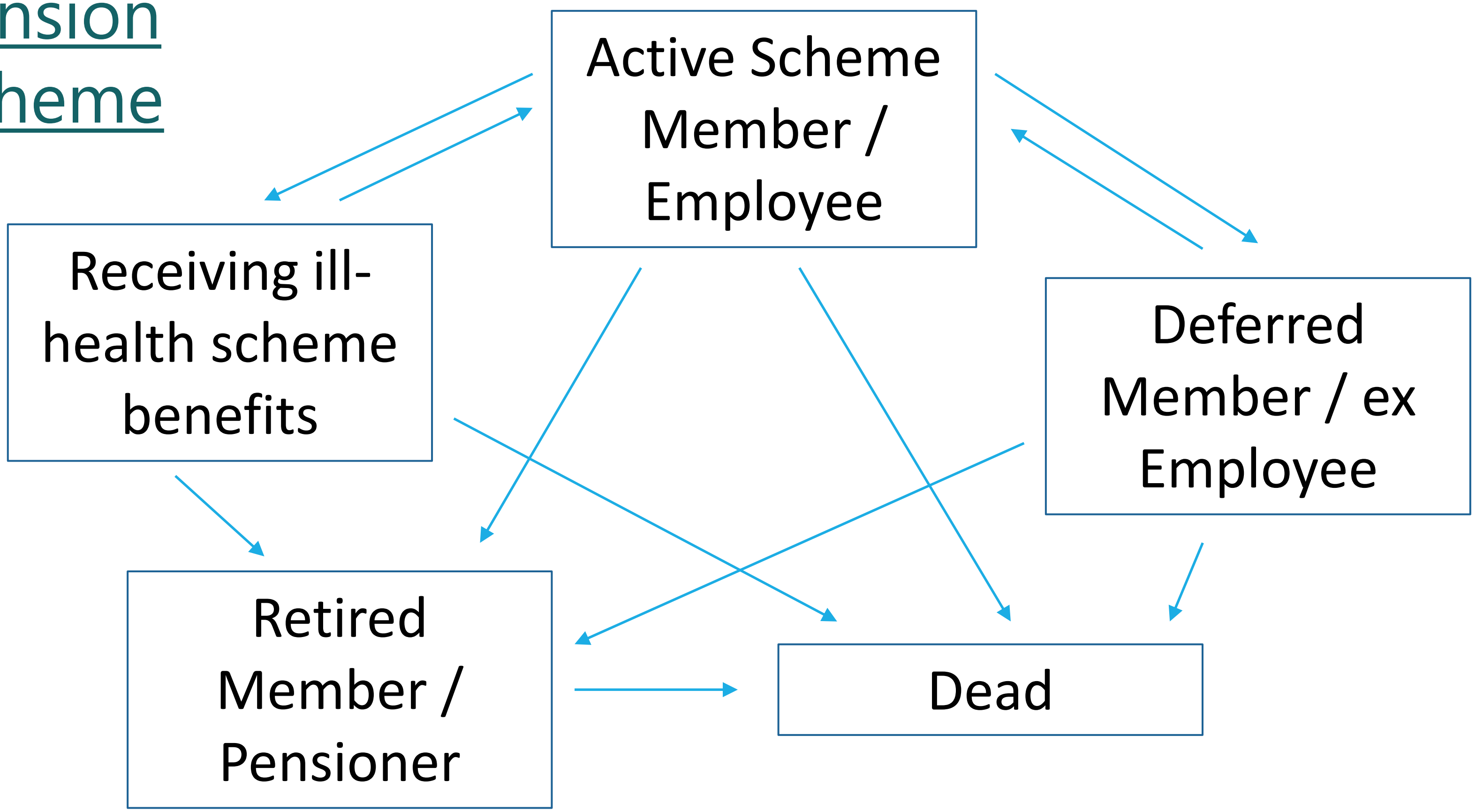
5. Markov Processes and Multi-State Models

CHRIS SUTTON, OCTOBER 2023

multi-state [Markov] models

where we are heading

Pension Scheme



link to Random Processes module

This week we will introduce a different way of thinking about survival models

- modelling transitions between different states
- we will begin with the simplest form (2 states) to establish the methodology
- then introduce the multi-state models that are more applicable in practice

This methodology uses what statisticians call Markov processes

- the theory behind this is part of the *Random Processes* module MTH6141

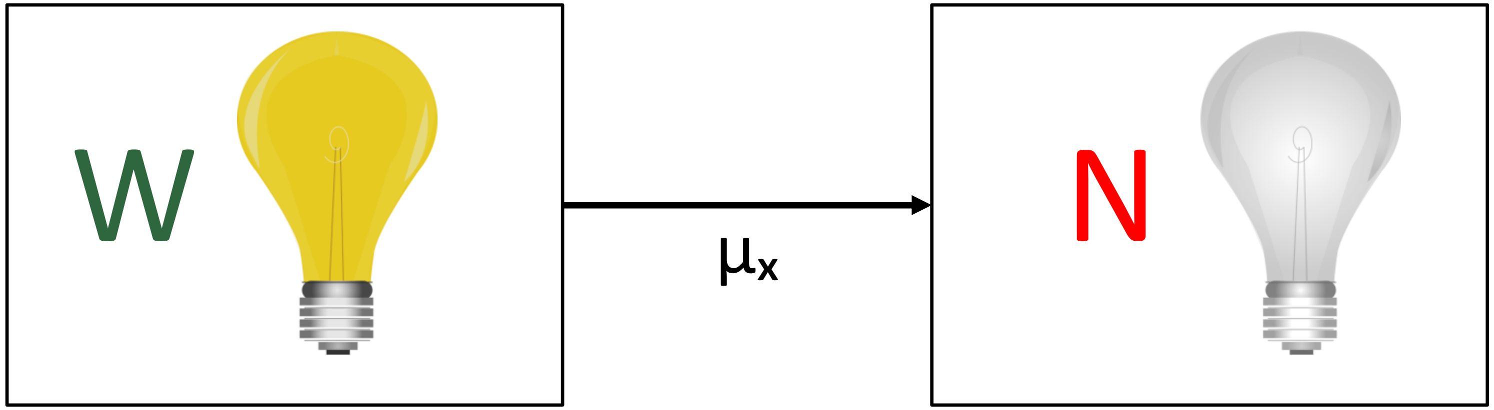
introducing the 2-state model

2 states: projector lightbulb

consider a lightbulb age x since first fitted to projector

2 **states** **W**orking and **N**ot Working

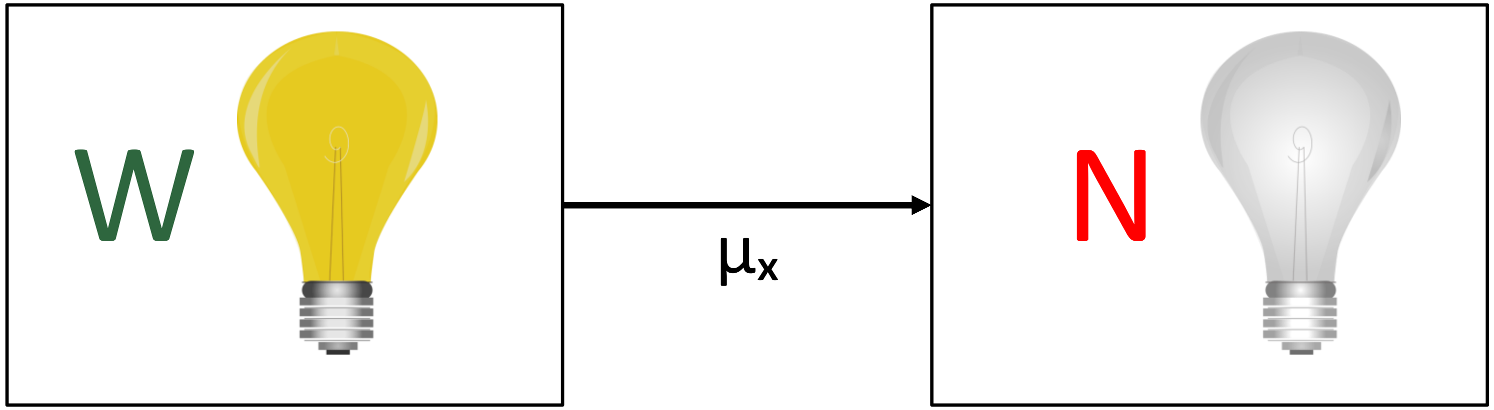
can only move in one direction **W** → **N** this is a **transition**



2 states: projector lightbulb

consider a lightbulb age x since first fitted to projector

the probability at age x that a bulb then **W**orking will be **N**ot Working at age x+t is governed by the age-dependant **<u>transition intensity</u>** μ_{x+t} $(t \geq 0)$



two key assumptions in this model

the probability of being in either state at some future date depends only on (i) age and (ii) the state currently occupied [the "Markov assumption"]

the probability of transition during time $t \geq 0$ is

$${}_{dt}q_{x+t} = \mu_{x+t}\,dt + o(dt)$$

transition probability in Markov process

$_{dt}q_{x+t} = \mu_{x+t}dt + o(dt)$

the remainder term $o(dt)$ is some function of the small time interval dt such that the p.d.f. $f_x(t) = o(dt)$ as $t \to 0$

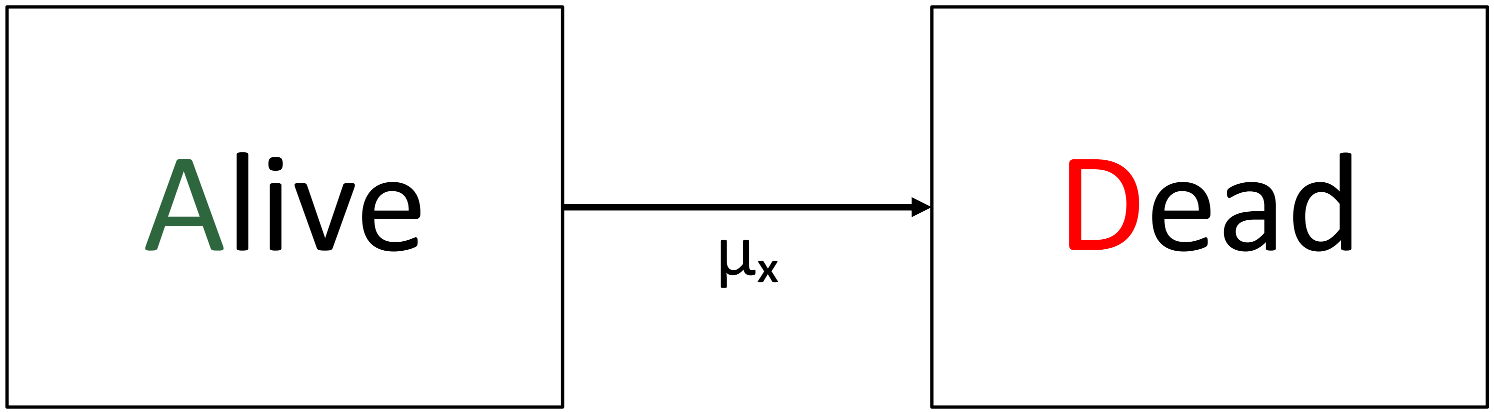
for small time interval dt the transition probability is approximately proportional to the length of time with the constant being the transition intensity μ_{x+t}

and $\lim_{t \to 0} \frac{o(dt)}{dt} = 0$

2 states: human life

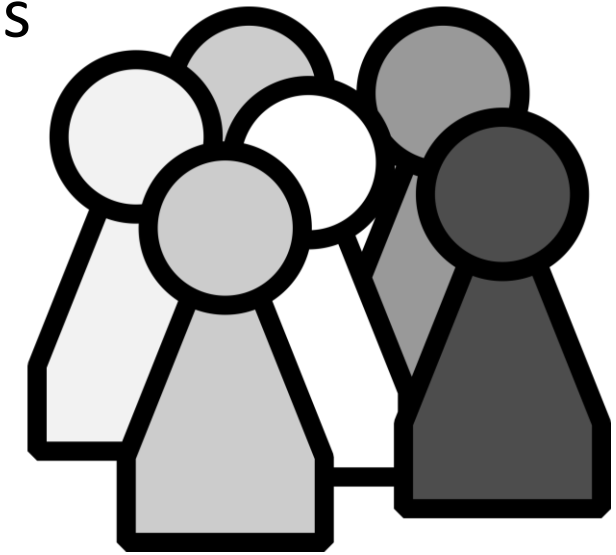
consider a life age x, 2 **states** **A**live and **D**ead

the probability at age x that a life then **A**live will be **D**ead at age x+t is governed by the **transition intensity** $\boldsymbol{\mu_{x+t}}$ (t≥0)



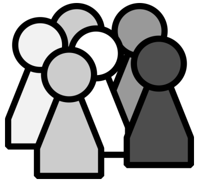
Warning – 2 very different models here

Previous weeks



This week

Week 2-4 and Week 5-6 models different



consider some population

lifetime of an individual in that population is a random variable T

that random variable has distribution function $F_x(t)$ and survival function $S_x(t)$

we seek methods for estimating these functions often using the hazard or force of mortality

consider an individual

that individual may be in one of two states (e.g. alive and dead)

we seek to understand how they might move between the two states dependent on the transition intensity

as we move to models with >2 states the differences in this approach will be magnified

Probabilities in the 2-state model

${}_{t+dt}p_x$

We begin with the survival probability ${}_{t+dt}p_x$

- let us condition on the state occupied at x+t

by the Markov assumption:

P[surviving from x to x+t+dt] =

P[alive at x+t] x P[survive from x+t to x+t+dt | alive at x+t]

+ P[dead at x+t] x P[survive from x+t to x+t+dt | dead at x+t]

= ${}_tp_x$ x P[survive from x+t to x+t+dt | alive at x+t]

+ ${}_tq_x$ x P[survive from x+t to x+t+dt | dead at x+t]

$= {}_tp_x \cdot {}_{dt}p_{x+t} + {}_tq_x \cdot 0 = {}_tp_x\,[1 - \mu_{x+t}dt - o(dt)]$

considering the 2 states at x+t

the 2nd term is nonsense and =0

from our 2nd model assumption

derivative of $_tp_x$

now separately, from the definition of a derivative we have

$$\frac{d}{dt}\,{}_tp_x = \lim_{dt\to 0} \frac{{}_{t+dt}p_x - {}_tp_x}{dt}$$

$$\frac{d}{dt}\,{}_tp_x = \lim_{dt\to 0} \frac{{}_tp_x\,[1 - \mu_{x+t}dt - o(dt)] - {}_tp_x}{dt} \qquad = -\,{}_tp_x\,\mu_{x+t} - \lim_{dt\to 0} \frac{o(dt)}{dt}$$

this term is zero by the definition of $o(dt)$

$$= -\,{}_tp_x\,\mu_{x+t}$$

leads to a familiar formula…

if $\frac{d}{dt}\, {}_tp_x = -\, {}_tp_x\, \mu_{x+t}$

then

$${}_tp_x = \exp\left[-\int_0^t \mu_{x+s}\, ds\right]$$

why get excited about uncovering our 'important formula' again? It is because we have done it entirely within the Markov framework here

in Markov processes this is known as a "Kolmogorov forward equation"

2-state model statistics

observations

We now consider the case where we have observed data which we assume comes from a 2-state model and are looking to derive statistics using this data

assume we observe N lives

- as we study these retrospectively we do not need to assume the observations are independent or even chosen at random
- we allow for censoring

let $x+a_i$ = age at which observation of the i^{th} life begins

and $x+b_i$ = age at which observation of the i^{th} life must cease if the life survives

for $i=1,2,\ldots,N$

indicator variable

we define a random variable D_i

- $D_i = 1$ if the i^{th} life observed to die
- $D_i = 0$ otherwise

D_i is an **indicator random variable** – here it indicates death occurring

we define a second random variable T_i where $x+T_i$ is the age at which the observation of the i^{th} life ends

- $D_i = 0 \rightarrow T_i = b_i$
- $D_i = 1 \rightarrow a_i < T_i < b_i$
- so the two variables D_i and T_i are not independent

waiting time

gives rise to a third variable, V_i the **waiting time** where

$V_i = T_i - a_i$

which has a mixed distribution with a probability mass at $b_i - a_i$

(D_i , V_i)

the pair (D_i , V_i) form a **statistic**

meaning the outcome of our N observations can be seen as a sample (d_i,v_i) taken from the distribution of (D_i , V_i)

let $f_i(d_i,v_i)$ be the joint distribution function of (D_i , V_i)

it is easiest to write $f_i(d_i,v_i)$ considering the two values for D_i separately

$$f_i(d_i,v_i) = \begin{cases} {}_{bi-ai}p_{x+ai} & \text{if } d_i=0 \\ {}_{vi}p_{x+ai} \cdot \mu_{x+ai+vi} & \text{if } d_i=1 \end{cases}$$

Demonstration

See separate PDF on QM Plus site Demonstration 8

if the transition intensity is constant μ

then the joint distribution function is

$$f_i(d_i,v_i) = \exp(-\mu v_i)\ \mu^{d_i}$$

and the joint probability function of all (Di,Vi) is $\exp(-\mu v)\ \mu^d$

where $d = \sum_{i=1}^{N} d_i$ (total number of deaths) and $v = \sum_{i=1}^{N} v_i$ (the total waiting time)

[if μ_{x+t} is a constant μ]

actuarial notation alert

actuaries often call the observed waiting time v the

"central exposed to risk" and denote it E_x^c

2-state model MLE

MLE for μ

with this arrangement and our probability distribution function, the maximum likelihood estimate for the transition intensity μ is quite straightforward

the likelihood function (of parameter μ; given observations for d and v) is

$$L(\mu; d,v) = \exp(-\mu v).\mu^d$$

which gives MLE $\hat{\mu}$ of

$$\hat{\mu} = \frac{d}{v}$$

See separate PDF file Demonstration 9 on QM Plus for more details of this

it can be shown that asymptotically $\hat{\mu} \sim \text{Normal}[\mu, \mu/E(v)]$

application of this model and MLE

remember that to derive our probability function and MLE we have assumed that the transition intensity μ_{x+t} is a constant μ in our range for t

- this assumption is most likely to be reasonable if we keep t short
- generally for practical work we will have $0 \leq t \leq 1$
- that is we are assuming μ_{x+t} is constant between ages x and x+1 (for 1 year)
- then we assume $\hat{\mu}$ is an estimate of $\mu_{x+½}$

we can then piece together these estimates $\hat{\mu}$ at different ages x to get a function for μ_x

- if we need to smooth this function for μ_x we can use a method called graduation which we'll introduce later in this module
- we can then calculate survival probabilities using ${}_tp_x = \exp(-\int \mu_{x+s} ds)$

the general multi-state model

general model

we can extend the 2-state model to any number of states

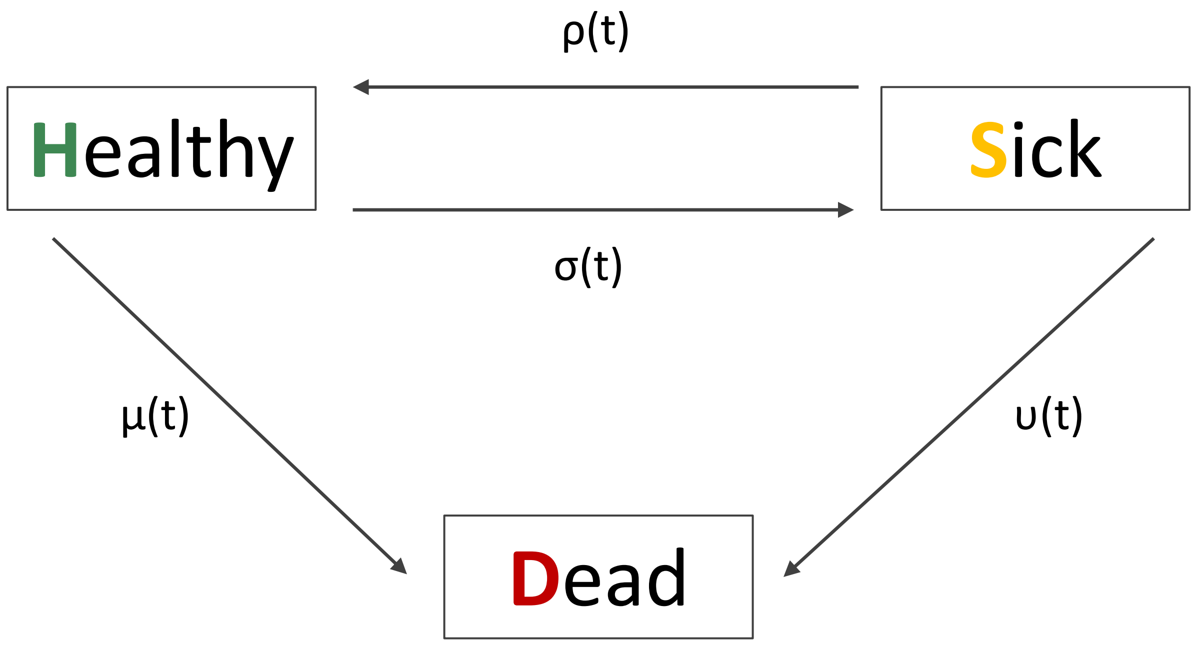
multi-state models are often suited to actuarial data sets

- pension fund membership
- health insurance premiums
- motor insurance claims and no-claims discounts

we observe:

- length of times between transitions (general case of v_i)
- number of transitions of each type (general case of d_i)
- some states will allow two-way movements, others only one-way

e.g. 3 state model



model set-up

3 states: H S D

4 transition intensities: $\sigma(t)$ $\rho(t)$ $\mu(t)$ $\upsilon(t)$

useful to assume for short t, the transition intensities are all constants σ ρ μ υ

we can now look to establish the likelihood function $L(\sigma, \rho, \mu, \upsilon)$

random variables

Random variable	definition	Observed sample
V_i	the waiting time of the i^{th} life in state H	v_i and $v = \sum v_i$
W_i	the waiting time of the i^{th} life in state S	w_i and $w = \sum w_i$
S_i	the number of H→ S transitions for the i^{th} life	s_i and $s = \sum s_i$
R_i	the number of S→ H transitions for the i^{th} life	r_i and $r = \sum r_i$
D_i	the number of H→ D transitions for the i^{th} life	d_i and $d = \sum d_i$
U_i	the number of S→ D transitions for the i^{th} life	u_i and $u = \sum u_i$

likelihood function

building on the work we did for the 2-state model we can see that the likelihood function is of the form

$L(\sigma, \rho, \mu, \upsilon) = \exp[-(\mu+\sigma)v].\exp[-(\upsilon+\rho)w].\mu^d.\upsilon^u.\sigma^s.\rho^r$

when taking log Likelihood this splits into the sum of 4 components (one for each of the parameters) of the form $\exp(-\mu d).\mu^d$ etc.

then the four MLEs are

$\hat{\mu} = d/v \qquad \hat{\upsilon} = u/w \qquad \hat{\sigma} = s/v \qquad \hat{\rho} = r/w$

we could show vector $(\hat{\sigma}, \hat{\rho}, \hat{\mu}, \hat{\upsilon})$ is asymptotically Normal with means $\sigma, \rho, \mu, \upsilon$ but these estimators are <u>not</u> independent (as e.g. D_i and U_i cannot both be 1)