MATH3085/6143 Survival Models – Worksheet 2 Solutions

1. i) For $0 < t \le t'$,

$$\begin{aligned} S_T(t) &= \exp\left(-\int_0^t h_T(s)\mathrm{d}s\right) \\ &= \exp\left(-\int_0^t \lambda_1\right) \\ &= \exp\left(-[\lambda_1 s]_0^t\right) \\ &= \exp(-\lambda_1 t). \end{aligned}$$

For $t' < t < \infty$,

$$\begin{aligned} S_T(t) &= \exp\left(-\int_0^t h_T(s)\mathrm{d}s\right) \\ &= \exp\left[-\left(\int_0^{t'} h_T(s)\mathrm{d}s + \int_{t'}^t h_T(s)\mathrm{d}s\right)\right] \\ &= \exp\left[-\left(\int_0^{t'} \lambda_1\mathrm{d}s + \int_{t'}^t \lambda_2\mathrm{d}s\right)\right] \\ &= \exp\left[-\left(\lambda_1 t' + \lambda_2(t-t')\right)\right]. \end{aligned}$$

Therefore, we have that

$$S_T(t) = \begin{cases} \exp(-\lambda_1 t) & , 0 < t \le t' \\ \exp\left[-\left(\lambda_1 t' + \lambda_2(t-t')\right)\right] & , t' < t < \infty \end{cases}.$$

ii)

$$\begin{aligned} f_T(t) &= S_T(t)h_T(t) \\ &= \begin{cases} \lambda_1\exp(-\lambda_1 t) & , 0 < t \le t' \\ \lambda_2\exp\left[-\left(\lambda_1 t' + \lambda_2(t-t')\right)\right] & , t' < t < \infty \end{cases}. \end{aligned}$$

iii) The likelihood function can be written as,

$$\begin{aligned} L(\lambda_1,\lambda_2) &= \prod_{i:d_i=1} f_T(t_i) \prod_{i:d_i=0} S_T(t_i) \\ &= \prod_{i:d_i=1,t_i<t'} f_T(t_i) \prod_{i:d_i=1,t_i\ge t'} f_T(t_i) \prod_{i:d_i=0,t<t'} S_T(t_i) \prod_{i:d_i=0,t\ge t'} S_T(t_i) \\ &= \prod_{i:d_i=1,t_i<t'} \left[\lambda_1\exp(-\lambda_1 t_i)\right] \prod_{i:d_i=1,t_i\ge t'} \left[\lambda_2\exp(-\lambda_1 t' - \lambda_2(t_i-t'))\right] \\ &\quad\times \prod_{i:d_i=0,t_i<t'} \left[\exp(-\lambda_1 t_i)\right] \prod_{i:d_i=0,t_i\ge t'} \left[\exp(-\lambda_1 t' - \lambda_2(t_i-t'))\right] \\ &= \lambda_1^{\sum_{i:d_i=1,t_i<t'}1} \lambda_2^{\sum_{i:d_i=1,t_i\ge t'}1} \exp\left[-\lambda_1\sum_{i:t_i<t'} t_i\right] \exp\left[-\sum_{i:t_i\ge t'}\left\{\lambda_1 t' + \lambda_2(t_i-t')\right\}\right] \\ &= \lambda_1^{m_1}\lambda_2^{m_2}\exp\left[-\lambda_1\left(n_2 t' + \sum_{i:t_i<t'} t_i\right)\right]\exp\left[-\lambda_2\sum_{i:t_i\ge t'}(t_i-t')\right] \text{ (shown).} \end{aligned}$$

iv) The log-likelihood is given by

$$l = \log L = m_1\log\lambda_1 + m_2\log\lambda_2 - \lambda\left[n_2 t' + \sum_{i:t_i<t'} t_i\right] - \lambda_2\sum_{i:t_i\ge t'}(t_i-t').$$

The 1^{st} order partial derivatives are then

$$\frac{\partial l}{\partial \lambda_1} = \frac{m_1}{\lambda_1} - \left(n_2 t' + \sum_{i:t_i<t'} t_i \right)$$

$$\frac{\partial l}{\partial \lambda_2} = \frac{m_2}{\lambda_2} - \sum_{i:t_i>t'} (t_i - t').$$

Now set $\frac{\partial l}{\lambda_1} = \frac{\partial l}{\lambda_2} = 0$ and solve for the MLEs $\hat{\lambda}_1$ and $\hat{\lambda}_2$ easily:

$$\begin{cases} \hat{\lambda}_1 = \frac{m_1}{n_2 t' + \sum_{i:t_i<t'} t_i} \\ \hat{\lambda}_2 = \frac{m_2}{\sum_{i:t_i \geq t'} (t_i - t')} \end{cases}.$$

2. i)

i	$t'_i \leq t < t'_{i+1}$	r_i	c_i	d'_i	$\frac{d'_i}{r_i}$	$1 - \frac{d'_i}{r_i}$	$\hat{S}(t) = \prod_{j=1}^{i} \left(1 - \frac{d'_j}{r_j}\right)$
0	$0 \leq t < 97$	12	0				1
1	$97 \leq t < 120$	12	1	2	$\frac{2}{12}$	$\frac{10}{12}$	$1 \times \frac{10}{12} = \frac{5}{6}$
2	$120 \leq t < 141$	9	0	3	$\frac{3}{9}$	$\frac{6}{9}$	$1 \times \frac{10}{12} \times \frac{6}{9} = \frac{5}{9}$
3	$141 \leq t < 150$	6	0	2	$\frac{2}{6}$	$\frac{4}{6}$	$1 \times \frac{10}{12} \times \frac{6}{9} \times \frac{4}{6} = \frac{10}{27}$
4	$150 \leq t < \infty$	4	3	1	$\frac{1}{4}$	$\frac{3}{4}$	$1 \times \frac{10}{12} \times \frac{6}{9} \times \frac{4}{6} \times \frac{3}{4} = \frac{5}{18}$

ii) Recall that the squared standard error of Kaplan-Meier estimate is given by

$$s.e.[\hat{S}(t)]^2 = \hat{S}(t)^2 \sum_{j=1}^{i} \frac{d'_j}{r_j(r_j - d'_j)}.$$

The following table can then be constructed:

i	$\frac{d'_i}{r_i(r_i-d'_i)}$	$s.e.[\hat{S}(t)]^2$	95% confidence intervals=$\hat{S}(t) \pm 1.96 \times s.e.[\hat{S}(t)]$
0	$-$	$-$	$-$
1	$\frac{2}{12(12-2)}$	$\left(\frac{5}{6}\right)^2 \left(\frac{1}{60}\right) = \frac{5}{432}$	$\frac{5}{6} \pm 1.96\sqrt{\frac{5}{432}} = (0.6225, 1.0442)$
2	$\frac{3}{9(9-3)}$	$\left(\frac{5}{9}\right)^2 \left(\frac{1}{60} + \frac{1}{18}\right) = \frac{65}{2916}$	$\frac{5}{9} \pm 1.96\sqrt{\frac{65}{2916}} = (0.2629, 0.8482)$
3	$\frac{2}{6(6-2)}$	$\left(\frac{10}{27}\right)^2 \left(\frac{1}{60} + \frac{1}{18} + \frac{1}{12}\right) = \frac{140}{6561}$	$\frac{10}{27} \pm 1.96\sqrt{\frac{140}{6561}} = (0.0841, 0.6567)$
4	$\frac{1}{4(4-1)}$	$\left(\frac{5}{18}\right)^2 \left(\frac{1}{60} + \frac{1}{18} + \frac{1}{12} + \frac{1}{12}\right) = \frac{215}{11664}$	$\frac{5}{18} \pm 1.96\sqrt{\frac{215}{11664}} = (0.0117, 0.5439)$

iii) Nelson-Aalen estimate relies on the quantity cumulative hazard function, $\hat{H}(t)$ which can be computed by summing the discrete hazard estimators. In particular,

i	$\frac{d'_i}{r_i}$	$\hat{H}(t) = \sum_{j=1}^{i} \frac{d'_i}{r_i}$
0	$-$	$-$
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{3}$	$\frac{1}{6} + \frac{1}{3} = \frac{1}{2}$
3	$\frac{1}{3}$	$\frac{1}{6} + \frac{1}{3} + \frac{1}{3} = \frac{5}{6}$
4	$\frac{1}{4}$	$\frac{1}{6} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} = \frac{13}{12}$

iv)

$t'_i \le t < t'_{i+1}$	$\hat{S}_{NA}(t) = \exp(-\hat{H}(t))$	$\hat{S}_{KM}(t)$
$0 \le t < 97$	1.0000	1.0000
$97 \le t < 120$	$\exp(-\frac{1}{6}) = 0.8465$	0.8333
$120 \le t < 141$	$\exp(-\frac{1}{2}) = 0.6065$	0.5556
$141 \le t < 150$	$\exp(-\frac{5}{6}) = 0.4346$	0.3704
$150 \le t < \infty$	$\exp(-\frac{13}{12}) = 0.3385$	0.2778

As seen in the lecture notes, Kaplan-Meier and Nelson-Aalen estimators of $S(t)$ will be close when $\frac{d'_j}{r_j}$ is small due to the Taylor's approximation of exponential function:

$$\exp\left(-\frac{d'_j}{r_j}\right) = 1 - \frac{d'_j}{r_j} - \frac{1}{2!}\left(\frac{d'_j}{r_j}\right)^2 - \cdots \approx 1 - \frac{d'_j}{r_j}.$$

Thus we have that

$$\begin{aligned} \hat{S}_{NA}(t) &= \exp\left(-\sum_{j=1}^{i} \frac{d'_j}{r_j}\right) \\ &= \prod_{j=1}^{i} \exp\left(-\frac{d'_j}{r_j}\right) \\ &\approx \prod_{j=1}^{i} \left(1 - \frac{d'_j}{r_j}\right) \\ &= \hat{S}_{KM}(t). \end{aligned}$$

By comparing the Kaplan-Meier and Nelson-Aalen estimates, it is found that the estimated $\hat{S}(t)$ is very close only for $97 \le t < 120$ because $\frac{d'_1}{r_1} = \frac{1}{6}$ is fairly small. The estimates for other durations differ considerably.

3. i) Interval censoring

- No: we are counting in days and we know which day each event occurred. (or yes: as it is a generalisation of right censoring and so is present if right censoring is present.)

Right censoring

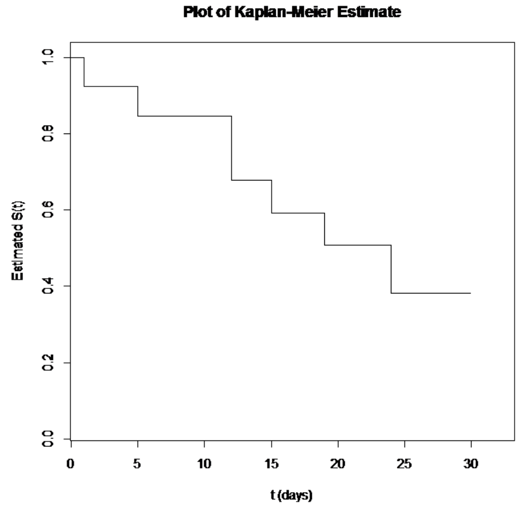
- Yes. The end of the course at day 30 cut short the investigation when not all candidates had qualified. The event that people left before day 30 without qualified is also a right censoring.

Informative censoring

- The end of the course at day 30 is clearly a non-informative censorring, as it is a prespecified time.
- Those who left during the 30 days will probably take longer to qualify than those who stayed (Or arguably more likely are bored to death), so this censoring is possible to be informative.

i	$t'_i \le t < t'_{i+1}$	r_i	c_i	d'_i	$\frac{d'_i}{r_i}$	$1 - \frac{d'_i}{r_i}$	$\hat{S}(t) = \prod_{j=1}^{i}\left(1 - \frac{d'_j}{r_j}\right)$
0	$0 \le t < 1$	13	0				1.000
1	$1 \le t < 5$	13	0	1	$\frac{1}{13}$	$\frac{12}{13}$	$\frac{12}{13} = 0.923$
2	$5 \le t < 12$	12	1	1	$\frac{1}{12}$	$\frac{11}{12}$	$1 \times \frac{12}{13} \times \frac{11}{12} = 0.846$
3	$12 \le t < 15$	10	0	2	$\frac{1}{5}$	$\frac{4}{5}$	$\cdots \times \frac{4}{5} = 0.677$
4	$15 \le t < 19$	8	0	1	$\frac{1}{8}$	$\frac{7}{8}$	$\cdots \times \frac{7}{8} = 0.592$
5	$19 \le t < 24$	7	2	1	$\frac{1}{7}$	$\frac{6}{7}$	$\cdots \times \frac{6}{7} = 0.508$
6	$24 \le t < 30(\infty)$	4	3	1	$\frac{1}{4}$	$\frac{3}{4}$	$\cdots \times \frac{3}{4} = 0.381$

ii)



iii)

4. (a) $S_T(0) = 1$, $S_T(t) \geq 0$ for all $t > 0$ and $S_T(t)$ is a non-increasing function of t. Hence $S_T(t)$ is a valid survival function.

(b)

$$h_T(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \log S_T(t) = \frac{\mathrm{d}}{\mathrm{d}t} \beta \log(1+t) = \frac{\beta}{1+t}$$

$$f_T(t) = h_T(t) S_T(t) = \frac{\beta}{(1+t)^{\beta+1}}$$

(c)

$$\begin{aligned} L(\beta) = \prod_{i:\text{failed}} f_T(t_i;\beta) \prod_{i:\text{censored}} S_T(t_i;\beta) &= \prod_{i=1}^{m} \frac{\beta}{(1+t_i)^{\beta+1}} \prod_{i=1}^{n-m} \frac{1}{(1+1000)^{\beta}} \\ &= \beta^m \frac{1}{1001^{(n-m)\beta}} \prod_{i=1}^{m} \frac{1}{(1+t_i)^{\beta+1}} \end{aligned}$$

Hence

$$\ell(\beta) = m \log \beta - (n-m)\beta \log 1001 - (\beta+1) \sum_{i=1}^{m} \log(1+t_i)$$

(d)

$$\frac{\partial}{\partial \beta} \ell(\beta) = \frac{m}{\beta} - (n-m) \log 1001 - \sum_{i=1}^{m} \log(1+t_i)$$

At $\hat{\beta}$,

$$\begin{aligned} \frac{\partial}{\partial \beta} \ell(\beta) = 0 \quad &\Rightarrow \quad \frac{m}{\hat{\beta}} = (n-m) \log 1001 + \sum_{i=1}^{m} \log(1+t_i) \\ &\Rightarrow \quad \hat{\beta} = \frac{m}{(n-m)\log 1001 + \sum_{i=1}^{m} \log(1+t_i)} \end{aligned}$$

For the data provided $\hat{\beta} = 3/(9 \log 1001 + \log 201 + 2 \log 801) = 0.0371$.

$$\frac{\partial^2}{\partial \beta^2} \ell(\beta) = -\frac{m}{\beta^2}$$

so

$$s.e.(\hat{\beta}) = \frac{\hat{\beta}}{m^{1/2}}$$

For the data provided $s.e.(\hat{\beta}) = 0.0214$.

(e) As the censoring occurs at a prespecified time (end of experiment) this is clearly noninformative.