Discrete Mortality Tables & Models

FINE4060: Advanced Personal Finance

Moshe A. Milevsky[1]

[1]Professor of Finance
Schulich School of Business &
Graduate Faculty of Mathematics and Statistics
York University, Toronto, Ontario

Lect. #4: [14 February 2019]

Modeling Retirement Mortality: Motivating Question

- You are the manager and trustee of a pension plan.
- Jane (or Joe) retired in 1980 at the age of 60.
- The pension plan is obligated to pay Jane (or Joe) a sum of $\$25,000$ per year for the rest of her life.
- How much money does the pension plan have to set-aside (or have) for this liability?
- Can you know this number with certainty? What does it depend on?

Today's Skills

Learn to...

- Import and work with (cohort) life tables.
- Extract mortality rates (and reconstruct life tables.)
- Analyze patterns within mortality (and life) tables.
- Create some simple mortality forecasts.
- Model Remaining Lifetime (RL). (Contrast with Portfolio Longevity.)
- No new functions, but will use: `diff(.)` `cumprod(.)` `append(.)`

Cohort Life Table (CLT): Hypothetical Group. Real People

- The source for this data is the Human Mortality Database (HMD), available for download at `http://www.mortality.org/`
- They collect (and clean) mortality data for many different countries. These numbers are for Canada.
- 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.)
- The standard cohort life table assumes a group of 100,000 people (male, female) born on January 1st, 1925. It tracks them over the next 85 years (as they die.) Technically speaking, it ignores immigration.
- Naturally, many are alive (today), but data ends in 2010 (age 85.)
- The HMD has (incomplete) cohort data going back to the year 1850!

First things first: Let's get the data into R.

The source is the Human Mortality Database (HMD), available online (after registration) for free. Import the CSV file (dataset) with life tables.

Script or Command Line

```
library(readr)
Canada_Cohort_Lifetable_1925
<- read_csv(".../Canada_Cohort_Lifetable_1925.csv")
```

Get to this...

Confirm you have this data/variable in R.

Go to file/funct

Untitled1 × | Canada_Cohort_Lifetable_1925 ×

Filter

	AGE	FEMALE	MALE	YEAR
1	0	100000.0	100000.0	1925
2	1	90769.3	87732.5	1926
3	2	89339.9	86143.8	1927
4	3	88753.9	85529.5	1928
5	4	88335.1	85090.0	1929
6	5	87990.6	84740.6	1930
7	6	87781.0	84506.0	1931
8	7	87612.2	84309.1	1932
9	8	87485.4	84156.7	1933
10	9	87386.1	84019.9	1934
11	10	87283.0	83895.8	1935
12	11	87174.7	83775.0	1936
13	12	87085.2	83658.2	1937
14	13	86975.2	83535.2	1938
15	14	86873.2	83418.0	1939

Showing 1 to 16 of 86 entries

Let's get some summary statistics

Command Line

```
> CLT<-Canada_Cohort_Lifetable_1925
> summary(CLT$AGE)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00   21.25   42.50   42.50   63.75   85.00
> summary(CLT$FEMALE)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  42021   75343   83754   78760   85971  100000
> summary(CLT$MALE)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  24183   64119   78750   70991   82245  100000
> summary(CLT$YEAR)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1925    1946    1968    1968    1989    2010
```

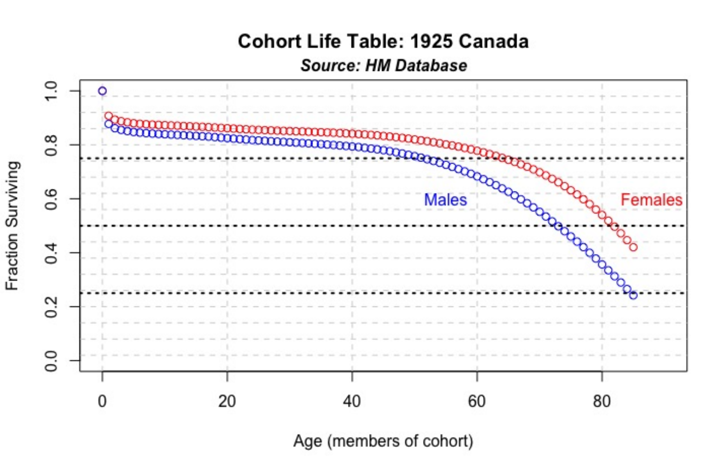
Time for a Good Figure

Script

```
plot(c(0,90),c(0,1),type="n",
xlab="Age (members of cohort)",ylab="Fraction Surviving")
title("Cohort Life Table: 1925 Canada")
mtext("Source: HM Database",side=3,line=0.3,font=4)
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)
text(88,0.6,"Females",col="red")
text(55,0.6,"Males",col="blue")
```

Figure of a Life Table

Every (cohort) Life Table (in the world) looks (something) like this



Questions we can answer:

You were born in the year 1925...

- At what (exact) age was 50% of the group (male vs. female) gone? This is the median lifespan for the cohort.
- What fraction (male vs. female) survive to age 30?
- What fraction (male vs. female) survive to age 60?
- What fraction (male vs. female) survive to age 80?
- Assuming you survive to the age of 60, what fraction (male vs. female) survive to age 80?
- Assuming you survive to the age of 30, what fraction (male vs. female) survive to age 60?

Before you compute, think about the concept of conditional probabilities.

Here are unconditional (from age zero) answers.

Command Line

```
> CLT$FEMALE[CLT$AGE==80]/100000
[1] 0.539938
> CLT$FEMALE[CLT$AGE==60]/100000
[1] 0.777939
> CLT$FEMALE[CLT$AGE==30]/100000
[1] 0.850935
> CLT$MALE[CLT$AGE==80]/100000
[1] 0.356446
> CLT$MALE[CLT$AGE==60]/100000
[1] 0.682913
> CLT$MALE[CLT$AGE==30]/100000
[1] 0.809378
```

It is common to interpret the (backward looking) *survival rate* as a (forward looking) *survival probability*, but one should be very careful.

And here are the (conditional) answers for females...

Command Line

```
 CLT$FEMALE[CLT$AGE==80]/CLT$FEMALE[CLT$AGE==60]
[1] 0.6940621
> CLT$FEMALE[CLT$AGE==60]/CLT$FEMALE[CLT$AGE==30]
[1] 0.9142167
> CLT$FEMALE[CLT$AGE==80]/CLT$FEMALE[CLT$AGE==30]
[1] 0.6345232
> (0.6940621)*(0.9142167)
[1] 0.6345232
```

Note #1: Survival rate to age 80 (for female born in 1925) **conditional** on live status at 60 is 69.4%, which is higher than **unconditional, at birth** survival rate to age 80, which was 54% (in the prior slide.) Make sure you understand this (now) or don't bother reading-on. Note #2: Also, make sure you understand why survival rate from 30 to 60 **times** survival rate from 60 to 80 is equal to survival rate from 30 to 80.

Did they get to the retirement (old) years?

Command Line

```
> CLT$MALE[CLT$AGE==70]/CLT$MALE[CLT$AGE==30]
[1] 0.6814542
> CLT$FEMALE[CLT$AGE==70]/CLT$FEMALE[CLT$AGE==30]
[1] 0.8203705
```

So, the conditional survival rate from age 30 to age 70, is 68% for males and 82% for females. Note. If the survival rate for males is 68% then the death rate before age 70, is 32%. Notice how many males (1/3) planned for an event (or maybe they didn't) that never happened.

Remember. This is the 1925 cohort

- Remember these computations are for the 1925 cohort. It's unclear (at this stage) what this might look like for the 1940 or 1960 cohort.
- For the 1925 cohort, the conditional (on age 30) survival rate to (old) age 85 was 30% for males and 50% for females. (Confirm this yourself.) Females obviously lived longer (in the past.)
- That is a long time to pay retirement pensions, especially for females.
- The number we computed is often denoted by $({}_tp_x)$, which is the conditional probability that an x-year-old will survive to age $(x+t)$.
- Do you think your survival probability will be higher or lower? Will you have better mortality experience? Or worse mortality?
- More on this to come (later.)

Remember. This is a population cohort

- Smokers might have different mortality patterns from non-smokers.
- Each province (state or city) might have it's own pattern of mortality.
- It's possible education, occupation, wealth and income play a role.
- You might want to know (much) more about the group you are modeling, before you try to price or value any retirement pensions.
- More on this to come (later.)

From Life to Death

We are interested in (more closely) examining the number of deaths (of the 1925 cohort) at any given age to see if we can uncover any patterns. Our objective is to forecast what fraction (of the 1925 cohort) will live to age 95, 100 or even 105.

Think of the following series. We know the first 25 elements (b/c we have data to age 85) of the series. What will the remaining 20 look like?

$$({}_0p_{60});\ ({}_1p_{60});\ ({}_2p_{60});\ ({}_3p_{60});\ \ldots\ ({}_{45}p_{60});$$

When does the value hit 5%? When does it hit 1%? When does it hit 0.1%? **Who cares?** Well, remember, you have to pay their pension.

Death Decrements via the diff Command

Built-in function we will use quite a bit. (It's the number dying each year.)

Command Line

```
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  -118.4  -130.4  -149.8
[43]  -152.7  -181.5  -207.0  -220.3  -249.7  -245.5  -273.6
[50]  -327.9  -301.6  -334.1  -360.1  -410.0  -420.6  -363.5
[57]  -468.4  -485.3  -497.3  -537.6  -587.6  -634.4  -680.4
[64]  -731.6  -723.3  -773.4  -884.6  -911.9  -978.3 -1080.2
[71] -1173.0 -1240.8 -1306.1 -1429.3 -1543.5 -1535.5 -1755.1
[78] -1794.0 -2016.7 -2020.4 -2131.2 -2209.1 -2461.7 -2524.8
[85] -2645.9
```

Mortality Rate (Table) from a Survival Rate (Table)

Define a new (and very important) variable q_x, which is the one-year death rate (a.k.a. probability) at age x.

Command Line

```
> qx_f<-(-diff(CLT$FEMALE))/CLT$FEMALE[-length(CLT$FEMALE)]
> qx_m<-(-diff(CLT$MALE))/CLT$MALE[-length(CLT$MALE)]
> summary(qx_f)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.0007435 0.0013303 0.0030059 0.0100196 0.0103900 0.0923070
> summary(qx_m)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.001394 0.001964 0.005139 0.016274 0.019497 0.122675
```

Notice how I standardized the death decrements by the number of survivors (at the start of) each year. The denominator includes the command `[-length(.)]`, which effectively drops the last element in the vector so numerator and denominator are the same length.

Plot the q_x curve

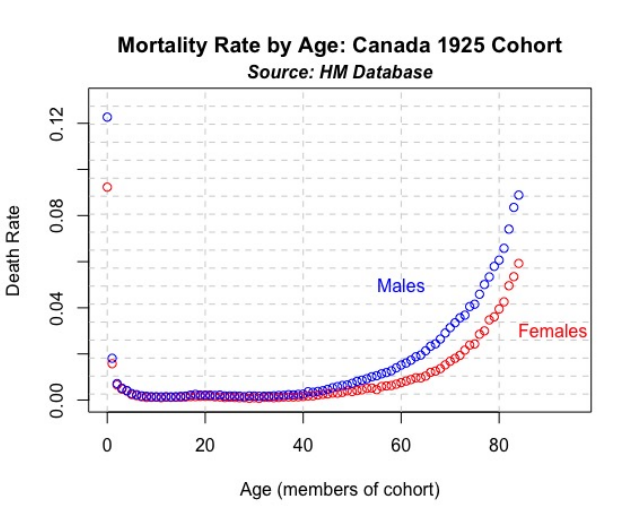
We generate a standard figure....

Script

```
plot(c(0,95),c(0,0.13),type="n",
     xlab="Age (members of cohort)",ylab="Death Rate")
title("Mortality Rate by Age: Canada 1925 Cohort")
mtext("Source: HM Database",side=3,line=0.3,font=4)
grid(ny=18,lty=20)
for (i in 1:86){
  points(i-1,qx_f[i],col="red")
  points(i-1,qx_m[i],col="blue")
}
text(91,0.03,"Females",col="red")
text(60,0.05,"Males",col="blue")
```

A Figure of Death

Every Mortality Table (in the world) looks (something) like this



Reconstructing a Life Table from Mortality Table

Starting with a mortality table (or vector) denoted by q_x, we can reconstruct a life table (or vector), based on the following algorithm.

$$({}_j p_x) = \prod_{i=0}^{j-1} (1 - q_{x+i})$$

Command Line

```
> LT<-100000*cumprod(1-qx)
> LT<-append(LT,100000,after=0)
```

The first command `cumprod(.)` multiplies the individual elements and the second command `append(.)` adds the initial 100,000 people to the vector. The `LT` vector is back where we started. **NOTE:** Sometimes you might be (i.) given a cohort life table and sometimes you might be (ii.) given the mortality rates themselves and have to reconstruct the life-table.

Patterns in the q_x values?

I will casually postulate that the q_x values are growing smoothly and exponentially between the age of (approximately) 35 and 85. So, let's test the following model.

$$q_{x+t} \;=\; q_x\, e^{gt}, \tag{1}$$

where t is measured in years, g is a growth rate and q_x is the initial age. This then implies that:

$$\ln[q_{x+t}] \;=\; \ln[q_x] + gt, \tag{2}$$

which we can easily test (graphically and then) statistically.

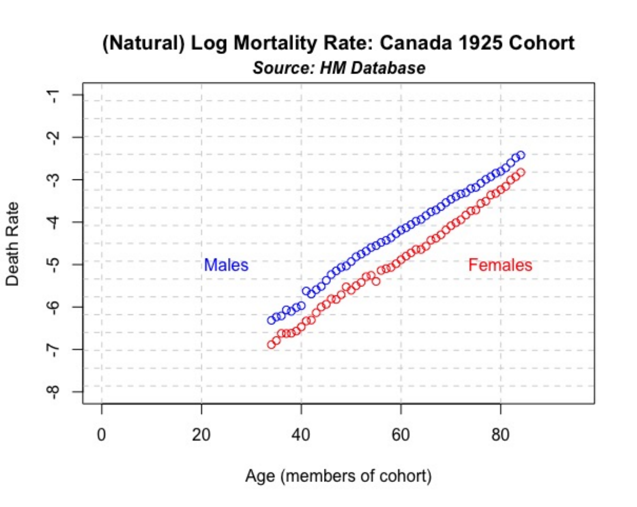
Plot the $\ln[q_x]$ curve

Script

```
plot(c(0,95),c(-8,-1),type="n",
     xlab="Age (members of cohort)",ylab="Death Rate")
title("(Natural) Log Mortality Rate: Canada 1925 Cohort")
mtext("Source: HM Database",side=3,line=0.3,font=4)
grid(ny=18,lty=20)
for (i in 35:86){
  points(i-1,log(qx_f[i]),col="red")
  points(i-1,log(qx_m[i]),col="blue")
}
text(80,-5,"Females",col="red")
text(25,-5,"Males",col="blue")
```

Compute the Natural Logarithm of the Mortality Rate

Yes. This is linear in age.



Let's do this formally: Females

Regress (log) mortality rate on age, from age $x = 35$ to age $x = 85$.

Command Line

```
> y<-log(qx_f[35:85])
> x<-35:85
> fit<-lm(y~x)
> summary(fit)
Residuals:
     Min        1Q    Median        3Q       Max
-0.16759 -0.05374 -0.00243  0.04297  0.17292
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.61583    0.04324  -222.4   <2e-16 ***
x            0.07835    0.00070   111.9   <2e-16 ***
Multiple R-squared:  0.9961,Adjusted R-squared:  0.996
```

The mortality rate q_x increases by approximately $g = 7.83\%$ per year.

Let's do this formally: Males

Command Line

```
> y<-log(qx_m[35:85])
> x<-35:85
> fit<-lm(y~x)
> summary(fit)
Residuals:
      Min        1Q    Median        3Q       Max
-0.187669 -0.055140  0.004994  0.056037  0.123211
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.9218500  0.0447970  -199.2   <2e-16 ***
x            0.0766056  0.0007251   105.6   <2e-16 ***
Residual standard error: 0.07622 on 49 degrees of freedom
Multiple R-squared:  0.9956,Adjusted R-squared:  0.9955
```

The mortality rate q_x increases by approximately $g = 7.66\%$ per year.

Intuitive Interpretation.

The mortality rate for the 1925 cohort at age x increases by approximately 7.66% for males and 7.83% for females. Project those values forward for the next 10 years ($q_{85}, q_{91} \ldots q_{94}$) and forecast the number of survivors to age 95. **Note:** Careful with indexing. qx[85] is mortality at age 84!

Command Line

```
> new_qm<-qx_m[85]*exp(c(1:10)*0.0766)
> new_qm
 [1] 0.09594691 0.10358525 0.11183169 ....
> new_qf<-qx_f[85]*exp(c(1:10)*0.0783)
> new_qf
 [1] 0.06406073 0.06927829 0.07492080 ....
> qx_m<-append(qx_m,new_qm)
> qx_f<-append(qx_f,new_qf)
```

I manufactured 10 additional mortality rate values for the ages from $x = 85$ to $x = 94$, and append them to $\mathtt{qx}_{\mathtt{m}}$ and $\mathtt{qx}_{\mathtt{f}}$.

Finally, re-build a life table and count survivors

Command Line

```
> LT_m<-100000*cumprod(1-qx_m)
> LT_m<-append(LT_m,100000,after=0)
> LT_f<-100000*cumprod(1-qx_f)
> LT_f<-append(LT_f,100000,after=0)
> LT_m[96]
[1] 5396.912
> LT_f[96]
[1] 15713.15
```

And we are done! Forecast of 5397 males and 15713 females alive at age 95, from the original 1925 cohort of 100,000. That is equivalent to a 15% survival rate for females and 5.3% or males, (conditional on age zero!)

Some Probability Theory: Remaining Lifetime (RL) Random Variable.

- Think of your remaining lifetime $\mathbf{T}_x$ as a random variable (at age x).
- $\mathbf{T}_x$ can take on any positive number in the range $(0, D]$, where D denotes the maximum remainder of life, and $\omega := x + D$ denotes the maximum length of life.
- The variable $\mathbf{T}_x$ can be modeled *discretely*, but *continuously* makes more sense.
- Here are examples of a discrete formulation:

$$\begin{aligned}
\mathbf{T}_x &\in \{1, 2, 3, 4, ...50\} \qquad \text{(In Years)} \\
\mathbf{T}_x &\in \{1/12, 2/12, 3/12, ...600/12\} \qquad \text{(In Months)} \\
\mathbf{T}_x &\in \{t_1, t_2, t_3, ...t_{N-1}, t_N\} \qquad \text{(General)}
\end{aligned}$$

- When time is discrete, $x + \mathbf{T}_x > x$.

Probability Mass Function (PMF)

- The probability mass function (PMF), which is the probability that $\mathbf{T}_x$ takes-on a value of exactly t_i, is denoted by:

$$f_x(t_i) \;=\; \Pr[\mathbf{T}_x = t_i],\; i = 1..N \tag{3}$$

- Obviously, $f_x(t_i) < 1$ and you also must die at some point so:

$$\sum_{i=1}^{N} f_x(t_i) = 1 \tag{4}$$

- Going back to our cohort life tables, the survival probability for a given number of years i, is actually:

$$({}_i p_x) \;=\; 1 - \Pr[\mathbf{T}_x < i], \tag{5}$$

which links the earlier work to the concept of $\mathbf{T}_x$

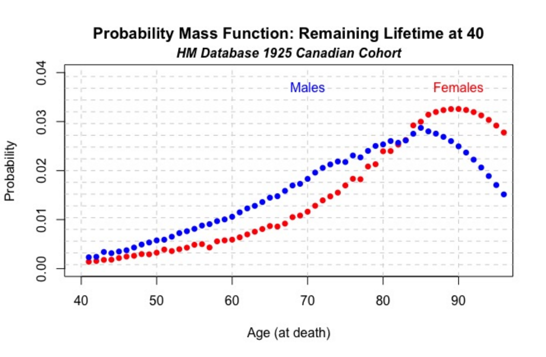
Constructing the Probability Mass Function (PMF)

Script

```
plot(c(40,95),c(0,0.04),type="n",
     xlab="Age (at death)",ylab="Probability")
title("Probability Mass Function: Remaining Lifetime at 40")
mtext("HMD 1925 Canadian Cohort",side=3,line=0.3,font=4)
grid(ny=18,lty=20)
for (i in 41:96){
  points(i,(LT_f[i-1]-LT_f[i])/LT_f[40],col="red",pch=16)
  points(i,(LT_m[i-1]-LT_m[i])/LT_f[40],col="blue",pch=16)
}
text(90,0.037,"Females",col="red")
text(70,0.037,"Males",col="blue")
```

Notice how we take the life table value at adjacent ages and scale by the initial (age 40) values

Use the CLT to Compute and Display a PMF



This enables you to compute the probability of dying in between any two given ages, conditional on being alive at age 40. Ignore the rough edges...

To Discuss: End with some (advanced) finance

- Assume the interest rate r is zero, that $\mathbf{T}_x$ is modeled in years, and you are a pension fund with only three clients, currently aged x, y, z You have to pay each person \$1 per year for the rest of their life.
- Question: What is the mathematical representation of the present value (PV) of your liability?
- Answer: It is a random variable (also) that can be written as:

$$\mathbf{L} := \mathbf{T}_x + \mathbf{T}_y + \mathbf{T}_z$$

- Question: Assume you manage a life insurance company with only three clients, and you have to pay \$1 upon to each of them (or their family) upon death. What is the PV of your liability?

Homework Questions: Question Sheet in Class

You must import the 1940 (Canadian) cohort, compare with the 1925 cohort, and compute a variety of metrics using the two mortality tables.