TMA4275 LIFETIME ANALYSIS

Slides 8: Parametric inference for the exponential model

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

In Slides 8 we consider parametric statistical inference for lifetime distributions with a single parameter θ. The exponential distribution is the example throughout, but most of the theory is valid for any lifetime distribution with a single parameter.

- Likelihood function for right censored data
- Likelihood analysis for exponential distribution
  - Likelihood function
  - Maximum likelihood estimation
  - Observed information
  - Standard error
  - Percentiles
  - MINITAB output
  - Confidence intervals (CI)
    - Standard CI
    - Standard CI for positive parameters
    - CI based on likelihood-function ("the 1.92-interval")
  - Testing hypotheses

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

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

Let $f(t;\theta) = \frac{1}{\theta} e^{-\frac{t}{\theta}}$. Then the likelihood function is

$$\begin{aligned} L(\theta) &= \prod_{i \in \delta_i = 1} f(y_i;\theta) \cdot \prod_{i \in \delta_i = 0} R(y_i;\theta) \\ &= \prod_{i;\delta_{i=1}} \frac{1}{\theta} e^{-\frac{y_i}{\theta}} \cdot \prod_{i;\delta_{i=0}} e^{-\frac{y_i}{\theta}} \\ &= \frac{1}{\theta^{\sum_{i=1}^n \delta_i}} e^{-\frac{\sum_{i=1}^n y_i}{\theta}} = \frac{1}{\theta^r} e^{-\frac{s}{\theta}}, \end{aligned}$$

where $r = \sum_{i=1}^n \delta_i = \textit{number of failures}$, $s = \sum_{i=1}^n y_i = \textit{total time on test}$.

The log-likelihood function is defined as

$$\ell(\theta) = \log L(\theta) = -r \ln \theta - \frac{s}{\theta}$$

The maximum likelihood estimator $\hat{\theta}$ is found from solving

$$\ell'(\theta) = -\frac{r}{\theta} + \frac{s}{\theta^2} = 0$$

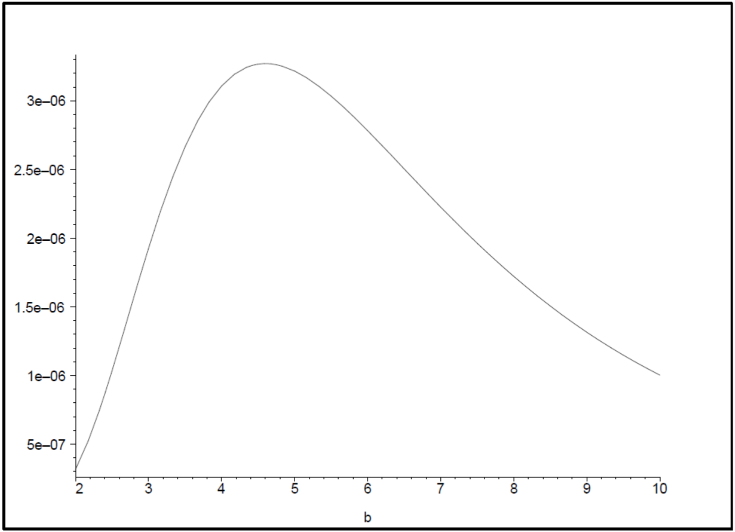
which implies $\hat{\theta} = s/r = \textit{“Exposure”}/\textit{“Occurrence”}$

LIFETIME OF A COMPONENT - RIGHT CENSORED FAILURE DATA FOR A SAMPLE OF 7 COMPONENTS:

y_i	δ_i
0.6	0
0.8	1
2.1	1
3.2	1
3.3	0
4.4	1
8.6	1

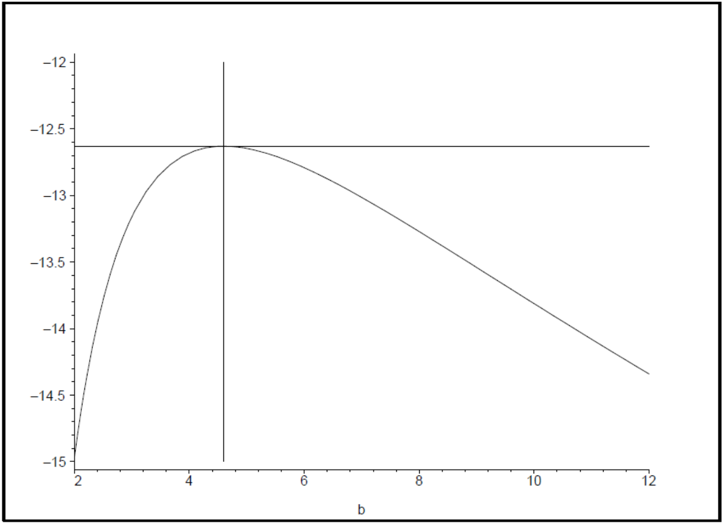
$s = \sum y_i = 23.0$ and $r = \sum \delta_i = 5$,

$\hat{\theta} = 23.0/5 = 4.6$ (*estimated MTTF for the component*)



$$L(\theta) = \frac{1}{\theta^5} e^{-\frac{23.0}{\theta}}; \quad \hat{\theta} = 4.6$$

LOG LIKELIHOOD FUNCTION



$$\ell(\theta) = -5 \ln \theta - \frac{23.0}{\theta}; \quad \hat{\theta} = 4.6$$

Question: **How good is the estimate $\hat{\theta}$?** Consider its **standard error!**

Suppose first there is **no** censoring. Then

$$\hat{\theta} = \frac{\sum_{i=1}^{n} T_i}{n} = \bar{T}$$

Standard computations from basic statistics course give:

$$\begin{aligned} E(\hat{\theta}) &= E(\bar{T}) = E(T_i) = \theta \\ Var(\hat{\theta}) &= Var(\bar{T}) = \frac{Var(T_i)}{n} = \frac{\theta^2}{n} \\ SD(\hat{\theta}) &= \frac{\theta}{\sqrt{n}} \end{aligned}$$

In practice one is interested in *estimating* the standard deviation of the estimator, called STANDARD ERROR = SE:

$$SE = \widehat{SD(\hat{\theta})} = \frac{\hat{\theta}}{\sqrt{n}}$$

Consider a situation with a one-dimensional parameter θ and a log-likelihood function $\ell(\theta)$.

The maximum likelihood estimator $\hat{\theta}$ is the solution of $\ell'(\theta) = 0$.

Let $\ell''(\theta)$ be the double derivative of $\ell(\theta)$.

The **observed information** about θ is defined in general by

$$I(\hat{\theta}) =_{def} -\ell''(\hat{\theta})$$

From statistical theory we can estimate $\text{Var}(\hat{\theta})$ by

$$\widehat{Var(\hat{\theta})} = \frac{1}{I(\hat{\theta})} = -\frac{1}{\ell''(\hat{\theta})}$$

so **standard error** becomes $SE = \widehat{SD(\hat{\theta})} = \sqrt{I(\hat{\theta})^{-1}} = \sqrt{-1/\ell''(\hat{\theta})}$

Recall that $\hat{\theta} = \frac{s}{r}$ and $\ell(\theta) = -r \ln \theta - \frac{s}{\theta}$. Now

$$\ell'(\theta) = -\frac{r}{\theta} + \frac{s}{\theta^2} \qquad \ell''(\theta) = \frac{r}{\theta^2} - \frac{2s}{\theta^3}$$

Thus the observed information is

$$I(\hat{\theta}) = -\ell''(\hat{\theta}) = -\frac{r}{\hat{\theta}^2} + \frac{2s}{\hat{\theta}^3} = -\frac{r}{\hat{\theta}^2} + \frac{2s}{\hat{\theta}^2 \cdot \frac{s}{r}} = \frac{r}{\hat{\theta}^2}$$

and hence

$$\widehat{Var(\hat{\theta})} = \frac{1}{I(\hat{\theta})} = \frac{\hat{\theta}^2}{r}, \qquad SE = \widehat{\text{SD}(\hat{\theta})} = \sqrt{\widehat{Var(\hat{\theta})}} = \frac{\hat{\theta}}{\sqrt{r}} \quad (*)$$

In example: $SE = \widehat{\text{SD}(\hat{\theta})} = \frac{4.6}{\sqrt{5}} = 2.05718$ (see MINITAB)

Note that $(*)$ *is compatible with the result for the non-censored case.*

Distribution Analysis: Y

```
Variable: Y

Censoring Information   Count
Uncensored value            5
Right censored value        2

Censoring value: D = 0

Estimation Method: Maximum Likelihood

Distribution:   Exponential


Parameter Estimates

                     Standard   95,0% Normal CI
Parameter  Estimate     Error    Lower    Upper
Mean            4,6   2,05718  1,91465  11,0516

Log-Likelihood = -12,630

Goodness-of-Fit
Anderson-Darling (adjusted) = 3,767


Characteristics of Distribution

                                         Standard   95,0% Normal CI
                              Estimate      Error      Lower    Upper
Mean(MTTF)                         4,6    2,05718    1,91465  11,0516
Standard Deviation                 4,6    2,05718    1,91465  11,0516
Median                         3,18848    1,42593    1,32713  7,66041
First Quartile(Q1)             1,32334   0,591815   0,550810  3,17936
Third Quartile(Q3)             6,37695    2,85186    2,65427  15,3208
Interquartile Range(IQR)       5,05362    2,26005    2,10346  12,1415
```

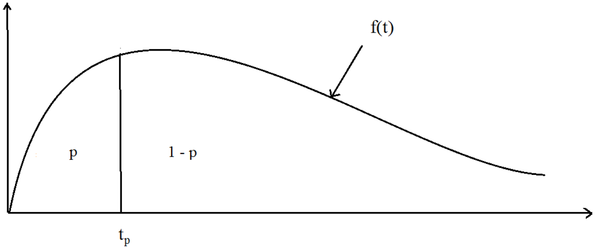
Let T be a lifetime with cdf given by $F(t) = P(T \leq t)$. Then the $100p\%$-percentile, t_p, is defined by

$$F(t_p) = p; \quad 0 < p < 1$$

This means that 100p% of the "components" fail before time t_p.

Special cases:

- Median: $p = 0.50$
- Lower quartile: $p = 0.25$
- Upper quartile: $p = 0.75$
- Interquartile range: $t_{0.75} - t_{0.25}$



$$\begin{aligned}
F(t_p) &= p \\
1 - e^{-\frac{t_p}{\theta}} &= p \\
-\frac{t_p}{\theta} &= \ln(1-p) \\
t_p &= -\theta \ln(1-p)
\end{aligned}$$

Thus, for example,

$$\text{median} = \theta \ln 2 = 0.693\,\theta$$

Estimates in example:

$$\begin{aligned}
\widehat{\text{median}} &= \hat{\theta} \ln 2 = 3.188 \\
\widehat{\text{first quartile}} &= -\hat{\theta} \ln(1-0.25) = 1.323 \\
\widehat{\text{third quartile}} &= -\hat{\theta} \ln(1-0.75) = 6.377 \\
\widehat{\text{interquartile range}} &= 6.377 - 1.323 = 5.054
\end{aligned}$$

For given $0 < p < 1$, we estimate t_p by

$$\hat{t}_p = -\hat{\theta} \ln(1-p)$$

Thus

$$\begin{aligned} \text{Var}(\hat{t}_p) &= (\ln(1-p))^2 \, \text{Var}(\hat{\theta}) \\ \text{SD}(\hat{t}_p) &= -\ln(1-p) \, \text{SD}(\hat{\theta}) \\ SE = \widehat{\text{SD}(\hat{t}_p)} &= -\ln(1-p) \, \widehat{\text{SD}(\hat{\theta})} \end{aligned}$$

In example, the standard error of the estimated median is

$$\text{SE(median)} = \ln 2 \cdot \widehat{\text{SD}(\hat{\theta})} = \ln 2 \cdot 2.05718 = 1.4259$$

THE STANDARD CONFIDENCE INTERVAL

Let $\hat{\theta}$ be the MLE of a (one-dimensional) parameter θ. Then for large n:

$$\frac{\hat{\theta} - \theta}{\widehat{\mathrm{SD}}(\hat{\theta})} \approx N(0,1) \qquad \textit{(asymptotically)}$$

From this,

$$P(-1.96 < \frac{\hat{\theta} - \theta}{\widehat{\mathrm{SD}}(\hat{\theta})} < 1.96) \approx 0.95$$

Rearranging the inequalities we get

$$P(\hat{\theta} - 1.96\, \widehat{\mathrm{SD}(\hat{\theta})} < \theta < \hat{\theta} + 1.96\, \widehat{\mathrm{SD}(\hat{\theta})}) \approx 0.95 \qquad (1)$$

which defines the *standard 95% confidence interval* for θ:

$$\hat{\theta} \pm 1.96\, \widehat{\mathrm{SD}(\hat{\theta})}.$$

(Of course we may change the 1.96 to obtain other percentages than 95%.)

This interval is used by MINITAB.

First, consider the standard confidence interval for $\ln\theta$ based on the estimate $\ln\hat{\theta}$. It is given by

$$P(\ln\hat{\theta} - 1.96\,\widehat{\mathrm{SD}(\ln\hat{\theta})} < \ln\theta < \ln\hat{\theta} + 1.96\,\widehat{\mathrm{SD}(\ln\hat{\theta})}) \approx 0.95.$$

Now express $\widehat{\mathrm{SD}(\ln\hat{\theta})}$ by $\widehat{\mathrm{SD}(\hat{\theta})}$. Taylor's theorem gives

$$\ln\hat{\theta} \approx \ln\theta + \frac{1}{\theta}(\hat{\theta} - \theta),$$

and therefore,

$$\mathrm{Var}(\ln\hat{\theta}) \approx \frac{\mathrm{Var}(\hat{\theta})}{\theta^2} \quad \text{and} \quad \mathrm{SD}(\ln\hat{\theta}) \approx \frac{\mathrm{SD}(\hat{\theta})}{\theta}.$$

Replacing the standard deviation in the previous slide we get

$$P(\ln \hat{\theta} - 1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta} < \ln \theta < \ln \hat{\theta} + 1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}) \approx 0.95.$$

The above inequalities are equivalent to the ones we get by taking the exponential of all terms, i.e. we have,

$$P(e^{\ln \hat{\theta} - 1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}} < e^{\ln \theta} < e^{\ln \hat{\theta} + 1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}}) \approx 0.95$$

or

$$P(\hat{\theta} e^{-1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}} < \theta < \hat{\theta} e^{1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}}) \approx 0.95.$$

This defines **the standard confidence interval for positive parameters**, which in short can be written

$$\hat{\theta}\ e^{\pm 1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}}$$

In example: $4.6 \exp(\pm 1.96 \cdot 2.05718/4.6) = (1.915, 11.052)$

For given $0 < p < 1$, we have

$$t_p = -\theta \ln(1-p)$$

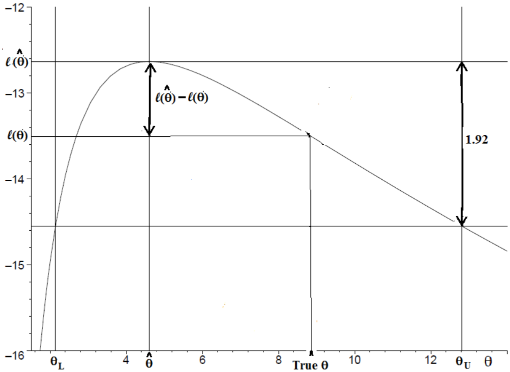
Recall the CI for θ:

$$P(\hat{\theta} e^{-1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}} < \theta < \hat{\theta} e^{1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}}) \approx 0.95.$$

Multiplying each term by $-\ln(1-p)$ we get a CI for t_p from

$$P(-\ln(1-p)\hat{\theta} e^{-1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}} < t_p < -\ln(1-p)\hat{\theta} e^{1.96\, \widehat{\mathrm{SD}(\hat{\theta})}/\hat{\theta}}) \approx 0.95.$$

Check yourself with MINITABs CI for the median and quartiles!



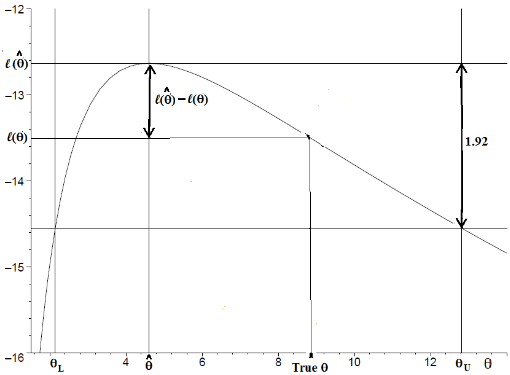
Likelihood theory: If θ is the true value of the parameter, then

$$W(\theta) = 2\big(\ell(\hat{\theta}) - \ell(\theta)\big) \approx \chi_1^2$$

Thus, from table of χ^2-distribution,

$$P(2\big(\ell(\hat{\theta}) - \ell(\theta)\big) \leq 3.84) \approx 0.95,$$

$$\textit{or equivalently, } P(\ell(\theta) \geq \ell(\hat{\theta}) - 1.92) \approx 0.95.$$



Recall that if θ is the true value of the parameter, then

$$P(\ell(\theta) \geq \ell(\hat{\theta}) - 1.92) \approx 0.95.$$

A confidence interval for θ can from this be found (see picture) as the set of θ for which $\ell(\theta) \geq -12.63 - 1.92 = -14.55$. This gives the interval from $\theta_L = 2.1$ to $\theta_U = 12.8$ which is hence an approximate 95% confidence interval.

TESTING HYPOTHESES ABOUT θ BY LOG LIKELIHOOD

Recall the general result:

$$W(\theta) = 2\big(\ell(\hat{\theta}) - \ell(\theta)\big) \approx \chi^2_1$$

Suppose we want to test the null hypothesis

$$H_0 : \theta = \theta_0 \text{ vs. } \theta \neq \theta_0$$

Then under the null hypothesis is $W(\theta_0)$ approximately χ^2_1.

Using significance level 5% we will then reject the null hypothesis if $W(\theta_0) > 3.84$.

With our data we get $W(10) = 2(-12.63 - (-13.81)) = 2.36$, so we do not reject the null hypothesis.

Note also that 10 is inside the 95% confidence interval that we just constructed. In general we can test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ with significance level 5% if we have at hand a 95% confidence interval for θ. *We then reject the null hypothesis if θ_0 is not in the confidence interval.* (Thus with $\theta_0 = 10$, we do not reject.)