B.4 Graduation, goodness of fit and Markov models

1. We compute $z_x$ for each age class by

$$z_x = \frac{d_x - E_x q_x^s}{\sqrt{E_x q_x^s(1 - q_x^s)}}.$$ 

<table>
<thead>
<tr>
<th>Age</th>
<th>Exposed to risk</th>
<th>Observed deaths</th>
<th>Expected deaths</th>
<th>$z_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–24</td>
<td>35000</td>
<td>35</td>
<td>33.95</td>
<td>0.1803</td>
</tr>
<tr>
<td>25–29</td>
<td>33000</td>
<td>30</td>
<td>29.04</td>
<td>0.1782</td>
</tr>
<tr>
<td>30–34</td>
<td>30000</td>
<td>31</td>
<td>35.1</td>
<td>-0.6924</td>
</tr>
<tr>
<td>35–39</td>
<td>30000</td>
<td>45</td>
<td>51.9</td>
<td>-0.9586</td>
</tr>
<tr>
<td>40–44</td>
<td>31000</td>
<td>84</td>
<td>80.6</td>
<td>0.3792</td>
</tr>
<tr>
<td>45–49</td>
<td>28000</td>
<td>138</td>
<td>128.8</td>
<td>0.8125</td>
</tr>
<tr>
<td>50–54</td>
<td>25000</td>
<td>229</td>
<td>212.5</td>
<td>1.1367</td>
</tr>
<tr>
<td>55–59</td>
<td>23000</td>
<td>360</td>
<td>345</td>
<td>0.8137</td>
</tr>
<tr>
<td>60–64</td>
<td>20000</td>
<td>522</td>
<td>500</td>
<td>0.9964</td>
</tr>
</tbody>
</table>

(a) The $X^2$ statistic is of the form $X^2 = \sum z_x^2$.

Substituting in the values from the table, we get $X^2 = 5.214$. Since this corresponds to $\chi^2$ with 9 degrees of freedom, we get a $p$-value of 0.815. This is consistent with the null hypothesis that the mortality rates are as given in the standard table.

(b) Sign test: We observe 7 positives out of 9 tries. Under the null hypothesis these 7 should be like $P = \text{Binom}(9, \frac{1}{2})$. The $p$-value is

$$P\{0, 1, 2, 7, 8, 9\} = \sum_{0,1,2,7,8,9} \left(\frac{1}{2}\right)^9 \binom{9}{k} = 0.18.$$ 

(c) The cumulative deviations test statistic is

$$Z = \frac{\sum (d_x - E_x q_x^s)}{\sqrt{\sum E_x q_x^s(1 - q_x^s)}},$$

which should have approximately $N(0, 1)$ distribution under the null hypothesis. We compute $Z = 1.27$, yielding a $p$-value (for the 2-sided $Z$ test) of 0.20.

The lower $p$-values for the tests for the signs test and the cumulative deviations test make sense, since there is some visible bias in the data, but they are still too high to be considered as significant evidence that the data did not come from the standard table.

Nonetheless we might suspect that the standard life table is underestimating mortality. This would be a bad thing for an insurance company, which would then set its premiums too low. Graduation, for example using $q_0^0 = b + aq_x^s$, could be appropriate.

2. (a) Crude estimates from the data are subject to stochastic fluctuation. Smoothing (graduating) the estimates may make more reliable predictions.

(b) We work with a model where mortality rates are constant on 5-year intervals.

The $z_x$ are calculated by $z_x = (d_x - E_x \hat{q}_x^c \hat{\mu}_{x+2.5})/\sqrt{E_x \hat{q}_x^c \hat{\mu}_{x+2.5}}$. We have approximately that the $z_x$ are i.i.d. $N(0, 1)$ random variables. So the likelihood is approximately

$$(\sqrt{2\pi})^{-n} \exp(-\sum z_x^2/2)$$

and maximising this over graduation parameters corresponds to minimising $\sum z_x^2$. 

$$\text{(B.1)}$$
For Gompertz-Makeham, $\mu_x = a + be^{\alpha x}$. (This is generally considered a reasonable model for the force of mortality from middle age onward.) To maximise the likelihood under the null hypothesis that the data comes from the Gompertz-Makeham family, we estimate $k = 3$ parameters. Meanwhile, the alternative hypothesis of general rates $\mu_x + \alpha x$, we have $m = 8$ free parameters. Still making the normal approximation, under the alternative hypothesis we can take $\hat{\alpha} \equiv \alpha$, leading to $\sum z^2_x \equiv 0$, so that the likelihood corresponding to (B.1) is simply $(\sqrt{2\pi})^{-n}$. The likelihood ratio between null hypothesis and alternative hypothesis is then $\lambda = \exp(-\sum z^2_x/2)$. The asymptotic theory for likelihood ratios tells us that $-2\log \lambda$ should have asymptotic distribution $\chi^2_{m-k}$. So here we have approximately $\sum z^2_x \approx \chi^2_5$. Here the $\chi^2$ statistic $\sum z^2_x$ is 5.105. on 8 observations. Comparing to $\chi^2_5$, we obtain a $p$-value 0.403. This does not present significant evidence of a bad fit.

### 3. $q(i) = d(i)/n$, the proportion of devices which broke down due to failure of component $i$. This is the maximum likelihood estimator of the underlying probability of breakdown due to failure of component $i$.

The columns $\hat{\mu}(i)$ give the corresponding maximum likelihood of the rate of failure of component $i$ (under the assumption of constant rates of failure and independence of failure times of the two components). We have

$$q^{(1)} = \left[1 - \exp(-\mu^{(1)} - \mu^{(2)})\right] \frac{\mu^{(1)}}{\mu^{(1)} + \mu^{(2)}}$$

$$q^{(2)} = \left[1 - \exp(-\mu^{(1)} - \mu^{(2)})\right] \frac{\mu^{(2)}}{\mu^{(1)} + \mu^{(2)}}$$

which leads to

$$\hat{\mu}^{(1)} = -\log \left(1 - q^{(1)} - q^{(2)}\right) \frac{q^{(1)}}{q^{(1)} + q^{(2)}}$$

$$\hat{\mu}^{(2)} = -\log \left(1 - q^{(1)} - q^{(2)}\right) \frac{q^{(2)}}{q^{(1)} + q^{(2)}}$$

The table is completed by

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>$d^{(1)}$</th>
<th>$d^{(2)}$</th>
<th>$q^{(1)}$</th>
<th>$q^{(2)}$</th>
<th>$\hat{\mu}^{(1)}$</th>
<th>$\hat{\mu}^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (dry)</td>
<td>200</td>
<td>18</td>
<td>34</td>
<td>0.090</td>
<td>0.170</td>
<td>0.104</td>
<td>0.197</td>
</tr>
<tr>
<td>B (humid)</td>
<td>500</td>
<td>31</td>
<td>138</td>
<td>0.062</td>
<td>0.276</td>
<td>0.076</td>
<td>0.337</td>
</tr>
<tr>
<td>C (normal)</td>
<td>200</td>
<td>18</td>
<td>42</td>
<td>0.090</td>
<td>0.210</td>
<td>0.107</td>
<td>0.250</td>
</tr>
</tbody>
</table>

The number of failures of component 1 in conditions A and C is the same, leading to $q_A^{(1)} = q_C^{(1)}$. However, the number of failures of component 2 is larger in C than in A. This suggests that the total exposure to risk in C is typically smaller than that in A, since the “population” decreases by more over the course of the observation time. So the same number of failures of component 1 corresponds to a higher estimated risk of failure in C than in A, and $\hat{\mu}_C^{(1)} > \hat{\mu}_A^{(1)}$.

To test the hypothesis that the failures rates do not depend on environmental conditions, it is simplest to work directly with the $q(i)$. This corresponds to a test of independence of rows and columns in a $3 \times 3$ contingency table. In each condition, the devices are classified as failure type 1, failure type 2, or no failure, giving the table of “observed values”:
For the null hypothesis of independence, we estimate column probabilities by 
\[ \hat{q}^{(1)} = \frac{18 + 31 + 18}{200 + 500 + 200} = 0.0744, \]
and
\[ \hat{q}^{(2)} = \frac{34 + 138 + 42}{200 + 500 + 200} = 0.2378, \]
leading to the table of “expected values”:

<table>
<thead>
<tr>
<th></th>
<th>( d^{(1)} )</th>
<th>( d^{(2)} )</th>
<th>( n - d^{(1)} - d^{(2)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>14.89</td>
<td>47.56</td>
<td>137.56</td>
</tr>
<tr>
<td>B</td>
<td>37.22</td>
<td>118.89</td>
<td>343.89</td>
</tr>
<tr>
<td>C</td>
<td>14.89</td>
<td>47.56</td>
<td>137.56</td>
</tr>
</tbody>
</table>

We can find
\[ X^2 = -2 \log \lambda \]
where \( \lambda \) is the likelihood ratio between the null hypothesis and a general alternative hypothesis, either precisely by
\[ X^2 = 2 \sum O_j (\log(O_j) - \log(E_j)) \]
\[ = 11.52 \]
or using the time-honoured approximation by
\[ X^2 = \sum \frac{(O_j - E_j)^2}{E_j} \]
\[ = 11.24. \]

The number of degrees of freedom under the null hypothesis is 2, and the number under the alternative hypothesis is 6. The difference is 4, so we compare \( X^2 \) to the \( \chi^2(4) \) distribution. This gives \( p \)-values of 0.023 using the first value or 0.024 using the second. Hence we have good evidence to reject the hypothesis that the failure rates are independent of environmental conditions.

4. (a) In lectures, we derived the likelihood of the path of a general continuous time Markov chain with generator \( Q \):
\[ \prod_{i} \prod_{j \neq i} \left( q_{ij}^{N_{ij}} \exp(-q_{ij}E_i) \right) \]
where \( N_{ij} \) is the number of transitions from \( i \) to \( j \), and \( E_i \) is the total time in state \( i \).

In this case the only transitions which are possible are from \( i \) to \( i + 1 \) for any \( i \) (with rate \( \lambda \)) or from \( i \) to \( i - 1 \) for \( i \geq 1 \) (with rate \( \mu \)). So the likelihood becomes
\[ \lambda^{N_+} \exp(-\lambda t) \mu^{N_-} \exp(-\mu E_-), \]
where \( N_+ = \sum_{i=0}^{\infty} N_{i,i+1} \) is the total number of arrivals observed, \( N_- = \sum_{i=1}^{\infty} N_{i,i-1} \) is the total number of departures observed, and \( E_- = \sum_{i=1}^{\infty} E_i = 1 - E_0 \) is the total “busy time” of the queue, i.e. the total amount of time when at least one customer is present.

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

Note that in fact \( N_+ \) simply has Poisson(\( \lambda t \)) distribution. For \( N_- \), the asymptotic theory gives an approximate variance.
We can consider approximate confidence intervals via the Fisher information matrix. Since the likelihood factorises, the matrix is diagonal. We obtain

\[ I_{\lambda\lambda} = \frac{E N_+}{\lambda^2} = \frac{t}{\lambda} \approx \frac{t^2}{N_+}, \]
\[ I_{\mu\mu} = \frac{E N_-}{\mu^2} \approx \frac{N_-}{\mu^2} = \frac{E^2_-}{N_-}. \]

The asymptotic joint distribution of \((\hat{\lambda}, \hat{\mu})\) is \(N((\lambda, \mu), I^{-1})\). So \(\hat{\lambda} - \lambda\) and \(\hat{\mu} - \mu\) are asymptotically independent, with distributions that we can estimate by \(N(0, \hat{\sigma}_\lambda^2)\) and \(N(0, \hat{\sigma}_\mu^2)\) respectively, where \(\hat{\sigma}_\lambda^2 = N_+/t^2\) and \(\hat{\sigma}_\mu^2 = N_-/E_-^2\).

So approximate \((1 - \alpha)\) confidence intervals for \(\lambda\) and \(\mu\) are given by

\[
(\hat{\lambda} - z_{\alpha/2}\hat{\sigma}_\lambda, \hat{\lambda} + z_{\alpha/2}\hat{\sigma}_\lambda),
(\hat{\mu} - z_{\alpha/2}\hat{\sigma}_\mu, \hat{\mu} + z_{\alpha/2}\hat{\sigma}_\mu),
\]

where \(z_{\alpha/2}\) is the \((1 - \alpha/2)\)-quantile of the standard normal distribution.

Our approximation for the joint distribution of \((\hat{\lambda}, \hat{\mu})\) yields approximate joint density proportional to

\[
\exp \left\{-\frac{1}{2} \left( \frac{(\lambda - \hat{\lambda})^2}{1/I_{\lambda\lambda}} + \frac{(\mu - \hat{\mu})^2}{1/I_{\mu\mu}} \right) \right\} \approx \exp \left\{-\frac{1}{2} \left( \frac{(\lambda - \hat{\lambda})^2}{\hat{\sigma}_\lambda^2} + \frac{(\mu - \hat{\mu})^2}{\hat{\sigma}_\mu^2} \right) \right\}.
\]

The minimal-area regions with given probability correspond to regions where the density is above some threshold. These are ellipsoids

\[ A_{\lambda,\mu} = \left\{ (x, y) : \frac{(x - \mu)^2}{\hat{\sigma}_\lambda^2} + \frac{(y - \mu)^2}{\hat{\sigma}_\mu^2} \leq c \right\}. \]

\((\hat{\lambda} - \lambda)^2/\hat{\sigma}_\lambda^2 + (\hat{\mu} - \mu)^2/\hat{\sigma}_\mu^2\) has approximately \(\chi^2(2)\) distribution, so for a \((1 - \alpha)\) confidence region, we should take \(c\) to be the \((1 - \alpha)\)-quantile of \(\chi^2(2)\).

In fact, \(\chi^2(2)\) is the same distribution as \(\text{Exp}(1/2)\), so we can compute \(c = -2\log(\alpha)\). Now set

\[ E_{\hat{\lambda},\hat{\mu}} = \left\{ (\lambda, \mu) : \frac{(\lambda - \hat{\lambda})^2}{\hat{\sigma}_\lambda^2} + \frac{(\mu - \hat{\mu})^2}{\hat{\sigma}_\mu^2} \leq c \right\}. \]

Then \(P((\lambda, \mu) \in E_{\hat{\lambda},\hat{\mu}}) \approx P((\hat{\lambda}, \hat{\mu}) \in A_{\lambda,\mu}) \approx 1 - \alpha\), and \(E_{\hat{\lambda},\hat{\mu}}\) is an approximate minimum-area \((1 - \alpha)100\%-\)confidence region for \((\lambda, \mu)\).

(b) \(i\). The state space is \(S = \{0, \ldots, m\}\). We can model using a chain with tridiagonal \(Q\)-matrix, with entries \(q_{i,i+1} = \lambda_i\) and \(q_{i+1,i} = \mu, i = 0, \ldots, m - 1\).

\(ii\). The likelihood function now becomes

\[
\prod_{i=0}^{m-1} \lambda_i^{N_{i,i+1}} \exp\{-\lambda_i E_i\} \mu^{N_-} \exp\{-\mu E_-\}.
\]

Therefore, the MLEs 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 covariance matrix. In particular, we have asymptotically

\[
\frac{\hat{\lambda}_0 - \hat{\lambda}_1}{\sqrt{\text{Var}(\hat{\lambda}_0) + \text{Var}(\hat{\lambda}_1)}} \approx N(0, 1) \quad \text{and} \quad \frac{(\hat{\lambda}_0 - \hat{\lambda}_1)^2}{\text{Var}(\hat{\lambda}_0) + \text{Var}(\hat{\lambda}_1)} \approx \chi^2_1.
\]

iv. Approximating the Fisher information matrix as we have done many times before, we get estimates

\[
\text{Var}(\hat{\lambda}_0) \approx N_0 / E_0 := \hat{\sigma}_0^2 \quad \text{and} \quad \text{Var}(\hat{\lambda}_1) \approx N_1 / E_1 := \hat{\sigma}_1^2.
\]

So using the previous part we can argue for the statistic

\[
\frac{\hat{\lambda}_0 - \hat{\lambda}_1}{\sqrt{\hat{\sigma}_0^2 + \hat{\sigma}_1^2}}
\]

which we would compare to the standard normal distribution for a test of the null hypothesis that \( \lambda_0 = \lambda_1 \).

5. (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

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

This can be maximised parameter by parameter. To maximise in \( \sigma \), 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 \( \ell \) is maximized for \( \hat{\sigma} = N_{HS} / E_H \). For \( N_{HS} = 150 \) and \( E_H = 6250 \) this is \( \hat{\sigma} = 150 / 6250 = 0.024 \).

(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

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

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

In particular, \( \sqrt{N_{HS}}/E_H = \sqrt{150}/6250 = 0.00196 \) is an estimate of the standard deviation of \( \hat{\sigma} \).

Then \( [\hat{\sigma} - 1.96\sqrt{N_{HS}}/E_H, \hat{\sigma} + 1.96\sqrt{N_{HS}}/E_H] = [0.0202, 0.0278] \) is an approximate 95% confidence interval.

(c) We assume that each policyholder enters the policy in state \( H \).

The holding time in state \( H \) has mean \( 1/(\sigma + \delta) \).

This is followed by a transition to state \( S \) with probability \( \sigma/(\sigma + \delta) \) (otherwise, a transition to the absorbing state \( D \)).

The mean holding time in state \( S \) is \( 1/(\rho + \gamma) \).
From $S$, we either transition to the absorbing state $D$, or go back to $H$. We could regard the transition back to $H$ as a new policyholder arriving, so overall it’s enough to balance the contributions from a single holding time in $H$ with the benefits from the holding time in $S$ which follows it, if any. In this way a balance between contributions and benefits corresponds to

$$\frac{1}{\sigma + \delta} C = \frac{\sigma}{\sigma + \delta} \frac{1}{\rho + \gamma} B,$$

so that $C/B = \sigma/\rho$. Of course in practice, the company would want to set a higher ratio in order to cover risk – we also ignore issues such as interest, expenses, etc.

For a large sample, $\hat{\sigma}$ and $\hat{\rho} + \hat{\gamma}$ should be approximately independent with normal distributions which, as above, we can estimate by $N(\sigma, s_1^2)$ and $N(\rho + \gamma, s_2^2)$ where $s_1^2 = N_{HS}/E_H^2$ and $s_2^2 = (N_{SH} + N_{SD})/E_S^2)$. For a large sample these variances are small and we can proceed by

$$\frac{\hat{\sigma}}{\hat{\rho} + \hat{\gamma}} \approx \left(\sigma + s_1 Z_1\right) \left(\rho + \gamma + s_2 Z_2\right)^{-1} \approx \frac{\sigma}{\rho + \gamma} + \frac{s_1}{\rho + \gamma} Z_1 - \frac{s_2 \sigma}{(\rho + \gamma)^2} Z_2$$

$$= \frac{\sigma}{\rho + \gamma} + sZ$$

where

$$s = \sqrt{\left(\frac{s_1}{\rho + \gamma}\right)^2 + \left(\frac{s_2 \sigma}{(\rho + \gamma)^2}\right)^2} \approx \sqrt{\left(\frac{s_1}{\hat{\rho} + \hat{\gamma}}\right)^2 + \left(\frac{s_2 \hat{\sigma}}{(\hat{\rho} + \hat{\gamma})^2}\right)^2}.$$

This leads to an approximate $(1 - \alpha)$-confidence interval of $\left(\frac{\hat{\sigma}}{\hat{\rho} + \hat{\gamma}} \pm s z_{1-\alpha}\right)$ where $z_{1-\alpha}$ is the $(1 - \alpha)$-quantile of $N(0, 1)$.

In this case for a 95% confidence interval, we would obtain $z = 1.96$, $s = 0.0118$, $\hat{\sigma} = 0.024$, $\hat{\rho} + \hat{\gamma} = 0.171$ giving the interval $(0.117, 0.163)$. 