SESA6085 – Advanced Aerospace Engineering Management

Lecture 5

2024-2025

Dr David Toal

Censored Data

Censored Data

- Previously we have assumed that we have a set of data with exact failures or measurements known
- Frequently data sets contain incomplete or censored data
- There are three main types of censoring which we will now consider in turn
  - Right censored
  - Left censored
  - Interval censored

Right-Censored Data

- Also called "type 1" censoring
- Consider a reliability test involving n components
- The test proceeds until a given time T
- During this time, T, the failure times of r individual components have been recorded
- When the test is terminated we have n-r components which have not failed
- This type of censoring is generally the most common when analysing reliability data

Right-Censored Data

- Consider tests of 10 servos over a 72 hour period

Failure Times (hrs)
30
32.5
40
41.0
43.0
50.6
57.2
67
>72

Exact failure times for 8 servos

2 servos have not failed after 72 hours (right-censored data)

Left-Censored Data

- This type of censoring is the opposite of right censoring
- Again consider a reliability test involving n components
- The test is commenced at t=0 and left for a period T, when the technician returns l components have failed
- We now have observations missing from the start of the data set, all we know is that the failures occurred before T

We know the number of failures in an interval, but not the exact times

Left-Censored Data

- Consider again tests of 10 different servos

Failure Times (hrs)
<24 hrs
31
35
43.7
50.3
56.1
65.8
70.3
80.1

2 servos have failed in the first 24 hours before observations begin (left-censored)

Exact failure times for 8 servos

Interval-Censored Data

- Interval-censored data is where failures are known to occur between known bounds
- Indeed a reliability test may be made up of only interval-censored data where the technician returns at set time intervals to record the number of failures
- Left-censoring may be considered as an extension of interval censoring to the lower bound $t=0$

MANY checks fall under this category, eg: regular maintenance checks/servicing

Interval-Censored Data

- Consider again tests for a set of different servos

Time Interval (hrs)	No. of Failures Observed
0-10	0
11-20	5
21-30	4
31-40	6
41-50	10
51-60	2

Exact failure times are unknown only the interval

Censored Data Notation

- Let's use MLE to fit a PDF to our censored data
- Let's begin by dividing our data up into censored and uncensored data
  - U – uncensored data subset
  - C – censored data subset
- And divide our censored data up amongst our different types of censoring
  - C_R – Right censored data subset
  - C_L – Left censored data subset
  - C_I – Interval censored data subset

Uncensored MLE

- For completely uncensored data our MLE is given by:

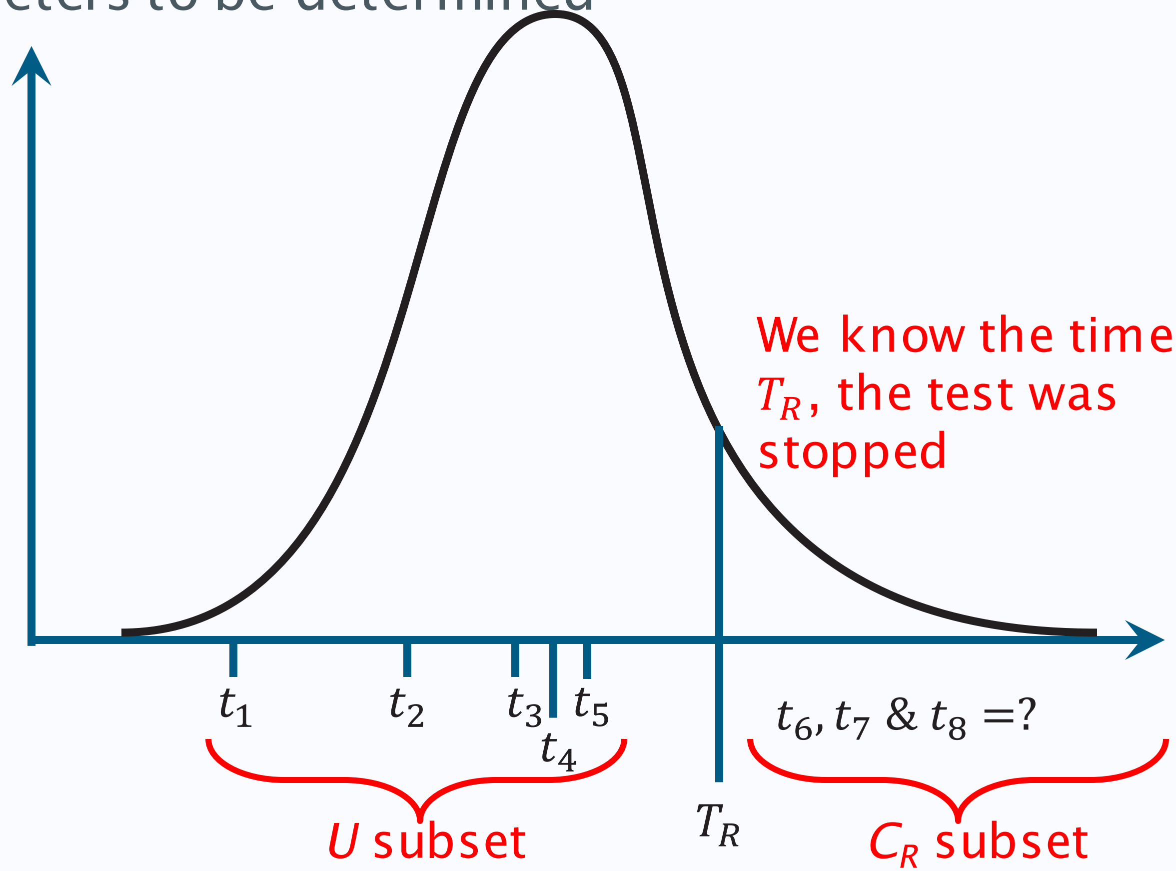
$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta})$$

- Where $f(t_i; \boldsymbol{\theta})$ is our PDF at time t_i whose shape is controlled by some parameters $\boldsymbol{\theta}$
- We saw previously how this can lead to closed form solutions for the parameters

Right-Censored MLE

As usual we start with our curve function we "know" it will fit to

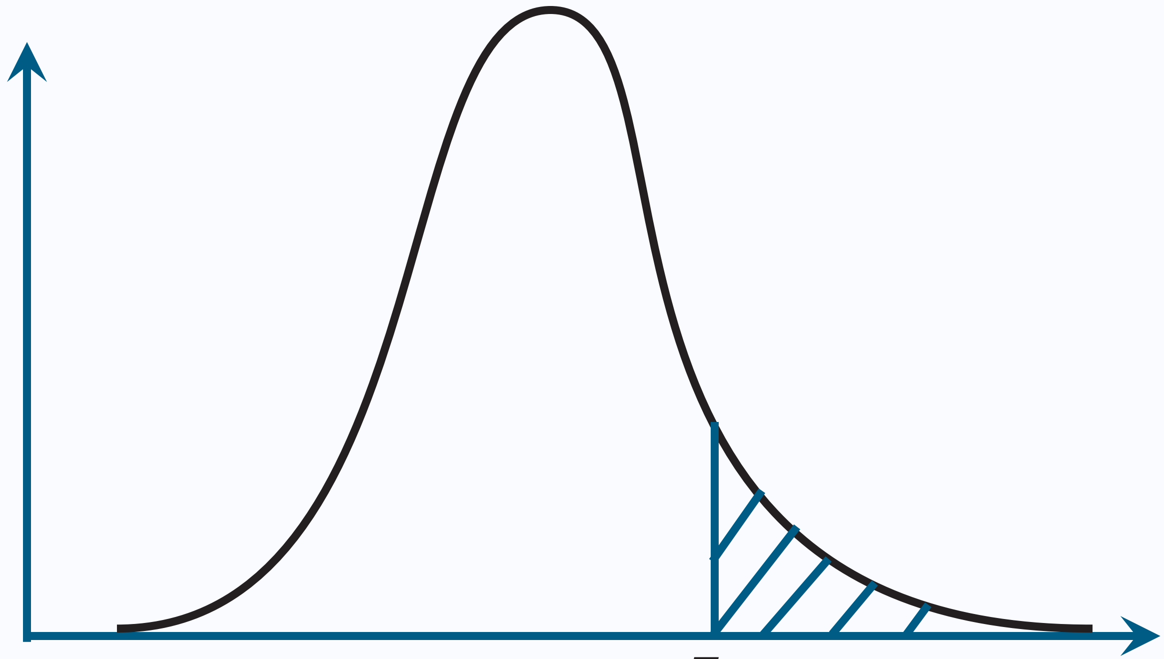
- Consider failure data from a test containing both a U and C_R subset. The test is assumed to follow a known distribution with parameters to be determined



University of Southampton

Right-Censored MLE

- What is the probability that our observations occurred after T_R?



$$P[t > T_R] = 1 - \int_0^{T_R} f(t)dt$$

<--- would be -inf for a guassian

Right-Censored MLE

$$P[t > T_R] = 1 - \int_0^{T_R} f(t)dt$$

- This equation should look familiar?
- This is our reliability function, hence

$$P[t > T_R] = \mathrm{R}(T_R)$$

- Rather than using the PDF for each of our right censored data points we instead use the reliability function for each data point

Right-Censored MLE

- The uncensored MLE

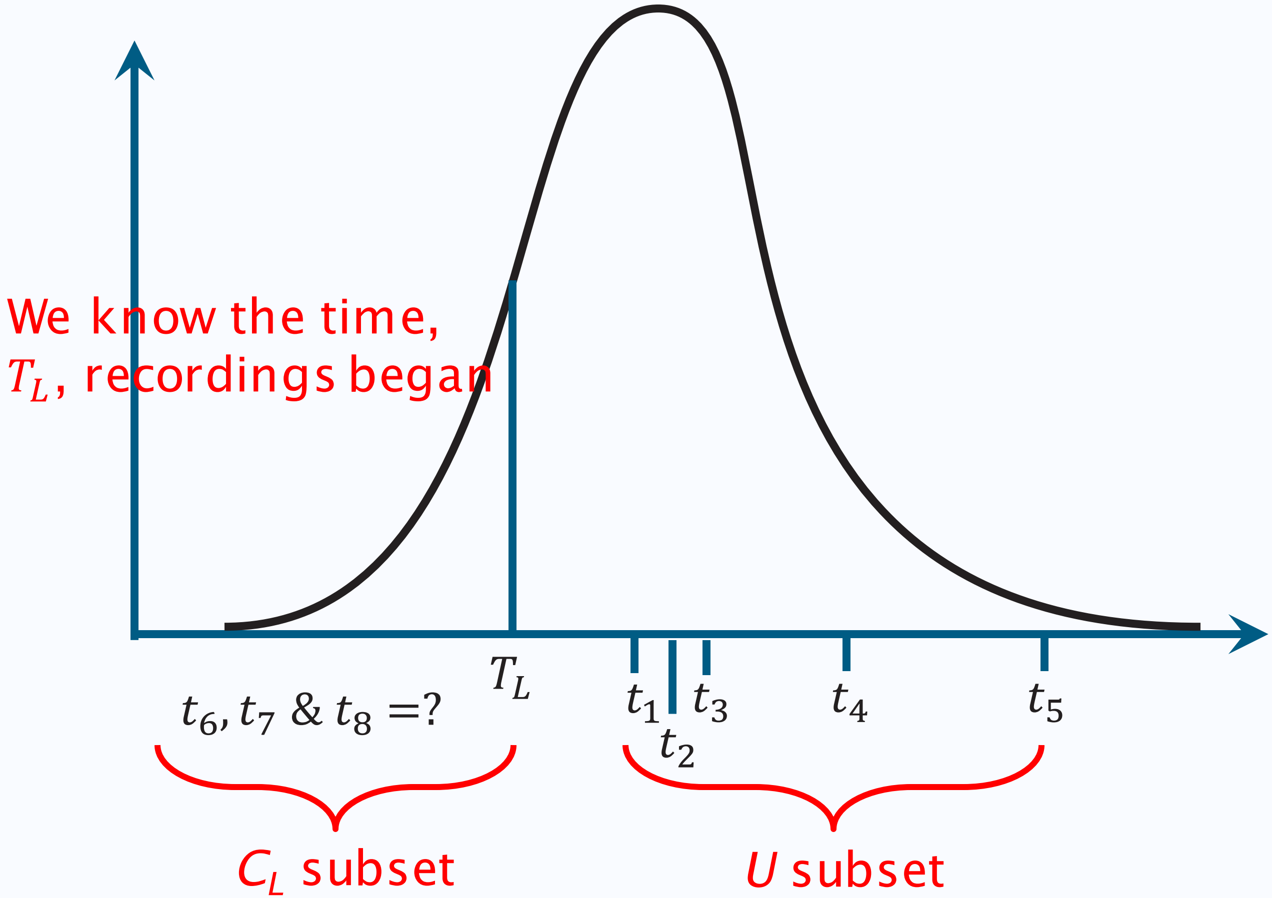
$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta})$$

- Is then modified to include $\mathrm{R}(T_R)$ for each censored point:

$$L(\boldsymbol{\theta}) = \underbrace{\left\{ \prod_{i \in U} f(t_i; \boldsymbol{\theta}) \right\}}_{\text{Uncensored}} \underbrace{\left\{ \prod_{i \in C_R} R(T_R; \boldsymbol{\theta}) \right\}}_{\text{Right-censored}}$$

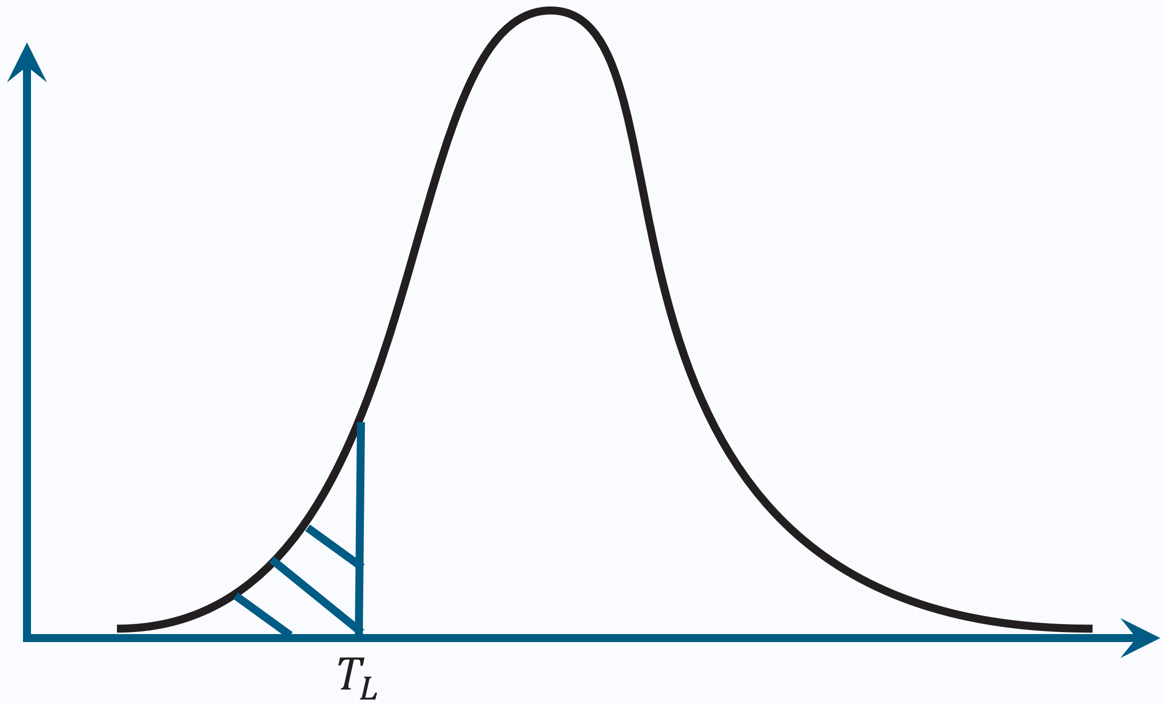
Left-Censored MLE

- Now that we've defined the MLE for right-censored data, how do we define the MLE for left-censored data



Left-Censored MLE

- What is the probability that our observations occurred before T_L?



$$P[t < T_L] = \int_0^{T_L} f(t)dt$$

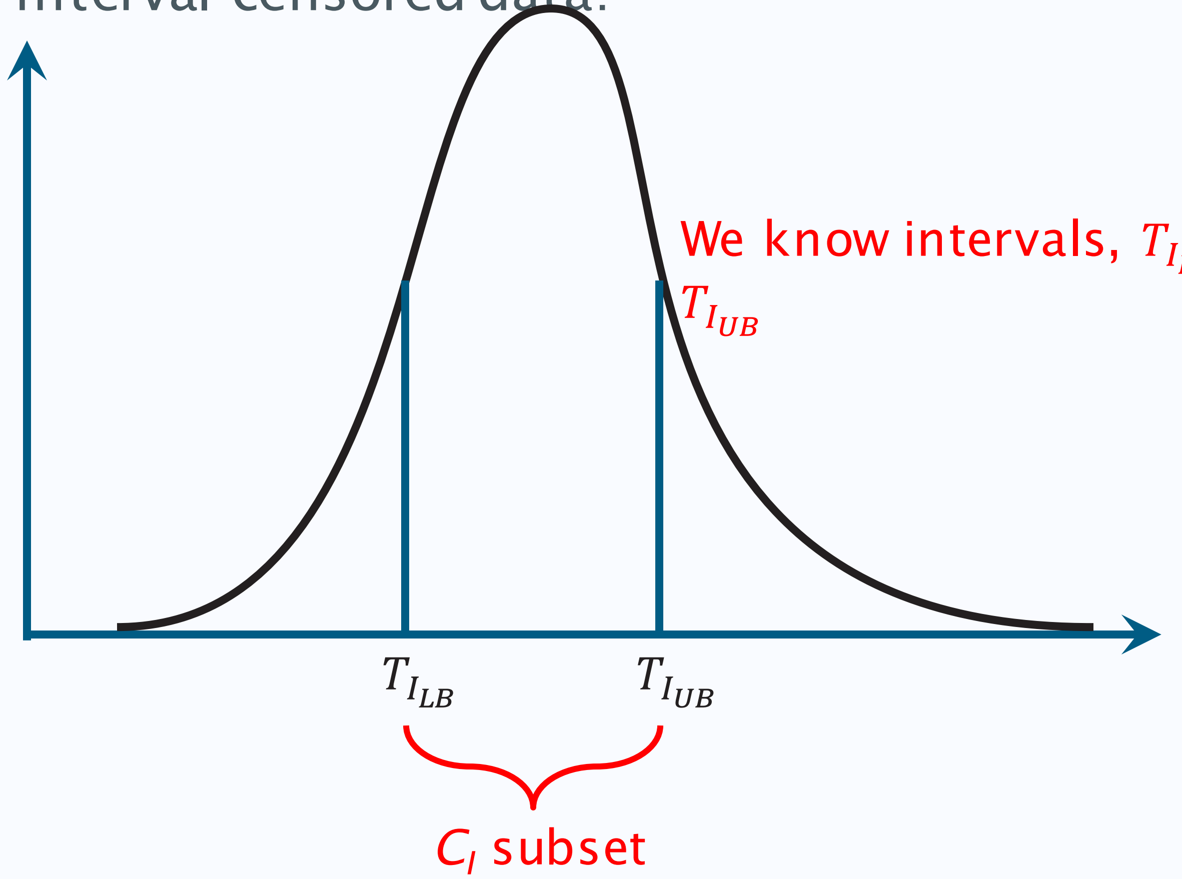
Left-Censored MLE

- Again this equation should look familiar to you as the CDF
- Hence instead of using the PDF for our left censored data we use the CDF
- Our MLE therefore becomes:

$$L(\boldsymbol{\theta}) = \underbrace{\left\{\prod_{i \in U} f(t_i; \boldsymbol{\theta})\right\}}_{\text{Uncensored}} \underbrace{\left\{\prod_{i \in C_L} F(T_L; \boldsymbol{\theta})\right\}}_{\text{Left-censored}}$$

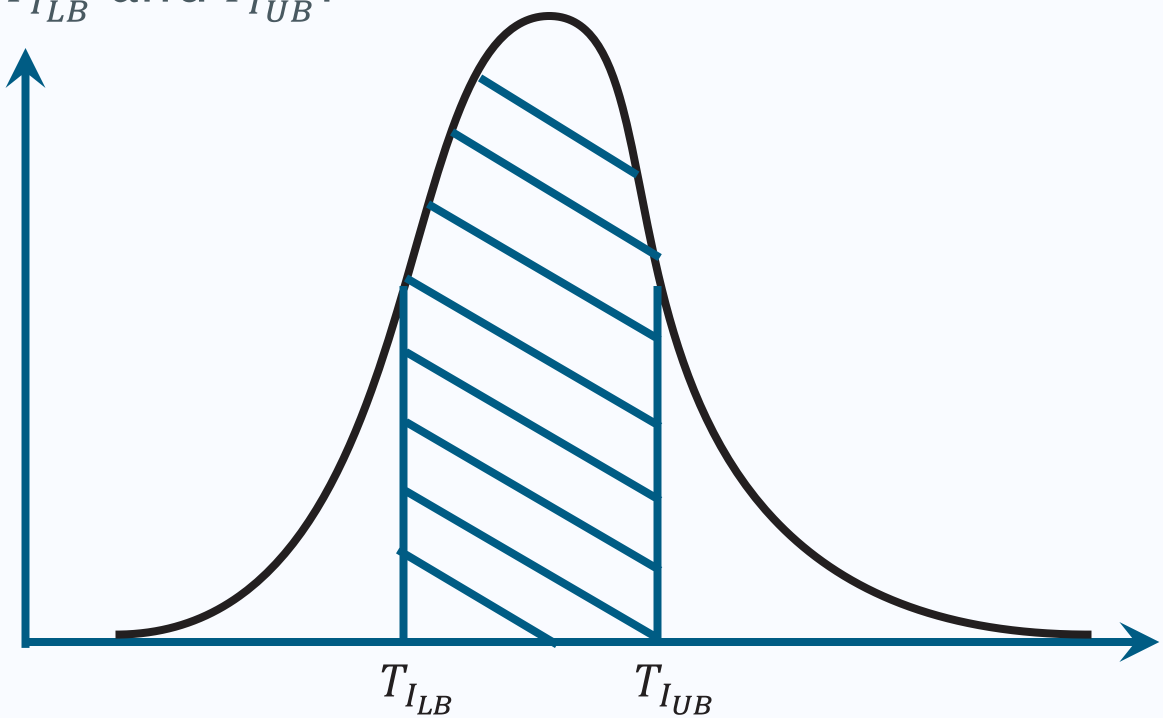
Interval-Censored MLE

- By now you should begin to see a pattern, how do we define the MLE for interval-censored data?



Interval-Censored MLE

- What is the probability that our observations occurred between $T_{I_{LB}}$ and $T_{I_{UB}}$?



$$P[T_{I1} < t < T_{I2}] = \int_0^{T_{I_{UB}}} f(t)dt - \int_0^{T_{I_{LB}}} f(t)dt$$

Interval-Censored MLE

- Again we're employing the CDF of our distribution
- However we now use the CDF for both the lower and upper bound of our interval
- For a single censored interval the MLE becomes:

$$L(\boldsymbol{\theta}) = \underbrace{\left\{\prod_{i\in U} f(t_i;\boldsymbol{\theta})\right\}}_{\text{Uncensored}} \underbrace{\left\{\prod_{i\in C_I} F\left(T_{I_{UB}};\boldsymbol{\theta}\right) - F\left(T_{I_{LB}};\boldsymbol{\theta}\right)\right\}}_{\text{Interval-censored}}$$

- What if there are multiple intervals?

Interval-Censored MLE

- Additional sets of lower and upper bounds can be defined and the probability of points lying within those intervals included in the expression for the likelihood

$$L(\boldsymbol{\theta}) = \left\{\prod_{i \in U} f(t_i; \boldsymbol{\theta})\right\} \times$$

$$\left\{\prod_{i \in C_{I_1}} F\left(T_{I1_{UB}}; \boldsymbol{\theta}\right) - F\left(T_{I1_{LB}}; \boldsymbol{\theta}\right)\right\} \times$$

$$\left\{\prod_{i \in C_{I_2}} F\left(T_{I2_{UB}}; \boldsymbol{\theta}\right) - F\left(T_{I2_{LB}}; \boldsymbol{\theta}\right)\right\} \times \cdots$$

Multiple Types of Censoring

- What if we have a situation where we need to include all three types of censoring?
- In this case the likelihood function is a combination of the three separate functions

$$L(\boldsymbol{\theta}) = \underbrace{\left\{\prod_{i\in U} f(t_i;\boldsymbol{\theta})\right\}}_{\text{Uncensored}} \underbrace{\left\{\prod_{i\in C_L} F(T_L;\boldsymbol{\theta})\right\}}_{\text{Left-censored}} \underbrace{\left\{\prod_{i\in C_I} F\left(T_{I_{UB}};\boldsymbol{\theta}\right) - F\left(T_{I_{LB}};\boldsymbol{\theta}\right)\right\}}_{\text{Interval-censored}} \underbrace{\left\{\prod_{i\in C_R} R(T_R;\boldsymbol{\theta})\right\}}_{\text{Right-censored}}$$

- Clearly depending on the distribution fitting a model to such data can become difficult and necessitates the use of an optimisation routine

Adding censored data to a uncensored dataset genrally means you cannot solve it analytically and need to use a numerical approach.

Censored Models

- Of course, even with multiple types of censoring....
  - We can still apply MLE to determine the optimal parameters for our model

$$L(\theta) = \prod_{i=1}^{n} f(t_i; \theta)$$

  - We can still use the Fisher information matrix to calculate confidence bounds in our parameters

$$I_{ij} = E\left[-\frac{\partial^2 l(t;\theta)}{\partial\theta_i \partial\theta_j}\right]$$

  - We could even have censored joint distribution functions

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} F_i(x_i)$$

University of
Southampton