TMA4275 LIFETIME ANALYSIS

Slides 10: Estimation in log-location scale families; threshold models; exact confidence interval for type II censoring

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

- Log-location-scale families
    - Likelihood function
    - Observed information
    - Standard error
    - Percentiles
    - Probability plots
    - Data example: Shock absorber data
- Parametric models with threshold parameters
- Exact CI for type II censoring

A lifetime T has a *log-location-scale* family of distributions if
$\ln T$ has a *location- scale* family i.e.

$$\ln T = \mu + \sigma U$$

where U has a "standardized" distributions centered around 0, with values in $(-\infty, +\infty)$.:

- if $U \sim N(0,1)$, then $T \sim \text{lognormal}(\mu, \sigma)$
- if $U \sim logistic(0,1)$, then $T \sim \text{log-logistic}(\mu, \sigma)$
- if $U \sim Gumbel(0,1)$, then $T \sim Weibull(\theta, \alpha)$ with

$$\ln \theta = \mu, \; 1/\alpha = \sigma$$

Below are given, respectively, cdf and pdf of some "standardized" distributions, for $-\infty < u < \infty$.

Generic: $\Psi(u) = P(U \leq u),\ \psi(u) = \Psi'(u)$

Normal: $\Phi(u) = \int_{-\infty}^{u} \phi(x)dx, \quad \phi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$

Logistic: $H(u) = \frac{e^u}{1+e^u},\ h(u) = \frac{e^u}{(1+e^u)^2}$

Gumbel: $G(u) = 1 - e^{-e^u},\ g(u) = e^{u-e^u}$

Let T be distributed as a log-location-scale family with

$$\ln T = \mu + \sigma U$$

Then

$$\begin{aligned} F_T(t) &= P(T \leq t) = P(\ln T \leq \ln t) \\ &= P(\mu + \sigma U \leq \ln t) = P\big(U \leq \frac{\ln t - \mu}{\sigma}\big) \\ &= \Psi\big(\frac{\ln t - \mu}{\sigma}\big) \end{aligned}$$

Thus

$$R_T(t) = 1 - \Psi\big(\frac{\ln t - \mu}{\sigma}\big)$$

and

$$f_T(t) = \psi\big(\frac{\ln t - \mu}{\sigma}\big) \cdot \frac{1}{\sigma t}$$

Likelihood for data from a general log-location-scale family:

$$L(\mu,\sigma) = \prod_{i:\delta_i=1} \psi\left(\frac{\ln y_i - \mu}{\sigma}\right) \cdot \frac{1}{\sigma y_i} \cdot \prod_{i:\delta_i=0} \left(1 - \Psi\left(\frac{\ln y_i - \mu}{\sigma}\right)\right)$$

and log-likelihood is

$$\ell(\mu,\sigma) = \sum_{i:\delta_i=1} \left(\ln \psi\left(\frac{\ln y_i - \mu}{\sigma}\right) - \ln \sigma - \ln y_i\right) + \sum_{i:\delta_i=0} \ln\left(1 - \Psi\left(\frac{\ln y_i - \mu}{\sigma}\right)\right)$$

Same theory as for Weibull (θ, α) basically holds for the MLE $\hat{\mu}, \hat{\sigma}$ as regards standard deviation, confidence intervals, etc.

Now the **observed information matrix** is

$$I(\hat{\mu}, \hat{\sigma}) = \begin{bmatrix} -\frac{\partial^2 l(\mu,\sigma)}{\partial \mu^2} & -\frac{\partial^2 l(\mu,\sigma)}{\partial \mu \partial \sigma} \\ -\frac{\partial^2 l(\mu,\theta)}{\partial \mu \partial \sigma} & -\frac{\partial^2 l(\mu,\sigma)}{\partial \sigma^2} \end{bmatrix}$$

and

$$\left[I(\hat{\mu}, \hat{\sigma})\right]^{-1} = \begin{bmatrix} \widehat{Var(\hat{\mu})} & \widehat{Cov(\hat{\sigma}, \hat{\mu})} \\ \widehat{Cov(\hat{\mu}, \hat{\sigma})} & \widehat{Var(\hat{\sigma})} \end{bmatrix}$$

These data are first reported in O'Connor (1985).

- Failure times, in *number of kilometers of use*, of vehicle shock absorbers.
- Two failure modes, denoted by MI and M2.
- One might be interested in the distribution of time to failure for mode MI, mode M2, or in the overall failure-time distribution of the part.

  Here we do not differentiate between modes MI and M2. We will consider estimation of the distribution of time to failure by either mode MI or M2.

Shock absorber data

Y = kilometers to failure, F = failure mode (0 is censoring)

Row	Y	F
1	6700	1
2	6950	0
3	7820	0
4	8790	0
5	9120	2
6	9660	0
7	9820	0
8	11310	0
9	11690	0
10	11850	0
11	11880	0
12	12140	0
13	12200	1
14	12870	0
15	13150	2
16	13330	0
17	13470	0
18	14040	0
19	14300	1
20	17520	1
21	17540	0
22	17890	0
23	18450	0
24	18960	0
25	18980	0
26	19410	0
27	20100	2
28	20100	0
29	20150	0
30	20320	0
31	20900	2
32	22700	1
33	23490	0
34	26510	1
35	27410	0
36	27490	1
37	27890	0
38	28100	0

Shock absorber data

```
Estimation Method: Maximum Likelihood Distribution:   Lognormal

Parameter Estimates

                      Standard    95,0% Normal CI
Parameter  Estimate     Error      Lower      Upper
Location    10,1448   0,144175   9,86219    10,4273
Scale      0,530068
                      0,112683   0,349447   0,804047


Log-Likelihood = -124,609

Goodness-of-Fit Anderson-Darling (adjusted) = 34,651

Characteristics of Distribution

                                      Standard   95,0% Normal CI
                          Estimate      Error     Lower     Upper
Mean(MTTF)                 29297,5    5455,91   20338,3   42203,2
Standard Deviation         16687,1    6787,01   7519,35   37032,5
Median                     25457,6    3670,36   19190,9   33770,7
First Quartile(Q1)         17805,2    2062,96   14188,1   22344,4
Third Quartile(Q3)         36399,0    7252,61   24631,2   53789,0
Interquartile Range(IQR)   18593,8    6115,60   9758,96   35426,9
```

SOME "HAND CALCULATIONS"

Estimates from MINITAB: $\hat{\mu} = 10.1448$, and $\hat{\sigma} = 0.530068$

$$\widehat{E(T)} \equiv \widehat{MTTF} = e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} = e^{10.1448+\frac{1}{2}\cdot 0.530068^2} = 29297.5$$

$$\widehat{SD(T)} = \sqrt{e^{2\hat{\mu}+\hat{\sigma}^2}(e^{\hat{\sigma}^2} - 1)} = 16687.1$$

$$\widehat{Median(T)} = e^{\hat{\mu}} = 25457.6$$

$$\hat{t}_{0.25} = e^{\hat{\mu}-0.67\hat{\sigma}} = 17805.2$$

$$\hat{t}_{0.75} = e^{\hat{\mu}+0.67\hat{\sigma}} = 36399.0$$

See next page: $\hat{t}_p = e^{\hat{\mu}+\hat{\sigma}\Phi^{-1}(p)}$.

Recall definition:

$$P(T \leq t_p) = p$$

$$p = P(T \leq t_p) = P(\ln T \leq \ln t_p) = \Psi\left(\frac{\ln t_p - \mu}{\sigma}\right)$$

From this,

$$\Psi^{-1}(p) = \frac{\ln t_p - \mu}{\sigma}$$

$$\ln t_p = \mu + \sigma\Psi^{-1}(p)$$

$$t_p = e^{\mu+\sigma\Psi^{-1}(p)}$$

where $\Psi^{-1}(p)$ has to be calculated for each model, see next page.

t_p FOR SPECIAL CASES

T is **lognormal**: $\Phi^{-1}(p)$ is in our tables of standard normal distribution.
Particular percentiles:
Median : $t_{0.5} = e^{\mu+\sigma\Phi^{-1}(0.5)} = e^{\mu}$ as $\Phi^{-1}(0.5) = 0$
$t_{0.25} = e^{\mu+\sigma\Phi^{-1}(0.25)} = e^{\mu-0.675\sigma}$
$t_{0.75} = e^{\mu+\sigma\Phi^{-1}(0.25)} = e^{\mu+0.675\sigma}$

T is **Weibull**: Here we need $G^{-1}(p)$. Solving $G(u) = 1 - e^{-e^{u}} = p$ we get $u = G^{-1}(p) = \ln(-\ln(1-p))$ and hence

$$\begin{aligned} t_p &= e^{\mu+\sigma\ln(-\ln(1-p))} = e^{\ln\theta+\frac{1}{\alpha}\ln(-\ln(1-p))} \\ &= e^{\ln\theta+\ln[(-\ln(1-p))^{1/\alpha}]} \\ &= \theta\cdot(-\ln(1-p))^{1/\alpha} \end{aligned}$$

(which we have derived earlier).

T is **log-logistic**: Here we need $H^{-1}(p)$. Solving $H(u) = \frac{e^u}{1+e^u} = p$ we get $u = H^{-1}(p) = \ln \frac{p}{1-p}$ and hence

$$t_p = e^{\mu+\sigma\cdot\ln\frac{p}{1-p}}$$

Median $= t_{0.5} = e^{\mu+\sigma\cdot\ln 1} = e^{\mu}$

$t_{0.25} = e^{\mu+\sigma\cdot\ln\frac{0.25}{0.75}} = e^{\mu-1.0986\sigma}$

$t_{0.75} = e^{\mu+1.0986\sigma}$

Shock absorber data:

Results for loglogistic (left), lognormal (middle), Weibull (right)

Table of Statistics	
Loc	10,1291
Scale	0,280982
Mean	28640,0
StDev	17608,6
Median	25062,8
IQR	15720,2
Failure	11
Censor	27
AD*	34,639

Table of Statistics	
Loc	10,1448
Scale	0,530068
Mean	29297,5
StDev	16687,1
Median	25457,6
IQR	18593,8
Failure	11
Censor	27
AD*	34,651

Table of Statistics	
Shape	3,16047
Scale	27718,7
Mean	24811,5
StDev	8605,90
Median	24683,6
IQR	12048,5
Failure	11
Censor	27
AD*	34,661

$F(t) = \Psi\left(\frac{\ln t - \mu}{\sigma}\right)$

$\Psi^{-1}(F(t)) = \frac{\ln t - \mu}{\sigma} = \frac{1}{\sigma} \ln t - \frac{\mu}{\sigma}$

Thus the points

$$(\ln t, \Psi^{-1}(F(t)))$$

are on the line

$$y = \frac{1}{\sigma} x - \frac{\mu}{\sigma}$$

For right-censored data we can estimate $F(t)$ by $1 - \hat{R}(t)$, where $\hat{R}(t)$ is the KM-estimator, and then plot the points

$$(\ln t_{(i)}, \Psi^{-1}(1 - \hat{R}(t_{(i)})))$$

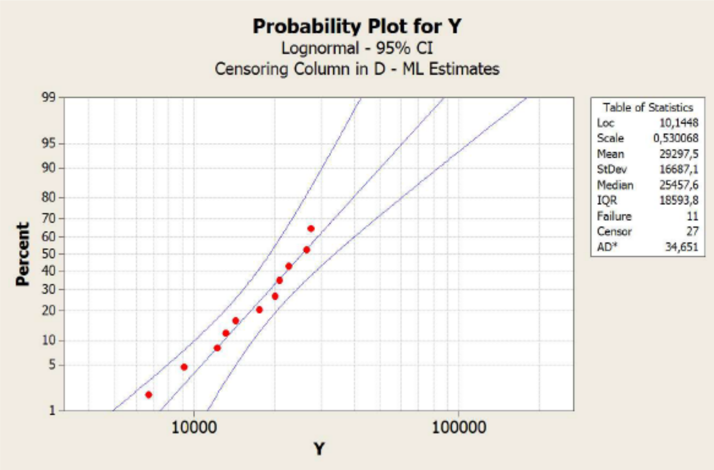
together with the line

$$y = \frac{1}{\hat{\sigma}} x - \frac{\hat{\mu}}{\hat{\sigma}}$$

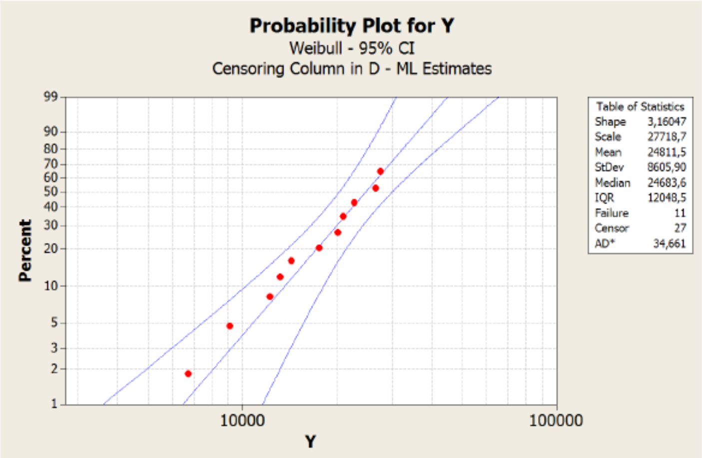
(As for Weibull, $\hat{R}(t)$ can be replaced by $\hat{\hat{R}}(t)$.)

- **Lognormal**: $\Phi^{-1}(p)$ is in statistical tables.
  Plot the points $(\ln t_{(i)}, \Phi^{-1}(1 - \hat{R}(t_{(i)})))$
- **Log-logistic**: $H^{-1}(p) = \ln \frac{p}{1-p}$
  Plot the points $(\ln t_{(i)}, \ln \frac{1-\hat{R}(t_{(i)})}{\hat{R}(t_{(i)})})$
- **Weibull**: $G^{-1}(p) = \ln(-\ln(1 - p))$
  Thus $G^{-1}(F(t)) = \ln(-\ln(1 - F(t))) = \ln(-\ln R(t))$,
  so plot the points
  $$(\ln t_{(i)}, \ln(-\ln \hat{R}(t_{(i)})))$$
  which is the same plot as derived earlier.

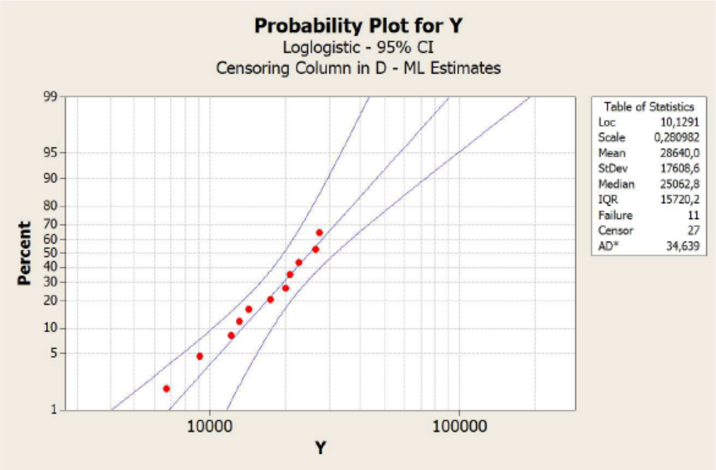
Shock absorber data



Shock absorber data



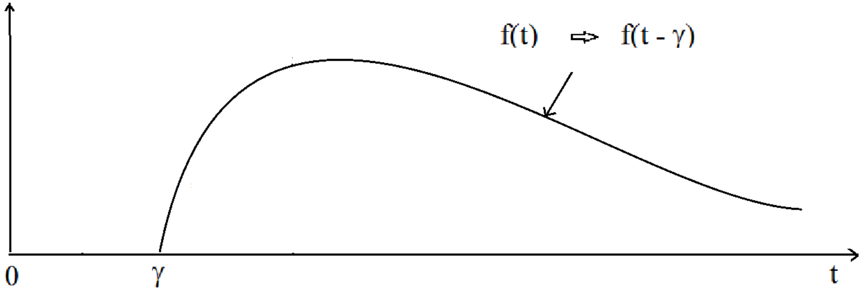
Shock absorber data

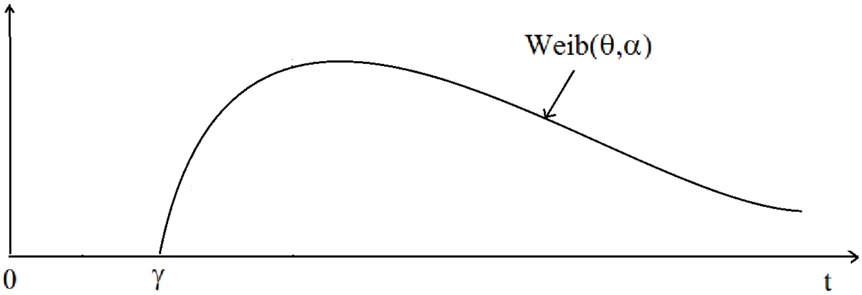


All distributions so far have been with positive densities from 0 and up. Threshold parameters $\gamma > 0$ can be added, so that "old" density $f(t); t > 0$, becomes "new" density

$$f(t - \gamma); \ t > \gamma$$

No failures can happen within the first γ time units, "guarantee time".





$$f(t;\theta,\alpha,\gamma) = \frac{\alpha}{\theta^\alpha}(t-\gamma)^{\alpha-1}e^{-(\frac{t-\gamma}{\theta})^\alpha}; \quad t > \gamma$$
$$= 0 \quad \text{otherwise}$$

$$R(t;\theta,\alpha,\gamma) = e^{-(\frac{t-\gamma}{\theta})^\alpha}; \quad t > \gamma$$

THREE-PARAMETER WEIBULL - LOG LIKELIHOOD

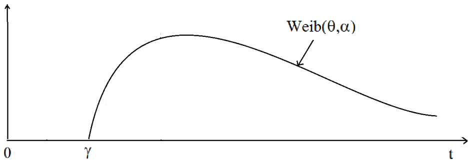
$$\ell(\theta,\alpha,\gamma) = r \ln \alpha - \alpha r \ln \theta + (\alpha - 1) \sum_{i:\delta_i=1} \ln(y_i - \gamma) - \sum_{i=1}^{n} \left(\frac{y_i - \gamma}{\theta}\right)^{\alpha}$$

where $r = \sum_{i=1}^{n} \delta_i$ is the number of failures, and where $\gamma \le \min y_i$.

Problem: log likelihood tends to ∞ if $\gamma = y_{(1)}$ (the smallest of the failure times) and $\alpha < 1$. Then there is no maximum likelihood estimate of the parameters.

So one usually assumes $\alpha \ge 1$, in which case there may be solutions obtained by differentiation as usual, but where one also needs to check the value of $l(\theta,\alpha,\gamma)$ on the boundary of the parameter space, i.e. $\alpha = 1$, in which case $\gamma = \min y_i$ is the maximizer for γ.

But - a profile log-likelihood may be the most "safe" procedure (see next slide).



Profile log-likelihood of γ:

$$\begin{aligned}\tilde{\ell}(\gamma) &= \max_{\theta,\alpha} \ell(\theta,\alpha,\gamma), \quad \gamma \text{ is fixed} \\ &= \ell(\hat{\theta}(\gamma),\hat{\alpha}(\gamma),\gamma)\end{aligned}$$

This is done for each γ by subtracting γ from all data and fitting an ordinary Weibull(θ,α).

Then the ML estimator $\hat{\gamma}$ is the one that maximizes $\tilde{\ell}(\gamma)$. The other ML estimates are $\hat{\theta}(\hat{\gamma}), \hat{\alpha}(\hat{\gamma})$.

Example: Pike (1966) data.

Pike (1966) cancer data for rats

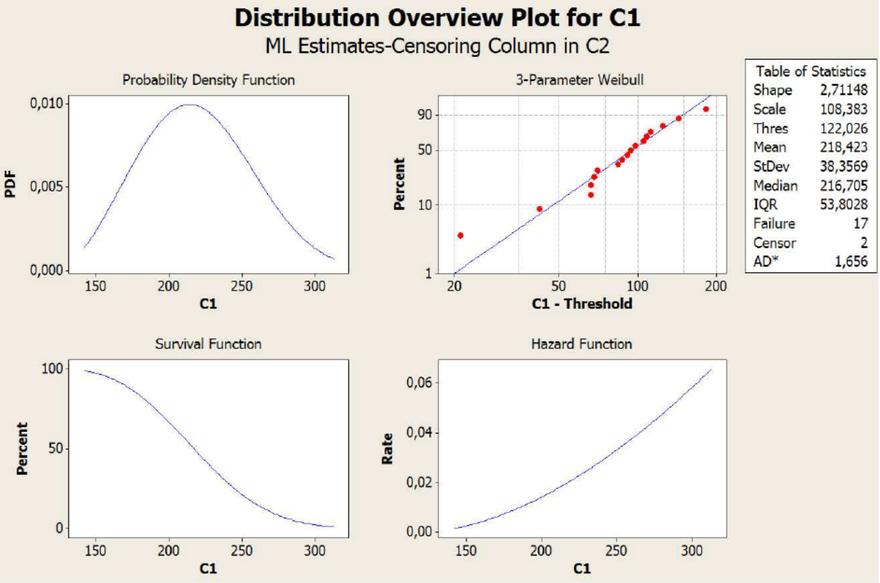
Row	Y	D
1	143	1
2	164	1
3	188	1
4	188	1
5	190	1
6	192	1
7	206	1
8	209	1
9	213	1
10	216	1
11	220	1
12	227	1
13	230	1
14	234	1
15	246	1
16	265	1
17	304	1
18	216	0
19	244	0

Distribution Analysis: C1

```
Variable: C1

Censoring Information   Count
Uncensored value           17
Right censored value        2

Censoring value: C2 = 0

Estimation Method: Maximum Likelihood

Distribution:   3-Parameter Weibull


Parameter Estimates

                     Standard    95,0% Normal CI
Parameter  Estimate     Error     Lower     Upper
Shape       2,71148   1,05876   1,26135   5,82878
Scale       108,383   32,5734   60,1367   195,335
Threshold   122,026   28,6924   65,7898   178,262

Log-Likelihood = -87,324

Goodness-of-Fit
Anderson-Darling (adjusted) = 1,656


Characteristics of Distribution

                                      Standard    95,0% Normal CI
                           Estimate      Error     Lower     Upper
Mean(MTTF)                  218,423    8,99156   201,492   236,777
Standard Deviation          38,3569    6,41597   27,6352   53,2383
Median                      216,705    9,89384   198,156   236,991
First Quartile(Q1)          190,481    9,63934   172,495   210,342
Third Quartile(Q3)          244,284    11,0118   223,627   266,849
Interquartile Range(IQR)    53,8028    8,97770   38,7945   74,6172
```

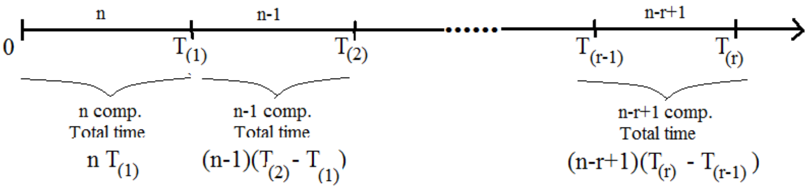
,

PIKE DATA (CONT.)



Pike 3-parameter Weibull: Profile log likelihood for γ

γ	$\hat{\theta}(\gamma)$	$\hat{\alpha}(\gamma)$	$\tilde{l}(\gamma)$
0	234.3	6.08	−88.233
60	173.2	4.49	−87.831
100	131.8	3.38	−87.467
110	121.2	3.08	−87.381
120	110.6	2.78	−87.327
122	108.4	2.71	−87.324
125	105.2	2.61	−87.330
130	99.7	2.44	−87.382
135	94.0	2.24	−87.542
140	88.0	1.99	−88.064
142	85.2	1.80	−88.773
143	81.1	1.00	−91.718



n units put on test at time $t = 0$. Stop after a given number r of failures.

$$\hat{\theta} = \frac{\sum_{i=1}^{n} Y_i}{r} = \frac{\sum_{i=1}^{r} T_{(i)} + (n-r)T_r}{r} = \frac{\text{" }TTT\text{"}}{r}$$

$$= \frac{\overbrace{nT_{(1)}}^{U_1 \sim expon(1/\theta)} + \overbrace{(n-1)(T_{(2)} - T_{(1)})}^{U_2 \sim expon(1/\theta)} + \cdots + \overbrace{(n-r+1)(T_{(r)} - T_{(r-1)})}^{U_r \sim expon(1/\theta)}}{r}$$

$$= \frac{U_1 + U_2 + \cdots + U_r}{r}$$

EXACT CONFIDENCE INTERVAL (CONT.)

From introductory courses it is known that for $U_i \sim \text{expon}(1/\theta)$ then

$$\frac{2U_i}{\theta} \sim \chi^2_2$$

Thus,

$$\frac{2r}{\theta}\hat{\theta} = \frac{2\sum_{i=1}^{r} U_i}{\theta} \sim \chi^2_{2r}$$

Hence, in table of χ^2_{2r}, we find a, b so that

$$P(a < \frac{2r}{\theta}\hat{\theta} < b) = 0.95$$

$$P(\frac{2r\hat{\theta}}{b} < \theta < \frac{2r\hat{\theta}}{a}) = 0.95$$

An exact 95% confidence interval for θ for type II censoring and exponential distribution is hence

$$\left(\frac{2r\hat{\theta}}{b}, \frac{2r\hat{\theta}}{a}\right), \quad or \left(\frac{2TTT}{b}, \frac{2TTT}{a}\right)$$

Confidence Interval for the Mean Life of a New Insulating Material

- A life test for a new insulating material used 25 specimens which were tested simultaneously at a high voltage of 30 kV.
- The test was run until 15 of the specimens failed.
- The 15 failure times (hours) were recorded as:

  1.08, 12.20, 17.80, 19.10, 26.00, 27.90, 28.20, 32.20, 35.90, 43.50, 44.00, 45.20, 45.70, 46.30, 47.80

  Then $TTT = 1.08 + \cdots + 47.80 + 10 \times 47.80 = 950.88$ hours.
- The ML estimate of θ and a 95% confidence interval are:

$$\begin{aligned} \widehat{\theta} &= 950.88/15 = 63.392 \text{ hours} \\ [\underset{\sim}{\theta}, \tilde{\theta}] &= \left[\frac{2(950.88)}{\chi^2_{(.975;30)}}, \frac{2(950.88)}{\chi^2_{(.025;30)}}\right] = \left[\frac{1901.76}{46.98}, \frac{1901.76}{16.79}\right] \\ &= [40.48, \quad 113.26]. \end{aligned}$$

Note: The interval is an exact 95% confidence interval in the case of type II censoring for given r.

It turns out that the interval is often a very good approximate 95% confidence interval also for general right censoring.

In our earlier example with $r = 5, \sum Y_i = 23$

$$\Big(\frac{2 \cdot 23}{\underbrace{20.483}_{0.025 \text{ in } \chi^2_{10}}}, \frac{2 \cdot 23}{\underbrace{3.247}_{0.975 \text{ in } \chi^2_{10}}}\Big)$$

$$(2.2458, 14.1669)$$