Exercise 1

The following R output shows part of the output obtained with function `summary.surfit`, which was used to estimate the survival functions in two treatment groups. The survival time of interest was measured in days:

```
> summary(survfit(Surv(stime, status) ~ treatment))

      Treatment=1
 time n.risk n.event survival
    1     38       1
   55     37       1
   74     36       1
   86     35       1
    :      :       :
  129     28       1
  172     27       1    0.684
  192     26       1
  194     25       1
  230     23       1
    :      :       :
  662     13       1    0.353
    :      :       :
 2081      1       0    0.353

      Treatment=2
 time n.risk n.event survival
  353     13       1
  365     12       1
  464      9       1
  475      8       1
  563      7       1
 1007      1       1
```

(a) Which are the values of the Kaplan-Meier estimator of the survival function of Treatment 1 after 74, 196, 700, and 2100 days, respectively?

(b) Which are the values of the Nelson-Aalen estimator of the cumulative hazard function of Treatment 1 after 74 and 80 days, respectively?

(c) Estimate the values of the hazard function of Treatment 1 after 74 and 80 days, respectively.

(d) Estimate the value of $S(1007)$ of Treatment 2 by means of both the Kaplan-Meier and the Nelson-Aalen estimator and compare both values. Which estimate do you think is more realistic?

Solution:

(a) Estimation of the survival probabilities using the Kaplan-Meier estimator:

$$\hat{S}(74) = \left(1 - \frac{1}{38}\right)\left(1 - \frac{1}{37}\right)\left(1 - \frac{1}{36}\right) = \frac{35}{38} = 0.921,$$

$$\hat{S}(196) = \hat{S}(172)\left(1 - \frac{1}{26}\right)\left(1 - \frac{1}{25}\right) = 0.631,$$

$$\hat{S}(700) = \hat{S}(662) = 0.353.$$

$\hat{S}(2100)$ Not defined, because the largest survival time observed (2081 days) is right-censored.

(b) Nelson-Aalen estimator for the cumulative hazard function:

$$\hat{\Lambda}_{\mathrm{NA}}(t) = \sum_{i:t_i \leq t} \frac{d_i}{n_i}.$$

Hence:

$$\hat{\Lambda}_{\mathrm{NA}}(74) = \hat{\Lambda}_{\mathrm{NA}}(80) = \frac{1}{38} + \frac{1}{37} + \frac{1}{36} = 0.081.$$

(c) The estimator of the hazard function at time t is the number of events at t divided by the number of individuals at risk at t:

$$\hat{\lambda}(74) = \frac{1}{36} = 0.03,$$
$$\hat{\lambda}(80) = \frac{0}{35} = 0.$$

(d)

$$\hat{S}_{\mathrm{KM}}(1007) = \big(1 - \frac{1}{13}\big) \cdots \big(1 - \frac{1}{1}\big) = 0,$$
$$\hat{S}_{\mathrm{NA}}(1007) = \exp\big(-\big(\frac{1}{13} + \cdots + \frac{1}{1}\big)\big) = 0.215.$$

According to the estimation obtained with the Kaplan-Meier estimator, the probability of surviving at least 1007 days is 0. Contrary to that, the estimated survival probability with the Nelson-Aalen estimator is positive, which seems to be more realistic.