Likelihood construction

Patrick Breheny

September 1

Introduction

- As we remarked at the outset, survival data is typically incompletely observed (censored); as a result, estimation of moments is not possible
- Likelihood, on the other hand, is a highly versatile tool for quantifying whether a parameter value is consistent with the data; this versatility makes it particularly well-suited to survival analysis
- For this reason, virtually all methods for analyzing survival data depend, at least to some extent, on likelihood principles

The virtues of likelihood

- In particular, an inevitable fact of survival data is that some failure times are observed, while others are only partially observed
- As we will see, the concept of likelihood is well-defined in both cases, and naturally captures the partial information contained in partial observations
- In addition, there is a simple, natural way of combining the information from different types of likelihood – essential for combining the information from fully- and partially-observed subjects

Likelihood: Definition

- Let X denote observable data, and suppose we have a probability model that relates potential values of X to an unknown parameter θ
- Given observed data $X = x$, the *likelihood function* for θ is defined as

$$L(\theta) = \mathbb{P}(x|\theta)$$

- Note that this is a function of θ, not x; now that we have observed the data, x is fixed
- Also, note that a likelihood function is not a probability distribution – for example, it does not have to integrate to 1

Likelihood for continuous distributions

- The definition on the previous slide implicitly assumes discrete data; for continuous distributions, $\mathbb{P}(X = x|\theta)$ is replaced by $f(x|\theta)$, where f is the density function
- Why is this reasonable?
- Suppose, instead of the density, we replaced $\mathbb{P}(X = x)$ with $\mathbb{P}\{X \in (x - \epsilon/2, x + \epsilon/2)\}$; then for small ϵ we have

$$\begin{aligned} L(\theta) &= \int_{x-\epsilon/2}^{x+\epsilon/2} f(u|\theta)du \\ &\approx \epsilon f(x|\theta) \end{aligned}$$

- Thus, at least in the limit $\epsilon \to 0$, the value of ϵ is just an arbitrary multiplicative constant and may be ignored

Fully observed data

- To get a sense of how likelihood works, particularly in the presence of censoring, let's work with the simple survival distribution we introduced in the previous lecture: the exponential distribution
- In particular, our probability model is

$$T_i \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$$

- Suppose we observe the following data:

$$\mathbf{t} = \{0.1, 0.5, 0.5, 1.6, 2.7\}$$

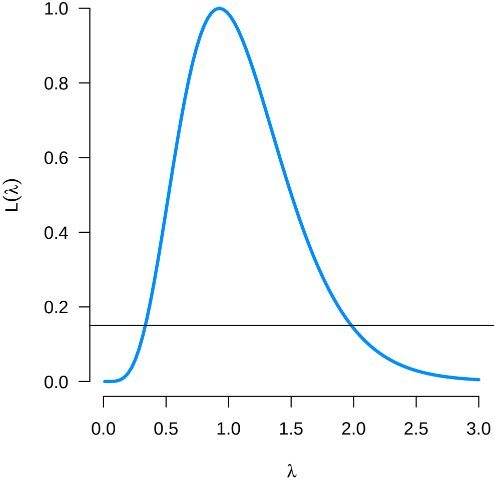
Fully observed data (cont'd)

- The likelihood is therefore

$$L(\lambda) = \prod_i f(t_i|\lambda),$$

  where $f(t_i|\lambda) = \lambda \exp(-\lambda t_i)$
- Likelihoods provide only a relative measure of preference for one parameter value vs. another
- In other words, the actual value of $L(\lambda)$ is not meaningful, but the relative quantity $L(\lambda_1)/L(\lambda_2)$ is meaningful
- For this reason, in all the plots for today, I will standardize L to have a maximum of 1

Likelihood: Fully observed data



Likelihood for censored data

- Now let's consider the situation in which some of that data is censored; in particular, suppose that the study was stopped at time $x = 1$
- For $\{t_1, t_2, t_3\} = \{0.1, 0.5, 0.5\}$, the likelihood remains the same
- For t_4 and t_5, however, the likelihood is now

$$\begin{aligned}\mathbb{P}(T > 1|\lambda) &= S(1|\lambda) \\ &= e^{-\lambda}\end{aligned}$$

Likelihood for censored data (cont'd)

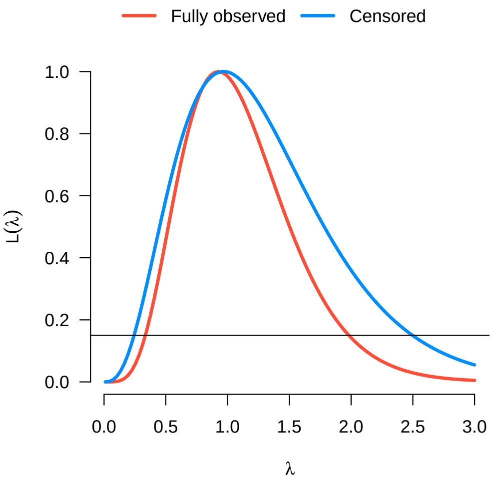
- Combining these likelihood is straightforward:

$$L(\lambda) = \prod_i L_i(\lambda),$$

  where $L_i(\lambda)$ is the contribution to the likelihood from the ith subject
- In other words,

$$L(\lambda) = \prod_{i=1}^{3} f(t_i|\lambda) \prod_{i=4}^{5} S(1|\lambda)$$

Likelihood with censored data



Comments

- Thus, what we learn in the two cases is more or less compatible, although the information is more concentrated in the fully observed case
- This makes perfect sense; as we lose information, the range of likely values of λ should become more broad

Right censoring

- This type of censoring, where it is only known that $T > t$ for some observations, is known as *right censoring*
- It is by far the most common type of censoring, and will be the primary focus of this course
- However, it is not the only of censoring possible; to see how likelihood works for other types of partially observed data, we will now examine various other possible types of censoring

Left censoring: Example

- The data could be *left censored*, meaning that for some observations, all we only know is that $T < t$
- For example, suppose we were studying the age at which teens start smoking, and suppose we start tracking students in high school
- Any student who started smoking before they entered high school would be left-censored

Left censoring: Contribution to likelihood

- In this case, the contribution to the likelihood from an observation left-censored at time t would be

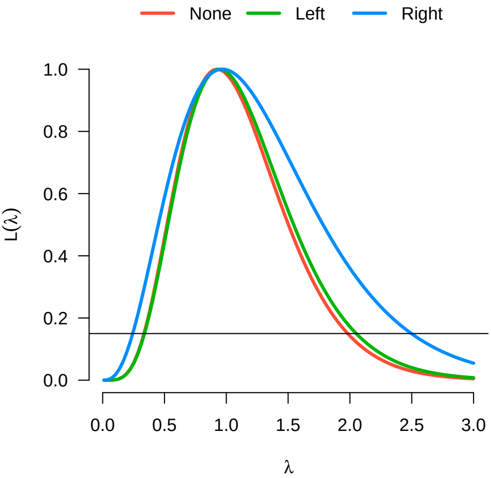
$$L_i(\lambda) = F(t|\lambda);$$

in the special case of the exponential distribution,

$$L_i(\lambda) = 1 - e^{-t\lambda}$$

- For our hypothetical exponential data, suppose observations 1-3 were left-censored at $t = 0.75$

Likelihood with left/right censored data



Interval censoring: Example

- Yet another possibility is that the data could be *interval censored*, meaning that for each time T, we only know an interval $[L, U]$ such that $L < T < U$
- For example, suppose a patient is regularly screened for cancer at 2-year intervals (age 60, 62, 64, . . .), and we first detect a tumor at age 64
- Obviously, the patient did not develop the tumor on the day of the screening; all we know is that the tumor developed sometime between ages 62 and 64

Interval censoring: Example

- In this case, the contribution to the likelihood is

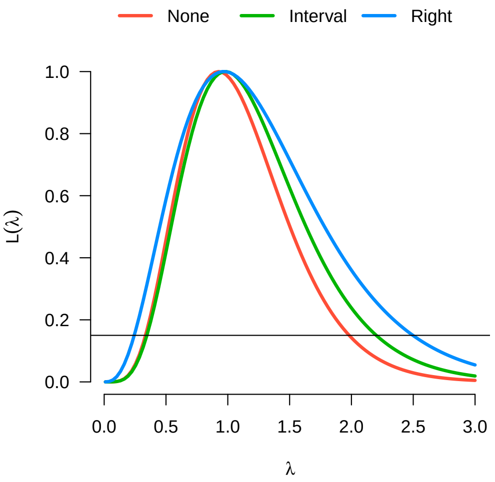
$$L_i(\lambda) = F(U|\lambda) - F(L|\lambda);$$

  in the special case of the exponential distribution,

$$L_i(\lambda) = e^{-L\lambda} - e^{-U\lambda}$$

- In our exponential example, suppose we only observe the times within intervals $[0, 1], [1, 2], [2, 3]$, and so on

Likelihood with interval censored data



Double censoring

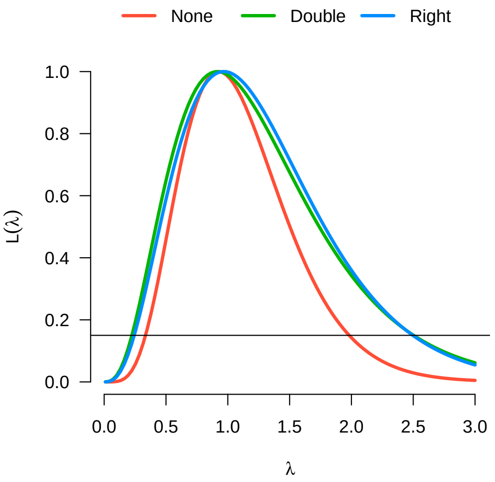
- As an alternative scenario, suppose that we only get to see whether $T < 1$ or not
- This is basically a special case of interval censoring, in that we only see whether an observation is in the interval $[0, 1]$ or the interval $[1, \infty)$
- This situation is known as *double censoring*
- As an example, suppose that in our smoking study from earlier, we only ask each subject once if they have tried smoking yet, and do not follow anyone over time; the data would be double censored

Double censoring: Contribution to likelihood

In the doubly censored case, the contributions to the likelihood are

$$L_i(\lambda) = F(1|\lambda) \qquad \text{for } i = 1, 2, 3$$
$$L_i(\lambda) = S(1|\lambda) \qquad \text{for } i = 4, 5$$

Likelihood with interval censored data, Scenario #2



Atomic radiation example

- Let us now consider a different type of phenomenon
- Suppose we were studying the survival of individuals exposed to radiation from the 1945 atomic bombings of Hiroshima and Nagasaki
- Ideally, of course, we would follow people immediately from 1945 onwards; obviously, that is a bit unrealistic in this case
- Suppose we were unable to enroll people in the study and begin to track their survival until 1950

Atomic radiation example (cont'd)

- In this scenario, anyone who died prior to 1950 would be missing from our sample
- This is different from left censoring, however
- In left censoring, we knew that there was a specific individual with a failure time $T < t$
- In this new scenario, however, people who die prior to 1950 are never enrolled in our study; indeed, we have no direct evidence that they exist at all

Truncation & Likelihood

- This new scenario is known as *truncation*; specifically, the case that we cannot observe an event if $T < t$ is known as *left truncation*
- What is the likelihood contribution in this case?

$$\begin{aligned} L_i(\lambda) &= f(t_i|T > u; \lambda) \\ &= \frac{f(t_i|\lambda)}{S(u|\lambda)}, \end{aligned}$$

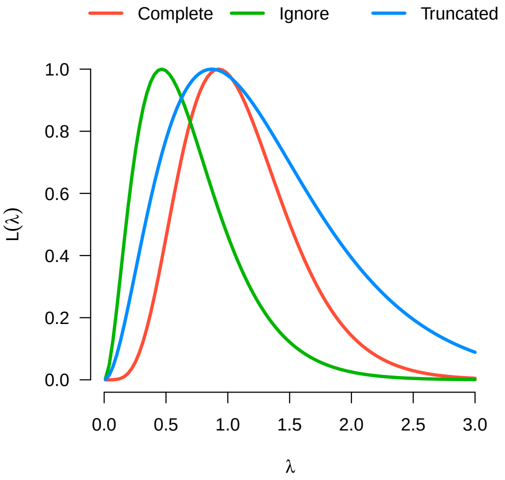
where u is the truncation time

- Note that each actual observation i gets inflated here (division by a number less than 1), because each observation implies a certain number of additional subjects that were unable to be observed

Truncation: Exponential example

- To get a sense of how truncation works, let's suppose our exponential data was truncated at $u = 1$
- Thus, we only have two observations: $\{1.6, 2.7\}$; we don't even know that subjects 1-3 exist
- Let's look at two likelihoods: the one that adjusts for truncation, as in the previous slide, and one that ignores the issue of truncation and just acts as if the observed sample was a simple random sample

Likelihood: Truncation



Remarks

- Adjusting for truncation does the appropriate thing
  - Inference remains more or less centered on the values it should be
  - But the range of likely values is broader since we have less information
- On the other hand, when we ignore truncation, our sample is clearly biased and our inference reflects that
- Left truncation is actually quite common outside of survival analysis as well, since there are often detection thresholds; for example, in astronomy, we cannot observe a star unless it is sufficiently bright

Right truncation

- Finally, *right truncation* is also possible; here, we cannot observe an event unless $T < t$
- For example, suppose we are studying the time until an HIV+ patient develops AIDS, but that we only become aware of such patients when they are actually diagnosed with AIDS
- Clearly, this sampling design will be skewed towards an over-representation of short incubation times

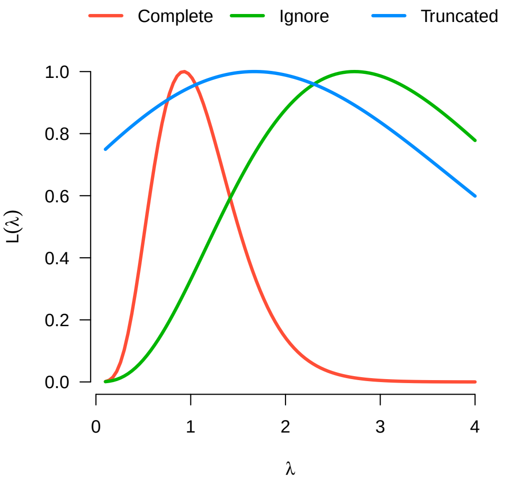
Likelihood for right truncation

- The derivation of the likelihood contribution is similar to the left truncation case:

$$\begin{aligned} L_i(\lambda) &= f(t_i|T < v; \lambda) \\ &= \frac{f(t_i|\lambda)}{F(v|\lambda)}, \end{aligned}$$

  where v is the right truncation time
- As an example, let's see what happens to the likelihood from our exponential example if survival times above 1 are truncated
- As before, we'll consider the ideal complete-data likelihood, the truncation-adjusted likelihood, and the likelihood we get from ignoring truncation

Right truncation



Comments

- Again, when we ignore truncation, we are stuck with the bias of the sampling design
- In this particular case, however, the data don't contain enough information to perform a meaningful adjustment for truncation – if we can't see samples with failure times over 1, we have no idea what λ is unless we collect a lot more data

Summary

Type	T	L_i
Direct observation	$T = t_i$	$f(t_i)$
Right censoring	$T > t_i$	$S(t_i)$
Left censoring	$T < t_i$	$F(t_i)$
Interval censoring	$l_i < T < r_i$	$F(r_i) - F(l_i)$
Left truncation	$T = t_i \| T > u$	$f(t_i)/S(u)$
Right truncation	$T = t_i \| T < v$	$f(t_i)/F(v)$

Final remarks

- Today we have seen how to construct a likelihood in the presence of various kinds of censoring and truncation
- Next time, we'll go into a bit more depth about the implicit assumptions we're making when we do this, and think about some situations in which they might be violated