TMA4275 Life time analysis
Exercise 5, Spring 2021

Problem 1: Problem 4.1 in ABG.

Problem 2: Let T_1 and T_2 be independent and exponentially distributed variables with

$$f_{T_j}(t) = \lambda_j e^{-\lambda_j t}, \quad t \geq 0,$$

and let

$$T = \min\{T_1, T_2\} \quad \text{and} \quad H = \begin{cases} 1 & \text{if } T_1 \leq T_2, \\ 2 & \text{if } T_1 > T_2. \end{cases}$$

Assuming T_1 and T_2 to be the potential times until failure of two different causes, T is the time until failure and D is a variable specifying the reason for the failure. The given setup is therefore a competing risks model with $k = 2$ possible causes for failure.

The cause-specific rates are one way to describe a competing risks model. An alternative way for describing such a model is to consider the cumulative incidence functions,

$$P(T \leq t, H = j) \text{ for } j = 1, \ldots, k.$$

a) Show that the cumulative incidence functions for the competing risks model specified above are

$$P(T \leq t, H = j) = \frac{\lambda_j}{\lambda_1 + \lambda_2} \left(1 - e^{-(\lambda_1 + \lambda_2)t}\right)$$

for $j = 1, 2$.

b) Show that the cause-specific hazard rates for the competing risks model specified above are

$$\alpha_j(t) = \lambda_j$$

c) Discuss the property that cause-specific hazard rates are not influenced by the distribution of the potential failures of other causes, while this is not true for the cumulative incidence function. Note: This is a general property for competing risks.

d) Choose reasonable values for λ_1 and λ_2 and simulate a number of realisations of (T, D). Illustrate the result in **c**) by using the simulated data to estimate the cause-specific hazard rates and cumulative incidence functions. Try with different sets of values for (λ_1, λ_2), changing only λ_1 or λ_2 at a time, and see the effect. Hint: Storing the simulated T's in a vector "obstime" and the correpsonding D's in a vector "d", you can in R estimate and plot the cause-specific hazard rate for cause 1 by first loading the "survival" package by the R command

library(survival)

and thereafter running the R commands

causespec1 <- survfit(Surv(obstime,d==1) 1,type="fh2")
plot(causespec1,fun="cumhaz",mark.time=F,main="Cum.haz. 1")

To estimate and plot the cimulative incidence functions you can use the R commands

cumincid <- survfit(Surv(obstime,d,type="mstate") 1)
plot(cumincid,conf.int=T,main="Cum.incidence").

Problem 3: At an intensive care unit at a hospital one is interested in whether the presence or no-presence of pneumonia for a patient at admission have an impact on the length of the hospital stay, T, i.e. time from entry to discharge from the hospital unit.

Below are times (in days) to discharfge from the hospital for 8 patients without pneumonia at admission ($x = 0$), and 7 patients with pneumonia at admission ($x = 1$).

No pneumonia at admission: $2, 3\star, 6, 6, 10, 11, 12\star, 23$
Pneumonia at admission: $4\star, 9, 12\star, 17, 24, 26\star, 32$

Here $\star$ means a right censored observation, and right censoring occured if a patient was still in hospital at the end of the study or if a patient died in hospital.

It was decided to analyse the data by a Cox regression model with the single covariate x defined above, representing the status of pneumonia at admission.

a) Write down an expression for the hazard function of a patient with pneumonia status x.

b) Write down an expression for the partial likelihood and use R to plot the partial likelihood as a function of β. Note: you may load the data into R by the command
pneu=read.table("http://www.math.ntnu.no/emner/TMA4275/2021v/Datasets/pneumonia.txt",header=T)

c) Estimate β using R (see commands below). Note the tied events at time 6. The simplest approach here is to use Breslow's approximation, which considers multiple event times as different events with the same risk set. Thus, each failure at a certain tied time produces a factor in the partial likelihood, with all having the same risk set (i.e. using the same denominator in the factors). You will have to write *method="breslow"* (see below), since the default method in R is slightly different (*method="efron"*)

You may use the following R commands:
library(survival)
fit.pneu = coxph(Surv(time,cens==1) x,method="breslow",data=pneu) summary(fit.pneu)

d) Is there a significant difference between the discharge times for patients without and with pneumonia? Formulate the question as a testing problem regarding β and derive the conclusion with the significance level is 5%.

Compute the estimate of the relative risk of a patient without pneumonia compared to a patient with pneumonia. What is the practical interpretation of this number in the current situation?

e) Discuss the difference between the test for $H_0 : \beta = 0$ as considered above, and a logrank test for the equality of the hazard functions of the two groups (with and without pnwumonia at admission).

Perform the logrank test by hand using the given data. Hint: the computed expected number of discharges under the null hypothesis should be 3.20 and 6.80 for the patients without and with pneumonia at adminission, respectively.

f) Perform the logrank test also by using R, for example using the command
survdiff(Surv(time,cens) x,data=pneu)
How does the conclusion of the logrank test fit with the conclusion of the corresponding problem in **d**)?