Exercise 1:

In **R** make use of the function `rgamma()` to generate 250 event times from a Gamma distribution with shape parameter 1.5 and scale parameter 0.75. Then, generate censored observations according to the following censoring schemes:

(a) For every individual $i = 1, \ldots, n$ there is a fixed observation period c_i (type I censoring). Assume that the fixed observation period is the same for all individuals and set $c = 1, 3, 5$.

(b) The study continues until the failure of the first k individuals, where k is some predetermined integer ($k < n$) (type II censoring). Set $k = 100, 150, 200$.

(c) Each unit has associated with it a potential censoring time C_i and a potential lifetime T_i, which are assumed to the independent random variables (type III censoring or random censoring). Assume that the censoring times are exponentially distributed with parameter $\lambda = 0.25, 0.5, 1$.

For (a)–(c) and varying parameters that govern the censoring mechanisms, calculate the Kaplan-Meier estimator $\hat{S}_{\text{KM}}(t)$ and the Nelson-Aalen estimator $\hat{H}_{\text{NA}}(t)$ for the censored data and compare the results with the corresponding estimations of the uncensored event times.

Exercise 2:

The file `melanoma.dat`, which is stored along with a description in the Stud.IP folder "Data", contains data of a clinical study at the Department of Plastic Surgery, University Hospital of Odense, Denmark. 205 patients were observed during the years 1962 to 1977 after a skin cancer operation (removal of the tumor and the surrounding skin), either up to the end of the year 1977 or their early death.

(a) Write a function `lifetable <- function(time,delta,grid){...}` in **R** to implement the life-table method. 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, the function should return the conditional probability of experiencing the event in each interval, the estimated survival function at the start of each interval, and the estimated density and hazard rate at the midpoint of each interval.

(b) Apply your function written in (a) to the skin cancer data. Define an appropriate censoring indicator `delta` and use equidistant time intervals (`grid`) between 0 and 6000 with interval lengths 100, 500, 1000 and 2000. Display each quantity that is asked for in (a) graphically and interpret the results.

(c) In **R**, compute and plot the Kaplan-Meier estimate of the survivor function for the data of survival after malignant melanoma.

(d) Graphically compare the Kaplan-Meier estimate obtained in (c) with the estimate of the survivor function obtained by the life-table method.

Exercise 3:

Several proposals for computing pointwise confidence intervals for values of the survival function do exist.

(a) Recall the data file `melanoma.dat` that was analysed in exercise 2. In **R**, use the `survfit()` function from the package `survival` to compute the Kaplan-Meier estimate of the survivor function and compare various pointwise 95% confidence intervals that are available in the option `conf.type`.

(b) Use the delta method that was introduced in the lectures to derive the asymptotic variance of $\ln(-\ln(\hat{S}(t))$.

(c) Based on the result obtained in (b) give a corresponding asymptotic $100(1-\alpha)\%$ confidence interval for $S(t)$.

Exercise 4:

In this exercise we consider data on the efficiency of bone marrow transplants for patients with acute myelotic leukemia (AML) and acute lymphoblastic leukemia (ALL). The data are reported in section 1.3 and appendix D of Klein and Moeschberger (2003) and details of this study are given in Copelan et al. (1991)[1]. The data are contained in the data frame `bmt` in the **R** package `KMsurv`. Make use of `help(bmt)` to become acquainted with the data set.

(a) We shall focus on the disease-free survival probabilities for the 38 patients with ALL. An individual is said to be disease free at a given time after transplant if that individual is alive without the recurrence of leukemia. In **R**, construct and graphically compare 95% pointwise confidence intervals, Equal-Precision (EP) confidence bands and Hall-Wellner confidence bands for the disease-free survival function in the range $100 \leq t \leq 600$. Hint: In the lecture, we used the function `confBands()` from the **R** package `OIsurv` to construct confidence bands. Confidence bands can also be computed using the `km.ci()` function in the **R** package `km.ci`.

(b) Estimate the mean disease-free survival time for the three disease groups.

(c) Estimate the median disease-free survival time for the three disease groups and construct a 95% confidence interval for the median.

[1] Copelan, E. A. et al. (1991): Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with Bu/Cy, *Blood*, Vol. 78, pp. 838-843.