Chapter 7
Modeling Human Longevity and Life Tables

Up to this point in the book, I have assumed a great fiction, namely that human longevity is known and finite. The success or failure of a retirement income plan was monitored and measured until a finite, e.g. 30 year, horizon. This chapter is the first to focus on the uncertainty or randomness in human longevity versus portfolio longevity. It begins with a detailed description and analysis of (historical) cohort life tables from the Human Mortality Database. The cohort life tables are used to extract population survival and death rates, which are then used to reconstruct life tables. The chapter concludes with a high-level discussion of mortality projections and improvements for future birth cohorts. The main emphasis is on gaining familiarity with the basic atomic structure \( (q_x) \) of actuarial life-science.

7.1 Functions Used and Defined

7.1.1 Sample of Native R Functions Used

- \texttt{diff(.)} creates a vector of differences of adjacent values.
- \texttt{which(v == q)} returns the index number for which \( v \) equals \( q \).
- \texttt{cumprod(.)} multiplies the vector elements.
- \texttt{abline(.)} plots horizontal lines.

7.1.2 User-Defined R Functions

- \texttt{SPQR(x, y, qx)} computes the survival rate (a.k.a. probability) to the age of \( y \), conditional on living at age \( x \), given the 1-year death rate vector \( qx \).
7.2 Death for Financial Economists: Motivation

Within the context of commerce, measuring life and death used to be the exclusive purview of insurance actuaries, while the serious business of managing capital between those two important dates was left to financial practitioners and economists. But today times have changed. In the last few decades the actuarial community has moved (aggressively) into the financial and economic industry as evidenced by curriculum changes for their stringent exams. In contrast, the community of financial economists and practitioners—very broadly defined—has remained somewhat ambivalent towards, and at times even hostile towards actuarial models of longevity and mortality. On an academic note, one is hard-pressed to locate a finance department in a business school, or even a proper economics department, in which life contingencies are taught in any serious way. To many financial economists, death is just a basic (Poisson) arrival process and as far as pricing is concerned, mortality is a diversifiable risk, and the rest is just noise. Yes, the recent COVID-19 crisis has seen an uptick of interest in models of mortality, even among economists and financial specialists, but for the most part it isn’t part of the formal canon of knowledge. Graduate and undergraduate students who are interested in something deeper than a toy model of life and death are directed to the mathematicians and statisticians in other faculties. I find that many economists echo Macbeth, who in Shakespeare’s famous play declares: There’s nothing serious in mortality, all is but toys, renown and grace is dead. And yet, how is a financial life-cycle economist (a.k.a. retirement income expert) supposed to have an intelligent conversation about the fair value of a pension annuity, or the role of longevity insurance in the retirement portfolio, without a proper quantitative understanding of life and death? Likewise, to have a meaningful and constructive dialogue with insurance actuaries one must understand their language, framework and world-view. In sum, the plan with for the next three Chaps. 7, 8 and 9 is to offer a practical need-to-know framework on how to approach, think about and construct models involving life and death, using R (obviously.)

7.3 Will You Live to Age 100?

The first step in a proper discussion of human longevity is to consider whether you personally will reach the age of 100. Do you have a family member, parent or grandparent who reached this age? It’s quite rare. In fact, I find it interesting that when audiences (i.e. large groups) are surveyed and asked to estimate how many people are alive (in their own country) above the age of 100, they tend to over-estimate the number (especially in Florida). Likewise, the scholarly research suggests that individuals are overly optimistic about their own survival odds to (very) advanced ages. Of course, the fact is no one really knows their true probabilities of reaching 100, or any other age. Such forecasts are not like predicting the outcome from spinning a roulette wheel or mixing a deck of cards. Those gaming
Table 7.1  Hitting 100: estimated number of centenarians around the world

| Location | Number (in 2015) | Rate per 10,000 |
|----------|------------------|-----------------|
| Australia | 4280             | 0.3             |
| Canada   | 8000             | 2.3             |
| China    | 48,000           | 0.3             |
| Japan    | 61,000           | 4.8             |
| U.S.     | 72,000           | 2.2             |
| World    | 451,000          | 0.6             |

Data source: pew research center (Accessed May 2018), [www.pewresearch.org](http://www.pewresearch.org)

From the perspective of retirement income planning, the main objectives in the next few chapters are to (1) quantify the uncertainty around human longevity, and learn how to (2) incorporate that uncertainty into your consumption, withdrawal and investment plans. To answer the first question I’ll begin by tapping into one of the best and most popular sources of historical mortality data.

### 7.4 The Human Mortality Database (HMD)

The main source of data used in the next few chapters is the Human Mortality Database (HMD), all of which can be downloaded (after you register for a free account) from [http://www.mortality.org/](http://www.mortality.org/). They collect, validate, and “clean” data on all aspects of death (yes, a bit morbid) for many different countries around the world. The numbers I will use for many of the examples are from Canada. The crude data can also be downloaded from [http://www.bdmc.umontreal.ca/](http://www.bdmc.umontreal.ca/) without registering (the last time I checked).
For ease of use, I have extracted a (very) small subset of the Canadian data and have made it available as a special .csv file, together with the other scripts and files. So, I’ll start by importing the file using the usual procedure (last described in Chap. 4). Technically speaking this is a synthetic cohort life table (CLT) for all Canadians born in the year 1925. I’ll explain exactly what these terms mean in a moment, but for now here is a reminder of the commands required to get the data into R and then stored as a local CLT variable, which is a shorter name and easier to work with. To be very clear though, the raw data available from the websites of the HMD and/or the Canadian HMD is stored in a slightly different (cumulative) format. I’ll return to the HMD formatting of the data at the very end of this chapter, in Sect. 7.13.

```r
CLT <- read.csv("./Canada_Cohort_Lifetable_1925.csv")
```

To begin with, notice the emphasis on the word COHORT, which is a group of people born in the exact same year and (technically) on the exact same date, which is why I use the word synthetic. The synthetic cohort life table tracks and monitors the life status of this group over the next 85 years (to the year 2010), and the only way to leave the table is to die. It’s important to emphasize that these (synthetic) cohort life tables ignore immigration (again, the only way out is to die) or migration, which is why the values never increase.

Now, this particular synthetic cohort life table assumes a group of 100,000 males and 100,000 females, all born on January 1st, 1925. To put that number in perspective, the total population of Canada in the year 1925 was slightly less than ten million, and was increasing by approximately 1.75% per year (albeit including immigration.) So, it’s not unreasonable to assume that a total of 200,000 people were born in the year 1925. But they certainly weren’t born on January 1st. That’s the synthetic part of the cohort. Not to get carried away by minutia, but the number of births expected in Canada during the year 2020 is approximately 400,000, which results in slightly more than 1000 births per day. And, for those readers in the USA, multiply the numbers by a factor of ten, resulting in approximately 10,000 births per day. This really is an astonishing number and places the 72,000 (American) centenarians in perspective.

Back to the synthetic cohort life table, which from here on I’ll abbreviate CLT. Some members of the original 1925 cohort are still alive today, but the data (I have used) ends in the year 2010 (at age 85). In fact, the HMD has (incomplete) cohort data going back to the year 1850, which you can (and I will ask you to) access and download for some of the end-of-chapter questions. Once you have downloaded and imported the data, ask R to summarize the CLT variable, and make sure you obtain these results. (Remember, this isn’t simulations. Your numbers should match mine.)

```r
summary(CLT)
```

| AGE | FEMALE | MALE | YEAR |
|-----|--------|------|------|
| Min. : 0.00 | Min. : 42021 | Min. : 24183 | Min. : 1925 |
| 1st Qu.:21.25 | 1st Qu.: 75343 | 1st Qu.: 64119 | 1st Qu.: 1946 |
Let me review and explain the four columns. The age of individuals represented in the CLT range from 0 to 85, located in the first column. That should be obvious by now. The second and third column contain the number of female and male survivors at every age (and year). The fourth and final column contains the years covered in this cohort life table. So, the last observation is (hypothetically, on January 1st, 2010) when a total of 42,021 females and 24,183 males were still alive at age 85 (which, hypothetically, was their birthday). Assuming your data summary matches mine, you can move on to plots and figures. In particular, I’ll graph the 2nd column CLT$FEMALE and the 3rd column CLT$MALE using the following script.

```r
plot(c(0,90),c(0,1),type="n",
     xlab="Age of Cohort (top), Year (bottom)",
     ylab="Fraction Surviving")
title("Cohort Life Table: 1925 Canada")
grid(ny=18,lty=20)
for (i in 1:86){
    points(i-1,CLT$FEMALE[i]/100000,col="red")
    points(i-1,CLT$MALE[i]/100000,col="blue")
}
abline(h=0.25,col="black",lty=3,lwd=2)
abline(h=0.50,col="black",lty=3,lwd=2)
abline(h=0.75,col="black",lty=3,lwd=2)
year<-1925+c(0:5)*20; age<-0+(0:5)*20
axis(side=1,line=1,at=age,labels=year,tick=F)
text(81,0.65,"Females",col="red")
text(58,0.65,"Males",col="blue")
```

Figure 7.1 plots the survival rate for males and females, which is the fraction of the initial cohort born on January 1st, 1925, still alive on subsequent birthdays. There is quite a bit going on in the script, including a few extra tricks I haven’t used before, so make sure you understand the new commands. As in all prior figures, I start with an empty plot type="n" and then layer on the data points. Notice the double x-axis covering the observation year as well as the chronological age of the cohort, via the `axis(.)` command. The `abline(.)` creates horizontal lines at 25%, 50%, and 75% survival, making it easier to discern ages at which those fractions survived. Finally, google the lty, lwd or col arguments to learn how they can be modified.

Notice the patterns in Fig. 7.1, which for the record are observed in every other country as well as Canada. At every age, the female survival curve (in red) is above the male survival curve (in blue). Initially the gap between the two curves is rather small, but it increases over age and time. By the year 2010, only 24% of the (1925 cohort) males have survived, but 42% of the females are still alive. Needless to say,
quite a lot of history took place during those 85 years, including World War II (in the 1940s) which reduced the size of the male cohort, who enlisted in their late teens and mid-20s. Notice the large precipitous declines in the very first year of life—from age 0 (in 1925) to age 1 (in 1926)—which captures the (very) high infant mortality rates in the first year (and few weeks) of life. That first-year drop plays a large role in the 24% (males) and 42% (females) who survive to age 85. But, if they (males or females) survived those first few years, their chances of reaching age 85 improved quite dramatically. I’ll get to that critical conditional on survival matter in just a bit. For now I’ll note that infant mortality rates (in Canada, 2020) are down to 5-per-1000 births. This number would hardly create a blip in the upper-left corner. So, Fig. 7.1 is a microcosm of the many issues addressed in the next few chapters. For example: What fraction of the (Canadian) 1925 cohort will survive to the year 2025? Needless to say, the full data hasn’t materialized yet. Or, what will Fig. 7.1 look like for the 1960s or 1980s cohort? Will you be one of those who survives to age 85 (or even 100!? ) What will your cohort’s full plot look like? Is there a continuous function that can approximate Fig. 7.1? I trust you appreciate that without some (approximate) answers to these questions it’s quite hard to build a proper retirement income plan. And, although the randomness of human longevity is fundamentally different from the randomness of portfolio longevity, the mathematical tools are quite similar.
I’ll now examine the data in CLT more closely, and start by answering the following questions. What fraction (male vs. female) of the birth cohort survived their first year of life, to reach the age of \( x = 1 \)? What fraction (male vs. female) survived to the age of \( x = 60 \)? What fraction (male vs. female) survived to the age of \( x = 75 \)? Given that we have data until the year 2010, there is no need to guess, estimate, or forecast these numbers. I count survivors and divide by the initial 100,000. Using the symbol \( S_x \) to denote the number of survivors at any age \( x \), I compute \( S_{75}/S_0 \), \( S_{60}/S_0 \) and \( S_1/S_0 \) using the following syntax in R:

```
CLT$FEMALE[CLT$AGE==75]/100000
[1] 0.631155
CLT$FEMALE[CLT$AGE==60]/100000
[1] 0.777939
CLT$FEMALE[CLT$AGE==1]/100000
[1] 0.907693
CLT$MALE[CLT$AGE==75]/100000
[1] 0.460165
CLT$MALE[CLT$AGE==60]/100000
[1] 0.682913
CLT$MALE[CLT$AGE==1]/100000
[1] 0.877325
```

So, 77.8% of the females and 68.3% of the males (in the 1925 birth cohort) survived to the age of 60, which you can think of as an *early* retirement date. The other 22.2% of females and 31.7% of males died sometime between birth and (early) retirement. Almost a third of males and a quarter of females never made it to (early) retirement, from the 1925 birth cohort. These numbers aren’t unique to Canada. The death rates are in fact higher—that is less people from the 1925 cohort survived to the year 1985—in other developed countries, including the USA and (especially) Europe. The situation (today) is much different, and I’ll discuss how and why in a moment. Moving on, the survival rate to age one was 87.7% for males, and implies that 12.3% of males born in 1925 never reached their first birthday. That is a staggering number and the source of the discontinuity in the upper corner of Fig. 7.1. I should note again that in 2020 the 1-year mortality rate at birth is a mere 0.5%, which is an average for males and females. So, that alone is one (very big) difference between the evolution of the 1925 cohort and the (future) 2020 cohort. Fewer people reached age 60 from the 1925 cohort, because many of them died in their first year of life. Naturally, a much larger fraction of the 2020 cohort will survive to the age of 60, because we lose fewer of them to infant mortality. But what fraction of the 2020 birth cohort will reach age 60? They should be (much) higher than the 77.8% (for females) and 68.3% computed in the above box, but how much higher? Keep that question on the back-burner for now.
I would like to focus on a slightly different question, namely, assuming a member of the 1925 cohort survived to the age of 40, what fraction survived to age 60? Please note the subtle conditioning. I am assuming or focusing on a subset of the group that actually reached age 40, and asking what fraction from that (smaller) group reached age 60. To answer this I must scale by the appropriate denominator, namely those who are alive at the age of 40. Mathematically, this is $S_{60}/S_{40}$, and the following script provides the answer.

\[
\frac{\text{CLT$\text{FEMALE}[\text{CLT$\text{AGE==60}] / \text{CLT$\text{FEMALE}[\text{CLT$\text{AGE==40}]}}}{\text{CLT$\text{MALE}[\text{CLT$\text{AGE==60}] / \text{CLT$\text{MALE}[\text{CLT$\text{AGE==40}]}}
\]

The 92.5% survival rate for females and 86% survival rate for males, to age 60, are much higher than the 77.8% (for females) and 68.3% computed in the earlier box. Both are associated with reaching the age of 60, but the earlier set of numbers is conditional on age zero, and the current set is conditional on age 40. That’s the main difference. This is why you should always get into the habit of asking: conditional on what? when given a survival rate. At the extreme, think of the 1-year survival rates contained in the next command.

\[
\begin{align*}
\frac{\text{CLT$\text{FEMALE}[\text{CLT$\text{AGE==60}] / \text{CLT$\text{FEMALE}[\text{CLT$\text{AGE==59}]}}}{\text{CLT$\text{MALE}[\text{CLT$\text{AGE==60}] / \text{CLT$\text{MALE}[\text{CLT$\text{AGE==59}]}}
\]
\]

Both number are well over 98%, and by this point in the narrative I hope you understand why. The fraction of 59 year-olds who survive to age 60 is quite high. The closer you get to some target age (60 for example) the higher the odds of getting there. It’s tautological, in a sense. Warning, though. It’s common to interpret the (backward looking, historical) survival rates as forward-looking survival probabilities, but one should be careful with that extrapolation. At this point all I’m saying is that the survival rates for the 1925 cohort, to any fixed age, increased as they get closer to that age. I’m not making statements about probability (yet.)

### 7.6 Another Birth Year: Another Cohort

The (Canadian) 1925 CLT tracks the longevity of babies born in 1925. But, as I mentioned a number of times already, cohorts born in other years (either before or after) experience (very) different trajectories of life and death. For example, a non-trivial segment of the 1925 cohort fought (and died) during World War II. The same can’t be said of the 1940 cohort, who would have been too young to fight. The 1940
7.6 Another Birth Year: Another Cohort

To examine the evolution of survival for the 1940 cohort versus the 1925 cohort, I have created a corresponding 1940 cohort life table, which you can import (and then plot) using the same procedure I described earlier. Figures 7.2 (females alone) and 7.3 (males alone) display the fraction of survivors at each age for both the 1925 and 1940 cohort, in the same picture. Obviously, the 1940 cohort only has data until age 70, which gets them to the year 2010.

Although both curves trend downward over time (a.k.a. people can die), there are a number of visual differences between the 1925 and 1940 birth cohort. First, the 1940 curve is consistently above the 1925 curve, which means that the survival rate to any given age is higher—and the implied death rate is lower—for the later cohort. This reduction or decline in mortality (a.k.a. improvement) isn’t unique to the 1940 versus the 1925 cohort, and is actually observed in (almost) all subsequent birth cohorts around the world. And, while it’s often difficult to see the change (i.e. the gap between the curves) from 1 year to the next, it’s quite evident when examined on a 5-year basis, as you can see from these figures. Now, whether or not this mortality improvement will continue into the future at the same pace, for example, the 2010 birth cohort versus the 2020 birth cohort is a subject of much debate and discussion in the medical, actuarial, and demographic community. Why the debate? Well, for example, look closely at the upper-left corner of both Figs. 7.2 and 7.3. One of the main differences between the 1925 and 1940 curves is the (above noted) first-year infant mortality rate. That creates a large gap between the curves and is a source of the improvements. But, over the last few decades this number (infant mortality) has been reduced quite dramatically (in the developed
Males: the 1925 versus the 1940 Canadian birth cohorts

world) and is unlikely to be improved much more. This then implies that it’s unlikely to have as large an influence on future cohorts and the gap between the two curves will shrink—or possibly even reverse itself. The same argument can be made about other causes of death (declining incidence of large-scale wars?) which might increase survival rates in future cohorts. Alternatively, deaths from opioid and pain medication overdoses (which were negligible until recently) might reduce the survival rate of (for example) the 1980 cohort versus the 1940 cohort.

In sum, I don’t want to get entangled or caught up in the demographic minutia, but the important takeaway is as follows. When computing, discussing, or quoting survival and death rates, it’s extremely important to be aware of (1) the year of birth of the relevant cohort as well as (2) the conditioning age. Actuaries have been trained to think (and talk) in this manner from quite early in their education, so it’s important you keep this in mind as well. The next time you hear someone predicting the chances of surviving to age 90, you should inquire about the assumed age (Is it from birth or retirement?) as well as the birth cohort. (Is it 1925 or 1940?) Otherwise, the statement is incomplete and rather meaningless. I’ll return to the conditioning point later on when I introduce and discuss tables for **Period Mortality** versus **Cohort Mortality**.

### 7.7 Extracting Death Rates from Life Tables

Back to the 1925 cohort, I am interested in (more closely) examining the actual number of deaths at any given age to see if I can uncover any patterns that can be exploited from prediction purposes. Remember that my practical objective here is
to forecast what fraction (of the 1925 cohort) might live to age 95, 100, or even 105. If there is a predictable pattern to deaths by age, then I might be able to assume that pattern will continue and use that to make an educated forecast. To start this process I calculate the 1-year death decrements via the `diff` command in R. Please execute the following command and ensure you get the same numbers.

```
> diff(CLT$FEMALE)
[1] -9230.7 -1429.4 -586.0 -418.8 -344.5 -209.6 -168.8
[8] -126.8 -99.3 -103.1 -108.3 -89.5 -110.0 -102.0
[15] -105.9 -104.6 -116.3 -130.2 -131.9 -138.9 -138.7
[22] -142.2 -121.5 -116.6 -94.8 -102.4 -96.8 -86.4
[29] -88.5 -64.0 -85.9 -63.2 -99.0 -97.2 -86.5
[36] -95.2 -112.5 -112.1 -112.7 -116.3 -130.2 -138.9
[43] -138.7 -103.1 -108.3 -89.5 -110.0 -102.0 -96.8
[50] -86.4 -112.7 -118.4 -130.4 -149.8 -152.7 -181.5
[57] -207.0 -220.3 -249.7 -245.5 -273.6 -327.9 -301.6
[64] -334.1 -360.1 -410.0 -420.6 -634.4 -497.3 -497.3
[71] -537.6 -587.6 -634.4 -680.4 -731.6 -723.3 -773.4
[78] -884.6 -911.9 -978.3 -1080.2 -1173.0 -1306.1 -1429.3
[85] -1543.5 -1535.5 -1755.1 -1794.0 -2016.7 -2020.4 -2131.2
```

I’ll get to the reason for the non-integer numbers in a moment. (How can 0.4 people die?) There are a total of 85 numbers, representing the 85 years until the cohorts 85th birthday. The `diff(.)` function in R subtracts any two adjacent numbers, technically \( S_{x+1} - S_x \) using the notation I introduced in the prior section, and displays the results. Between the age \( x = 0 \) and \( x = 1 \), a total of \( S_1 - S_0 = 9230.7 \) females died (during the year 1925) from the 1925 birth cohort. Between age \( x = 1 \) and \( x = 2 \), a total of \( S_2 - S_1 = 1429.4 \) females died (during the year 1926) from the 1925 birth cohort, etc. Let’s see if there is any discernible pattern. The first year contains the most deaths, and the smallest number of deaths were between the ages of \( x = 31 \) and \( x = 32 \), for a total of \( S_{32} - S_{31} = 63.2 \) deaths. The number of deaths appear to increase from age \( x = 32 \) onwards, although they jump around and occasionally do decline. Now, I’ll get back to 0.4 deaths. The reason you see fractions is because the cohort life tables themselves have fractional lives. Why? Because the cohort life tables started with 100,000 (synthetic) people who were then killed based on the 1-year death rates denoted by \( q_x \leq 1 \) and reported in the Human Mortality Database. Stated formally:

\[
S_0 = 100,000 \\
S_1 = S_0 \times (1 - q_0) \\
S_2 = S_1 \times (1 - q_1) \\
S_x = S_{x-1} \times (1 - q_{x-1}) \\
S_{x+1} - S_x = -S_x q_x \tag{7.1}
\]
To be very clear, the \( S_x \) values in the csv file you imported were manufactured (by the author) from the \( q_x \) values reported and stored in the HMD. So, the \texttt{diff(.)} command subtracts \( S_{x+1} \) from \( S_x \) and reports the values of \((-S_x \times q_x)\), per Eq. (7.1). Another way to present and think about the cohort life table is as follows:

\[
S_x = S_0 + \sum_{i=1}^{x} (-S_{i-1} \times q_{i-1})
\] (7.2)

The number of survivors at some age \( x \) is the initial number of individuals in the cohort, \( S_0 = 100,000 \), minus all those who died in between those two ages. So, for example, at the age of \( x = 85 \), the number of survivors \( S_{85} \) is equal to the initial 100,000 minus the sum of deaths during the ages \( x = 0 \) to \( x = 84 \). This can be computed and confirmed in R as follows:

```r
100000+sum(diff(CLT$FEMALE))
[1] 42021.1
CLT$FEMALE[CLT$AGE==85]
[1] 42021.1
```

So, although we know (in the year 2010) that there were \( S_{85} = 42,021.1 \) survivors from the 1925 (synthetic) birth cohort, we will not learn the 1-year death rate \( q_{85} \) for this group, until January 1st, 2011, when the survivors reach age 86. Notice that the table \( \text{CLT} \) contains 86 rows, but the vector created via the \texttt{diff(.)} contains 85 elements. These pesky things (e.g. 85 versus 86 elements) can make the difference between a script that works, versus one that refuses to co-operate.

### 7.8 The One-Year Death Rate \( q_x \)

The symbol \( q_x \) plays a very important role in this book and more generally in actuarial life-sciences. Rearranging equation (7.1), the central \( q_x \) value can be extracted directly from the cohort life table \( S_x \), via the relationship \( q_x = (S_x - S_{x+1})/S_x \). To be precise, in the context of the 1925 cohort, \( 0 \leq q_x \leq 1 \) represents the \textit{realized} 1-year death rate between age \( x \) and age \( (x + 1) \). Although it’s a number between zero and one, it isn’t a probability (yet) and at this point should be interpreted as a decrement or proportional reduction in the number of survivors at a given age \( x \). This \( q_x \) vector is (obviously) gender specific and can be computed in R using the following syntax for the 1925 cohort.

```r
qx_f<(-diff(CLT$FEMALE))/CLT$FEMALE[-length(CLT$FEMALE)]
qx_m<(-diff(CLT$MALE))/CLT$MALE[-length(CLT$MALE)]
summary(round(qx_f,digits=6))
  Min. 1st Qu. Median Mean 3rd Qu. Max.
0.000743 0.001330 0.003006 0.010020 0.010390 0.092307
summary(round(qx_m,digits=6))
```
Notice how I standardized the death decrements: \((S_x - S_{x+1})\) by the number of survivors: \(S_x\) in each year. The denominator of the defining function for \(q_x\) includes the new command \([-\text{length(.)}]\), which drops the last (age 85) element in the relevant CLT vector so that the numerator and denominator are of the same length. (The negative sign removes an index and \text{length} returns the length of the vector, which is the last index.) The 1-year death rate vector \(q_x\) takes on its highest value of 9.23% (females) and 12.26% (males) in the first year of life (a.k.a. the infant mortality rate.) The lowest value can be located via the command:

\[
\text{which}(\text{qx}\_\text{f}==\text{min}(\text{qx}\_\text{f})) \\
[1] 32 \\
\text{which}(\text{qx}\_\text{m}==\text{min}(\text{qx}\_\text{m})) \\
[1] 12 \\
\text{qx}\_\text{f}[32] \\
[1] 0.00074346 \\
\text{qx}\_\text{m}[12] \\
[1] 0.00139421
\]

Notice how the \text{which(.)} command locates the index of the smallest number in the \(qx\) vector, which is a useful command to know when working in \textit{R}. At the index value of 32, which is between age \(x = 31\) and age \(x = 32\), the 1-year death rate was 0.07%. For males it’s at age \(x = 11\), at 0.1%. Note again that in \textit{R} the element \(qx[1]\) is a death rate between age \(x = 0\) and \(x = 1\), etc. Finally, here is a script that plots the \(q_x\) values from the age of \(x = 0\) to the age of \(x = 84\). Remember that I don’t have \(q_{85}\), and only \(S_{85}\) is known.

\[
\text{plot}(c(0,90),c(0,0.14),\text{type}="n", \\
xlab="Age of Cohort (top), Year (bottom)", ylab="Death Rate") \\
\text{grid(ny=18,lty=20)} \\
\text{for (i in 1:81){ \\
\text{points(i-1,\text{qx}\_\text{f}[i],col="red") \\
\text{points(i-1,\text{qx}\_\text{m}[i],col="blue")}} \\
\text{year<-1925+c(0:5)*20; age<-0+(0:5)*20} \\
\text{axis(side=1,\text{line}=1,at=age,labels=year,tick=F)} \\
\text{text(88,0.05,"Females",col="red")} \\
\text{text(70,0.05,"Males",col="blue")}
\]

Figure 7.4 displays the 1-year death rates for the 1925 cohort, but is actually a good picture and general indication of 1-year death rates for almost any cohort and in any country. They start out (very) high in the first few years of life, begin to decline and reach a minimum (at age \(x = 31\) for females and \(x = 11\) for males, in the 1925 cohort) and then begin steadily climbing by age, to reach values of between 8 and 10% around the age of \(x = 80\) (again, for the 1925 cohort.) At the risk of stating the obvious, the older you are the more likely it is you will not survive to your next birthday.
7.9 Reversing the Process: From Death Rates to Life Tables

Starting with a mortality table (or vector) denoted by $q_x$, and keeping Eq. (7.1) in mind, I can reconstruct a life table, and the value of $S_x$ at any age $x$, based on the following algorithm.

$$S_x = S_0 \prod_{i=0}^{x-1} (1 - q_i) \quad (7.3)$$

The intuition here is that I begin with $S_0 = 100,000$ and then kill members of the cohort at the rate of $q_i$ per year, thus leaving $(1 - q_i)$ survivors. In R this can be achieved via:

```r
LifeTable<-100000*cumprod(1-qx_f)
LifeTable<-c(100000,LifeTable)
all.equal(LifeTable,CLT$FEMALE)
```

The first command `cumprod(.)` multiplies the individual elements of the $q_x$ vector, and the second command adds the initial 100,000 people to the beginning of the vector. In some sense, the LifeTable vector is back where I started: CLT, and the final command `all.equal` tests to confirm that indeed they are the same. The benefit of going full circle is that if you have any vector of $q_x$ values, you
should be able to compute survival rates, from any age \( y \) to any age \( z > y \). You would compute those numbers by creating the entire cohort life table \( S_x \), and then dividing \( S_z / S_y \), etc.

## 7.10 Death Rates and Survival Rates: n-Years

More generally, if the 1-year death rate \( q_x \) is defined as one minus the ratio: \( S_{x+1} / S_x \), then the \( n \)-year death rate, that is the fraction of people who are alive at age \( x \), who die before age \( (x + n) \), can be computed via \( S_{x+n} / S_x \), where \( n \) is an integer number of years. I will adhere to the actuarial convention of denoting the \( n \)-year death rate by placing a subscript on the lower-left of the \( q_x \) symbol. This may seem odd and cumbersome, but carved in stone (by actuaries) over a century ago. The \( n \)-year death rate is as follows:

\[
(nq_x) := 1 - \left( \frac{S_{x+n}}{S_x} \right) = 1 - \prod_{i=x}^{x+n-1} (1 - q_i)
\]  

(7.4)

Here is an example of an \( n \)-year death rates vector, using the (female) numbers from the 1925 cohort life table. I’ll compute 20-year and 30-year values, conditioning of the age of \( x = 50 \).

|                  |       |
|------------------|-------|
| 1 - CLT$ FEMALE $ [CLT$ AGE==70] / CLT$ FEMALE $ [CLT$ AGE==50] | 0.1483938 |
| 1 - CLT$ FEMALE $ [CLT$ AGE==80] / CLT$ FEMALE $ [CLT$ AGE==50] | 0.3413173 |

The fraction of 50-year-old females (in the year 1975) who died before their 70th birthday (the year 1995) was a mere 14.8%. That is the value of: \((20q_{50})\), according to the above definition. But, the fraction of 50-year-old females who died before their 80th birthday, before the year 2005, was a heftier \((30q_{50}) = 34.1\%\). This should be intuitive given the longer (30 year) window. Finally, I conclude this section by defining a function in R that computes \( n \)-year survival rates for a given vector of 1-year death rates.

```r
SPQR<-function(x,y,qx){
  LT<-c(1,cumprod(1-qx))
  LT[x+1]/LT[y+1]
}
```

For example, the command \( SPQR(50, 80, qx) \), where the \( qx \) vector is the relevant vector for females \((q_{x,f})\) in the 1925 cohort, should return the value
0.6586, which is a 65.85% survival rate to age 80, for those who made it to age 50. This is exactly: $1 - 0.3415$, which is one minus the 30-year death rate computed in the prior box. Note, the $y$ value in the $SPQR(x, y, qx)$ function is technically $x + n$ in the definition of $nqx$, because a person that is $x$ years old must live another $n$ years to reach age $y$.

### 7.11 A First Look at Natural Laws Governing Death

![Fig. 7.5 Linear relationship between log of mortality rate and age](source: HMD. Analysis and Graphics by Author in R)

At this point I will casually postulate that the $qx$ values grow exponentially between the age of (approximately) 35 and 85. I use the words *casually postulate* because at this point it’s really a conjecture. In the next chapter #8 I’ll provide theoretical support for this assertion, but for now I’ll examine whether it’s a reasonable approximation. Remember that my goal is to predict or forecast the 15 or so missing values of $q_{85}, q_{86} \ldots q_{99}$ so I can predict what fraction of the 1925 cohort will ever get to age 100. From a mathematical point of view (approximate) exponential growth in death rates implies that

$$q_{x+n} = qx e^{gn},$$  \hspace{1cm} (7.5)

where $qx$ is the death rate at some baseline age $x$, the index $n$ is measured in years, and $g$ is the growth rate of death rates (notice the double use of the word *rate*). Then, taking natural logarithms of both sides implies that
\[ \ln[q_{x+n}] = \ln[q_x] + gn \] (7.6)

which is a testable hypothesis. And, although I do plan to do this formally (and rigorously) in the next chapter, for now I’ll simply plot the natural logarithm of the 1-year death rates \( q_x \), and display them in Fig. 7.5. Alas, we have found a pattern to death: a straight line! Again, the \( \ln[q_x] \) values are approximately linear in age, after the age of 35 or so. Moreover, using standard regression (or least square) techniques, I can estimate the slope of the line in Fig. 7.5, using various starting ages for \( x \), for both the 1925 and the 1940 cohort. In fact, the next table displays 95% confidence intervals for the coefficient of the linear regression of the log of the death rate when the regression is conducted from three different starting ages 35, 45, 55, to age 85. The 95% confidence intervals in Table 7.2 are obtained by adding (and subtracting) 1.96 standard errors to the point estimate of the regression coefficient. I realize that I’m being vague and somewhat elusive, for the important takeaway is that the growth rate of the death rate is between 7 and 8%, for the 1925 cohort.

Table 7.2  Estimated growth rates in 1-year death rates (Canada)

| From age | Cohort born in 1925 | Cohort born in 1940 |
|----------|---------------------|---------------------|
|          | Female | Male           | Female | Male           |
| \( x = 35 \) | 0.0773, 0.0802 | 0.0751, 0.0779 | 0.0739, 0.0769 | 0.0697, 0.0738 |
| \( x = 45 \) | 0.0766, 0.0808 | 0.0717, 0.0734 | 0.0721, 0.0768 | 0.0737, 0.0775 |
| \( x = 55 \) | 0.0829, 0.0869 | 0.0714, 0.0739 | 0.0687, 0.0761 | 0.0676, 0.0729 |

Data source: human mortality database. analysis by author

7.12  Projecting Survival Rates to Age 100

Given the 1-year death rates at age \( x = 84 \) for males and females, I can project those values forward for the next 15 years to obtain values of \( (q_{85}, q_{87} \ldots q_{99}) \) and finally forecast survivors to age 100. I will manufacture (forecast) these 15 additional death rate values based on the “theory” that mortality rates (starting at \( q_{85} \)) continue to grow at 7.6% for males and 7.8% for females, based on the table. Then, I’ll simply use those 15 numbers to create new \( S_x \) values, and stitch them together into the life table. Here is the script that does all of that.

```r
qx_m.new<-qx_m[85]*exp(c(1:15)*0.076)
qx_m<-append(qx_m,qx_m.new,after=85)
LT<-100000*cumprod(1-qx_m)
LT<-append(LT,100000,after=0)
LT[101]
[1] 1372.044
```
What do I get from all this? My prediction is that 1372 males from the 1925 cohort will survive to age 100, to the year 2025. This is a 100-year survival rate (forecast) of 1.37% from age zero. Doing the same for females, but assuming the 1-year death rate $q_x$ will grow at 7.8% per year (instead of the male 7.6%), results in 6410 females who survive to age 100. This is a 100-year survival rate (forecast) of 6.4% for females from the 1925 cohort. Remember one last time, that $LT[101]$ is the number of survivors at age $x = 100$, because you can’t get R to accept an $LT[0]$. The lowest index value is [1], which is age zero.

```r
qx_f.new <- qx_f[85] * exp((1:15) * 0.078)
qx_f <- append(qx_f, qx_f.new, after=85)
LT <- 100000 * cumprod(1-qx_f)
LT <- append(LT, 100000, after=0)
LT[101]
[1] 6410.31
```

Figure 7.6 plots these values and shows the projected fraction of survivors of the 1925 cohort up to age 100. Just to be clear, The first 85 years in Fig. 7.6 are based on realized mortality rates from the Human Mortality Database (HMD). I extrapolated the values from age 86 to age 99 by growing the 1-year death rates by $g$ percent per year, and then appended those values to the cohort life table. So, end this chapter with my first statistical prediction. The fraction of survivors.
7.13 Final Notes

- You might have noticed that the word *probability* didn’t appear anywhere in this chapter. Rather, this chapter offers a first stab at *thinking* about death *rates* from a historical population perspective. Forward-looking probabilities—that is thinking about your own human longevity as a random variable $T_x$, with its own probability density function, will come in the next chapter.

- Note that the Human Mortality Database (HMD)—of which a screenshot appears in Fig. 7.7—stores the cohort mortality tables (the so-called death rate in the column labeled 1x1), in $q_x$ format, instead of $S_x$ format. In other words, you have to create the life tables from the death rates, which is the opposite of how I introduced the topic in this chapter. Be careful when you download the data, to use the *cohort* and not the *period* values. The exact URL is provided in reference [2]. For more on longevity see reference [3, 4], or reference [1] for a classic textbook on actuarial methods.

- Towards the end of this chapter, when I was working with $q_x$ values, I should have been (more) careful to (constantly) remind readers these are 1-year death rates for the 1925 cohort, or occasionally the 1940 cohort. If and when I need to forcefully remind readers that the $q_x$ vector is from a specific cohort birth year, I will use the notation $q_x[1925]$, just to ensure everyone is on the same page.

- See the following Table 7.3 and make sure you understand the difference between $q_{60}[1940]$ versus $q_{40}[1960]$. In particular, the elements that have a question mark, can’t be known at the end of the year 2020, because those deaths haven’t occurred yet. Stated differently, in any given year (for example 2020), deaths at all ages will complete the diagonal of the table.

**Fig. 7.7** Human mortality database: Period vs. Cohort
Table 7.3  Cohort vs. Period mortality: which \( q_x \) is known by end-of 2020?

| Birth year | Age 0 | Age 20 | Age 40 | Age 60 | Age 80 |
|------------|-------|--------|--------|--------|--------|
| 1920       | \( q_0[1920] \) | \( q_{20}[1920] \) | \( q_{40}[1920] \) | \( q_{60}[1920] \) | \( q_{80}[1920] \) |
| 1940       | \( q_0[1940] \) | \( q_{20}[1940] \) | \( q_{40}[1940] \) | \( q_{60}[1940] \) | \( q_{80}[1940] \) |
| 1960       | \( q_0[1960] \) | \( q_{20}[1960] \) | \( q_{40}[1960] \) | \( q_{60}[1960] \) | ? |
| 1980       | \( q_0[1980] \) | \( q_{20}[1980] \) | \( q_{40}[1980] \) | ? | ? |
| 2000       | \( q_0[2000] \) | \( q_{20}[2000] \) | ? | ? | ? |

Questions and Problems

7.1 Referring again to Fig. 7.7, download the 1960 Canadian cohort (not period!) death rate vector, \( q_x[1960] \), and compute the fraction of those born in the year 1960, who survived to the age of 50 (in the year 2010.) Using our notation, that is \( S_{50}/S_0 \). Compare the values to the corresponding numbers for the 1925 and 1940 cohort.

7.2 Estimate the slope and intercept of \( \ln[q_x] \) for the 1940 cohort (both males and females), using data from the age of 35–70. You can either do this using the built-in least squares (a.k.a. linear model) function in R, called \( \text{ lm}( . ) \), or look up the formula for the slope and intercept (in any textbook on statistics) and compute these values using brute-force. Please discuss some of the methodological concerns with such a (small) set of numbers.

7.3 Forecast or estimate the fraction of the 1960 cohort (who are 50 years-old in the year 2010), who will survive to the age of 100, based on the methodology described in the chapter. How does it compare to the 1925 cohort?

7.4 Please estimate values of \( S_{90}/S_{60} \), which is the 30-year survival rate conditional on being age \( x = 60 \), for the 1925, 1940, and 1960 cohort.

7.5 Pick one other country (any one) from the Human Mortality Database (HMD), which should all look quite similar to Fig. 7.7, and download 1-year (cohort) death rates, create a 1940 cohort life table, and compare projected survival rates \( S_{90}/S_{60} \), conditional on age 60.

References

1. Dickson, D. C. M., Hardy, M. R., & Waters, H. R. (2010). *Actuarial mathematics for life contingent risks*. Cambridge: Cambridge University Press.
2. Human Mortality Database (CHMD). [http://www.bdlc.umontreal.ca/CHMD/prov/ont/ont.htm](http://www.bdlc.umontreal.ca/CHMD/prov/ont/ont.htm)
3. Maier H., Gampe J., Jeune B., Robine J.M., & Vaupel J.W. (Eds.) (2010). *Supercentenarians*. Heidelberg: Springer.
4. Olshansky, S. J., & Carnes, B. A. (2001). *The quest for immortality: Science at the Frontiers of aging*. New York: W.W. Norton & Company.