NUS National University of Singapore

Censoring and truncation

ST3242: Introduction to Survival Analysis

Alex Cook

August 2008

Definitions

Censoring

Censoring is when an observation is incomplete due to some random cause. The cause of the censoring must be independent of the event of interest if we are to use standard methods of analysis.

Truncation

Truncation is a variant of censoring which occurs when the incomplete nature of the observation is due to a systematic selection process inherent to the study design.

Definitions

NUS National University of Singapore

Censoring

Censoring is when an observation is incomplete due to some random cause. The cause of the censoring must be independent of the event of interest if we are to use standard methods of analysis.

Truncation

Truncation is a variant of censoring which occurs when the incomplete nature of the observation is due to a systematic selection process inherent to the study design.

Examples

NUS National University of Singapore

Examples of censored data

Lung cancer patients are recruited to a study to test the effect of a drug on their survival from lung cancer.

- A takes part in the study until her death at time T_A. Her survival time is uncensored.
- B takes part in the study until time T_B. He then leaves the study. His survival time is censored: we know it is *at least* T_B but we don't know it precisely.
- C takes part in the study until time T_C. She then is hit by a car and dies. Her survival time *with regard to the event of interest*, namely death through lung cancer, is also censored: we know it is *at least* T_C.

Examples

Examples of censored data

Lung cancer patients are recruited to a study to test the effect of a drug on their survival from lung cancer.

- A takes part in the study until her death at time T_A. Her survival time is uncensored.
- B takes part in the study until time T_B. He then leaves the study. His survival time is censored: we know it is *at least* T_B but we don't know it precisely.
- C takes part in the study until time T_C. She then is hit by a car and dies. Her survival time *with regard to the event of interest*, namely death through lung cancer, is also censored: we know it is *at least* T_C.

Examples

Examples of censored data

Lung cancer patients are recruited to a study to test the effect of a drug on their survival from lung cancer.

- A takes part in the study until her death at time T_A. Her survival time is uncensored.
- B takes part in the study until time T_B. He then leaves the study. His survival time is censored: we know it is *at least* T_B but we don't know it precisely.
- C takes part in the study until time T_C. She then is hit by a car and dies. Her survival time *with regard to the event of interest*, namely death through lung cancer, is also censored: we know it is *at least* T_C.

Types of censoring

Right censored data

- Commonest form of censoring is right censoring.
- Subjects followed until some time, at which the event has yet to occur, but then takes no further part in the study.
- This may be because:
  - the subject dies from another cause, independently of the cause of interest;
  - the study ends while the subject survives; or
  - the subject is lost to the study, by dropping out, moving to a different area, etc.

Types of censoring

Right censored data

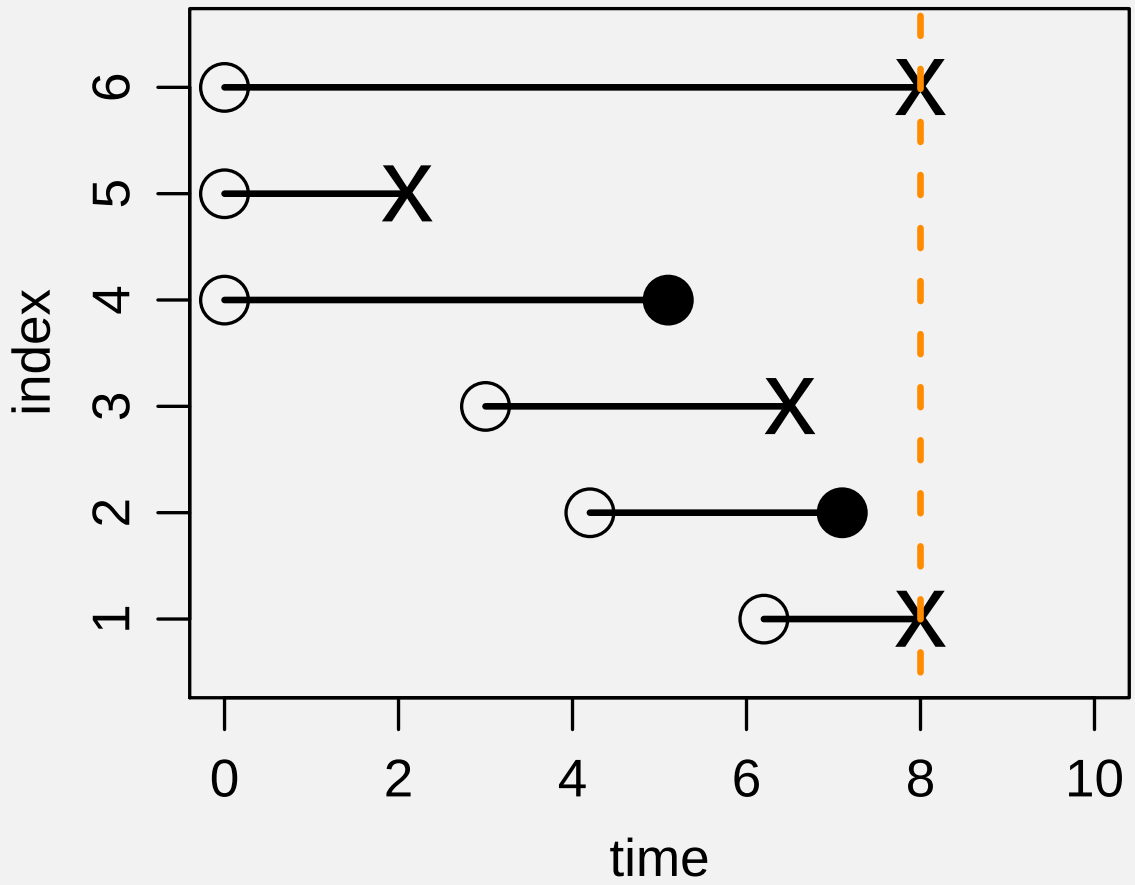
If our data contain only uncensored and right-censored data, we can represent all individuals by the triple (i, t_i, δ_i):

- i indexes subjects,
- t_i is the time at which the death or censoring event occurs to individual i, and
- δ_i is an indicator: $\delta_i = 1$ if i is uncensored and $\delta_i = 0$ if censored.

Most methods in the course apply to right censored data only.

Types of censoring

Right censored data



Types of censoring

Left censored data

- Left censoring is much rarer.
- Event of interest already occurred at the observation time, but it is not known exactly when.
- Examples of left censoring include:
  - infection with a sexually-transmitted disease such as HIV/AIDS;
  - onset of a pre-symptomatic illness such as cancer; and
  - time at which teenagers begin to drink alcohol.

Types of censoring

Interval censored data

- Interval censoring: exact time event occurs is not known precisely, but an interval bounding this time is known.
- If interval is very short, e.g. 1d, it is common to ignore this form of censoring and pick one end point of the interval consistently.
- Examples of interval censoring include:
  - infection with a sexually-transmitted disease such as HIV/AIDS with regular testing (e.g. annually); and
  - failure of a machine during the Chinese New Year.

Types of truncation

Left truncated data

- Left truncation: due to structure of the study design, we can only observe those individuals whose event time is greater than some truncation threshold.
- This threshold may be the same for all individuals or may be random.
- For example, in actuarial life tables, it is common to ignore those who die in the womb.
- Another example: imagine you wish to study how long people who have been hospitalised for a heart attack survive taking some treatment at home. The start time is taken to be the time of the heart attack. Only those individuals who survive their stay in hospital are able to be included in the study.

Types of truncation

Right truncated data

- Right truncation: only individuals with event times less than some threshold are included in the study.
- If you ask a group of smoking school pupils at what age they started smoking, you neccessarily have truncated data, as individuals who start smoking after leaving school are not included in the study.
- Similarly, if a drug developed in 1990 is compared with a placebo to analyse the survival times of (dead) cancer patients, all times to death must be $\leq$ 18y.

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

Leukæmia patients are given a drug or a placebo. Survival time is the duration from remission to relapse. The study ends at 52 weeks with some patients yet to relapse.

Answer:

Right censoring

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

Leukæmia patients are given a drug or a placebo. Survival time is the duration from remission to relapse. The study ends at 52 weeks with some patients yet to relapse.

Answer:

Right censoring

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

The age at which children are able to count from 1–10 at school. Some children are already able to count before joining school.

Answer:

Left censoring

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

The age at which children are able to count from 1–10 at school. Some children are already able to count before joining school.

Answer:

Left censoring

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

Death times at a retirement community.

Answer:

Left truncation

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

Death times at a retirement community.

Answer:

Left truncation

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

Early in the AIDS outbreak, patients with AIDS were recruited to study the time from infection with HIV to development of AIDS. At the time of the study, many people were infected with HIV but had not yet developed symptoms of AIDS.

Answer:

Right truncation

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

Early in the AIDS outbreak, patients with AIDS were recruited to study the time from infection with HIV to development of AIDS. At the time of the study, many people were infected with HIV but had not yet developed symptoms of AIDS.

Answer:

Right truncation

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

NUS students are asked the age at which they first tried marijuana.

- Some answer never.
- Some answer with an exact age.
- Some report using it but forget when.

Answer:

- Right censoring
- No censoring!
- Left censoring

Types of censoring and truncation

NUS National University of Singapore

Quiz: what kind of censoring or truncation?

NUS students are asked the age at which they first tried marijuana.

- Some answer never.
- Some answer with an exact age.
- Some report using it but forget when.

Answer:

- Right censoring
- No censoring!
- Left censoring

An alternative formulation

Let. . .

$\mathcal{X}$ be the set of possible failure times, $|\mathcal{X}| = n$
x_i be the potential failure time of individual i
y_i be the potential censoring/truncation time of individual i
$\mathcal{T}$ be the set of observed failure/censoring times, $|\mathcal{T}| = m \leq n$
t_j be the failure time of individual j

Importance of independence

NUS National University of Singapore

- Vital that cause of censoring or truncation be independent of failure.
- If not the methods in the course cannot be used.

Example of violation of independence

Patients are recruited following operation for lung cancer. "Survival" is time after diagnosis to death from lung cancer. Some patients may die of other cancers before dying of lung cancer. Those dying of other cancers more likely to have died from lung cancer.

Importance of independence

NUS National University of Singapore

- Vital that cause of censoring or truncation be independent of failure.
- If not the methods in the course cannot be used.

Example of violation of independence

Patients are recruited following operation for lung cancer. "Survival" is time after diagnosis to death from lung cancer. Some patients may die of other cancers before dying of lung cancer. Those dying of other cancers more likely to have died from lung cancer.

If in doubt...

National University of Singapore

- Try to recast the problem in some way. E.g. redefine failure as death from any cancer.
- Alternatively, try sensitivity analysis. Imagine all deaths due to other cancer are:
  - independent censoring events
  - "deaths due to lung cancer"

  Then compare consistency of results.

If in doubt...

- Try to recast the problem in some way. E.g. redefine failure as death from any cancer.
- Alternatively, try sensitivity analysis. Imagine all deaths due to other cancer are:
  - independent censoring events
  - "deaths due to lung cancer"

  Then compare consistency of results.

Fitting to censored data

Although censored and truncated data are only partially observed, we do not throw them away. Excluding them:

- increases the variance in our estimates; and
- leads to biased results.

Fortunately, for parametric models, it is easy to construct the likelihood function in the presence of censoring.

Fitting to censored data

NUS National University of Singapore

Likelihood

- If T is distributed according to some known family of distributions with unknown parameters θ, then:
  - density function is $f(t|\theta)$
  - distribution function is $F(t|\theta) = \int_0^t f(\tau|\theta)\,\mathrm{d}\tau$

Fitting to censored data

Likelihood

- Let $\mathcal{R}$ be the set of right-censored, $\mathcal{L}$ of left-censored, $\mathcal{I}$ of interval-censored and $\mathcal{U}$ of uncensored observations.
- Denote the data by $\mathcal{D} = (\mathcal{R}, \mathcal{L}, \mathcal{I}, \mathcal{U}, \mathcal{T})$ where $\mathcal{T} = \{t_i : i = 1, \ldots, n\}$.
- The log-likelihood function with no truncation is

$$\log L(\mathcal{D}|\theta) = \ldots \tag{1}$$

- The log-likelihood function with truncation is

$$\log L(\mathcal{D}|\theta) = \ldots \tag{2}$$

Conclusions

- In this section, we introduced the concepts of censoring and truncation.
- These are very frequently occuring phenomena in survival analysis.
- The most common is right censoring.
- In the next section we introduce the survival and hazard functions, which are popular ways to describe survival data in the presence of censoring.