QMWS - Survival Analysis
Censoring and truncation

Instructor: Kevin McGregor

York University
Department of Mathematics and Statistics

Censoring

The reason survival analysis exists is because of **censoring**.

An observation is **censored** if the individual is in the study, but their time to event is **unknown**.

There are different types of censoring; the type depends on whether the event happens before or after the censoring time.

- Right censoring
- Left censoring
- Double censoring
- Interval censoring

Truncation is a similar concept (though different from censoring) that we will also cover in this lecture.

Right censoring

Right censoring is the most common type of censoring.

In this type of censoring, we know that an event has not happened up to time t. We do not observe the individual after that time.

Right censoring: examples

Examples of right censoring might include:

- A study ending after 1 year. Some individuals will not have had an event before then. In this case, right censoring happens at the same time for all individuals.
- Dropouts: some patients may stop showing up for appointments. In this case, right censoring happens at different times among individuals.
- Competing risks: something happens to an individual that prevents the event of interest from happening.
  - E.g. In a study on heart disease mortality, an individual dies of a different cause.

Note: we can have different kinds of right censoring in the same study.

Right censoring: mechanism

The censoring mechanism itself can be thought of as a random variable.

Let X be the time to event, and C_r be the censoring time; both X and C_r are random variables.

The **observed** time T is then:

$$T = \min(X, C_r)$$

We also have an event/censoring indicator δ, such that:

$$\delta = \begin{cases} 0 & \text{if } T = C_r \\ 1 & \text{if } T = X \end{cases}$$

Left censoring

In **left censoring**, the event time is **less** than the censoring time.

Left censoring: example

Examples of left censoring might include:

- In a study on time to first cannabis use in high school students, a student might answer "I've used cannabis, but I don't remember exactly when the first time was".
- When estimating time to first COVID-19 infection, we may know that some individuals have had COVID-19 before a certain date (e.g. the beginning of the study), but we don't know exactly when.

Left censoring: mechanism

The left censoring mechanism is similar to before:

Let X be the time to event, and C_l be the censoring time; both X and C_l are random variables.

The **observed** time T is then:

$$T = \max(X, C_l)$$

The event/censoring indicator δ, is defined as before:

$$\delta = \begin{cases} 0 & \text{if } T = C_l \\ 1 & \text{if } T = X \end{cases}$$

Doubly censored data

We can have a study with **both** left and right censoring. That is, some observations are left censored, and some are right censored. This is called **doubly censored** data:

Doubly censored data: example

Example of doubly censored data:

- In a study on time to first cannabis use in high school students, a student might answer "I've used cannabis, but I don't remember exactly when the first time was" (left censoring). Another student might answer "I've never used cannabis"; if they use it in the future, then we have right censoring.
- When estimating time to first COVID-19 infection, we may know that some individuals have had COVID-19 before a certain date (e.g. the beginning of the study), but we don't know exactly when (left censoring). Individuals who did **not** get COVID-19 by the end of the study would be right censored.

Doubly censored data: mechanism

Here's the mechanism for doubly censored data:

Let X be the time to event, C_l the *left* censoring time, and C_r the *right* censoring time.

The **observed** time T is then:

$$T = \max(\min(X, C_r), C_l)$$

Now, the event/censoring indicator δ, is defined differently:

$$\delta = \begin{cases} -1 & \text{if } T = C_l \\ 0 & \text{if } T = C_r \\ 1 & \text{if } T = X \end{cases}$$

Interval censoring

In **interval censoring**, we know the event happens **between** two times. That is, they fall inside some interval (L, R).

Interval censoring: example

Examples of interval censoring might include:

- We want to estimate time to disease onset. We might know that a patient developed the disease between two doctor's appointments a year apart, but we might not know the *exact* date.
- A fault is discovered in an industrial machine. We don't know exactly when the fault happened, we only know it happened sometime before the inspection during which it was discovered and after the last inspection.

Inverval censoring: mechanism

In a interval censored observation, the mechanism is as follows:

The time to event is X. We know that $X \in (L, R)$; i.e. we know the event time falls within this range.

For all individuals, we observe whether the event has occurred as of time R.

Left truncation

A very important quality that is sometimes present in time to event data is **truncation**.

The first type of truncation is **left truncation**. An individual is left truncated at time Y_l if they have an event **before** time Y_l and they do **not** appear in the study.

- This is **not** left censoring. In left censoring, the event happens before some censoring time C_l, but **the individual still appears in the study**.
- With left truncation, **we never see this individual in our data**.
- The existence of left truncated observations depends on domain knowledge.

Left truncation

Example of left truncation:

- A researcher wants to estimate survival times by considering a sample of retirement home residents. However, certain residents will have died before they can even move in to the retirement home. The move-in date is the truncation event.

Left truncation mechanism

In left truncation, we only see individuals in the study when:

$$X > Y_l$$

The observed data are **conditional** on the above statement. When we construct likelihood for survival data (later), we will have to take this conditioning into account.

Right truncation

The second type of truncation is **right truncation**. An individual is right truncated at time Y_r if they have an event **after** time Y_r and they do **not** appear in the study.

- Again, this is **not** right censoring.
- In right censoring, we still see the individual in the study; in right truncation, we never see the individual.

Right truncation: example

Example of right truncation:

- Consider a study from Lagakos et al. (1988) that estimated time to AIDS onset after HIV acquired from a blood transfusion.
- Individuals were included in the study only if they had developed AIDS by June 30, 1986.
- Individuals who had received a blood transfusion **before** June 30, 1986, but developed AIDS **after** this date would never have appeared in the study. This is **right truncation**.

Right truncation mechanism

In right truncation, we only see individuals in the study when:

$$X < Y_r$$

The observed data are **conditional** on the above statement. When we construct likelihood for survival data (later), we will have to take this conditioning into account.

Final thoughts

There are many types of censoring and truncation.

In fact, we can have **multiple** types of censoring and truncation present in the **same study**!

This makes analysis of survival data **very** complicated in some cases!

Example 1

The Multiple Sleep Latency Test (MSLT) is a diagnostic test for narcolepsy. In the MSLT, patients take 4 naps throughout the day. In each session, a technician records the time to sleep onset; if the patient doesn't fall alseep within 20 minutes, the session is ended.

Example 2

In a marathon, a time keeper is hired to record the runners' times at the finish line. However, the time keeper arrives late and 4 people have already finished the marathon and left.

Example 3

A researcher wants to estimate the mean time to death among skiers buried in an avalanche. In each case, the researcher has the time of the avalanche, the time that the individual was extracted, and the status of the individual when they were extracted (alive/deceased).