TMA4275 LIFETIME ANALYSIS

Slides 7: Introduction to parametric inference in lifetime models

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

https://www.ntnu.edu/employees/bo
bo.lindqvist@ntnu.no

NTNU, Spring 2020

- *Parametric distributions*
- *Types of censoring*
- *Representation of censored data*
- *Likelihood construction for arbitrary censored data*

- Smallest extreme (Gumbel)
- Weibull
- 3-parameter Weibull
- Exponential
- 2-parameter exponential
- Normal
- Lognormal
- 3-parameter Lognormal
- Logistic
- Log logistic
- 3-parameter Loglogistic

A model for a lifetime T is called *parametric* if it is given on the form $f(t;\theta)$, $F(t;\theta)$, etc., for functions which are “fixed” except for a parameter value θ which is allowed to vary in some prespecified interval or area. *Note:* θ may be a vector of several parameters.

Examples:

- $f(t;\theta) = \frac{1}{\theta} e^{-t/\theta}$, $F(t;\theta) = 1 - e^{-t/\theta}$; $\theta > 0$
- $f(t;\alpha,\theta) = \frac{1}{\theta}\left(\frac{t}{\theta}\right)^{\alpha-1} e^{-(t/\theta)^{\alpha}}$, $F(t;\alpha,\theta) = 1 - e^{-(t/\theta)^{\alpha}}$; $\alpha > 0, \theta > 0$

Aim: *To estimate or test hypotheses about the true* θ *in a sample (possibly censored) of observations of* T*.*

On the other hand, a model for a lifetime T is called *nonparametric* if it allows any shape of the functions f and F.

- Complements nonparametric techniques.
- Parametric models can be described concisely with just a few parameters, instead of having to report an entire curve.
- It is possible to use a parametric model to extrapolate (in time) to the lower or upper tail of a distribution.
- Parametric models provide smooth estimates of failure time distributions. In practice it is often useful to compare various parametric and nonparametric analyses of a data set.

Lifetime data typically include *censored* data, meaning that:

- some lifetimes are known to have occurred only within certain intervals.
- The remaining lifetimes are known exactly.

Categories of censoring:

- right censoring
- left censoring
- interval censoring

Right censoring is the most common way of censoring. Different subtypes of right censoring can be considered. A common way of presenting right-censored data is as follows:

n units are observed, with potential i.i.d. lifetimes $T_1, T_2, \cdots, T_n$. For each i, we observe a time Y_i which is either the true lifetime T_i, or a censoring time $C_i < T_i$, in which case the true lifetime is "to the right" of the observed time C_i.

The observation from a unit is the pair (Y_i, δ_i) where the *censoring indicator* δ_i is defined by

$$\delta_i = \begin{cases} 1 & \text{if} \quad Y_i = T_i \\ 0 & \text{if} \quad Y_i = C_i, \text{ in which case it is known that } T_i > Y_i \end{cases}$$

LEFT CENSORING

The lifetime T_i for the ith individual is *left censored* if it is *less than* a censoring time C_i, that is, the event of interest has already occurred for the individual before that person is observed in the study at time C_i.

In this case we *observe* the pair (Y_i, ϵ_i), where

$$
\begin{aligned}
Y_i &= \max(T_i, C_i) \\
\epsilon_i &= \begin{cases} 1 & \text{if} \quad T_i \geq C_i \\ 0 & \text{if} \quad T_i < C_i \end{cases}
\end{aligned}
$$

.

(From *Klein and Moeschberger).* In a study to determine the distribution of the time until first marijuana use among high school boys in California, the question was asked, "When did you first use marijuana?" One of the responses was "I have used it but can not recall just when the first time was." A boy who chose this response is indicating that the event had occurred prior to the boy's age at interview but the exact age at which he started using marijuana is unknown. This is an example of a left-censored event time.

If, on the other hand, the answer was "I never used it". What kind of censoring would this correspond to?

When the lifetime is only known to occur *within an interval*, the lifetime is said to be *interval censored.*

Example: Suppose that patients in a clinical trial have *periodic follow-up* and the patient's event time T_i is only known to fall in an interval $(L_i, R_i]$.

Interval censoring may also occur in industrial experiments where there is periodic inspection for proper functioning of equipment items.

Assume we have data for n units with *potential lifetimes* $T_1, T_2, \cdots, T_n \sim_{i.i.d.} f(t;\theta)$.

Noncensored unit: Record the failure time T_i (*ideal case*)

Censored unit: Exact lifetime T_i is not recorded; all we know is that $T_i \in [a, b]$ for an interval of times.

Here

- a is the observed time, and $b = \infty$ for *right censorings*
- $a = 0$, while b is the observed time for *left censorings*
- $0 < a < b < \infty$ for an interval censoring between the observed interval limits a and b

Data for censored data are entered as follows:

Unit no	start variable	end variable	Frequency (optional)
1	a_1	b_1	f_1
2	a_2	b_2	f_2
3	a_3	b_3	f_3
⋮	⋮	⋮	⋮

NOTE: An uncensored observation is entered by letting both a_i and b_i equal the observed lifetime.

LIKELIHOOD CONSTRUCTION: EXAMPLE

Obs. type	Lower bound a_i	Upper bound b_i	Likelihood contribution
Exact lifetime	1.7	1.7	$f(1.7; \theta)$
Right cens.	2.0	∞	$1 - F(2.0; \theta)$
Left cens.	0	0.5	$F(0.5; \theta)$
Interval cens.	1.0	1.5	$F(1.5; \theta) - F(1.0; \theta)$

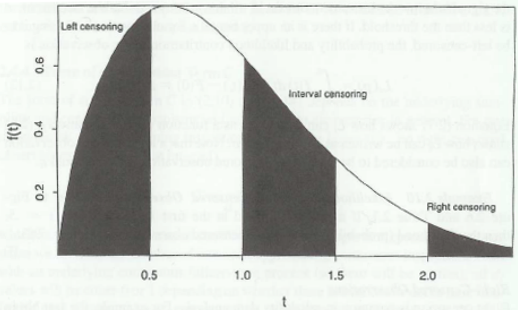


Figure 2.6. Likelihood contributions for different kinds of censoring.

With general censoring, each observed time T_i is represented in our data as $\{T_i \in [a_i, b_i]\}$.

The likelihood function is then defined as

$$\begin{aligned}
L(\theta) &= \text{Probability of gettting the observed data under parameter } \theta \\
&= P_\theta(T_1 \in [a_1, b_1] \cap \cdots \cap T_n \in [a_n, b_n]) \\
&= P_\theta(T_1 \in [a_1, b_1]) \cdots P_\theta(T_n \in [a_1, b_1]) \\
&= (F(b_1; \theta) - F(a_1; \theta)) \cdots (F(b_n; \theta) - F(a_n; \theta)) \\
&= \prod_{i=1}^{n} (F(b_i; \theta) - F(a_i; \theta))
\end{aligned}$$

Generally, the information $T_i \in [a_i, b_i]$ contributes by $F(b_i; \theta) - F(a_i; \theta)$.

Special cases:

- *Right censoring:* Here $b_i = \infty$, so the contribution to likelihood function is

$$F(\infty; \theta) - F(a_i; \theta) = 1 - F(a_i; \theta) = R(a_i, \theta)$$

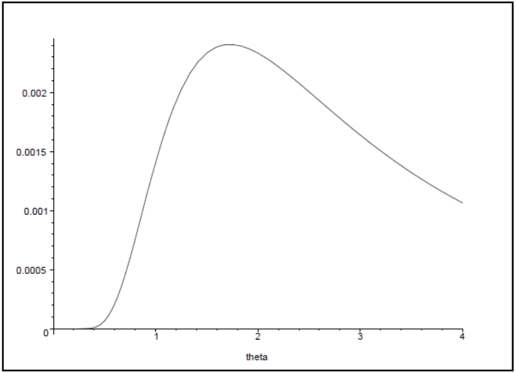
- *Left censoring:* Here $a_i = 0$, so contribution to likelihood is

$$F(b_i; \theta) - F(0; \theta) = F(b_i, \theta)$$

- *Interval censoring:* Contribution is $F(b_i; \theta) - F(a_i; \theta)$
- *Exact observed lifetime:* Then $a_i = b_i$. Write instead $b_i = a_i + h$, so contribution is $F(a_i + h; \theta) - F(a_i; \theta) \approx f(a_i)h$. So we let the contribution be be just $f(a_i)$.

LIKELIHOOD FOR MODEL $f(t;\theta) = (1/\theta)e^{-t/\theta}$

$$L(\theta) = (\frac{1}{\theta}e^{-1.7/\theta}) \cdot (e^{-2.0/\theta}) \cdot (1 - e^{-0.5/\theta}) \cdot (e^{-1.0/\theta} - e^{-1.5/\theta})$$



Maximum likelihood estimate: $\hat{\theta} = 1.725$

Distribution Analysis, Start = A and End = B

```
Variable Start: A  End: B

Censoring Information     Count
Uncensored value              1
Right censored value          1
Interval censored value       1
Left censored value           1

Estimation Method: Maximum Likelihood

Distribution:   Exponential


Parameter Estimates

                      Standard   95,0% Normal CI
Parameter  Estimate      Error      Lower     Upper
Mean        1,72529   0,998421   0,554978   5,36353

Log-Likelihood = -6,029

Goodness-of-Fit
Anderson-Darling (adjusted) = 4,933


Characteristics of Distribution

                                         Standard   95,0% Normal CI
                            Estimate        Error      Lower     Upper
Mean(MTTF)                   1,72529     0,998421   0,554978   5,36353
Standard Deviation           1,72529     0,998421   0,554978   5,36353
Median                       1,19588     0,692053   0,384682   3,71771
First Quartile(Q1)          0,496336     0,287228   0,159657   1,54299
Third Quartile(Q3)           2,39177      1,38411   0,769363   7,43543
Interquartile Range(IQR)     1,89543      1,09688   0,609706   5,89244
```

LIKELIHOOD FOR RIGHT CENSORED DATA

Recall notation: The observations are (y_i, δ_i) for $i = 1, 2, \ldots, n$, where the *censoring indicator* δ_i is defined by

$$\delta_i = \begin{cases} 1 & \text{if} \quad y_i \text{ is the lifetime } t_i \\ 0 & \text{if} \quad y_i \text{ is a censoring time, so } t_i > y_i \end{cases}$$

Then the likelihood function becomes

$$L(\theta) = \prod_{i:\delta_i=1} f(y_i;\theta) \cdot \prod_{i:\delta_i=0} R(y_i;\theta)$$