B.4 Multiple decrements and general Markov models

1. (a) We are not given the size n of the group. However, this is not essential since the factorised form of the likelihood does not depend on n. We obtain

$$\prod_{i \in \mathbb{S}} \prod_{j \neq i} q_{ij}^{N_{ij}} \exp\{-q_{ij}E_i\} = \sigma^{N_{HS}} e^{-\sigma E_H} \mu^{N_{H\Delta}} e^{-\mu E_H} \rho^{N_{SH}} e^{-\rho E_S} \nu^{N_{S\Delta}} e^{-\nu E_S}.$$

This can be maximised parameter by parameter. To maximise in σ, we maximise $\sigma^{N_{HS}} e^{-\sigma E_H}$ or, passing to logs,

$$\ell(\sigma) = N_{HS} \log(\sigma) - \sigma E_H \Rightarrow \ell'(\sigma) = \frac{N_{HS}}{\sigma} - E_H \Rightarrow \ell''(\sigma) = -\frac{N_{HS}}{\sigma^2} < 0$$

and ℓ is maximized for $\hat{\sigma} = N_{HS}/E_H$. For $N_{HS} = 15$ and $E_H = 625$ this is $\hat{\sigma} = 15/625 = 0.024$.

(b) For the asymptotic distribution, we require the Fisher Information. The likelihood factorises, so the log likelihood is the sum of functions of single parameters, so the Fisher Information matrix is diagonal, and we calculate, approximating the Fisher Information by the observed information and its estimate

$$I_{\sigma\sigma} = -\mathbb{E}(\ell''(\sigma)) \approx \frac{N_{HS}}{\sigma^2} \approx \frac{E_H^2}{N_{HS}}.$$

From the asymptotic theory, $\hat{\sigma} \sim \mathcal{N}(\sigma, N_{HS}/E_H^2)$.

(c) In particular, $\sqrt{N_{HS}}/E_H = \sqrt{15}/625 = 0.0062$ is an estimate of the standard deviation of $\hat{\sigma}$.

(d) Then $[\hat{\sigma} - 1.96\sqrt{N_{HS}}/E_i, \hat{\sigma} + 1.96\sqrt{N_{HS}}/E_i] = [0.012, 0.036]$ is an approximate 95% confidence interval.

(e) For solvency the company requires aggregate contributions to be greater than or equal to aggregate benefits, i.e. here $CE_H \geq BE_S$, so $C/B \geq E_S/E_H = 35/625$. Second order considerations (variances) can be used to give confidence intervals around such a value. We neglect here administrative expenses and further risk considerations.

(f) Particularly, if the age range is wide, suggestion i. can improve the predictive power. ii. is likely to be of minor importance since the effects will average out over the year anyway. iii. must be expected to have impact on the variances since long-term illnesses cause a lot of benefit payments from only few individuals.

Markov models with age-dependent transition rates such as i. were done in the lectures. For ii., the techniques can be easily adapted if appropriate data are recorded (birthdays or numbers at risk every season, say, and transitions and waiting times every season). For iii., a bit more work is required to adapt the methods, since the Markovian character is lost.

We probably have about 650 subjects in the group. With a total of 27 transitions, this is not anywhere near enough to fit a refined model, neither to distinguish ages nor, say, the four seasons to discretize the time of the year or rates changing with duration of illness in months.

2. (a) i. Since only transitions from i to $i+1$ and $i-1$ are possible, the likelihood can be written as

$$\begin{aligned} &\lambda^{N_{01}} \exp\{-\lambda E_0\} \prod_{i=1}^{\infty} \lambda^{N_{i,i+1}} \mu^{N_{i,i-1}} \exp\{-(\lambda+\mu)E_i\} \\ &= \lambda^{N_+} \exp\{-\lambda E_+\} \mu^{N_-} \exp\{-\mu E_-\}, \end{aligned}$$

where

$$N_+ = \sum_{i=0}^{\infty} N_{i,i+1}, \quad N_- = \sum_{i=1}^{\infty} N_{i,i-1}, \quad E_+ = \sum_{i=0}^{\infty} E_i, \quad E_- = \sum_{i=1}^{\infty} E_i.$$

From the form of the likelihood we easily compute $\hat{\lambda} = N_+/E_+$ and $\hat{\mu} = N_-/E_-$.

ii. We use the observed information matrix $\tilde{I}$ to estimate the Fisher information matrix I. It is diagonal, since the likelihood fully factorises. We obtain

$$\tilde{I}_{\lambda\lambda} = \frac{N_+}{\lambda^2}, \qquad \tilde{I}_{\mu\mu} = \frac{N_-}{\mu^2}.$$

By the asymptotic theory of maximum likelihood estimators, $(\hat{\lambda}, \hat{\mu}) \sim \mathcal{N}((\lambda, \mu), \tilde{I}^{-1})$, approximately, for large n. Since Normally distributed random variables are independent if and only if they are uncorrelated, we deduce that $\hat{\lambda}$ and $\hat{\mu}$ are asymptotically independent.

iii. Asymptotically, $\lambda - \hat{\lambda}$ is normal with mean 0 and variance $\sigma_1^2 := \hat{\lambda}^2/N_+$ and $\mu - \hat{\mu}$ is normal with mean 0 and variance $\sigma_2^2 := \hat{\mu}^2/N_-$. Thus, a (1-$\alpha$)-CI for λ is $\hat{\lambda} \pm u_{\alpha/2}\sigma_1$, and a 95% CI for μ is $\hat{\mu} \pm u_{\alpha/2}\sigma_2$.
We can represent these errors as approximately $\sigma_1 Z_1 + \sigma_2 Z_2$, where Z_1 and Z_2 are independent standard normal random variables. The curves of constant probability are thus ellipsoids, and the approximate (1-α)-confidence region of minimal area will be

$$\left\{ (\lambda, \mu) : \frac{(\lambda - \hat{\lambda})^2}{\sigma_1^2} + \frac{(\mu - \hat{\mu})^2}{\sigma_2^2} \le c_\alpha \right\},$$

where c_α is the $1 - \alpha$ quantile of a χ^2 distribution with 2 degrees of freedom.

(b) i. The state space is $\mathbb{S} = \{0, \ldots, m\}$. We allow tridiagonal Q-matrices with parameters $q_{i,i+1} = \lambda_i$ and $q_{i+1,i} = \mu$, $i = 0, \ldots, m-1$.

ii. The likelihood function is now

$$\begin{aligned} &\lambda_0^{N_{01}} \exp\{-\lambda E_0\} \mu^{N_{m,m-1}} \exp\{-\mu E_m\} \prod_{i=1}^{m-1} \lambda_i^{N_{i,i+1}} \mu^{N_{i,i-1}} \exp\{-(\lambda_i + \mu) E_i\} \\ &= \left(\prod_{i=0}^{m-1} \lambda_i^{N_{i,i+1}} \exp\{-\lambda_i E_i\} \right) \mu^{N_-} \exp\{-\mu E_-\}. \end{aligned}$$

Therefore, the maximum likelihood estimators are $\hat{\lambda}_i = N_{i,i+1}/E_i$, $i = 0, \ldots, m-1$, and $\hat{\mu} = N_-/E_-$.

iii. As in (a), the asymptotic distribution of the maximum likelihood estimators is multivariate Normal with diagonal variance-covariance matrix. In particular, for $m = 2$,

$$\frac{\hat{\lambda}_0 - \hat{\lambda}_1}{\sqrt{\mathrm{Var}(\hat{\lambda}_0) + \mathrm{Var}(\hat{\lambda}_1)}} \sim \mathcal{N}(0,1) \qquad \text{and} \qquad \frac{(\hat{\lambda}_0 - \hat{\lambda}_1)^2}{\mathrm{Var}(\hat{\lambda}_0) + \mathrm{Var}(\hat{\lambda}_1)} \sim \chi_1^2.$$

iv. The problem is to compute the variance of N_i/E_i. It is reasonable to treat E_i as fixed, and simply compute the conditional variance. There are several ways to approach this. We might first say that the conditional distribution of N_i, conditioned on a particular value of E_i, is Poisson with parameter λE_i, so the conditional variance of N_i is also $\lambda_i E_i$, and the conditional variance of $\hat{\lambda}$ is λ_i/E_i. This is not exactly true, since we have also assumed that there are exactly n transitions, so $N_- + N_0 + N_1 = n$. This means that the transitions during a period of time E_i do not really have a Poisson distribution. How much does this matter?
The answer is, not much, asymptotically. We would expect that, since the dependence between two transitions will be very small when the total number is very large. The correct calculation is as follows: Consider just $\hat{\lambda}_0$, and assume n is even. Conditioned

on N_0, there is a gamma distribution for E_i, with parameters N_0 and λ_1. We have then

$$\begin{aligned}\mathbb{E}\big[(E_i)^{-1}\big] &= \frac{\lambda_0}{N_0 - 1}\\ \mathbb{E}\big[(E_i)^{-2}\big] &= \frac{(\lambda_0)^2}{(N_0-1)(N_0-2)}\\ Var(E_i^{-1}) &= \frac{(\lambda_0)^2}{(N_0-1)^2(N_0-2)} \approx (N_0)^{-3}(\lambda_0)^2.\end{aligned}$$

Thus, conditioned on N_0, we have $Var(\hat{\lambda}_0) \approx \lambda_0^2/N_0$. Since $N_0/E_0 \to \lambda_0$, this is very close to λ_0/E_0 for large n, which is the result we had before.

v. *

3. (a) Let T_j be the total time spent in state j before death, and W the total value of a life. Then letting $X(t)$ be the state of the individual at age t (and $w_D = 0$ for the absorbing state),

$$W = \int_0^\infty w_{X(t)} dt = \int_0^\infty \sum_j w_j \mathbf{1}_{\{X(t)=j\}} dt = \sum_j w_j T_j.$$

Thus $E[W] = \sum w_j E[T_j]$.
From Theorem 10.5.1 we know that $E_i[T_j] = (Q_*^{-1})_{i,j}$. If we start in state i with probability p_i, this expected total time becomes $\sum_i p_i (Q_*^{-1})_{i,j}$, and

$$E[W] = \sum_i p_i (Q_*^{-1})_{i,j} w_j = p^T Q_*^{-1} w.$$

(b) Suppose the model is the simple Healthy-Sick-Dead model described in section 10.4.2, with $\sigma = 0.1$, $\rho = 0$, $\delta = 0.01$, and $\gamma = 0.2$. What is the life expectancy of someone initially healthy?
We have

$$Q = \begin{pmatrix} -0.11 & 0.1 & 0.01\\ 0 & -0.2 & 0.2\\ 0 & 0 & 0\end{pmatrix} \qquad Q_* = \begin{pmatrix} -0.11 & 0.1\\ 0 & -0.2\end{pmatrix} \qquad -Q_*^{-1} = \begin{pmatrix} 9.1 & 4.55\\ 0 & 5\end{pmatrix}$$

So, the healthy person has, on average 13.65 remaining years.

(c) The healthy person has, on average, 9.1 remaining healthy years and 4.55 sick years, which count for 2.3 QALYs. Thus, 11.4 QALYs altogether. The sick person has 2.5 QALYs on average.

(d) We now have

$$Q = \begin{pmatrix} -0.11 & 0.1 & 0.01\\ 0.2 & -0.4 & 0.2\\ 0 & 0 & 0\end{pmatrix} \qquad Q_* = \begin{pmatrix} -0.11 & 0.1\\ 0.2 & -0.4\end{pmatrix} \qquad -Q_*^{-1} = \begin{pmatrix} 16.7 & 4.2\\ 8.3 & 4.6\end{pmatrix}$$

Thus, we would be increasing the QALY expectancy for healthy people to 18.8, and for sick people to 10.6.
Obviously, we could achieve unlimited life-expectancy increase by reducing γ down to 0. We can compute this exactly by

$$Q_* = \begin{pmatrix} -0.11 & 0.1\\ 0 & -\gamma\end{pmatrix} \qquad -Q_* = \begin{pmatrix} 9.1 & \frac{0.91}{\gamma}\\ 0 & \frac{1}{\gamma}.\end{pmatrix}$$

Thus, to get sick people up to 10.6 expected QALYs we need $\gamma = 1/21.2 = 0.047$; and to get healthy people up to 18.8 QALYs we need $0.91/\gamma = 19.4$, which is also $\gamma = 0.047$.

(e) We write the state space with absorbing states D_H and D_S, representing what state the individual died from. The generator then becomes

$$Q = \begin{pmatrix} -0.11 & 0.1 & 0.01 & 0 \\ 0.2 & -0.4 & 0 & 0.2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

By Theorem 10.5.2 we need to find the eigenvectors with eigenvalue 0. We obtain

$$\begin{pmatrix} 0.17 \\ 0.083 \\ 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 0.83 \\ 0.917 \\ 0 \\ 1 \end{pmatrix}.$$

(f) We want to estimate the average number of years spent in the healthy and sick states. We begin at the end: The remaining life of a healthy or sick "old" person. We get the matrix

$$-Q_*^{-1} = \begin{pmatrix} 3.85 & 4.81 \\ 1.92 & 4.90 \end{pmatrix},$$

so that the worth of the remaining life is 6.25 or 4.37 when starting from H or S respectively. We can diagonalise Q_* by

$$Q_* = \begin{pmatrix} -0.51 & 0.5 \\ 0.20 & -0.40 \end{pmatrix} = \begin{pmatrix} -0.883 & -0.799 \\ 0.470 & -0.601 \end{pmatrix} \begin{pmatrix} -0.776 & 0 \\ 0 & -0.134 \end{pmatrix} \begin{pmatrix} -0.663 & 0.882 \\ -0.518 & -0.974 \end{pmatrix}$$

It's clear that the resulting distribution after 30 years will depend only on the eigenvector with eigenvalue -0.134. The exact computation is

$$e^{30Q_*} = \begin{pmatrix} -0.883 & -0.799 \\ 0.470 & -0.601 \end{pmatrix} \begin{pmatrix} e^{-0.776\cdot 30} & 0 \\ 0 & e^{-0.134\cdot 30} \end{pmatrix} \begin{pmatrix} -0.663 & 0.882 \\ -0.518 & -0.974 \end{pmatrix} = \begin{pmatrix} 0.0074 & 0.014 \\ 0.0056 & 0.0105 \end{pmatrix}$$

This gives the probability that, after 30 years, a person is in states H and S (columns) when starting from H or S (rows).
From Theorem 10.5.1 we see that

$${}_yE_x(30) \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} = Q_*^{-1}\left(e^{30Q_*} - I\right) \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 6.1 \\ 4.3 \end{pmatrix}$$

Thus, the expected number of QALYs before age 30 for a healthy person is 6.1, and after age 30 is

$$0.0074 \cdot 6.1 + 0.014 \times 4.3 = 0.11,$$

so 6.2 in all. For a sick person, the corresponding numbers are 4.3 before age 30, and

$$0.0056 \cdot 6.25 + 0.01 \times 4.37 = 0.08,$$

so 4.4 in all.