Session 3: Lecture notes exercise solutions

Exercise 3.1
Find the maximum likelihood estimates for λ and β

The likelihood is

$$L = \prod_{i=1}^{n} \left\{\lambda e^{\beta x_i}\right\}^{\delta_i} \exp\left\{-\lambda t_i e^{\beta x_i}\right\}.$$

Therefore the log likelihood is

$$l = \log(\lambda) \sum_{i=1}^{n} \delta_i + \beta \sum_{i=1}^{n} \delta_i x_i - \lambda \sum_{i=1}^{n} t_i e^{\beta x_i}.$$

We are focusing on a binary exposure, so x takes value 0 or 1. It is convenient to introduce some notation at this point. We let $\sum_{i=1}^{n} \delta_i = n_1$ (this is the number of events), and we let $\sum_{i=1}^{n} \delta_i x_i = n_{11}$ (this is the number of events among individuals with $x = 1$). We also let $\sum_{x_i=1} t_i = T_1$ (this is the sum of the survival and censoring time among individuals with $x = 1$), and $\sum_{x_i=0} t_i = T_0$ (this is the sum of the survival and censoring time among individuals with $x = 0$). The log likelihood can now be written as follows:

$$l = n_1 \log(\lambda) + n_{11}\beta - \lambda(T_0 + T_1 e^{\beta}).$$

To find the maximum likelihood estimates for λ and β we find the derivatives of the log likelihood and set equal to zero:

$$\frac{dl}{d\lambda} = \frac{n_1}{\lambda} - (T_0 + T_1 e^{\beta}) = 0,$$

$$\frac{dl}{d\beta} = n_{11} - \lambda T_1 e^{\beta} = 0.$$

Solving these simultaneous equations gives the MLEs:

$$\hat{\lambda} = \frac{(n_1 - n_{11})}{T_0}, \hat{\beta} = \log\left\{\frac{T_0 n_{11}}{T_1(n_1 - n_{11})}\right\}.$$

Exercise 3.2
An exponential model and a Weibull model were fitted to the breast cancer example data. The results from fitting models in Stata are shown below, where im is the indicator of IM positive status (taking value 0 if IM negative and 1 if IM positive). Interpret the results.

From the exponential model, the estimate of the baseline hazard (the hazard for a person with im=0) is $e^{-6.58} = 0.0014$ and the log hazard ratio (HR) estimate is $e^{1.116} = 3.05$, with 95% CI $(e^{0.267}, e^{1.964}) = (1.31, 7.13)$. The hazard in the IM positive group is just over three times that in the IM negative group. The 85% CI excludes 1 and the p-value is 0.010 - there is strong evidence against the null hypothesis that the hazard is the same in the two groups. The interpretation of the baseline hazard is that, for example, if we followed 10,000 women in the IM negative group for 1 year we would expect to see 14 deaths.

From the weibull model the estimate of λ (this is the 'scale parameter') is $e^{-7.758} = 0.00043$ and the estimate of κ (this is the 'shape parameter') is 1.226. The hazard ratio estimate is $e^{1.201} = 3.32$ with 95% CI is $(1.40, 7.88)$. The HR estimate is very similar in the exponential and Weibull models.

Exercise 3.3
What would expect a a plot of $\log H(t|x)$ against $\log t$ in the two exposure groups to look like if an exponential model is valid?

The exponential is a special case of the Weibull with $\kappa = 1$. Setting $\kappa = 1$ in equation 3.16 in the notes gives

$$\log H(t|x) = \log\{-\log S(t|x)\} = \log \lambda + \log t + \beta x$$

Under the exponential distribution, a plot of $\log H(t|x)$ against $\log t$ in the two exposure groups ($x = 0, 1$) will be straight lines with intercept $\log \lambda$ in the $x = 0$ group and intercept $\log \lambda + \beta$ in the $x = 1$ group, with both lines having a slope of 1.

Exercise 3.4
Figure 3.4 shows plots of Kaplan-Meier estimates of $\log H(t|x) = \log\{-\log S(t|x)\}$ against $\log t$. Use the plot and the output from fitting exponential and Weibull models to the breast cancer data (given above) to assess the fit of the Exponential and Weibull models.

Looking at the plot, the lines in the two groups are approximately parallel over time, and they are straight. This suggests that the Weibull model is appropriate for these data.

We can also assess the fit of the exponential and Weibull models by looking at the estimate of κ and it's confidence interval/p-value. A test of the null hypothesis that $\kappa = 1$ (or $\log \kappa = 0$) is a test of whether the Weibull provides a better fit than the exponential. here the p-value associated with $\log \kappa$ is 0.256, so there is no evidence here that the Weibull model provided a better fit to the data than the exponential. In other words, there is no evidence against the null hypothesis that the baseline hazard is constant over time.

We could alternatively perform a likelihood ratio test, which gives the same conclusion.