowing the parameters per interval to differ, the model was being used in a manner other than that which was originally intended. The committee appears to be retaining the old mortality law, but the actual modelling process is different. Presently the law returns as a vague starting point which says that an interval as large as possible must be approximated with Makeham's law of mortality. The mortality law is no law any more, but just a convenient way of smoothing the observed mortality probabilities. Thus, over time the graduation has come to look like the modelling itself.

The reparametrisation of Gompertz' law of mortality as presented in the previous chapter leads to another consideration. When changing the parameters at certain points, the assumptions about life expectancy and variance also change. So reaching a certain age results in a sudden change of life expectancy. Besides this interpretation problem, only quantitative comparisons between mortality probabilities of different populations are possible when dealing with different parameters for each interval.

Although it is theoretically possible, by repeating the construction process, to recalculate the life table itself, this still does not demonstrate that the life table is in fact the best possible. The construction process of the committee is secret. Thus, one cannot determine whether the life tables of the Actuarial Association are the most feasible, but by repeating the modelling process every five years the model remains realistic and accepted.

Zadeh's embeddability and Maciejowski's programmes can be seen as examples of computational intelligence. In contrast of the first method, there is no obvious way to automate the last method because the human experts are an essential part of the modelling process. The

skilful hand of the actuary is indispensable for the modelling process. It would be tempting explain to this expert knowledge as subjective and therefore to make any automatic model discovery impossible. However, often it is possible to formalise this kind of knowledge in order to automate it.

Chapter 4

Applications

It is important for insurers and pension funds to gain insight into the future development of mortality. However, virtually no good methods exist that take into account the expected increase of the life expectancy. The Central Bureau of Statistics in the Netherlands has made a forecast for the future development of mortality probabilities (4.1.4). In the first section (4.1), methods are developed in order to forecast the increase in life expectancy.

Application in actuarial science includes (among other things) calculating the level of premium and reserves. In (4.2), a method for calculating different types of insurances on the basis of the mortality model developed in (2.5) is explored by applying computational intelligence. First, the knowledge domain, life insurance mathematics, is described (4.2.1). After that, an expert system that uses this knowledge is presented. Furthermore, stop-loss transformations of the developed mortality model will be described in (4.2.4).

4.1 Mortality forecasts

An important application of the described models focuses on forecasting mortality in the future. Two important purposes can be distinguished in doing this. Firstly, one uses forecasting models in order to identify trends. A second purpose, deriving from the first, is the construction of forecasting models so that they can be used for actuarial applications. An important problem here is determining the magnitude of the so-called 'longevity' risk.

In order to forecast mortality, the time factor will play a role.

4.1.1 Time and age

One result of the traditional use of the life tables is that no distinction is made between time and age. Many of the models discussed in chapter 2 were developed on the assumption that age and time are interchangeable. The first mortality models assume an initial population with size $S(0)$. As time goes on, individuals will die. At time (or age) x the size of the population will be $S(x)$.

Many actuarial calculations are based on this assumption. This is related to two types of life tables: the cohort table and the period table. The first is constructed from a population (beginning with $S(0)$) which is observed for a number of years until all of the individuals from this population have died. Due to the long observation that is required, a pure cohort table is not easy to make. Because the environment changes constantly, the mortality probabilities of the cohort will be influenced. The advantage for actuarial applications is negligible.

Period tables are constructed by observing a population during a brief period. A complete life table is then derived from the measured data. Due to the brief observation period, time dependent effects are negligible. Period life tables are widely used in actuarial calculations. In fact, the Dutch Actuarial Association makes a period life table every five years.

The difference between the two types can be explained as follows. Assume that $q(x,t)$ is the mortality probability of someone with age x at moment t. With a cohort table, the number of the living with age x is based on the probabilities $q(x-1,t-1)$, $q(x-2,t-2)$, ..., $q(0,0)$. Cohort tables primarily measure the effects of time on a population. With a period table, the size of the population depends on the probabilities $q(x-1,t)$, $q(x-2,t)$, ..., $q(0,t)$. Therefore, the period table primarily measures the effects of age on mortality. If the mortality probabilities of a population vary over time, the period and the cohort table will differ.

The origin of the assumption is not surprising. The course of time is the same as the course of age. The assumption however entails that it is difficult to model the influences of an increasing life expectancy over time with the current models. Population heterogeneity can also arise as a result of different birth years.

It is expected that life expectancy will rise in the future. Considering this, the use of period tables measured in the past leads to wrong forecasts. It is not known just how far the period tables are off. In the Netherlands, this problem is referred to by the term 'longevity risk'.

4.1.2 Life expectancies

Research on mortality trends is not only interesting from a demographic point of view; in the insurance world, people are interested in future developments of mortality. However not much is known about this subject yet. There still is no information which is purely descriptive in nature, subdividing the mortality trends into a number of characteristics. There is still no sound theoretical basis, and yet there are only a number of ad hoc explanations of trends.

Various techniques exist for discovering trends and derivatives of trends; from simple graphical methods to increasingly complex statistical and mathematical models. Examples of this are: contour maps, descriptive trend analysis, cohort and period life tables, mortality models which describe the age profile of mortality, age-period-cohort model and projection models, see for example Tabeau, Poppel and Willekens in [TPW94].

Two problems make the analysis of trends difficult. Firstly, there is too little information available for the analysis; and secondly, there are some methodological problems. Methodological problems arise because the temporal factors ('white noise ') must be distinguished from the structural factors. In fact, there are three categories of factors, which influence the mortality probability:

- Age effects
- Periodic effects
- Cohort- or generation-effects

The first effect is self-explanatory. The mortality probability depends heavily on age. By 'periodic effects' effects on the mortality probability which arise at a determined time or within a period are meant. Cohort- or generation-effects have an impact on the individuals with the same birth year or those who belong to the same generation.

Decomposition between white noise and structural factors requires the specification of an APC-model and the quantification of effects of age, period and cohort. Information about the temporal factors can be derived from the period to which the mortality refers. Historical factors can be approximated from the cohort or birth year. Unfortunately, it is not possible to separate periodic effects from cohort effects and measure them independently. When estimating the parameters for periodic and cohort effects an identifiability problem arises.

Due to a lack of data, it is often impossible to derive effects of age, period and cohort from a time series of age-structured data. In most applications, one uses mortality data per age and

calendar year. It is impossible to derive a cohort from this. In order to make an unambiguous separation, the age, the year and the cohort to whom the person belongs must be determined.

For making forecasts, the cohort-effects and period-effects are often extrapolated into the future. It often appears to be a problem to make a clear distinction between the two effects.

In recent years, the fact that people are increasingly living to an older age is above all due to the decline of acute mortality due to infectious and cardiovascular diseases. The health care system is providing no further decline in mortality. How long life expectancy will continue to rise in the future is a matter of speculation. It is most likely that one cannot continue to extrapolate the observed increase in life expectancy. As there were in the past, there are scientists today who believe that the human life span has a certain maximum biological limit, so that the increase in life expectancy will ultimately grind to a halt regardless of further medical developments. Although from time to time there are people who live even longer, it is generally believed that the limit lies around 118-120 years.

Although the general trend is a rising one, temporal changes are observable in the life expectancy of certain groups. There have been periods in which the age- or sex-dependent probability of a cause of death fell (or, on the contrary, rose) [MSC97]. Since the 1950's, the mortality patterns of men and women differ significantly. Until 1985, the life expectancy for women was growing, but since that year, the increase of the life expectancy is diminishing. During the same period, the life expectancy for men also rose, but to a lesser degree than for women. Because of this, the difference in life expectancy increased between men and women until 1980. Since that year, the difference has fallen to 5.5 years.

Recent developments in the Netherlands reveal a decline in the life expectancy for women between 25 and 40 years. Mortality probabilities for men reveal an increase within this interval. As of 80 years, the mortality probabilities are decreasing slightly. The CBS [Cbs97a] thinks that a selection took place during the preceding years, for example as a result of influenza, which resulted in premature death for frail elderly persons. Because of this, the stronger are still alive, whereby a decline in the mortality probability was observed. The CBS suspects a negative selection as opposed to a positive period-effect. Provisionally the increase of the life expectancy will continue and the CBS even expects that the maximum life expectancy will still not have been reached by the year 2010.

Actuarial models based on the assumption of an increasing life expectancy do exist, although they are often based on simplifying assumptions. The Collective Reference Tariff Committee (CRC) introduced a population mortality model that forecasts mortality probabilities. This model resulted in an age-structured life table, which would represent the conditions in the near

future (about twenty years from now). For each age, the observed probabilities are smoothed out with aid of a graduation algorithm and extrapolated to the near future. By extrapolating mortality probabilities instead of the life expectancy, it is possible to forecast mortality developments for specific age groups.

The Coll '93 table depends only on age and not time, and thus forms a snapshot. The most recent table is made for the period 2006-2010. Application of the Coll '93 table for pension contracts will lead to mortality profits being made before 2010 (because current mortality probabilities are higher when compared to the table) and mortality losses after 2010.

4.1.3 Heterogeneous populations

This section describes how cohort- and period-effects can be processed within mortality models.

In a previous section, we saw that population heterogeneity can arise when a population consists of homogeneous sub-populations with mutually differing (mortality-)properties. Heterogeneity also arises if a population is modelled over a long period. With the prospect of a rising life expectancy, newborns at the beginning of the period probably have a lower life expectancy than those who are born at the end of the period. The point in time of birth, or the cohort, to which an individual belongs, is important for the life expectancy. Below, the change of the mortality probability of a population during a period is modelled using the theory of heterogeneous populations.

We will make use of the earlier-described vitality v_a. By choosing for every time a different distribution of vitality, a future development of the mortality probabilities can be modelled. The conditioned mortality-intensity is defined by

$$\mu_{v_a}(x) = \mu(x - v_a) \tag{4.1}$$

with v_a ranging over the interval $(-\infty, \infty)$. In order to make the mortality probability not just age-dependent but also time-dependent, the mortality-intensity must be expanded. We define $\overline{\mu}(x,t)$ as the heterogeneous mortality-intensity of age x at time t.

The population can be divided into homogeneous sub-populations in two ways. The first method is to assume homogeneous cohorts or generations. That means that within each cohort the distribution of vitality is the same. For a cohort the distribution remains unchanged over the course of time (and age). A second possibility is to allow the distribution of vitality to vary over time. In this case, the distribution is the same for each age at a given point in time. Both of these methods are examined below.

Cohort-dependent increase of life expectancy

With a cohort- or generation-dependent distribution of vitality, the distribution is conditioned by the cohort $t_0 = t - x$, the birth year of an individual with age x at time t. The probability $f_{t_0}(v_a)$ represents the probability of vitality v_a given a cohort t_0. Because an age correction v_a is used, the vitality is defined for $(-\infty, \infty)$. The heterogeneous mortality-intensity is then defined by

$$\overline{\mu}(x,t) = \int_{-\infty}^{\infty} \mu_{v_a}(x) f_{t_0}(v_a)\, dv_a \tag{4.2}$$

The equation is a transformation of the age-cohort domain to the age-time domain. Assuming the exponential mortality-intensity

$$\mu(x) = \frac{1}{b} e^{\frac{x-a}{b}}$$

then substitution of 4.1 in 4.2 yields the following

$$\overline{\mu}(x,t) = \mu(x) \int_{-\infty}^{\infty} e^{-\frac{v_a}{b}} f_{t_0}(v_a)\, dv_a = \mu(x)\, \psi_{f_{t_0}}\left(-\frac{1}{b}\right)$$

with $\psi_{f_{t_0}}$ the moment-generating function of $f_{t_0}(v_a)$.

We find for the heterogeneous life probability of cohort t_0

$$\overline{S}(x,t_0) = S(x)^{\psi_{f_{t_0}}\left(-\frac{1}{b}\right)}$$

The life expectancy $\overline{e}(x,t_0)$ of a zero-year old for a cohort born at time t_0 is defined by

$$\overline{e}(x,t_0) = \int_x^{\infty} \overline{S}(y,t_0)\, dy$$

The increase of the life expectancy for cohort-dependent vitality is attributed solely to the cohort where the increase takes place, not to the entire population. This appears (roughly) to amount to a correction of the age.

Above we assumed that the age correction was the same for every age. By defining the conditioned mortality-intensity as an age correction times an age-dependent function $\varphi(x)$ we obtain

$$\mu_{v_a}(x) = \mu(x - \varphi(x)\, v_a)$$

This leads to

$$\overline{\mu}(x,t) = \mu(x) \int_{-\infty}^{\infty} e^{-\frac{\varphi(x)v_a}{b}} f_{t_0}(v_a)\, dv_a = \mu(x - w(x,t))$$

with $w(x,t) = -b \ln \psi_{f_{t_0}}(-\varphi(x)/b)$.

When assuming that the probability density is a point estimate at point $z(t_0)$ (i.e. $f_{t_0}(v_a) = 1$ for $v_a = z(t_0)$ and $f_{t_0}(v_a) = 0$ for $v_a \neq z(t_0)$) then this yields

$$\overline{\mu}(x,t) = \mu(x - \varphi(x)\, z(t_0))$$

where $z(t_0)$ is the increase of the life expectancy for cohort t_0 and $\varphi(x)$ the age-dependent function.

Furthermore, an appropriate function for $\varphi(x)$ would be the base-line population $S(x)$. In this way the increase of life expectancy decreases with an increasing age. When $\varphi(x) = S(x)$ then the model which the CBS uses in order to create future life tables is obtained [Cbs97a].

Period-dependent increase of life expectancy

A realistic assumption is to allow an increase in the life expectancy to influence the entire population at a specific moment. With a period-dependent distribution of vitality, the distribution is conditioned by the moment or period t. The probability $f_t(v_a)$ represents the probability of vitality v_a given a time t. The heterogeneous mortality-intensity is then defined by

$$\overline{\mu}(x,t) = \int_{-\infty}^{\infty} \mu_{v_a}(x)\, f_t(v_a)\, dv_a \tag{4.3}$$

And the following also holds

$$\overline{\mu}(x,t) = \mu(x)\, \psi_{f_t}\left(-\frac{1}{b}\right)$$

The life expectancy for a person aged x at time t is defined by

$$\overline{e}(x,t) = \int_{x}^{\infty} \overline{S}(y, t+y-x)\, dy$$

is not readily calculated because the moment-generating function has no constant value, but depends on all future expectations and variances of the probability densities $f_t(v_a)$.

The life expectancy for a fixed time t (indicated with subscript f) can be calculated by

integrating over the age and not changing the time

$$\overline{e}_f(x,t) = \int_x^\infty \overline{S}(y,t)\,dy$$

If the life expectancy rises for newborns, then this still does not indicate whether and to what extent the increase benefits the existing population. It is likely that the remaining life expectancy after reaching adulthood at group level was in part already determined in youth. However, alongside a cohort effect there is also a period-effect. For a forecast, one must indicate to what extent this is the case.

Also important is whether a cohort effect is the same for each age, or whether it extinguishes as the age gets higher. When empirical figures are studied [Cbs97b], it appears that at a lower age there can be a significant difference compared to population figures. This difference declines as age increases. It is possible that the increase in life expectancy displays a similar behaviour. Empirical figures about income shows a declining effect of the income on the life expectancy [Cbs95].

4.1.4 Forecast of the CBS

This section describes how, by using the theory of heterogeneous populations, a model for mortality prognosis can be supplemented with the mortality development trends described earlier. In [RT97], Kaufmann and Hillgärtner use in a study about extrapolating life tables a regression technique to fit a time-dependent Gompertz distribution. Here we will also model the uncertainty with respect to the forecast. The CBS has done a forecast of the life expectancy of newborns. We will therefore proceed from the assumption of a cohort-dependent increase of life expectancy.

The CBS describes three variants for the development of future mortality. The Middle-variant describes the most plausible development. The Low variant proceeds on the assumption of a low increase in life expectancy. In this variant, the population growth is also low. The opposite applies to the High variant. It is assumed that there is a 95% probability that the population size lies between the Low and the High variant. Compared to preceding forecasts, the life expectancy rises. In the current forecast, the CBS assumes that life expectancy will rise until 2050. In preceding forecasts, constant life expectancy was assumed after 2010.

To determine the expected increase, three conditions were assumed. Firstly, a linear spline model determines the short-term increase. Secondly, the life expectancy in 2050 is equal to 80 years for men and 83 years for women. Thirdly, it is assumed that the increase in 2050

is somewhat smaller than around 2020 (due to the non-linearity of the life expectancy). A quadratic function cannot do a good job of describing the slight acceleration of the trend after 2000 for women. The life expectancy values for the period between 1996 and 2050 are determined with a curve of a third-degree polynomial. The farther the prediction looks into the future, the less reliable it is. Therefore, the difference between the High- and Low-variant and the Middle-variant increases with time. The dispersion (or standard deviation) thus increases in time.

We use vitality by lowering the age, so

$$\mu_{v_a}(x) = \mu(x - v_a)$$

In the following, we explore the class of exponential mortality-intensities for which the life expectancy and dispersion are dependent on cohort t_0. The distribution $f_{t_0}(v_a)$ of the vitality is conditioned by cohort t_0. The heterogeneous mortality-intensity can then be written as

$$\overline{\mu}(x,t) = \mu(x)\,\psi_{f_{t_0}}\left(-\frac{1}{b}\right)$$

where $t_0 = t - x$. Assuming an exponentially rising mortality-intensity and from

$$a(t_0) = -b \ln \psi_{f_{t_0}}\left(-\frac{1}{b}\right)$$

this yields

$$\overline{\mu}(x,t) = \mu(x - a(t_0))$$

or a cohort-dependent distribution $G(x \mid \mu + a(t_0), \sigma)$.

Below $a(t_0)$ is determined in such a way that the development of mortality such as the CBS has predicted it for the coming fifty years is correctly modelled. Because uncertainty exists about the increase in life expectancy, we model the increase with a standard normal distribution. The average of the normal distribution is equal to the increase of the life expectancy from 0 to cohort t_0, so

$$s(t_0) = \overline{e}(0,t_0) - \overline{e}(0,0)$$

Because the High- and Low-variants over time differ from one another increasingly, there is an increasing uncertainty with respect to the prediction. The variance $\beta(t_0)$ for cohort t_0 can be derived from the Low- and High-variants.

If it is assumed that the distribution of the vitality for cohort t_0 is normally distributed with $f_{t_0}(v_a) \sim N(s(t_0), \beta(t_0))$, then the moment-generating function is equal to

$$\psi_{f_{t_0}}\left(-\frac{1}{b}\right) = e^{-\frac{s(t_0)}{b}+\frac{1}{2}\frac{\beta^2(t_0)}{b^2}}$$

Substitution yields

$$a(t_0) = s(t_0) - \frac{1}{2b}\beta^2(t_0) \tag{4.4}$$

The more recent a prediction is, the higher the standard deviation and, because it is a recent prediction, the lower the increase of life expectancy. Figure 4-1 illustrates for a number of years the corresponding standard normal distribution.

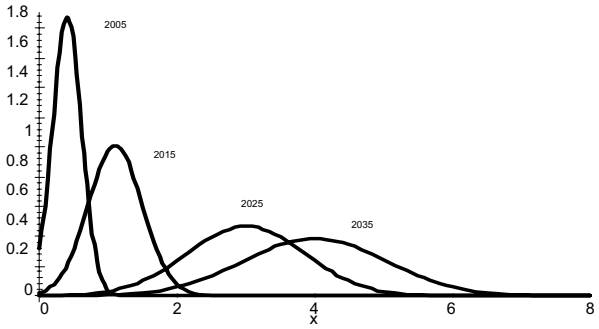


Figure 4-1: The raise of life expectancy represented by normal distributions for several years

The uncertainty relating to the increase of the life expectancy is expressed in a rising variance $\beta(t_0)$ and results in the age correction $a(t_0)$ becoming smaller.

If only a cohort-effect is considered, then the increase can be easily modelled with an age correction. With a period-effect, this is more complicated. Now, the increase per period, when totalled, must lead to the desired increase of the life expectancy.

4.2 Computational intelligence

As described in section 1.2, the task of logic according to Carnap [Car50] is summarised as providing a universal language for the description of all possible worlds (systems, structures,

models, situations) and the study and description of such a language. The notation system of such an artificial language (formal language system) has an empty but comprehensive symbolic framework and must offer space for the symbolic representation of all sciences.

The artificial language that could be applied to all fields of knowledge was never successfully developed. Instead, 'languages' remained which were specially oriented towards separate fields of knowledge. Life insurance mathematics is such a language of life insurances, see for example [Wol96].

Here we will follow the strategy of Suppes and describe an axiomatic explanation of life insurance mathematics. It will be shown that this yields a simple description and leads to an actuarial knowledge system (AKS) which contains the axioms as knowledge. Using the axioms, the system is capable of automatically representing insurances symbolically and algorithmically and thereby applying computational intelligence in two areas.

The automatic derivation of formulas is important because insurance companies are increasingly confronted with a changing demand for insurance products. The traditional solution, the imperative programming of formulas by computer experts when demanded, generates at least three problems. The first is the communication problem between computer programmers and actuaries because they each think in their own knowledge domain. Secondly, the maintenance problem caused by the traditional implementations and, thirdly, the problem of introducing new insurance products quickly in a competing market [Hen96].

4.2.1 Life insurance mathematics

Given the problem of deriving the formula of a certain life insurance, we will define a general function for the expected value of the present value of a basic life insurance. From this general function, other life insurances can be derived. After that, a representation of this knowledge will be given, resulting in a description of the Actuarial Knowledge System (AKS). Because the representation of mortality provides new actuarial applications, the last section of this chapter will describe closed form solutions for the stop-loss order of the n-th degree by using Meijer's G-functions (see the Appendix for this). This makes it possible to compare mortality distributions with regard to their risk. The following description is taken from [WK00b].

Suppose that $T(x)$ is the random variable for the remaining lifetime for an insured with age x (the insured dies at $x + T(x)$). Then, the survival distribution $F_x(t)$ is defined by

$$F_x(t) = 1 - \Pr[T(x) \leq t], \text{ for } x, t \geq 0 \tag{4.5}$$

If we assume that time causes no difference in the life expectancy of a population then we have

$$F_x(t) = \frac{F_0(x+t)}{F_0(x)}, \text{ for } x, t \geq 0$$

where $F_0(x)$ is the survival distribution of the random variable for the lifetime of a newborn. Note that $F_x(t)$ corresponds with the well known probability $p(x; x+t)$ to survive the age interval $[x, x+t)$ and that $F_0(x)$ corresponds to the survival function $S(x)$. The derivative of $1 - F_{x+k}(u)$ equals the density of the distribution

$$\frac{d}{du}\left(1 - F_{x+k}(u)\right) = p(x+k; x+k+u)\,\mu(x+k+u)$$

In the following, we will use also T for $T(x)$.

A basic element of the life insurance industry is the whole life insurance. This insurance can be defined as a payment of 1 unit, paid at time t by the insurer to the beneficiaries of the insured, given that the insured dies at time t. The following random variable for the present value of a payment at time $x+T$ can be obtained

$$Z_x(t) = \begin{cases} e^{-\delta T(x)} & \text{for } T(x) \geq t \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

where δ is the fixed interest intensity. By taking the expected value of $Z_x(t)$ the value of a whole life insurance is obtained.

The random variable $Z_x(t)$ defined by (4.6) can be seen as a building block for all life insurances. This will be explained below. In the following corollary, we introduce a general function for the k-th moment of $Z_x(t)$.

Corollary 15 *Given a random variable $T(x)$, a fixed value of δ and a random variable $Z_x(t)$ defined in (4.6) then, with a general function $G_x(t, r, T)$ defined by*

$$G_x(t, r, T) = e^{rt} F_x(t)\,\psi_{T(x+t)}(r) \tag{4.7}$$

where $\psi_{T(x)}(r)$ is the moment generating function of $T(x)$ evaluated at r, the k-th moment of $Z_x(t)$ is

$$E\left[Z_x^k(t)\right] = G_x(t, -k\delta, T) \tag{4.8}$$

Proof. Take $\psi_{T(x)}(r) = E\left[e^{rT(x)}\right]$, i.e. the moment generating function of $T(x)$ evaluated at

r, hence

$$\begin{aligned}\psi_{T(x+t)}(r) &= E\left[e^{rT(x+t)}\right] \\ &= -\int_0^\infty e^{ru} dF_{x+t}(u)\end{aligned}$$

With (4.5), we have

$$F_{x+t}(u) = \frac{F_0(x+t+u)}{F_0(x+t)} = \frac{F_0(x+t+u)/F_0(x)}{F_0(x+t)/F_0(x)} = \frac{F_x(t+u)}{F_x(t)}$$

and

$$\psi_{T(x+t)}(r) = -\int_0^\infty e^{ru} d\frac{F_x(u+t)}{F_x(t)}$$

Then by substitution we find

$$\begin{aligned}E\left[Z_x^k(t)\right] &= -\int_t^\infty e^{-k\delta u} dF_x(u) \\ &= -e^{-k\delta t} F_x(t) \int_0^\infty e^{-k\delta u} d\frac{F_x(u+t)}{F_x(t)} \\ &= e^{-k\delta t} F_x(t)\, \psi_{T(x+t)}(-k\delta) \\ &= G_x(t, -k\delta, T)\end{aligned}$$

which proves the corollary.

■

Below, the function $G_x(t, -\delta, T)$ will be abbreviated to $G_x(t)$, i.e. the case $k = 1$. The reason why (4.8) is defined for moments higher than 1 is that this general definition also allows calculating variances of life insurances.

Given a decreasing remaining lifetime, the following property can be derived. For $t > 0$, the general function $G_x(t)$ will ultimately decrease and become zero. This is caused by the fact that t is the deferred period of the insurance and the higher the age the more probable it is that the insured dies within this period. Therefore, the value of $G_x(t)$ converges to zero. So, for a real valued t,

$$\lim_{x\to\infty} G_x(t) = \begin{cases} 1 & \text{for } t = 0 \\ 0 & \text{for } t > 0 \end{cases}$$

and

$$\lim_{x \to -\infty} G_x(t) = 0$$

With (4.8) we can obtain

$$G_x(0) = \overline{A}_x$$

where $\overline{A}_x$ relates to a whole life insurance for an insured aged x. In the next section, we will show how to derive other insurances from the basic building block $Z_x(t)$.

The general function $G_x(t)$ will show its power: to represent and to deduce. It is a 'germ' for the deductive power of computationally intelligent systems for life insurances. To illustrate this, the next section will show how insurances can be expressed with the germ function $G_x(t)$.

Life insurance examples

In this section, we will show how to deduce insurances from the germ function $G_x(t)$. We will start with examples of temporary and deferred insurances.

Example 16 (temporary) *From (4.6) it is easy to see that*

$$Z_x(t) - Z_x(t+n) = \begin{cases} e^{-\delta T(x)} & \text{for } t \le T(x) < t+n \\ 0 & \text{otherwise} \end{cases}$$

It follows that the expected values of n-years temporary insurance (i.e. cases $k = 1$) can be deduced from the germ function $G_x(t)$

$$E[Z_x(t) - Z_x(t+n)] = G_x(t) - G_x(t+n)$$

Example 17 (deferred) *To represent a m-years deferred insurance with help of $G_x(t)$ is easy*

$$Z_x(t+m) = \begin{cases} e^{-\delta T(x)} & \text{for } T(x) \ge t+m \\ 0 & \text{otherwise} \end{cases}$$

$$E[Z_x(t+m)] = G_x(t+m)$$

Increasing insurances exist in linear and geometric form. Both forms can be deduced from the germ function $G_x(t)$. This is exemplified in examples 18 and 19.

Example 18 (linear increase) *In case of linear increase a payment of $T(x) - t$ is done if the insured dies, so the increase starts after time t. The discounted value of this payment is*

$(T(x)-t)\,Z_x(t)$. *If considering the random variable* $Z_x(t)$ *as defined in (4.6), the expected value of a linearly increasing insurance can be expressed in terms of the germ function* $G_x(t)$ *by*

$$E\left[(T(x)-t)\,Z_x(t)\right] = -\frac{d}{d\delta}G_x(t) - tG_x(t) \tag{4.9}$$

With

$$\begin{aligned} E\left[T(x)\,Z_x(t)\right] &= -\int_t^\infty ue^{-\delta u}dF_x(u) \\ &= \frac{d}{d\delta}\int_t^\infty e^{-\delta u}dF_x(u) \\ &= -\frac{d}{d\delta}G_x(t) \end{aligned}$$

we prove (4.9).

Example 19 (geometric increase) *With geometric increase a payment of* $e^{sT(x)}$ *is done. Insurances with a geometric increase are simple to compute. With increasing intensity* s *the expected value is*

$$E\left[e^{sT(x)}Z_x(t)\right] = G_x(t, s-\delta, T)$$

Note that $E\left[Z^k\right]$ *has meaning if* k *is non-integer. For* $k = 1-\frac{s}{\delta}$*, we find for* $E\left[Z^k\right] = G_x(t, -k\delta, T)$ *the geometric increasing insurance* $G_x(t, s-\delta, T)$.

The insurances above are endowments. Now, we turn to a t years deferred annuity with a random variable $Y_x(t)$ defined by

$$Y_x(t) = \begin{cases} \int_{T(x)-T(x+t)}^{T(x)} e^{-\delta u}du & \text{for } T(x) \geq t \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

Example 20 (annuity) *If an annuity starts after* t *years we have with (4.10)*

$$Y_x(t) = \frac{1}{\delta}Z_x(t)\left(Z_{x+t}^{-1}(0)-1\right)$$

concluding

$$E\left[Y_x(t)\right] = \frac{1}{\delta}G_x(t)\left(G_{x+t}^{-1}(0)-1\right) \tag{4.11}$$

Example 21 (increasing annuity) *A linear increasing annuity can be found by using the increasing property with (4.11)*

$$E\left[T\left(x\right)Y_{x}\left(0\right)\right]=\frac{1}{\delta}\left(1+\frac{d}{d\delta}G_{x}\left(0\right)\right)$$

When an actuary calculates the insurance, he actually performs the operations described above, albeit that he is unaware of the exact definitions of the operations. Therefore the knowledge of the operations can be seen as deep procedural and primitive knowledge.

The operations above are computationally independent. An insurance can be temporary, deferred and increasing at the same time. The properties are deducted from the same germ function $G_x\left(t\right)$. Note that the order of deductions is not indifferent.

Life insurance reserves

Besides expressing many insurance lump sums, the germ function is also helpful in making actuarial reserve calculations. For example, if we want to calculate the remaining lump sum reserve of a whole life insurance.

Corollary 22 *The net lump sum reserve of a whole life insurance can be expressed by the germ function in the following way*

$${}_{k}V=\frac{G_{x}\left(t\vee k\right)}{e^{-\delta k}F_{x}\left(k\right)}=G_{x+k}\left(0\vee t-k\right)$$

where the infix operator in $a\vee b$ identifies the maximum of a and b.

Proof. Because the whole life insurance is t years deferred, no payments are done before this time. Therefore, the reserve before time t is given by $G_x\left(t\right)$ divided by the probability to survive $[x,x+k)$ and the interest over this period, or $e^{-\delta k}F_x\left(k\right)$. After time t, the remaining reserve is $G_x\left(k\right)/\left(e^{-\delta k}F_x\left(k\right)\right)$. This first part of the corollary can be seen as a net retrospective reserve. The lump sum G is divided by a ${}_{k}E_{x}$, in actuarial notation. What remains to be proven is the second part of the corollary. We have

$$G_{x+k}\left(0\vee t-k\right)=e^{-\delta(0\vee t-k)}F_{x+k}\left(0\vee t-k\right)E\left[e^{-\delta T(x+(k\vee t))}\right]$$

and with

$$F_{x+k}\left(0\vee t-k\right)=\frac{F_{x}\left(k\vee t\right)}{F_{x}\left(k\right)}$$

we find

$$G_{x+k}\left(0 \vee t-k\right)=\frac{e^{\delta k}}{F_x\left(k\right)}G_x\left(t \vee k\right)$$

Which proves the corollary.

■

The second part can also be rewritten to a net prospective reserve. We find the following integrals for ${}_kV$

$${}_kV=\begin{cases} -\int_t^\infty e^{-\delta u}dF_{x+k}\left(u\right) & \text{for } t\leq k \\ -\int_t^\infty e^{-\delta(u-k)}dF_{x+k}\left(u-k\right) & \text{for } t>k \end{cases}$$

Having defined a computational and analytical tool for making actuarial calculations, the question remains how to calculate the germ function given a certain mortality assumption. At this point, normally a discrete life table is introduced which represents, possibly graduated, observed mortality probabilities of a certain period. Consequently, a quantitative model is derived without any knowledge of how the properties of the life table influence the germ function. Analytical mortality laws for life tables are not often used because of the mathematical and computational complexity. However, with the introduced germ function it becomes possible to use well known mortality laws in an easy way. Below, two analytical insurance germs are derived: the Gompertz insurance germ and the Makeham insurance germ.

4.2.2 Life insurance germs

To be able to calculate the insurances described above, assumptions about the mortality probabilities are necessary. In this section, closed form solutions for life insurances contracts such as annuities and whole life insurances are defined by using special functions. The mortality distribution used in this section is based on the general model of mortality described in section 2.5.

The Gompertz Insurance Germ

In the following, we will assume that $F_0\left(x\right)$ is a Gompertz distribution. Then, based on the earlier-described and reformulated distribution of Gompertz, life probabilities and life expectancies can be calculated with the aid of the incomplete Gamma function. The heterogeneous life distribution follows from this and generalisation of the model to a model with more variables is possible, for example by using a multiplicative hazard and a log linear vitality function resulting in the distribution $G\left(x \mid \mu+b\boldsymbol{\beta}^T\mathbf{v},\sigma\right)$ (as described in section 3.6.2) or by using a

cohort-dependent distribution $G\left(x \mid \mu + s\left(t_0\right) - \beta^2\left(t_0\right)/2b, \sigma\right)$ (as described in section 4.1.4). Using Gompertz' distribution, it is possible to give analytical formulas of traditional life insurances and insurances based on Cox's regression model [Cox72].

Because all insurances can be expressed by the germ $G_x\left(t\right)$ as defined in (2.36), it is necessary to derive the moment generating function of $T\left(x+t\right)$ evaluated at point r. The moment generating function of Gompertz' distribution is a Gamma function multiplied with a factor. Then, the n-th derivative of the moment generating function is a derivative of the Gamma function. These derivatives are called Meijer's G-functions.

Corollary 23 *Let $T\left(x\right)$ be a random variable for the remaining lifetime with a Gompertz distribution, then the moment generating function of $T\left(x\right)$ for a real variable x equals*

$$\psi_{T(x)}\left(r\right) = z^{-rb}e^{z}\Gamma\left(1+rb, z\right) \tag{4.12}$$

with $z = e^{(x-a)/b}$ and its n-th derivative equals

$$\frac{d^n}{dr^n}\psi_{T(x)}\left(r\right) = n!b^n z^{1-rb}e^{z}G^{n+2,0}_{n+1,n+2}\left(z\left|\begin{array}{cccc} 0 & \ldots & 0 & \\ -1 & \ldots & rb & -1 \end{array}\right.\right)$$

Proof. If X has a Gompertz distribution then

$$\psi_{I_{[x,\infty)}(X)}\left(r\right) = e^{ra}\Gamma\left(1+rb, e^{\frac{x-a}{b}}\right)$$

and because

$$F_0\left(x\right)\psi_{T(x)}\left(r\right) = e^{-rx}\psi_{I_{[x,\infty)}(X)}\left(r\right)$$

equation (4.12) can be obtained. In section 4.2.4 a general result based on stop-loss order is obtained.

■

The Gompertz Insurance Germ $GIG_x\left(t\right)$ can now be found by substituting the moment generating function of $F_0\left(t\right)$ in (4.7).

Definition 24 *The Gompertz Insurance Germ (GIG) is defined by*

$$GIG_x\left(t\right) = z^{\delta b}e^{z}\Gamma\left(1-\delta b, \omega z\right)$$

with $z = e^{(x-a)/b}$, $\omega = e^{t/b}$, and δ the interest intensity and a and b the Gompertz parameters.

An example of the Gompertz Insurance Germ is given in Figure 4-2. For $t = 0$ the germ converges to one, i.e. the probability to die within a small period becomes larger. For a one year deferred insurance the germ decreases after an age, i.e. the probability to die within one year becomes larger, so no payment had to be done.

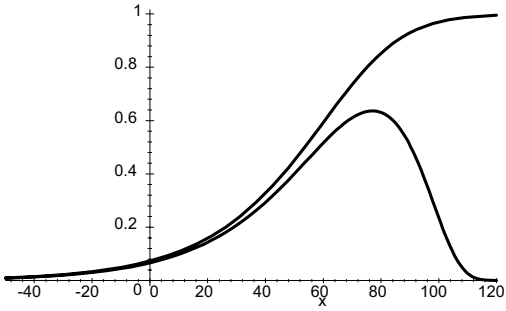


Figure 4-2: The Gompertz Insurance Germ for $a = 75$, $b = 10$ and $\delta = 0.04$ for $t = 0$ and $t = 1$

The Makeham Insurance Germ

In the following, we will assume that $F_0(x)$ is a Makeham distribution. If we assume that the random variable is according to Makeham's law of mortality then we have, with the reformulation of Gompertz' law of mortality in mind, $F_0(x) = e^{sx-z}$ and $F_x(t) = e^{st+z(1-\omega)}$ with $z = e^{(x-a)/b}$ and $\omega = e^{t/b}$. Note that normally the logarithmic of s is taken, i.e. $e^{sx}g^{c^x}$ is used instead of $s^x g^{c^x}$.

Corollary 25 *Let $T(x)$ be a random variable for the remaining lifetime with a Makeham distribution, then the moment generating function of $T(x)$ for a real variable x equals*

$$E\left[e^{rT(x)}\right] = z^{-ub}e^z\left(\Gamma(bu+1,z) - bs\Gamma(bu,z)\right)$$

with $u = (s+r)$ and $z = e^{(x-a)/b}$.

Proof. Generally, the following is true

$$\begin{aligned} E\left[e^{rT(x)}\right] &= \int_0^\infty e^{ru} d\left(1 - F_x\left(u\right)\right) \\ &= \frac{e^{-rx}}{F_0\left(x\right)} \int_x^\infty e^{ru} d\left(1 - F_0\left(u\right)\right) \end{aligned}$$

For the Makeham survival distribution, we have $F_0\left(u\right) = e^{su-w}$, with $w = e^{(u-a)/b}$, and for the density, defined by $f_0\left(u\right) = \frac{d}{du}\left(1 - F_0\left(u\right)\right)$, we have

$$f_0\left(u\right) = \left(\frac{1}{b}w - s\right) e^{su-w}$$

Substituting this leads to

$$\begin{aligned} E\left[e^{rT(x)}\right] &= \frac{e^{-rx}}{F_0\left(x\right)} \int_x^\infty \left(\frac{1}{b}w - s\right) e^{(s+r)u-w} du \\ &= \frac{e^{-rx}}{F_0\left(x\right)} \left(\frac{1}{b}\int_x^\infty we^{(r+s)u} e^{-w} du - s\int_x^\infty e^{(r+s)u} e^{-w} du\right) \\ &= \frac{e^{-rx}}{F_0\left(x\right)} \left(\frac{1}{b} e^{\frac{-a}{b}} \int_x^\infty e^{\left(r+s+\frac{1}{b}\right)u} e^{-w} du - s\int_x^\infty e^{(r+s)u} e^{-w} du\right) \end{aligned}$$

By substituting $s = e^{\frac{-a}{b}}$, $t = e^{\frac{u}{b}}$, $a = br$, $w = e^{(u-a)/b}$, $z = e^{(x-a)/b}$ and $p = e^{x/b}$ in the (upper) incomplete Gamma function

$$\int_p^\infty t^{a-1} e^{-st} dt = \frac{\Gamma\left(a, sp\right)}{s^a}$$

we have

$$\int_x^\infty e^{ur} e^{-w} du = be^{ar} \Gamma\left(br, z\right)$$

Applying this leads to

$$E\left[e^{rT(x)}\right] = z^{-ub} e^z \left(\Gamma\left(bu + 1, z\right) - sb\Gamma\left(bu, z\right)\right)$$

which proves the corollary.

■

For the moments we find

$$\begin{aligned} & E\left[T^n(x)\right]|_{r=0} \\ = \; & n!b^n z^{1-sb} e^z \left(G^{n+2,0}_{n+1,n+2}\left(z \left| \begin{matrix} 0 & ... & 0 & \\ -1 & ... & bu & -1 \end{matrix} \right. \right) - bsG^{n+2,0}_{n+1,n+2}\left(z \left| \begin{matrix} 0 & ... & 0 & \\ -1 & ... & bu & -1 \end{matrix} \right. \right) \right) \end{aligned}$$

The Makeham Insurance Germ $MIG_x(t)$ can now be found by substituting the moment generating function of $F_0(t)$ in (4.7).

Definition 26 *The Makeham Insurance Germ (MIG) is defined by*

$$MIG_x(t) = z^{-ub} e^z \left(\Gamma(bu+1, \omega z) - bs\Gamma(bu, \omega z)\right)$$

with $u = (s-\delta)$, $z = e^{(x-a)/b}$, $\omega = e^{t/b}$, δ *the interest intensity and* a, b, s *the Makeham parameters.*

The Gompertz insurance germ can be derived by substituting $s = 0$, nullifying the last part.

Because the mortality model can be used in a large number of different types of insurances, a method that simplifies application of the model is important. A basic function is introduced from which other types of insurances can be automatically derived. This automatic derivation is executed by the AKS expert system (Actuarial Knowledge System). This system contains expert knowledge from the actuarial field in order to automate derivation of closed form solutions for insurance contracts. The design of AKS will be described below.

4.2.3 Representing insurance knowledge

Traditional life insurances are represented in three global ways: natural language, symbolic and algorithmic. The specialised intellectual field of life insurance mathematics describes the interrelations between the types of representation.

The first type of representation is the natural language with which an insurance is described. There are annuities (with a temporary or life-long period of benefits) which generate a periodic payment while an insured is alive, and whole life insurances, which pay a capital at the time the insured dies. As more properties play a role, the description of the insurance becomes more complex and difficult to explain clearly.

The symbolic representation gives a brief and reliable representation of an insurance. Symbols can be quickly read, and experts are capable of quickly expressing one symbolic repre-

sentation into another. The significance of a symbol does require that one has mastered the essentials of the specialised intellectual field.

The algorithmic representation describes how an insurance should be calculated. This representation can usually be expressed in a computer language. Different algorithmic representations have different properties concerning the efficiency of processing.

In the following two sections, the representation by symbols and the representation by algorithms is described.

Symbolic representation

The symbol of a life insurance is based on the standard International Actuarial Notation and described in [Wil95], which is implemented by using a 'template' in [Hor96]. The template

$$\begin{array}{c} {}^{5}\,{}^{(4)} \\ {}_{1|}\,\Phi_{2:\overline{3}|} \end{array}$$

indicates the general form of a traditional life insurance, whereby the numbers indicate the positions of the properties of the insurance. The symbol Φ indicates whether an insurance for life ($\Phi = a$) or for death ($\Phi = A$) is considered.

The other properties are:

1 This position indicates whether the insurance is deferred, i.e. that the insurance commences at a later point in time. If omitted, then the insurance commences immediately.

2 This position indicates what person or persons the insurance covers. Generally, the identification of the person is the age, because only the age is relevant for the risk, but when the person is in a certain group or state, then that is also indicated here.

3 This position indicates the duration of the insurance. If omitted, then it is a lifetime insurance.

4 This position, with brackets, indicates the payment frequency. If omitted, then payment is once a year.

5 This position indicates the moment of payment with $\{\emptyset, -, \cdot\cdot\}$. If omitted, then payment is at the end of the year. If there is a dash, then it is continuous payment, while two points is indicates payment in advance.

For example, the symbol

$$ {}_{m|}\ddot{a}^{(12)}{}_{x} $$

indicates a m years deferred life insurance for an individual aged x with a monthly payment in advance.

With the representation by symbols, one still does not have the algorithm for calculating the insurance. The algorithm is described below.

Algorithmic representation

Below all properties of the insurances are associated with short elementary formulas. By applying the formulas associated with the properties, the corresponding formulas can be derived. Two cases are considered: continuous and discrete payment.

We start the process with a whole life insurance $G_x(t)$. The properties of insurances can be translated into formulas. The following table shows the properties of continuous insurances with corresponding substitutions.

Nr.	Operation	Deduction
1	endowment	map $G_x(t)$ onto $G_x(t)$
2	annuity	map $G_x(t)$ onto $\frac{1}{\delta}G_x(t)\left(G^{-1}_{x+t}(0)-1\right)$
3	n years temporary	map $G_x(t)$ onto $G_x(t)-G_x(t+n)$
4	m years deferred	map $G_x(t)$ onto $G_x(t+m)$
5	geometric increase with s	map $G_x(t)$ onto $G_x(t, s-\delta, T)$
6	linear increase	map $G_x(t)$ onto $-tG_x(t)-\frac{d}{d\delta}G_x(t)$
7	age correction with α	map $G_x(t)$ onto $G_{x-\alpha}(t)$

By applying the formulas by substituting them into each other, the actuarial formulas can be automatically derived. Take for example a temporary deferred life insurance

$$ {}_{m|}\overline{A}^{1}_{x:n} \quad = \boxed{\begin{matrix}\text{temporary}\\ \text{(n)}\end{matrix}} \leftarrow \boxed{\begin{matrix}\text{deferred}\\ \text{(m)}\end{matrix}} \leftarrow \quad G_x(0) $$

Substitution of the formulas of the properties yields the formula

$$ {}_{m|}\overline{A}_{x:n} = G_x(m) - G_x(m+n) $$

A temporary annuity is found by

$$\overline{a}_{x:n} \quad = \quad \boxed{\begin{matrix}\text{temporary}\\(\text{n})\end{matrix}} \leftarrow \boxed{\begin{matrix}\text{annuity}\\(\text{x})\end{matrix}} \leftarrow \quad G_x\left(0\right)$$

Substitution of the formulas of the properties yields the formula

$$\overline{a}_{x:n} = \frac{1}{\delta}\left(1 - \left(G_x\left(0\right) - G_x\left(n\right)\right)\right)$$

The life expectancy of the female population is usually higher than the life expectancy of the male population. When a single explanatory variable g representing the gender of the insured is introduced with $g = 0$ if the insured is a man and $g = 1$ if the insured is a woman, and the difference in life expectancy is for example three years, then $b\boldsymbol{\beta}^T\mathbf{v} = 3$ with $\mathbf{v} = [g]$. In case of a whole life insurance for a female this results in

$$\overline{A}_x \quad = \quad \boxed{\begin{matrix}\text{age correction}\\(\text{female})\end{matrix}} \leftarrow \quad G_x\left(0\right)$$

Substitution of the properties yields the formula

$$\overline{A}_x = G_{x-3}\left(0\right)$$

Other explanatory variables can be used.

The explicit functions for insurances above make it possible to perform actuarial calculations quickly and efficiently. Only the basic function $G_x\left(t\right)$ is used to derive all other often-used insurance formulas.

Application of computational intelligence makes it possible to automatically derive virtually all traditional insurances (without exhaustive search for all possible combinations). Furthermore, it is possible to use explanatory variables for calculating insurances for special groups. Very fast algorithms already exist for calculating the Gamma function. The above can moreover be applied in new actuarial techniques as will be described below.

Deriving premiums and variances of life insurances

The net premium P for a whole life insurance equals the expected value of $Z_x(0)$

$$P = E[Z_x(0)] = G_x(0)$$

Often, to derive the gross premium, a linear load is used to increase the premium and, although more frequently used in risk theory, sometimes the load is related to the risk of the insurance, i.e. the variance, resulting in

$$P' = (1+\alpha) E[Z_x(0)] + \beta \operatorname{Var}[Z_x(0)] \tag{4.13}$$

where α is the linear load for the expected value and β the load related to the risk of the insurance.

The theory presented above simplifies these calculations. For the variance of $Z_x(t)$ we find

$$\operatorname{Var}[Z_x(t)] = G_x(t, -2\delta, T) - G_x^2(t, -\delta, T)$$

The variance of an increasing insurance can be calculated with

$$E\left[T^2 e^{2tT}\right] = \frac{d^2}{dt^2}\psi_T(2t)$$

so we have

$$\operatorname{Var}[T(x) Z_x(t)] = \frac{d^2}{dt^2}\psi_{T(x)}(2t) - \left(\frac{d}{dt}\psi_{T(x)}(t)\right)^2$$

If the insured population behaves according a Gompertz distribution then increasing insurances show the premium $E[T(x) Z_x(t)] \pm \sigma$, with $\sigma = \sqrt{\operatorname{Var}[T(x) Z_x(t)]}$, as the figure below shows.

Nowadays, more complex insurances than whole life insurances exist. The following example describes how a Universal Life can be calculated by using ordinary functions. The following example shows how these new insurances can be expressed in traditional ones with an age addition.

Corollary 27 *Take the following insurances at time $t = 0$*

- *payment of one unit at time n if the insured is alive at time n with an additional payment of $\theta\ {}_tV$ units at time t if the insured dies before time n, with $0 \leq \theta < 1$ and ${}_tV$ the reserve*

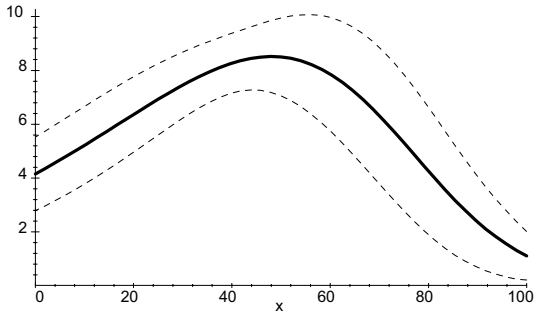


Figure 4-3: The premium at age x of an increasing insurance +/- σ

at time t, where $X \sim G(x \mid \mu, \sigma)$

- *payment of one unit at time n if the insured is alive at time n, where*

  $X^* \sim G(x \mid \mu - b \ln(1-\theta), \sigma)$

If the premiums equal the expected value of the insurance, then the premiums will be the same

$$P = P^*$$

Proof. First we derive premium P from Thiele's reserve differential equation

$$\frac{d}{dt}\,{}_tV = \delta\,{}_tV - \left(u^O(t) - {}_tV\right)\mu(x+t) + p(t)$$

where $u^O(t)$ is the amount paid when the insured dies, and $p(t)$ the premium intensity at time t. Without premium payment $p(t) = 0$ for $t \geq 0$, the reserve ${}_tV$ increases with an interest intensity and decreases with the risk premium of $(\theta - 1)$ times the reserve ${}_tV$, so $u^O(t) = \theta\,{}_tV$. Then the following differential equation holds

$$\frac{d}{dt}\,{}_tV = (\delta - (\theta-1)\mu(x+t))\;{}_tV$$

With ${}_0V = 1$ and solving this equation, one obtains

$$ {}_tV = e^{\delta t} \left(F_x\left(t\right)\right)^{\theta-1} $$

so the present value of the first insurance equals $e^{-\delta n} \left(F_x\left(n\right)\right)^{1-\theta}$.

For (continuous) premium payment, a similar method results in the following premium

$$ P = \frac{e^{-\delta n} \left(F_x\left(n\right)\right)^{1-\theta}}{\int_0^n e^{-\delta t} \left(F_x\left(t\right)\right)^{1-\theta} dt} = \left(\int_0^n e^{\delta(n-t)} \left(F_{x+t}\left(n-t\right)\right)^{\theta-1} dt \right)^{-1} \tag{4.14} $$

We have for the mortality intensity of the second population

$$ \mu_{X^*}\left(x\right) = \frac{1}{b} e^{\frac{x-a^*}{b}} = \left(1-\theta\right) \frac{1}{b} e^{\frac{x-a}{b}} = \left(1-\theta\right) \mu_X\left(x\right) $$

with $a^* = a - b \ln\left(1-\theta\right)$. So, by substitution this results in

$$ \left(F_{x+t}\left(n-t\right)\right)^{\theta-1} = F^*_{x+t}\left(n-t\right) $$

and therefore $P = P^*$.

■

Note that case $\theta = 1$ is no insurance. Case $\theta > 1$ cannot be calculated with an age correction but can be found by evaluating (4.14).

Example 28 *With the parameters values* $n = 20$, $a = 81.751, b = 11.695$ *and* $\delta = \ln 1.04$ *the following table can be derived for different ages* x *and reserve payments* θ

θ	$P\ (\times 1000)$ x=20	$P\ (\times 1000)$ x=40	$P\ (\times 1000)$ x=60
0	32.402	30.112	19.817
0.25	32.533	30.796	22.566
0.5	32.664	31.493	25.648
0.75	32.796	32.203	29.091
1	32.928	32.928	32.928
1.25	33.060	33.666	37.188
1.5	33.193	34.418	41.904

The premium for an insured aged x for first insurance is equal to the premium for an insured aged $x + b \ln(1-\theta)$ for the second insurance. Also, for the same premium, some one who is 20 years older can get almost 82% from the reserve if he dies.

Extensions of the AKS

It is possible to represent the basic function $G_x(t)$ as a matrix, distinguishing along with conditions like 'life' and 'death' other conditions can also be distinguished. When n populations are assumed with life expectancy μ_i and standard deviation σ_i for sub-population $i = 1, ..., n$, and an absorbing state d (death) is assumed then the following transition probabilities can be derived.

The density in state i is the probability to survive age x and not to move to another state

$$S_{ii}(x) = G(x \mid \mu_i, \sigma_i) - \sum_{k \geq 1, k \neq i, k \neq d} S_{ik}(x)$$

for $i = 1, ..., n$, and $S_{ik}(x)$ is the density moving from i to k.

The probability to move to state d is

$$S_{id}(x) = 1 - G(x \mid \mu_i, \sigma_i)$$

Then $\sum_{k \geq 1} S_{ik}(x) = 1$, for $i = 1, ..., n$ and $S_{dd}(x) = 1$. The formulas above remain essentially the same, with the possibility of making actuarial calculations based on Markov chains.

An example of a Markov process is the KAZO disability model where, along with death, the disability risk is also included. The model was developed by an actuarial committee of Dutch life insurers and its purpose was to design a tariff for individual disability insurances. The committee used data from the Centre for Insurance Statistics over the period 1984-1988, as well as other material. The result is presented in a publication from ASTIN Netherlands [Ast90].

The model distinguishes a condition a (active) and d (died) and the conditions i_s, with s the duration of disability. The distinction is important because the recovery probabilities drop as the duration of the disability increases. The transition probabilities are calculated without using mortality-intensities. The weighted least squares method was used as an approximation method for disability probabilities and recovery probabilities.

Two assumptions were made in order to simplify the model. The first entails that the

probability of death for a disabled is the same as for an active individual. The second assumption is that no one can recover and then become disabled again in the same year.

The life tables of the AG are used for the mortality probability. The disability probability is approximated by an exponential function, which is dependent on the attained age

$$p_{ia}\left(t+1,t\right)=a\cdot b^{x+t}$$

The recovery probability is approximated with a linear decrement function of the attained age. The parameters in the function are dependent on the duration of disability s

$$p_{ai_s}\left(t+1,t\right)=a_s-b_s\left(x+t\right)$$

For calculation of this sort of insurances, the same representation as above can be used and the same operations.

With the new formulation of Gompertz' law of mortality it is not only possible to compute the value of insurances easily but also mortality risks can be compared with each other. In the following section, it is described how mortality risks based on Gompertz' law of mortality can be compared.

4.2.4 Ordering of actuarial risks

In utility theory, the behaviour of risk-averse decision-makers is modelled by applying a non-decrement concave utility function to a given risk. A utility function can be seen as an efficiency criterion for making choices between risks. Generalisations of this criterion improved the standard methods such as average and variance analysis familiar from financial theory (whereby one prefers a risk whose first moments are the smallest). Actuarial studies of ordering of risks are often based on efficiency criteria. Research is aimed at finding and applying orders, proving that orders are preserved under certain manipulations and methods for defining safe distributions starting from only a few characteristics [Hee91].

A certain order of risks can be expressed with actuarial inequalities. For this, the stop-loss order and the accompanying stop-loss transformations are often used. A risk X precedes risk Y in stop-loss order of the n-th degree, represented by $X\leq_{(n)}Y$, if $E\left[X^n\right]$ and $E\left[Y^n\right]$ are finite and satisfy

$$E\left[X^k\right]\leq E\left[Y^k\right],\text{ for }k=1,2,...,n-1$$

and

$$E\left[\left((X-x)_+\right)^n\right] \leq E\left[\left((Y-x)_+\right)^n\right], \text{ for all } x \geq 0$$

The stop-loss of the n-th degree can be written as the sum of moments of $I_{[x,\infty)}(X)$

$$E\left[\left((X-x)_+\right)^n\right] = \sum_{0\leq k\leq n} \binom{n}{k} (-x)^{n-k} E\left[\left(I_{[x,\infty)}(X)\right)^k\right] \tag{4.15}$$

The k-th moment of $I_{[x,\infty)}(X)$ can be found by evaluating the moment-generating function of $I_{[x,\infty)}(X)$ at $t=0$. The moment-generating function $I_{[x,\infty)}(X)$ of a Gompertz distribution can be easily found by using the (above) incomplete Gamma function

$$\psi_{I_{[x,\infty)}(X)}(t) = e^{at}\Gamma\left(bt+1, e^{\frac{x-a}{b}}\right) \tag{4.16}$$

whereby the incomplete Gamma function is equal to $\Gamma(a,z) = \int_z^\infty e^{-t}t^{a-1}dt$.

The stop-loss order of the n-th degree of the Gompertz distribution is derived below [WK00a]. There is a strong relationship between stop-loss order of the n-th degree and the Meijer's G-functions, as the following theorem describes.

Theorem 29 *Given a random variable X with the Gompertz distribution, then the following can be said for the stop-loss order of the n-th degree*

$$E\left[\left((X-x)_+\right)^n\right] = n!zb^n G^{n+2,0}_{n+1,n+2}\left(z \left| \begin{matrix} 0 & \ldots & 0 & 0 & \\ -1 & \ldots & -1 & 0 & -1 \end{matrix} \right.\right) \tag{4.17}$$

with $z = e^{\frac{x-a}{b}}$.

Proof. Because, with

$$E\left[(X-x)_+\right] = F_0(x)\,\psi_{T(x)}(r)$$

one knows that

$$F_0(x)\,\psi_{T(x)}(r) = e^{-rx}\psi_{I_{[x,\infty)}(X)}(r)$$

The n-th derivative is

$$F_0(x)\frac{d^n}{dr^n}\psi_T(r) = e^{-rx}\sum_{0\leq k\leq n}\binom{n}{k}(-x)^{n-k}\frac{d^k}{dr^k}\psi_{I_{[x,\infty)}(X)}(r) \tag{4.18}$$

By evaluation at point $r=0$, the stop-loss of the n-th degree can be written as a sum of

moments of $I_{[x,\infty)}(X)$, or, generally,

$$E\left[\left((X-x)_+\right)^n\right] = \sum_{0\le k\le n}\binom{n}{k}(-x)^{n-k}E\left[\left(I_{[x,\infty)}(X)\right)^k\right]$$

The k-th moment of $I_{[x,\infty)}(X)$ can be found by evaluation of the k-th derivative of the m.g.f. of $I_{[x,\infty)}(X)$ at $t=0$. The m.g.f. of $I_{[x,\infty)}(X)$ can easily be found taking the Incomplete Gamma Function as is shown above

$$\psi_{I_{[x,\infty)}(X)}(r) = e^{ar}\Gamma\left(1+rb, e^{\frac{x-a}{b}}\right)$$

The k-derivative of the m.g.f. then equals

$$\frac{d^k}{dr^k}\psi_{I_{[x,\infty)}(X)}(r) = e^{ar}\sum_{0\le m\le k}\binom{k}{m}a^{k-m}\frac{d^m}{dr^m}\Gamma\left(1+rb, e^{\frac{x-a}{b}}\right) \tag{4.19}$$

Substitution (4.19) in (4.18) results in

$$F_0(x)\frac{d^n}{dr^n}\psi_T(r) = e^{r(a-x)}\sum_{0\le k\le n}\binom{n}{k}(a-x)^{n-k}\frac{d^k}{dr^k}\Gamma\left(1+rb, e^{\frac{x-a}{b}}\right) \tag{4.20}$$

The Meijer's G-function satisfies the following equation

$$\frac{d^k}{dr^k}\Gamma\left(1+rb, e^{\frac{x-a}{b}}\right) = e^{\frac{x-a}{b}}\sum_{0\le m\le k}\frac{k!}{(k-m)!}(x-a)^{k-m}b^m C_m\left(e^{\frac{x-a}{b}}, rb\right)$$

with $C_m\left(e^{\frac{x-a}{b}}, rb\right) = G^{m+2,0}_{m+1,m+2}\left(e^{\frac{x-a}{b}}\middle|\begin{array}{ccccc} 0 & ... & 0 & 0 & \\ -1 & ... & -1 & rb & -1\end{array}\right)$.

Substituting this in (4.20) and simplifying leads to

$$F_0(x)\frac{d^n}{dr^n}\psi_T(r) = e^{r(a-x)}e^{\frac{x-a}{b}}n!b^n C_n\left(e^{\frac{x-a}{b}}, rb\right)$$

Evaluating both sides at point $r=0$ gives us with $z = e^{\frac{x-a}{b}}$

$$E\left[\left((X-x)_+\right)^n\right] = zn!b^n G^{n+2,0}_{n+1,n+2}\left(z\middle|\begin{array}{ccccc} 0 & ... & 0 & 0 & \\ -1 & ... & -1 & 0 & -1\end{array}\right)$$

■

Numerically evaluating Meijer's G-functions can be done with mathematical software packages.

The remaining life span $e_X(x)$ at age x is defined by $\pi_X(x)/(1-G_X(x))$ with the stop-loss transformation of X defined by $\pi_X(x) \equiv E\left[(X-x)_+\right]$ and $(X-x)_+ \equiv \max\{X-x, 0\}$. By using (4.17), the remaining life span can be given

$$e_X(x) = b\frac{G_{2,3}^{3,0}\left(z\left|\begin{array}{cccc} 0 & 0 & 0 & \\ -1 & -1 & 0 & -1 \end{array}\right.\right)}{G_{1,2}^{2,0}\left(z\left|\begin{array}{ccc} 0 & 0 & \\ -1 & 0 & -1 \end{array}\right.\right)} \tag{4.21}$$

Figure 4-4 represents the remaining life span of a random variable X whereby $\mu = 75$ and $\sigma = 15$ (with $a = 81.751$ and $b = 11.695$)

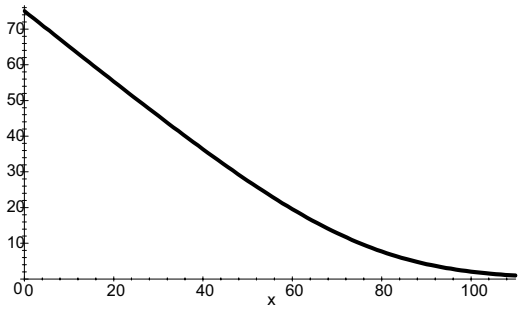


Figure 4-4: Remaining life expectancy of X, with $\mu = 75$ and $\sigma = 15$

The remaining standard deviation $(X-x)_+$ can be found by (4.17). Figure 4-5 represents the remaining standard deviation $\sqrt{Var\left[(X-x)_+\right]/(1-G_X(x))}$ whereby $\mu = 75$ and $\sigma = 15$.

Below two orders are defined which are extensively used in the ordering of mortality risks [Hee91].

Definition 30 *The life span of a random variable X precedes Y in mortality order, represented by $X \leq_m Y$, if $\mu_X(x) \geq \mu_Y(x)$ for all $x \geq 0$.*

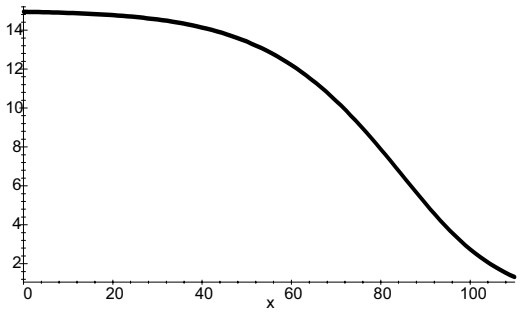


Figure 4-5: Remaining standard deviation of X, with $\mu = 75$ and $\sigma = 15$

Definition 31 *A random variable X precedes Y in average remaining life span order, represented by $X \leq_{mrl} Y$, if $e_X(x) \leq e_Y(x)$ for all $x \geq 0$.*

It can be readily seen that a random variable X with distribution $G(x \mid \mu_1, \sigma_1)$ in mortality order precedes random variable Y with $G(x \mid \mu_2, \sigma_2)$ if

$$\frac{x - a_1}{b_1} \geq \frac{x - a_2}{b_2}$$

for all $x \geq 0$. From this it follows that sufficient conditions are $b_1 a_2 \geq b_2 a_1$ and $b_2 \geq b_1$, which is equal to the constraints

$$\sigma_1 \mu_2 \geq \sigma_2 \mu_1 \text{ and } \sigma_2 \geq \sigma_1$$

The mortality order implies $X \leq_{mrl} Y$ which can be demonstrated by using (4.21).

Example 32 *A well-known method for approximating the mortality probability for women is to set back the mortality probability of men by a number of years. From the most recent AG life tables [Ag97], the following life expectancy and standard deviation can be derived*

Life expectancy μ and standard deviation σ

	μ	σ
Men	74.81	13.37
Women	80.73	13.00

Now take the following random variables (with the Gompertz distribution):

$X \sim G(x \mid 74.81, 13.37)$, $Y \sim G(x \mid 80.73, 13.37)$ *and* $Z \sim G(x \mid 80.73, 13.00)$. *They yield the following order*

$$X \leq_m Y$$

and

$$X \leq_m Z$$

However, it is impossible to order Y *with respect to* Z *with the given mortality and remaining life span order. The following stop-loss order is valid*

$$Z \leq_{(2)} Y$$

The example above shows that, when the mortality risk of females is modelled with a Gompertz distribution, it is riskier to use an age corrected male mortality risk (such that the life expectancy equals the life expectancy of females) than using life expectancy and variance of the female population. This follows from the observation that the variance of the female population is less than the male population (during the observed period) and because a correction of the age does not change the variance.

Example 33 *Assume that a heterogeneous population consists of* n *sub-populations, each with a density of* f_i *for* $i = 1, ..., n$. *Furthermore, assume that the mortality-intensity of each sub-population is derived from a random variable* X *with distribution* $G(x \mid \mu, \sigma)$ *and age correction* c_i. *The life expectancy of each sub-population at age 0 then is* $\mu + c_i$. *The standard deviation* σ *is constant. Then, the heterogeneous mortality-intensity* $\overline{\mu}_X(x)$ *is given by*

$$\overline{\mu}_X(x) = \frac{\sum_{1 \leq i \leq n} f_i \mu(x - c_i) G(x \mid \mu + c_i, \sigma)}{\sum_{1 \leq i \leq n} f_i G(x \mid \mu + c_i, \sigma)}$$

This yields the following mortality order for all $x \geq 0$

$$\overline{\mu}_X(x) \geq \mu_X(x - \overline{c}) \tag{4.22}$$

with average $\overline{c} = \sum_{1 \leq i \leq n} f_i c_i$. *This can be proven by using Jensen's inequality* $E[f(X)] \geq f(E[X])$ *with* $f(x)$ *a convex function.*

Chapter 5

Results

This thesis unifies mortality models proposed through the years. This unification is done with aid of a new parameter vitality and by relating the vitality of a population to the mortality intensity. The possibility arises to relate physiological variables to and incorporate them into vitality. The unifying model arranged as a sum model embeds the first and second law of mortality of Makeham, arranged as a multiplicative model frailty and Cox' regression models are embedded, and the third possible structure embeds Thiele's law of mortality.

Many models implicitly attempt to describe the effect of population heterogeneity on mortality. This observation has resulted in a general model (see section 2.5) which is based on the vitality concept. Physiological variables at an individual level can be used in the model. Along with the existing class of heterogeneous frailty models, which are used by Vaupel and Manton, and where the heterogeneous mortality-intensity is defined by

$$\overline{\mu}(x) = \overline{v}\mu(x)$$

a derived class of models is studied. Summing up the average vitality by

$$\overline{\mu}(x) = \mu(x) + \overline{v}$$

and into which Makeham's first and second laws of mortality can be integrated.

The observability of the new parameter v is discussed.

A second result is a new way of modelling Gompertz' law of mortality

$$S(x) = g^{c^x} \tag{5.1}$$

is endeavoured in this thesis. By introduction of the life expectancy μ and standard deviation σ, with

$$a = \mu + \sigma\gamma\frac{\sqrt{6}}{\pi} \text{ and } b = \frac{\sqrt{6}}{\pi}\sigma \tag{5.2}$$

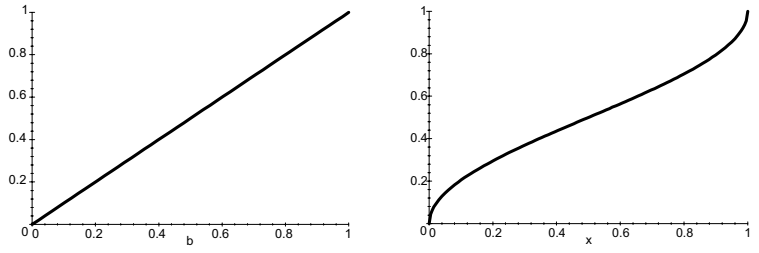
the mortality intensity

$$\mu(x) = \frac{1}{b}e^{\frac{x-a}{b}}$$

leads to the relation with the parameters used in Gompertz' law as (5.1) in $g = e^{-e^{-a/b}}$ and $c = e^{1/b}$.

A third result is a spin-off of the new way of modelling Gompertz' law of mortality. This led to a general model class with the possibility to describe explicitly the heterogeneity of a population. Multiple decrement survival functions form a sub-class of this general class. Also, the multiplicative hazard model can be seen as a sub-class and it is shown that the well known Cox's regression model with a log linear vitality function results in a heterogeneous population with sub-populations with life expectancy $\mu + b\boldsymbol{\beta}^T\mathbf{v}$.

A fourth result is a new demographic foundation for interpolation methods. It appears (see section 3.4.3) that the actuarial method about the average population during a year (which is used for calculating the mortality probability) is applicable (when all possible population permutations are taken into account). The discovered demographic foundation is named after Baxter's arcsin-law and demonstrates that the distribution of the population change during a year is not straight (Figure 5.3.a) but curved (Figure 5.3.b).



a) The uniform distribution b) The arcsin-law (5.3)

In the 'Applications' chapter, finally, instrumentation is developed to employ previously obtained theoretical results in this thesis. A first actuarial application is computational prediction of insurance data by explicitly modelling the increase of life expectancy as predicted by

the CBS (see section 4.1.4). This is achieved via theory of heterogeneous populations endowed with the vitality concept (see section 2.5.2).

A second actuarial application also builds upon previously obtained theoretical results in this thesis. Section 4.2 offers tools for accelerating computations in life insurance mathematics with help of computational intelligence. By using the obtained general model (2.5), it appears possible to automatically derive and efficiently calculate a large number of different life insurances. In order to create a 'germ' for this automatic derivation, a general function $G_x(t,r,T)$, defined by

$$G_x(t,r,T) = e^{rt} F_x(t)\, \psi_{T(x+t)}(r)$$

for the expected value of the present value of a basic life insurance is defined (section 4.2.1). Based on the reformulated law of Gompertz two insurance germ functions are derived: the Gompertz Insurance Germ $GIG_x(t)$ and the Makeham Insurance Germ $MIG_x(t)$. These germs offer an easy computational formulation of the general function $G_x(t)$ for both laws of mortality. From the general function, other life insurances (for example deferred, increasing and Universal Life insurances) are derived which result in a representation of insurance knowledge in the Actuarial Knowledge System (AKS).

Finally, the new representation provides two new actuarial applications: Easy calculation of variances of life insurances and solutions for the stop-loss order of the n-th degree in terms of Meijer's G-functions (section 4.2.4). For a random variable X with the Gompertz distribution with parameters μ and σ, the stop-loss order of the n-th degree is defined

$$E\left[\left((X-x)_+\right)^n\right] = n! z b^n G^{n+2,0}_{n+1,n+2}\left(z \left| \begin{array}{ccccc} 0 & ... & 0 & 0 & \\ -1 & ... & -1 & 0 & -1 \end{array} \right.\right)$$

with (5.2) and $z = e^{(x-a)/b}$.

In addition, the stop-loss transformation is used for ordering mortality risks and it is shown how remaining life expectancy and variance are calculated with Meijer's G-functions and enable comparison of mortality distributions with regard to their risk.

Chapter 6

Conclusions and discussion

History has shown that mortality models have gradually been improved while the modelling strategy has gradually been modified. From purely data-oriented methods, the switch was made to the use of theories or mortality laws with which the observed mortality was described. Today one can still discern this knowledge-oriented approach in the contemporary mortality models. A theory is indispensable for the description of mortality, but often inadequate for describing non-homogeneous populations.

Reformulation of Gompertz' law of mortality allowed the law to be easily used within the theory of heterogeneous populations. With the new explanatory model for heterogeneous populations (2.5.2), which makes segmentation into subpopulations possible by using a continuous vitality parameter, individual physiological properties can be summarised. Existing mortality laws (2.2) and multiple decrement survival functions (2.3) can be integrated into this general model for mortality (2.5).

An important question is how specific mortality figures can be used by pension funds and insurance companies for predicting the reserve development. The improvement of the interpolation methods (see section 3.4.3 and repeated in the results) has resulted in the possibility of determining survival functions for subpopulations. By using CBS data on incomes [Cbs95] (from which it appears that a higher income leads to a higher life expectancy and a lower variance), income-dependent mortality can be modelled with the method presented. This allows realistic forecasts to be made of the development of an insured portfolio.

Moreover, the general model presented generates possibilities for applying techniques such as neural networks. Neural networks are ideally appropriate for discovering causal relations between individual characteristics and summarising them in (for example) a 'vitality' factor.

Applications of preceding theoretical innovations are mortality forecasts for heterogeneous populations (4.1). Modelling the future increase of life expectancy deserves recommendation. The actuary is obliged to build up sufficient reserves in order to be able to cover future payments. The increase in life expectancy is a trend which will continue and which will lead to people making use of their pensions for a longer period than in the past. This development will therefore have to be taken into account when determining the reserves. For the prediction of mortality based on an estimated vitality, a time parameter is introduced in (3.6). By improving the forecast, it is possible to make a realistic forecast of the longevity risk. This is important for the pension funds, who, in the present situation, are confronted with a life table with a higher life expectancy every 5 years. The method described puts pension funds in a better position to determine the longevity risk.

There is a number of differences between life and non-life insurances. Within life actuarial science, the variable that is used to model the underlying insurance risk is usually the expected life span. Within non-life actuarial science, the financial risk itself (the level of the claim) is seen as the variable. The practice of life actuarial science differs sharply from the practice of risk theory and (a last point) determination of the premium differs as well.

Because knowledge of mortality models has improved (for example by identifying the parameters of Gompertz' law of mortality as moments of a probability distribution) tools can be designed for simplifying life insurance mathematics (4.2). Because of this, it becomes possible to bridge the differences between life and non-life methods.

The new explanatory modelling with the aid of the vitality parameter suffers from the 'Garden of Eden' modelling syndrome: this syndrome arises if it is not clear, how the original measurements leading to it must be performed.

Secondly, Makeham's first and second mortality laws and (partially) Thiele's mortality law have successfully been integrated into one general model, but it was not possible in any simple way to integrate Gompertz' mortality law in a general model as well. Therefore, the basic assumption remains the exponentially rising mortality-intensity.

One problem with forecasts is that they are always based on extrapolated survival functions from the past. Life expectancy is highly dependent on social factors, which are difficult to extrapolate. A qualitative approximation of mortality forecast, which would have to be able to separate the period-effect and the cohort-effect, is vital for a good forecast.

Appendix: the Gamma function and Meijer's G-function

The complete Gamma function is an extension of the factorial to complex and real values. The function is analytical everywhere, except at $z = 0, -1, -2, ...$ The function can be defined as definition integral for $\mathrm{Re}\,(z) > 0$ with

$$\Gamma(z) \equiv \int_0^\infty t^{z-1} e^{-t} dt \tag{6.1}$$

or

$$\Gamma(z) \equiv \int_0^1 \left[\ln\left(\frac{1}{t}\right)\right]^{z-1} dt$$

By partially integrating (6.1), it can be proven that

$$\Gamma(x) = (x-1)\,\Gamma(x-1)$$

For the entire value $n = 1, 2, 3, ...$ one can prove that

$$\Gamma(n) = (n-1)!$$

Binet's formula is

$$\ln\Gamma(a) = \left(a - \frac{1}{2}\right)\ln a - a + \frac{1}{2}\ln(2\pi) + 2\int_0^\infty \frac{\tan\left(\frac{z}{a}\right)}{e^{2\pi z} - 1} dz$$

for $\mathrm{Re}\,(a) > 0$. The Gamma function can also be written as an infinite product

$$\Gamma(z) \equiv \left[z e^{\gamma z} \prod_{r \geq 1} \left(1 + \frac{z}{r}\right) e^{-\frac{z}{r}} \right]^{-1}$$

with the Euler - Mascheroni constant γ defined by

$$\gamma = \lim_{n \to \infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln n \right) \approx 0.5772156649$$

Differentiating yields

$$-\frac{\Gamma'(z)}{\Gamma(z)} = \frac{1}{z} + \gamma + \sum_{n \geq 1} \left(\frac{1}{n+z} - \frac{1}{n} \right)$$

and

$$\Gamma'(z) = \Gamma(z)\,\Psi(z) = \Gamma(z)\,\psi_0(z)$$

with $\Psi(z)$ the Digamma function and $\psi_0(z)$ the Polygamma function. The n-th derivatives can be expressed as terms of the Polygamma functions ψ_n, $\psi_{n-1}, ..., \psi_0$. Furthermore,

$$\Gamma'(n) = -(n-1)! \left(\frac{1}{n} + \gamma - \sum_{1 \leq k \leq n} \frac{1}{k} \right)$$

The Lanczos approximation for $z > 0$ is

$$\Gamma(z+1) = \left(z + \gamma + \frac{1}{2}\right)^{z+\frac{1}{2}} e^{z+\gamma+\frac{1}{2}} \sqrt{2\pi} \times \left[c_0 + \frac{c_1}{z+1} + \frac{c_2}{z+2} + ... + \frac{c_n}{z+n} + \epsilon \right]$$

The Gauss multiplication formula is useful for the multiplication of Gamma functions

$$\Gamma(z)\,\Gamma\left(z + \frac{1}{n}\right) ... \Gamma\left(z + \frac{n+1}{n}\right) = (2\pi)^{\frac{n-1}{2}}\, n^{\frac{1}{2} - nz}\, \Gamma(nz)$$

The (above) incomplete Gamma function is defined by

$$\Gamma(a, x) = \int_x^{\infty} t^{a-1} e^{-t} dt$$

For an entire value $n = 1, 2, 3, ...$, it can be stated that

$$\Gamma(n, x) = (n-1)! e^{-x} \sum_{0 \le s < n} \frac{x^s}{s!}$$

A very general function is Meijer's G-function (see [MS73]) defined by

$$G_{p,q}^{m,n}\left(z \mid \begin{matrix} a_1, ..., a_p \\ b_1, ..., b_q \end{matrix}\right) \equiv \frac{1}{2\pi i} \int_{\gamma_L} \frac{\prod_{1 \le j \le m} \Gamma(b_j + z) \prod_{1 \le j \le n} \Gamma(1 - a_j - z)}{\prod_{m+1 \le j \le q} \Gamma(1 - b_j - z) \prod_{n+1 \le j \le p} \Gamma(a_j + z)} z^{-s} ds$$

where $\Gamma(z)$ is the Gamma function, $i = \sqrt{-1}$, $z \neq 0$ and

$$z^s = e^{s(\ln|z| + i \arg z)}$$

in which $\ln|z|$ denotes the natural logarithm of $|z|$ and $\arg z$ is not necessarily the principal value. Furthermore, m, n, p and q are integers with $0 \le n \le p$ and $0 \le m \le q$. Parameters a_k, $k = 1, ..., p$ and b_j, $j = 1, ..., p$ are complex numbers such that

$$a_k - b_j \neq 0, 1, 2, 3, ...$$

for $j = 1, ..., n$ and $h = 1, ..., m$. The contour γ_L lies between the poles of $\Gamma(1 - a_i - z)$ and the poles of $\Gamma(b_i + 1)$.

This function has applications in numerous fields such as statistical distributions, functional equations, theoretical physics and hydrodynamics. Special cases include

$$\begin{aligned} G_{1,2}^{2,1}\left(z \left| \begin{matrix} 1, 1 \\ 1, 0 \end{matrix}\right.\right) &= \ln(z+1) \\ G_{1,0}^{0,2}\left(\frac{1}{2}z \left| \begin{matrix} \\ 0, \frac{1}{2} \end{matrix}\right.\right) &= \frac{\cos\left(\sqrt{2z}\right)}{\sqrt{\pi}} \\ G_{0,1}^{1,0}(z \mid 1 - a) &= e^{-\frac{1}{z}} z^{-a} \end{aligned}$$

References

[AH93] Aarssen, K., Haan, L. de, On the Maximal Life Span of humans, TI 93, Nr. 116, Tinbergen Institute, Amsterdam, 1993

[Acz69] Aczel, J., On Applications and theory of Functional Equations, Birkhäuser Verlag, Basel/Stuttgart, 1969

[Ag92] Actuarieel genootschap, AG-tafels 1985-90, Amsterdam, 1992

[Ag97] Actuarieel genootschap, AG-tafels 1990-95, Amsterdam, 1997

[Ast90] ASTIN Nederland, Het KAZO-Tarief 1990, AOV-Individueel, ASTIN Nederland, 1990

[BM93] Bedford, T., Meilijson, I., A characterisation of marginal distributions of (possibly dependent) lifetime variables which right censor each other, Reports of the Faculty of Technical Mathematics and Informatics, Vol. 93, Nr. 116, Delft, 1993

[BM96] Bedford, T., Meilijson, I., A new approach to censored lifetime variables, Reliability Engineering and System Safety, Vol. 51, pp. 181-187, 1996

[BP86] Benjamin, B., Pollard, J.H, The analysis of mortality and other actuarial statistics, Heinemann, London, 1980

[BGH86] Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., Nesbitt, C.J., Actuarial mathematics, Society of Actuaries, 1986

[Bun67a] Bunge, M., Scientific research I, the search for system, Springer-Verlag, Berlin, 1967

[Bun67b] Bunge, M., Scientific research II, the search for truth, Springer-Verlag, Berlin, 1967

[Car50] Carnap, R., Logical foundations of probability, University of Chicago Press, Chicago, 1950

[Cbs95] CBS-rapport, Sterfte, doodsoorzaak en inkomen, Voorburg, Heerlen, 1995

[Cbs97a] CBS-rapport, Bevolkingsprognose 1996-2050, Mndstat Bevolking Vol. 97, Nr. 1, pp. 62-76, 1997

[Cbs97b] CBS-rapport, Sterfte: trends, achtergronden en prognose, Mndstat Bevolking Vol. 97, Nr. 5, pp. 10-16, 1997

[Cbs97c] CBS-rapport, Overledenen naar doodsoorzaak 1995, Series A1, Voorburg, Heerlen, 1997

[Chi68] Chiang, C.L., Introduction to stochastic processes in biostatistics, Wiley, New York, 1968

[Coo93] Cooke, R.M., The total time on test statistic and age-dependent censoring, Statistics and Probability Letters, Vol. 18, pp. 307-312, 1993

[Cox72] Cox, D.R., Regression models and life tables (with discussion), Journal of the Statistical Society, Vol. 34, Series B, pp. 506-527, 1972

[CO84] Cox, D.R., Oakes, D., Analysis of Survival Data, Chapman and Hall, Cambridge, 1984

[Cro94] Crowder, M., Identifiability crisis in competing risks, International Statistical Review, Vol. 62, Nr. 3, pp. 379-391, 1994

[Daw80] Daw, R. H., Johann Heinrich Lambert, Journal of the Institute of Actuaries, Vol. 107, pp. 345-363, 1980

[DLS49] Dublin, L.I., Lotka, A.J., Spiegelman, M., Length of life, a study of the life table, Ronald Press Company, New York, 1949

[Fra93] Franses, P.H., The Gompertz Curve: Estimation and Model Selection, TI 93, Nr. 74, Tinbergen Institute, Amsterdam, 1993

[Ga75] Gail, M., A review and critique of some models used in competing risk analysis, Biometrics, Vol 31, pp. 209-222, 1975

[GK78] Gans, H. de, Korzec, M., Mathematische demografie, Wolters-Noordhoff, Groningen, 1978

[Ges97] Geskus, R.B., Estimation of smooth functionals with interval censored data, and something completely different, Thesis Delft University of Technology, Delft, 1997

[Gom1825] Gompertz, B., On the nature of the function expressive of the Law of Human Mortality and on a new mode of determining the Value of Life Contingencies, Philosophical Transactions Royal Society 36, pp. 513-585, W. Nicol, London, 1825

[Gra1662] Graunt, J., Natural and political observations mentioned in a following index, and made upon the Bills of Mortality, London 1662, republished in: Journal of the Institute of Actuaries, Vol. 90, pp. 1-61, 1964

[Gre28] Greenwood, M., Laws of Mortality from the biological Point of View, Journal of Hygiene, Vol. 28, pp. 267-294, 1928

[Hab88] Haberman, S., Measuring relative mortality experience, Journal of the Institute of Actuaries, Vol. 115, Nr. 2, pp. 271-298, 1988

[Hal1693] Halley, E., An estimate of the degrees of the mortality of mankind, Philosophical Transactions Nr. 17, pp. 596-610, pp. 653-656, 1693

[Hee91] Heerwaarden, A.E. Van, Ordering of actuarial risks - theory and actuarial applications, thesis, Universiteit van Amsterdam, Amsterdam, 1991

[HP80] Heligman, L., Pollard, J.H., The age pattern of mortality, Journal of the Institute of actuaries, Vol. 107, pp. 49-80, 1980

[Hen96] Henskens, M.M.C., Kennissysteem voor de actuaris, Pensioen Advies, Vol. 96, Nr. 3, pp. 28-32, 1996

[Hic87] Hickman, J.C., Connections between graduation and forecasting, Advanced in the statistical scenes, Vol. VI, Actuarial science edited by MacNeill, I.B., Umphrey, G.J., D. Reidel Publishing Company, Dordrecht, pp. 1-17, 1987

[Hor96] Hordijk, M., SHEECS voor een Actuarieel Kennissysteem, doctoraalscriptie, TU Delft, 1996

[Hou84] Hougaard, P., Life table methods for heterogeneous populations: Distributions describing the heterogeneity, Biometrika, Vol. 71, Nr. 1, pp. 75-83, 1984

[Ina89] Inaba, H., Functional analytic approach to age-structured population dynamics, Thesis, University of Leiden, Leiden, 1989

[Jab69] Jacobs, Das kombinatorische Äquivalenzprinzip und das arcsin-Gesetz von E. Sparre Andersen, Selecte Mathematica, Vol. 1, 1969

[JS95] Janssen, J., Skiadas, C.H., Life table data dynamic modelling, Applied stochastic models and data analysis, Vol. 11, pp. 35-50, 1995

[Keu94] Keuzenkamp, H.A., Probability, econometrics and inference: a treatise on the foundations of economic inference, thesis, Katholieke Universiteit Brabant, Tilburg, 1994

[Key81] Keyfitz, N., Choice of function for mortality analysis; effective forecasting depends on a minimum parameter representation, In: methodologies for the collection and analysis of mortality data. Proceedings of a seminar at Dakar, Senegal by the committee on factors affecting mortality and the length of life, Vallin, J., Pollard, J.H., Heligman, L. (Editors), pp. 225-243, 1981

[KG96] Kaas, R., Goovaerts, M.J., Inleiding risicotheorie, Instituut voor Actuariaat en Econometrie, Universiteit van Amsterdam, Amsterdam, 1996

[Kli93] Kling, B., Life insurance, a non-life approach, thesis, Tinbergen Instituut, Amsterdam, 1993

[Kun98] Kunimura, D., The Gompertz distribution-estimation of parameters, Actuarial Research Clearing House, Society of Actuaries, Illinois, 1998

[Lan90] Lancaster, H.O., Expectations of life: a study in the demography, statistics and history of world mortality, Springer, New York, 1990

[LPQ81] Landberg, N., Proschan, F., Quinzi, A.J., Estimating dependent life lengths, with applications to the theory of competing risks, The annals of statistics, Vol. 9, Nr. 1, pp. 157-167, 1981

[Lee92] Lee, E.T., Statistical methods for survival data analysis, Second edition, Wiley, New York, 1992

[Mac78] Maciejowski, J.M., Lecture Notes in Control and Information Science : The Modelling of Systems with Small Observation Sets, Vol. 10, Springer-Verlag, Berlin, 1978

[Mak1867] Makeham, W.M., On the Law of Mortality, Journal of the Institute of Actuaries, Vol. 13, pp. 335-340, 1867

[MAL77] Mode, C.J., Avery, R.C., Littman, G.S., Potter, R.G. Jr., Methodological issues underlying multiple decrement life table analysis, Demography, Vol. 14, Nr. 1, 1977

[MS84] Manton, K.G., Stallard, E., Recent trends in mortality analysis, Center for Demographic Studies, Duke University, Durham, North Carolina, Academic Press, London, 1984

[MSC97] Manton, K.G., Stallard, E., Corder, L., Changes in the age dependence of mortality and disability: cohort and other determinants, Demography, Vol. 34, Nr. 1, 1997

[MSV81] Manton, K.G., Stallard, E., Vaupel, J.W., Methods for comparing the mortality experience of heterogeneous populations, Demography, Vol. 18, Nr. 3, 1981

[MSV86] Manton, K.G., Stallard, E., Vaupel, J.W., Alternative models for the heterogeneity of mortality risks among the aged, Journal of the American Statistical Association, Vol. 8, Nr. 395, 1986

[MS73] Mathai, A.M., Saxena, R.K., Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences, Springer-Verlag, Berlin, 1973

[Mil77] Miller, D., A note on independence of multivariate lifetimes in competing risks models, The annals of Statistics, Vol. 5, Nr. 3, pp. 576-579, 1977

[Mis39] Mises, R. von, Kleines lehrbuch des positivismus, W.P. van Stockum & Zoon, Holland, 1939

[Nag61] Nagel, E., The structure of science, Routledge & Kegan Paul Ltd., London, 1961

[Nam87] Namboodiri, K., Suchindran, C.M., Life Table techniques and their applications, Academic Press, London, 1987

[OC97] Olshansky, S.J., Carnes, B.A., Ever since Gompertz, Demography, Vol. 34, Nr. 1, 1997

[Ode96] Oden, J.T., Demkowicz, L.F, Applied Functional Analysis, CRC Press, 1996

[Pap65] Papoulis, A., Probability, random variables, and stochastic processes, McGraw-Hill, New York, 1965

[Pet75] Peterson, A.V., Bounds for a joint distribution function with fixed sub-distribution functions: application to competing risks, Proc. Nat. Acad. Sci. USA, Vol. 73, Nr. 1, pp. 11-13, 1975

[Pet77] Peterson, A.V., Expressing the Kaplan-Meier Estimator as a function of empirical subsurvival functions, Journal of the American Statistical Association, Vol. 72, Nr. 360, 1977

[Pop59] Popper, K.R., The logic of scientific discovery, Basic Books, New York, 1959

[PK78] Prentice, R.L., Kalbfleisch, J.D., Regression analysis of grouped survival data with application to breast cancer, Biometrics, Vol. 34, Nr. 1, pp. 25-39, 1978

[Pre72] Pressat, R., Demographic analysis, First edition, Aldine Atherton, Chicago, 1972

[PKS72] Preston, S.H., Keyfitz, N., Schoen, R., Causes of death: Life tables for national populations, Seminar Press, New York and London, 1972

[Pro98] Promislow, D.E.L., Longevity and the barren aristocrat, Nature, Vol. 396, 1998

[Ren91] Renshaw, A.E., Actuarial graduation practice and generalised linear and non-linear models, Journal of the Institute of Actuaries, Vol 118, Nr. 1, pp. 295-312, 1991

[Ren88] Renshaw, A.E., Modelling excess mortality using GLIM, Journal of the Institute of Actuaries, Vol 115, Nr. 2, pp. 299-315, 1988

[RT97] Reiss, R.D., Thomas, M., Statistical Analysis of Extreme Values, Birkhäuser, Germany, 1997

[Szc99] Szczepaniak, P.S., Computational intelligence and application, Physica-Verlag, Germany, 1999

[Shi84] Shiryayev, A.N., Probability, Graduate Texts in Mathematics, Springer-Verlag, Berlin, 1984

[SK77] Smith, D., Keyfitz, N., Mathematical Demography, Selected papers, Springer-Verlag, Berlin, 1977

[Ste62] Stevens, S.S., On the theory of scales of measurement, Philosophy of science, edited by Danto, A., Morgenbesser, S., The World Publishing Company, Ohio, 1962

[Ste79] Stegmüller, W., The structuralist view of theories, Springer-Verlag, New York, 1979

[TPW94] Tabeau, E.W.A., Poppel. F. van, Willekens, F., Mortality in the Netherlands; the database, NIDI-report 36, Den Haag, 1994

[Tsi74] Tsiatis, A., A nonidentifiability aspect of the problem of competing risks, Proc. Nat. Acad. Sci. USA, Vol. 72, Nr. 1, pp. 20-22, 1975

[Tur90] Turney, P., The curve fitting problem: a solution, Brit. J. Phil. Sci., Vol. 41, pp. 509-530, 1990

[VMS79] Vaupel, J.W., Manton, K., Stallard, E., The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography 16, pp. 439-454, 1979

[VY83] Vaupel, J.W., Yashin, A.I., The deviant dynamics of death in heterogeneous populations, Research Report of the International Institute for Applied Systems Analysis, Vol 83, Nr. 1, 1983

[Wil97] Willemse, M., Bibliografie van sterftetafels, not published, Practis, 1997

[Wil95] Willemse, W.J., Actuariële kennisrepresentatie, doctoraalscriptie, TU Delft, 1995

[WK00a] Willemse, W.J., Koppelaar, H., Knowledge elicitation of Gompertz's law of mortality, Scandinavian Actuarial Journal, Nr. 2, pp. 168-179, 2000

[WK00b] Willemse, W.J., Koppelaar, H., Computational intelligence in life insurances, to be published

[Wol94] Wolthuis, H., Life insurance mathematics (the Markovian model), Caire Education Series 2, CAIRE, Brussels, 1994

[Wol96] Wolthuis, H., Levensverzekeringswiskunde Deel 1, Instituut voor Actuariaat en Econometrie, Amsterdam, 1996

[YI97] Yashin, A.I., Iachine, I.A., How frailty models can be used for evaluating longevity limits: taking advantage of an interdisciplinary approach, Demography, Vol. 34, Nr. 1, 1997

[Zad56] Zadeh, L.A., On the identification problem, IRE Transactions on Circuit Theory, Vol. 3, pp. 277-281, 1956

Samenvatting

'Computational Intelligence: Sterftemodellen voor de actuaris', proefschrift van W.J. Willemse

In dit proefschrift wordt de wisselwerking tussen sterftemodel en sterftedata langs twee hoofdlijnen uitgewerkt. Ten eerste de al of niet onderliggende modelbouw. De tweede hoofdlijn is de meetmethode en de al of niet expliciete data bewerkingsmethode. De strategie om verklarende modellen te bouwen is in 1935 bij gebrek aan resultaat gestaakt (2.2).

In (2.4) wordt aangetoond dat de discussie weer levendig zou kunnen worden door een herformulering van de bekende sterftewet van Gompertz zodat de wet eenvoudig gebruikt kan worden in de theorie van heterogene populaties. Aangetoond wordt dat de parameters in de sterftewet van Gompertz in feite de levensverwachting en variantie van de onderliggende populatie zijn. Een nieuw verklarend model voor heterogene populaties wordt dan in (2.5.2) ingevoerd. De verklaring bestaat uit een continue parameter vitaliteit die segmentatie mogelijk maakt in homogene subpopulaties. De parameter vitaliteit maakt het mogelijk om individuele fysiologische eigenschappen samen te vatten en te laten resulteren in een sterfteprognose (4.1). Het bleek mogelijk de bestaande sterftewetten (2.2) en meervoudig dalende overlevingsfuncties (2.3) in te passen in een algemener model (2.5) voor sterfte.

De bepaling van sterftekansen op metrische wijze leidde tot het nieuwe inzicht dat de gebruikelijke lineaire interpolatiemethoden voor gegevens onjuist zijn (3.4). Omdat de voorspelling van sterfte op grond van een geschatte vitaliteit een belangrijke is wordt in (3.6) een tijdsparameter ingevoerd.

De toepassing van voorgaande theoretische vernieuwingen zijn sterfteprognoses voor heterogene populaties (4.1), gereedschap voor versnelling van berekeningen in de levensverzekeringswiskunde (4.2) en ten slotte een methode om stop-loss transformaties van de sterftewet van Gompertz te berekenen.

Summary

'Computational Intelligence: Mortality models for the actuary', PhD-thesis by W.J. Willemse

In this thesis the interaction between mortality model and mortality data is elaborated along two main lines. Firstly, the (possibly) underlying modelling. The second main line is the measurement method and the explicit (or implicit) data processing method. The strategy of building explanatory models came to a halt in 1935 due to lack of results (2.2).

In (2.4) it is demonstrated that the discussion could be revived by reformulating the well-known Gompertz' law of mortality so that the law can be easily used in the theory of heterogeneous populations. It is demonstrated that the parameters in Gompertz' law of mortality are in fact the life expectancy and standard deviation of the underlying population. A new explanatory model for heterogeneous populations is then introduced in (2.5.2). The explanation is a continuous vitality parameter which makes segmentation possible in homogeneous subpopulations. The vitality parameter makes it possible to summarise individual physiological properties and to have them result in a mortality prognosis (4.1). It appears to be possible to integrate the existing mortality laws (2.2) and multiple decrement survival functions (2.3) into a general model (2.5) for mortality.

The metric determination of mortality probabilities led to the new insight that the usual linear interpolation methods for data are inaccurate (3.4). It is because the prediction of mortality on the basis of an estimated vitality is important in (3.6) that a time parameter is introduced.

The application of previous theoretical innovations are mortality forecasts for heterogeneous populations (4.1), tools for accelerating calculations in life insurance mathematics (4.2) and finally a method to calculate stop-loss transformations of Gompertz' law of mortality.

Curriculum Vitae

Willem Jan Willemse was born in Leidschendam on 8 March 1973. He studied Technical Informatics in Delft from 1991 to 1995, graduating cum laude on the subject 'Actuarial knowledge technology', a study performed within the engineering group of Delft Univerity of Technology. For this he received the prize of the Delft University Fund for best graduating student.

After that he started work as a junior consultant at Practis B.V., a company which does consulting and provides services for insurance companies, pension funds and social security institutions. In October 1996 he began a research project at the Delft University of Technology sponsored by Practis and the Ministry of Economic Affairs. It was this project that led to this thesis.

He currently works at Practis as a consultant in the field of system design and implementation of systems for life and disability insurances.