

# Survival Analysis

## Lecture 2

Marta Fiocco & Hein Putter

Department of Medical Statistics and Bioinformatics  
Leiden University Medical Center

# Outline

## Recall

## Right censoring

- Type I censoring

- Type II censoring

## Left or interval censoring

- Left censoring

- Interval censoring

## Truncation

- Left truncation

- Right truncation

## Likelihood

- Likelihood construction

- Likelihood for for general censoring case

## Likelihood for left truncated data

# Outline

## Recall

### Right censoring

Type I censoring

Type II censoring

### Left or interval censoring

Left censoring

Interval censoring

### Truncation

Left truncation

Right truncation

### Likelihood

Likelihood construction

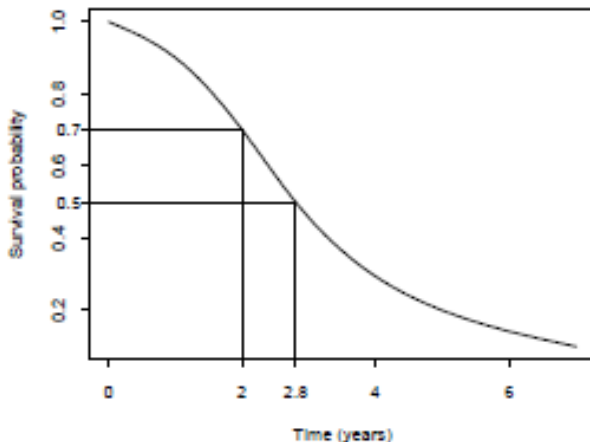
Likelihood for for general censoring case

### Likelihood for left truncated data

# Basic quantities

- ▶ **Survival function:**  $S(x) = P(X > x) = \int_x^\infty f(t)dt$
- ▶ **Mean survival time:**  $\mu = \int_0^\infty S(t)dt$
- ▶ **Median survival time:**  $S(x_{0.5}) = 0.5$
- ▶ **Mean Residual Life Time(mrl):**  $mrl(x_0)E(X - x_0|X \geq x_0)$   
(average remaining survival time given the population has survived beyond  $x_0$ )
- ▶ **Hazard rate:**  $h(x) = \lim_{\Delta x \rightarrow 0} \frac{Pr[x \leq X < x + \Delta x | X \geq x]}{\Delta x}$   
(instantaneous rate of failure -experiencing the event- at time  $t$  given that an individual is alive at time  $t$ )
  - ▶ Note: The hazard rate is **NOT** a probability, it is a probability rate; Therefore it is possible that a hazard rate can exceed one

# The survival function for a hypothetical population



- ▶ in the population 70% of the individuals survive 2 years
- ▶ median survival time: 2.8 years (i.e., 50% of the population will survive at least 2.8 years)

# Outline

## Recall

## Right censoring

- Type I censoring

- Type II censoring

## Left or interval censoring

- Left censoring

- Interval censoring

## Truncation

- Left truncation

- Right truncation

## Likelihood

- Likelihood construction

- Likelihood for for general censoring case

## Likelihood for left truncated data

## Right Censoring

- ▶ **Type I censoring:** event is observed if it occurs prior to some prespecified time
- ▶ Some individuals are still alive at the end of the study or analysis so the event of interest, (death, relapse etc) has not occurred
- ▶ In addition to censoring because of insufficient follow-up (i.e., end of study censoring due to staggered entry), other reasons for censoring includes
  - ▶ loss to follow-up: patients stop coming to clinic or move away
  - ▶ deaths from other causes: competing risks



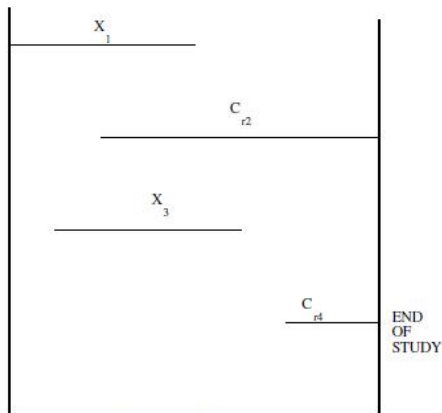
## Type I censoring

- ▶  $X$ : lifetime;  $X$ 's independent and identically distributed (i.i.d.)
- ▶  $C_r$ : right censoring time
- ▶ The lifetime  $X$  of an individual is known iff:  $X \leq C_r$ ;
- ▶  $X > C_r$ : individual is a survivor; the event time is censored at  $C_r$
- ▶ data representation:  $(T, \delta)$ , in which  $T = \min(X, C_r)$  and  $\delta$  the event indicator (0/1)
- ▶  $T$  unknown if larger than  $C_r$

$$\delta = \begin{cases} 1 & T = X \\ 0 & T = C_r. \end{cases} \quad (1)$$

## Type I censoring

- ▶ progressive Type I censoring: different, fixed-sacrifice censoring times
- ▶ generalized Type I censoring: different entry times, terminal point of the study predetermined (censoring times known when a subject enters the study)



## Type I censoring

- ▶ progressive Type I censoring: different, fixed-sacrifice censoring times
- ▶ sacrifice times are chosen to reduce the cost of maintaining the animals while allowing for limited information on the survival experience of longer lived mice

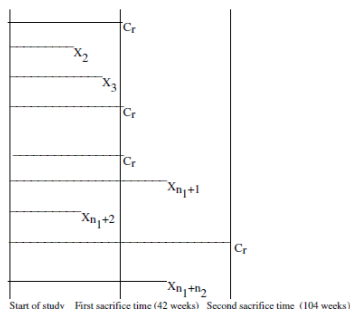
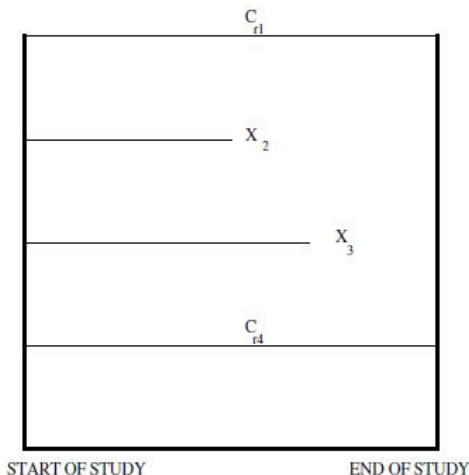


Figure 3.2 *Type I censoring with two different sacrifice times*

## Type I censoring

Generalized Type I censoring for the four individuals with each individuals starting time backed up to 0



## Representation type I censoring for the four individuals

$$T_1 = X_1, \delta_1 = 1$$

$$T_1 = C_{r_2}, \delta_2 = 0$$

$$T_3 = X_3, \delta_3 = 1$$

$$T_4 = C_{r_4}, \delta_4 = 0$$

- ▶  $X_1$ : death time for first subject
- ▶  $X_2$ : right censored time for second subject

- ▶ **Type II censoring**: Study continues until failure of the first  $r$  individuals
- ▶ Progressive Type II censoring: after the first  $r_1$  failures, part of the rest removed
- ▶ **Competing risks censoring**: special case random censoring
- ▶ later during the course more details
- ▶ censoring has to be **non-informative/independent**

# Outline

## Recall

## Right censoring

Type I censoring

Type II censoring

## Left or interval censoring

Left censoring

Interval censoring

## Truncation

Left truncation

Right truncation

## Likelihood

Likelihood construction

Likelihood for for general censoring case

## Likelihood for left truncated data

# Left censoring

- ▶ censoring occurs when the event of interest is only known to happen before a specific time point
- ▶ example 1.17: study of *time to first marijuana use*; 191 high school boys were asked: "when did you first use marijuana?"
  - ▶ "I have used it but cannot recall when the first time was"  $\Rightarrow$  their *time to first marijuana use* is **left censored** at their current age
  - ▶ boys who never used marijuana: *time to first marijuana use* is **right censored** at their current age
  - ▶ exact *time to first use marijuana* only for those who remembered



## Notation

- ▶  $C_l$ : censoring time (the event of interest has already occurred for the individual before that person is observed in the study at time  $C_l$ )
- ▶  $X < C_l$  ( $X$ : lifetime)
- ▶ data representation:  $(T, \delta)$ , where  $T = X$  if the lifetime is observed;  $\delta$  event indicator

$$\delta = \begin{cases} 1 & \text{if exact lifetime } X \text{ is observed} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

- ▶  $T = \max(X, C_l)$
- ▶ often: data doubly censored

- ▶ the event of interest is only known to take place in an interval
- ▶ e.g. clinical event not immediately diagnosed
- ▶ or equipment checked at intervals
- ▶ example 1.18:
  - ▶ in a study to compare *time to cosmetic deterioration* of breasts for breast cancer patients treated with radiotherapy and radiotherapy + chemotherapy
  - ▶ patients were examined at each clinical visit for breast retraction and the breast retraction is only known to take place between two clinical visits or right censored at the end of the study
- ▶ event in interval  $(L_i, R_i]$

# Outline

## Recall

## Right censoring

Type I censoring

Type II censoring

## Left or interval censoring

Left censoring

Interval censoring

## Truncation

Left truncation

Right truncation

## Likelihood

Likelihood construction

Likelihood for for general censoring case

## Likelihood for left truncated data

## Left truncation

- ▶ don't confuse it with censoring!
- ▶ part of the individuals is not observed because event is **outside observational window**
- ▶ example 1.16:
  - ▶ study of life expectancy (survival time measured from birth to death) using elderly residents in a retirement community: the individuals must survive to a sufficient age to enter the retirement community
  - ▶ their survival time is **left truncated** by their age entering the community
  - ▶ ignoring the truncation will lead to a biased sample and the survival time from the sample will over estimate the underlying life expectancy (later estimate survival function to look at the bias)

- ▶ the study sample consists of only those individuals who have already experienced the event
- ▶ example 1.19:
  - ▶ incubation period between infection with AIDS virus and the onset of clinical AIDS
  - ▶ ideal approach: collect a sample of patients infected with AIDS virus and then follow them for some period of time until some of them develop clinical AIDS
  - ▶ disadvantages: too lengthy and costly
  - ▶ alternative: study patients who were infected with AIDS from a contaminated blood transfusion and later developed clinical AIDS
  - ▶ but in this case: total number of patients infected with AIDS is unknown

- ▶ A similar approach can be used to study the induction time for pediatric AIDS
- ▶ Children were infected with AIDS in utero or at birth and later developed clinical AIDS
- ▶ the sample consists of children only *known* to develop AIDS

# Outline

## Recall

## Right censoring

Type I censoring

Type II censoring

## Left or interval censoring

Left censoring

Interval censoring

## Truncation

Left truncation

Right truncation

## Likelihood

Likelihood construction

Likelihood for for general censoring case

## Likelihood for left truncated data

## Likelihood construction

- ▶ random sample of individuals of size  $n$  from a specific population whose true survival times are  $X_1, X_2, \dots, X_n$
- ▶ due to right censoring (such as end of the study, loss to follow-up, competing risks (death from other causes) or any combination of these) we don't always observe these survival times,
- ▶  $C_1, C_2, \dots, C_n$ : potential censoring time
- ▶ potential data:  $(X_i, C_i), i = 1, \dots, n$
- ▶ observed data:  $(T_i, \delta_i), i = 1, \dots, n$  where

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

$$\delta_i = I(T_i \leq C_i) \begin{cases} 1 & \text{if } X_i \leq C_i \text{ (observed failure (lifetime))} \\ 0 & \text{if } X_i > C_i \text{ (observed censoring).} \end{cases}$$



- ▶ The potential data are:  $(X_1, C_1), \dots, (X_n, C_n)$
- ▶ The actual observed data are:  $(T_1, \delta_1), \dots, (T_n, \delta_n)$
- ▶ interested in **making inference on the random variable  $X$**   
i.e. in any of the following functions
  - ▶  $f(x)$ : density function;
  - ▶  $F(x)$ : distribution function;
  - ▶  $S(x)$ : survival function;
  - ▶  $h(x)$ : hazard function;

## Contributions to the likelihood:

- ▶  $D$ : set of death times;
- ▶  $R$  the set of right-censored observations,
- ▶  $L$ : set of left-censored observations;
- ▶  $I$ : set of interval censored observations
- ▶ the likelihood can be constructed by putting together the component parts as

$$L \propto \prod_{i \in D} f(x_i) \prod_{i \in R} S(C_r) \prod_{i \in L} (1 - S(C_l)) \prod_{i \in I} (S(L_i) - S(R_i))$$

# Outline

## Recall

## Right censoring

Type I censoring

Type II censoring

## Left or interval censoring

Left censoring

Interval censoring

## Truncation

Left truncation

Right truncation

## Likelihood

Likelihood construction

Likelihood for for general censoring case

## Likelihood for left truncated data

- ▶  $(T, \delta)$ : pair of random variables representing the right censoring case
- ▶  $\delta$ : indicates whether the lifetime  $X$  has been observed ( $\delta = 1$ ) or not ( $\delta = 0$ )
- ▶  $T = \min(X, C_r)$ ;  $T = X$  if the lifetime is observed;  $T = C_r$  if it is right censored
- ▶ for  $\delta = 0$

$$P(T, \delta = 0) = P(T = C_r | \delta = 0)P(\delta = 0) =$$

$$P(\delta = 0) = P(X > C_r) = S(C_r)$$

- ▶ for  $\delta = 1$

$$P(T, \delta = 1) = P(T = X | \delta = 1)P(\delta = 1) =$$

$$P(X = T | X \leq C_r)P(X \leq C_r) =$$

$$\left\{ \frac{P(X = T, X \leq C_r)}{P(X \leq C_r)} \right\} P(X \leq C_r) = f(t)$$

- ▶ combining the two expressions together we get

$$P(t, \delta) = [f(t)]^\delta [S(t)]^{1-\delta}$$

- ▶ likelihood for sample of pairs  $(T_i, \delta_i)$ ,  $i = 1, \dots, n$

$$L = \prod_{i=1}^n P(t_i, \delta_i) = \prod_{i=1}^n [f(t_i)]^{\delta_i} [S(t_i)]^{1-\delta_i}$$

- since  $f(t_i) = h(t_i)S(t_i)$  the likelihood is equivalent to

$$L = \prod_{i=1}^n [h(t_i)]^{\delta_i} S(t_i) = \prod_{i=1}^n [h(t_i)]^{\delta_i} \exp[-H(t_i)]$$

- ▶  $D$ : set of death times;  $R$ : set of right censored times
- ▶  $L$ : set of left censored observations
- ▶  $I$ : set of interval censored observations (only knowledge: the real survival time  $X_i \in [L_i, R_i]$ )
- ▶ generalization to **any** kind of censoring

$$L \propto \prod_{d \in D} f(x_d) \prod_{r \in R} S(x_r) \prod_{l \in L} (1 - S(x_l)) \prod_{i \in I} (S(L_i) - S(R_i))$$

- ▶ note that  
 $S(L_i) - S(R_i) = P(L_i \leq X_i \leq R_i) = P(X_i \in [L_i, R_i])$

# Outline

## Recall

## Right censoring

- Type I censoring

- Type II censoring

## Left or interval censoring

- Left censoring

- Interval censoring

## Truncation

- Left truncation

- Right truncation

## Likelihood

- Likelihood construction

- Likelihood for for general censoring case

## Likelihood for left truncated data



- ▶ the survival time  $X_i$  is left truncated at  $Y_i$
- ▶ consider the conditional distribution of  $X_i$  given that  $X_i \geq Y_i$

$$g(x|X_i \geq Y_i) = \frac{f(x)}{P(X_i \geq Y_i)} = \frac{f(x)}{S(Y_i)}$$

- ▶ probability survival time  $X_r$  is *right* censored at  $t_r$

$$P(X_r \geq t_r | X_r \geq Y_r) = \frac{S(t_r)}{S(Y_r)}$$

- ▶ probability survival time  $X_l$  is *left* censored at  $t_l$

$$P(X_l \leq t_l | X_l \geq Y_l) = \frac{S(Y_l) - S(t_l)}{S(Y_l)}$$

- ▶ probability survival time  $X_i$  is in  $[L_i; R_i]$  where  $(L_i \geq R_i)$

$$\begin{aligned} P(L_i \leq X_i | X_i \geq Y_i) &= P(X_i \geq L_i | X_i \geq Y_i) - P(X_i \geq R_i | X_i \geq Y_i) \\ &= \frac{S(L_i) - S(R_i)}{S(Y_i)} \end{aligned}$$

- likelihood function for left truncated observations

$$L = \prod_{d \in D} \frac{f(t_d)}{S(Y_d)} \prod_{r \in R} \frac{S(t_r)}{S(Y_r)} \prod_{l \in L} \frac{S(Y_l) - S(t_l)}{S(Y_l)} \prod_{i \in I} \frac{S(L_i) - S(R_i)}{S(Y_i)}$$

$$= \frac{\left[ \prod_{d \in D} f(t_d) \prod_{r \in R} S(t_r) \prod_{l \in L} [S(Y_l) - S(t_l)] \prod_{i \in I} [S(L_i) - S(R_i)] \right]}{\prod_{i=1}^n S(Y_i)}$$

- ▶ likelihood function for right truncated observations
- ▶ only deaths are observed

$$L = \frac{f(Y_i)}{1 - S(Y_i)}$$