Likelihood Function for Censored Data

Likelihood function for right censored data

- For right censored data, the contribution to the likelihood function are
  - $f(t)$ for uncensored observation
  - $S(t)$ for right censored observation
- The likelihood function is given as

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

Left censoring

- The data could be left censored with only information we know is $T^0 < c$.
- Suppose we are interested in age at which high school students start smoking.
- Any student who started smoking before the study could be left-censored.

Left censoring

- For left censored data, the contribution to the likelihood is
  $$P(T_0 \leq c) = 1 - S(c).$$
- It is the probability that the event occurs before censoring time c.

Interval censoring

- If the data is interval censored, we only know
  $$C_l \leq T^0 \leq C_r.$$
- For example, HIV patients are tested annually.
- If a patient was tested positive in year 2017 and negative in year 2016, all we know is that the infection occurred between two visits.

Interval censoring

- For interval censored data, the contribution to the likelihood is
  $$P(C_l \leq T^0 \leq \ C_r) = S(C_l) - S(C_r).$$

Likelihood construction for other types of censoring

- Contribution to the likelihood function for various types of censoring
  - Exact lifetime $f(t)$
  - Right censored observation $S(C_r)$
  - Left censored observation $1 - S(C_l)$
  - Interval censored observation $S(C_l) - S(C_r)$

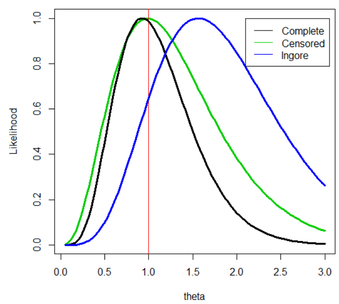
Likelihood construction for other types of censoring

- If data contains different types of censoring, the full likelihood is the product of likelihood contribution from different soures.
- For example, if data contains observed events, right censored and left censored data, then $L \propto \prod_{i \in D} f(t_i) \prod_{i \in R} S(c_i) \prod_{i \in L} [1 - S(c_i)]$
  - D is the set of events
  - R is the set of right censored observations
  - L is the set of left censored observations.

An example

- Suppose true survival times T_i^0 i.i.d.$\sim Exp(1)$.
- The true survival times are
  {0.2, 0.4, 0.5, 1.8, 2.5}.
- Suppose any observation is left censored if less than 0.3, and is right censored if larger than 1.
- We observe the following censored data
  {(0.3, L), 0.4, 0.5, (1,R), (1,R)}.
- The likelihood function is given as

$$L(\theta) = (1 - S_\theta(0.3)) f_\theta(0.4) f_\theta(0.5) S_\theta(1)\, S_\theta(1)$$
$$= (1 - \exp(-0.3\theta))\theta^2 \exp(-2.9\theta)$$



Summary

- Have showed how to construct a likelihood in presence of different types of censoring.
- Our likelihood correctly adjusts for censoring.
  - Remains more or less centered on the true value
  - But the range of likely values is broader since less information contained in censored data
- Ignoring censoring can lead to biased inference.