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

http://www.math.ntnu.no/~bo/
bo@math.ntnu.no

NTNU, Spring 2014

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

- if $\ln T \sim \mathcal{N}(\mu, \sigma)$ then T is lognormal(μ, σ), i.e. $\ln T = \mu + \sigma U$ and where $U \sim \mathcal{N}(0, 1)$
- if $\ln T \sim$ *logistic*(μ, σ) then T is log-logistic(μ, σ), i.e. $\ln T = \mu + \sigma U$ and where $U \sim$ *logistic*$(0, 1)$
- if T is Weibull(θ, α), i.e. $lnT = \ln \theta + \frac{1}{\alpha} U$ and where $U \sim Gumbel(0, 1)$, then $\ln T \sim Gumbel(\ln \theta, 1/\alpha)$

Typically location- scale family correspond to a $\Phi(x)$, centered around 0. such that

$$P(\ln T \leq y) = \Phi(\frac{y-\mu}{\sigma})$$

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

$$\text{Logistic} : \Phi(x) = \frac{e^x}{1+e^x} \quad \text{where} \quad \phi(x) = \frac{e^x}{1+e^x}^{2}$$

$$\text{Gumbel} : \Phi(x) = 1 - e^{-e^x} \quad \text{where} \quad \phi(x) = e^{x-e^x}$$

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

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

so

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

and

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

Thus for all log- location-scale models:

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

and log-likelihood is

$$l(\mu,\sigma) = \sum_{i:\delta_i=1} \left(\ln \phi\left(\frac{\ln y_i - \mu}{\sigma}\right) - \ln \sigma - \ln y_i \right) + \sum_{i:\delta_i=0} \ln \left(1 - \Phi\left(\frac{\ln y_i - \mu}{\sigma}\right)\right)$$

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

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

and

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

- Lognormal, $\hat{\mu} = 10.1448$, and $\hat{\sigma} = 0.530068$
- $\widehat{E(T)} \equiv \widehat{MTTF} = e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2} = e^{10.1448+\frac{1}{2}\cdot 0.530068^2} = 29297.5$
- $\widehat{SD(T)} = \sqrt{e^{2\hat{\mu}+\hat{\sigma}^2}(e^{\hat{\sigma}^2} - 1)} = 16687.1$
- $\widehat{Median(T)} = e^{\hat{\mu}} = 25457.6$, $\hat{t}_{0.25} = e^{\hat{\mu}-0.67\hat{\sigma}} = 17805.2$, $\hat{t}_{0.75} = e^{\hat{\mu}+0.67\hat{\sigma}} = 36399.0$ and $\hat{t}_p = e^{\hat{\mu}+\hat{\sigma}\Phi^{-1}(p)}$ in general.

Definition:

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

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

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

$$\ln t_p = \mu + \sigma\Phi^{-1}(p) \quad \text{analogous with} \quad \ln T = \mu + \sigma U$$

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

where $\Phi^{-1}(p)$ differs from model to model.

- T is **Lognormal**:
  $\Phi^{-1}(p)$ is in our tables of standard normal distribution. In particular 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**:
  $\Phi^{-1}(p) = \ln(-\ln(1-p))$

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

- T is **loglogistic**:

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

$$\text{Median} = t_{0.5} = e^{\mu + \sigma \cdot \ln 1} = e^{\mu}$$
$$t_{0.25} = e^{\mu + \sigma \cdot \ln \frac{0.25}{0.75}} = e^{\mu - 1.0986\sigma}$$
$$t_{0.75} = e^{\mu + 1.0986\sigma}$$

- $F(t) = \Phi(\frac{\ln t - \mu}{\sigma})$
- So, $\Phi^{-1}(F(t)) = \frac{\ln t - \mu}{\sigma} = \frac{1}{\sigma} \ln t - \frac{\mu}{\sigma}$
- Thus the points
  $(\ln t, \Phi^{-1}(F(t)))$ are on the line $y = \frac{1}{\sigma} x - \frac{\mu}{\sigma}$
- Thus estimating $F(t)$ by $1 - \hat{R}(t)$, we get the plot of the points
  $(\ln t_{(i)}, \Phi^{-1}(1 - \hat{R}(t_{(i)})))$ or $(\ln t_{(i)}, \Phi^{-1}(1 - \hat{\hat{R}}(t_{(i)})))$

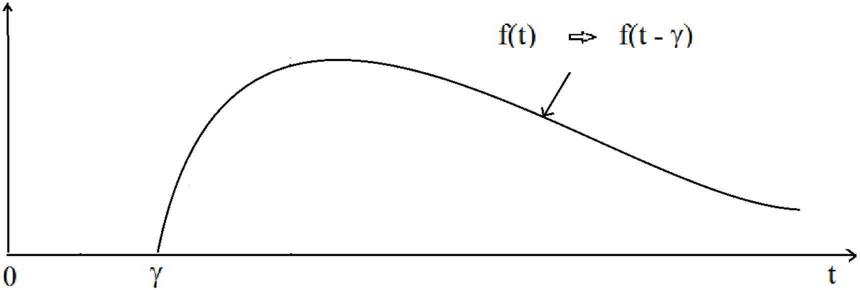
where $t_{(i)}$ are the observed *failure* times.

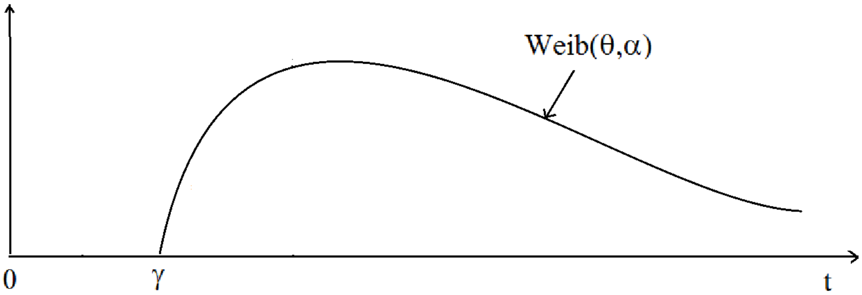
- **Lognormal**: $\Phi^{-1}(p)$ is in our tables.
- **Loglogistic**: $\Phi^{-1}(p) = \ln \frac{p}{1-p}$
- **Weibull**: $\Phi^{-1}(p) = \ln(-\ln(1-p))$h
  so the Weibull-plot obtained this way will be based on $(\ln t, \ln(-\ln(1-F(t)))) = (\ln t, \ln(-\ln R(t)))$, which is the one we have derived earlier.

All distibuion 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".





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

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

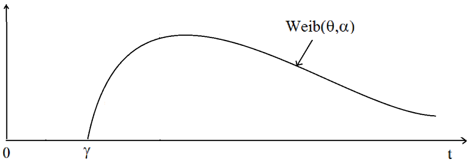
$$l(\theta, \alpha, \gamma) = r \ln \alpha - \alpha r \ln \theta + (\alpha - 1) \sum_{i:\delta_i=1} \ln(y_i - \gamma) - \sum_{i=1}^{n} \left(\frac{y_i - \gamma}{\theta}\right)^{\alpha}$$

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

Problem : likelihood tends to ∞ if $\gamma_1 = t_{(1)}$ (the smallest of the failure times) and $\alpha < 1$. Then there is no maximum likelihood estimate of the parameters.

So one usually assumes $\alpha \geq 1$, 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:

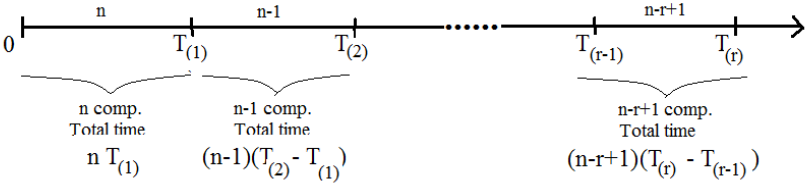
$$\begin{aligned}\tilde{l}(\gamma) &= \max_{\theta,\alpha} l(\theta,\alpha,\gamma), \quad \gamma \text{ is fixed} \\ &= l(\hat{\theta}(\gamma),\hat{\alpha}(\gamma),\gamma)\end{aligned}$$

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 $\hat{l}(\gamma)$. The other ML estimates are $\hat{\theta}(\hat{\gamma}),\hat{\alpha}(\hat{\gamma})$.

Example: Pike (1966) data.

EXACT CONFIDENCE INTERVAL FOR EXPONENTIAL DISTRIBUTION AND TYPE II CENSORING



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

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

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

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

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

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

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 an approximate 95% confidence interval also for general right censoring.

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

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

$$(2.2458, 14.1669)$$