

Lecture 7

Multiple decrements model

Reading: Gerber Sections 7.1-7.3 and 11.7, Cox-Oakes Sections 9.1-9.2, Norris Section 1.10, CT4 Units 9,10-3,10-4
Further reading: Cox-Oakes Section 9.3

7.1 The Poisson model

Under the assumption of a constant hazard rate (force of mortality) $\mu_{x+\frac{1}{2}}$ over the year $(x, x+1]$, we may view the estimation problem as a chain of separate hazard rate estimation problems, one for each year of life. Each individual lives some portion of a year in the age interval $(x, x+1]$, the portion being 0 (if he dies before birthday x), 1 (if he dies after birthday $x+1$), or between 0 and 1 if he dies between the two birthdays. Suppose now we lay these intervals end to end, with a mark at the end of an interval where an individual died. It is not hard to see that what results is a Poisson process on the interval $[0, E_x^c]$, where E_x^c is the total observed years at risk.

Suppose we treat E_x^c as though it were a constant. Then if D_x represents the numbers dying in the year the model uses

$$P\{D_x = k\} = \frac{\left(\mu_{x+\frac{1}{2}} E_x^c\right)^k e^{-\mu_{x+\frac{1}{2}} E_x^c}}{k!}, \quad k = 0, 1, 2, \dots$$

which is an approximation to the 2-state model, and which in fact yields the same likelihood.

The estimator for the constant force of mortality over the year is

$$\tilde{\mu}_{x+\frac{1}{2}} = \frac{D_x}{E_x^c}, \quad \text{with estimate } \frac{d_x}{E_x^c}.$$

Under the Poisson model we therefore have that

$$\text{var} \tilde{\mu}_{x+\frac{1}{2}} = \frac{\mu_{x+\frac{1}{2}} E_x^c}{(E_x^c)^2} = \frac{\mu_{x+\frac{1}{2}}}{E_x^c}.$$

So the estimate will be

$$\text{var} \tilde{\mu}_{x+\frac{1}{2}} \approx \frac{d_x}{(E_x^c)^2}.$$

If we compare with the **2-state stochastic model** over year $(x, x + 1)$, assuming constant $\mu = \mu_{x+\frac{1}{2}}$, then the likelihood is

$$L = \prod_1^n \mu^{\delta_i} e^{-\mu t_i},$$

where $\delta_i = 1$ if life i dies and $t_i = b_i - a_i$ in previous terminology (see the binomial model). Hence

$$L = \mu^{d_x} e^{-\mu E_x^c}$$

and so

$$\hat{\mu} = \frac{D_x}{E_x^c}.$$

The estimator is exactly the same as for the Poisson model except that both D_x and E_x^c are random variables. Using asymptotic likelihood theory we see that the estimate for the variance is

$$\text{var} \hat{\mu} \approx \frac{\mu^2}{d_x} \approx \frac{d_x}{(E_x^c)^2}.$$

Are we justified in treating E_x^c as though it were fixed? Certainly it's not exactly the same: The numerator and denominator are both random, and they are not even independent. One way of looking at this is to ask, how different would our estimate have been in a given realisation, had we fixed the total time under observation in advance. If we observe m lives from the start of year x , we see that D_x is approximately normal with mean $m q_x$ and variance $m q_x (1 - q_x)$, while E_x^c is normal with mean $m - m q_x (1 - e_*)$, where e_* is the expected remaining length of a life starting from age x , conditioned on its being less than 1; and variance $m \sigma^2$. (If μ_x is not very large, this is close to $\frac{1}{2}$.) Looking at the first-order Taylor series expansion, we see that the ratio D_x/E_x^c varies only by a normal error term times $m^{-1/2}$, plus a bias of order m^{-1} . For large m , then, the estimate on the basis of fixed m (number of individuals) is almost the same as the estimate we would have made from observing the Poisson model for the fixed total time at risk $m - m q_x (1 - e_*)$.

7.2 Rates in the single decrement model

The rate parameter in the two-state Markov model with $L \sim \text{Exp}(\lambda)$ has the infinitesimal interpretation

$$\mathbb{P}(X_{t+\varepsilon} = 1 | X_t = 0) = \mathbb{P}(L \leq t + \varepsilon | L > t) = 1 - e^{-\lambda \varepsilon} = \lambda \varepsilon + o(\varepsilon),$$

and for a general L with right-continuous density and hence right-continuous force of mortality $t \mapsto \mu_t$, we have

$$\mathbb{P}(X_{t+\varepsilon} = 1 | X_t = 0) = \mathbb{P}(L \leq t + \varepsilon | L > t) = 1 - \exp \left\{ - \int_t^{t+\varepsilon} \mu_s ds \right\} = \mu_t \varepsilon + o(\varepsilon),$$

since, by l'Hôpital's rule

$$\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \left(1 - \exp \left\{ - \int_t^{t+\varepsilon} \mu_s ds \right\} \right) = \lim_{\varepsilon \downarrow 0} \mu_{t+\varepsilon} \exp \left\{ - \int_t^{t+\varepsilon} \mu_s ds \right\} = \mu_t.$$

It is therefore natural to express the two-state model by a time-dependent Q -matrix

$$Q(t) = \begin{pmatrix} -\lambda(t) & \lambda(t) \\ 0 & 0 \end{pmatrix}, \quad \text{where } \lambda(t) = \mu_t = h_L(t).$$

For estimation purposes, it has been convenient to add as additional assumption that $\lambda(t) = \mu_t = \mu_{x+\frac{1}{2}} = \lambda(x + \frac{1}{2})$ is constant on $x \leq t < x + 1$, $x \in \mathbb{N}$.

We have expressed the process $X = (X_t)_{t \geq 0}$ as $X_t = 0$ for $0 \leq t < L$ and $X_t = 1$ for $t > L$, where L is the transition time. Given the observed transition times y_1, \dots, y_n of n independent copies of X (corresponding to n different ‘individuals’), we have constructed two different sets of maximum likelihood estimates

$$\begin{aligned} \hat{\mu}_{x+\frac{1}{2}}^{(0)}(y_1, \dots, y_n) &= -\ln \left(\hat{q}_x^{(0)}(y_1, \dots, y_n) \right) = \ln \left(1 - \frac{d_x(y_1, \dots, y_n)}{\ell_x(y_1, \dots, y_n)} \right), \\ \hat{\mu}_{x+\frac{1}{2}}(y_1, \dots, y_n) &= \frac{d_x(y_1, \dots, y_n)}{\tilde{\ell}_x(y_1, \dots, y_n)}, \quad 0 \leq x \leq \max\{y_1, \dots, y_n\}. \end{aligned}$$

If we furthermore assume that $\lambda(t) \equiv \lambda$ for all $t \geq 0$, then the maximum likelihood estimator is simply

$$\hat{\lambda} = \frac{n}{y_1 + \dots + y_n} = \frac{d_0 + \dots + d_{[\max\{y_1, \dots, y_n\}]}}{\tilde{\ell}_0 + \dots + \tilde{\ell}_{[\max\{y_1, \dots, y_n\}]}}.$$

7.3 Multiple decrement models

The simplest (and most immediately fruitful) way to generalise the single-decrements model is to allow transitions to multiple absorbing states. Of course, as demographer Kenneth Wachter has put it, it may seem peculiar to introduce multiple “dead” states into our models since there is only one way of being dead; but (as he continues), there are many ways of getting there. Further, there are many other settings which can be modelled by a single nonabsorbing state transitioning into one of several possible absorbing states. Some examples are

- A working population insured for disability might transition into multiple different possible causes of disability, which may be associated with different costs.
- Workers may leave a company through retirement, resignation, or death.
- A model of unmarried cohabitations, which may end either by separation or marriage.
- Unemployed individuals may leave that state either by finding a job, or by giving up looking for work and so becoming “long-term unemployed”.

An important common element is that calling the states “absorbing” does not have to mean that it is a deathlike state, from which nothing more happens. Rather, it simply means that our model does not follow any further developments.

7.3.1 An introductory example

This example is taken from section 8.2 of [Wac].

According to United Nations statistics, the probability of dying for men in Zimbabwe in 2000 was ${}_5q_{30} = 0.1134$, with AIDS accounting for approximately $4/5$ of the deaths in this age group. Suppose we wish to answer the question: what would be the effect on mortality rates of a complete cure for AIDS?

One might immediately be inclined to think that the mortality rate would be reduced to $1/5$ of its current rate, so that what the probability of dying of some other cause in the absence of AIDS, which we might write as ${}_5q_{30}^{OTHER*}$, would be 0.02268. On further reflection, though, it seems that this is too low: This is the proportion of people aged 30 who *currently* die of causes other than AIDS. If AIDS were eliminated, surely some of the people who now die of AIDS would instead die of something else.

Of course, this is not yet a well-defined mathematical problem. To make it such, we need to impose extra conditions. In particular, we impose the **competing risks** assumption: Individual causes of death are assumed to act independently. You might imagine an individual drawing lots from multiple urns, labelled “AIDS”, “Stroke”, “Plane crash”, to determine whether he will die of this cause in the next year. The fraction of black lots among the white is precisely q_x , when the individual has age x . If he gets no black lot, he survives the year. If he draws two or more, we only get to see the one drawn first, since he can only die once. The probability of surviving is then the product of the survival probabilities:

$${}_i q_x = 1 - (1 - {}_i q_x^{CAUSE1})(1 - {}_i q_x^{CAUSE2}) \dots \quad (1)$$

What is the fraction of deaths due to a given cause? Assuming constant mortality rate over the time interval due to each cause, we have

$$1 - {}_t q_x^{CAUSE1} = e^{-t\lambda_x^{CAUSE1}}.$$

Given a death, the probability of it being due to a given cause, is proportional to the associated hazard rate. Consequently,

$$\lambda_x^{CAUSE1} = \text{fraction of deaths due to CAUSE 1} \times \lambda_x,$$

which implies that

$${}_t q_x^{CAUSE1} = 1 - (1 - {}_t q_x)^{\text{fraction of deaths due to CAUSE 1}}.$$

(Note that this is the same formula that we use for changing lengths of time intervals: ${}_t q_x = 1 - (1 - {}_1 q_x)^t$.) This tells us the probability of dying from cause 1 in the absence of any other cause. The probability of dying of any cause at all is then given by (1).

Applying this to our Zimbabwe AIDS example, treating the causes as being either AIDS or OTHER, we see that the probability of dying of AIDS in the absence of any other cause is

$${}_5 q_{30}^{AIDS*} = 1 - (1 - {}_5 q_{30})^{4/5} = 1 - 0.8866^{4/5} = 0.0918,$$

while the probability of dying of any other cause, in the absence of AIDS, is

$${}_5 q_{30}^{OTHER*} = 1 - (1 - {}_5 q_{30})^{1/5} = 1 - 0.8866^{1/5} = 0.0238.$$

Appropriately, we have the total cause of death $0.1138 = 1 - (1 - 0.0918)(1 - 0.0238)$.

Is the competing risks assumption reasonable? Another way of putting this is to ask, what circumstances would cause the assumption to be violated? The answer is: Competing risks is violated when a subpopulation is at higher than average risk for multiple causes of death simultaneously; or conversely, when those at higher than average risk for one cause of death are protected from another cause of death. For example, smokers have more than 10 times the risk of dying from lung cancer than nonsmokers have; but they also have substantially higher mortality from other cancers, heart disease, stroke, and so on. If a perfect cure for lung cancer were to be found, it would not save nearly as many lives as one might suppose, from a competing-risks calculation like the one above, because the lives that would be saved would be almost all those of smokers, and they would be more likely to die of something else than an equivalent number of saved lives from the general population.

7.3.2 Basic theory

We consider here more general $1+m$ -state Markov models with state space $\mathbb{S} = \{0, \dots, m\}$ that only have one transition from 0 to j , for some $1 \leq j \leq m$, with absorption in j . We can write down an – in general time-dependent – Q -matrix

$$Q(t) = \begin{pmatrix} -\lambda_+(t) & \lambda_1(t) & \cdots & \lambda_m(t) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \text{where } \lambda_+(t) = \lambda_1(t) + \dots + \lambda_m(t).$$

Such models occur naturally where insurance policies provide different benefits for different causes of death, or distinguish death and disability, possibly in various different strengths or forms. This is also clearly a building block (one transition only) for general Markov models, where states $j = 1, \dots, m$ may not all be absorbing.

Such a model depends upon the assumption that different causes of death act independently — that is, the probability of dying is the product of what might be understood as the probability of dying from each individual cause acting alone.

7.3.3 Multiple decrements – time-homogeneous rates

In the time-homogeneous case, we can think of the multiple decrement model as m exponential clocks C_j with parameters λ_j , $1 \leq j \leq m$, and when the first clock goes off, say, clock j , the only transition takes place, and leads to state j . Alternatively, we can describe the model as consisting of one $L \sim \text{Exp}(\lambda_+)$ holding time in state 0, after which the new state j is chosen *independently* with probability λ_j/λ_+ , $1 \leq j \leq m$. The likelihood for a sample of size 1 consists of two ingredients, the density $\lambda_+ e^{-t\lambda_+}$ of the exponential time, and the probability λ_j/λ_+ of the transition observed. This gives $\lambda_j e^{-t\lambda_+}$, or, for a sample of size n of lifetimes t_i and states j_i , $1 \leq i \leq n$,

$$\prod_{i=1}^n \lambda_{j_i} e^{-t_i \lambda_+} = \prod_{j=1}^m \lambda_j^{n_j} e^{-\lambda_j(t_1 + \dots + t_n)},$$

where n_j is the number of transitions to j . Again, this can be solved factor by factor to give

$$\hat{\lambda}_j = \frac{n_j}{t_1 + \dots + t_n}, \quad 1 \leq j \leq m.$$

In particular, we find again $\hat{\lambda}_+ = n/(t_1 + \dots + t_n)$, since $n_1 + \dots + n_m = n$.

In the competing-clocks description, we can interpret the likelihood as consisting of m ingredients, namely the density $\lambda_j e^{-\lambda_j t}$ of clock j to go off at time t , and probabilities $e^{-\lambda_k t}$ of clocks C_k , $k \neq j$, to go off after time t .