Probability function for two-state Markov models

• T_i is random variable describing future lifetime of life i

• V_i is the time life i is observed, $V_i = \min(T_i, \text{censoring point})$

• D_i is the death indicator, i.e. $D_i = 1$ if life dies, $D_i = 0$ if life censored

• How to derive the joint distribution of (D_i, V_i)?

• If there were no censoring ($d_i = 1$), the probability density of V_i is the same as T_i, namely:

$$f_i(d_i = 1, v_i) = S_i(v_i)\mu_{x_i+v_i}$$

where $S_i()$ is the survivor function for future lifetime.

• Where there is censoring ($d_i = 0$), the probability function is simply the survivor function up to time v_i:

$$f_i(d_i = 0, v_i) = S_i(v_i)$$

• Since $\mu^0_{x_i+v_i} = 1$, we can write a single probability function for the joint distribution of (D_i, V_i):

$$f_i(d_i, v_i) = S_i(v_i)\mu^{d_i}_{x_i+v_i}$$

• We have the general result $S_i(v_i) = e^{-H_i(v_i)}$, where $H_i()$ is the integrated hazard function:

$$f_i(d_i, v_i) = e^{-H_i(v_i)}\mu^{d_i}_{x_i+v_i}$$

• This is a general result. For most of our examples we will assume a constant force of mortality, so:

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

• As an aside, we note that $0! = 1! = 1$ and $v_i^0 = 1$, so:

$$f_i(d_i = 0, v_i) = e^{-\mu v_i}.\frac{(\mu v_i)^{d_i}}{d_i!}$$

$$f_i(d_i = 1, v_i) = e^{-\mu v_i}.\frac{(\mu v_i)^{d_i}}{d_i!}.\frac{1}{v_i}$$

• The first probability is that of zero events in a Poisson process with parameter μv_i.

• The second probability is that of a single event in a Poisson process with parameter μv_i, but where this probability is scaled by $\frac{1}{v_i}$.

• On a log scale, the scale factor $\frac{1}{v_i}$ would be an addition of $-\log v_i$, i.e. an *offset* of $\log v_i$

Data and estimation — maximum likelihood

• Sample of n lives with waiting times $v' = (v_1, v_2, \ldots, v_n)$ and death indicators $d' = (d_1, d_2, \ldots, d_n)$

• The likelihood function, L, is the product of the probability functions evaluated at the data points:

$$\begin{aligned} L(\mu; d, v) &= \prod_{i=1}^{n} f_i(d_i, v_i) \\ &= \prod_{i=1}^{n} e^{-\mu v_i}.\mu^{d_i} \\ &= e^{-\mu \sum v_i}.\mu^{\sum d_i} \end{aligned}$$

where $\sum v_i$ is the total waiting time and $\sum d_i$ is the total number of deaths.

• Actuaries call $\sum v_i$ the *central exposed to risk.*

• Maximum-likelihood estimate of μ, $\hat{\mu}$, is given by $\sum d_i / \sum v_i$

Comparison with Cox model

• The above is a full likelihood function

• Contrast with Cox *partial likelihood*

• Markov model is using all information

• Markov model has no trouble with ties in death times

• Markov model can estimate shape of hazard function

Example in R

Consider a small sample of six lives (based on Example 5.1 on page 46 of the Red Book):

i	v_i	d_i
1	1.00	0
2	0.50	1
3	0.50	0
4	0.25	1
5	0.25	1
6	0.25	0

where $\sum v_i = 2.75$ and $\sum d_i = 3$, so $\hat{\mu} = 1.090909$.