Actuarial Models: solutions example sheet 8

Answer to 1

(a) The table indicating the values of $t_i$, $d_i$ and $r_i$ are given in Table 1. We can then compute the Kaplan-Meier estimator at the times $t_i$ using $\hat{S}(t_0) = \hat{S}(0) = 1$ and the recursion,

$$\hat{S}(t_i) = \hat{S}(t_{i-1}) \left( 1 - \frac{d_i}{r_i} \right).$$

This gives the fifth column of Table 1. Further, for $i = 1, 2, 3, 4$ we have that for $t \in [t_i, t_{i+1})$, $\hat{S}(t) = \hat{S}(t_i)$ and $\hat{S}(t) = 0$ for $t \geq t_5 = 9.5$. The plot of the Kaplan-Meier estimator is given in Figure 1.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$t_i$</th>
<th>$d_i$</th>
<th>$r_i$</th>
<th>$\hat{S}(t_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>6.5</td>
<td>2</td>
<td>10</td>
<td>0.800</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>0.686</td>
</tr>
<tr>
<td>3</td>
<td>8.5</td>
<td>1</td>
<td>5</td>
<td>0.549</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>3</td>
<td>4</td>
<td>0.137</td>
</tr>
<tr>
<td>5</td>
<td>9.5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1: Table corresponding to Exercise 1.

(b) We know from the lecture notes that $E[T] = \int_0^\infty S(t)dt$. Hence we can estimate $E[T]$ using the Kaplan-Meier estimator via $\int_0^\infty \hat{S}(t)dt$, i.e. we need to compute the surface
under the graph of \( \hat{S}(t) \). This is given by
\[
\int_0^\infty \hat{S}(t) \, dt = \sum_{i=0}^{4} \hat{S}(t_i)(t_{i+1} - t_i) = 8.27.
\]

(c) We need to estimate \( \mathbb{E}[T_7] \) (which is equal to \( \mathbb{E}[T - 7 | T > 7] \)). From the lecture notes, we have that
\[
\mathbb{E}[T_7] = \frac{1}{S(7)} \int_0^\infty S(7 + u) \, du = \frac{1}{S(7)} \int_7^\infty S(u) \, du.
\]
Hence, using the Kaplan-Meier estimator, we can estimate expected residual life time at time 7 by
\[
\frac{1}{\hat{S}(7)} \int_7^\infty \hat{S}(u) \, du = \frac{1}{\hat{S}(7)} \left( \hat{S}(7) * (8.5 - 7) + \hat{S}(8.5) * (9 - 8.5) + \hat{S}(9) * (9.5 - 9) \right)
= 2.0.
\]

(d) The curtate expected (residual) burn time at time 7 is, see lecture notes, given by
\[
\frac{1}{S(7)} \sum_{j=1}^\infty S(7 + j),
\]
which can be estimated, using the Kaplan-Meier estimator, by
\[
\frac{1}{S(7)} [\hat{S}(8) + \hat{S}(9)] = 1.2.
\]

**Answer to 2**

We now want to estimate the survival function \( t \mapsto S_{60.0}(t), \, t \geq 0 \) and the cumulative hazard function \( t \mapsto A_{60.0}(t), \, t \geq 0 \) of the residual life time at age 60.0. The observed genuine (residual) failure times (after age 60.0) are given by the observed ages at which a person died minus 60.0 and form the first column of Table 2. The values of \( d_i \), the observed number of deaths at the genuine failure times are easy to get, but care has to be taken with determining the values of \( r_i \), the number of people at risk just before the genuine failure times. Namely, a person only counts to the number at risk just before \( t_i \) if he/she was being observed (and being observed to be alive) just before \( t_i \).

Considering, for example, the first genuine residual failure time \( t_1 = 9.7 \), we see from the data that all lives were alive at age 69.7. However, only 6 out of 10 were being observed at the age of 69.7, the other 4 started to be observed at a later age. This means that \( r_1 \) is equal to 6 and not 10, because we should not use a person who started being observed at age \( x \) for estimating the hazard of dying at age \( y < x \). Since if this person would have died at age \( y < x \), he/she would not have been part of the study. To take this argument to the extreme, if you want to estimate the survival probability of surviving until at least 40 years, you are
not going to use a study with people who are all over 50 years old at the start, but you are
going to use persons who can actually die before they turn 40.

Once the \( r_i \)'s are determined (note that they can now increase over time as well), one
can then compute the Kaplan-Meier and Nelson-Aalen estimators using the same formulas
as in the notes, see Table 2. Note further that \( \hat{S}_{60.0}(t) = \hat{S}_{60.0}(t_i), \hat{A}_{60.0}(t) = \hat{A}_{60.0}(t_i) \) for
\( t \in [t_i, t_{i+1}) \), \( i = 1, \ldots, 5 \) and \( \hat{S}_{60.0}(t) = 1, \hat{A}_{60.0}(t) = 0 \) for \( t \in [0, t_1) \).

<table>
<thead>
<tr>
<th>( t_i )</th>
<th>( d_i )</th>
<th>( r_i )</th>
<th>( \hat{S}_{60.0}(t_i) )</th>
<th>( \hat{A}_{60.0}(t_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.7</td>
<td>1</td>
<td>6</td>
<td>0.833</td>
<td>0.167</td>
</tr>
<tr>
<td>12.2</td>
<td>1</td>
<td>5</td>
<td>0.667</td>
<td>0.367</td>
</tr>
<tr>
<td>13.2</td>
<td>1</td>
<td>4</td>
<td>0.500</td>
<td>0.617</td>
</tr>
<tr>
<td>17.8</td>
<td>1</td>
<td>5</td>
<td>0.400</td>
<td>0.817</td>
</tr>
<tr>
<td>19.6</td>
<td>2</td>
<td>4</td>
<td>0.200</td>
<td>1.317</td>
</tr>
</tbody>
</table>

Table 2: Table corresponding to Exercise 2.

Answer to 3

(a) We first need to translate the information in the table into failure times. This can
be easily done as the failure time corresponding to a patient is just the difference
between the time observation ended and the time of operation. Hence we have the
following set of failure time where \( * \) indicates a censored value:

\[ 30, 30, 30*, 35, 40, 40, 50, 68, 71, 100*, 116*, 120*. \]

We can now easily deduce the values for \( t_i \) (ith ordered observed genuine failure time),
\( d_i \) (the observed number of failures at \( t_i \)) and \( r_i \) (observed number of patients at risk
of failing just before \( t_i \)). The values are given in Table 3. Then we can compute the
Kaplan-Meier and the Nelson-Aalen estimator \( \hat{S}(t) \) and \( \hat{A}(t) \) at the failure times \( t_i \)
using \( \hat{S}(t_1) = (1 - d_1/r_1), \hat{A}(t_1) = d_1/r_1 \) and the recursion,

\[
\hat{S}(t_i) = \hat{S}(t_{i-1})(1 - d_i/r_i), \quad i = 2, 3, \ldots \\
\hat{A}(t_i) = \hat{A}(t_{i-1}) + d_i/r_i, \quad i = 2, 3, \ldots .
\]

This results are given in the last column of Table 3. Note that the value of \( \hat{S}(t) \) and
\( \hat{A}(t) \) for \( t \in [t_i, t_{i+1}) \) is given by \( \hat{S}(t) = \hat{S}(t_i) \) and \( \hat{A}(t) = \hat{A}(t_i) \).

(b) We want to find an estimate for \( \Pr(T \geq 70) \). Since \( \hat{S}(t) \) is an estimate for \( \Pr(T > t) \)
and there are no observed failures at \( t = 70 \), the required estimate is given by \( \hat{S}(70) =
\hat{S}(68) = 0.370 \).
<table>
<thead>
<tr>
<th></th>
<th>$t_i$</th>
<th>$d_i$</th>
<th>$r_i$</th>
<th>$\hat{S}(t_i)$</th>
<th>$\hat{A}(t_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>2</td>
<td>12</td>
<td>0.833</td>
<td>0.167</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>1</td>
<td>9</td>
<td>0.741</td>
<td>0.278</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>2</td>
<td>8</td>
<td>0.556</td>
<td>0.528</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>1</td>
<td>6</td>
<td>0.463</td>
<td>0.694</td>
</tr>
<tr>
<td>5</td>
<td>68</td>
<td>1</td>
<td>5</td>
<td>0.370</td>
<td>0.894</td>
</tr>
<tr>
<td>6</td>
<td>71</td>
<td>1</td>
<td>4</td>
<td>0.278</td>
<td>1.144</td>
</tr>
</tbody>
</table>

Table 3: Table corresponding to Exercise 3(a).