SURVIVAL ANALYSIS Fall 2008 L3

Sampling schemes with incomplete data

1. Type I Censoring

2. Type II Censoring

3. Random Censoring:
   3.1 Right-Censoring
   3.2 Left-censoring

4. Interval censoring

5. Truncation

Homework: *Klein, J. & Moeschberger, M. (2003):*
Ch. 3, Sec. 1-4,

L3

SURVIVAL ANALYSIS

Introduction: Censorship

The censored observations contain only partial information (incomplete observations) about the random variable of interest.
Let us consider different types of censoring.

1. Type I Censoring

Let

$$T_1, \ldots, T_n$$

be independent , identically distributed (i.i.d.) random variables each with cdf F.
Assume also that t_c is a some (pre-assigned) fixed **censoring** time.
Instead of observing $T_1, \ldots, T_n$, the variables of interest, we can only observe

$$Y_1, \ldots, Y_n,$$

where

$$Y_i = \begin{cases} T_i, & \text{if } T_i \le t_c, \\ t_c, & \text{if } T_i > t_c. \end{cases}$$

Remark 1. The distribution of Y has positive mass at $y = t_c$:

$$P(T > t_c) > 0 .$$

2. Type II Censoring

Let $r < n$ be fixed, and let

$$T_{(1)} < \ldots < T_{(n)}$$

be the order statistics of $T_1, \ldots, T_n$. Observation ceases after r-th failure, so we can only observe

$$T_{(1)}, \ldots, T_{(r)}.$$

The full ordered observed sample is

$$Y_{(1)} = T_{(1)}$$
$$Y_{(2)} = T_{(2)}$$
$$\cdot$$
$$Y_{(r)} = T_{(r)}$$
$$Y_{(r+1)} = T_{(r)}$$
$$\cdot$$
$$Y_{(n)} = T_{(r)}$$

Remark 2. In Type II Censoring model we have instead t_c the random time = $T_{(r)}$,
Both the Type I and the Type II censoring arise in engineering applications.

3. Random Censoring

3.1 Right-censoring

Let $C_1, \ldots, C_n$ be i.i.d. random variables each with cdf G. Here C_i is the censoring time associated with T_i.
We can only observe the pairs :

$$(Y_1, \delta_1), (Y_2, \delta_2), \ldots, (Y_n, \delta_n),$$

where

$$Y_i = \min(T_i, C_i),$$

and

$$\delta_i = I(T_i \le C_i).$$

$\delta_1, \ldots, \delta_n$ contain the censoring information.

Remark 3. Usually it is assumed that T's and C's are independent. So that $Y_1, \ldots, Y_n$ are i.i.d. random variables as well.
This type of censoring is called the **right- censoring**.

3.2 Left -censoring

There is also another model, the so-called **left-censoring** model, when we can only observe

$$(Y_1, \varepsilon_1), (Y_2, \varepsilon_2), \ldots, (Y_n, \varepsilon_n),$$

where

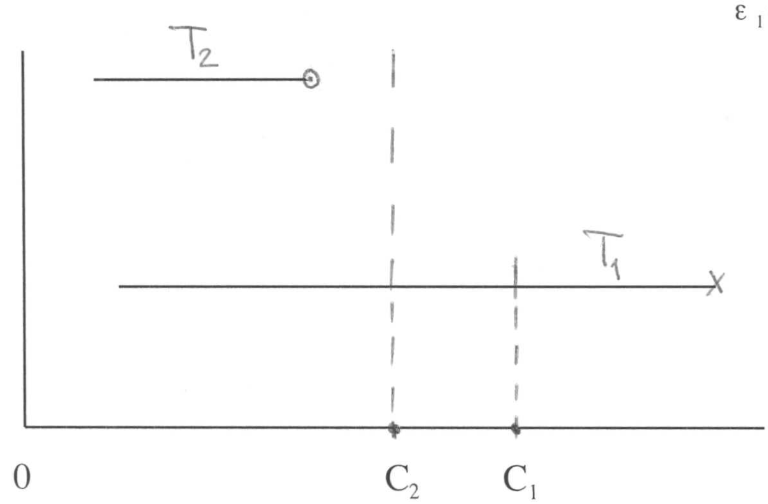
$$Y_i = \max(T_i, C_i),$$

and

$$\varepsilon_i = I(C_i \le T_i).$$

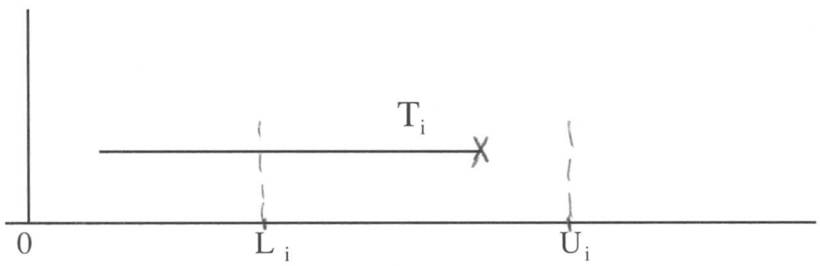
$\varepsilon_1, \ldots, \varepsilon_n$ contain again the censoring information.

$\varepsilon_1 = 1, \quad \varepsilon_2 = 0,$



4. Interval Censoring

The combination of the left-censoring and rith-censoring leads to the so-called Interval- censoring model when we observe T_i only on a set of the form $[L_i, U_i]$



Here T_i is observed.

5. Random Truncation

In contrast to the interval censoring there is a **random truncation** model in which : if the random variable of interest falls outside some interval it is not recorded. The Left-truncation model is very common in the fields like demography and epidemiology.

6. Examples:

Example #1. Type I Censoring:
In the period 1962 –1972, n=225 patients with malignant melanoma (cancer of the skin) had radical operation performed at the Department of Plastic Surgery, University Hospital of Odense, Denmark.
All patients were followed until the end of 1977, that is, it was noted if and when any of the patients died.

The time variable – is *time since operation*.
Among the possible **risk factors** (or *covariates*) screened for significance were the *sex* and *age* at operation.

The *covariates* could be both qualitative and quantitative, say, characteristic of the tumor, such as its width, thickness and location.

Note that survival time T_i is known only for those who died before the end of 1977. Thus, patients alive on Dec. 31, 1977 were censored at that day,
i.e. t_c is fixed . We have Type I censoring.

14 patients had died earlier from causes other than malignant melanoma, so in the study of the death intensity from disease only, theses patients are censored.
134 – alive at Jan. 1 , 1978, so they are censored as well.

Example #2.
There is a batch of transistors; we put them all on test at t=0, and record their times to failure. Some transistors may take a long time to burn out. We will not want to wait that long to end the experiment. Therefore we might stop the experiment at a pre-specified time t_c .
So, we have here the Type I censoring.

Type II Censoring: if we decide to wait until a pre-specified fraction
r/n
of the transistors has burned out.

Random Censoring:
arises in medical applications with animal studies or clinical trials.

In a clinical trials, patients may enter the study at different (random) times;
Then each is treated with one of several possible therapies.
We want to observe their life times, but censoring occurs in one of the following forms:
(a) Loss to follow-up: the patient may decide to move elsewhere
(b) termination of the study
(c) drop out: the patient refuse to continue the treatment

Truncation:
Example#3.
Left-truncation is very common in epidemiology. For, example, suppose we want to get the distribution and expected size of a certain organelle in the cell. Because of limitations on the measuring equipment, if an organelle is below a *certain size* it can not be detected.