Lecture

INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

Class: SY BSc

Subject : Statistical and Risk Modelling 1

Subject Code: PUSASQF

Chapter: Unit 2 Chapter 2

Chapter Name: Age-dependent transition intensities

Today's Agenda

1. Survival Models
   1. Simple Binomial model
   2. Multiple State models
   3. Poisson Model

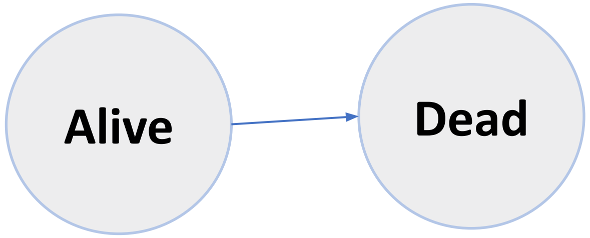
2. Exposure to Risk
   1. Central and Initial Exposed to Risk
   2. Homogeneity
   3. Principle of Correspondence

3. Calculation of Exposed to Risk
   1. Exact Calculation of Central Exposed to Risk
   2. Census Approximation
   3. Trapezium Approximation

4. Different definitions of age
   1. Correspondence between Census and Death Data

1 Survival Models

1.1 Simple Binomial Model



Consider the basic survival model with two states, Alive and Dead.

Suppose that there are 'n' independent and identical lives aged x in the observation. We observe the lives for say, one year. We denote the number of deaths occurred during the year by 'd'. We have q_x as the probability of death of a person aged x in one year.

If we assume that Number of Deaths follows Binomial distribution then the model is as:

Binomial Model: $\mathbf{D \sim \mathit{Bin}(n, q_x)}$
where estimate of $q_x = \frac{d}{n}$

Probability of exactly d deaths:
P(D=d) = ${}^{n}C_{d} \text{ x } (q_x)^d \text{ x } (1 - q_x)^{n-d}$

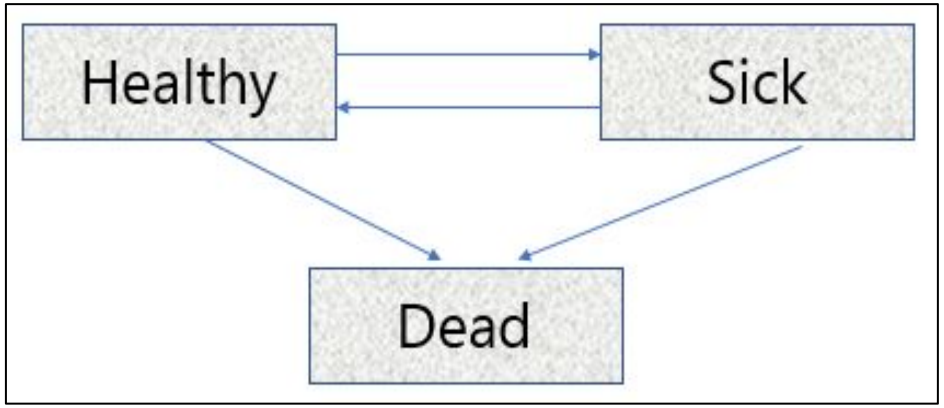
1.2 Multiple States Model

In a two state model a person who is alive; either remains alive or transitions to the dead state at any time t in the age label of x.

We can have multiple state models also, with three or more states.
Eg: Consider a simple multiple-state model with three states: healthy, sick and dead. In that particular model, a life in the healthy state can move to the sick state or the dead state. Similarly, a life in the sick state can move to the healthy state or the dead state.

The transition probability from state i to state j is given by:

$$p_{ij} = \frac{number\ of\ transitions\ from\ state\ i\ to\ state\ j}{number\ of\ transitions\ out\ of\ state\ i}$$



1.3 The Poisson Model

Consider that we observe N independent and identical lives. The lives are observed for a total of E_x^c person-years. This is the total exposed to risk i.e. the total waiting time.
Denote the Number of deaths by D. If we assume that Number of deaths follows Poisson distribution, then we get the model as:

The Poisson Model: D ~ Poi (μE_x^c)
Probability of exactly d deaths:

$$P(D = d) = \frac{e^{-\mu E_x^c} \cdot (\mu E_x^c)^d}{d!}$$

The Poisson model is not an exact model, since it allows a non-zero probability of more than *N* deaths, but it is often a very good approximation. The probability of more than *N* deaths is usually negligible.

2 Exposure to Risk

In this section we will basically deal with how to calculate mortality rates using the observed data or investigation.

We intend to observe how many deaths have occurred, the period of investigation, the death times and waiting times and using all this information we want to compute the mortality rates.

In this chapter ahead we consider the problem of heterogeneity, the central and initial exposed to risk, exact calculations of exposed to risk and the census approximation.

2.1 Central and Initial Exposed to Risk

Exposed to Risk

- Central Exposed to Risk
- Initial Exposed to Risk

Central exposed to risk (waiting times) are natural and intrinsically observable-just record the time spent under observation by each life. It is denoted as E_x^c . Since we observe waiting times the Poisson model is suitable. Central exposed to risk is generally preferable.

Initial exposed to risk is more complicated unless we can use the idealized binomial model in which N lives are observed for a whole year without censoring. Initial exposed to risk is denoted as E_x.

The initial exposed to risk can be reasonably approximated by: $\boldsymbol{E_x \approx E_x^c + \frac{1}{2} d_x}$

2.2 Homogeneity

In the binomial model or multiple state and Poisson models, **one important assumption made is that we can observe a group of identical lives**. It means the lives under observation are identical in various aspects, they have homogeneous characteristics.

However this is not completely possible in reality. Lives have different attributes which makes the group heterogeneous.

If we conduct analysis using such heterogeneous group of lives, our estimate of mortality rate will be more so like an average rate for the lives with different characteristics. If such average rates are used in various calculations, it can lead to inaccuracy and create problems.

Mortality also varies with age, which is an example of heterogeneity itself.

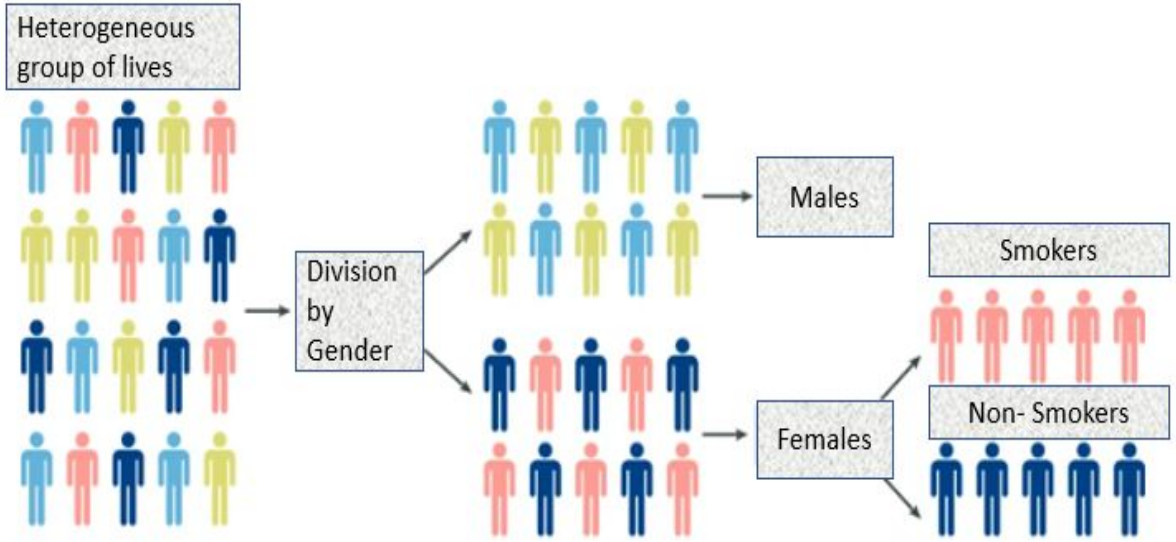
So what is the solution?

2.2 Homogeneity

The answer to the problem of heterogeneity is that we sub-divide the data into homogeneous groups based on the characteristics known, from experience, to have a significant effect on mortality. This ought to reduce the heterogeneity of each class so formed, although much will probably remain.

Various factors based on which sub-division is made are:

- **Age**
- **Gender**
- **Type of policy (in insurance)**
- **Smoking status**
- **Marital status, etc**



2.3 Principle of Correspondence

To calculate q_x and μ_x estimators, data for deaths and exposed-to-risk is used. These 2 data sets need to be consistent, which is obvious, without which the ratios are meaningless.

However in insurance companies they often come from 2 different sources:

- Deaths from claims data
- And exposed-to-risk from premiums collected data

We need to take care to ensure that these two use the same definition of age x.

The Principle of Correspondence
A life alive at time t should be included in the exposure at age x at time t if and only if, were that life to die immediately, they would be counted in the deaths data d_x at age x.

3 Calculation of Exposed to Risk

There are two methods used to calculate the central exposed to risk.

1] Exact Calculation.

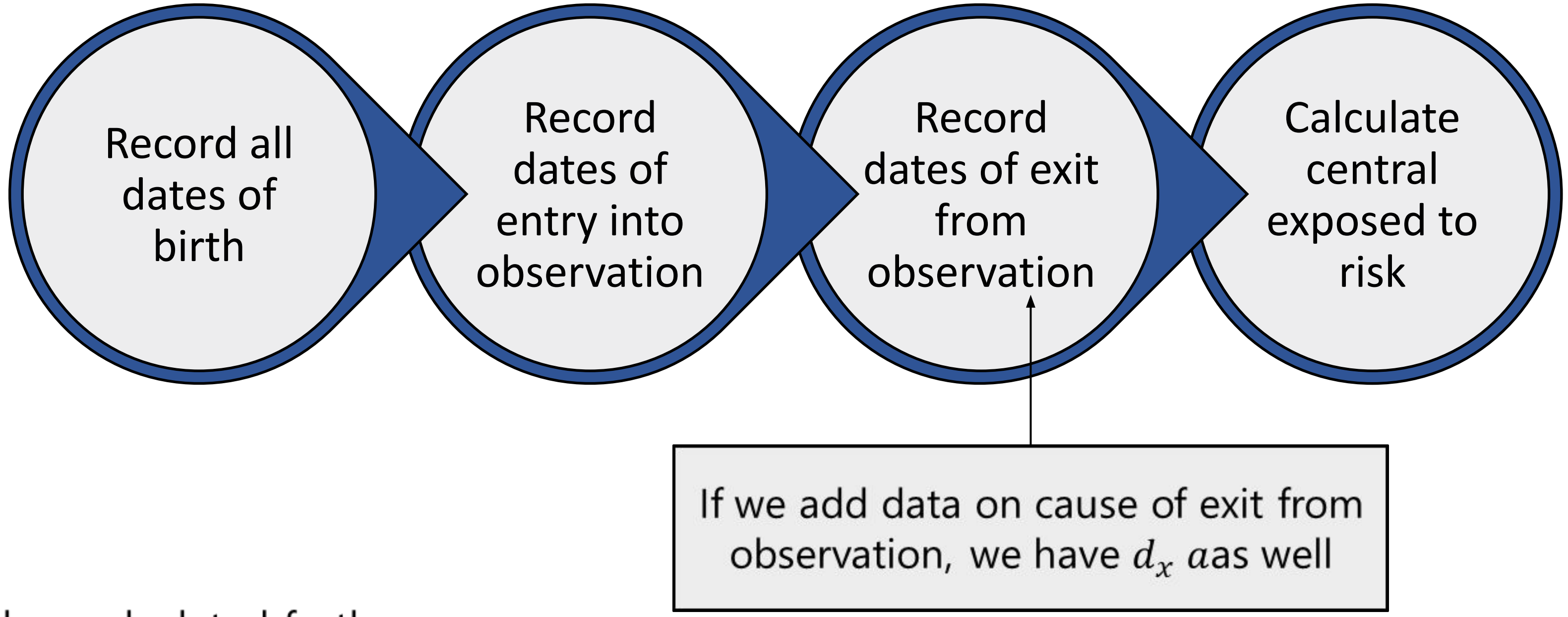
When we have complete precise data we use this method. However it is often difficult to get a precise data in all forms, hence this method is less often.

2] Census Approximation.

When we have incomplete data or the age definition does not correspond to [x, x+1]and the exact calculation is not possible, in such cases we use the census approximation.

3.1 Exact Calculation of Central Exposed to Risk

The procedure for exact calculation follows a simple flowchart as below:



The E_x^c is then calculated further.

3.1 Exact Calculation of Central Exposed to Risk

The central exposed to risk E_x^c for a life with age label x is the time from Date A to Date B where:

Date A:
The latest of

- the date of reaching age label x
- the start of the investigation
- the date of entry (joining)

Date B:
The earliest of

- the date of reaching age label x+1.
- the end of the investigation
- the date of exit (whatever reason

Note that: Calculation allows for all movements into and out of the study for all causes. All decrements (exits) contribute a fraction of the year of exit and all increments (new entrants) contribute a fraction of a year in the year of entry.
E_x^c does not depend on the decrement under study (death, lapse, surrender, sickness). Exposure is measured in year.

3.2 Census Approximation to E_x^c

Death data is often in the form:

- d_x = total number deaths age x last birthday in the calendar years K, K+1, … K+N. So N+1 calendar years of data for deaths between ages x and x+1.

Also consider that we don't have exact entry and exit dates, and rather we have a census data as:

- $P_{x,t}$ = number of lives under observation, aged x last birthday at time t where t in this case is 1st January in calendar years K, K+1, … K+N. So N+1 calendar years of total number policies in-force on 1st January.

Now define $P_{x,t}$ to be the number of lives under observation, aged x last birthday, at any time, then:

The Census Approximation to E_x^c is: $\boldsymbol{E_x^c = \int_K^{K+N+1} P_{x.t}\, dt}$

3.3 Trapezium Approximation to E_x^c

In the census approximation that we derived, one problem is that we do not know the value of $P_{x,t}$ for all *t*, so we cannot work out the exact value of the integral.

Hence we make an approximation.
The trapezium approximation assumes that $P_{x,t}$ is linear between census dates.

By the Trapezium Approximation, we have :

$$E_x^c \approx \sum_{t=K}^{K+N} \frac{1}{2}\left(P_{x,t} + P_{x,t+1}\right)$$

Question

Man Life is an insurance company which only sells life insurance to males. It has recently bought another smaller company called Mixed Life which sells business to both males and females. The company is reviewing the premium rates it charges for life insurance.

Man Life has records of the number of policies in force at their year end, which is 30 September, recorded by age last birthday. Mixed Life has records of the number of policies in force on 31 December each year recorded by age last birthday for males and age nearest birthday for females.

The data for the most recent years are presented on the next slide.

Question

Man Life

Age last birthday	*Number of policies 30 Sept. 2015*	*Number of policies 30 Sept. 2016*	*Number of policies 30 Sept. 2017*
49	4,789	4,296	4,367
50	4,953	5,009	4,809
51	5,300	5,186	5,902

Mixed Life Males

Age last birthday	*Number of policies 31 Dec. 2015*	*Number of policies 31 Dec. 2016*	*Number of policies 31 Dec. 2017*
49	1,832	1,650	1,698
50	1,800	1,750	1,550
51	1,966	1,756	1,569

Mixed Life Females

Age nearest birthday	*Number of policies 31 Dec. 2015*	*Number of policies 31 Dec. 2016*	*Number of policies 31 Dec. 2017*
49	1,602	1,568	1,639
50	1,506	1,497	1,508
51	1,610	1,587	1,411

Question

i. Calculate the central exposed to risk of the combined portfolio for males aged 50 last birthday for the calendar year 2016, stating each assumption you make at the point where you make it.

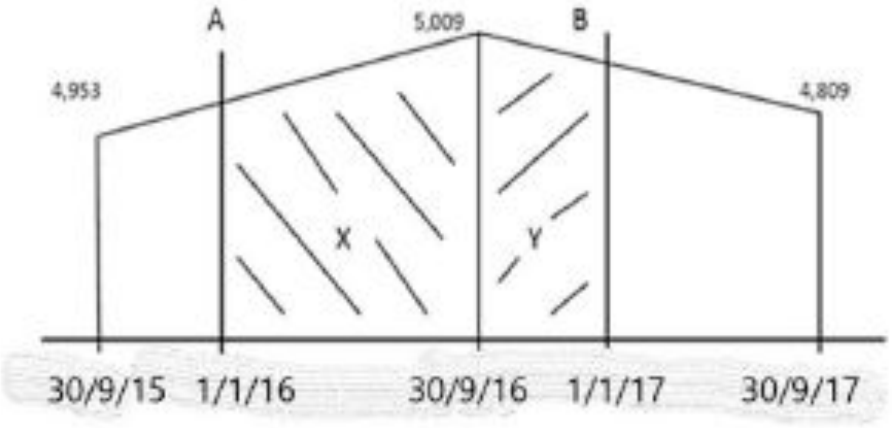
ii. Calculate the central exposed to risk of the combined male and female portfolio for persons aged 50 last birthday for the calendar year 2016, stating each assumption you make at the point where you make it.

Legislation has been brought in which means that males and females must be charged the same premium rates for life insurance. The company is considering basing its future premium rates on the number of deaths across the whole male and female portfolio of the two companies at each age divided by the exposed to risk across the combined male and female portfolio at each age.

iii. Discuss the appropriateness of the company's approach to determining its future premium rates.

Solution

i. Man Life

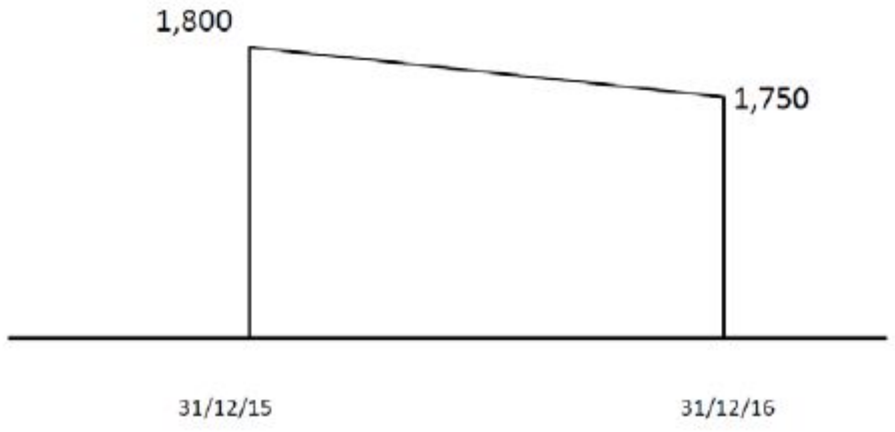


See the diagram above. The required exposed to risk is represented by Area X + Area Y
Assuming that the population varies linearly over inter-census periods, and that the data for 31 December in a year can be taken to represent the data for 1 January the following year

Number of policies in force on 1 January 2016 (A) = (¾ * 4,953) + (¼ * 5,009) = 4,967
Number of policies in force on 1 January 2017 (B) = (¾ * 5,009) + (¼ *4,809) = 4,959
Area X = 9/24 * (4,967 + 5,009) = 3,741
Area Y = 3/24 * (5,009 + 4,959) = 1,246
Exposed to risk = 3,741 + 1,246 = 4,987

Solution

i. Mixed Life



Assuming that the population varies linearly over inter-census periods, and that the data for 31 December in a year can be taken to represent the data for 1 January the following year

Exposed to risk = ½ (1,800 + 1,750) = 1,775

Total male exposed to risk = 4,987 + 1,775 = 6,762

Solution

ii.
We need to adjust the age definition for the female lives.

Assuming birthdays are spread evenly over calendar years, and that the data for 31 December in a year can be taken to represent the data for 1 January the following year, the number of policies in force aged 50 last birthday is equal to 0.5 * number of policies in force aged 50 nearest birthday + 0.5 * number of policies in force aged 51 nearest birthday

on 31 December 2015 this is ½ (1,506 + 1610) = 1,558
on 31 December 2016 this is ½ (1,497 + 1,587) = 1,542

so the exposed to risk for the female lives at age 50 last birthday is ½ (1,558 + 1,542) = 1,550\

Total exposed to risk of the combined portfolio is therefore 6,762 + 1,550 = 8,312

Solution

iii.
This approach will only work if the mix of males and females remains the same. [+½]

It is not clear whether this will happen in the future.

Need to know what competitors are doing.

Other companies may base their rates on a different mix of in force business, or some estimate of future mix.

Consider mortality improvements going forward, and in particular the future development of the ratio between male and female death rates.

What demographic does the company want to target, e.g. only males?

Some selection effects are nullified by the fact that all companies are required to charge unisex rates.

The overall mix of business by gender may alter temporarily as those who are likely to lose out by the introduction of the new legislation may make a dash to get cover before the legislation comes into force.

4 Different Definitions of Age

Earlier we used “age last birthday” in d_x which gives year of age [x, x+1].
Other age definitions are possible:

- d_x : Number of deaths at age x nearest birthday
- d_x : Number of deaths at age x next

Rate Interval
A *rate interval* is a period of one year during which a life’s recorded age remains the same, *eg* the period during which an individual is ‘aged 36 last birthday’.

Resulting estimates are as follows:

Definition of x	Rate interval	q estimates	
Age last birthday	[x, x + 1]		
Age nearest birthday			
Age next birthday	[x-1, x]		

4.1 Correspondence between Census and Death data

With different age definitions we need to check that the principle of correspondence is satisfied.

Census data {P} is consistent with death data {d} if and only if any of the lives counted in P were to die on the census date itself then they would be included in {d}.

Death Data has Priority

If we find death and census date with different age definitions we must adjust the census data, and not the death data because as mortality rates are usually small each piece of death data carries more information and should be preserved intact.

Recap

- We've learnt the three basic survival models , namely:
    1. Simple Binomial model, where a life transitions just between two states (eg: alive & dead)
    2. Multiple states model, where a life can transition between three or more states (eg: alive, sick & dead)
    3. The Poisson model, where ***number of deaths*** follow the poisson distribution with parameter (μE_x^c)

- Next we deal with ways of how to calculate mortality rates using the observed data or investigation:
    1. Central Exposed to Death, where we just record the time spent under observation by each life.
    2. Initial exposed to risk, is more complicated , unless we can use the idealized binomial model.

- Heterogeneity is a problem we face while dealing with this data, ie lives have different attributes which makes the group heterogeneous which goes against our homogeneity assumption while using the binomial and poisson models. This can be tackled by dividing the group into further sub groups based on various factors like age, gender, smoking status, etc

- By the Principle of Correspondence, A life alive at time t should be included in the exposure at age x at time t if and only if, were that life to die immediately, they would be counted in the deaths data d_x at age x.

- There are two ways of calculating central exposed to risk, the exact calculation, used when we have complete data, and the census approximation, used when we have incomplete data.
- The Census Approximation to E_x^c is: $\boldsymbol{E_x^c = \int_K^{K+N+1} P_{x.t}\, dt}$
- If $P_{x,t}$ is not known for all ages, we use the Trapezium Approximation: $\boldsymbol{E_x^c} \approx \sum_{t=K}^{K+N} \frac{1}{2}(\boldsymbol{P_{x,t} + P_{x,t+1}})$

Homework Question

CT4 April 2011 Q8

(i) Explain the difference between the central and the initial exposed to risk, in the context of mortality investigations. [2]

An investigation studied the mortality of infants aged under 1 year.

The following table gives details of 10 lives involved in the investigation. Infants with no date of death given were still alive on their first birthday.

(ii) Calculate the maximum likelihood estimate of the force of mortality, using a
two-state model and assuming that the force is constant. [3]

Life	*Date of birth*	*Date of death*
1	1 August 2008	-
2	1 September 2008	-
3	1 December 2008	1 February 2009
4	1 January 2009	-
5	1 February 2009	-
6	1 March 2009	1 December 2009
7	1 June 2009	-
8	1 July 2009	-
9	1 September 2009	-
10	1 November 2009	1 December 2009

Homework Question

(iii) Hence estimate the infant mortality rate, $q0$. [1]

(iv) Estimate the infant mortality rate, $q0$,using the initial exposed to risk. [1]

(v) Explain the difference between the two estimates. [2]

[Total 9]