Exercise 4.7: The partial likelihood is a profile likelihood

Assume that $N_i(t)$ are counting processes with intensity processes $\lambda_i(t) = Y_i(t)\alpha_0(t)r(\beta, x_i)$.

Then the *total* log-likelihood can be written as

$$l(\alpha_0, \beta) = \sum_{i=1}^{n} \int_0^{\tau} \log(\lambda_i(t))dN_i(t) - \int_0^{\tau} \lambda_\bullet(t)dt$$

where $\lambda_\bullet(t) = \sum_{i=1}^{n} \lambda_i(t)$.

For instance for right censored data $(\tilde{T}_i, D_i)$ this expression becomes the more familiar

$$\sum_{i=1}^{N} \left[\log(\alpha_i(\tilde{T}_i))D_i - \int_0^{\tilde{T}_i} \alpha_i(s)ds \right] = \log(\prod_{i=1}^{n} f_i(\tilde{T}_i)^{D_i} S(\tilde{T}_i)^{1-D_i})$$

where $\alpha_i(t) = \alpha_0(t)r(\beta, x_i)$, $S_i(t) = \exp(-\int_0^t \alpha_i(s)ds)$ and $f_i(t) = \alpha_i(t)/S_i(t)$.

4.7a: But then

$$
\begin{aligned}
l(\alpha_0, \beta) &= \textstyle\sum_{i=1}^n \int_0^\tau [\log(\alpha_0(t)) + \log(Y_i(t)) + \log(r(\beta, X_i)] dN_i(t) \\
&\quad - \textstyle\int_0^\tau \alpha_0(t) \sum_{i=1}^n Y_i(t) r(\beta, x_i) dt \\
&= \textstyle\int_0^\tau \log(\alpha_0(t)) dN_\bullet(t) + \sum_{i=1}^n N_i(\tau) r(\beta, x_i) \\
&\quad - \textstyle\int_0^\tau S^{(0)}(\beta, t) \alpha_0(t) dt
\end{aligned}
$$

where $S^{(0)}(\beta, t) = \sum_{i=1}^n Y_i(t) r(\beta, x_i)$ and $N_\bullet(t) = \sum_{i=1}^n N_i(t)$.)

This since $r(\beta, x_i)$ does not depend on t and since $\log(Y_i(t)) dN_i(t) = 0$ (if $Y_i(t) = 1$ this holds true and if $Y_i(t) = 0$ then necessarily $dN_i(t) = 0$.

Clearly the term $\int_0^\tau \log(\alpha_0(t)) dN_\bullet(t)$ can be made arbitrarily large by letting $\alpha_0(t)$ becoming arbitrarily large for t such that $dN_\bullet(t) = 1$.

4.7a: But then

The expectation of the term $\int_0^\tau S^{(0)}(\beta,t)\alpha_0(t)dt$ equals the expectation of $N_\bullet(\tau)$ since $N_\bullet(t) - \int_0^t S^{(0)}(\beta,s)\alpha_0(s)ds = M_\bullet(t)$, i.e. a martingale, and so the this last term needs to be kept finite.

To maximize the log-likelihood over all possible models, including discrete distributions, we will obtain maximal values if we extend the model by substituting the integrals by sums over event times given as

$$\begin{aligned} l(A_0,\beta) = \; & \sum_{T_j} \log(\Delta A_0(T_j)) + \sum_{i=1}^n N_i(\tau) r(\beta, x_i) \\ & - \sum_{T_j} S^{(0)}(\beta, T_j)\Delta A_0(T_j) \end{aligned}$$

where the T_j are the observed event times and $A_0(t) = \sum_{T_j \le t} \Delta A_0(T_j)$.

4.7b: Then for given β

$l(A_0, \beta)$ is maximized for

$$\frac{\partial l(A_0, \beta)}{\partial \Delta A_0(T_k)} = 0,$$

thus for

$$\frac{1}{\Delta A_0(T_k)} - S^{(0)}(\beta, T_k) = 0$$

and so we get

$$\Delta \hat{A}_0(T_k, \beta) = \frac{1}{S^{(0)}(\beta, T_k)}$$

if there are no equal event times.

4.7c: Then inserting $\Delta\hat{A}_0(T_j,\beta)$ into $l(A_0,\beta)$ we get

$$\begin{aligned} l(\hat{A}_0,\beta) &= \sum_{T_j}\log\left(\frac{1}{S^{(0)}(\beta,T_j)}\right) + \sum_{i=1}^n N_i(\tau) r(\beta,x_i) \\ &\quad - \sum_{T_j} S^{(0)}(\beta,T_j)\frac{1}{S^{(0)}(\beta,T_j)} \\ &= -\sum_{T_j}\log(S^{(0)}(\beta,T_j)) + \sum_{T_j} r(\beta,x_j) - N_\bullet(\tau) \end{aligned}$$

Here $N_\bullet(\tau)$ is a constant and does not influence the maximum. Thus

$$\exp(l(\hat{A}_0,\beta)+N_\bullet(\tau)) = \prod_{T_j}\frac{r(\beta,x_j)}{S^{(0)}(\beta,T_j)} = \prod_{T_j}\frac{r(\beta,x_i)}{\sum_{i\in\mathcal{R}(T_j)} r(\beta,x_i)}$$

where $\mathcal{R}(T_j)$ is the risk set at time T_j. Thus

$$\exp(l(\hat{A}_0,\beta) + N_\bullet(\tau)) = L(\beta)$$

where $L(\beta)$ the partial likelihood.

4.7d: Since $N_\bullet(\tau)$ does not depend on β

we then have that maximizing profile likelihood $l(\hat{A}_0, \beta)$ with respect to β is the same as maximizing the the partial likelihood $L(\beta)$.