For the three exercises on this sheet we assume that n independent realisations $D_n = \{(t_i, \delta_i),\ i = 1, \ldots, n\}$ are a mixture of event times and right censored observations, where t_i denotes the observed time for the ith individual and δ_i is the corresponding censoring indicator ($i = 1, \ldots, n$), so that $\delta_i = 1$ if t_i is an event time and $\delta_i = 0$ if the time is censored. Assuming a random censoring scheme, the likelihood function for the right censored data, D_n, takes the form

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

Exercise 1:

Let

$$\begin{aligned} D_{15} = \{&(1.05, 1), (2.16, 0), (2.37, 1), (3.22, 0), (3.58, 0), (3.78, 1), (3.86, 1), (4.01, 0), \\ &(4.23, 1), (4.42, 1), (4.64, 1), (4.99, 0), (5.85, 1), (6.03, 0), (6.31, 1)\} \end{aligned}$$

be the observed data. Assume that the true event times are exponentially distributed with parameter λ, that is, $T \sim \mathcal{E}(\lambda)$.

(a) For the data D_n derive the log-likelihood $l(\lambda)$, the score function $s(\lambda)$, the observed Fisher information $I(\lambda)$, and the maximum likelihood estimator (MLE) $\hat{\lambda}_{ML}$ of λ.

(b) Derive $\hat{\lambda}_{ML}$ under the assumption that all realisations t_i from D_n are uncensored.

(c) Derive $\hat{\lambda}_{ML}$ and $I(\hat{\lambda}_{ML})$ for the data D_{15}.

(d) In **R** graphically compare the log-likelihood for the data D_{15} with the log-likelihood obtained from using the following result:

$$\hat{\lambda}_{ML} \overset{a}{\sim} \mathcal{N}(\lambda, I^{-1}(\hat{\lambda}_{ML})) \ .$$

For the plot, use a range of $[0, 1]$ for λ and mark the location at which $\lambda = \hat{\lambda}_{ML}$.

(e) Compute a 95% confidence interval for λ based on the asymptotic normal distribution.

Exercise 2:

Assume that the true event times are distributed as $T \sim \mathcal{WB}(\alpha, \lambda)$. In this case, the MLEs of α and λ cannot be obtained analytically.

(a) For the data D_n derive the log-likelihood $l(\alpha, \lambda)$, the score function $s(\alpha, \lambda)$ and the observed Fisher information $I(\alpha, \lambda)$ and implement these functions in **R**. Build the following functions

```
loglike <- function(theta,time,delta){...}
score <- function(theta,time,delta){...}
fisher <- function(theta,time,delta){...}
```

where the argument `theta` contains the parameter vector $\theta = (\alpha, \lambda)^\top$, `time` are the observed event times of the n individuals and `delta` is the censoring indicator.

(b) Use the functions written in (a) to implement the Newton-Raphson algorithm for computing the ML estimators $\hat{\alpha}_{ML}$ and $\hat{\lambda}_{ML}$ in the Weibull model. Use the function

```
newton.raphson <- function(time,delta,start,maxiterations){...}
```

where the argument `time` contains the observed event times of the n individuals, `delta` is the censoring indicator, `start` is the vector of initial values for $\theta = (\alpha, \lambda)^\top$ and `maxiterations` corresponds to the maximum number of iterations at which the algorithm terminates.
Hint: The Newton-Raphson algorithm is an iterative method for the numerical optimization of functions. Based on initial values $\hat{\theta}^{(0)}$ new values of the unknown parameters are determined iteratively as

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + s\left(\hat{\theta}^{(k)}\right) I^{-1}\left(\hat{\theta}^{(k)}\right) \ .$$

The iterative procedure continues until a stopping criterion is met. As a stopping criterion one can use

$$\frac{||\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}||}{||\hat{\theta}^{(k)}||} < \epsilon \ ,$$

where ϵ is a small positive number.

(c) Generate $n = 200$ event times according to the distribution $\mathcal{WB}(2, 1)$ with the first 50 observations being censored. First execute the command `set.seed(1234)`. Make use of your function in (b) to compute the MLEs for the censored and uncensored observations. Use the arguments `start=c(1,1)` and `maxiterations=50`.

Exercise 3:

In the piecewise constant exponential model for survival data, the time axis is partitioned into intervals with cut-points $0 = a_0 < a_1 < \ldots < a_q = \infty$. The cut-points could be e.g. equidistant or correspond to percentiles of the observed data. For example, if $q = 10$, we can take $a_1, a_2, \ldots, a_9$ as the tenth to ninetieth percentiles of all observed uncensored event-free times, respectively. We will then assume that the hazard rate is constant within each interval, so that

$$h(t) = h_k \ , \quad \text{for} \quad a_{k-1} \leq t < a_k \ , \qquad k = 1, \ldots, q \ .$$

(a) Assuming that the survival time T is distributed according to the piecewise exponential model, derive the density $f_T(t)$.

(b) Consider uncensored realisations $t_1, \ldots, t_n$ arising from the piecewise exponential model. Determine the maximum likelihood estimators (MLEs) of the parameters $h_1, \ldots, h_q$.

(c) Determine the MLEs of the parameters $h_1, \ldots, h_q$ arising from a piecewise exponential model with right censored data D_n.

(d) Write a function `piecewise.exponential <- function(time,delta,grid){...}` in **R** to implement the piecewise exponential model. The argument `time` corresponds to the event times of the n individuals, `delta` is a censoring indicator and `grid` contains the points used for discretization of the time axis. For a given set of data D_n, the function should calculate the MLEs of the parameters $h_1, \ldots, h_q$.

(e) Recall the data `melanoma.dat` that were analyzed in previous tutorials. Fit a piecewise constant exponential survival model to this set of data. First define an appropriate censoring indicator. Then determine the MLEs of $h_1, \ldots, h_q$ under the following scenarios: (i) assuming that all observations are uncensored and (ii) taking account of the censored observations in the data. For this purpose, use equidistant time intervals with lengths 500, 1000 and 2000. Compare both scenarios graphically.