Exercises Survival Analysis Lecture 2

Marta Fiocco & Hein Putter

1 Introduction

Look at the following study protocols and describe the type of censored data (in case censored data are present). In case of censoring, argue if you expect the censoring mechanism to be informative. In case of truncation, argue conditional on what you can make inference.

Represent some of the persons in the dataset graphically.

1. In a randomized clinical trial, one group of patients receives a new treatment and the other group placebo. We want to know if the treatment prolongs survival.

2. Time to relapse after diagnosis of breast cancer is studied for treatment A and B (assigned on the basis of clinical considerations).

3. A study is designed to investigate the relationship between the duration of breast-feeding and several explanatory variables.

4. A cohort of children is followed to study when they can identify at least 10 letters correctly. They are followed between their 5th and 7th birthday. During this period they are tested every 3 months. For practical reasons, the children are only followed as long as they attend one of the participating schools (all normal primary schools).

5. Time to failure of a new type of hip prothesis is compared to the standard type. The study is stopped after 50 events.

6. Time from Diagnosis to Death for Leukemia patients is estimated on the basis of the EBMT registry (in which patients are entered after transplant; their date of diagnosis is included). Data are collected for patients transplanted between 1995 and 2000, with follow-up (if available) until December 2009.

7. Time from conception to birth (for humans).

8. We want to estimate the survival of elderly people. For that purpose, we follow all people living in old people's homes in Amsterdam between 1/1/2005 and 1/1/2009. The time scale is time since birth.

9. Consider a study in which interest centers on the time to recurrence of a particular cancer following surgical removal of the primary tumor. (a) Three months after their operation, the patients are first examined to determine if the cancer has recurred. At this time, some of the patients may be found to have a recurrence.

(b) Which type of censoring mechanism is present in the data if a patient is observed to be disease free at three months, but a recurrence is detected when examined six months after surgery?

10. In a school six years old children who can read four-letter words without difficulty are selected. Data on these children are used to determine how long it takes from learning the first (separate) letter to being able to read four-letter words.

11. Simulate survival data by assuming independent censoring following:
    (a) an exponential distribution;
    (b) a Weibull distribution.
    (c) Plot the survival functions.
    Hint: 1) Simulate survival time from the exponential distribution; 2) Simulate censoring time from exponential distribution; 3) Make the indicator variable δ.

12. Suppose the underlying survival time $T \sim exp(\lambda)$. We observe the following data: (t_i, δ_i), $i = 1, \ldots, n$. Write down the likelihood function and find the MLE (maximum likelihood estimator) of λ in case the data are (i)right censored; (ii): ignore now the censoring mechanism, compute MLE. (iii) Compare the estimators obtained in (i) and (ii). What are your conclusions about the bias?

13. The table shows an example taken from a clinical trial of 146 patients treated after they had a myocardial infarction (MI). Estimate the survival at 5 years: $S(5) = P(T \geq 5)$; motivate your answer.

Year since entry into study	Number alive and under observation at beginning of interval	Number dying during interval	Number censored or withdrawn
$[0, 1)$	146	27	3
$[1, 2)$	116	18	10
$[2, 3)$	88	21	10
$[3, 4)$	57	9	3
$[4, 5)$	45	1	3
$[5, 6)$	41	2	11
$[6, 7)$	28	3	5
$[7, 8)$	20	1	8
$[8, 9)$	11	2	1
$[9, 10)$	8	2	6