Practical session 5:
Illness-death model and left-censored data

Analysis of Stanford heart transplant data

Between July 1967 and April 1974, 103 end-stage heart disease patients who could not respond to conventional medical or surgical therapy were admitted to the Stanford Heart Transplant Study at Stanford University Hospital, Palo Alto, California. A donor heart, matched on blood type, was then sought for each patient. This search lasted anywhere from a few days to almost a year. Some patients died before a suitable heart was found. At the end of the study, the 103 patients were divided into two subgroups: those not receiving cardiac allografts because of the unavailability of a suitable donor within the study period, and those who were matched with an appropriate donor and received an allograft. Along with the survival status (1 for dead, 0 for alive) and the transplant status (1 = transplanted, 0 = not transplanted), the patient's age at acceptance in the study (in years) was also recorded.

1. Inspect the data set *heart*, available in the library(survival). How are the data formatted? How many patients are there for heart transplant? How has the variable *event* been coded? Which transition intensities in the health-illness model can be estimated using these (formatted) data? Why are there two rows for some individuals? When is there late entry?

   The original data is provided in data set *jasa* in library survival. This may be helpful.

2. The question of interest is to investigate whether transplantation of a new heart changes the mortality rate, i.e. to compare the death rates between patients without and with transplantation. What is the suitable time scale?

3. Denote the mortality before transplantation by $\alpha_{02}(t)$ and the after transplantation by $\alpha_{12}(t)$, where time t is the time since acceptance in the heart transplant study. Assuming the same baseline hazard for the two death rates, a possible model can be specified through $\alpha_{02}(t) = k\alpha_{12}(t)$. Fit this model to the data using the Cox proportional hazard model. Use age as a covariate.

Analysis of Baboon descent data

The file *baboon.csv* contains the data on observed descent times for a troop of baboons. Here, the descent time is defined as the time when half of the members of the study troop had descended from the trees they spend overnight. The observers often arrived later in the day than the descent, and for such days they could only ascertain that the descent had taken place before their own arrival time, i.e. the descent time was left-censored. The problem is to describe the distribution of the descent time based on the left-censored data.

The data file contains four data items: obs, date, time, and status (1 = uncensored, 0 = left-censored). A non-parametric (Kaplan-Meier) estimate of the descent time can be determined by 'time reversal'. Use "24–time" (hours) as your time (i.e. time before midnight). You can then analyse the data as right-censored. Plot the resulting survival curve in the original time scale.

N.B. Time is coded somewhat oddly. For example, 630 means 6 hours and 30 minutes. You need to transform time into decimals first.

N.B. This example is not one of the health-illness model.