Quantifying biometric life insurance risks with non-parametric smoothing methods

Tomas, J.

Citation for published version (APA):
Chapter 1

Context and motivations

1.1 Context

Outside of the world of property or liability insurance, life insurance occupies a separate place that it deserves in more ways than one. It emerges as an atypical island teeming with singularities. We can report for example a legal environment of its own, dedicated accounting rules, a specific technical approach, and more generally, principles of functioning that diverge from the foundational philosophy of other branches. In a life insurance contract, the concepts of injury, repair or compensation remain absent in the contractual terms. The guarantees are fixed and freely consented in advance at the time of subscription. Benefits are paid without reference to a financial damage sustained or caused. This positioning also leads to the idea that one can give a value to life and this heretic idea was not easy to admit.

In the following, we present the heuristic evolution of the analysis of mortality. We discuss briefly the mathematical developments and mental changes toward viewing death as a proper subject of human and mathematical investigation and not the concern of god alone. With few exceptions it was mathematicians and astronomers who built the mortality table that deserves to be considered as one of the crowning achievements of the scientific revolution.

1.1.1 The origins of life tables and population dynamics studies

«From these Considerations I have formed the Adjoined Table, whose Uses are manifold, and give a more just Idea of the State and Condition of Manking, than nay thing yet extant that I know of. It exhibits the Number of People in the City of Breslaw of all Ages, from the Birth to extream Old Age, and thereby shews the chances of mortality at all Ages, and likewise how to make a certain Estimate of the value of Annuities for Lives, which
hitherto has been only done by an imaginary valuation: Also the
Chances that there are that a Person of any Age proposed does
live to any other Age given; with many more, as I shall hereafter
shew.».

Halley (1693, p.600)

The idea that one can give a value to life runs up through the history
against ethical, religious and political considerations leading to prohibit this
life insurance, viewed as intrinsically immoral, *malum omen non est provid-
endum*.

In the late Middle Ages, the traditional christian conception of death forbids
speculation about it, and thus the idea that there may be laws - other than
god - that can explain it. This christian view of a *divine order* - which was
that a man died by the will of god who offered the paradise as a reward or hell
as a damnation - seemed to respond to an older belief for which death was
following physical and deterministic laws. As recalled by Charpentier (2007)
the first civilization of Mesopotamia believed in the concept of climacteric
age, meaning a critical year marked by fatal accidents in which astrologists
claim that considerable alterations appear in the body that leads to illness
and death. The climacteric ages are multiples of seven or nine where the
danger of death is much larger than the others. This idea, born from astro-
logists, is found as well in Europe and Japan, and among philosophers and
mathematicians like Gottfried Wilhem Leibniz, see Rohrbasser and Véron
(1998, p.32). Briefly, the idea that there are physical *laws* for the death or
accidents, although contested by the christians, is relatively old.

Insurances linked to life expectancy requires the existence of tables. How-
ever at the beginning, such tables have appeared to answer other needs.
The idea has sprouted in Rome. In the early 3rd century, the jurist Ulpian
(Dometius Ulpianus), perhaps to be considered as the father of actuaries, de-
vised a table for the legal conversion of a life annuity to an annuity certain
and identified that the values of annuities should be based on the age of the
beneficiaries. But it was much later that these tables were created. To build
a table, one needed a census to know the distribution of a population by age
(with reliable years of birth). If some brilliant mathematicians have done
much for the conceptualization of probabilities, we must remember that is a
merchant, John Grant and his friend William Petty, one of the founders of
the Royal Society in London, who first conceived the notion of a mortality
table. However, Le Bras (2000) asks who between John Grant, and William
Petty has first conceived this notion? The question would be insignificant
but for a philosophical issue about the role of demography. Le Bras (2000)
explains that John Grant represents the *plebeian* who works with a scientific
method away from the oligarchy. While William Petty is close to the political
power, he has succeeded in the oligarchy instead of following a modest and
detached existence such as expected from scientists. In other words, by the
choice of its founding hero, demography is defined either as a pure science or as an instrument at the service of a state, because we should remember that since the 17th century, the population represents the wealth of nations and the power of the states. William Petty understood that this new science referred to a political project and not the converse.

The political origin of the life table is English, but the economical origin appeared in the Netherlands. Johan de Witt, in 1671, implemented a method rather pragmatic and empirical to calculate the annuities. His method allowed many mathematicians to address the issue by introducing probabilities on the duration of human life.

If the notion of life expectancy has arisen for the first time in 1746 in the work of Antoine Deparcieux, "Essai sur les probabilités de la durée de la vie humaine" see Charpentier (2007), Véron and Rohrbasser (2000, p.11) note that calculations done by Lodewijk Huygens appear in his mail in 1669 with his brother Christiaan where he estimated that his brother will live until the age of 56 and half, and him only until age 55.

At Breslau (belonging to the Habsburg empire, now in Poland and called Wroclaw), registers of births and deaths according to gender and age had been kept since the end of the 16th century. Hald (1990) recalls that a prominent evangelical pastor and scientist, Caspar Neumann, used the list, in 1687 and the following years, in his attempts to fight popular superstitions about the influence on health of the phases of the moon and the climacteric ages.

Neumann sent his results to Leibniz, who in 1689 informed Henry Justell, secretary of the Royal Society in London, of Neumann researches, see Dupâquier (1985). Justell therefore wrote to Neumann who responded by sending his observations for each of the years 1687-1691. The Society asked Edmond Halley to analyze the data and Halley (1693) presented a table with the number of people living in an age group. From this material, some figures of modern science hypothesized the first age patterns of adult mortality and deduced the associated life tables, i.e, the corresponding survivors.

In 1740, Nicolaas Struyck pointed out that the value of annuities should be calculated from life tables based on observations (as done by Halley) and not from hypotheses (as done by de Witt) see Hald (1990, p.395). However, he considered the construction of Halley’s table as unsatisfactory because Halley had access to the number of deaths only and not to the corresponding number of living. He wished to provide a reliable life table for annuitants. His observations comprise 794 male and 876 female annuitants who bought their annuities in Amsterdam in 1672-1674 and 1686-1689. For each five-year group, he tabulates the number of annuitants entering at a given age and the number of survivors at any later age. Assuming that mortality at a given age did not change over time, he summed the number exposed to risk and the number of deaths for each age group. He calculated the rates of mortality from which he derived a table that corresponds to the form still used today. He stressed that the mortality of females was smaller than that of males and presented the first life tables for males and females separately.
Perhaps the first statistical results to be taken seriously were the Northampton tables of 1780, devised by Richard Price. He worked from parish registers in Northampton, and produced corresponding tables. Price’s tables were not very conservative for the annuities. In 1808 the British government, hard-pressed by war and inflation, decided to issue annuities based on Price’s Tables. Hence it lost millions of pounds because people lived longer than was implied by the table, see Hacking (1975, p113-114). But the first table that became the usual standard of British and American insurance companies for nearly a century is the table known as the Carlisle table, built in 1815 by Joshua Milne on the basis of statistics from parishes in Carlisle.

This element of the panoply of the perfect actuary is so essential today that it is sometimes hard to imagine that it has only more than two centuries of existence. In fact the invention was not so simple, as we have seen. It is the result of the meeting of two favorable events. The first is the scientific invention of probabilities. The second, much more down to earth, is the growing need for actuaries to refine their calculation of annuities. Thus, from 1662 to 1766, from Grant and Petty to Depracieux and Milne, through Leibniz and the Huygens brothers, Halley and Struyk, actuaries on one side and mathematicians and astronomers on the other tackled the same questions about the duration of life, each bringing his stone to the edifice, and finally built in 100 years, after many hesitations, the mortality table.

Figure 1.1 compares the survival functions at birth issued from the different tables. We note that the survival curves move towards a rectangular shape. We use the term rectangularization to describe this feature: the more the time passed, the more the probability of death becomes flat at younger ages (one died rarely before 60 years), then much more brutal one the end. The point of maximum downward slope of the survival curve progressively moves toward the very old ages. This feature is called the expansion of the survival function, see Pitacco et al. (2009, p.53).

![Figure 1.1: Survival functions at birth issued from the different tables. Source: Hald (1990), Halley (1693), Gompertz (1825) and Gompertz (1871)](image-url)
Around 1870, demographers particularly in Germany felt the need for a simple chart to present population dynamics, especially in view of establishing life table formulas. This chart is known as the *Lexis Diagram*, but it is a misnomer according to Vandeschrick (2001).

To be useful, this chart must allow for location on one plane of three coordinates used to classify deaths and survivors, namely: the date, the age and the moment of birth.

Briefly, there were three solutions for this problem: In 1869, Gustav Zeuner worked out a first solution. In 1870, Otto Brasche proposed a second one with networks of parallels; his version is the most currently used now. In 1874, Karl Becker proposed a third one. In 1875, Wilhelm Lexis took back Zeuner’s diagram and just added networks of parallels. In spite of all this, the name *Lexis Diagram* is now used universally.

**Figure 1.2:** Left panel: *Lexis diagram containing life-times for birth cohorts of* $t-1$ and $t$. Each individual is presented as a line in a time-age plane, and points denote the death for a given individual. **Right panel:** *Lexis diagram containing counts of events pertaining to birth cohorts of* $t-1$ to $t$.

Figure 1.2, left panel, shows a simplified version of a Lexis diagram. In this diagram, an individual life history is drawn as a line segment with slope 1. This line starts on the horizontal axis at the time of birth and ends at the time of death. The value on the vertical axis is the individual’s age. Hence a life-time starts at zero (birth) and ends at the age of death. In this way data are properly represented according to the three demographic coordinates. The individual life-time can be regrouped and hence the *Lexis Diagram* also allows a summary of aggregated death and population data by age, period and cohort. For instance, in Figure 1.2, right panel, from the birth cohort of six births during period $t$: (1) death in $t$ and five survivors to the beginning of the following period $t+1$; (2) deaths at age 0 in $t+1$ and three survivors to age 1; (1) death to the cohort at age 1 during $t+1$ and two survivors to the beginning of the period $t+2$. The *Lexis Diagram* has become a standard tool for summarizing population dynamics.
1.1.2 Measures of mortality: Notation

Probabilities of survival and death

This section makes precise the notation used in this dissertation to quantify the biometric life insurance risks. We refer to Pitacco et al. (2009) for more details. The age at which a person will die is obviously unknown. At most we can evaluate, for a particular population, the risk of death in a given time interval. Death is then viewed as an event whose occurrence is probabilistic in nature and it is natural to resort to a mathematical framework and probabilities calculus to describe the life time of individuals.

We consider a person aged $x$, and denote by $T_x$ the random variable representing his/her remaining lifetime. In actuarial notation, probabilities like $\Pr[T_x > h]$ and $\Pr[h < T_x \leq h + k]$ are usually involved. When a life table is available, these probabilities can be immediately derived from the life table itself, provided that the ages and durations are integers.

In life insurance mathematics, a specific notation is commonly used for the probabilities of survival or death. The notation for the survival probability is as follows,

$$h p_x = \Pr[T_x > h],$$

where $h$ is an integer. (1.1)

In particular $1 p_x$ is simply denoted $p_x$. Trivially $0 p_x = 1$.

The notation for the probability of death is as follows,

$$h|k q_x = \Pr[h < T_x \leq h + k].$$

(1.2)

If $h = 0$, the notation $k q_x$ is used. In particular, when $h = 0$ and $k = 1$, the symbol $q_x$ is commonly adopted. Clearly, $0 q_x = 0$.

Note that in all symbols, the right-hand side subscript denotes the age being considered. Conversely, the left-hand side subscript refers to the duration, whose meaning depends on the specific probability addressed. The purpose of measuring the life span or conversely the mortality is to enable inferences to be drawn about the likelihood of death occurring within a specific population during a specific period of time. It is natural, therefore, for the basic measure to be expressed in proportional terms. The denominator (of which the numerator is the relevant number of deaths) is commonly referred to as population at risk or exposed to risk. To be specific, let us assume that we are given the number of deaths recorded, $d_x$, and the number of individuals initially exposed to the risk of death, $l_x$, all aged $x$ last birthday, and that our experience, for simplicity, is limited to this single age $x$, where $x = 1, 2, \ldots, n$. The observed estimate of the one-year probability of death is denoted by $q_x$,

$$q_x = 1 - \frac{l_{x+1}}{l_x} = \frac{d_x}{l_x}. \quad (1.3)$$

In Figure 1.2, let $Z_{AD}$ the number of life-lines crossing segments $AD$ and $Y_{ABCD}$ the number of deaths in the square $ABCD$, then equation (1.3) is $Y_{ABCD}/Z_{AD}$. For the observed annual survival probability, we have

$$p_x = 1 - q_x.$$
In general for the observed survival probability, we have,

\[
h p_x = p_x \frac{p_{x+1} \ldots p_{x+h-1}}{l_x} = \frac{l_{x+h}}{l_x},
\]

while for the observed probability of dying,

\[
k q_x = 1 - k p_x = 1 - \frac{l_x + k}{l_x},
\]

and

\[
h | k q_x = h p_x k q_x + h = \frac{l_{x+h} - l_{x+h+k}}{l_x}.
\]

### Survival function

Suppose that we have to evaluate the probability of survival and of dying when age and times are real numbers. Tools other than the life table are then needed. We move now to an age-continuous context. We call \(S(t)\) the survival function and define it for \(t \geq 0\) as follows,

\[
S(t) = \mathbb{P}[T_0 > t],
\]

where \(T_0\) denotes the random lifetime for a newborn. Considering the probability (1.1), we have

\[
\mathbb{P}[T_x > h] = \mathbb{P}[T_0 > x + h | T_0 > x] = \frac{\mathbb{P}[T_0 > x + h]}{\mathbb{P}[T_0 > x]},
\]

and thus

\[
h p_x = \frac{S(x + h)}{S(x)}.
\]

For the probability (1.2), we obtain

\[
h | k q_x = \frac{S(x + h) - S(x + h + k)}{S(x)},
\]

and in particular,

\[
k q_x = \frac{S(x) - S(x + k)}{S(x)}.
\]

Turning back to the mortality table, we note that since \(l_x\) is the expected number of people alive at age \(x\) out of a cohort initially composed of \(l_0\) individuals, we have

\[
l_x = l_0 \mathbb{P}[T_0 > x],
\]

and in terms of the survival function, \(l_x = l_0 S(x)\), provided that all individuals have the same age-pattern of mortality described by \(S(x)\). Thus, the \(l_x\)’s are proportional to the values which the survival function takes on integer ages \(x\), and so the mortality table can be interpreted as a tabulation of the survival function, see Pitacco et al. (2009, p.52).
Forces of mortality

We consider the probability of an individual age $x$ of dying before age $x + t$ (with $x$ and $t$ real numbers), namely $tq_x$. The force of mortality (or mortality intensity) is defined as

$$
\varphi_x = \lim_{t \to 0} \frac{\mathbb{P}[T_x \leq t]}{t} = \lim_{t \to 0} \frac{tq_x}{t},
$$

hence it represents the instantaneous rate of mortality at a given age $x$. In terms of the survival function,

$$
\varphi_x = -\frac{d}{dx} \ln S(x),
$$

so

$$
S(x) = \exp \left( -\int_0^x \varphi_u \, du \right).
$$

Central death rates

The behavior of the force of mortality in the interval $(x, x + 1)$ can be summarized by the central death rate at age $x$,

$$
m_x = \frac{\int_0^1 S(x + u)\varphi_{x+u}du}{\int_0^1 S(x + u)du} = \frac{S(x) - S(x + 1)}{\int_0^1 S(x + u)du}.
$$

The integral $\int_0^1 S(x + u)du$ can be approximated using the trapezoidal rule. In Figure 1.2, let $Z_{AD}$ and $Z_{BC}$, the number of life-lines crossing segments $AD$ and $BC$ respectively, and $Y_{ABCD}$ the number of deaths in the square $ABCD$, then the central death rate is approximated by $Y_{ABCD}/((Z_{AD} + Z_{BC})/2)$, and

$$
m_x = \frac{S(x) - S(x + 1)}{(S(x) + S(x + 1))/2}.
$$

With the assumption of constant force of mortality - frequently adopted in actuarial science calculations - which assumes $\varphi_{x+t} = \varphi_x$ for $0 \leq t < 1$, we obtain, from (1.4),

$$
m_x = \varphi_x.
$$

1.1.3 Portraying mortality over age and over age and time

Portraying mortality over age

Figure 1.3 displays the one-year transformed crude probabilities of death (year 2008), logit scale, for ages $x = 0, 1, \ldots, 98$ and each gender for the dutch population provided by the Human Mortality Database (2012). The Human Mortality Database (HMD) was initiated by the Department of Demography at the University of California Berkeley, USA, and the Max Planck Institute
for Demographic Research, Rostock, Germany. This international project provides detailed mortality and population data that can be accessed online for research purposes.

Figure 1.3: Transformed crude one-year probabilities of death, logit scale, for Dutch Male (left panel) and Dutch females (right panel) in 2008. Source: HMD.

From Figure 1.3, we recognize the typical shape of a mortality curve. Mortality is highest at the extremes of age. Once the newborn infant has survived the hazard of the first days of life, the rate of mortality falls rapidly. Most of the deaths after the first days are due to exogenous causes, mainly infections and until recent times when this component has shrunk to very small proportions, the rate was a sensitive index of social conditions and of public health progress. During childhood the risk of death is very small, being very largely confined to that of the occasional lethal infection, which modern treatments have made extremely rare, and severe accidental injuries to which child risk recklessness or lack of adult care sometimes leads. In adolescence, the impact and strain of industrial and urban life bring a rise in mortality. These and other factors, inherent in the social and economic environment and individual ways of life, reacting upon constitutional weakness, lead to a continuing increase in the risk of death as age advances. At later ages, the wearing out of the human frame rather than inimical qualities of the environment becomes the dominant cause of mortality, see Benjamin and Pollard (1980).

We show in Figure 1.3 the difference in the patterns of mortality for the two genders. The death rates for females are lower than those for males at all ages. (Before 1890 there was an excess in the death rate of females at adolescence and early adult ages mainly associated with the heavier mortality from tuberculosis in girls). Briefly, the higher mortality of males may be explained in medical terms as follows, see Benjamin and Pollard (1980) for more details.

In infancy and early childhood, boys are generally more vulnerable to some birth hazards (prematurity, malformation, birth injury), to infection (possibly as a result of some biological factors) and to injuries (possibly as a
result of more vigorous and venturesome activities). These are the principal causes of death at those ages.

In early and middle adult life, the principal causes of death are accidents and violence, heart diseases and cancers. The higher risk for accidents must be regarded as occupational in the broader sense of including, as compared with females, more outdoor movement in traffic for instance, as well as greater industrial hazards.

At more advanced ages, the process of physical deterioration and lessening resistance to disease associated with general wear and tear appear to proceed faster in men. Age for age, cerebral hemorrhages, arterial diseases, cancers (especially of the lungs) and bronchitis take a heavier toll of males than females. Some, at least, of this excess mortality has been self inflicted by cigarette smoking. The contemporary increase in industrial countries of mortality cancer of the lung and coronary arterial disease (especially for men) has been exercising considerable influence on the shape of the curve of death rates with age.

**Portraying mortality over age and time**

Figures 1.4 and 1.5 display the mortality surfaces and level plots for the Dutch males and females respectively. We see that the surface is subjected to period shocks corresponding to wars, epidemics, summer heat waves, and so on. It is apparent that dramatic changes in mortality have occurred over the 20th century, as illustrated by the downward trends and variations in shape.

![Figure 1.4](image)

**Figure 1.4:** Surface and level plot of the observed one-year probabilities of death, logit scale, for Dutch males, period 1850-2008. Source: HMD.

Figures 1.6 and 1.7 depict the observed annual probabilities of death, for some selected periods. The mortality has decreased for both sexes and all ages without interruption, primarily due to the control of infectious diseases.
This reduction is stronger for the young ages. The decrease over time at ages 20-30 for the females reflects the rapid decline in childbearing mortality. However, the hump in mortality around ages 18-25 has become increasingly important especially for the young males. The increase of life expectancy has continued to the late 20th century with the decline in mortality at the highest ages, mainly due to the reduction of mortality from cardiovascular diseases.

The trend in the observed annual probabilities of death are displayed in Figures 1.8 and 1.9, for Dutch males and females respectively. When we examine Figure 1.8, we see different behavior for age-specific probabilities of death affecting Dutch males. At age 20, a rapid reduction in mortality took place after a peak in the early 1940s due the World War II. However,

since the 1950s, only modest improvements have occurred. This is typical for ages around the accident hump, as explained in Pitacco et al. (2009, p.98); male mortality has not really decreased since the 1970s. We even observe an increase of the mortality. This unfavorable evolution is due to the increase of traffic accidents particularly acute in the 1960s. Between 1980 and the mid-1990s, the apparition of AIDS had a negative influence on the reduction of mortality. At age 40, the same decrease is present after the World War II, followed by a much slower reduction in mortality after 1960. The decrease after 1970 is more marked than at age 20. At age 60, the mortality rates have declined rapidly after 1970, whereas the decreasing during 1850-1970 was more moderate. At age 80, this decrease appears after 1990.

Figure 1.8: Trend in the observed probabilities of death, logit scale, for Dutch males at ages 20, 40, 60 and 80, period 1850-2008. Source: HMD.

The analysis for the Dutch females is similar to the one for the male population for age 20 and 40, but with several differences. At age 20, a structural
break seems to have occurred, with a relatively high level of mortality before the second world war and a much lower one after 1950. Then after the mid-1950s modest improvements are visible. At age 40, the decline is more pronounced after 1960 than for the male population. At age 60, the rate of decrease is more regular. At age 80, after 1950, the trend in the reduction of mortality has tended to accelerate.

Until 1980, females have benefited more from the reduction of mortality than males, and the gap in life expectancy has widened significantly between the genders. Nevertheless, in the last three decades, the gap has stabilized and begun to decline. This reduction is essentially due to an acceleration in the improvement among the males and some slowing of the improvement among females under age 60. At the later ages, on the other hand, improvement continued to be more rapid for females than males. Although cancer mortality is falling for both men and women, cancer is now the leading cause of death, overtaking cardiovascular diseases, for which mortality has considerably reduced, see Meslé (2006). Future improvement will depend on success in the control of cancer and neuron-degenerative diseases.

1.1.4 The irregularities in the progression of the observed rates

The symbol $q_x$ represents the one-year observed probability of death for a particular population at age $x$. It lies above or below the true underlying value. From Figures 1.4 and 1.5, the roughness of the surface indicates volatility. In estimating mortality, the actuary knows that the past experience from which the observed mortality rates and the life table have been derived will never be exactly reproduced in the future. Thus a certain random element of fluctuation will be inherent in the observations and the smaller the group, the greater will be the relative random errors in the deaths and the less reliable will be the resulting estimates.
These deviations from the true underlying rates may be assumed to be random and to fluctuate from age to age both in size and sign. These irregularities in the progression of the observed rates of mortality could be reduced by increasing the number $l_x$ of persons observed. If the number of individuals in the group had been considerably larger, the set of observed probabilities, $q_x$, would have displayed a much more regular progression with $x$. In the limit, it would have exhibited a smooth progression, as explained in Copas and Haberman (1983).

The idea of a group of persons attaining age $x$ and being gradually reduced in numbers, until they are all dead, in such a way that the rates of mortality at successive ages form a smooth series is a purely theoretical conception. It is nevertheless a very useful conception, as recalls Alistair (1989), which forms the basis of the theory of life contingencies and has been shown by long use to be suitable for solving most actuarial problems in life insurance. This is not to suggest that measurement can be allowed to be inexact. On the contrary, as Benjamin and Pollard (1980) mention, if judgment has to be introduced in any final estimation, it is likely to be sounder when on the basis of adequate analysis of past experience.

Provided these errors are random in nature, they may be reduced by increasing the size of the sample and thereby extending the scope of the investigation. A simpler, cheaper and more practicable alternative is often to use graduation to partly remove these random errors. Thus, by graduating the mortality rates, we aim to concentrate on the underlying mortality pattern (high mortality at birth, low infant mortality, accident hump, senescence effect) avoiding the erratic departures from it.

Various approaches to graduation can be adopted. In particular, two broad categories can be recognized:

i. Parametric approaches, involving the use of mortality laws; Hannerz (2001) defines a mortality law as a mathematical expression that describes mortality as a function of age.

ii. Non-parametric approaches.

1.2 Motivations

1.2.1 Getting out of a procrustean bed of fixed parametrization: From parametric to smooth models

Assume $n$ pairs of observations $\{(x_i, q_i)\}_{i=1}^n$ have been collected, then the regression relationship can be modeled as

$$q_i = f(x_i) + u_i, \quad i = 1, 2, \ldots, n;$$

with the unknown regression function $f$ and an error term $u_i$, representing random errors in the observations or variability from sources not included in the $x_i$. 
The aim of a regression analysis is to produce a reasonable analysis of the unknown response function \( f \). This task of approximating the mean function can be done essentially in two ways. The quite often used *parametric* approach is to assume that the mean curve \( f \) has some pre-specified functional form, for instance, a line with unknown slope and intercept. As an alternative one could try to estimate \( f \) *non-parametrically* without reference to a specific form.

The first approach to analyze a regression relationship is called parametric since it assumes that the functional form (for example, Thiele law, Perks laws, Gompertz-Makeham class of models, and so on) is fully described by a finite set of parameters. A tacit assumption of the parametric approach though is that the curve can be represented in terms of the parametric model or that, at least, it is believed that the approximation bias of the best parametric fit is a negligible quantity. Such laws simplify the calculation of mortality functions and allow to extrapolate at the highest ages for instance. But to be useful, they have to reproduce closely the data. According to Alistair (1989) it is now thought that it is unlikely that a law can be found that represents the mortality rate over a large range of ages, although some complicated expressions have been used in the attempt.

By contrast, non-parametric modeling of regression relationship does not project the observed data into a Procrustean bed of a fixed parametrization. A preselected parametric model might be too restricted or too low-dimensional to fit unexpected features, whereas the non-parametric approach offers a flexible tool in analyzing unknown regression relationship. The term *non-parametric* thus refers to the flexible functional form of the regression curve. Like parametric methods, they too are liable to give biased estimates, but in such a way that it is possible to balance an increase in bias with a decrease in sampling variation.

The question of which approach should be taken in data analysis was a key issue in a bitter fight between Pearson and Fisher in the 1920’s, as recalls Härdle (1990). Fisher pointed out that the non-parametric approach gave generally poor efficiency whereas Pearson was more concerned about the specification question. Both points of view are interesting in their own right. Pearson pointed out that the price we have to pay for pure parametric fitting is the possibly of gross misspecification resulting in too high model bias. On the other hand, Fisher was concerned about a too pure consideration of parameter-free models which may result in more variable estimates, especially for small sample size.

### 1.2.2 *Natura non agit per saltum*: The basic idea of smoothing

We have previously seen that the crude rates can be seen as a sample from a larger population of lives and thus they contain some random fluctuations.
If we believed that the true rates were independent, then the crude rates would be our final estimate of the true underlying mortality rates. However, a common prior opinion about the form of the true rates is that each true rate of mortality is closely related to its neighbors, that is the observations $q_j$ in the neighborhood of the target point $q_i$ should contain information about the value of $f$ at $x_i$. Gavin et al. (1993) explain that this relationship is expressed by the belief that the true rates progress smoothly from one age to the next.

Benjamin and Pollard (1980) recall the saying, *Natura non agit per saltum*, which expresses the fact that natural forces operate gradually and their effects become apparent continuously and not in sudden jumps. It follows that the data for several ages $x_j$ on either side of age $x_i$ can be used to augment the basic information we have at age $x_i$, and an improved estimate of $q_i$ can be obtained by smoothing the individual estimates.

So the next step is to graduate the crude rates in order to remove any random fluctuation. This procedure of approximation of the mean response curve $f()$ is commonly called *smoothing*. Hence, the mortality is not summarized by a small number of parameters, but described by the $n$ annual probabilities of dying. It may be considered as a compromise between faith in the data and reduced roughness caused by the noise. In the actuarial literature, the process of smoothing a mortality table was known as graduating the data, the little hills and valleys of the rough data were to be graded into smoothness, just as in building a road over rough terrain.

The concept of smoothness has been used in the previous paragraphs without actually being defined. We deliberately avoid a detailed presentation here. The interested reader can have a look at Bizley (1958) and Diewert and Wales (2006). We all have an intuitive idea about what we mean by smooth, as for instance the road building analogy. Formal mathematical analysis may state the smoothness condition as a bound on derivatives of $f$. Bizley (1958) observes that smoothness is intimately concerned with predictability, and proposes the following definition of smoothness: a continuous curve is smooth at the points for which the absolute value of the rate of change of curvature with respect to distance measured along the curve is small. For Benjamin and Pollard (1980), the Bizley’s requirements of small change of curvature turns out to be equivalent in the mortality context to requiring that third-order differences are small.

### 1.2.3 Smoothers and parameters selection

Smoothing alone, however, is not graduation. Graduated rates must be representative of the underlying data. The two qualities, *smoothness* and *goodness of fit*, tend to conflict, in the sense that smoothness may not be improved beyond a certain point without some sacrifice of goodness of fit and vice versa. Thus, a graduation will often turn out to be a compromise between optimal fit and optimal smoothness.
To be useful, the method should allow the graduator some latitude in choosing the relative emphasis to place smoothness and fit.

Special attention has to be paid to the fact that smoothers, by definition, average over observations with different mean values. The amount of averaging is controlled by a weight sequence which is tuned by a smoothing parameter, denoted $\lambda$. This smoothing parameter regulates the size of the neighborhood around the target point $x_i$.

A local average over a too large neighborhood would cast away the good with the bad. In this situation an extremely over-smooth curve would be produced, resulting in a wrong estimate $\hat{f}$. On the other hand, defining the smoothing parameter so that it corresponds to a very small neighborhood would not sift the chaff from the wheat. Only a small number of observations would contribute non-negligibly to the estimate $\hat{f}(x_i)$ at $x_i$ making it very rough and wiggly. In this case the variability of $\hat{f}(x_i)$ would be inflated. Finding the choice of smoothing parameter that balances the trade-off between over-smoothing and under-smoothing is called the smoothing parameters selection problem. To give insight into the smoothing parameters selection problem, consider Figure 1.10 below.

![Figure 1.10: Estimated curve and transformed crude mortality rates (dots), logit scale, for Dutch Male 2008. Source: HMD.](image-url)
The curves represent non-parametric estimates of the mortality rates. The more wiggly curve has been computed using a local polynomials estimate with a very small neighborhood. By contrast, the flatter curve has been computed using a very large neighborhood. Which smoothing parameter is correct? The question will be discussed in Section 2.5.

1.2.4 Historical review of the development of smoothing approaches

The problem of smoothing sequences of observations is relevant in many branches of sciences. In the following, we review the development of smoothing methods starting in the late 18th to the 21st centuries, leading up to the development of the use of local polynomial regression and afterward local likelihood methods.

Early work

Local regression is a natural extension of parametric fitting, so natural that local regression arose independently at different points in time and in different countries. The setting for this early work was univariate and involved equally spaced $x$. It was simple enough that good-performing smoothers could be developed and were computationally feasible by hand calculation. Also, most of the early work arose in actuarial studies, as remark Cleveland and Loader (1996). Mortality and sickness rates were smoothed as a function of age.

Haberman (1996, p.40) reports that smoothing was used as early as 1765 by the Swiss mathematician and physicist Johann Lambert. He was born in Mülhausen, now Mulhouse in Alsace, France; then an exclave of Switzerland. Daw (1980, p.347) explains in his 1765’s work (volume 1) that he graduated the value $l_x$, at decennial ages, which he had calculated from the deaths recorded in the London Bills of Mortality for 1753-1758. He does not read off the graduated values of $l_x$ at all ages from his graph, but gives two methods of graduation and/or interpolation. The first was a graphical method for introducing osculating parabolas between two points. The second was a method of fitting a polynomial of fifth degree to represent a section of the curve which was then able to hang together with the corresponding polynomials for the immediately preceding and succeeding sections of the curve. This methodology is effectively what has come to be known as osculatory interpolation, and was re-invented more than 100 years later by Thomas Sprague.

John Finlaison, subsequently first president of the Institute of Actuaries in January 1823, started preparing the mortality data that were to provide the first life table consisting of graduated observations at individual ages. His 1829 work is described by Seal (1982, p.89). His formula is based on overlapping piecewise linear arcs extending over nine successive values, with eight of the nine being used in the next arc, and thus represents the first published
example of a graduation by the adjusted-average method. This piecewise approach to smoothing was extended in 1866 by the Italian meteorologist and astronomer Giovanni Schiaparelli who assumed a cubic polynomial to extend to a stretch of consecutive observed values.

In the same year (1866), Wesley Woolhouse presented a detailed exposition of graduation of mortality rates using summation formulae, stressing the conceptual differences between graduation and interpolation. He considered the case where the fourth differences of the corrections $v_x = \hat{q}_x - q_x$ to an observed series of rates had small values and proposed to minimize $\sum v_x^2$ in terms of $\Delta^4 v_x$ and thus obtain estimates of $v_x$ and hence $\hat{q}_x$. Seal (1982, p.93) demonstrates that the equations for $\hat{q}_x$ are equivalent to those which arise from fitting piecewise cubic polynomials by least squares to equidistant observations.

The use of symmetrical moving weighted average formulae to smooth equally spaced observations of a function of one variable, which generalized Woolhouse’s summation formulae, was systematically investigated in a series of papers by the American statistician Erastus De Forest, as reports Haberman (1996, p.41). De Forest’s principal innovation was to introduce optimality criteria into the problem of estimating the coefficients.

In 1887, Thomas Sprague’s paper on the graphic method of graduation appeared. This paper rediscovered (following Lambert) osculatory interpolation showing how formulae could be devised to ensure continuity of the first derivatives of overlapping interpolation curves. Osculatory interpolation was used as a method of graduation for the English life table in the early nineteenth century.

A new style of summation graduation and its testing had started with Spencer, in 1904 and 1907, and had blossomed in Vaughan’s 1933, 1934 and 1935 articles. The method developed by Spencer in his 1904 article had become popular because it was computationally efficient and had good performance. We note three crucial properties. First, the smoother exactly reproduces cubic polynomials as explained in Cleveland and Loader (1996). Second, the smoothing coefficients are a smooth function of length of the bandwidth, and decay smoothly to zero at the ends. Third, the smoothing can be carried out by applying a sequence of smoothers each of which is simple; this was done to facilitate hand computation. Achieving all three of these properties is remarkable.

Whittaker (1923) suggested an alternative method of graduation. This can be regarded as what would now be called a Bayesian approach to graduation, see Taylor (1992). It results in the minimization of the combination of a measure of goodness of fit of the graduation to the observation and a measure of smoothness.

Modern work

We have seen that the methods presented in the introduction are inherited from a long actuarial tradition. However local regression methods received
little attention in the statistical literature until the late 1970’s.

For Cleveland and Loader (1996), the modern view of smoothing by local regression has origins in the 1950’s and 1960’s, with kernel methods introduced in the density estimation setting (Rosenblatt (1956), Parzen (1962)) and the regression setting (Watson (1964)). Kernel methods are a special case of local regression; it amounts to choosing the parametric family to consist of constant functions. Kernel methods have found actuarial application by Copas and Haberman (1983), followed by Gavin et al. (1993) and Gavin et al. (1995).

However, recognizing the weaknesses of a local constant approximation, the more general local regression enjoyed a reincarnation beginning in the late 1970’s. It includes the mathematical development of Stone (1977), Stone (1980), and the lowess procedure of Cleveland (1979). It provides a number of important insights about the choices of the smoothing parameters. For example it was nearly a given that for most applications the weight function needed to be smooth, that local constant fitting was inadequate, and that smoothers needed to reproduce exactly (and not just asymptotically) at least a quadratic.

Among other features, the local regression method and linear estimation theory trivialize problems that have proven to be major stumbling blocks for more widely studied kernel methods. The kernel estimation literature contains extensive work on bias correction methods: finding modifications that asymptotically remove dependence of the bias on the slope, curvature, and so on. Examples include boundary kernels, see Müller (1987), and higher order kernels, see Gasser et al. (1985) and Schucany (1989). Local regression methods can then be viewed as an extension of kernel methods and an attempt to extend the theory of kernel methods. This treatment has become popular in the 1990s, for example Hastie and Loader (1993) and to some extent Loader (1999b). The approach has its uses: small bandwidth asymptotic properties of local regression, such as rates of convergence and optimality theory, rely heavily on results for kernel methods. But for practical purposes, the kernel theory is of limited use, since it often provides poor approximations and requires restrictive conditions.

Furthermore, while the early smoothing work was based on an assumption of a near-Gaussian distribution, the modern view extended smoothing to other distributions. Cleveland (1979) developed robust smoothers. Later, Tibshirani and Hastie (1987) took local fitting one step further; in any situation where a dependent variable depends on independent variables, a local likelihood procedure can be carried out. Hence they substantially extended the domain of smoothing to many distributional settings such as logistic regression, and developed general fitting algorithms. The extension to new settings has continued in the 1990’s with Fan et al. (1998) and Loader (1996).
1.3 Outline of the thesis

In Chapter 2, a non-parametric graduation method is discussed. We introduce local polynomial regression. We discuss the choice of the smoothing parameters and criteria used for models selection. We graduate the data through the choice of the smoothing parameters. The graduation and corresponding confidence intervals are carried out over the entire age range. Tests are used to compare the graduated rates obtained with those obtained by the Whittaker-Henderson smoothing.

Our contribution Tomas (2012a) - that presents extensively local polynomial technique in view of graduating experience data originating from life insurance - can be viewed as the prolongation of the kernel estimation for graduation proposed by Gavin et al. (1993). It is completed in this chapter with Tomas (2012b) analyzing the influence of the boundaries on the choice of the smoothing parameters.

In Chapter 3, our aim is to extend the local smoothing technique described in Chapter 2 to model situations where a non Gaussian likelihood is appropriate. We incorporate the concepts of the non-parametric regression technique of local polynomials to localized generalized linear models and local likelihood contexts.

Related work is in Delwarde et al. (2004) and Debón et al. (2006), but our work examines the statistical properties of the estimators and the choice of the smoothing parameters by classical selectors as well as the plug-in methodology. The applications cover the graduation of both the probabilities of death and the forces of mortality over the entire age range involving historical data from the Netherlands. In addition we provide a method for constructing pointwise confidence intervals that are not depending on the estimates using the variance stabilizing link. This method allows us to produce confidence intervals in presence of zero-responses.

In Chapters 2 and 3, the weight functions have always had a fixed or global bandwidth. Rather than restricting the smoothing parameters to a fixed value, Chapter 4 discusses more flexible approaches allowing the smoothing parameters to vary across the observations. An application involving individuals subscribing long-term care insurance is presented. We analyze the incidence of mortality as a function of both the age of occurrence of the pathology and the duration of the care. We distinguish the intersection of confidence intervals rule and local bandwidth correction factors.

Part of our work is an extension of the adaptive kernel methods proposed by Gavin et al. (1995) to adaptive local kernel-weighted log-likelihoods techniques. We vary the amount of smoothing in a location dependent manner and allow adjustments based on the reliability of the data. Tests and single indices summarizing the lifetime probability distribution are used to compare the graduated series to \( p \)-splines smoothing and local likelihood models.
Chapter 5 illustrates the construction and validation of portfolio specific prospective mortality tables. We are interested in the variation of mortality with attained age and calendar year. From portfolios of several insurance companies we construct, in a first step, a global prospective reference table summarizing the mortality experience of these portfolios. We investigate the divergences in the mortality surfaces generated by a number of previously proposed models. We focus on the model risk and, to a lesser extent, on the risk of expert judgment related to the choice of the external references used. We use non-parametric method, namely local kernel-weighted log-likelihood and semi-parametric relational models, to graduate and extrapolate the surfaces. The extrapolation of the smoothed surface, obtained by local likelihood methods, is performed by identifying the mortality components and their importance over time using functional principal component analysis. Then time series methods are used to extrapolate the time-varying parameter, while semi-parametric relational models have the advantage of integrated estimation and forecasting. Tests and indices summarizing the lifetime probability distribution are used to measure the impact of model choices. The mortality of the entire population is not specific to any subpopulation. The second step of our approach is then to build entity specific prospective mortality tables by adjusting the reference table validated in the first step to the mortality of each portfolio. A Poisson generalized linear model including interactions with age and calendar year gives a solution to this problem.