Università degli Studi di Padova

Dipartimento di Scienze Statistiche
CORSO DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXIX

A MIXTURE MODEL TO DISTINGUISH MORTALITY COMPONENTS

Coordinatore del Corso: Prof. Monica Chiogna

Supervisore: Prof. Stefano Mazzuco

Co-supervisore: Prof. Vladimir Canudas-Romo

Dottoranda: Lucia Zanotto

Padua, 11–15, 2016

If you haven't measured something,
you really don't know very much about it.

Karl Pearson

Abstract

We propose a new parametric model for the life table distribution of deaths. The model is a mixture of three distributions: for infant and child mortality, for accidental and premature mortality, and for adult mortality. This permits to study the characteristics of the different components analyzing the changes of the shape of their distributions. In particular we focused on the features of accidental and premature mortality, since this component, for several reasons, is not often taken into account. The model offers the advantage of computing, in explicit form, the three component's contributions to life expectancy. This is useful to understand the transformations of the distribution of deaths. Moreover, all parameters in the model have a demographic interpretation, so it is easy to study trends and compare different populations. The mixture model was tested using Swedish raw data from the Human Mortality Database and then fitted for different countries. Our results about infant, childhood and adult mortality confirm what we generally know about the tendencies of these components: decrease of deaths at age 0, disappearance of childhood mortality, compression and shifting of adult distribution. For what regards premature mortality, we observed that, over time, its function becomes flatter and more symmetric, and its mode shifts progressively. This implies that almost entirely the accidental mortality has disappeared, while the premature mortality continuous to exist. We also noted that its contribution to explain life expectancy decreased during the last century, but in recent years, for some countries, especially France, it started to grow. This means that the shape of the distribution of deaths is changed: after a compression around the adult mode, an increase of the number of deaths outside the adult distribution is registered.

The model does not perform satisfactory estimates when the interval of the last open age class is too wide, in particular when the late mode of the distribution is not clearly recognizable in the death distribution. To improve the estimates we employed EM algorithm, which proved to be a good method to solve this problem. EM algorithm can also be applied to reconstruct incomplete birth cohort data. To compute confidence intervals bootstrap resampling was adopted.

Sommario

In questo studio proponiamo un nuovo modello parametrico con cui analizzare la distribuzione dei decessi per età e le sue componenti. Poiché i valori dei decessi della tavola di mortalità possono essere considerati una funzione di probabilità, il modello è composto da una mistura di distribuzioni: una per la mortalità infantile, una per la mortalità accidentale e prematura ed un'altra per la mortalità adulta. L'analisi dei parametri ci ha permesso di studiare le caratteristiche delle singole componenti e la loro trasformazione nel tempo. In particolare ci siamo focalizzati sull'approfondimento della mortalità accidentale e prematura, che spesso, per diverse ragioni, non viene affrontato. Il principale vantaggio di questo metodo è la possibilità di quantificare, in forma esplicita, il contributo di ciascuna componente nel calcolo della speranza di vita alla nascita. Inoltre, tutti i parametri possono essere interpretati da un punto di vista demografico, facilitando l'analisi dei risultati ottenuti e il confronto sia tra anni che tra nazioni. Il modello è stato prima testato sui dati grezzi della Svezia dello Human Mortality Database e poi stimato per differenti paesi. I risultati ottenuti riguardanti la mortalità infantile ed adulta sono in linea con quanto già sappiamo da precedenti studi: riduzione dei decessi in età 0, scomparsa della mortalità durante l'età infantile, compressione e spostamento della distribuzione della mortalità adulta. Invece, per quanto riguarda la mortalità accidentale e prematura, abbiamo notato che la sua distribuzione diventa sempre più piatta e simmetrica, e la sua moda si sposta progressivamente verso destra. Questo suggerisce che nella maggior parte dei paesi considerati, la mortalità accidentale sia andata progressivamente scomparendo. Tale considerazione non vale per la mortalità prematura che, in alcuni casi, sembra addirittura in aumento. Infatti, per un'accurata stima della speranza di vita alla nascita, il suo contributo risulta fondamentale. Queste conclusioni sembrano essere confermate da recenti studi sulla mortalità: dopo aver osservato una maggiore concentrazione dei decessi attorno all'età modale ed al loro progressivo spostamento verso le età più avanzate, stiamo assistendo ad un aumento del numero di decessi che si verificano al di fuori della distribuzione della mortalità adulta.

Abbiamo notato che le stime del modello non risultano soddisfacenti quando l'ultima classe di età coinvolge un vasto intervallo e la moda della distribuzione non è chiaramente visibile. Per migliorare le stime abbiamo utilizzato l'algoritmo EM, che si è dimostrato essere una valida soluzione. Questa metodologia può anche essere impiegata per ricostruire i decessi di una coorte di nati non ancora arrivata ad esaurimento. Gli intervalli di confidenza sono stati calcolati mediante il ricampionamento bootrap.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Professor Stefano Mazzucco who helped me in all the time of my research, even when I was kilometers away from home, who reassures me and remembers me that I am a "Doctor" when I look on the dark side of things. I would like to thank my co-supervisor Vladimir Canudas Romo who goaded me (and he does still now) to face the research with dedication and passion with his irrepressible enthusiasm and his thousand ideas. I would also to thank his family which welcomed me and made me feel less foreign. I would like to thank Professor Bruno Scarpa, the first one who believed in me and he always tells me that I can do everything I want. I would like to express my sincere gratitude to Professor A. Azzalini who answered all my questions patiently making me become less ignorant. My sincere thanks also goes to Professor R. Rau e Professor S. Drefahl who read carefully my work and gave me precious comments which incentivized me to widen my research from various perspectives. I thank Claudia who from colleague has become a precious friend who supported me throughout my years of study with her affection and sincerity; I can relay on her. I would also to thank Khanh for his "Luci work!" and "Luci, don't worry"; without him it would not be the same. I thank my dear friend Francesca who gives me hope with her kindness and simpleness. My sincere thanks also goes to Adam, Maarten, Virginia, Julia, Marie-Pier, Marius, Anthony, Silvia, Jonas and Catalina for their support (for many memorable evenings out and in) and help: they made me feel special. Most importantly, none of this could have happened without my parents who glimpsed of my way and encouraged me to follow it. I thank my adorable sister who loves me and always supports me. I thank Giordano with all my heart: you, more than all, make me feel happy. I am lucky to have you.

Table of Contents

Abstract iii

Riassunto v

Acknowledgments vii

Table of Contents vii

List of Figures xi

List of Tables xiii

1 Introduction 1

1.0.1 Background . . . 5

1.1 Life Table . . . 7

1.2 Mortality components . . . 9

1.3 Main models to study mortality . . . 11

2 Model and method 15

2.1 Definition of the mixture model . . . 16

2.1.1 Estimation of the model parameters . . . 22

2.1.2 Demographic interpretation of the parameters . . . 23

2.1.3 Advantages of the model . . . 25

2.2 A simplification of the mixture model . . . 27

3 Results 31

3.1 Model testing . . . 31

3.2 Infant Mortality . . . 36

3.3 Adult Mortality . . . 38

3.4 Accidental and Premature mortality . . . 41

3.5 Premature mortality and educational levels . . . 46

4 EM algorithm 51

4.1 Definition of EM algorithm . . . 51

4.1.1 Bootstrap confidence intervals . . . 52

4.2 Results . . . 53

4.3 Cohort data . . . 55

5 Conclusion **57**

A Model for Sweden in 1910 - 2010 **59**

B Mortality and educational level for women **61**

C EM estimates for USA in 1980 **63**

D Functions implemented in R **65**

BIBLIOGRAPHY **69**

List of Figures

1.1 Distribution of death by age for Sweden in 1751. 2
1.2 Distribution of deaths by age for Sweden in 1935 and in 2010 (source: Human Mortality Database). 3
1.3 Logarithm of male mortality rate of different education levels (source: ISTAT). . . 3
1.4 Survival curve for different year in Sweden (source: Human Mortality Database). . 6
1.5 Data for period death rates. 8
1.6 Data for cohort death rates. 9
1.7 Lexis' components for the distribution of death by age (Source: Lexis, 1879). . . . 10
1.8 Pearson's five components for the distribution of death by age (Source: Pearson, 1897). 10

2.1 Probability density function of the Half Normal distribution as the scale parameter changes. 17
2.2 Examined functions to model infant and childhood mortality. 17
2.3 Comparison between the values estimated for infant mortality with Half Normal and Half Cauchy distributions. 19
2.4 Estimated Cauchy distribution for infant mortality for 2000 in Sweden. 19
2.5 Probability density function of the Skew Normal distribution as the shape parameter changes (with ξ and ω are fixed and equal to 0 and 1, respectively). 21
2.6 Errors committed by the model without the accidental and premature distribution. 21
2.7 Probability density function of the mixture model considering different values for η and α, respectively. 22
2.8 Maximum likelihood estimates for the mixture model. 23
2.9 The different areas that compose the death distribution. 26
2.10 Probability density function of the Skew Bimodal Normal distribution as the parameter α^* changes. 28
2.11 Model fitted for France 1977 and Russia 2006. 29
2.12 Relation between α^* and λ^* in the SBN distribution. 30

3.1 Model fit for Sweden in 1935, 1973 and 2010. 32
3.2 Boxplot of the distribution of the errors committed by the mixture model. 32
3.3 Models fit for Sweden in 1935, 1973 and 2010. 33
3.4 AIC values for Siler, Heligman and Pollard and Mixture model from 1910 to 2911. 34
3.5 Errors committed by the mixture model for the examined countries in different years. 35
3.6 Fit of the mixture model for the USA in 2009. 36
3.7 Fit of the mixture model for Ukraine in two neighbor years. 37

3.8 Trend of the mixture parameter η. 37
3.9 Trend of the variance of the childhood mortality. 38
3.10 Trend of the skewness parameter λ_M. 39
3.11 Trend of the late mode h_M. 40
3.12 Trend of the shape parameter ω_M. 40
3.13 Trend of the percentage of deaths related with adult mortality component. 41
3.14 Trend of the accidental and premature mode h_m. 42
3.15 Trend of the skewness parameter λ_m. 43
3.16 Trend of the scale parameter ω_m. 43
3.17 Percentage od death related with accidental and premature mortality. 44
3.18 The e_0 contribution of the three components: in pink adult mortality, in green the accidental and premature, in light blue the infant and the child one. 45
3.19 Distribution of death by age and related survival curves for different years in France. 45
3.20 External causes of death (black solid line) in France and the estimate f^m function (blue line). 49
3.21 Restricted external causes of death (black solid line) in France and the estimate f_m function (blue line). 49

4.1 EM estimate and their confidence intervals for USA in 1980 when the last open age class is 85+, 74+ and 69+, respectively. 54
4.2 EM estimate for the USA in 2009. 54
4.3 Reconstruction of Danish birth cohort 1940 (male) and estimates of life expectancy for not concluded cohorts (male). 55

A.1 Estimated model for different years in Sweden (1910-1960). 59
A.2 Estimated model for different years in Sweden (1970-2010). 60

B.1 Decomposition of the area for levels of education and quantification of lost years due to premature mortality in the calculus of life expectancy at birth. 61

C.1 EM estimates and confidence intervals for USA 1980. 63

List of Tables

2.1 Estimated values for the infant mortality using three different probability distribution. 18
2.2 Areas of the different components for mortality components. 27

3.1 Parameter values for different levels of education (male). 46
3.2 Decomposition of the area for different levels of education (male). 47
3.3 Lost years due to premature mortality in the calculus of life expectancy at birth. . 47

4.1 Parameter values estimated via maximum likelihood for USA in 2009 and 2010. . . 51

B.1 Parameter values for different levels of education (female). 61

Chapter 1

Introduction

Together with fertility and migration, mortality is one of the three components that defines a population. It is a complex object because it changes across time, countries and groups with different socio-economic features. It is continuously developing, its force decides the number of survivors at each age and characterized countries.

In Medieval age the image of Death was a woman skeleton, who was often portrayed with her sickle. She cut the grass in the fields without any differentiation. In Promessi Sposi, Alessandro Manzoni wrote:

"Come il fiore già rigoglioso sullo stelo cade insieme col fiorellino in boccio, al passar della falce che pareggia tutte l'erbe del prato",

Without metaphor, there was the belief that Death did not depend on social status, age and sex. This certainty led to the representation of "Dance of Death" (1400-1500) in which men, women and children of different class origin (labourers, king, pope) was painted dancing with skeletons. The aim of these pictures was to remind people the fragility of life. Particularly in England, Death was associated with a gambler who played with dices. In the "The Rime of the Ancient Mariner" by Samuel Taylor Coleridge (1797-98) we can find the description of Death playing to win the souls of the sailors:

"The naked hulk alongside came,
And the twain were casting dice:
-The game is done! I've won, I've won!-
Quoth she, and whistles thrice."

The belief that Death did not follow rules derived from the uncertainty of life, caused by famines, epidemics, wars and bad healthy conditions. However, at the end of 1800, different authors explained that the probability of dying was not the same at every age. In Figure 1.1 the distribution of deaths by age is drawn for Sweden in 1751 [1]. There is a strong concentration of events after birth and during childhood. The remain curve is very flat with a little increase between 40 and 80 years old. The life expectancy at birth is estimate equal to 36.8 years old. This means that during 1751 in Sweden, people were not exposed to the same risk of dying.

Mortality has changed over time and it is different across countries, however its elementary structure is preserved. To analyze and understand mortality it is necessary to distinguish among its

[1] Sweden is the countries which has the longest time series in the Human Mortality Database. Even if the quality of its data between 1751 and 1860 is poor, we include this graph only to give an idea of the distribution of deaths during the eighteenth century.

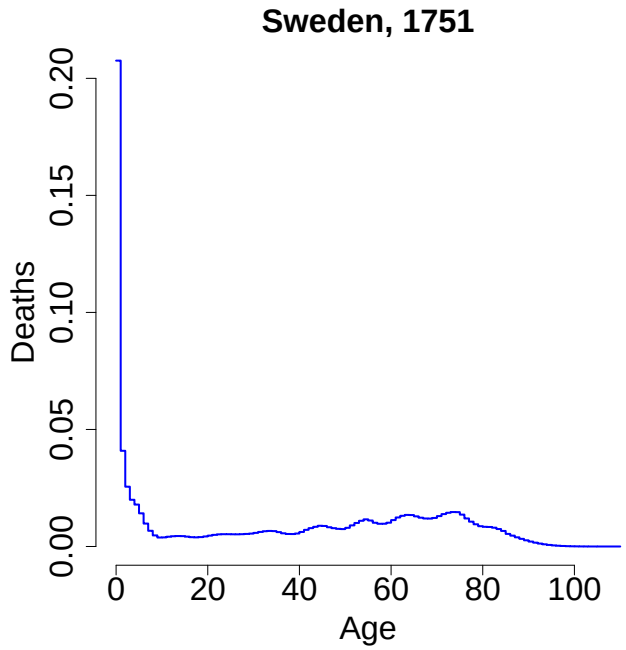


Figure 1.1: Distribution of death by age for Sweden in 1751.

different components. Infant mortality is identified with the deaths at age 0, childhood mortality with the deaths occur between age 1 and the beginning of teenager age. Accidental mortality indicates the hump that sometimes is visible in male distribution around 20 years old. Premature mortality can be defined as the deaths occurring in an intermediate region between childhood and old ages. Adult mortality is made up of deaths in the last part of the life span. Adult and premature mortality regions might be partially overlapping and this can make them hard to distinguish.
Using these categories, for example, it is possible to describe the changes in Sweden through 1935 and 2010 (see Figure 1.2). We observe a high reduction of infant mortality; childhood and accidental mortality decreased so much that, in 2010, are quite negligible; adult mortality undergoes a compression and a shifting. As a consequence premature mortality reduces its intensity.
Mortality components are also useful to compare groups with different characteristics. In Figure 1.3 the logarithm of male mortality rate for different education levels is plotted (source: ISTAT 2016). The educational qualification is a proxy of social-economic status and the problem of the gap between life style and mortality is not new in the literature (Hattersley, 1997; Huisman et al., 2004; Dalstra et al., 2006; Marmot and McDowall, 1986; Shkolnikov et al., 2011; Strand et al., 2010; Valkonen, 2001; Valkonen et al., 1993, Zarulli et al., 2013, Zarulli et al., 2012). Accompany with the increment of school years, we observe a well-defined decrease of mortality rate. The group with higher education shows a lower mortality trend, while the people without qualification or with primary school diploma has the higher one. The other two groups are proportionally included in the range of the previous curves. Individuals with middle school diploma collect a rate mortality pattern higher than men with high school diploma. It is evident that the main differences are concentrated among 25 and 65 years old, which means that premature mortality plays a significant role in the study of this phenomenon.

In order to identify and quantify the different mortality components, we propose a new model with 8 parameters which fits the distribution of deaths by age (Section 1.1). Our target is to intro-

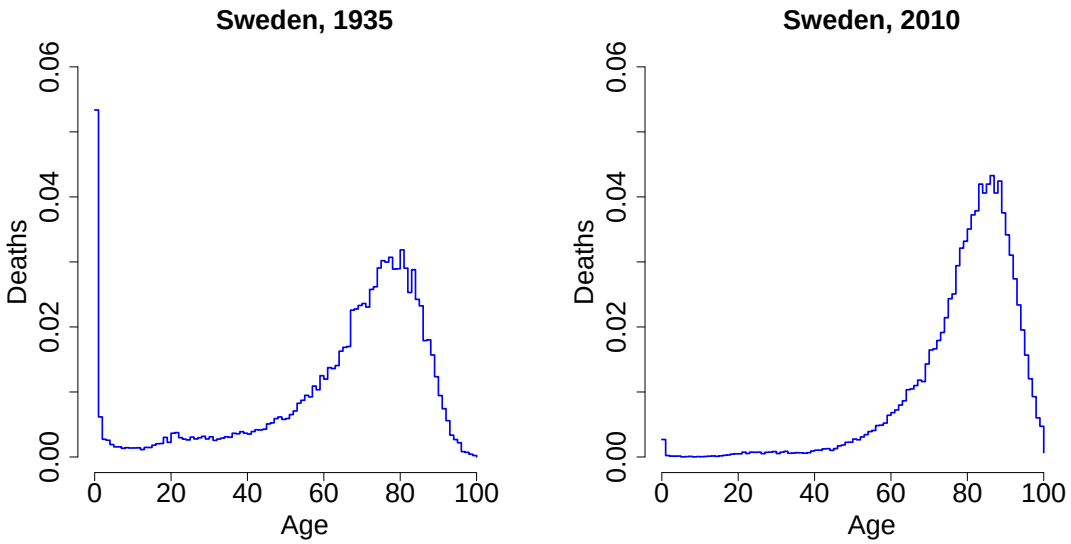


Figure 1.2: Distribution of deaths by age for Sweden in 1935 and in 2010 (source: Human Mortality Database).

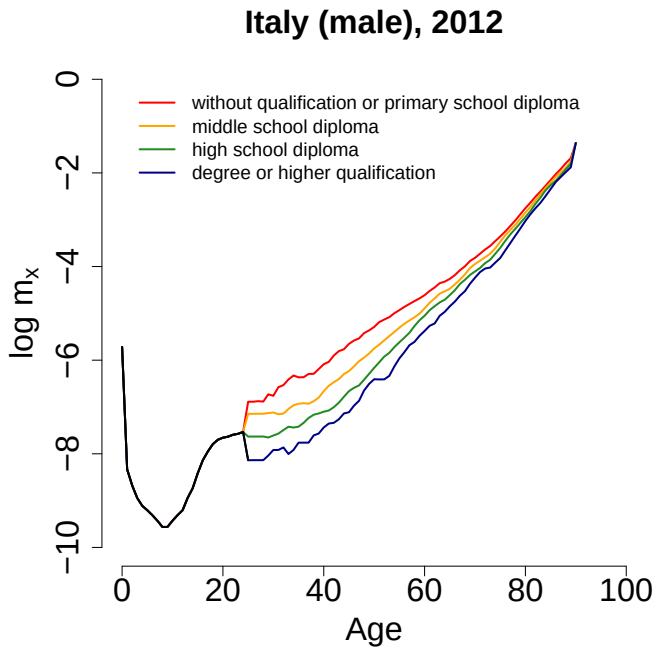


Figure 1.3: Logarithm of male mortality rate of different education levels (source: ISTAT).

duce a new tool which is able to identify mortality components and, in particular, to differentiate premature and adult mortality. In fact, premature mortality is frequently not evaluated because there is not a clear definition of it and, often, its identification depends on the description of adult mortality [2].

Even if the distribution of deaths is very useful for the analysis of mortality, it is not often take into consideration in the modelling process because of its shape, which is not easy to approximate. Someone may object that, even if we are working with probability, hazard or survival functions, it is possible to easily reconstruct the entire life table and also the distribution of deaths. This way of think leads to the consequence that the starting point is not important and the best choice is the function which has a simple shape to model. This can not be true if the target of the study is to divide among mortality components and, in particular, to observe premature mortality. In this case the distribution of deaths by age can be more appropriate to answer the question. Moreover it is important to remark that the identification of premature mortality can be done only consider the distribution of deaths by age. Indeed, considering the patter of the force of mortality (see Section 2) the senescent mortality, the last part of the curve, is basically modelled with one function. Using two mathematical equations is not suitable to fit an almost linear curve, without any break points. This prevents from the distinction between premature and adult mortality: how to split two components that are modelled with the same function? Also in Heligman and Pollard's model (1980), which fits the probability of dying (Section 2), there is not this differentiation. They use a function to fit the accidental hump, but accidental mortality is only a part of premature mortality. Also with the analysis of the survival curve the distinction in not possible. The indexes constructed to study horizontalization, verticalization, and longevity extension (Section 1.0.1) can give us some knowledges about the shape of the distribution of deaths, but they are insufficient to reconstruct the incidence of each mortality component.

The choice of the use of the distribution of deaths by age is due to two reasons. First the distribution of deaths can be see as a probability density function (see Chapter 2) so we can use a mixture of probability distributions to fit it: we desire that every mortality component have it own distribution, such as it is possible to compute all statistician indexes (like mean, mode, variance, skewness,...) and better analyze their trends (Section 2). Second the employment of this curve simplify the identification of mortality components, as many authors showed and discussed (Section 1.2). In fact, most of the authors who worked in this field focused on this distribution because the identification of mortality components is easier than with others life table functions. Moreover, in our work we decided to follow Pearson's theory, which origins from the observation of the distribution of deaths. The obvious consequence is to implement a model to fit this curve. Anyway, also with our model it is possible to reconstruct life tables using, as starting point, the estimated number of deaths by age and the amount of births.

The distinctive trait of Pearson's idea is to assign to every mortality components a distribution. In particular for infant and childhood we set an Half Normal distribution, for adult mortality a Skew Normal is employed and another one is used for accidental and premature mortality (see Section 2.1). In this way, every components has a precise definition. No one component is defined in relation to another one, even if their distribution are correlated. We studied the parameters trends and, for each, we found a demographic interpretation. Moreover their values can be used

[2] In fact, taking into account the distribution of deaths by age, it is common to consider adult mortality as the deaths identified by a symmetric curve with its maximum on the modal age and its sides equal to the last part of death cure (Lexis, 1879). All the deaths out of this zone are considered premature mortality.

both as indicator of the mortality pattern and to explore the differences between populations. An advantage of our model is that it allows to compute in explicit form life expectancy at birth, considering the contribution of each mortality components. Recently some authors exposed the importance to use the modal age at death in addition to life expectancy at birth in the study of mortality (Cheung et al., 2009, Kannisto, 2001, Wilmoth and Horiuchi, 1999). In our model it is achievable to compute the mode of the death distribution by age, even if with numeric method. With our model we can also decompose the overall area under the death curve into single areas, that indicate the percentages of deaths related with each mortality components.

1.0.1 Background

In demographic history "demographic transition" represents a fundamental change. This term defines the transition from high birth and death rates to lower birth and death levels. The theory was first proposed by Thompson (1930), who observed the trends of births and death rates in industrialized societies. Later on, Landry (1934) analyzed these transformations, identifying three demographic stages: before, during and after the demographic transition. In the first phase, before the nineteen century (Notestein, 1953; Landry, 1934), the distribution of deaths was very similar for all European counties and was characterized by a very hight infant mortality and a slow decreasing in childhood. The life expectancy at birth was around 25-35 year old (Coale, 1989; Flinn, 1981; Livi-Bacci, 1999) and there was a substantial equality between births and deaths (Malthus, 1926). Around 1800, with the improvements in personal hygiene, in medicine, in technology and then, with the possibility of regular food supply (Coale, 1989) and the increase in nutrition (Lee, 2003) a reduction of mortality started. The decline of mortality, at the fist stage was mainly characterized by a decrease of infant and in childhood mortality. The deaths began to be postponed at older ages and concentrated around the adult mode of the distribution, but still with a big variance and asymmetry on the left. The result was an extension in life expectancy. The mortality decline was followed by fertility reduction. Between 1890 and 1920, in most European countries, marital fertility became to decline (Lee, 2003). The pretransition fertility was hight within marriage and characterized by a total fertility rate greater than four children per woman (Livi-Bacci, 1999). The decline was due to many factors: the improvement in child survival, the parental decision to invest more in children health and welfare, the women entrance in labor market, the increase of contraceptive behaviors (Lee, 2003). The result was that fertility fell far below replacement level in many industrial nations (Van de Kaa, 1987). The combination of fertility and mortality trends led to population growth. In recent years we observe new trends both in mortality and in fertility. For the last one a light increase of the total fertility rate is registered and the U.N. projection forecasts to return fertility toward replacement levels (2.1). Regarding mortality, life expectancy at birth continuous to grow. These increments are particularly ascribed to a compression of the deaths around the late mode at death (Bongaarts, 2005; Canudas-Romo, 2008; Kannisto, 1996, Lan Karen Cheung and Robine, 2007) and a shifting of the adult mode towards older ages (Cheung, 2003; Lan Karen Cheung and Robine, 2007; Cheung et al., 2008; Cheung et al., 2005).
The entire process of the mortality reduction is well explained observing the survival curve, which is made up of the percentage of living at every age. As we already mentioned, the mortality change follows two sequential steps: first a large mortality reduction in early ages (first half of twentieth century), then a reduction among elderly in life (Kannisto, 1994, Robine, 2001, Wilmoth and Horiuchi, 1999, Wilmoth et al., 2000). According to Fries (1980) the human survival curve tends to become more and more rectangular as mortality declines because an upper boundary

to life expectancy at 85 years old is determined by fixed genetic limits. This concept is called rectangularization of survival curve (see the left graph in Figure 1.4). It is a theoretical principle:

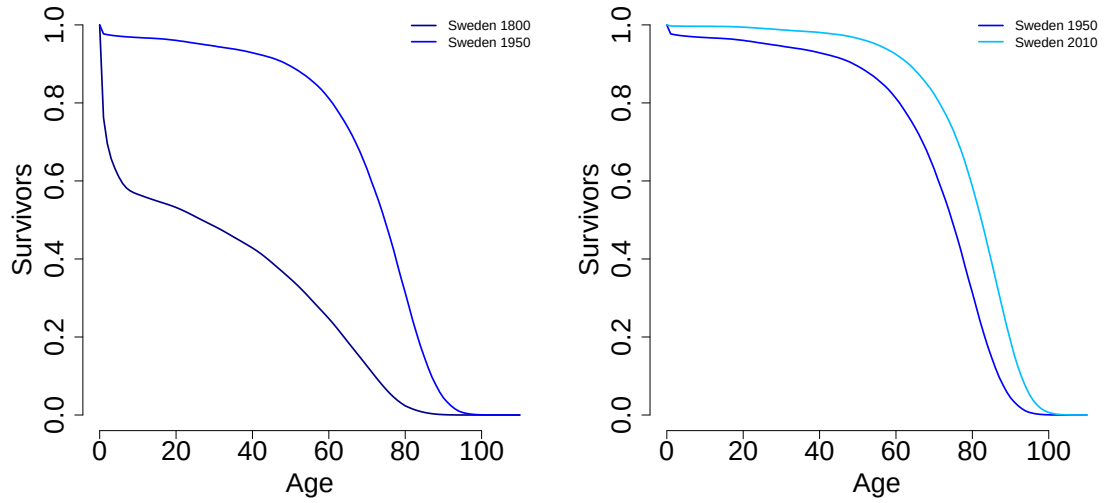


Figure 1.4: Survival curve for different year in Sweden (source: Human Mortality Database).

survival curve will never become perfectly rectangular, since it would imply zero variability in the age at death (Wilmoth and Horiuchi, 1999). Many authors found evidences of rectangularization during the epidemiological transition (Cheung, 2001; Eakin and Witten, 1995; Go et al., 1995; Hill, 1993; Levy, 1996; Manton and Stallard, 1996; Martel and Bourbeau, 2003; Nagnur, 1986; Nusselder and Mackenbach, 1996, Nusselder and Mackenbach, 1997; Paccaud et al., 1998; Pelletier et al., 1997; Robine, 2001; Rothenberg et al., 1991 Wilmoth and Horiuchi, 1999). In order to identify and quantify this concept Cheung et al. (2005) use three definitions to analyze the complexity of rectangularization:

- horizontalization corresponds to the proportion of individuals surviving before the significant decrease of the survival curve;
- verticalization describes the concentration of events around modal age at death;
- longevity extension which denotes how far is the right tail of the survival curve with respect to the modal age at death.

Nevertheless with the increase of life expectancy at birth the rectangularization theory became more controversial (Barbi et al., 2003; Oeppen and Vaupel, 2002; Olshansky et al., 2001). Instead of an increase of rectangularization trend, an almost-parallel shift of the survival curve to the right is emerging in France, Japan, Sweden and USA (Lynch and Brown, 2001; Horiuchi and Wilmoth, 1997, 1998; Robine, 2001; Yashin et al., 2001), as it is possible to see in the left graph of Figure 1.4. These shift increases the mean value of the life span and the number of centenarians. In some countries the shift of the modal age at death is accomplished with a slightly decrease of the variance of the life span, mostly because of the continuing mortality decline in youth and adult ages (Yashin et al., 2001). While, in other countries this phenomenon is associated with a derectangularization of survival curve, which means that a deconcentration of deaths around the modal age is registered (Gavrilov and Gavrilova, 1991).

Following the paths of adult and premature mortality is essential to explain these transformations, to indicate which is the component responsible for this process and to understand the relation between their paths.

1.1 Life Table

The main instrument to statistically study mortality is the life table. It describes the elimination by death of a group of individuals. Since all people die, the intensity of the phenomenon is always equal to 1. This method has the advantage to provide compared result between populations and different years. In fact the outcomes are not affected by the amount of individuals and the age structure of the population (Blangiardo, 1997). The life table describes the force of mortality that characterizes and distinguishes populations. It is composed by several functions.

- q_x is the probability of dying in age interval $[x, x+1)$.
- p_x is the probability of survive in age interval $[x, x+1)$. It is the complement of q_x:

$$p_x = 1 - q_x. \tag{1.1}$$

- l_x are the survivors at age x. To compute the life table l_0, the number of births, is required. This number is called root of the table and it is an arbitrary value. Usually $l_0 = 10^k$ with $k \in \mathbb{N}$. The survival curve is obtained plotting the values of l_x/l_0 of the life table. The probability of die and the survivors are tied by the following equation

$$l_{x+1} = l_x - l_x q_x. \tag{1.2}$$

- d_x are the deaths in age interval $[x, x+1)$, so

$$d_x = l_x q_x = l_x - l_{x+1}. \tag{1.3}$$

  Thus, for the last open age class $d_\Omega = l_\Omega$.

- L_x are the years lived between x and $x+1$ and they can be computed

$$L_x = \frac{l_x + l_{x+1}}{2} = l_x - 0.5 d_x, \tag{1.4}$$

  With Equation (1.4), we are assuming that people die in the age interval $[x, x+1)$ live 6 months on average. This is not true in particular for age 0, where deaths are concentrated near the birth. For this reason

$$L_0 = l_1 + a_0 d_0, \tag{1.5}$$

  where $a_0 \in (0, 1)$ is the coefficient to correct the number of years lived and it is computed as follow

$$a_0 = \begin{cases} 0.350 \text{ for female,} \\ 0.330 \text{ for male} \end{cases} \quad \text{if } m_0 \geq 0.107 \tag{1.6}$$

$$a_0 = \begin{cases} 0.053 + 2.800 m_0 \text{ for female,} \\ 0.045 + 2.684 m_0 \text{ for male} \end{cases} \quad \text{if } m_0 < 0.107, \tag{1.7}$$

  where m_0 is the mortality rate for age 0. For the last open age class, we set $L_\Omega = 0.5 l_\Omega$.

- T_x indicates the overall number of years lived by surviving at age x until the last age Ω:

$$T_x = L_x + L_{x+1} + \cdots + L_{\Omega-1}. \tag{1.8}$$

  For the last open category $T_\Omega = L_\Omega$

- e_x is the remaining life expectancy for people at age x:

$$e_x = \frac{T_x}{l_x}. \tag{1.9}$$

  Often the life expectancy at birth, e_0, is the main index to summarize the life table.

The life table can be computed for cohort data or period data. With the term "cohort" we identify a group of births who experiment the same event (born) in a specific time interval (Preston et al., 2001), so a cohort life table depicts the life history of a specific group of individuals (Wilmoth et al., 2007). Instead a period life table is constructed considering individuals of different generation, who coexist in the same time interval, even thought with different age (Blangiardo, 1997). Then, the period life table is supposed to represent the mortality conditions at a specific moment in time. However, observed period death rates are only one result of a random process for which other outcomes are possible (Wilmoth et al., 2007). Both cohort and period life tables are composed by the same functions, but the starting point to compute the probability of die is different. When we are working with period data the first step is to compute the age specific mortality rates

$$m_x = \frac{D_x}{\bar{P}_x}, \tag{1.10}$$

where D_x are the death counts occur in a calendar year and $\bar{P}_x$ the average of the population at the beginning and the end of the considered period. In Figure 1.5 the Lexis diagram is drawn to better understand the quantity involved. In particular: $D_x = D_L(x,t) + D_U(x,t)$ and $\bar{P}_x =$

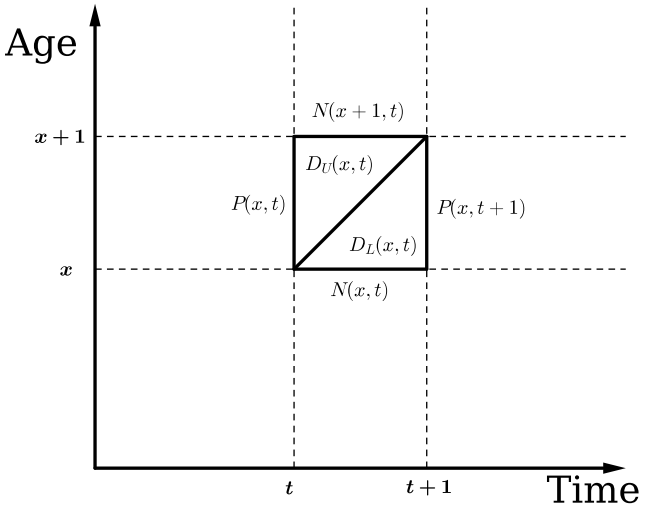


Figure 1.5: Data for period death rates.

$0.5[P(x,t) + P(x,t+1)]$, so that the death rate m_x involves the deaths of two neighbor cohort. To migrate from death rates to probabilities and compute the life table we can use the follow relation

$$q_x = \frac{2m_x}{2 + m_x}. \tag{1.11}$$

This Equation follows the consideration that $m_x = \frac{d_x}{L_x}$ and $Lx = lx - 0.5d_x$. Hence

$$m_x = \frac{d_x}{lx - 0.5d_x} \quad \Longrightarrow \quad \frac{1}{mx} = \frac{1}{q_x} - \frac{1}{2}, \tag{1.12}$$

and Equation (1.11) is obtained (Cox, 1950). Equation (1.11) can not be employed to compute q_0 because the assumption that d_0 randomly occurs in the time interval is not realistic. To reconstruct the probability of die at age 0 it is convenient to consider (Coale et al., 2013; Preston et al., 2001)

$$q_0 = \frac{m_0}{1 + (1 - a_0)m_x}. \tag{1.13}$$

For the last open age class we set $q_\Omega = 1$.
To reconstruct the life table for cohort data we can compute directly the probability of die by

$$q_x = \frac{D_x}{\bar{P}_x}, \tag{1.14}$$

where D_x are now the deaths of individual who are the same biological age and $\bar{P}_x$ is the alive population at the halfway point of the considered time interval. In Figure 1.6 it is possible to see that the amount of both the numerator ($D_x = D_U(x, t+1) + D_L(x, t)$) and the denominator ($\bar{P}_x = P(x, t+1)$) belong to the same birth cohort.

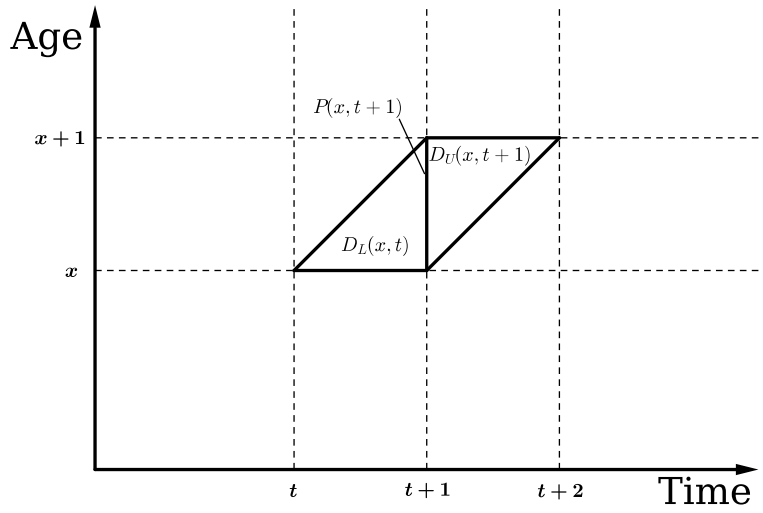


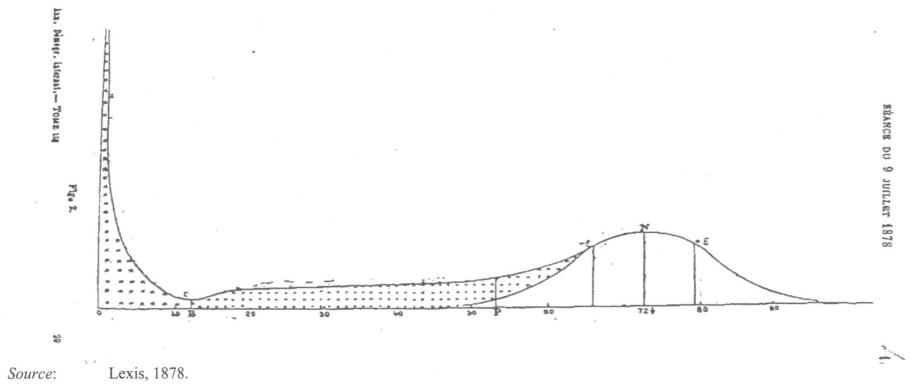
Figure 1.6: Data for cohort death rates.

1.2 Mortality components

In demographic literature there are two main theories about mortality components. Both of them take place from the observation of the distribution of deaths by age (d_x of a life table). This curve is chosen because it permits an easier identification of mortality component thanks to the presence of breaking points which are both particular ages (like age 0 for infant mortality or age 20 for accidental mortality) and maximum and minimum in their shape (late mode of the distribution for adult deaths, minimum in the first part of the graph to identify childhood mortality).

Lexis (1879) decomposes the distribution of deaths by age in three parts: a "J curve" after birth corresponding to infant deaths, the "normal deaths" around the late modal age at death

that obey the law of accidental errors and reflect a natural lifetime, and a transition region where premature deaths and normal deaths partly overlap (see Figure 1.7). Lexis considers the mode of



Source: Lexis, 1878.

Figure 1.7: Lexis' components for the distribution of death by age (Source: Lexis, 1879).

the death curve the central characteristic of human longevity. He defines the area of natural death flipping the curve from the late mode to the end of the last age Ω. All the deaths outside this symmetric distribution are premature mortality, which is related to reproduction, risk behavior, accidents and infection diseases, while the deaths under the symmetric distribution are considered adult mortality. This theory was widely shared (Benjamin, 1959; Bodio, 1887; Elderton, 1903, Gumbel, 1937, Levasseur, 1889; Pareto, 1964, Perozzo, 1879) and used to develop indexes to study mortality changes (Cheung et al., 2005; Kannisto, 2001).

Pearson (1897) evaluates the problem of mortality components on a more statistical point of view. In his opinion mortality curve is not one simple frequency curve, but is made up of several components. He considers the distribution of deaths composed by five functions with different degrees of skewness and he maps them in Figure 1.8. He distinguishes between infancy and childhood

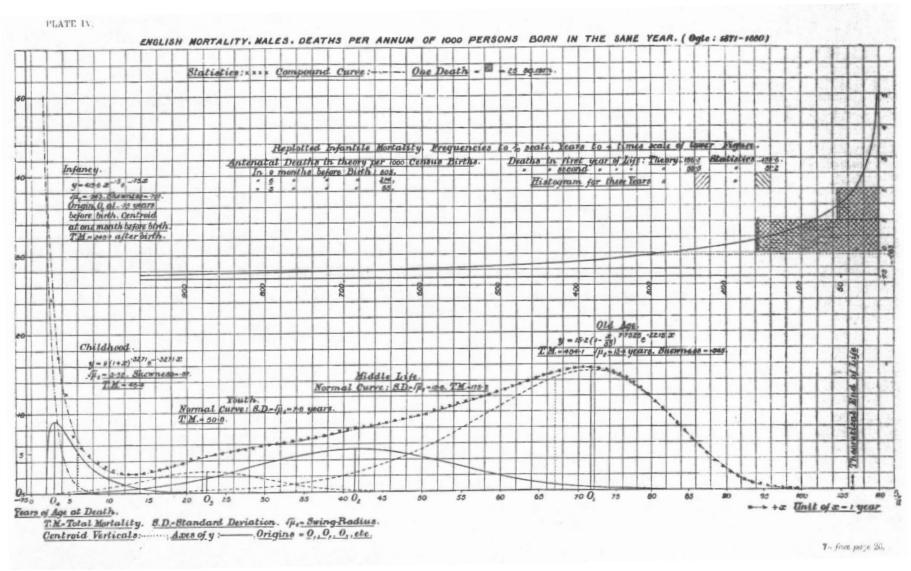


Figure 1.8: Pearson's five components for the distribution of death by age (Source: Pearson, 1897).

mortality. He proposes for the first an exponential curve, covering also antenatal period (to model the foetus probability of dying) and, for the second, a very skew distribution. He also differentiates

Lexis' transitional region between youth (accidental) and middle life (premature) deaths. For both of them he draws a symmetric distribution with a mode around 25 years old and another one around 40 years old, respectively. Pearson evaluates the possibility that premature mortality is due to a a specific or a group of specific diseases. He analyses the trend of cancer and accidents but he concludes that there is not a specific cause to explain middle life deaths. Finally he identifies old mortality like an asymmetric distribution with skewness towards youth. The reasons why he made this choice are double. Pearson's theoretical argumentation was that adult mortality depends on the incidence of deaths at earlier ages, so it can not be symmetric, while the practical motive was that without a skew curve, he was not able to obtain a satisfactory fit of the overall curve. Pearson's theory had not the same popularity of Lexis' one, probably for its complexity, even if it is more complete and statistically supported.

Even if the two theories have the same goal (identifying mortality component and propose rules to distinguish them) the tools and the results are not the same and, for some aspect, in contrast. Fist the methods are different: Lexis' analysis is based on the observation of the shape of the curve and its break points, so that the components are related to specific shapes; Pearson's theory used distribution to model the curve and he related each function to one component. One consequence is that Pearson introduces the concept of probability in his argumentation, so that, at a certain age, a death is related with all the components, but with different probabilities. Instead, Lexis divided the curve in three defined areas, therefore, at a certain age, the associated mortality component is clear identifiable except for the transition region in which the area of premature and adult mortality are partially overlapping. The other effect of these points of view is that in Lexis' formulation, premature mortality is a consequence of the shape of normal deaths and it can be define only as their consequence, while, with Pearson, it has its own mathematical expression. Another difference lies in the idea of adult mortality. For Lexis it a symmetric area, while for Pearson it is a skew distribution. This difference depends on the concept of independence (or dependence) of deaths. For Pearson there is a relation between deaths, because the events we observe after, inevitably can not occur before. Instead, Lexis suggested that there is no connection among deaths, so a symmetric function is suitable to capture adult mortality: adult deaths randomly occur so the shape we observe from the mode of the death distribution to its end must to be equal to what happens before the mode.

1.3 Main models to study mortality

Even if nonparametric methods permit a more accurate fit of the mortality data, parametric models are often employed because they facilitate interpretability, comparison and forecasting (Congdon, 1993). The analysis of parameter values can be used as indicator of mortality pattern and to quantify the difference among groups of individuals. Moreover the time series of the parameter estimates can be studied in order to predict future mortality scenarios.
Congdon (1993) delineates the guidelines to assess model performances in demographic research. They can be considered the ideal characteristics that a model should have: smoothness, parsimony, interpolation, comparison, trends and forecasting, analytic manipulation.

- Smoothness: the ability of the model to produce smoothed estimates, which are not affected by data random fluctuations.

- Parsimony: the number of parameters needs to be as small as possible, but at the same time

a suitable number to guarantee the best estimate. There is a trade off between the amount of parameters and the best fit possible: usually, fit increases when the quantity of parameters grows. The best model is necessarily a compromise between these two aspects.

- Interpolation: the model can be applied also when the data are clustered in age classes or are incomplete (very common situations when we are working with developing countries or with historical data).
- Comparison: the model needs to be flexible. It can be applied to different groups in space, time or subset of populations.
- Trends and forecasting: the series of parameters estimates identify patterns over time. The extension of them allows forecasts for the future.
- Analytic manipulation: if a function involved has known properties, they need to be considered when the characteristics of the overall model are analyzed.

To study the age pattern of human mortality, some models ware proposed. In this section we recall only the main models, but many others are proposed in the literature (Beard, 1971; Basellini et al., 1971; de Beer and Janssen, 2014; Coale and Kisker, 1990; Horiuchi and Wilmoth, 1998, Thatcher, 1999). Historically, the exponential distribution was first adopted to analyze mortality. It presents a constant hazard function (Lawless, 2011). This characteristic facilitates the mathematical tractability of the model, but in many contexts it proves to be inappropriate because of this strong restriction (Klein and Moeschberger, 2005).
In 1825 Gompertz introduced a model to fit the instantaneous rate of mortality at age x, called force of mortality μ_x. It is defined as follows:

$$\mu_x = a\mathrm{e}^{bx}, \tag{1.15}$$

$$\text{with} \quad \mu_x = -\frac{\partial}{\partial x}\ln(l_x). \tag{1.16}$$

The model can be interpreted as the deterioration of body age by age. It is generally employed to describe the increase of mortality at older ages. It has two parameters that describe the initial size of mortality (a) and the rate of mortality increase (b). The model is able to fit the increase of mortality at older ages, but it seems to underestimate young adult mortality rates and, for compensation, overestimate the oldest ages rates (Vaupel et al., 1998, Bongaarts, 2005).
Since this model was not able to capture accurately the mortality pattern, Makeham (1860) added a constant (a_1), representing deaths that occur randomly with respect to age (Thatcher, 1999).

$$\mu_x = a_1 + a_2\mathrm{e}^{b_2 x}. \tag{1.17}$$

Subsequently, Siler (1979) generalized the latter model in order to include infant and childhood mortality. His model has five parameters and three components.

$$\mu_x = a_1\mathrm{e}^{-b_1 x} + a_2 + a_3\mathrm{e}^{b_3 x}. \tag{1.18}$$

The additional component has two parameters representing the initial magnitude of infant mortality (a_1) and its decline rate during the childhood (b_1).
Another strategy to analyze adult mortality is the logistic model, which was first proposed by Perks (1932):

$$\mu_x = a_1 + \frac{a_2\mathrm{e}^{b_2 x}}{1 + a_3\mathrm{e}^{b_2 x}}. \tag{1.19}$$

This model can be considered a generalization of both Makeham and Gompertz model. In fact we obtain Equation (1.17) when $a_3 = 0$, while if $a_1 = 0$ and $a_3 = 0$ Equation (1.15) is recovered. Recently, this model was employed to estimate μ_x at older ages (Wilmoth and Horiuchi, 1999).
Kannisto (1992) have proposed logistic-type models to better estimate the force of mortality at older ages (over 85) in industrialized countries (see Thatcher et al., 1998). It has the following equation:

$$\mu_x = a_1 + \frac{a_2 e^{b_2 x}}{1 + a_2 e^{b_2 x}}. \tag{1.20}$$

In the Human Mortality Database (HMD), the smooth estimates of the probability of dying at ages greater than 80 years old, are computed with Equation (1.20), that can be rewritten considering q_x as response variable (Wilmoth et al., 2007).
Parallel with the develop of Gompertz model, the Weibull distribution (1951) was introduced in demographic research to fit adult mortality hazard rate:

$$\mu_x = a x^b. \tag{1.21}$$

Even if it was introduced to explain failure of a technical system, it can be transposed to mortality considering the death as the results of body failure (for example cells damage) (Thatcher, 1999). Weibull distribution is a monotonic function and this can be inappropriate in some situations (Bennett, 1983). To bypass this problem log-logistic or log-normal distribution can be chosen (Bennett, 1983).
Heligman and Pollard (1980) started from the observation there was not a model able to capture accidental mortality. They developed an eight-parameter model incorporating the hump between 10 and 40 years. They model the probability of dying q_x:

$$\frac{q_x}{1 - q_x} = A^{(x+B)^C} + D e^{-E(\ln x - \ln F)^2} + G H^x. \tag{1.22}$$

The model is a mixture of three components. The first function describes the infant mortality and its reduction. A is very close to q_1, the probability of die during the first year, C fits the decline of infant mortality and B measures the location of child mortality. The second polynomial is similar to a log-normal function and reflects the accidental mortality. F indicates the location, E the spread and D the force of accidental mortality. The third part reflects old age mortality being G the base level of old age mortality and H the rate of increase of mortality with age. Nevertheless, in order to fit well the little hump around age 20 when it is very flat, an extra parameter is required (Congdon, 1993; Heligman and Pollard, 1980; Kostaki, 1992).
Most of the models presented focus only on a specific part of mortality pattern. Using these methods, it is not possible to analyze the overall mortality, but only part of it. This can be a limit if the target of the research is to take into account the transformations of all mortality components and their relations. Moreover, in no one of these models there is the differentiation among premature and adult mortality. In some sense, in Siler model (and in Makeham too) there is a constant to estimate the accidental mortality, but, as a constant, it can not change across age. Moreover the last part of the curve is considered as a unique block and there is no way to separate premature mortality from adult mortality. Indeed, the separation between accident and premature mortality is not possible if the response variable is the force of mortality μ_x. Its curve is characterized by a strong decrease of the hazard straight after age 0 and an increase around 10 years old. There are no mathematical reasons to use two functions to fit this increase. The same in Heligman and Pollard model: the authors indicate a specific function for the accidental hump,

and another one for the last part of the curve. Also in this case, the split between premature and adult mortality is not possible because the last part of the curve is approximate with only one function. There is the possibility to recognize accidental mortality, but premature and adult mortality remain inseparable.

Chapter 2

Model and method

The goal of our work is to implement a model which is able to distinguish among mortality components, and, in particular to differentiate adult and premature mortality. The direct consequence of this is the employment of the distribution of deaths by age. This life table function, in fact, is the more suitable to reach our target because it facilitates the identification of mortality components, as many authors already showed (see Section 1.2). Even if the distribution of deaths by age (d_x) has a not easy shape to modelled, it has the advantage that it can be seen as a probability density function. Instead of the models presented before, which use the hazard or the probability of dying, we want to create a model that can fit directly the number of deaths. In fact the distribution of deaths (conveniently divided by the root of the life table l_0) give rise to a probability density function:

$$\sum_{x=0}^{\Omega} \frac{d_x}{l_0} = 1. \tag{2.1}$$

This means that, instead of the use of mathematical function to fit the shape of the target distribution, we can use a probability density function. This leads to some "statistical" advantages. In particular we can compute in explicit form the likelihood function (see Paragrapher 2.1.1). The first advantage is the possibility to estimate the parameters of the model using maximum likelihood procedure. The property of Equation (2.1) is not respect when μ_x or q_x are taken into consideration and more (or different) elaborations are required to find the parameter values of the models. This means that the property of maximum likelihood estimates can be directly applied only when we are working with probability density functions. The existence of an explicit form for the likelihood equation guarantees the possibility to employ several statistical method that require this function to be applied. An example is the EM algorithm, which will be debate in Chapter 4 and it is useful to improve the parameter estimates when there are some missing data. In our case we use it to upgrade the model fit when the last open age class is too wide, so that the frequency of the death counts at older ages is missing.

Unfortunately the shape of the distribution of deaths is complex and, in statistical literature, there is not a probability density function which has a similar form of the examined curve. The problem can be solved collecting several distributions and implementing a mixture distribution. The issues are: how many distributions? which functions?
We decided to follow Pearson's theory regarding two argumentations. Pearson's idea was to assign to every mortality component its own distribution. In this way it is possible to understand their characteristics and their transformations. In Lexis' theory, premature mortality was only a transition region specified in relation to normal death and infant mortality. It has not a proper

definition, so it has not individual features (as mean, mode, variance,...). The other key point of Pearson's theory is the shape of the distribution assigning to adult mortality. Instead of a symmetric function he uses a skew one. We are strongly convinced that the frequency of deaths at later ages must depend on the incidence of deaths at earlier ages, as Pearson wrote. In fact the deep we observe from the late mode is the consequence of the number of deaths we have registered before. For instance, when there is a concentration of deaths at young ages due to war, the distribution is different not only for premature mortality, but also at older ages. Even if events are independent, the shape of the distribution depends on their frequency: if some deaths occur at age 0, they can not occur at other ages.
Obviously if we strictly follow Pearson's approach we will obtain a mixture with not less than 13 parameters. This leads to identification problems: we have too many parameters to estimate. We need to find a compromise between the number of functions involved and the detail we require to describe mortality components. We end with three functions: one for infant and childhood mortality, one for accidental and premature mortality and one for adult mortality.

2.1 Definition of the mixture model

To fit infant and childhood mortality a Half Normal (HN) distribution is employed. Its shape and its theoretical characteristics are close to the features of the first part of the distribution of deaths by age. This distribution, defined only for values greater or equal to 0, has the following probability density function (pdf):

$$f_I(x;\sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad x \geq 0, \tag{2.2}$$

where the variable x is the age at death and $\sigma \in (0, +\infty)$ is the scale parameter of the distribution. If the parameter σ grows, the distribution becomes more flat (see Figure 2.1). The Half Normal distribution is not the only function which can be adapted to fit the first part of the distribution of deaths. We took into consideration also Exponential distribution, Half t distribution and Half Cauchy distribution, which have a similar shape as it possible to see in Figure 2.2, and they can be compatible with infant and childhood mortality. We immediately exclude the Half t distribution because its pdf involves two parameters. Considering we have other two distributions to involve in the model we avoid to have too many parameters to estimate. We estimated the model attaching the different functions and we observed the predicted value for d_0, which can be consider the target point for the evaluation. The result are show in Table 2.1. Exponential distribution is completely outside of the admissible range, probably because of identification problems. Half Normal and Half Cauchy distributions produce similar estimates. In Figure 2.3, we plot the predicted values against the real value d_0. As it is possible to see the Half Normal distribution presents more smooth and stable values for infant mortality. Even if the gap between the true d_0 and the Half Cauchy predictions is smaller, there are a lot of variations in its trend. This instability reflects the variability of the parameter estimates. We decided to chose the Half Normal distribution precisely because we desire to avoid instability in the estimates and to privilege a smoothed trend. The problem of the Half Normal distribution is that it always underestimates d_0, as it is possible to se in Figure 2.4. Since capturing infant mortality is a fundamental requirement when we want to model the entire schedule of the distribution of deaths, we studied the behavior of this function in order to obtain satisfactory estimate of the first part of the distribution of deaths. After some examinations, we decided to fix $\sigma = 1$ following two considerations. First, this allows that the

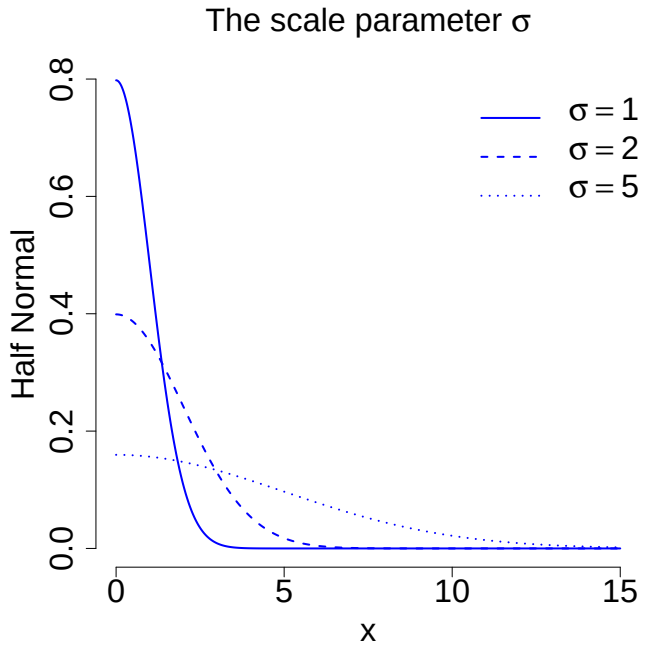


Figure 2.1: Probability density function of the Half Normal distribution as the scale parameter changes.

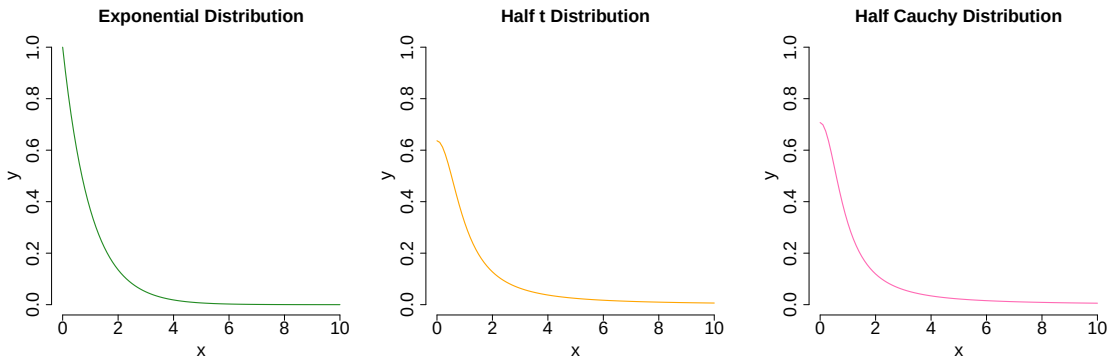


Figure 2.2: Examined functions to model infant and childhood mortality.

year	d_0	Half Normal	Exponential	Half Cauchy
1970	0.0128	0.0014	0.8615	0.0185
1971	0.0128	0.0013	19.9232	0.0221
1972	0.0128	0.0013	23.7020	0.0228
1973	0.0109	0.0012	22.5172	0.0179
1974	0.0106	0.0012	38.0546	0.0171
1975	0.0098	0.0012	25.9709	0.0152
1976	0.0091	0.0011	30.4012	0.0130
1977	0.0090	0.0010	37.9104	0.0139
1978	0.0089	0.0011	21.7291	0.0148
1979	0.0085	0.0010	1.0453	0.0134
1980	0.0081	0.0009	33.5609	0.0138
1981	0.0074	0.0009	1.3500	0.0122
1982	0.0071	0.0011	29.1572	0.0120
1983	0.0070	0.0009	23.3642	0.0014
1984	0.0072	0.0009	18.1488	0.0148
1985	0.0072	0.0008	25.8872	0.0124
1986	0.0066	0.0009	26.6990	0.0012
1987	0.0067	0.0011	30.1672	0.0113
1988	0.0066	0.0009	27.4116	0.0104
1989	0.0066	0.0008	26.8988	0.0127
1990	0.0066	0.0008	27.0197	0.0011
1991	0.0066	0.0008	25.2491	0.0131
1992	0.0060	0.0007	23.2269	0.0095
1993	0.0055	0.0007	34.5276	0.0009
1994	0.0049	0.0007	0.8835	0.0008
1995	0.0047	0.0006	25.1667	0.0008
1996	0.0042	0.0006	29.9299	0.0007
1997	0.0041	0.0006	28.9013	0.0007
1998	0.0040	0.0006	35.4364	0.0007
1999	0.0041	0.0006	22.4405	0.0007
2000	0.0040	0.0006	20.4089	0.0007
2001	0.0040	0.0008	0.0337	0.0007
2002	0.0035	0.0008	27.7988	0.0007
2003	0.0036	0.0006	41.1585	0.0006
2004	0.0033	0.0006	19.2383	0.0006
2005	0.0025	0.0005	47.8166	0.0005
2006	0.0030	0.0005	34.4219	0.0006
2007	0.0027	0.0005	26.8723	0.0005
2008	0.0025	0.0005	34.1590	0.0005
2009	0.0026	0.0005	48.8063	0.0006
2010	0.0027	0.0005	31.9587	0.0006
2011	0.0022	0.0005	43.2158	0.0005

Table 2.1: Estimated values for the infant mortality using three different probability distribution.

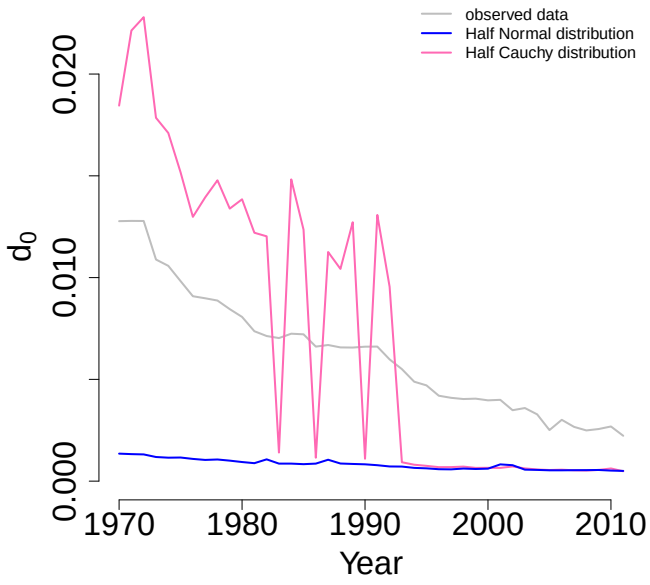


Figure 2.3: Comparison between the values estimated for infant mortality with Half Normal and Half Cauchy distributions.

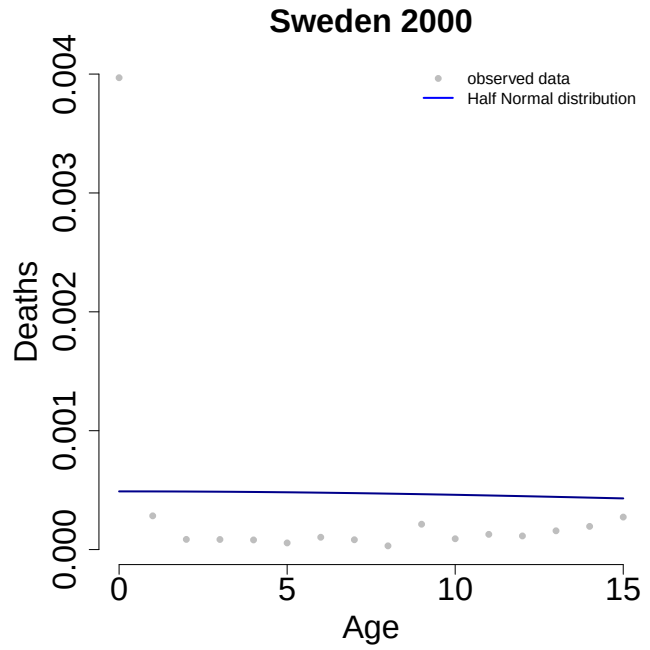


Figure 2.4: Estimated Cauchy distribution for infant mortality for 2000 in Sweden.

model can adequately capture the steep decline in the initial probability of death right after birth. In fact, there is only one data point, the number of deaths occurring between age 0 and 1, that can influence the slope of the Half Normal distribution. When its value is big it has a very strong influence in the estimation process, so the shape parameter will be well estimated. In low infant mortality contest, instead, its value is small and not very different from the number of deaths registered during childhood. In these cases we obtain large values for the parameter, that leads to a too flat distribution, which is unable to catch correctly d_0. Fixing $\sigma = 1$, we set the shape of the function and we are able to capture the initial magnitude of the distribution of deaths.
Second, we aim to reduce the number of parameters to avoid identification problems. This issue comes up when "for a given number of parameter and a given functional form, widely different values of one or more parameters may be associated with an approximately similar measure of goodness of fit" (Congdon, 1993, p. 240).

To capture accidental and premature mortality (m) and adult mortality (M) two Skew Normal (SN) distributions (Azzalini, 1985) are employed. This class of distributions was developed by Azzalini in 1985 and the two pdfs are the following:

$$f_m(x;\xi_m,\omega_m,\lambda_m) = \frac{2}{\omega_m}\phi\left(\frac{x-\xi_m}{\omega_m}\right)\Phi\left(\lambda_m\frac{x-\xi_m}{\omega_m}\right), \tag{2.3}$$

$$f_M(x;\xi_M,\omega_M,\lambda_M) = \frac{2}{\omega_M}\phi\left(\frac{x-\xi_M}{\omega_M}\right)\Phi\left(\lambda_M\frac{x-\xi_M}{\omega_M}\right), \tag{2.4}$$

where the subscript m indicates the function for accidental and premature mortality, while M the function for adult mortality; $\phi(\cdot)$ is the standard normal pdf, $\Phi(\cdot)$ the standard normal cumulative distribution function (cdf), $\xi_{(\cdot)} \in \mathbb{R}$ the location parameter, $\omega_{(\cdot)} \in (0,+\infty)$ the scale parameter and $\lambda_{(\cdot)} \in \mathbb{R}$ the skewness parameter. If $\lambda_{(\cdot)} = 0$, a Normal density function is obtained. In Figure 2.5 is shown the shape of the SN according to different values of λ. For the sake of simplicity, in the graphs, ξ and ω are fixed and equal to 0 and 1, respectively.

As Pearson theorized, the shape of the distribution for adult mortality needs to be skew. According to this schedule we choose a SN distribution. In statistic literature there are others distributions which define an asymmetric shape, also with less parameters. Our choice follows two considerations. First, an advantage of this class of functions is that it includes the normal distribution as particular case. In fact the parameter λ explicit controls the skewness of the SN distribution. Taking advantage of this we can examine the value of λ_M to verify if adult mortality component really needs an asymmetric distribution to allow a good overall fitting (as Pearson wrote) or a symmetric distribution is suitable. In this last case the parameters is required to be close to 0. The second point is that this distribution was already satisfactorily employed to fit the second part of the distribution of deaths by Mazzuco et al. (2016). In particular they use a Bimodal Skew Normal (SBN) distribution, which is essentially a mixture of two distributions, as we explain in Section 2.2. The authors show that the value of the mixture parameter is close to 0, in particular for recent years and in the situation in which the number of deaths, that occur at younger ages, is few. When the mixture parameter is close to 0, the SBN is essentially a SN distribution. In their work the authors tested the degree of adaptation of the model and found that the model is an adapted choice to model adult mortality path.

To decide the distribution to model accidental and premature mortality we consider the shape of the death curve in its middle part, that is the sum of accidental and premature mortality. Obviously the distribution will be a compromise between the two symmetric curves Pearson designed. To have a better idea of the configuration of this part of the curve, we predict the death counts d_x^*

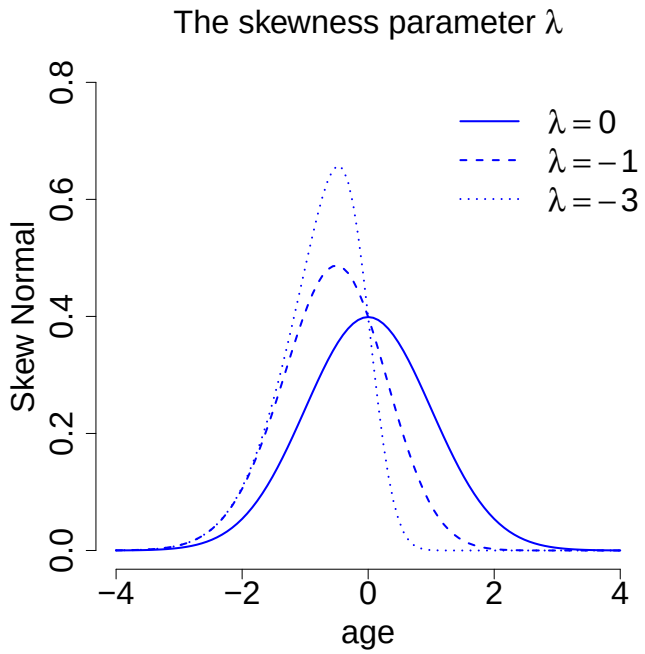


Figure 2.5: Probability density function of the Skew Normal distribution as the shape parameter changes (with ξ and ω are fixed and equal to 0 and 1, respectively).

with only Equation (2.2) and (2.4). We subtracted to the real d_x the computed values d_x^* and we observed the pattern of the errors committed. In Figure 2.6 some years are reported, focusing on the range of accidental and premature mortality. Both the graph of the left and the one on the

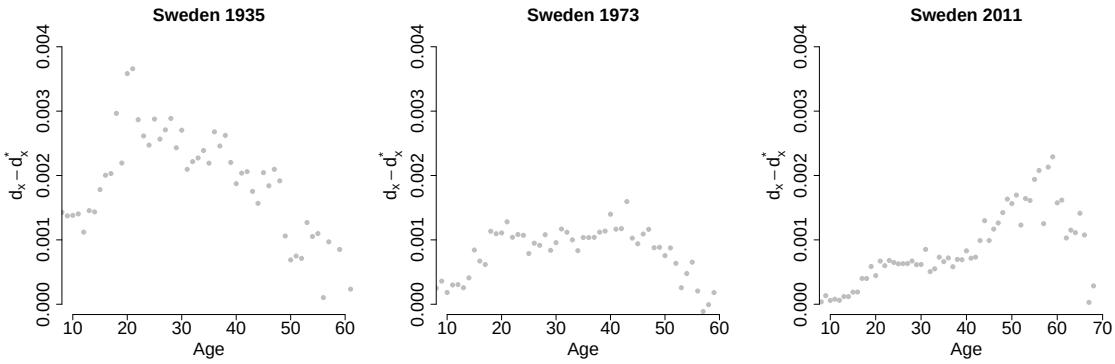


Figure 2.6: Errors committed by the model without the accidental and premature distribution.

right presents an evident asymmetry. The first one, on the left and the second, on the right. The left asymmetry is due to the presence of the accidental hump, in fact its mode is around age 20. In recent years the shape of premature mortality appears discordant from Pearson's theory because it shows an asymmetric distribution. The points in the middle plot seem to assume a symmetric shape, since the accidental mortality reduces its intensity and premature mortality continuous to exist. This means that the required distribution needs to model all the illustrated situations. We want that the new distribution fits the accidental hump (left graph) and premature mortality both when its shape is symmetric or not (middle and right graphs). For these reasons we employed another SN: it is a flexible distribution because the parameter λ controls the skewness and can fit

all three situations.
Combining the three Equation (2.2), (2.3) and (2.4) with the mixture (or weighting) parameters η and α, a model with 8 parameters is obtained:

$$f(x,\theta) = \eta \cdot f_I(x;1) + (1-\eta) \cdot \Big[\alpha f_m(x;\xi_m,\omega_m,\lambda_m) + (1-\alpha) f_M(x;\xi_M,\omega_M,\lambda_M)\Big], \tag{2.5}$$

where θ is a vector of 8 parameters. η is the first mixture parameter with value ranging in $[0,1]$ and α is the second mixture parameter which also varies in the interval $[0,1]$. The shape of the mixture model is drawn in Figure 2.7. If the value of η decreases, the infant mortality declines; with the increase of α, the relevance of accidental and premature mortality grows.

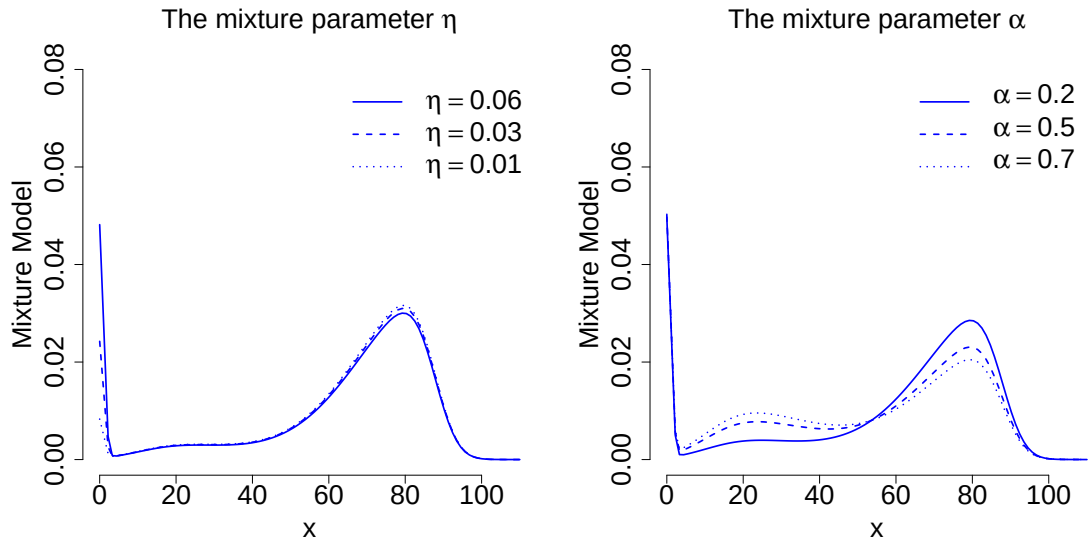


Figure 2.7: Probability density function of the mixture model considering different values for η and α, respectively.

The distribution of deaths is only defined for positive values (the ages x) and it is bounded in the interval $x \in [0, 110]$, but the functions we used for the model spread their probability in a larger interval: the support of the Half Normal distribution is $x \in [0, +\infty]$, while for the Skew Normal distribution we have $x \in \mathbb{R}$. Therefore, we need to verify if the mixture model can be considered a proper distribution for our data. This means that, in the interval $[0, 110]$, the probability of the estimate curve has to be close to 1. We compute

$$P(0 \leq X \leq 110) = \int_0^{110} f(x,\theta)\, dx \tag{2.6}$$

for different countries and years. The minimum value we found is 0.9976, so we can use our model assuming it is a proper distribution for the data, so that

$$\int_0^{\Omega} f(x,\theta)\, dx = 1. \tag{2.7}$$

.

2.1.1 Estimation of the model parameters

We use Maximum Likelihood to estimate the parameters of the mixture model. The data available are in aggregate form: we do not know the age of death for every individual, but the number

of deaths in every age interval (see Figure 2.8). Likely the intervals are disjointed and mutually

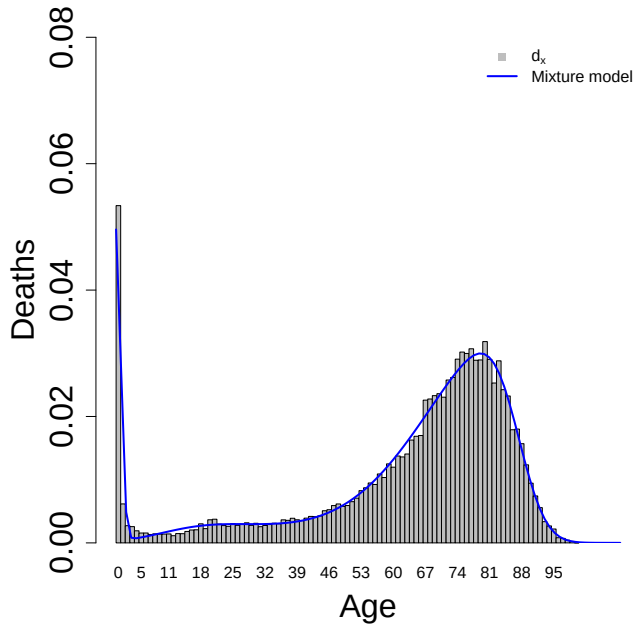


Figure 2.8: Maximum likelihood estimates for the mixture model.

exclusives (individuals die only once) and space partitioning (they cover all the life span). Therefore, since we are modeling the probability of the number of deaths, that occur in the age interval $[x, x+1)$, the multinomial distribution is appropriate. The likelihood that follows is:

$$L(\theta; d_x) = \prod_{x=0}^{\Omega} p(x;\theta)^{d_x},$$

where d_x are the deaths of the mortality table, Ω is the last age at death and $p(x;\theta)$ corresponds to the number of deaths in the interval x and $x+1$

$$p(x;\theta) = \int_x^{x+1} f(t;\theta)\, dt.$$

2.1.2 Demographic interpretation of the parameters

The first mixture parameter η is the intensity of infant mortality and it is related to q_0. Moreover, this value is also associated with the variance of the first part of the distribution of deaths, which explains how quickly the child mortality decreases. Considering Equation (2.2) the variance of the Half Normal distribution is

$$V_{f_I}(X) = \sigma^2 \left(1 - \frac{2}{\pi}\right). \tag{2.8}$$

Since $\sigma = 1$ and $V(\eta \cdot X) = \eta^2 V_{(X)}$, the decrease of mortality during childhood can be computed with

$$\eta^2 \left(1 - \frac{2}{\pi}\right), \tag{2.9}$$

which depends only on the parameter η: if it value is large also the variance is large, which means that the decrease of childhood mortality is slow, the opposite if its value is small.

The second mixture parameter α indicates the importance of the premature mortality: if it is close to 0, accidental and premature mortality are not so relevant.

The three parameters of f_m are: ξ_m (location parameter), which is related to the value of the second mode; ω_m (scale parameter), that is correlated with the variance of the distribution, so if its value is small the premature mortality is concentrated on some ages, while if its value is big, we obtain a vary flat function, which means that the premature deaths affect an ample interval age. The last parameter is λ_m: if its value is positive, we obtain a skewness on the right, otherwise the skewness is on the left. This parameter permits to understand where accidental and premature deaths are concentrated. Moreover the combine analysis of ξ_m and λ_m allows to determine if in the death distribution there is the accident hump: ξ_m small and λ_m not equal to 0.

We also have three parameters for f_M: ξ_M which is related to adult mode; ω_M says how much adult deaths are concentrated around the late mode and it can be seen as a measure of adult mortality compression. The parameter of skewness is λ_M. We expect to observe negative values for this parameter because, usually, the adult distribution of deaths shows an asymmetry towards the youth ages (left). Furthermore, if its value is near to 0 the adult distribution is close to the symmetry, otherwise we observe skewness. Moreover the value of λ_M permits to verify Pearson's theory: if it is significantly different from 0, the distribution of adult deaths is skew as Pearson wrote, otherwise Lexis' theory of normality death is validated.

Premature mortality in the mixture model

The study of premature mortality is a difficult issue. This component is not easy to distinguish in the distribution of deaths by age. There is not a visible breaking point or a range of ages, which give as some indication about the position of this distribution. For instance, for adult mortality we have the late mode and for childhood mortality we know that it starts at age 1, but for premature mortality there are no indications. The problem is that the area of the premature mortality partially overlaps the area of adult mortality, so that the two components seem to be an unique distribution.

Some authors, that modelled the force of mortality μ_x, improved their models adding a function to capture the accidental hump around 20 years old (Heligman and Pollard, 1980; Kostaki, 1992; Rogers and Little, 1994; Thiele and Sprague, 1871). Unfortunately, this hump is only a part of premature mortality: the accidental one. In fact, taking into account, the function μ_x it is not possible to distinguish between premature and adult mortality: the last part of the curve is treated as a unique block and fitted with one distribution, even if its force is decided by the sum of two components.

Indeed, only with the study of the distribution of deaths by age, we can try to distinguish these two components. But, now, the problem is how to determinate it. According to Lexis, premature mortality is described as a transition region, defined in relation to infant and adult mortality. So it is considered as consequence of two others components, without proper features. This makes hard its study (which characteristics can we investigate if premature mortality has no proper features?) and not real interesting (why should we analyze something that depends on something else? We can just study infant and adult mortality).

The novelty idea of our model is to characterized premature mortality using a distribution, as Pearson suggested. In this way we can characterized premature mortality with its mode, its mean, its variance, its skewness... and study it independently. What the model does not investigate are the causes of premature mortality. This problem is still an open question. The answer can be

useful to examine better this mortality component and to understand its transformations in the future.

2.1.3 Advantages of the model

An advantage of this model is that we can decompose, in explicit form, the contribution to life expectancy at birth of the three different components: infant, adult, accidental and premature mortality. In fact e_0 is the mean of the distribution and it should be divided into the weighted average of the Half Normal distribution and the means of the Skew Normal distributions multiplied by the constants η and α:

$$\begin{aligned} e_0 &= \int_0^\Omega x \cdot f(x,\theta)\, dx \\ &= \eta \int_0^\Omega x \cdot f_I(x;1)\, dx + (1-\eta)\alpha \int_0^\Omega x \cdot f_m(x;\xi_m,\omega_m,\lambda_m)\, dx + \\ &\qquad (1-\eta)(1-\alpha) \int_0^\Omega x \cdot f_M(x;\xi_M,\omega_M,\lambda_M)\, dx \\ &= \eta \left(\frac{\sqrt{2}}{\sqrt{\pi}}\right) + (1-\eta)\alpha \left(\xi_m + \omega_m \frac{\lambda_m}{\sqrt{1+\lambda_m^2}}\sqrt{\frac{2}{\pi}}\right) + \\ &\qquad (1-\eta)(1-\alpha) \left(\xi_M + \omega_M \frac{\lambda_M}{\sqrt{1+\lambda_M^2}}\sqrt{\frac{2}{\pi}}\right) \\ &= e_{0_I} + e_{0_m} + e_{0_M}, \end{aligned} \tag{2.10}$$

where e_{0_I}, e_{0_m} and e_{0_M} denote respectively the contribution of infant, accidental and premature and adult mortality in the calculus of life expectancy at birth.

The total variance of the model can be computed with the following formula:

$$\begin{aligned} V_f(X) &= V\Big[\eta f_I(X) + \alpha(1-\eta) f_m(X) + (1-\alpha)(1-\eta) f_M(X)\Big] \\ &= \eta^2 V\Big[f_I(X)\Big] + \alpha^2(1-\eta)^2 V\Big[f_m(X)\Big] + (1-\alpha)^2(1-\eta)^2 V\Big[f_M(X)\Big], \end{aligned} \tag{2.11}$$

and it can be a measure of rectangularization of the survival curve. No variance implies perfect rectangularization (Horiuchi and Wilmoth, 1998), so bigger is the value resulting from Equation (2.11), smaller is the degree of rectangularization. In addition the variance of the two Skew Normal distributions,

$$V_m = \omega_m^2 \left((1-\eta)\alpha\right)^2 \left(1 - \frac{2\left(\frac{\lambda_m}{\sqrt{1+\lambda_m^2}}\right)^2}{\pi}\right) \qquad \text{and} \tag{2.12}$$

$$V_M = \omega_M^2 \left((1-\eta)(1-\alpha)\right)^2 \left(1 - \frac{2\left(\frac{\lambda_M}{\sqrt{1+\lambda_M^2}}\right)^2}{\pi}\right), \tag{2.13}$$

can be interpreted respectively in terms of horizontalization and verticalization (Cheung et al., 2005) of the survival curve. In fact Equation (2.12), together with the variance of infant mortality (Equation 2.9), indicates how many deaths occur before adult mortality, while the second shows how concentrated the deaths are around the modal age at death.

An important measure of longevity used to understand mortality changes is the modal age at death (Canudas-Romo, 2008; Cheung et al., 2005; Horiuchi et al., 2013; Missov et al., 2015; Bergeron-Boucher et al., 2015). For our model it is possible to identify 3 different modes: h_I related to infant mortality, h_m for accidental and premature mortality, and h_M the adult modal age at death. The HN distribution, describing infant and child mortality, has always its mode at age 0, and its level is very easy to compute:

$$h_I = \eta \cdot f_I(0,1) = \eta \frac{\sqrt{2}}{\sqrt{\pi}} \exp\left(-\frac{0}{2}\right) = \eta \frac{\sqrt{2}}{\sqrt{\pi}}. \tag{2.14}$$

For the modes of the Skew Normal distribution, which describe respectively accidental and premature, and adult mortality, we have

$$h_m = (1-\eta) \cdot \alpha \cdot f_m(x_{h_m}; \xi_m, \omega_m, \lambda_m), \tag{2.15}$$

$$h_m = (1-\eta) \cdot (1-\alpha) \cdot f_M(x_{h_M}; \xi_M, \omega_M, \lambda_M), \tag{2.16}$$

which required numerical method to calculate $f_m(x_{h_m})$ and $f_M(x_{h_M})$.

It is possible to split the area under the distribution of deaths using the functions involved in the mixture model (see Figure 2.9). This permits to calculate the percentages of deaths related to

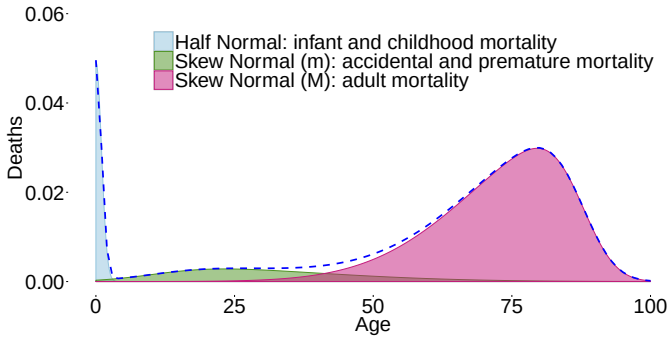


Figure 2.9: The different areas that compose the death distribution.

infant and child, adult, accidental and premature mortality. The infant mortality area (A_I) can be measured with the integral:

$$A_I = \int_0^{\Omega} \eta \cdot f_I(0,1)\, dx = \eta, \tag{2.17}$$

where Ω is last age at death. We can assume that the HN distribution spreads all its probability in the interval $[0, \Omega]$, so its integral will be equal to 1. The premature mortality area (A_m) is calculated as the integral of f_m between 10 and Ω.

$$A_m = \int_{10}^{\Omega} \alpha(1-\eta) \cdot f_m(x; \xi_m, \omega_m, \lambda_m)\, dx. \tag{2.18}$$

We chose 10 years old as a boundary between childhood and adolescence. In fact around this age we usually observe the minimum number of deaths in the first part of the curve (Livi Bacci, 1983). Looking at Figure 2.9 we can see that the probability mass of f_m can cover also part of childhood. This happens in particular when infant mortality is high. To avoid the problem to include childhood deaths in the count of accidental and premature mortality, we decided to used

Mortality	Area
Infant	$A_I = \eta$
Premature	$A_m = 1 - (A_I + A_M)$
Adult	$A_M = (1-\eta)(1-\alpha)$

Table 2.2: Areas of the different components for mortality components.

age 10 as marker.
The adult mortality area (A_M) is:

$$A_M = \int_0^{\Omega} (1-\eta) \cdot (1-\alpha) \cdot f_M(x; \xi_M, \omega_M, \lambda_M)\, dx = (1-\eta)(1-\alpha), \tag{2.19}$$

because, again, we consider the integral of f_M between 0 and Ω equal to 1. In this case the lower boundary in Equation (2.19) is equal to 0 because the probability mass of f_M does not involve infant and childhood mortality.

We summarize the results in Table 2.2.

2.2 A simplification of the mixture model

In order to reduce the number of parameters we propose another mixture model, very close to the previous one in some contexts, but also with some limitations. Again for infant mortality we employ the HN distribution. In this case its scale parameter is free because, reducing the number of parameters, there will be less identification problems. The second part of the curve (accidental, premature and adult mortality) is modelled using a "generalization" of the Skew Normal distribution, called Skew Bimodal Normal (SBN) distribution (Rocha et al., 2013). It has the following pdf

$$f_A(x; \alpha^*, \xi^*, \omega^*, \lambda^*) = 2\omega^{*-1} \left[\frac{1 + \alpha^* \left(\frac{x-\xi^*}{\omega^*} \right)^2}{1+\alpha^*} \right] \phi\left(\frac{x-\xi^*}{\omega^*} \right) \Phi\left[\lambda^* \left(\frac{x-\xi^*}{\omega^*} \right) \right], \tag{2.20}$$

where ξ^* is the location parameter, ω^* the scale parameter, λ^* the shape parameter and α^* regulates the number of the modes in the distribution. In fact, the SBN can have at most two modes: if $\alpha^* \geq 0$ the pdf has two modes, otherwise ($\alpha^* < 0.5$) the pdf is unimodal (Elal-Olivero et al., 2009). In Figure 2.10 the behavior of the SBN is shown for different values of α^*, considering $\xi^* = 0$ and $\omega^* = 1$ fixed. When α^* increases, the second mode becomes more visible. The form of the function is compatible with the second part of the distribution of deaths. Moreover if the mode of accidental mortality is not present in the data, the model ignores it. This is an advantage: in Heligman and Pollard model accidental mortality is always estimated, even if there is not the necessity. In this model, with only one parameter we can estimate extra deaths, if their frequency is relevant enough. By mixing Equation (2.1) and (2.20) we obtain the following model (Mazzuco et al., 2016)

$$f*(x; \theta^*) = \eta^* f_I(x, \sigma) + (1-\eta^*) f_A(x; \xi^*, \omega^*, \lambda^*, \alpha^*), \tag{2.21}$$

where $\theta^* = (\eta^*, \sigma, \xi^*, \omega^*, \lambda^*, \alpha^*)$ is a 6 parameter vector. The mixture parameter η^* indicates the weigh of the two functions, so it is related with the importance of infant mortality. Maximum

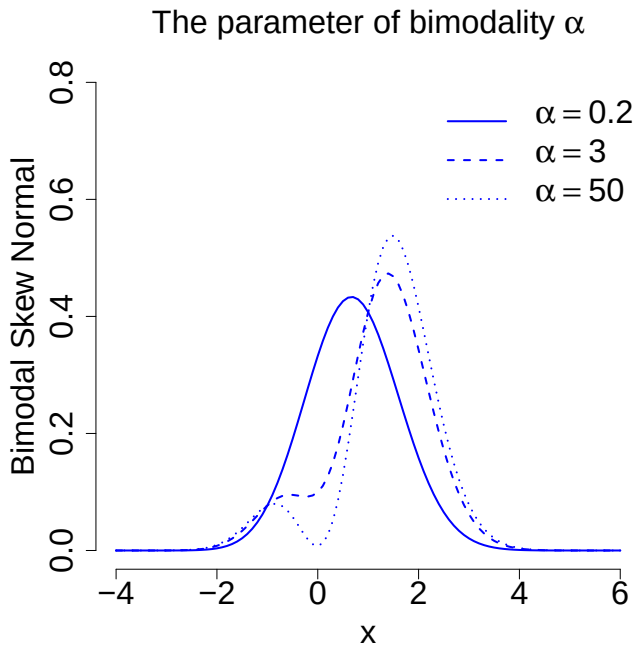


Figure 2.10: Probability density function of the Skew Bimodal Normal distribution as the parameter α^* changes.

likelihood is used to estimate the vector θ^*, and the strategy adopted to calculate parameter values is the same expressed before in Equation (2.1.1).

To facilitate the interpretation of the model, the author link every parameter to demographic features. The mixture parameter η^* indicates the relative weight of infant and child mortality. The scale parameter of the Half Normal is associated with the decline mortality in childhood: high values of σ imply that child mortality (and not only infant mortality) is high. The value of parameter α^* is connected with the presence of the accidental hump: if it is greater than 0.5 the hump is detected by the mixture model and the higher the value of α, the more pronounced the hump. The SBN scale parameter ω^* controls the concentration of deaths around the late modal age at death: if its value is small the deaths are concentrated around the late modal age at death. The other shape parameter of SBN monitors the distribution skewness: if λ^* is positive the pdf is skewed on the right, while if λ^* is negative the skew is on the left. As for interpretation, skewness indicates where the mass of distribution is more concentrated: a left skew means that most of the deaths are concentrated before the modal age. Finally, ξ^* is the location parameter of SBN and it is strictly related with average life duration of adults. Considering that life expectancy at birth is particularly sensitive to infant mortality level, ξ^* is associated with life expectancy at 10 years old.

Moreover the model can be useful to detect the level of rectangularization in the survival curve. In particular, using the concept of horizontalization and verticalization (Cheung et al., 2005), we can relate the mixture parameter η^*, the HN scale parameter σ and the SBN α^* parameter with the degree of horizontalization. Whereas the SBN scale parameter ω^* can be taken as a measure of verticalization: the lower the variance, the steeper the decline of survival curve in old age.

The model has been fitted to several mortality data with very different shapes of death distribution. It is flexible enough to fit mortality both in pre and post transitional contexts, with or without

accidental hump. Moreover Equation (2.21) is more parsimonious than the complete mixture model and this makes it easier to fit. Even if the goodness of fit is satisfactory in many cases, there are some limitations and disadvantages. In Figure 2.11 the curve estimated using Equation (2.21) is plotted. Sometimes the model is unable to fit correctly d_0: in France 1977 infant mortality is

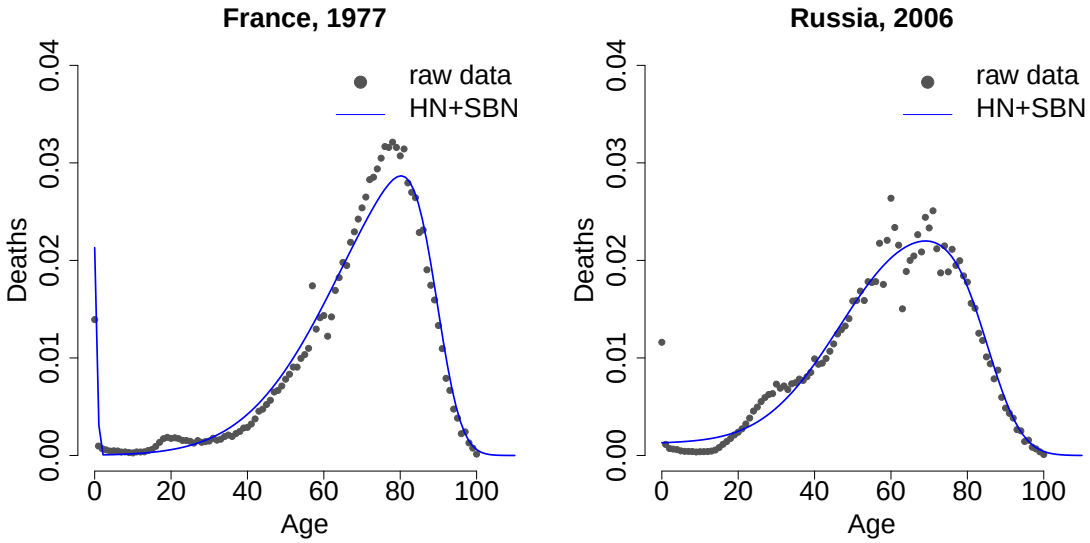


Figure 2.11: Model fitted for France 1977 and Russia 2006.

overestimated, while in Russia 2006 it is underestimated. The difficulty to model in a satisfactory way the first point of the death distribution was also found for the mixture model we presented before. An easy solution can be to fix the values of the shape parameter of HN distribution. A good candidate should be 1: this value was employed in the mixture model (2.5) with good results, so also in this case it can be useful. The second problem regards the estimate of accidental and premature mortality. In the graph of France the accidental hump is not accurately fitted and also the mode of the distribution is not well modeled, in particular on the left side. In Russia we observe that accidental mortality is not captured, even if it is visible. Also the mode of the distribution seems to be underestimated. The reason why this happens is that the SBN distribution is not a generalization of the SN distribution, but it is a mixture of SN distributions. Considering $\xi^* = 0$ and $\omega^* = 1$ we can write

$$\begin{aligned} f(x;\alpha^*,\lambda^*) &= 2\left(\frac{1+\alpha^* x^2}{1+\alpha^*}\right)\phi(x)\Phi(\lambda^* x) \\ &= \frac{1}{1+\alpha^*}2\phi(x)\Phi(\lambda^* x) + \frac{\alpha^*}{1+\alpha^*}x^2 2\phi(x)\Phi(\lambda^* x) \\ &= \frac{1}{1+\alpha^*}SN(\lambda^*) + \frac{\alpha^*}{1+\alpha^*}x^2 SN(\lambda^*). \end{aligned} \tag{2.22}$$

We can see that the SBN is composed by a SN distribution and a x^2SN distribution. The mixture parameter is α^*. It is clear that the second mode is due to x^2SN: if $\alpha^* = 0$ we obtain only a SN distribution. Moreover, analyzing Equation (2.22) the two functions involved have the same values for all the parameters. This restriction is a disadvantage because it limits the model flexibility and its adaptability. Moreover the skewness parameter λ^* has the same sign for both the functions, so the asymmetry side is the same. This implies that if λ^* is negative and $\alpha^* \geq 0$ we will obtain an extra mode on the right side, the contrary if λ^* is negative (see Figure 2.12). Unfortunately the

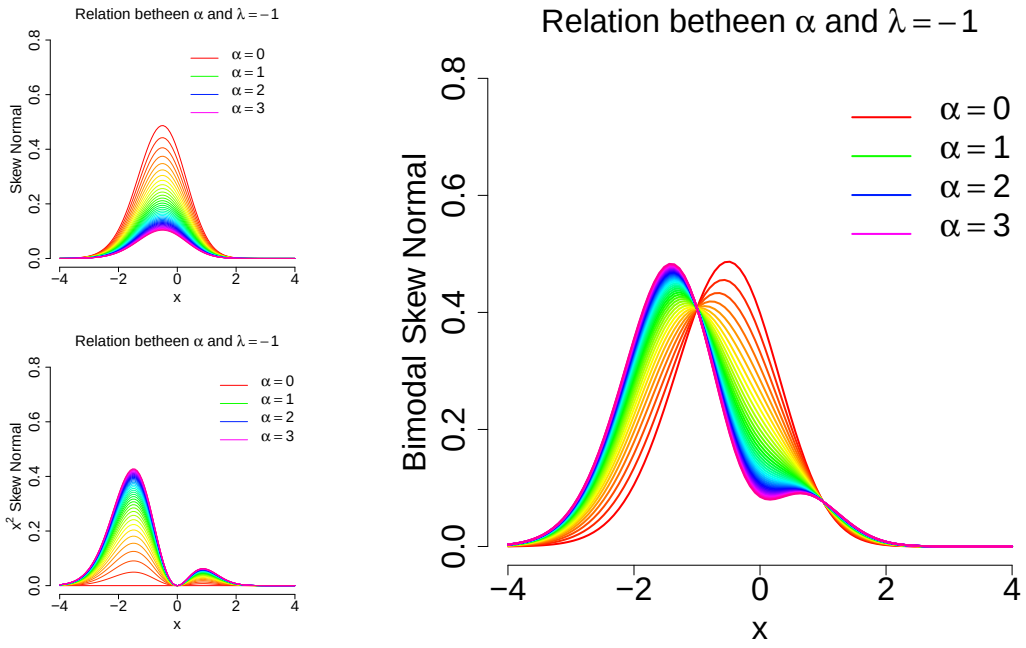


Figure 2.12: Relation between α^* and λ^* in the SBN distribution.

death curve shows a left asymmetry ($\lambda^* < 0$) for adult mortality, but accidental and premature mortality are arranged on the left, so they need a positive value for λ^*. This required can not be satisfied by the model and it is the reason why, in some situations, the fit for accidental and premature mortality is inadequate. Moreover, when the second hump is pronounced and the model fit it, the estimates of SBN parameters are necessarily a compromise: the values in the two functions involved are the same. Finally if the aim of the study is to quantify premature mortality it is better to employ the model we discussed above.

Chapter 3

Results

We employed the data of the Human Mortality Database (HMD) to test the mixture model. This has been fitted to Sweden, France, Italy, Spain, Portugal, East Germany, Russia, Ukraine, Czech Republic, Hungary and USA. These countries have been chosen because they are representative of different mortality patterns we observe for developed countries. For all of them we compute the d_x using raw data as we want to avoid to fit the model to already modelled data. For Spain and Portugal, the raw data of population size is available only for few years (census years). In these cases we determined the d_x employing the population estimated by the HMD.

3.1 Model testing

To test the model, the raw death counts of Sweden from 1910 to 2011 are used. This country has a long time series and good raw data, even in the past century. It is a furerunner country in terms of reduction of mortality, so it well represents the entire mortality history of most European countries. For this reason, we can suppose that the model can be used for other countries which have experimented the same transformation in mortality (even if in different years than Sweden) and show similar Sweden mortality trend. However, data from other countries have been used, in order to check the model behaviour with some particular patterns that we cannot find in Swedish data. For example, a high level of mortality due to external causes can be found in several eastern European countries, leading to a particularly skewed distribution of deaths. In Figure 3.1 we show the fitted model. With the black solid line the overall mixture model is drawn. The blue dotted line highlights the fit of the Skew Normal employed to estimate accidental and premature mortality. The big dots point out the three modes of the distribution. In Appendix A more estimated years are reported. For each year, the curve of the mixture model is very close to the observed data. There is not a part of the deaths distribution which is not well approximated by the mixture. In addition, the curves are smooth, so the mixture model is able to capture the mortality trend without excessively suffering for random data oscillations.
To evaluate the model in a more formal way, we compute the absolute value of the differences between the real death counts d_x and the estimated ones $\hat{d_x}$, as follows:

$$Errors_{year} = \sum_{x=0}^{\Omega} |d_x - \hat{d_x}|. \tag{3.1}$$

The boxplot of the computed errors is shown in Figure 3.2. We can conclude that our model is able to capture different mortality paths because the errors committed are in general small:

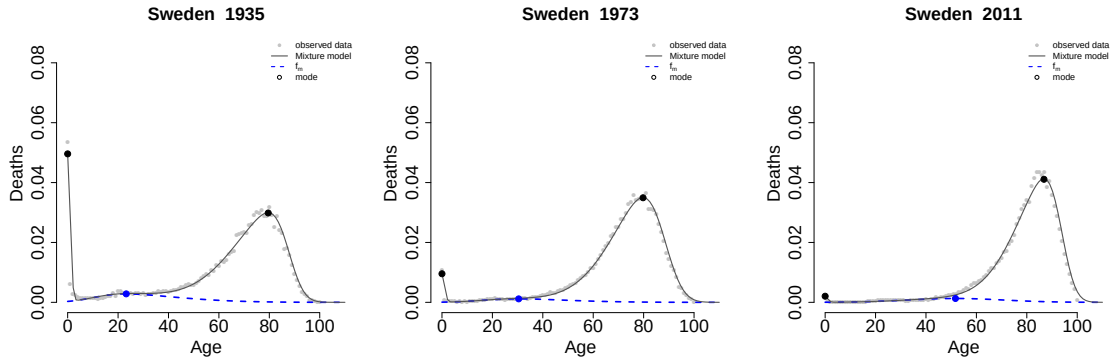


Figure 3.1: Model fit for Sweden in 1935, 1973 and 2010.

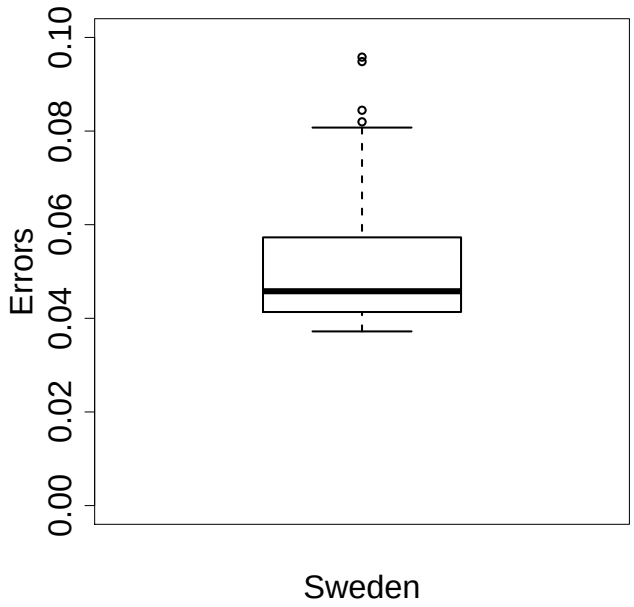


Figure 3.2: Boxplot of the distribution of the errors committed by the mixture model.

interquartile range includes into (0.041; 0.052) and median equal to 0.046. The maximum error is registered in 1910 and its value corresponds to 0.096, which is still small.

In order to check the performance of the mixture distribution, we compare it with Siler and Heligman and Pollard functions, which are also mortality models for the whole age range. Since for every outcome of the considered models the reconstruction of the life table is possible, we calculate the estimated death distribution and we compare them in Figure 3.3. In 1935 Siler and the mixture

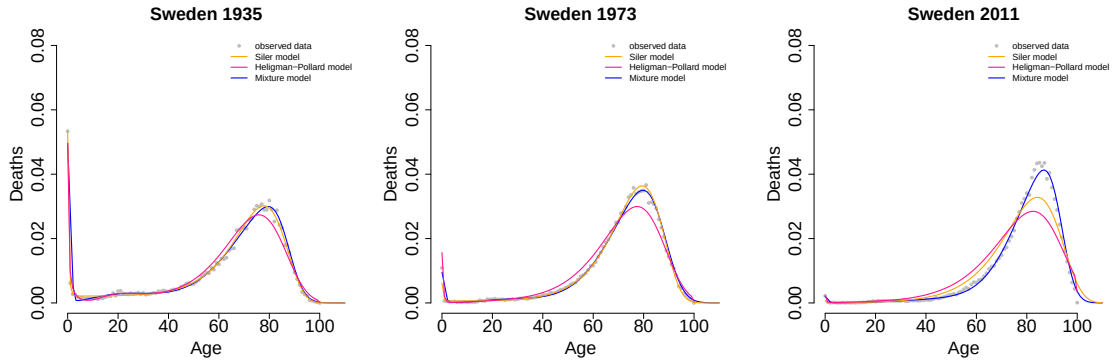


Figure 3.3: Models fit for Sweden in 1935, 1973 and 2010.

curve are close. Heligman and Pollard model shows a problem of balancing of the distribution, which is not right centered on the real adult mode. In this year, the mixture model is the worst in term of infant and childhood mortality, in particular considering the fit of childhood deaths. For 1973, Siler and the mixture model are again very similar, while Heligman and Pollard's curve underestimates the deaths of premature and adult mortality, even if it is centered on the late mode. In 2011 there is a clear distinction between the three curves and only the mixture model is able to fit the peak of adult mortality. Only looking at the graphs, we can say that Siler and our model have similar performance except for recent years, when Siler is not able to right estimates the last part of the distribution of deaths.

The three models are different in terms of mathematical expression, outcome and number of parameters, so a direct test in term of log likelihood is not correct. The Akaike information criterion (AIC) provides the comparison among not nested models (Akaike, 1974, 1998). Its expression takes into consideration the value of the likelihood function and the number of parameters employed:

$$AIC = 2p - 2\log L(x, \hat{\theta}) \tag{3.2}$$

where p is the number of parameters and $L(x, \hat{\theta})$ is the value of the likelihood computed using the estimated parameters $\hat{\theta}$ (maximum likelihood). Considering the amount of parameters is important in comparison because there is a trade off between their counts and the goodness of fit: when the number of parameters grows also the goodness of fit increases, even if the addition of parameters does not provide a statistic improvement of the fit. In AIC criterion, the number of parameters penalizes maximum likelihood estimates, so it discourages overfitting. The preferred model is the one with the minimum AIC value. For every model we calculate the AIC criterion in the considered year. The results are plotted in Figure 3.4. In the first part of the graph the model performances are very similar and there is not one better then another. Since 1970 we start to note some differences and Heligman and Pollard model proves to be the worst, even if the distance among curves is very small. The performances of the mixture model appears very close to Siler

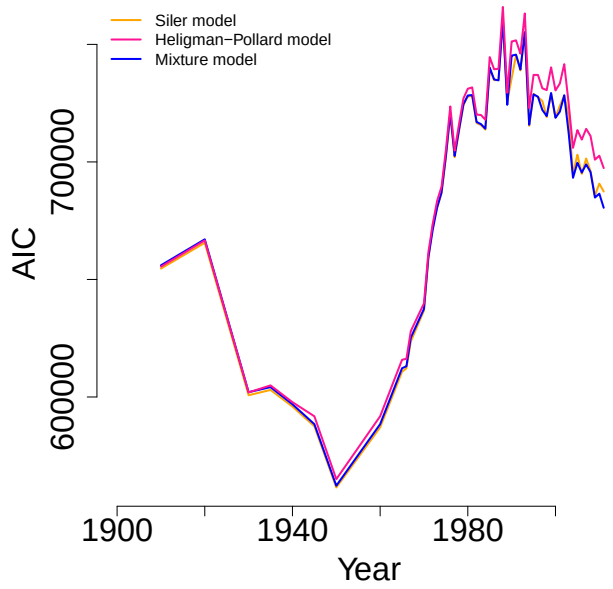


Figure 3.4: AIC values for Siler, Heligman and Pollard and Mixture model from 1910 to 2911.

model. This means that our model really improve the fit adding 3 parameters and it is better than Heligman and Pollard model, which has the same number of parameters.

In addition to Sweden, we estimate the mixture model for France, Italy, Spain, Portugal, East Germany, Russia, Ukraine, Czech Republic, Hungary and USA. For these countries the following years are available:

- Sweden: 1910, 1920, 1930, 1935, 1940, 1945, 1950, 1960, 1965, 1966, 1967 and from 1970 to 2011;
- France: from 1900 to 2013;
- Italy: form 1915 to 1920, 1931, 1936, from 1940 to 1951 and from 1982 to 2012.
- Spain: from 1961 to 2012;
- Portugal: from 1970 to 2012;
- East Germany: from 1956 to 1969 and from 1971 to 2011;
- Russia: from 1970 to 2010;
- Ukraine: from 1970 to 2013;
- Czech Republic: from 1947 to 2014;
- Hungary: from 1950 to 2009;
- USA: from 1959 to 2013.

For every year and every country the model performance was evaluated comparing the the estimated $\hat{d_x}$ with the life table deaths provided by the HMD (d_x^{HMD}):

$$Errors_{year} = \sum_{x=0}^{\Omega} |d_x^{HMD} - \hat{d_x}|. \tag{3.3}$$

In Figure 3.5 the trends of the selected countries are plotted. In general, the errors committed

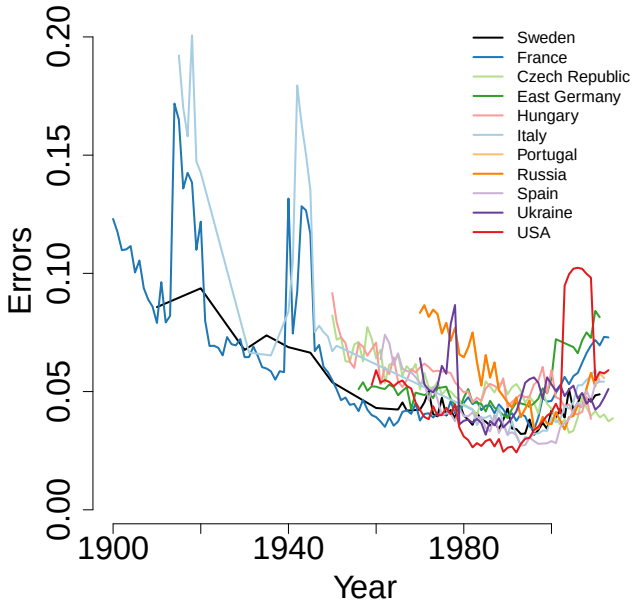


Figure 3.5: Errors committed by the mixture model for the examined countries in different years.

are small with a decreasing trend. The pattern of all the countries is similar, which means that the overall model fit is in line with the HMD. In the last years we observe a light increment of the overall errors, however very contained. In the first part of the graph, in correspondence of the years of the two world wars, we observe two peaks both for France and for Italy. These increments depend to the large discrepancy between the row death counts and the one reconstructed by the HMD, in correspondence of the age involved in the military draft (18-40 years old). In particular, the HMD evaluates a lower number of deaths than the one found in the row data. Between 2000 and 2009 an increase of the error is registered for the USA. In fact, in these years, the model is not able to capture the right shape of the distribution of deaths because the last open age class is to wide (85+). In particular, we observe a bad fit of the Skew Normal distribution which models the adult mortality component, as it is possible to see in Figure 3.6. In Chapter 4 we will discuss how to increase the parameter estimates with the employment of the EM algorithm. This technique is useful when in the data there are some missing value: in our case the missing values are the frequencies of the number of deaths after age 84. Step by step, the algorithm increments the model fit, estimating the missing value and computing the maximization of the likelihood function, until convergence.

Even if for all East European countries the trends of the errors are not different from the others, we find identification problem. In fact, although the overall fit is good, there are gaps in the values of parameter estimates, which can not be compatible with their trends. This is particularly evident

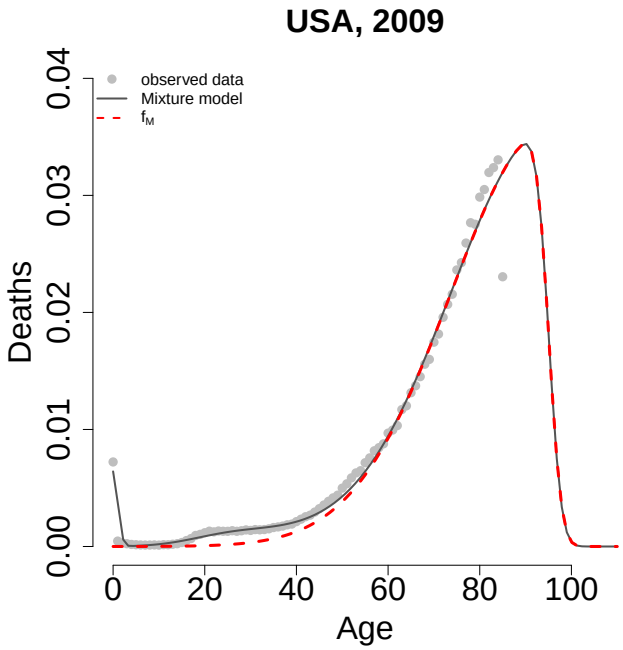


Figure 3.6: Fit of the mixture model for the USA in 2009.

for the distribution of accidental and premature mortality. An example is reported in Figure 3.7, which show how the shape of f_m changes for two neighbor years. This problem seems to be related to the presence of two modes in the middle part of the distribution of deaths: the accidental hump and an increase of premature mortality before the late mode. This origins a bimodal distribution, which the SN f_m is not able to cover.

In the next sections we are going to show the result obtained for male population of Sweden, France, Est Germany and Czech Republic. Italy, Spain and Portugal are very close to France estimate. Unfortunately, Italy has not good raw data: in some years the number of deaths are greater than the exposure population, so we conveniently fix the denominator. Furthermore the estimates are very fragmentary because, to compute a satisfactory life tables, some years have been removed.

3.2 Infant Mortality

During the demographic transition, most of developed countries show a reduction of infant mortality. In the course of the last century the incidence of deaths at age 0 continues to decrease. We find the same trend considering the estimates of the mixture parameter η, which is related to the importance of infant mortality. All the countries present the same trend (see Figure 3.8). Even in countries where, at the beginning of the period considered, the estimate of η was already low, we notice a further reduction of infant mortality. As one might expect during the two world wars, in France the level of infant mortality slightly increases: the bad socio-economic conditions cause an increment of the number of deaths at age 0. For recent years, η is very close to 0, which means that infant mortality has a very small incidence in the overall mortality. The reduction of infant mortality is accomplished with a decrease of childhood mortality. To measure this feature

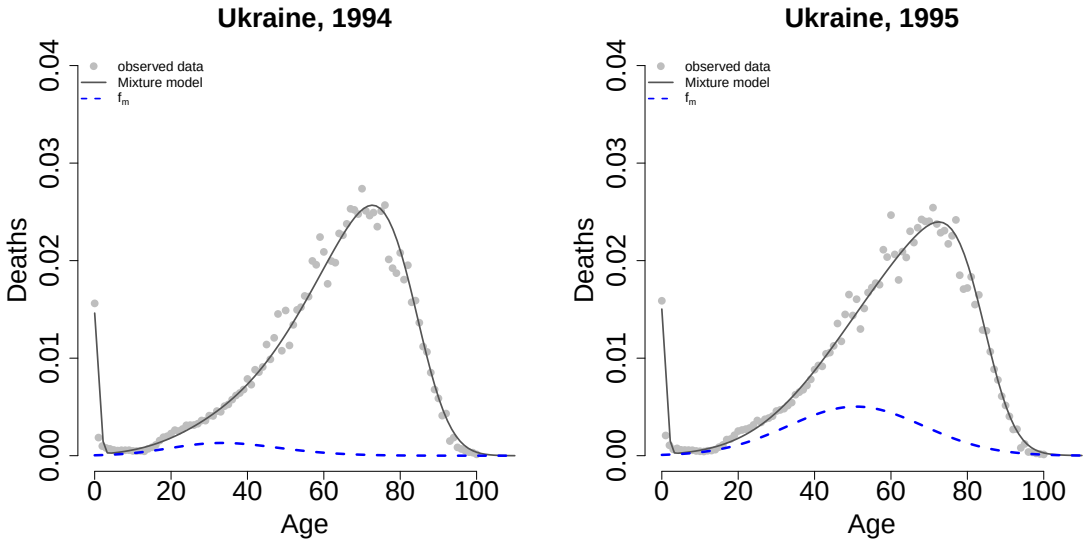


Figure 3.7: Fit of the mixture model for Ukraine in two neighbor years.

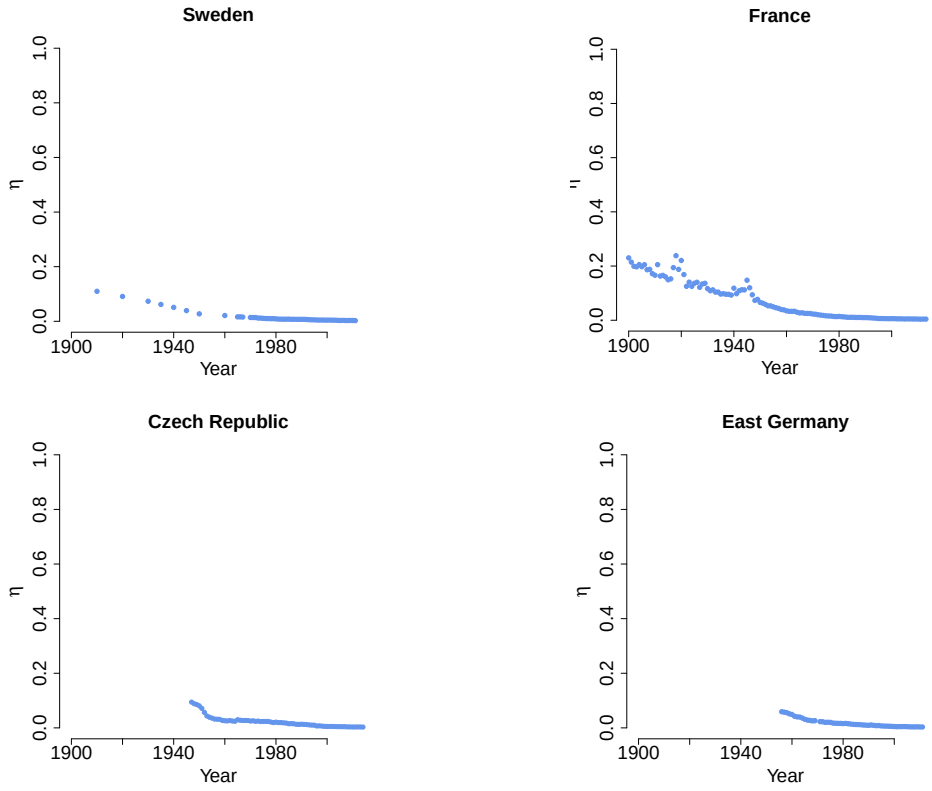


Figure 3.8: Trend of the mixture parameter η.

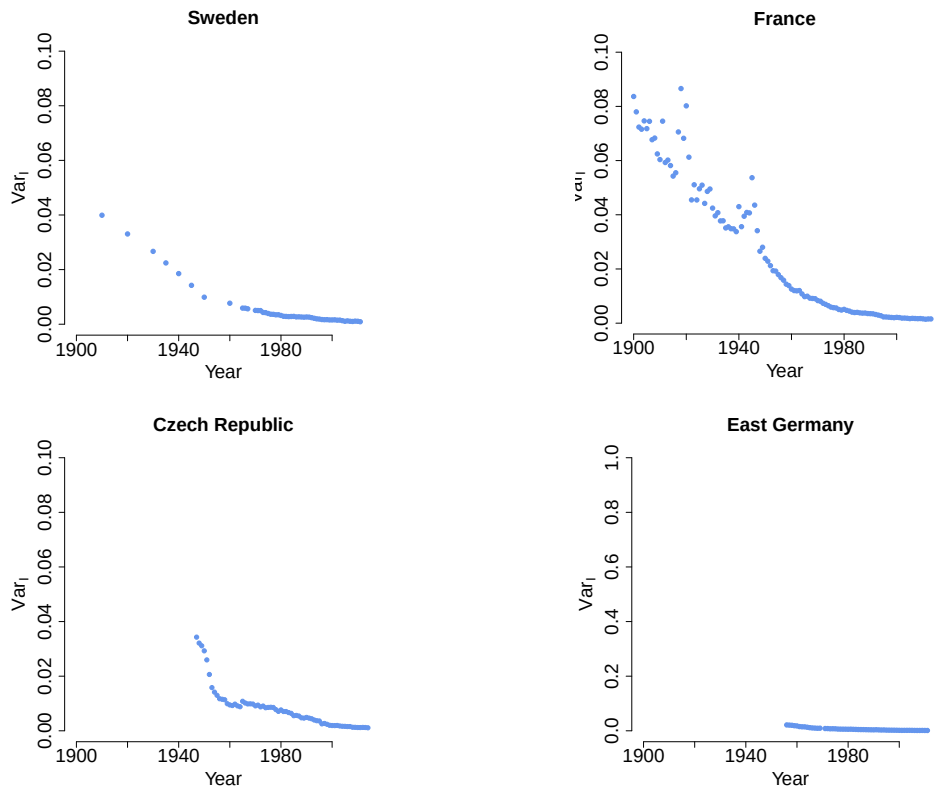


Figure 3.9: Trend of the variance of the childhood mortality.

we use Equation (2.9). As we expected (see Figure 3.9), we observe a diminution of the variance of the first part of the curve, so that its value now is very close to 0. This means that deaths are concentrated on the same point: age 0. Again in France we observe the effects of the world wars on childhood mortality: its level comes back 10 yeas before. Moreover, even if the trend is the same for each country, there are some differences. In Sweden the reduction is not so pronounced as in France. Indeed in Sweden the decrease of mortality began before than in France, so its starting value was already a result of this process. In France the reduction of mortality started after some years, but was faster. In Czech Republic the decrease is very fast during the first years of the time series and then slower. In Germany the level is already low since the first year of observation (1956).

3.3 Adult Mortality

To verify Pearson's theory, we analyze the values of the parameter λ_M. Pearson asserts that adult mortality is a skew distribution. In the SN f_M the parameter responsible of the skewness is λ_M. If it is close to 0 we obtain a symmetric curve. In Figure 3.10 the trend of this parameter is plotted. For all the countries its value is negative and included in the interval $(-5, -3)$, except for France during the two world wars, where the skewness is bigger. It is interesting to see that its value is quite stable in the observed period, which means that the left asymmetry of the adult mortality component is not a changing feature or it changes very slowly. These results confirm Pearson's theory: we need to take into consideration that the deep we observe at the end of death distribution depends on the incidence of the phenomenon before its late mode. The normal death

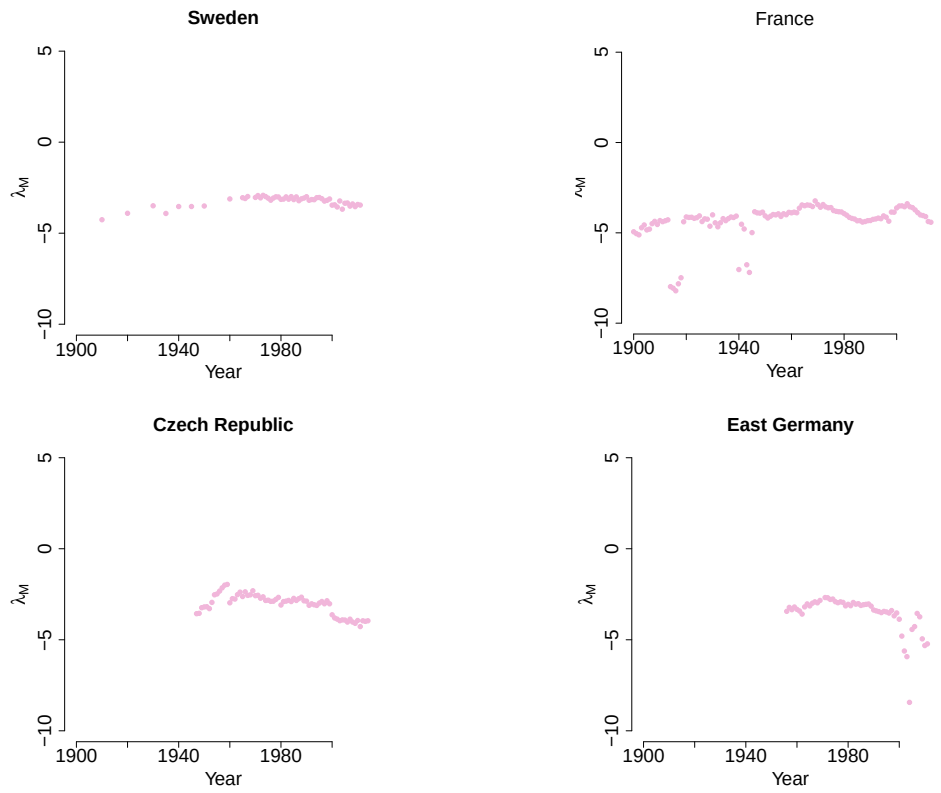


Figure 3.10: Trend of the skewness parameter λ_M.

assumption can be useful in many situations, but it is not confirmed by the data.

The features of adult mortality are diffusely studied. In particular demographers show a shifting and a compression of this component in the last 15-20 years. With our mixture model these characteristics can be studied using the values of the late mode at death h_M, and the trend of the shape parameter ω_m, respectively. In Figure 3.11 we observe that after a period of stability (Sweden, Czech Republic and Germany) or light increase (France), in recent years the mode of the distribution is shifting to the right. In Czech Republic we see that the last value registered is still smaller than the others, which are quite similar. This means that mortality improvements are delayed compared to the other countries. In general, looking at the results we reach the same conclusion of a shifting of the modal age of death.

We use the time series of the values of ω_M to quantify the compression of the adult deaths around the modal age. In Figure 3.12 we observe that there are two trends. For Sweden and France, we can see an almost linear decrease; for Czech Republic after a reduction, the values of ω_M increases; in Germany the path is almost stable. It is true that the variance of the adult mortality depends on ω_M, λ_M and their related mixture parameters (see Equation 2.13), but it is also true that we observed approximately the same trend for all the parameters except ω_M. So, we can say that not all the countries experiment the same contraction degree of mortality during adulthood. Czech Republic and Germany do not seem to experiment this phenomenon, while in Sweden and in France the contraction is relevant.

Finally, with Equation (2.18) we compute the amount of deaths related with adult mortality component. The results are reported in Figure 3.13. As we expected the great part of the deaths occur in adulthood. This proportion increases over time, quickly at the beginning (see Sweden

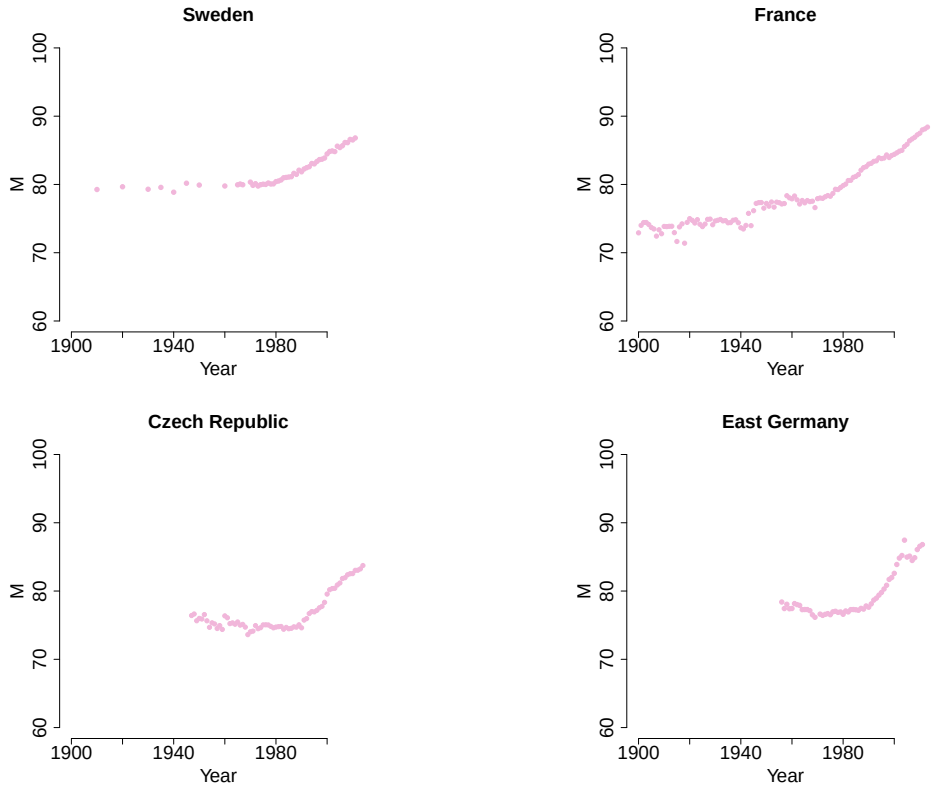


Figure 3.11: Trend of the late mode h_M.

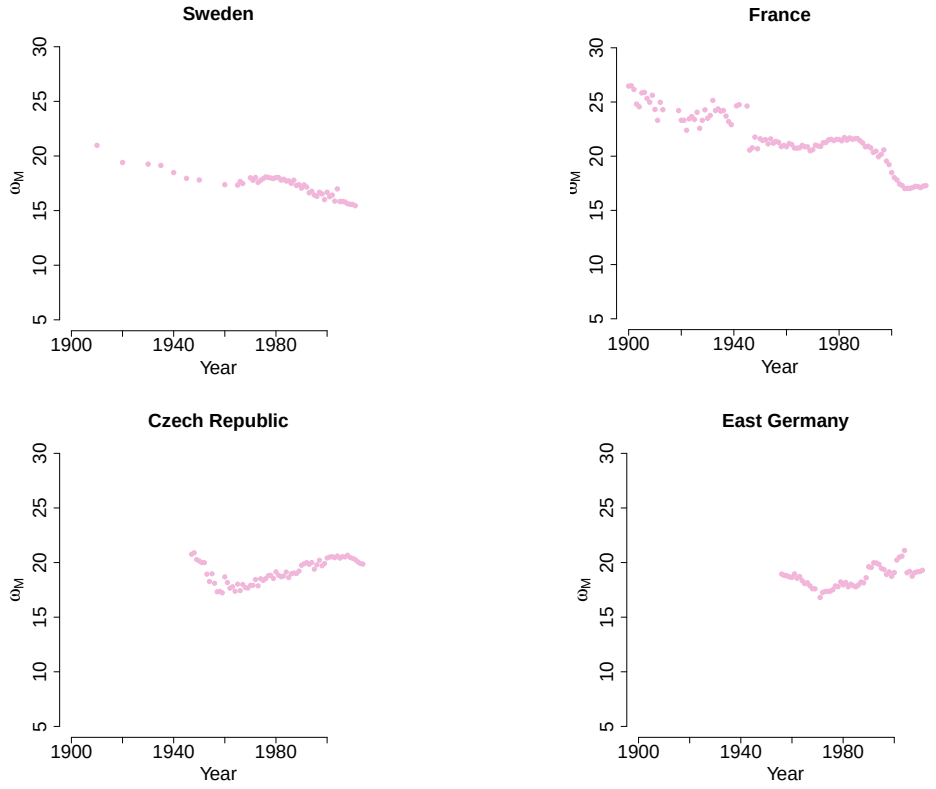


Figure 3.12: Trend of the shape parameter ω_M.

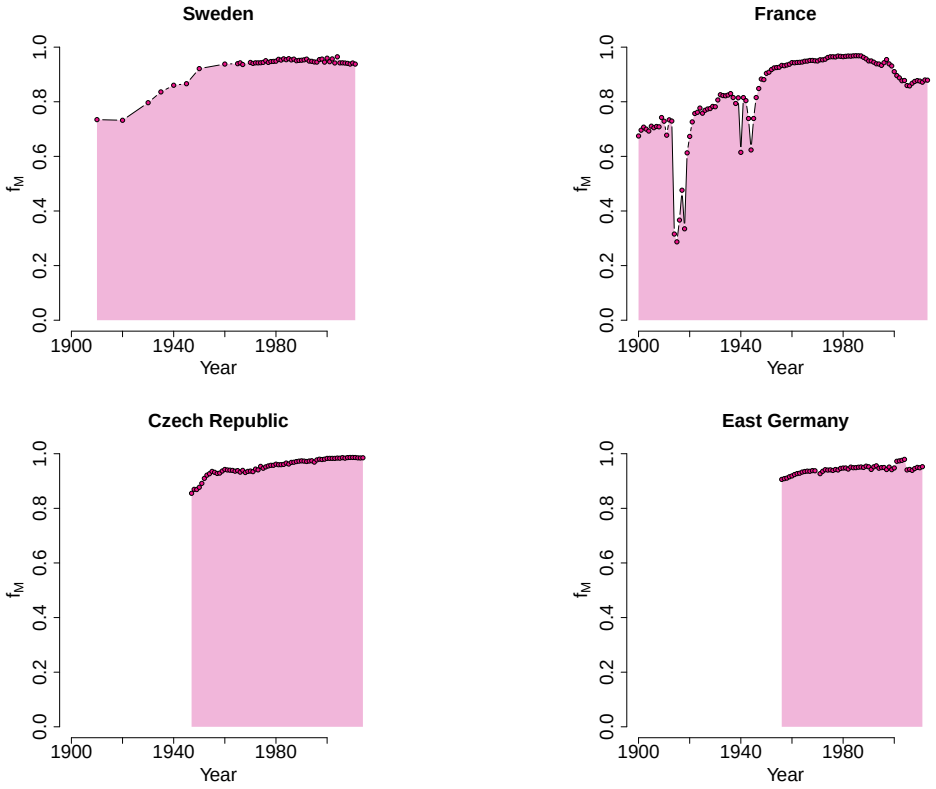


Figure 3.13: Trend of the percentage of deaths related with adult mortality component.

and France) and then slowly. In France the two negative peaks are due to the world wars: in fact in these years the number of premature deaths increases. In particular during the first world war the number of victims was enormous. France shows a particular trend also at the end of the time series: the percentage of adult deaths decreases. This is due to an increase of premature mortality. In the next section we will discuss about this particular feature.

3.4 Accidental and Premature mortality

We focus our attention on f_m, the function that corresponds to accidental and premature mortality. Analyzing its parameters we can study the features of this mortality component, understanding its transformations and the difference among countries. In Figure 3.14 the mode of f_m is presented. For all countries, except Czech Republic, we note an increase of the modal value with an acceleration starting during the last twenty years of the last century. The increment is particularly evident in France. Moreover, for these countries the range of variation is very similar. This means that at the beginning of the observation period the SN f_m captures the mode of the accidental mortality. Progressively, the accidental hump disappears, while the incidence of premature mortality grows. Its mode shifts on the right of the distribution following the same trend of the adult mode. For Czech Republic the value of the mode is quite constant around 25 years old. In fact, across all the considered period, the accidental hump is clearly recognizable and well distinct from the adult mode. Moreover the increase of premature mortality is not registered.
The parameter related with the shape of the distribution is λ_m, indicating the presence of skewness and its side. For all countries, as we can view in Figure 3.15, its values are close to 0, suggesting

Figure 3.14: Trend of the accidental and premature mode h_m.

that the curve is quite symmetric. In France we observe some outliers during the period across the two world wars: in fact, during those years, the distribution of deaths between 18 and 40 year old was very asymmetric because of the increase number of deaths.
For the scale parameter ω_m (see Figure 3.16) we observe different trends. It decreases for Sweden and Czech Republic, for Est Germany is approximately constant, while for France we can see a decline until the end of the 1990, then a strong increment in very few years, followed by a stabilization. For Sweden, France and East Germany the SN f_m becomes less concentrated and more flat: we pass from a situation in which its probability was amassed between 20 and 40 years old, to a context in which its probability is spread across middle life. The decrease of ω_m for Sweden and Germany is explained considering that at the beginning, f_m fits the accidental mortality and also the premature one, so it requires a big variance: it covers a large age interval (youth and first part of adulthood). With the disappearance of accidental mortality, f_m becomes more concentrated because premature mortality is spread in a smaller interval. For Czech Republic ω_m decreases: in this case the probability mass becomes more concentrated around the accidental mode.
To conclude we plotted in Figure 3.17 the percentage of deaths related with accidental and premature mortality. In Sweden we observe a reduction of this percentage until 1960. After a period of stability, from 2000 this rate seems lightly increasing. For France, removing the two peaks due to the world war, we can see a general decline of accidental and premature deaths since 1990. From this year, the trend reverses its trajectory and it reaches bigger values than those observed at the beginning of the time series. Czech Republic shows a decreasing trend with very small percentage of accidental and premature deaths. Germany presents stable number of deaths related to this mortality component.

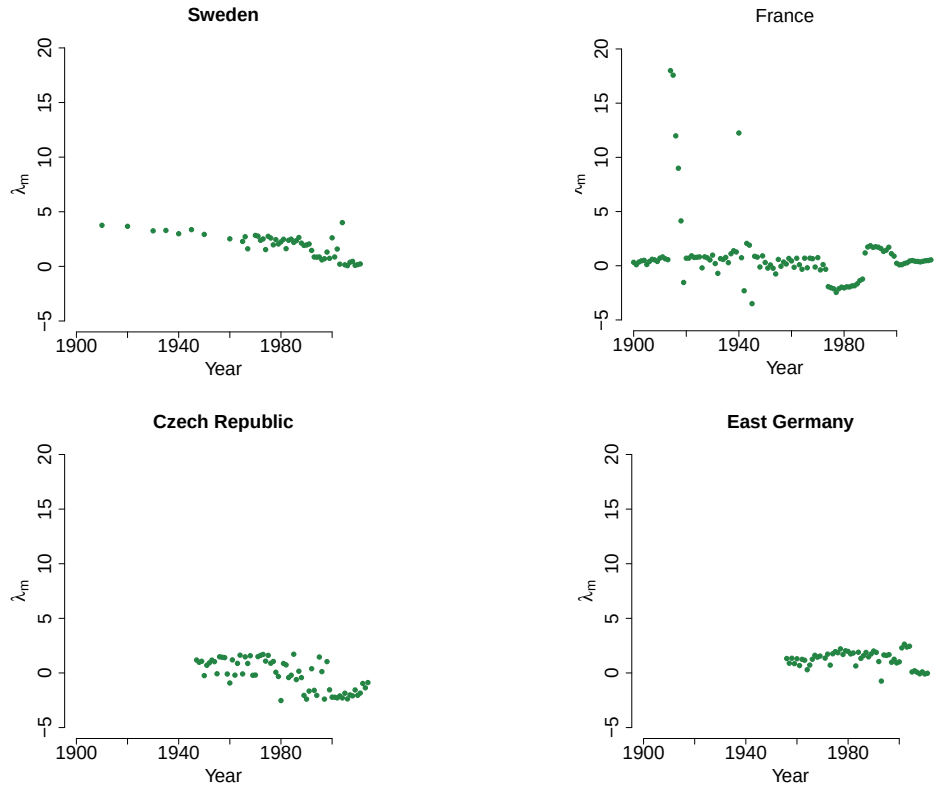


Figure 3.15: Trend of the skewness parameter λ_m.

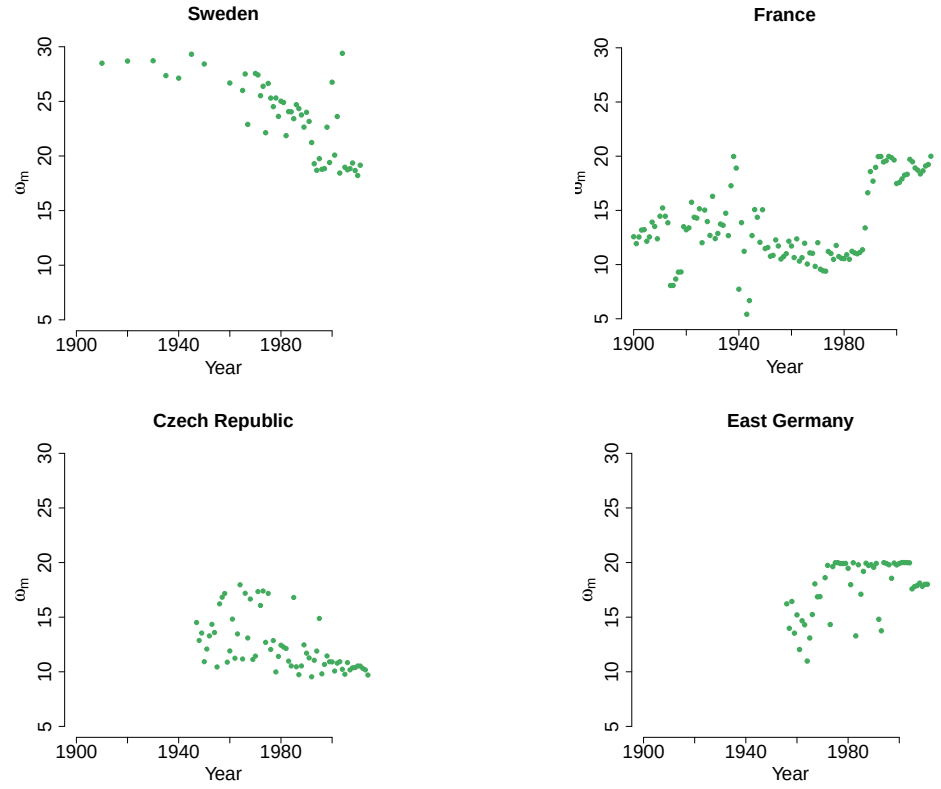


Figure 3.16: Trend of the scale parameter ω_m.

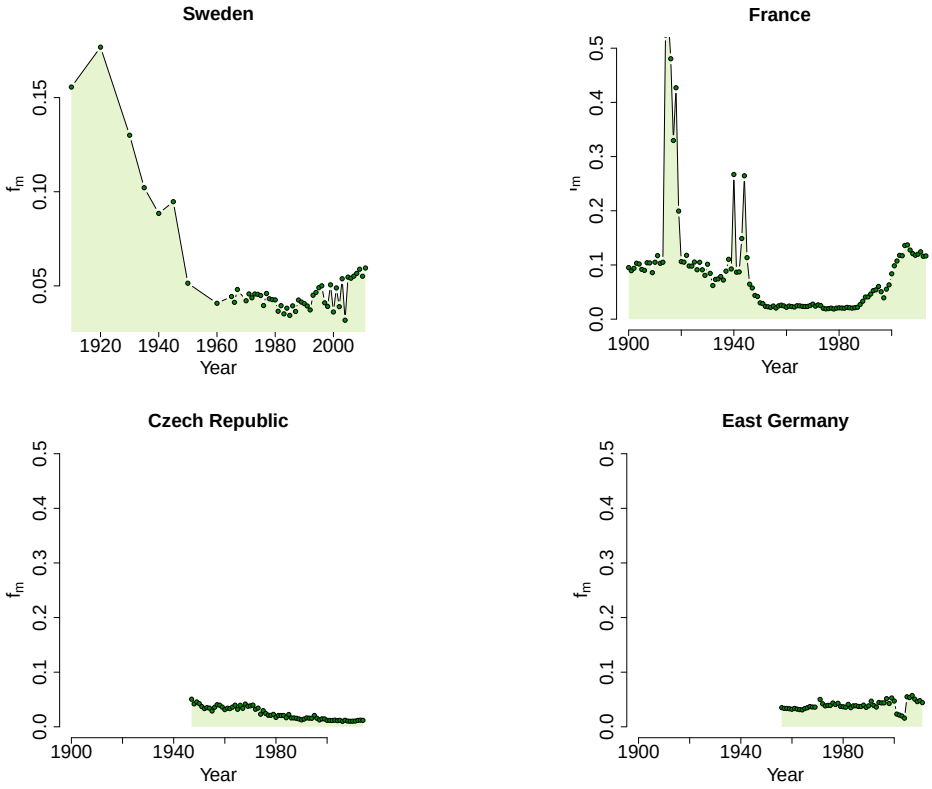


Figure 3.17: Percentage od death related with accidental and premature mortality.

We compute the decomposition of e_0 for all the countries. In Figure 3.18, we can see the contribution of the three components e_{0_I}, e_{0_m} and e_{0_M}. The great contribution in the calculus of the life expectancy at birth comes from adult mortality. Its importance increases across time. For all the countries, infant and childhood contribution is so small that is not visible in the graphs. In Czech Republic and in Germany the contribution of accidental and premature mortality is very small, quite negligible, in particular during the last years. In Sweden and in France accidental and premature contribution reduces between 1930 and 1950 (except in France during the world war years, as we expected), and, then, it becomes quite constant, without disappearing. In the last few years (1990-2011) it increases in particular for France.

A focus on France

France is an emblematic case in the study of accidental and premature mortality. During the demographic transition we observe a gradually disappearance of the accident hump and a greater flattening of the deaths distribution in its middle part. If we compare the red curve with the black one in graph on the left in Figure 3.19 this process is more clear: the distribution of deaths experiments a shift and a compression in adult mortality. This phenomenon can be also expressed in term of rectangularization of the survival curve: in the right graph of Figure (3.19) it is possible to see that the survival curve of 1980 is more similar to a rectangular shape then the one in 1910. The change is due both to the reduction of infant and childhood mortality and to the decrease of accidental and premature mortality. In the last twenty years, instead of a greater rectangularization of the survival curve we observe a almost-parallel shift on the right. Instead to observe only an increase of the compression of the deaths at modal age, we also observe a continuous shift of the

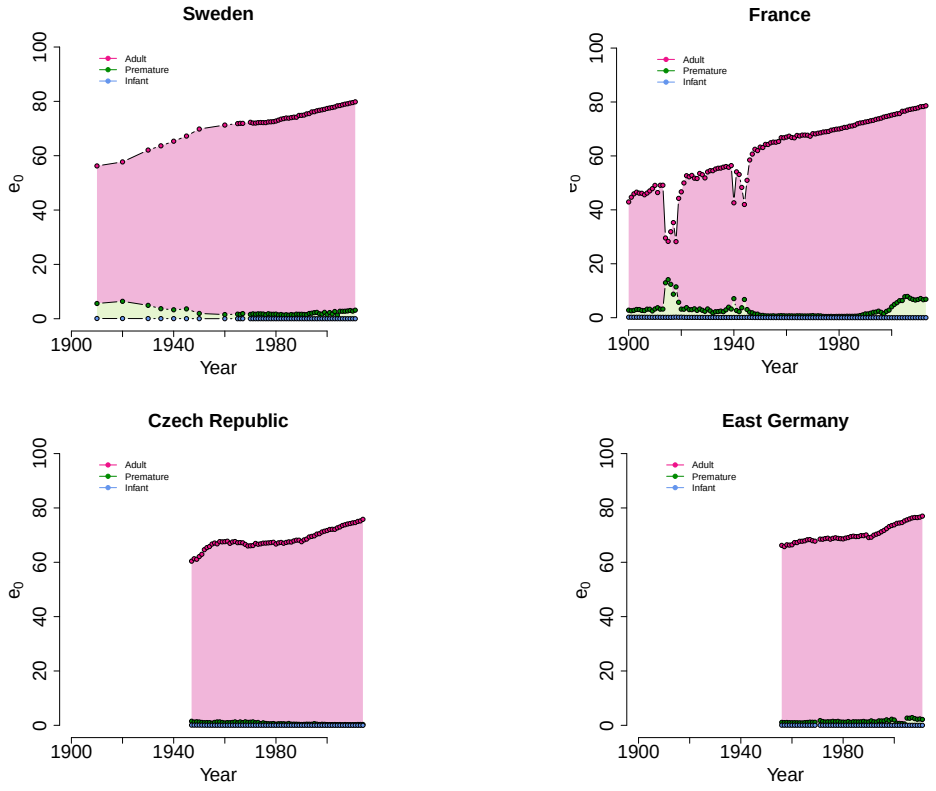


Figure 3.18: The e_0 contribution of the three components: in pink adult mortality, in green the accidental and premature, in light blue the infant and the child one.

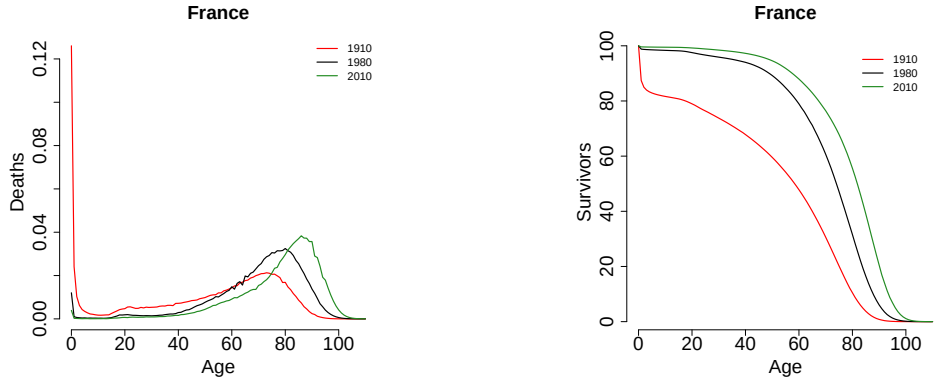


Figure 3.19: Distribution of death by age and related survival curves for different years in France.

Ed. level	ξ_m	ω_m	λ_m	ξ_M	ω_M	λ_M	α
l_0	43.93	13.07	0.66	93.13	16.53	-4.8	0.09
l_1	47.73	13.09	0.36	93.74	15.55	-6.3	0.06
l_2	54.39	13.57	0.03	94.27	14.84	-6.8	0.05
l_3	59.14	14.18	0.02	94.75	13.59	-9.1	0.06

Table 3.1: Parameter values for different levels of education (male).

modal age. This process is accomplished with an increment of the overall variance caused by deaths that, for some unknown reasons, does not follow the general trend of mortality and they occur outside the shape of adult mortality distribution, as it is possible to see observing the difference between the distribution of deaths in 1980 and 2010. This leads to an increase of the transition area between infant and adult mortality, and then into an grater incidence of premature mortality. Also in Sweden we observe the same phenomenon, even if less pronounced.

3.5 Premature mortality and educational levels

In the introduction we presented the problem of the gap between educational levels and mortality. In Figure 1.3 the logarithm of the male mortality rate is plotted for groups with different educational levels:

- without qualification or with primary school diploma (l_0),
- with middle school diploma (l_1),
- with high school diploma (l_2),
- with degree of higher qualification (l_3).

As we have already observed, individuals with the lower mortality rate are the more educated, while the higher mortality rate belongs to less educated people. We also noted that the bigger differences are concentrated between 25 and 65 years old, so during premature mortality. Our aim is to measure the percentage of deaths that occur in this interval and to quantify the contribution of premature mortality in the calculation of life expectancy at birth for different groups. To reach the targets, we estimate the parameter values of the mixture model for the four groups. Until age 25 there are no differences in the risk of dying, so we decided to reduce the mixture model including only the function related to accidental and premature mortality and adult mortality:

$$\tilde{f}(x;\theta) = \alpha f_m(x;\xi_m,\omega_m,\lambda_m) \ + (1-\alpha) f_M(x;\xi_M,\omega_M,\lambda_M) \tag{3.4}$$

The results are displayed in Table 3.5. As we can see, looking at the value of the parameter α, the incidence of premature mortality is more relevant for people with the lowest educational level (greater value). Moreover for these individuals, the distribution f_m is more concentrated on younger ages. On the contrary, the group with the highest educational level presents the lowest value for the mixture parameter and premature mortality distribution shifted at older ages: there is a difference of almost 15 years between $\xi_{m_{l_0}}$ and $\xi_{m_{l_3}}$. In the SN related to adult mortality, the values of the parameters ξ_M and ω_M are close to each other, although the increase of educational leads to a shifting and a compression of the distribution. The differences among the adult distributions are due to the skewness parameter λ_M: we observe a simultaneous increase of the asymmetry and the educational level.

Ed. level	25-40	40-65	65^+
l_0	0.02	0.14	0.83
l_1	0.01	0.10	0.88
l_2	0.01	0.08	0.91
l_3	0.00	0.06	0.94

Table 3.2: Decomposition of the area for different levels of education (male).

Ed. level	$\hat{e}_0$	$\hat{e}_0^*$
l_0	77.2	79.0
l_1	79.4	80.2
l_2	80.9	81.2
l_3	82.5	82.5

Table 3.3: Lost years due to premature mortality in the calculus of life expectancy at birth.

Using these values we compute the percentage of deaths in the intervals 25-40, 40-65 and 65+ (see Table 3.5). For male without qualification or primary school diploma the percentage of deaths after 65 years old is equal to 83%. This percentage increases with the increment of educational level and for individuals with degree or higher qualification it reaches the value 94%.
These results are reflected in life expectancy at birth ($\hat{e}_0$). We compute the estimate of e_0 sum up the mean of f_m and f_M multiplied by α and $(1-\alpha)$, respectively. Since we are considering deaths until 25 years old equal to 0, our estimates are overestimates. We correct them subtracting 0.11847, which is the lost life expectancy in the range (0, 25). The results show that this index increases with educational level, as it is possible to see in Table 3.5. In order to quantify the lost years due to premature mortality, we recompute the life expectancy at birth ($\hat{e}_0^*$) assuming premature mortality fixed and equal to the level we registered for men with higher educational level ($\alpha_{m_{l_3}}, \omega_{m_{l_3}}, \alpha_{m_{l_3}}, \lambda_{m_{l_3}}$). Whit this calculation, less educated people benefit from an increase of 3 years. Also for the other two groups there is an improvement in life expectancy at birth of about one year.
The same analysis are performed for women. Even if the difference among groups are less pronounced, we reach the same conclusions. In Appendix B we report the tables, which show these results.
These estimates are useful to understand the characteristics of mortality for groups with different socio-economic conditions. In particular the cause of the great part of deaths in the age intervals 25-44 and 45-64 are due to car crashes, suicides and breathing apparatus cancer (Mazzuco et al., 2016). These causes are related with the life style: the increase of educational level leads to more awareness choses in term of life protection.

Premature mortality and causes of death

Usually accidental and premature mortality are associated with external causes of death (car crashes, poisoning, homicides and suicides). The disappearance of the accidental hump and the increase of premature mortality, lead us to investigate if there are some changes in the shape of the external causes of deaths, that can justify the estimated shape of f_m. To answer this question we use the data of World Health Organization (WHO). In this database we can find the number of deaths in five-year classes, aggregated by different causes. We show the results of France because it is the country with the greatest increase of premature mortality (we perform the same analyses for all the countries we have taken into account, reaching the same conclusions).

In the WHO database we can study the trend of external causes of death for France from 1954 to 1998. Since the overall number of deaths in the WHO does not correspond with the number of deaths in the HMD, we use the percentage of external causes of death and we reconstruct the specific mortality rate using the m_x computed with HMD data. The percentage of deaths of external causes for each age class is computed as

$$ {}_5p_x^e = \frac{{}_5D_x^e}{{}_5D_x}, \tag{3.5} $$

where ${}_5D_x^e$ are the numbers of death due to external causes and ${}_5D_x$ are the total numbers of deaths in WHO. To disaggregate the data and obtain p_x^e, we use smoothing spline technique .

The term "spline" is used in mathematics to indicate piecewise polynomials for approximating functions whose values are known only at certain points. In order to do this, it is first necessary to fix K points $\eta_1 < \eta_2 < \cdots < \eta_K$ (our ${}_5p_x^e$), called knots, along the horizontal axis. A function $x(s)$ may then be constructed so that it passes exactly through the knots and is free at the other points, provided that it presents regular overall behaviour. Between two successive knots (η_i, η_{i+1}), the curve $x(s)$ coincides with a suitable polynomial of prefixed degree d (usually, $d = 3$ and we speak of cubic splines). These sections of polynomials meet at points $\eta_i, i = 2, \ldots, K-1$, and so the resulting function has continuous derivatives from degree 0 to degree $d-1$ in each η_i.
"B-splines" are a set of special spline functions which derive their computational convenience from the fact that they are non-zero over at most $d+1$ adjacent intervals, so they have compact support. Every spline function of degree d associated with the knots sequence $\eta_1, \ldots, \eta_K$ can be uniquely represented as a linear combination of B-spline basis functions of the same degree, that is

$$ x(s) = \sum_{i=1}^{n_B} \gamma_i B_{i,d}(s), \tag{3.6} $$

where n_B is the total number of B-splines of order d $B_{i,d}(s)$,. The smoothing spline is a method to fit a smooth curve to a set of noisy observations. For each unit, the estimated curve $\hat{x}(s) = \sum_{i=1}^{n_B} \hat{\gamma}_i B_{i,d}(s)$ is obtained with

$$ \hat{\gamma} = \left(B^T B + \lambda \Omega_B\right)^{-1} B^T z, \tag{3.7} $$

where $\lambda > 0$ is a fixed penalization parameter which controls smoothing and can be chosen via cross-validation, B is the matrix $p_2 \times n_B$ whose rows are the B-spline basis functions for each s_j, Ω_B is the $n_B \times n_B$ matrix of which the generic element is $\int B''_{i,d}(t) B''_{j,d}(t) dt$, and z is the $p_2 \times 1$ vector containing the recorded values of the function.

Applying this technique we obtain the probabilities p_x^e and we can compute the age-specific external death rates

$$ m_x^e = p_x^e \cdot m_x \ , \tag{3.8} $$

where, m_x is the age-specific death rates obtained from the HMD. With the specific mortality rates we can compute the life table so that $d_x = \sum_i d_x^i$. We apply again smoothing spline to get a shape that does not depend too much on the randomly variation in the original m_x (smoothing parameter equal to 0.5). Since the mixture model was estimated on d_x, we can verify if there is a relation between d_x^e and f_m. In Figure 3.20 the distribution of external causes of deaths and the function f_m are showed. The two curves do not overlap and they are different both for shape and for range. This means that, accidental and premature mortality can not be explained in terms of external causes of death. However, we need to take into consideration that, in the WHO database,

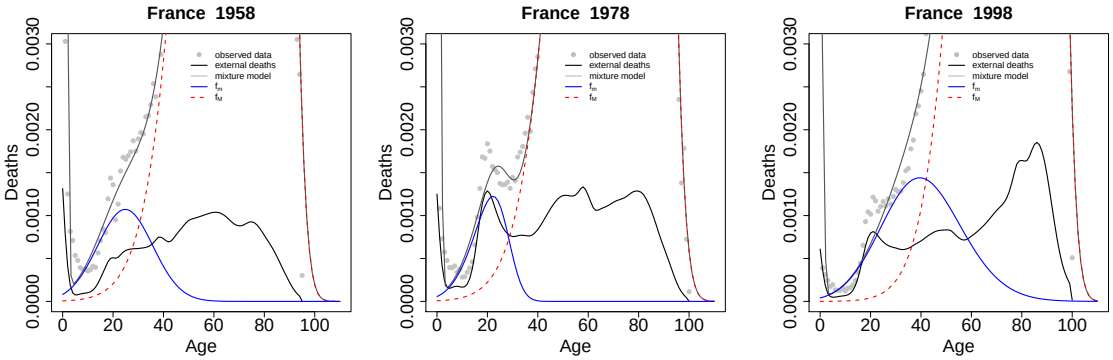


Figure 3.20: External causes of death (black solid line) in France and the estimate f^m function (blue line).

external causes of death include not only car crashes, poisoning, suicides and homicides, but also general accidents. This first analysis is not able to take into consideration this heterogeneity. To restrict the analysis, we employ the data of The Human Cause-of-Death Database, in which more details about the specific causes of death are provides from 2000 to 2013. We consider only:

- alcohol abuse,
- drug abuse,
- accidental poisoning,
- suicide and self-inflicted injury,
- homicide and injury purposely inflicted by other person (not war),
- accident with transportation accident.

The new result are reported in Figure 3.21. Again no relation between f_m for accidental and

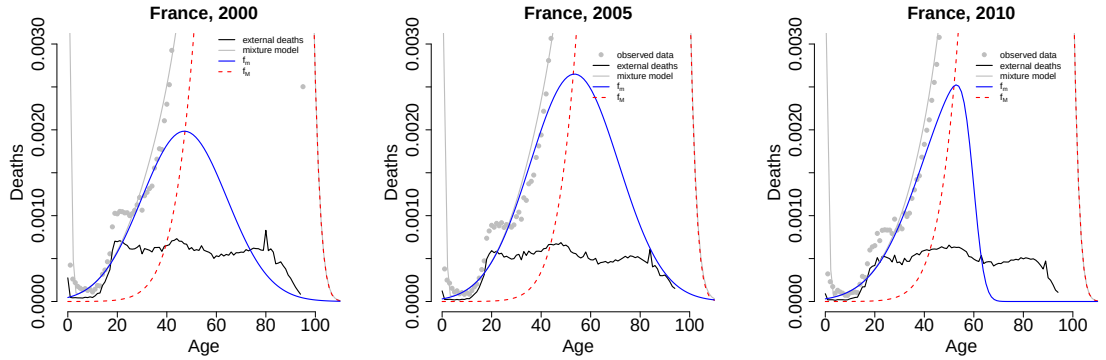


Figure 3.21: Restricted external causes of death (black solid line) in France and the estimate f_m function (blue line).

premature mortality and external deaths is found. This means that car crashes, poisoning, suicides and homicides do not appear to be associated with our function, so the biological justification of the model seem to be missed. However Pearson never spoke about mortality components in terms of different causes of death, and, on the contrary, he wrote that he did not find any association between premature mortality and specific diseases.

Chapter 4

EM algorithm to improve maximum likelihood estimate and reconstruct cohort data

Using maximum likelihood approach, sometimes the model fit is difficult when the last open age class covers a large age interval, so that is not possible to see the natural decline of the death distribution. In these cases the parameter estimated of the f_M distribution related with adult mortality are not appropriate. We find this issue in the USA raw data: between 2000 and 2009 the last class is 85+. In particular for 2009, at age 84 the number of deaths are still increasing and the position of the mode of the death curve is not clearly defined. The fitted model is plotted in Figure 3.6. Both the HN and the first SN present a good approximation, but it is not the same for the last part. The lack of all the right hand-side of the death curve leads to a bad fit: the mode is too shifted on the right and the skewness is excessively steep. Moreover in 2010, deaths until class 100+ are available, so we can calculate the estimates of the parameters for this year. We can reasonably assume that the differences between 2009 and 2010 are very small, so also the two estimated curves should be very close. In Table (4.1) the parameter estimate for these two years are reported. The parameter ξ_M, which is related to the late mode of the distribution, and

USA	η	α	ξ_M	ω_M	λ_M	ξ_m	ω_m	λ_m
2009	0.01	0.04	95.07	21.33	-9.85	19.09	0.04	2.40
2010	0.02	0.03	88.75	21.14	-3.02	30.13	0.03	-1.66

Table 4.1: Parameter values estimated via maximum likelihood for USA in 2009 and 2010.

λ_M, which indicates the skewness, are very different in the two years. The problem is that, after 84 years old, in 2009 the number of deaths for each age class is missing, so that the estimates are based only on the data points on the left side of the adult hump. To fix this problem and to obtain satisfactory estimates we employ the Expectation Maximization (EM) algorithm.

4.1 Definition of EM algorithm

[TM]he EM algorithm is useful when there are incomplete or missing data in the sample. We can suppose that we have observed data Y as incomplete because we have missing data Z. In this situation we can not compute the "complete" likelihood $L_c(\theta|Y, Z)$ because we do not know Z. The idea of the algorithm is to use an interactive procedure to obtain likelihood maximization when the

equations cannot be solved directly because of unknown data observations. The basic structure of algorithm consists of two sequential steps. First (Estimation step), assuming $\theta^{(v)}$ as the true parameter value, the conditional expectation of $L_c(\theta|Y,Z)$, given Y, is computed

$$Q(\theta,\theta^{(v)}) = \mathbb{E}_{\theta^{(v)}}\left[\log L_c(\theta|Y,Z)|Y\right] = \int L_c(\theta|Y,Z)p_{z|Y}\left(z|Y,\theta^{(v)}\right)\,dz, \tag{4.1}$$

then, considering $\theta^{(v)}$ fixed, $Q(\theta,\theta^{(v)})$ is maximized respecting θ (Maximization step). The result is a new value of $\theta = \theta^{(v+1)}$, that is reinserted in the Estimation step and so on. At the first iteration the value of θ^1 is arbitrary chosen. The loop stops when the difference between the value of the likelihood at step (v+1) and at step (v) is smaller than a fixed number. Dempster et al. proved that likelihood increases at each step of the iterations and its convergence is guarantee, while Wu shows that the sequence of $\theta^{(v)}$ converges to maximum likelihood estimate.

We rearrange EM algorithm to solve our problem. In fact we can consider as missing data the frequencies of deaths which occur after age 85. We observe the number of deaths until age x^*, consequently we can compute only $d_x^K = (d_0,\dots,d_{x^*})$. The values of $d_x^U = (d_{x^*+1},\dots,d_\Omega)$ are unknown, but V, the number of deaths that will occur in the age interval (x^*,Ω), is fixed

$$\sum_{i=x^{*+1}}^{\Omega} d_i^U = V. \tag{4.2}$$

In the Estimation step we calculate the values of the unknown deaths

$$d_i^{(v)} = \int_t^{t+1} f(t,\theta^{(v)})\,dt \text{ with } t = (x^*,\dots,\Omega) \tag{4.3}$$

$$\hat{d}_i^{(v)} = V\cdot\frac{d_i^{(v)}}{\sum_{i=x^*+1}^{\Omega} d_i^{(v)}}. \tag{4.4}$$

We obtain a new vector of response variables

$$d_x^{est(v)} = (d_0,\dots,d_{x^*},\hat{d}_{x^*+1}^{(v)},\dots,\hat{d}_\Omega^{(v)}), \tag{4.5}$$

that can be inserted in Equation (2.8) to compute maximum likelihood estimate with respect to θ (Maximization step). We obtain a new value for $\theta = \theta^{(v+1)}$ and we reinsert it in the Equation (4.3), so the algorithm restarts. The loop stops when

$$\left|\log L(\theta^{(v+1)};d_x) - \log L(\theta^{(v)};d_x)\right| \leq 10^{-6}. \tag{4.6}$$

In general, as initial value we use $\theta^{(1)} = (0.1, 0.09, 90, 25, -3, 20, 20, 2)$, that are the same initial values we use to compute maximum likelihood estimates when data are complete. If there are more suitable values, for instance the estimates of an year close to the target one, they can be used as initial values.

4.1.1 Bootstrap confidence intervals

EM algorithm is numerically stable and it is easy to implement in many situations (Jamshidian and Jennrich, 2000). Unfortunately, it does not permit to calculate confidence intervals, which are widely employed because they combine point estimation and hypothesis testing (Di Ciccio and Efron, 1996). Moreover they can be used as a measure of variability of the estimates. In

our particular case they measure the error we commit when EM algorithm is applied. In order to produce confidence intervals we apply bootstrap resampling. This technique was introduced by Efron in 1978. The key idea is to estimate the variation of statistics using a single sample of data. We obtain a random sample (called bootstrap sample), which is computed by sampling, with replacement, from the original dataset. Using the bootstrap sample we estimate the parameter of interest. This procedure (extraction of the random sample and computation of the estimate) is repeated many times in order to create an empirical distribution of the statistic. This distribution is a good approximation of the true (and unknown) probability distribution (Delleji et al., 2007) and can be used to compute confidence interval.
More formally: considering the data $x = (x_1, x_2, \cdots, x_n)$, we obtain an empirical bootstrap sample randomly extracting n elements from the original dataset $x^* = (x_1^*, x_2^*, \cdots, x_n^*)$. We compute the statistic S on the new dataset x^* and we register the result s^*. We repeat the procedure many times so that we collect $s^{*1}, s^{*2}, \cdots, s^{*m}$ from the $x^{*1}, x^{*2}, \cdots, x^{*m}$ samples. The distribution of s^* approximates the real distribution of s. To obtain the 95% confidence interval of s we calculate the empirical quantile $Q_{2.5}(s^*)$ and the empirical quantile $Q_{97.5}(s^*)$ as lower and upper boundary, respectively.

We apply bootstrap technique to obtain confidence intervals both for the unknown d_x and for the estimates of life expectancy at birth. Since we are working with aggregate data we need to reconstruct the starting sample. To do this the total number of deaths is required. With this value we can calculate the amount of individuals that die at age x and create an imitation of the original sample (the real sample is composed by the exact age of death for every subject). From this imitation sample we extract the bootstrap samples and we aggregate again the data in order to apply the EM algorithm and collect the estimates of θ. With each parameter vector we estimate the values of every d_x with Equation (4.3), so that the confidence interval will be $\left[Q_{2.5}(d_x); Q_{97.5}(d_x)\right]$. Instead, to estimate the confidence interval for e_0, we apply Equation (2.10) to every parameter vector, we obtain the estimate distribution of e_0, and again with the quantile $\left[Q_{2.5}(e_0); Q_{97.5}(e_0)\right]$ we get the confidence interval.

4.2 Results

The EM algorithm is tested to analyze its performances with incomplete information data. We consider a complete death distribution and, via maximum likelihood we estimate the parameters of the mixture model. These values are viewed as target estimates. Progressively removing the last age class, we applied EM algorithm to estimate the parameters and reconstruct the missing d_x. With bootstrap method, we compute the confidence intervals of d_x (500 replications).
In Figure 4.1 we show how the algorithm works. (In Appendix C more estimates are reported.) In most cases, the estimates (black bars) are close to the target distribution (red points) and they include the true d_x (grey points). With the increase of the range of the last open class the confidence intervals become larger. This is particularly evident looking at the trend of the estimate variation on the right of the curve. A critic point is the mode of death distribution. When it is not present in the data, the confidence intervals show a bigger variability and they do not always contain the real estimates, even if they are very close. In these situations the model overestimates the mode of the death distribution. The class 60+ can be considered the age limit to reconstruct the curve: if the last age class ends before this age the estimate curve can be unsuitable. However this limit point can not be taken as fix rule: the goodness of fit depends on the percentage of

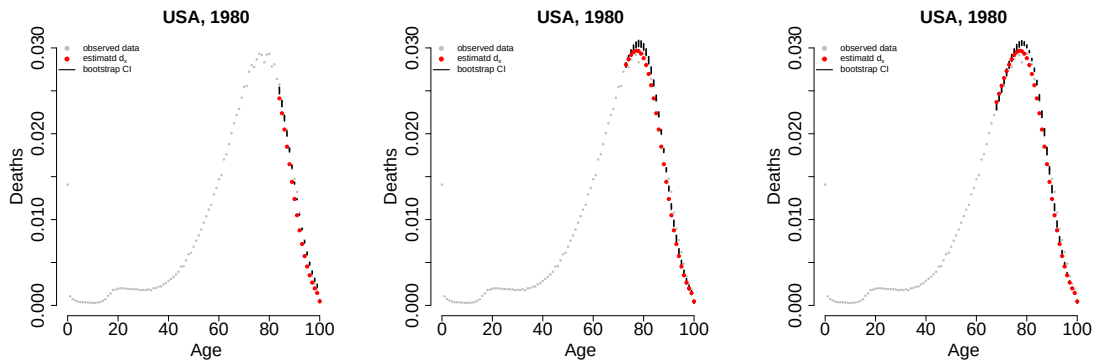


Figure 4.1: EM estimate and their confidence intervals for USA in 1980 when the last open age class is 85+, 74+ and 69+, respectively.

deaths included in the last age interval (lower the percentage, better the estimate) and the shape of the death distribution. In fact, if the adult hump is well defined also before 60 years old, the estimates will be good, otherwise we can have some problems.
In Figure 4.2 the mixture model estimate via EM algorithm for the USA in 2009 (orange color) is plotted. The new curve is preferable than the one estimate only via Maximum Likelihood: it

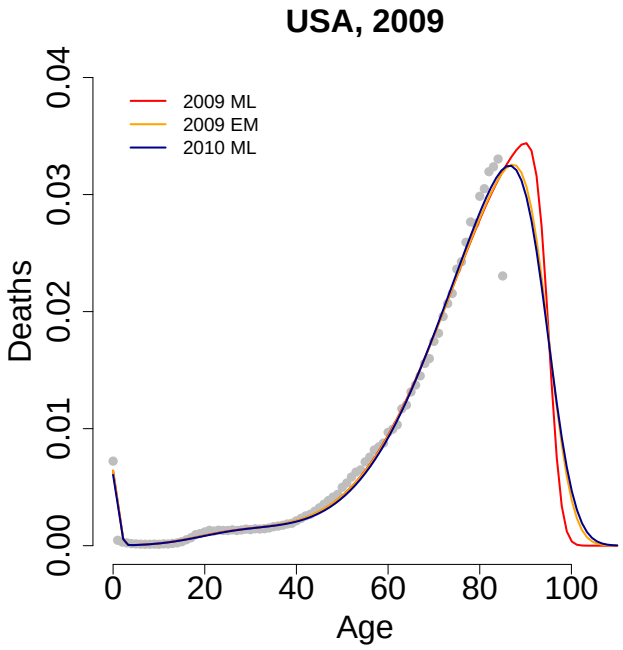


Figure 4.2: EM estimate for the USA in 2009.

is almost overlapping the one for 2010, that we consider our target distribution. This means that the EM algorithm is a good solution when the interval of the last open class is too wide and the maximum likelihood estimate fails.

4.3 Cohort data

The problem of studying mortality using birth cohort data is not new in the literature. Preston et al. (2001) define a birth cohort as all the people born within the same time period, passing through life together, and reaching specific ages at the same time. The main advantage of longitudinal data is the homogeneity of the observations (Livi Bacci, 1983), i.e. individuals experiment the same environmental and socio-economic conditions during the same period. Goldstein and Wachter (2005) shows the existence of a gap between period life expectancy and life expectancy experienced by cohorts while Canudas-Romo and Guillot (2015) wrote that period data are limited because they show only one aspect, their current mortality levels, of the health of the populations. Moreover the year of birth is an important variable to explain mortality (Willets, 1999) so much that Willets (2004) coins the term "cohort effect" and Richards et al., 2006 establishes that cohort has a more significant impact than period in mortality estimations. Unfortunately, not always, historical observations can be accessible and, in a lot of cases, they are available only for a small group of developed countries. Moreover to produce reports we need to wait the cohort extinction, so that there is a big gap between the year of birth and the calculation of mortality rates. For these reasons the use of period life table continuous to be the main instrument to produce reports (Robine et al., 2007), even if there are some disadvantages. In order to reduce the inconvenience of the waiting time to compute statistics from incomplete cohort data, we estimate the number of deaths for a generation still not concluded using EM algorithm. The reconstruction of a birth cohort can be treated as a missing data problem: for a not complete cohort we can register only the number of deaths until age x^* and we can consider the interval (x^*+1, Ω) as the last open age class, where the deaths are equal to $N - \sum_{i=1}^{x^*} D_i$, with N the number of births in the considered cohort.

We apply EM algorithm to reconstruct Danish male birth cohort 1940. From the HMD, we know the number of births in 1940 (B_{1940}) and the number of deaths until 2013 ($D_{1940} = (D_0, \cdots, D_{72})$). We calculate the life table considering as last age class $73+ = N_{1940} - \sum_{i=1}^{72} D_i$. In graph on the left in Figure 4.3 the estimate cure via EM algorithm is shown. The estimate curve seems to have

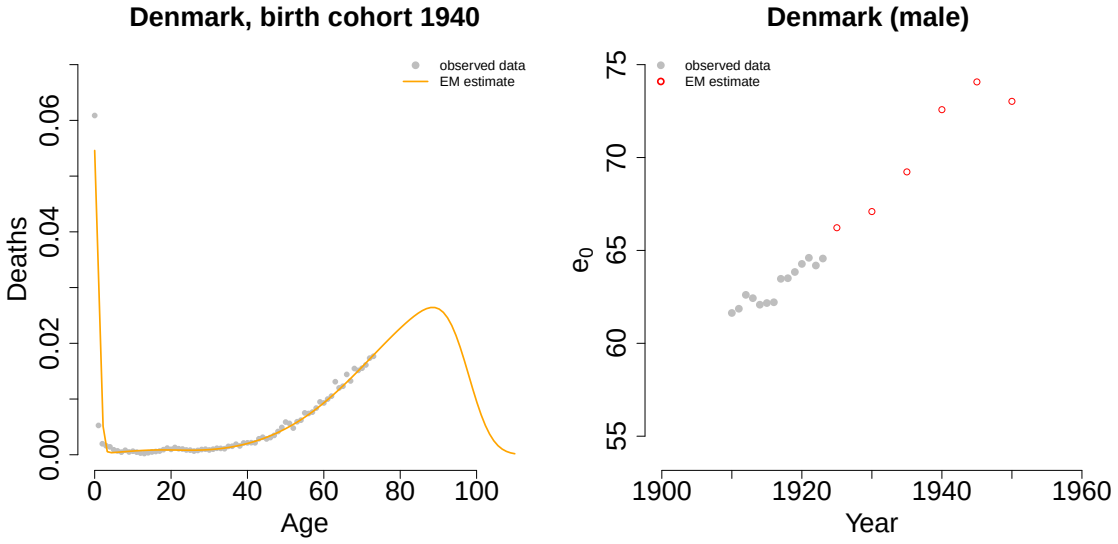


Figure 4.3: Reconstruction of Danish birth cohort 1940 (male) and estimates of life expectancy for not concluded cohorts (male).

a good fit for the cohort data. Its shape is compatible with the mortality trend and the skewness of f_M appear correct. In order to understand if our estimates are credible, we compute the life expectancy at birth both for complete cohort data (via Maximum Likelihood) and for incomplete cohort data (via EM algorithm). The results are shown on the right graph in Figure 4.3: in gray the value of e_0 for complete cohorts, in red the estimate of e_0 for incomplete cohorts. We note an increase tendency, which means that our estimates are generally correct. In fact, as we expected, the trend indicates an increment in life expectancy. However for the last point (cohort 1950) the computed life expectancy at birth is to small. For this cohort the last open age class is 59+, that can be too wide also for EM algorithm.

Chapter 5

Conclusion

We proposed a new mortality model to fit the distribution of deaths by age in the life table. This model is based on a mixture of an Half Normal distribution and two Skew Normal distributions. These functions are chosen following Pearson's theory about mortality components. The Half Normal models infant and childhood mortality, one Skew Normal is adopted to fit accidental and premature mortality, and the last Skew Normal is employed for adult mortality. The model has been fitted in several contexts providing a good fit for a wide range of mortality schedules. The errors committed forecasting the number of deaths in relation to the input data are small and, in general, the results are close to HMD distribution of deaths. For East European countries, even if the overall fit is good, some identification problems was found. This prevent the study of the parameters thens (in particular the ones for accidental and premature mortality) because several gaps are present in the estimates.
The performance of mixture distribution was evaluated in relation to Siler and Heligman and Pollard models, which also fit the entire mortality schedule. Our results show that our model permits a better fit than Heligman and Pollard model, which has the same number of parameters. For recent years, we observe that only the mixture model is able to capture the entire hump of adult mortality. In term of AIC, we found that the mixture model is lower than Heligman and Pollard model and close to Siler one. This means that the addition of three parameters really decrease the information loss.

All parameters of the model have a demographic interpretation and they can be studied to analyze the characteristics and the transformations of mortality components. The results we obtained for infant and childhood mortality show a reduction and a concentration of the incidence of deaths at age 0: the risk of dying after birth is very small and the mortality during childhood has almost disappeared (Kannisto, 1994; Robine, 2001; Wilmoth and Horiuchi, 1999; Wilmoth et al., 2000). For adult mortality we find a general compression of the deaths around the modal age at death (Bongaarts, 2005; Canudas-Romo, 2008; Kannisto, 1996; Lan Karen Cheung and Robine, 2007). In most of the countries this phenomenon is accomplished with a shifting of the late mode on the right of the distribution (Cheung, 2003; Lan Karen Cheung and Robine, 2007; Cheung et al., 2008, 2005). The conclusions about infant, childhood and adult mortality are consistent with what we already know about the trends of these components: decrease of infant mortality, shifting and compression of adult mortality. This supports the validity of the model and the interpretation of the parameters. A characteristic which is not taken into account to study adult mortality, is its skewness. With our model we showed that the adult mortality distribution is skew and that this feature is quite stable in time. Instead of define accidental and premature mortality as a

consequence of the others two components, in our model it is defined with an own distribution. Taking advantage of this, we analyzed its characteristics. We observed that during the last century, the accidental hump around 20 years old is disappeared. In the course of this transition the death curve becomes more flat in its middle part, so that the incidence of premature mortality is small and quite stable. Recently, for the countries which experiment a shifting and a compression of the adult mortality (Sweden and France), we observed an increase of premature mortality due to deaths that occurs near, but outside the adult distribution. The same results was found by several authors, who studied the survival curve and observe its almost-parallel shift on the right (Gavrilov and Gavrilova, 1991; Lynch and Brown, 2001; Manton and Tolley, 1991; Robine, 2001; Yashin et al., 2001). This consideration suggests the trend of adult mortality is correlated with the trend of premature mortality because the transformations of adult distribution influence what happens to premature distribution.
The definition of a own distribution for accidental and premature mortality permits to study mortality considering educational levels. Taking into account the trend of mortality rates, we note that the differences are concentrated during youth and first adulthood. Hypothetically filling this the gap, we calculate that, in Italy, the life expectancy at birth of lower educated male increases by three years old.

This calculus was achievable because our model provides the possibility to compute in explicit form the life expectancy at birth, that is the pondered average of the means of the three distributions involved. In this way it it easy to understand the contribution of the single mortality components in the estimate of life expectancy at birth. As we expected, in general the main contribution in its calculus comes from adult mortality. However we note that accidental and premature component is useful in the overall computation because it permits a more accurate estimate.

Moreover, the possibility to disaggregate the total area under the distribution of deaths into the single areas permits to calculate the percentage of deaths related to each mortality components. Since the sum of these percentages is always equal to 1, we can compare different mortality data and quantify these differences in terms of incidence of mortality components.

When the last open age class covers a too wide interval, we find some problems in the estimates, as it happened for USA in 2009. In particular the adult mortality was bad fitted because its distribution was too skew. To fix this issue, we proposed to implement EM algorithm (Dempster et al., 1977), which is an easy and not excessively expensive, from a computational prospective, procedure (Jamshidian and Jennrich, 2000). Our results showed that the improvements in the estimate are significant.
This method can also be applied to reconstruct a birth cohort data and evaluate life expectancy at birth almost 40 years before its natural ending. This can increase the use of cohort data to study mortality features and speed up the comparison between period and longitudinal data, which often suggest differences in term of mortality prospectives (Canudas-Romo and Guillot, 2015; Goldstein and Wachter, 2005; Richards et al., 2006; Willets, 1999, 2004). Unfortunately EM algorithm does not provide confidence intervals, that are important to evaluate the variability of the estimates (Di Ciccio and Efron, 1996). To solve this issue we used bootstrap technique (Efron, 1979). However the employment of bootstrap resampling is time expensive, in particular if the target is more than a single year. Also the boundary of 60+ as last age class can be a limitation for researchers that would work with more recent data.

Appendix A

Model for Sweden in 1910 - 2010

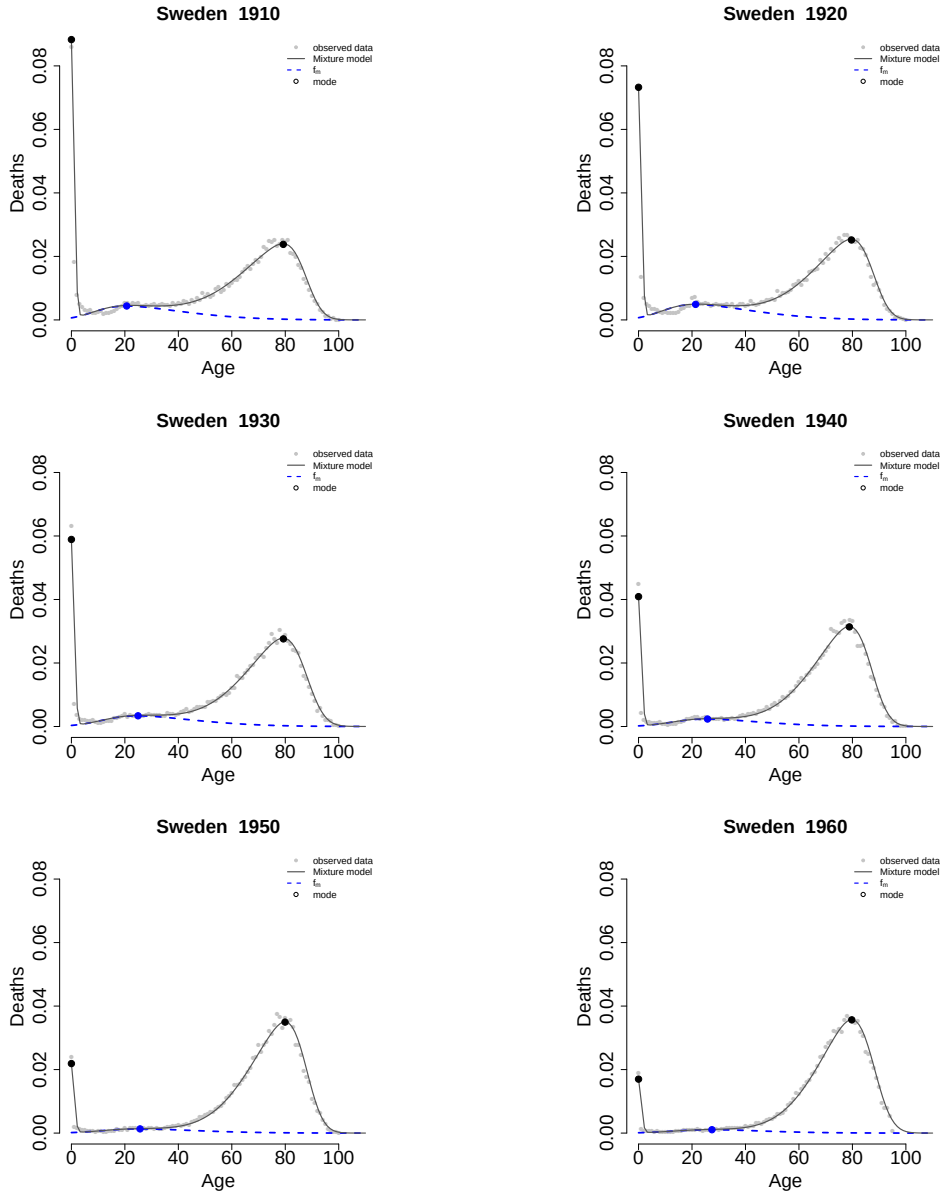


Figure A.1: Estimated model for different years in Sweden (1910-1960).

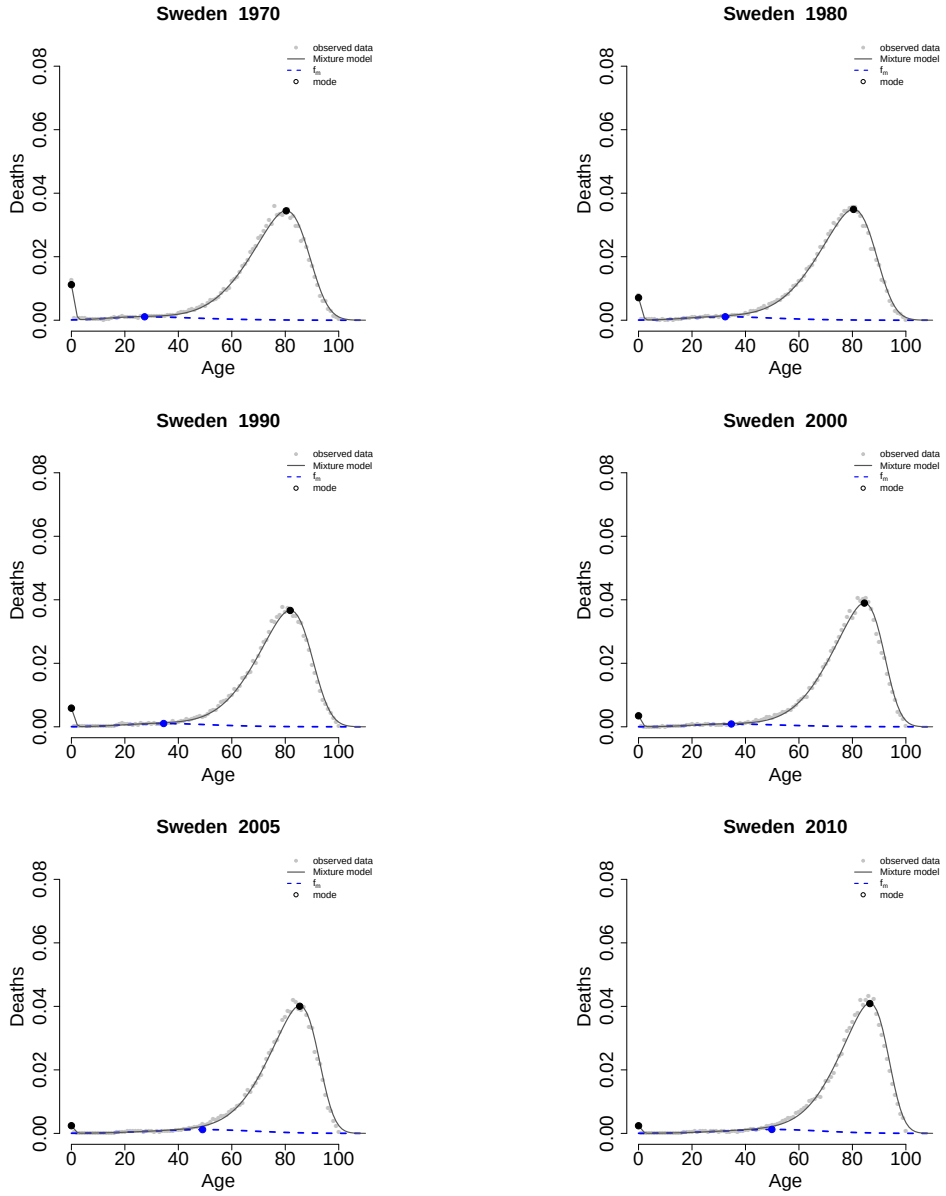


Figure A.2: Estimated model for different years in Sweden (1970-2010).

Appendix B

Mortality and educational level for women

FEMALE	ξ_m	ω_m	λ_m	ξ_M	ω_M	λ_M	α
l0	59.36	14.73	0.05	96.01	12.14	-10.08	0.12
l1	59.77	15.06	0.46	96.75	12.09	-10.08	0.11
l2	66.89	14.34	0.01	96.98	11.68	-10.08	0.11
l3	66.87	13.92	0.06	97.29	11.50	-10.08	0.10

Table B.1: Parameter values for different levels of education (female).

Ed. level	25-40	40-65	65^+
l_0	0.01	0.08	0.91
l_1	0.00	0.06	0.94
l_2	0.00	0.05	0.95
l_3	0.00	0.05	0.95

Ed. level	$\hat{e}_0$	$\hat{e}_0^*$
l_0	83.1	84.5
l_1	84.6	85.2
l_2	85.3	85.7
l_3	86.0	86.0

Figure B.1: Decomposition of the area for levels of education and quantification of lost years due to premature mortality in the calculus of life expectancy at birth.

Appendix C

EM estimates for USA in 1980

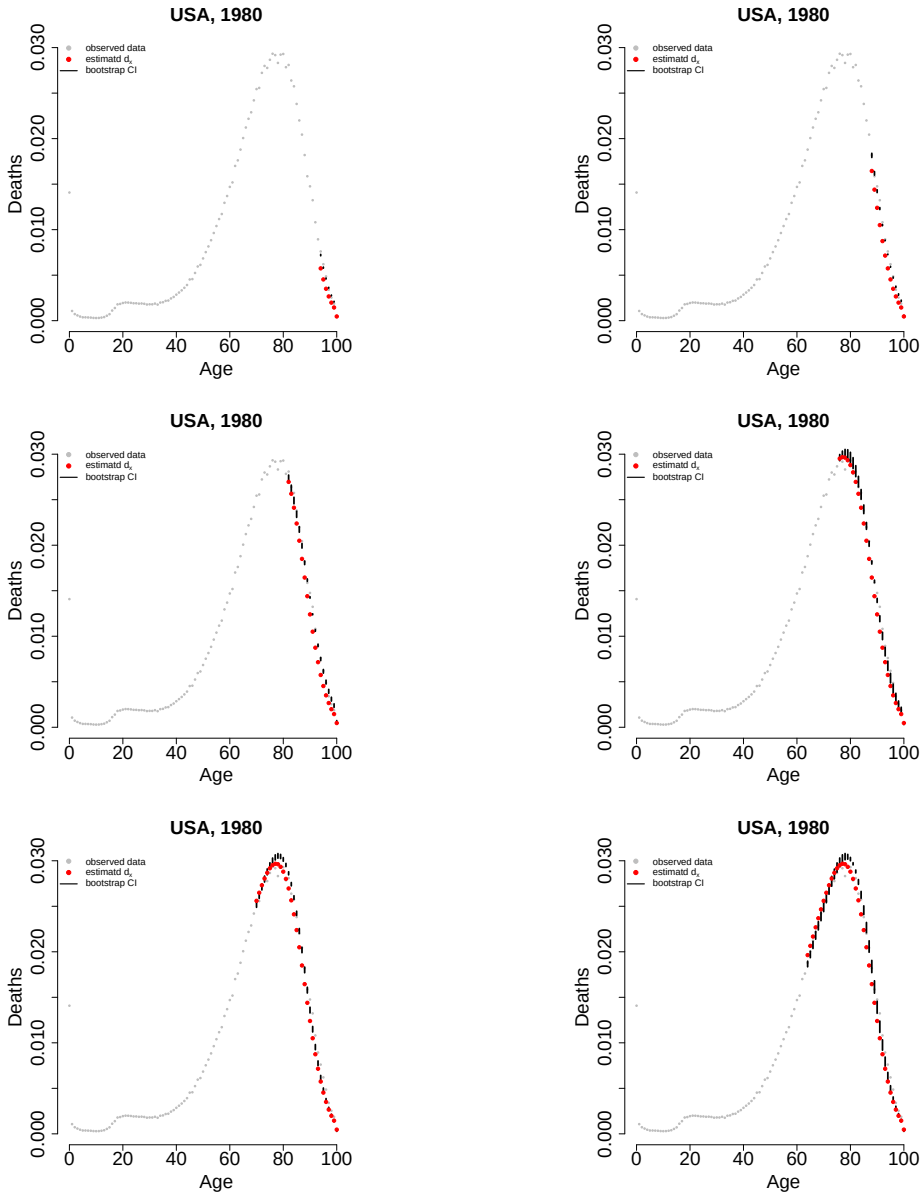


Figure C.1: EM estimates and confidence intervals for USA 1980.

Appendix D

Functions implemented in R

```
### Mixture model

lx.sbn<-function(x,omega1,omega2,alpha,lambda1,lambda2,xi1,xi2,m){
ntsbn<-function(x){((2/omega1)*(1-alpha)*dnorm((x-xi1)/omega1)*pnorm(lambda1*(x-xi1)/omega1))+
((2/omega2)*(alpha)*dnorm((x-xi2)/omega2)*pnorm(lambda2*(x-xi2)/omega2))}
f<-m*ntsbn(x)+(1-m)*as.numeric(x>=0&x<=105)*((sqrt(2)/(sqrt(pi)))*exp(-(x^2)/(2)))
return(f)
}



### Maximum Likelihood estimate

lx.sbn.mle<-function(Age, d.x, param){

ww1<-param[1]
ww2<-param[2]
aa<-param[3]
ll1<-param[4]
ll2<-param[5]
xx1<-param[6]
xx2<-param[7]
m<-param[8]
nrow<-NROW(d.x)
Age2 <- c(Age, 115)

sbn<-function(x){
ntsbn<-function(x){((2/ww1)*(1-aa)*dnorm((x-xx1)/ww1)*pnorm(ll1*(x-xx1)/ww1))+
((2/ww2)*((aa))*dnorm((x-xx2)/ww2)*pnorm(ll2*(x-xx2)/ww2))}
#tsbn<-function(x){as.numeric(x>=0&x<=105)*ntsbn(x)/(integrate(ntsbn,lower=0,upper=(max(Age2)-1))$
f<-m*ntsbn(x)+(1-m)*as.numeric(x>=0&x<=105)*((sqrt(2)/(sqrt(pi)))*exp(-(x^2)/(2)))
return(f)
}
```

```
smv <- 0
for (i in 1:length(d.x)){
f.hat<-integrate(sbn,Age2[i],Age2[i+1])$value/(Age2[i+1]-Age2[i])
smv<-smv-d.x[i]*(ifelse(f.hat>0,log(f.hat),0))
}
return(smv)
}



### EM function

EM <- function(dx,Age,V,param = c(25,20,0.10,-3,2,90,20,0.9)){
omega1 <- param[1]
omega2 <- param[2]
alpha <- param[3]
lambda1 <- param[4]
lambda2 <- param[5]
xi1 <- param[6]
xi2 <- param[7]
m <- param[8]

EMpar <- matrix(nrow = 1000, ncol = 8,
dimnames = list(c(1:1000),c("w1","w2","a","l1","l2","x1","x2","m")))
EMpar[1,] <- param

loglik<- rep(NA, 1000)
loglik[1]<-0
loglik[2]<- log(lx.sbn.mle(Age=Age, d.x=dx, param=param))

k <- 2

# loop nlminb
while(abs(loglik[k]-loglik[k-1]) >= 0.00001){

# E step
age.est <- c(Age[length(Age)]:105)
pr.est <- rep(NA,length(age.est))
dx.est <- rep(NA,length(age.est))
n <- 105-length(Age) +1
for(i in c(1:n)){
pr.est[i] <- integrate(lx.sbn,omega1,omega2,alpha,lambda1,lambda2,xi1,xi2,m,
lower = age.est[i],upper = age.est[i+1])$value
}
pr.est[length(age.est)] <- integrate(lx.sbn,omega1,omega2,alpha,lambda1,lambda2,xi1,xi2,m,
```

```
lower = 105,upper = 120)$value
dx.est <- V * (pr.est/sum(pr.est))

# M step
d.x <- c(dx[-length(dx)],dx.est)
age <- c(0:105)
par<- nlminb(start = c(param[1],param[2],param[3],param[4],param[5],param[6],param[7],param[8]),lx
Age=age, d.x=d.x,control=list(iter.max=15000,abs.tol=1e-20),
lower=c(0.1,0.1,0.0001,-5,-5,30,18,0.001),upper=c(50,30,10,5,20,100,60,1))

omega1 <- par$par[1]
omega2 <- par$par[2]
alpha <- par$par[3]
lambda1 <- par$par[4]
lambda2 <- par$par[5]
xi1 <- par$par[6]
xi2 <- par$par[7]
m <- par$par[8]
EMpar[k,] <- par$par
loglik[k+1] <- log(par$objective)
k<- k+1
}
EMpar.ok <- na.omit(data.frame(EMpar))
best.par <- EMpar.ok[length(EMpar.ok[,1]),]
best.lik <- loglik[length(EMpar.ok[,1])+1]
return(list(parameter=best.par,likelihood=best.lik,
iterations=length(EMpar.ok[,1])))
}



### Bootstrap

boot <- function(dx,age,nboot,death=100000){
nobs <- death*dx
minage <- min(age)
maxage <- max(age)
V <- dx[length(dx)]
data <- rep.int(age,round(nobs))
boot=foreach(i=1:nboot, .combine=cbind) %dopar% {
newdata <- sample(data,replace=TRUE,size=death)
Deaths2 <- as.numeric(table(factor(newdata,levels=minage:maxage)))
dx <- Deaths2/sum(Deaths2)
EMest <- EM(dx=dx,Age=age,V=V)
as.numeric(EMest$parameter)
```

```
}
}
```

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.

Barbi, E., Caselli, G., and Vallin, J. (2003). Trajectories of extreme survival in heterogeneous populations. *Population (english edition)*, pages 43–65.

Basellini, U., Canudas-Romo, V., and Lenart, A. (1971). Location-scale models in demography: a generalization of the parametric models of mortality.

Beard, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. *Biological aspects of demography*, 999:57–68.

Benjamin, B. (1959). Actuarial aspects of human lifespans. In *Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing), Volume 5*, pages 2–20. Wiley Online Library.

Bennett, S. (1983). Log-logistic regression models for survival data. *Applied Statistics*, pages 165–171.

Bergeron-Boucher, M.-P., Ebeling, M., and Canudas-Romo, V. (2015). Decomposing changes in life expectancy: Compression versus shifting mortality. *Demographic Research*, 33:391–424.

Blangiardo, G. C. (1997). *Elementi di demografia*. Il mulino.

Bodio, L. (1887). Quelques renseignements sur les conditions hygiéniques et sanitaires de l'italie. *Bulletin de l'Institut international de statistique*, 2(part 1):264–84.

Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. *Demography*, 42(1):23–49.

Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research*, 19:1179–1204.

Canudas-Romo, V. and Guillot, M. (2015). Truncated cross-sectional average length of life: A measure for comparing the mortality history of cohorts. *Population studies*, 69(2):147–159.

Cheung, J. (2001). The long-term trend of non-maori mortality and its more recent compression effect. In *Population Association of New Zealand biannual conference, Wellington*.

Cheung, K. S. L. (2003). Scalar expansion and normal longevity in hong kong.

Cheung, S. L. K., ROBINE, J. M., and Caselli, G. (2008). The use of cohort and period data to explore changes in adult longevity in low mortality countries. *Genus*, pages 101–129.

Cheung, S. L. K., Robine, J.-M., Paccaud, F., and Marazzi, A. (2009). Dissecting the compression of mortality in switzerland, 1876-2005. *Demographic research*, 21:569–598.

Cheung, S. L. K., Robine, J.-M., Tu, E. J.-C., and Caselli, G. (2005). Three dimensions of the survival curve: Horizontalization, verticalization, and longevity extension. *Demography*, 42(2):243–258.

Coale, A. J. (1989). Demographic transition. In *Social Economics*, pages 16–23. Springer.

Coale, A. J., Demeny, P., and Vaughan, B. (2013). *Regional Model Life Tables and Stable Populations: Studies in Population*. Elsevier.

Coale, A. J. and Kisker, E. E. (1990). Defects in data on old-age mortality in the united states: new procedures for calculating mortality schedules and life tables at the highest ages.

Congdon, P. (1993). Statistical graduation in local demographic analysis and projection. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 237–270.

Cox, P. R. (1950). *Demography*. Cambridge university press.

Dalstra, J., Kunst, A., Mackenbach, J., on Socioeconomic Inequalities in Health, E. W. G., et al. (2006). A comparative appraisal of the relationship of education, income and housing tenure with less than good health among the elderly in europe. *Social science & medicine*, 62(8):2046–2060.

de Beer, J. and Janssen, F. (2014). Netherlands interdisciplinary demographic institute (nidi) po box 11650 2502 ar the hague phone: 070-3565200 e-mail: beer@ nidi. nl for all correspondence.

Delleji, T., Zribi, M., and Hamida, A. B. (2007). On the em algorithm and bootstrap approach combination for improving satellite image fusion. *International Journal of Signal Processing*, 4(1).

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Di Ciccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical science*, pages 189–212.

Eakin, T. and Witten, M. (1995). How square is the survival curve of a given species? *Experimental Gerontology*, 30(1):33–64.

Efron, B. (1979). Computers and the theory of statistics: thinking the unthinkable. *SIAM review*, 21(4):460–480.

Elal-Olivero, D., Gómez, H. W., and Quintana, F. A. (2009). Bayesian modeling using a class of bimodal skew-elliptical distributions. *Journal of Statistical Planning and Inference*, 139(4):1484–1492.

Elderton, W. P. (1903). Graduation and analysis of a sickness table. *Biometrika*, 2(3):260–272.

Flinn, M. W. (1981). The european demographic system 1500-1820. In *Johns Hopkins Symposia in Comparative History*, number 12. Johns Hopkins University Press Baltimore Md. United States 1981.

Fries, J. F. (2002). Aging, natural death, and the compression of morbidity. *Bulletin of the World Health Organization*, 80(3):245–250.

Gavrilov, L. A. and Gavrilova, N. S. (1991). The biology of life span: a quantitative approach.

Go, C. G., Brustrom, J. E., Lynch, M. F., and Aldwin, C. M. (1995). Ethnic trends in survival curves and mortality. *The Gerontologist*, 35(3):318–326.

Goldstein, J. R. and Wachter, K. W. (2005). Gaps and lags: Relationships between period and cohort life expectancy. *unpublished paper. March*, 22.

Gompertz, B. (1825). 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 of the Royal Society of London*, 115:513–583.

Gumbel, E. J. (1937). *La durée extrême de la vie humaine*. Number 520. Hermann et cie.

Hattersley, L. (1997). Expectation of life by social class. *Health Inequalities. Office for National Statistics (Series DS No 15), TSO: London*.

Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(01):49–80.

Hill, G. (1993). The entropy of the survival curve: An alternative measure. *Canadian Studies in Population*, 20(1):43–57.

Horiuchi, S., Ouellette, N., Cheung, S. L. K., and Robine, J.-M. (2013). Modal age at death: lifespan indicator in the era of longevity extension. *Vienna Yearbook of Population Research*, pages 37–69.

Horiuchi, S. and Wilmoth, J. R. (1997). Age patterns of the life table aging rate for major causes of death in japan, 1951–1990. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 52(1):B67–B77.

Horiuchi, S. and Wilmoth, J. R. (1998). Deceleration in the age pattern of mortality at olderages. *Demography*, 35(4):391–412.

Huisman, M., Kunst, A. E., Andersen, O., Bopp, M., Borgan, J.-K., Borrell, C., Costa, G., Deboosere, P., Desplanques, G., Donkin, A., et al. (2004). Socioeconomic inequalities in mortality among elderly people in 11 european populations. *Journal of Epidemiology and Community Health*, 58(6):468–475.

Jamshidian, M. and Jennrich, R. I. (2000). Standard errors for em estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):257–270.

Kannisto, V. (1992). Presentation at a workshop on old-age mortality.

Kannisto, V. (1994). Development of oldest-old mortality 1950-1990: evidence from 28 developed countries.

Kannisto, V. (1996). The advancing frontier of survival: life tables for old age.

Kannisto, V. (2001). Mode and dispersion of the length of life. *Population: An English Selection*, pages 159–171.

Klein, J. P. and Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Kostaki, A. (1992). A nine-parameter version of the heligman-pollard formula. *Mathematical population studies*, 3(4):277–288.

Lan Karen Cheung, S. and Robine, J.-M. (2007). Increase in common longevity and the compression of mortality: The case of japan. *Population studies*, 61(1):85–97.

Landry, A. (1934). *La révolution démographique: études et essais sur les problèmes de la population*. Ined.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.

Lee, R. (2003). The demographic transition: three centuries of fundamental change. *The journal of economic perspectives*, 17(4):167–190.

Levasseur, E. (1889). *La population française: Histoire de la population avant 1789 et démographie de la France comparée à celle des autres nations au XIXe siécle, précédée d'une introduction sur la statistique*, volume 1. Arthur Rousseau.

Levy, M. L. (1996). La rectangularisation de la courbe des survivants. *Rectangularization of the survival curve*, pages 576–79.

Lexis, W. H. R. A. (1879). *Sur la durée normale de la vie humaine et sur la théorie de la stabilité des rapports statistiques*. Vve. F. Henry.

Livi Bacci, M. (1983). Introduzione alla demografia.

Livi-Bacci, M. (1999). *The population of Europe*. Blackwell Oxford.

Lynch, S. M. and Brown, J. S. (2001). Reconsidering mortality compression and deceleration: An alternative model of mortality rates. *Demography*, 38(1):79–95.

Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, 8(6):301–310.

Malthus, T. R. (1926). First essay on population 1798.

Manton, K. G. and Stallard, E. (1996). Longevity in the united states: Age and sex-specific evidence on life span limits from mortality patterns 1960–1990. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 51(5):B362–B375.

Manton, K. G. and Tolley, H. D. (1991). Rectangularization of the survival curve implications of an ill-posed question. *Journal of Aging and Health*, 3(2):172–193.

Marmot, M. G. and McDowall, M. E. (1986). Mortality decline and widening social inequalities. *The Lancet*, 328(8501):274–276.

Martel, S. and Bourbeau, R. (2003). Compression de la mortalité et rectangularisation de la courbe de survie au québec au cours du xxe siècle. *Cahiers québécois de démographie*, 32(1):43–75.

Mazzuco, S., Scarpa, B., and Zanotto, L. (2016). a mortality model based on a mixture distribution function.

Missov, T. I., Lenart, A., Nemeth, L., Canudas-Romo, V., and Vaupel, J. (2015). The gompertz force of mortality in terms of the modal age at death. *Demographic Research*, 32:1031–1048.

Nagnur, D. (1986). Rectangularization of the survival curve and entropy: The canadian experience, 1921-1981. *Canadian Studies in Population*, 13(1):83–102.

Notestein, F. W. (1953). *Economic problems of population change*. Oxford University Press London.

Nusselder, W. J. and Mackenbach, J. P. (1996). Rectangularization of the survival curve in the netherlands, 1950-1992. *The Gerontologist*, 36(6):773–782.

Nusselder, W. J. and Mackenbach, J. P. (1997). Rectangularization of the survival curve in the netherlands: an analysis of underlying causes of death. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 52(3):S145–S154.

Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570):1029–1031.

Olshansky, S. J., Carnes, B. A., and Désesquelles, A. (2001). Prospects for human longevity. *Science*, 291(5508):1491–1492.

Paccaud, F., Pinto, C. S., Marazzi, A., and Mili, J. (1998). Age at death and rectangularisation of the survival curve: trends in switzerland, 1969-1994. *Journal of Epidemiology and Community Health*, 52(7):412–415.

Pareto, V. (1964). *Cours d'économie politique*, volume 1. Librairie Droz.

Pearson, K. (1897). *Chances of Death, and Other Studies in Evolution*, volume 1. CUP Archive.

Pelletier, F., Légaré, J., and Bourbeau, R. (1997). Mortality in quebec during the nineteenth century: From the state to the cities. *Population Studies*, 51(1):93–103.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries (1886-1994)*, 63(1):12–57.

Perozzo, L. (1879). Distribuzione dei morti per eta. *Annali di Statistica*, pages 75–93.

Preston, S. H., Heuveline, P., and Guillot, M. (2001). Demography: measuring and modeling population processes. *Pop. Dev. Rev*, 27:365.

Richards, S. J., Kirkby, J., and Currie, I. D. (2006). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, 12(01):5–38.

Robine, J.-M. (2001). Redefining the stages of the epidemiological transition by a study of the dispersion of life spans: The case of france. *Population: An English Selection*, pages 173–193.

Robine, J.-M., Crimmins, E. M., Horiuchi, S., and Zeng, Y. (2007). *Human longevity, individual life duration, and the growth of the oldest-old population*, volume 4. Springer Science & Business Media.

Rocha, G. H., Loschi, R. H., and Arellano-Valle, R. B. (2013). Inference in flexible families of distributions with normal kernel. *Statistics*, 47(6):1184–1206.

Rogers, A. and Little, J. S. (1994). Parameterizing age patterns of demographic rates with the multiexponential model schedule. *Mathematical Population Studies*, 4(3):175–195.

Rothenberg, R., Lentzner, H. R., and Parker, R. A. (1991). Population aging patterns: the expansion of mortality. *Journal of gerontology*, 46(2):S66–S70.

Shkolnikov, V. M., Andreev, E. M., Jdanov, D. A., Jasilionis, D., Kravdal, Ø., Vågerö, D., and Valkonen, T. (2011). Increasing absolute mortality disparities by education in finland, norway and sweden, 1971–2000. *Journal of epidemiology and community health*, pages jech–2009.

Siler, W. (1979). A competing-risk model for animal mortality. *Ecology*, 60(4):750–757.

Strand, B. H., Grøholt, E.-K., Steingrímsdóttir, Ó. A., Blakely, T., Graff-Iversen, S., and Næss, Ø. (2010). Educational inequalities in mortality over four decades in norway: prospective study of middle aged men and women followed for cause specific mortality, 1960-2000. *Bmj*, 340:c654.

Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):5–43.

Thatcher, A. R., Kannisto, V., and Vaupel, J. W. (1998). The force of mortality at ages 80 to 120.

Thiele, T. N. and Sprague, T. (1871). On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the danish life insurance company of 1871. *Journal of the Institute of Actuaries and Assurance Magazine*, 16(5):313–329.

Thompson, W. S. (1930). *Population problems*. New York: McGraw-Hill.

Valkonen, T. (2001). Trends in differential mortality in european countries. *Trends in mortality and differential mortality*, 36:185–321.

Valkonen, T., Martelin, T., Rimpela, A., Notkola, V., and Savela, S. (1993). Socio-economic mortality differences in finland 1981-90.

Van de Kaa, D. J. (1987). Europe's second demographic transition. *Population bulletin*, 42(1):1–59.

Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., Iachine, I. A., Kannisto, V., Khazaeli, A. A., Liedo, P., et al. (1998). Biodemographic trajectories of longevity. *Science*, 280(5365):855–860.

Weibull, W. (1951). Wide applicability. *Journal of applied mechanics*, 103:293–297.

Willets, R. (1999). The cohort effect: insights and explanationsmortality in the next millennium. *Paper presented to the Staple Inn Actuarial Society.*

Willets, R. (2004). The cohort effect: insights and explanations. *British Actuarial Journal*, 10(04):833–877.

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D. A., Boe, C., Bubenheim, M., Philipov, D., Shkolnikov, V., and Vachon, P. (2007). Methods protocol for the human mortality database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality. org [version 31/05/2007]*, 9:10–11.

Wilmoth, J. R., Deegan, L. J., Lundström, H., and Horiuchi, S. (2000). Increase of maximum life-span in sweden, 1861-1999. *Science*, 289(5488):2366–2368.

Wilmoth, J. R. and Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography*, 36(4):475–495.

Wu, C. J. (1983). On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95–103.

Yashin, A. I., Begun, A. S., Boiko, S. I., Ukraintseva, S. V., and Oeppen, J. (2001). The new trends in survival improvement require a revision of traditional gerontological concepts. *Experimental Gerontology*, 37(1):157–167.

Zarulli, V., Jasilionis, D., and Jdanov, D. A. (2012). Changes in educational differentials in old-age mortality in finland and sweden between 1971-1975 and 1996-2000. *Demographic research*, 26:489–510.

Zarulli, V., Marinacci, C., Costa, G., and Caselli, G. (2013). Mortality by education level at late-adult ages in turin: a survival analysis using frailty models with period and cohort approaches. *BMJ open*, 3(7):e002841.

Lucia Zanotto

Curriculum Vitae

Dept. of Statistical Science,
University of Padova, Italy
(+39) 392 791 5452
zanotto@stat.unipd.it

"The best thing about being a statistician is that you get to play in everyone's backyard" - John Tukey

Personal Information

Date of Birth	22/12/1987
Address	Vic. A. Vivaldi8, 35020 Albignasego (PD), Italy
Nationality	Italian
Gender	F
Skype Contact	lucy.zanotto

Current position

Since Jannuary 2017	**Postdoctoral Research Fellow**, *Departement of Statistical Science*, University of Padova, Italy
Since Jannuary 2014	**Ph.D Candidate in Statistical Science, admitted to the final exam**, *Departement of Statistical Science*, University of Padova, Italy

Education

2009–2012	**Master Degree in Statistical Science**, *Departement of Statistical Science*, University of Padova, Italy. Final mark: 110/110 cum laude
2006–2009	**Bachelor Degree in Statistics, Population and Society**, *Faculty of Statistical Science*, University of Padova, Italy. Final mark: 110/110
2001–2006	**Hight Shool**, *Liceo Socio-Psico Pedagogico*, Amedeo Duca d'Aosta, Padova, Italy. Final mark: 90/100

Masters Thesis

Title	*Kurtosis Indices for the Multinomial Skew Normal Distribution*
Supervisors	Professor Bruno Scarpa

Description Usually, to study a distribution it is useful to compute median, variance, skewness and kurtosis. The last one is the weight measure of tails comparing with the Gaussian function. Unlike univariate case, in multivariate field is difficult to have a unique formula to calculate it, so several authors have proposed different measures. We compute and compare this indices for the Multinomial Skew-Normal distribution, that includes the Normal one, as particular case. We note that, when the symmetry is modified, the probability function is always leptocurtic. Moreover we show in which the measures are different and suggest when is suitable employed them.

Experience

Academic

2014–Present **Ph.D Student**, Departement of Statistical Science, University of Padova (PD), Italy. *Research project: Development a new model for the distribution of deaths Supervisor: Prof. Stefano Mazzuco.*
A mixture parametric model is employed to fit the distribution of deaths in the life table, which is caracterized by three different component: infant, accidental and adult mortality. The model is the sum of an Half Normal distribution and a Bimodal Skew-Normal Distribution.

2015–2016 **Visiting Period**: Max-Planck Odense Center on the Biodemography of Aging (MaxO), Southern Danish University, Odense (DK).
Supervisor: Prof. Vladimir Canudas-Romo

March–June 2016 **Research Assistant**: Max-Planck Odense Center on the Biodemography of Aging (MaxO), Southern Danish University, Odense (DK)

2012–2013 **Research Assistant**, Departement of Statistical Science, University of Padova (PD), Italy. *Research project: Generalization of Skew-Normal Distribution to estimate mortality data.*
We use a Bimodal Skew-Normal distribution conbined with a Half Normal to fit the Italian data of deaths. The parameters of the model are estimated with the maximum likelihood.

Teaching

2016-2017 **Teacher**, Departement of Statistical Science, University of Padova (PD), Italy
For the course "Popolazione e mutamento socio-economico", lectures with R about the main demographic aspects: rate (age specific and not), sex pyramide, life table, total fertility rate, population forecasts.

June–July 2016 **Teacher**, School for advanced statistics courses, University of Asti (AT), Italy
Lectures and exercises with R software about different methods for data anlalysis.

2009–2008 **Tutor**, Departemet of Statistical Science, University of Padova (PD), Italy
Exercises and small lectures for undergraduate students.

Business

2013–2014 **Assisistant Key Account Manager**, GENERTELLIFE, Mogliano Veneto (TR), Italy.
Activities: trimestral report to analyse sale trends, competitor analysis, marketing, direct assistance to Banca Generali about insurance policies, development of new products.
Supervisor: Dr. Bianca Maria Spinato

.

Pubblications

Papers in preparation

2016 Zanotto L, Canudas-Romo V., Mazzuco S., *A mixture-function mortality model: illustration of the evolution of Premature Mortality* (in preparation).

2016 Mazzuco S., Scarpa B., Zanotto L., *A mortality model based on a mixture distribution function* (in preparation).

Papers published

2016 Mazzuco S. and Zanotto L., *Lo studio allunga la vita o la salva?*, Neodemos http://www.neodemos.info/lo-studio-allunga-la-vita-o-la-salva/.

2012 Multiple Authors, *Crisi dei distretti industriali veneti*, Quaderni di ricerca, Unioncamere Veneto
http://www.unioncameredelveneto.it/userfiles/ID191-QdR16xweb.pdf

Talk in a conferences

2016 Zanotto L, Canudas-Romo V., Mazzuco S., *Evolution of Premature Mortality*, EPC conference, Mainz (Germany).

Poster in a conferences

2016 Mazzuco S., Scarpa B., Zanotto L., *A mortality model based on a mixture distribution function*, EPC conference, Mainz (Germany).

Awards

October 2016 Best 3 minutes presentation – Cycle XXXII Opening, Departement of Statistical Science, University of Padova

Languages and Computer skills

Languages **Italin** (mothertongue), **English** (intermediate)

Computer skills **Statistical Software**: R (good), SAS (basic).
Suite Software: Office (excelent).
Operative System: Mac OS X, Windows.
Document markup language: Latex (good).

About me

Statistics is a powerful way to understand, know and choose. I like practical questions and issues to solve because they are a source of motivation to find out new things.

I am a very curious, resourceful and I try to do everything with passion and perseverance, in particular when results do not came immediately. Respect and honesty are fundamental values for me. I am able to work in team and by myself, also in stressful situation. I like visit places and meet people: I am fascinated from different cultures.
On my night table a book never misses.

Autorizzo il trattamento dei miei dati personali, ai sensi del D.lgs. 196 del 30 giugno 2003.

November 28^{th}, 2016

Lucia Zanotto