Session 1: Lecture notes exercise solutions

Exercise 1.1: Write the survivor function $S(t)$ in terms of the hazard $h(t)$ and then in terms of the cumulative hazard $H(t)$

Before doing the exercise, we go back and give the derivations leading to the relationships shown just above the exercise (pg 8 in the notes):

$$f(t) = \frac{d}{dt}F(t) = \frac{d}{dt}\{1 - S(t)\} = -\frac{d}{dt}S(t)$$

The survival function can be written:

$$S(t) = 1 - F(t) = 1 - \int_0^t f(u)du = \int_t^\infty f(u)du$$

The hazard function can be written:

$$\begin{aligned} h(t) &= \lim_{\delta \to 0} \frac{1}{\delta} \Pr(t \leq T < t + \delta | T > t) \\ &= \lim_{\delta \to 0} \frac{1}{\delta} \Pr(t \leq T < t + \delta, T > t) / \Pr(T > t) \qquad \text{(by Bayes' Theorem)} \\ &= \lim_{\delta \to 0} \frac{1}{\delta} \Pr(t \leq T < t + \delta) / \Pr(T > t) \\ &= f(t)/S(t) \end{aligned}$$

Next we show that $S(t) = \exp\{-\int_0^t h(u)du\} = \exp\{-H(t)\}$. One way of showing this is as follows. Using what we know about the derivatives of logs of functions:

$$h(t) = \frac{f(t)}{S(t)} = \frac{dF(t)/dt}{S(t)} = \frac{-dS(t)/dt}{S(t)} = -\frac{d}{dt}\log S(t),$$

Integrating both sides of the above over the range from 0 to t, gives:

$$\int_0^t h(u)du = -\int_0^t \frac{d}{du}\log S(u)du.$$

The right hand side simplifies to

$$-\left[\log S(u)\right]_{u=0}^{u=t} = -\{\log S(t) - \log S(0)\}.$$

Now we use the fact that $S(0) = 1$, because at time 0 nobody has yet had the outcome of interest. Hence the above expression simplifies to $\log S(t)$. Plugging this back into the previous expression we get:

$$\int_0^t h(u)du = -\log S(t).$$

Rearranging this gives

$$S(t) = \exp\{-\int_0^t h(u)du\} = \exp\{-H(t)\}.$$

Exercise 1.2: What is the connection between the exponential distribution and the Weibull distribution?

When the shape parameter κ is set to have value 1, the Weibull distribution reduces to the exponential distribution. That is, the exponential distribution is a special case of the Weibull distribution with $\kappa = 1$.

Exercise 1.3: Using what we know about the relationships between the survivor, hazard and probability density functions, write the likelihood above in terms of the hazard function and the survivor function.

In equation (1.12) we have the general form for the likelihood for survival data:

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

Using the result that $h(t) = f(t)/S(t)$, this can be written:

$$\begin{aligned} L &= \prod_{i=1}^{n} \{h(t_i)S(t_i)\}^{\delta_i} S(t_i)^{1-\delta_i} \\ &= \prod_{i=1}^{n} h(t_i)^{\delta_i} S(t_i) \end{aligned}$$

Exercise 1.4:
(a) Write down the likelihood for survival data (including censoring) which is assumed to follow an exponential distribution.
(b) Find the maximum likelihood estimate for λ

(a) Using equation (1.12) in the notes, or the result from Exercise 1.3, the likelihood under the exponential distribution is

$$L = \prod_{i=1}^{n} \lambda^{\delta_i} e^{-\lambda t_i}$$

where δ_i tales value 1 for individuals who have the event of interest and value 0 for individuals who are censored and t_i is the event time ($\delta_i = 1$) the the censoring time ($\delta_i = 0$). n is the total number of individuals.

(b) The log likelihood is

$$l = \sum_{i=1}^{n} \delta_i \log \lambda - \lambda \sum_{i=1}^{n} t_i.$$

The derivative with respect to λ is

$$\frac{dl}{d\lambda} = \sum_{i=1}^{n} \delta_i/\lambda - \sum_{i=1}^{n} t_i.$$

Setting this equal to zero and solving for λ gives the maximum likelihood estimator:

$$\hat{\lambda} = \sum_{i=1}^{n} \delta_i / \sum_{i=1}^{n} t_i,$$

which is the number of events divided by the total follow-up time.