STOCKHOLMS UNIVERSITET
DEPT. OF MATHEMATICS
DIV. OF MATHEMATICAL STATISTICS

MT7052
EXAM
October 25 2024

Sketch solutions: Inference and prediction for life and health processes 7.5 ECTS Credits

October 25 2024

Note: All references to "the book" refers to Aalen et al. (2008) Survival and Event History Analysis.

Problem 1

Let $A^*(t) := \int_0^t I(Y(s) > 0)\alpha(s)\mathrm{d}s$ and introduce

$$\widehat{A}(t) := \int_0^t \frac{I(Y(s) > 0)}{Y(s)}\mathrm{d}N(s). \tag{1}$$

a) - b) See the derivation in the book, see Ch. 3, and the discussion about how the stochastic integral can be replaced by a sum over the jump times.

c) By noting that $I(Y(s) > 0) \leq 1$ it directly follows that $A^*(t)$ results in under estimation.

(10p)

Problem 2

a) Let $Y(t)$ denote the number of animals at risk of dying. It then follows that $N_j(t)$ is governed by

$$\lambda_j(t) := Y(t)\alpha_j(t),\ j = 1, 2$$

and that

$$N(t) = N_1(t) + N_2(t)$$

is governed by

$$\lambda(t) = \lambda_1(t) + \lambda_2(t) = Y(t)(\alpha_1(t) + \alpha_2(t)).$$

For more details, see the book.

b) - c) From **a)** we have that

$$\widehat{A}_j(t) := \sum_{k:T_k \leq t} \frac{\Delta N_j(T_k)}{Y(T_k)},$$

where $\Delta N_j(T_k) = 1$ if it is death by cause j at jump time T_k, and $\Delta N_j(T_k) = 0$ otherwise. From this we see that we can write the cumulative intensity estimator neglecting cause of death according to

$$\widehat{A}(t) := \sum_{k:T_k \leq t} \frac{\Delta N_1(T_k) + \Delta N_2(T_k)}{Y(T_k)} = \widehat{A}_1(t) + \widehat{A}_2(t).$$

Finally, we have

T_k	$Y(T_k)$	$\Delta A_1(T_k)$	$\Delta A_2(T_k)$	$A_1(T_k)$	$A_2(T_k)$
0.451	15	0	1/15	0	1/15
0.463	13	1/13	0	1/13	1/15
0.510	10	0	1/10	1/13	1/15 + 1/10
0.562	9	0	1/9	1/13	1/15 + 1/10 + 1/9
0.602	7	1/7	0	1/13 + 1/7	1/15 + 1/10 + 1/9

(10p)

Problem 3

a) The definition of the risk set when the ith event occurs, $\mathcal{R}_i$, is the set of indices of the individuals that are at risk at the time point of the i:th observed event, see e.g. p. 135. From the data we see that there are three observed events at time points 1, 3 and 10 which gives us the following three risk sets $\mathcal{R}_1 = \{1, 2, 3, 4, 5\}, \mathcal{R}_1 = \{2, 3, 4, 5\}, \mathcal{R}_3 = \{5\}$.

b) Cox's partial likelihood is given by (4.7) on p. 135, that is

$$L(\beta) = \prod_{\forall T_j} \frac{r(\beta, x_j)}{\sum_{i \in \mathcal{R}_j} r(\beta, x_i)},$$

where we have used that $x_j(t) \equiv x_j$ and where x_j refers to the covariate vector of the individual that experiences the jth jump. For the given data we only have three contributions to the likelihood:

$$\frac{e^\beta}{e^\beta + 1 + e^\beta + 1 + e^\beta}, \frac{1}{1 + e^\beta + 1 + e^\beta}, \frac{e^\beta}{e^\beta},$$

which gives us that

$$\begin{aligned} L(\beta) &= \frac{e^\beta}{3e^\beta + 2} \cdot \frac{1}{2e^\beta + 2} \cdot 1 \\ &= \frac{e^\beta}{6e^{2\beta} + 10e^\beta + 4}. \end{aligned}$$

Thus

$$L'(\beta) = \frac{e^\beta}{(6e^{2\beta} + 10e^\beta + 4)^2}(4 - 6e^{2\beta}),$$

which gives us that

$$\hat{\beta} = \frac{1}{2}\log\frac{2}{3}.$$

c) From the problem formulation we have that

$$\chi^2_{LR} = 2(\log L(\hat{\beta}) - \log L(\beta_0)) \sim \chi^2(1)$$

and we are interested in testing $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ at 95% confidence level, i.e. we reject H_0 if

$$2(\log L(\hat{\beta}) - \log L(0)) > \chi^2(1)_{0.05} = 3.84.$$

First we have that

$$\log L(\beta) = \beta - \log(6e^{2\beta} + 10e^\beta + 4),$$

which gives us that

$$\begin{aligned} \log L(\hat{\beta}) &= \frac{1}{2}\log\frac{2}{3} - \log(8 + 10\sqrt{\frac{2}{3}}), \\ \log L(0) &= -\log(20), \end{aligned}$$

that is

$$\chi^2_{LR} \approx 2 \cdot 0.0102 < 3.84,$$

and we can not reject H_0 at the 95%-level.

(10p)

Problem 4

a) From the book we know that

$$S(t) := \mathrm{E}[\exp\{-\int_0^t \alpha(u \mid Z)\mathrm{d}u\}] = \mathcal{L}_Z(A(t)),$$

where $A(t) := \int_0^t \alpha(u)\mathrm{d}u$. The proof of

$$\mu(t) = -\frac{S'(t)}{S(t)}$$

can be found in the book, and by combining with the above we arrive at

$$\mu(t) = \alpha(t)\frac{-\mathcal{L}_Z'(A(t))}{\mathcal{L}_Z(A(t))},$$

which after simplifications result in

$$\mu(t) = \frac{\alpha(t)}{1+\delta A(t)}.$$

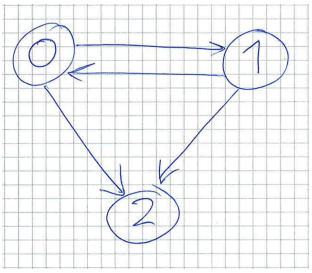
b) By setting $\alpha(t) = \exp\{\beta t\}$, it follows that $A(t) = (\exp\{\beta t\} - 1)/\beta$, and that

$$\begin{aligned}\mu(t) &= \frac{\exp\{\beta t\}}{1 + \frac{\delta}{\beta}(\exp\{\beta t\} - 1)} \\ &= \frac{1}{\frac{\delta}{\beta} + (1 - \frac{\delta}{\beta})\exp\{-\beta t\}} \to \frac{\beta}{\delta},\end{aligned}$$

as $t \to \infty$ if $\beta > 0$ and the limit tends to 0 if $\beta \leq 0$, where one can note that $\beta = 0$ yields $\alpha(t) = 1$. For the case with $\beta > 0$ the population hazard rate is approximately constant for large t, implying that if you have survived for a sufficiently long time your remaining life time will be approximately Exp-distributed.

(10p)

Problem 5



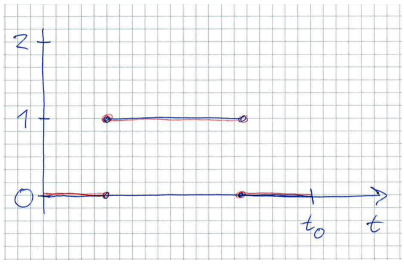


Figure 1: Left panel: Solution to Problem 5**a)**. Right panel: Solution to Problem 5**c)**

a) - **c)** See Fig. 1.

b) To start off, we have that

$$\begin{aligned}\alpha_0(t) &:= \alpha_{01}(t) + \alpha_{02}(t),\\ \alpha_1(t) &:= \alpha_{10}(t) + \alpha_{12}(t),\\ \alpha_2(t) &:= 0.\end{aligned}$$

This, together with the boundary conditions $P_{ij}(s,s) = \delta_{ij}$, $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, gives us that $P_{22}(s,t) = 1$ and $P_{2j}(s,t) = 0$ for $j \neq 2$.

By using the above and simplifying the KFEs gives us that

$$\begin{aligned}\frac{\partial}{\partial t}P_{00}(s,t) &= P_{01}(s,t)\alpha_{10}(t) - P_{00}(s,t)\alpha_0(t),\\ \frac{\partial}{\partial t}P_{01}(s,t) &= P_{00}(s,t)\alpha_{01}(t) - P_{01}(s,t)\alpha_1(t),\\ \frac{\partial}{\partial t}P_{02}(s,t) &= P_{00}(s,t)\alpha_{02}(t) + P_{01}(s,t)\alpha_{12}(t),\\ \frac{\partial}{\partial t}P_{10}(s,t) &= P_{11}(s,t)\alpha_{10}(t) - P_{10}(s,t)\alpha_0(t),\\ \frac{\partial}{\partial t}P_{11}(s,t) &= P_{10}(s,t)\alpha_{01}(t) - P_{11}(s,t)\alpha_1(t),\\ \frac{\partial}{\partial t}P_{12}(s,t) &= P_{10}(s,t)\alpha_{02}(t) + P_{11}(s,t)\alpha_{12}(t).\end{aligned}$$

(10p)