Exercise 1:

Freireich et al. (1963)[1] report the results of a clinical trial of a drug 6-mercaptopurine (6-MP) versus a placebo in 42 children with acute leukemia. The trial was conducted at 11 American hospitals. Patients were selected who had a complete or partial remission of their leukemia induced by treatment with the drug prednisone. (A complete or partial remission means that either most or all signs of disease had disappeared from the bone marrow.) The trial was conducted by matching pairs of patients at a given hospital by remission status (complete or partial) and randomizing within the pair to either a 6-MP or placebo maintenance therapy. Patients were followed until their leukemia returned (relapse) or until the end of study. The data are given in the data frame `drug6mp` in the **R** package `KMsurv`. Make use of `help(drug6mp)` to become acquainted with the data set. Note that the remission times in Freireich et al. (1963) were measured in weeks and not, as indicated in `help(drug6mp)`, in months!

(a) Consider only the 6-MP patients. Compute (using only your pocket calculator) the Kaplan-Meier estimate of the survivor function along with the standard errors.

(b) In **R**, using the `survfit` function from the package `survival`, compute and plot the Kaplan-Meier estimates of $S(t)$ for the two treatment groups.

(c) Consider again only the 6-MP patients. Compute (using only your pocket calculator) the Nelson-Aalen estimate of the cumulative hazard function and its estimated variance.

(d) Write a function `nelson.aalen <- function(time,delta)` in **R** that returns, for a given set of data, the Nelson-Aalen estimator of the cumulative hazard and its estimated variance. The argument `time` corresponds to the event times of the n individuals and `delta` is a censoring indicator.

(e) For the 6-MP group, compute and plot in **R** the Nelson-Aalen estimator along with 95% pointwise asymptotic confidence intervals.

(f) Graphically compare in **R** the Nelson-Aalen estimator $\hat{H}_{\mathrm{NA}}(t)$ with the Breslow estimator $\hat{H}_{\mathrm{B}}(t) = -\ln(\hat{S}_{\mathrm{KM}}(t))$.

[1] Freireich, E. J. et al. (1963): The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: a model for the evaluation of other potential useful therapy, *Blood*, Vol. 21, pp. 699-716.

Exercise 2:

Let T be a non-negative continuous random variable denoting the time until the event of interest. Show that the mean survival time, $\mu = \mathrm{E}(T)$, is

$$\mu = \int_0^\infty S(t)\,\mathrm{d}t \ . \tag{1}$$

Given the Kaplan-Meier estimator $\hat{S}_{\mathrm{KM}}(t)$ for $S(t)$, use (1) to think of a suitable estimator $\hat{\mu}$ for μ. You may use your proposed estimator to estimate the mean survival time for the 6-MP patients in exercise 1.

Exercise 3:

It follows from the definition of the hazard rate of a continuous random variable T that

$$\mathrm{P}(t \le T < t + \Delta t \mid T \ge t) = h(t)\Delta t + o(\Delta t) \ ,$$

where Δt is a small time interval and $o(\Delta t)/\Delta t \to 0$ as $\Delta t \to 0$. Hence, for small Δt,

$$\mathrm{P}(t \le T < t + \Delta t \mid T \ge t) \approx h(t)\Delta t =: \mathrm{P}_1 \ .$$

Compare the probability $\mathrm{P}(t \le T < t + \Delta t \mid T \ge t) =: \mathrm{P}_2$ with its approximation P_1, by giving an expression for its difference $\delta = \mathrm{P}_1 - \mathrm{P}_2$, in the following two cases:

(a) $T \sim \mathcal{E}(\lambda)$,

(b) $T \sim \mathcal{WB}(\alpha, \lambda)$ and $\alpha = 2$.

In **R**, plot P_1, P_2 and $\delta = \mathrm{P}_1 - \mathrm{P}_2$ versus $\Delta t \in [0, 2]$ for $T \sim \mathcal{E}(1)$ and $T \sim \mathcal{WB}(2, 1)$. Use $t = 0.5, 1$.