TMA4275 LIFETIME ANALYSIS

Slides 15: Parametric estimation in NHPPs

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 2015

- Parametric models for the ROCOF of NHPPs
  - Power law NHPP
  - Log linear NHPP
- Likelihood function for parametric NHPPs
  - Single system
  - m systems
  - Profile log likelihood
- Maximum likelihood estimation
  - Observed information matrix
  - Confidence intervals
- Examples
  - A simple example with 3 systems
  - Grampus data

Power law NHPP (also called Weibull-process)

$w(t) = \lambda\beta t^{\beta-1}$ for $\lambda, \beta > 0$

$\searrow$ if $\beta < 1$

$\nearrow$ if $\beta > 1$

HPP if $\beta = 1$

$W(t) = \int_0^t w(u)du = \int_0^t \lambda\beta u^{\beta-1} du = \lambda t^{\beta}$

Similar to Weibull hazard $z(t) = \frac{\alpha}{\theta}(\frac{t}{\theta})^{\alpha-1}$

Log linear NHPP

$w(t) = e^{\alpha+\beta t}$ for $-\infty < \lambda, \beta < \infty$

$\searrow$ if $\beta < 0$

$\nearrow$ if $\beta > 0$

HPP if $\beta = 0$

$W(t) = \frac{e^{\alpha}}{\beta}(e^{\beta t} - 1)$

Suppose first that we observe a *single* NHPP with ROCOF $w(t;\theta)$ on the time interval $[0,\tau]$, where τ is fixed.

Then we know that if N is the number of events, then $N \equiv N(\tau) \sim Poisson(W(\tau;\theta))$, where $W(\tau;\theta)) = \int_0^\tau w(u;\theta)du$.

Let the times of events be: $(0 \leq)\ \ s_1 \leq s_2 \leq \cdots \leq s_N\ \ (\leq \tau)$.

Define also

$$W(s,t;\theta) = \int_s^t w(u;\theta)du = E[N(s,t)]$$

On the next slides we derive the likelihood function for these observations.

Divide time axis at $h_0 = 0 \leq h_1 < h_2 < \cdots < h_r \equiv \tau$.

Let $D_i = \#$events in $(h_{i-1}, h_i]$. Then $D_i \sim Poisson(W(h_{i-1}, h_i; \theta))$, and $D_1, D_2, \cdots, D_r$ are independent (properties of NHPP), so the likelihood is

$$
\begin{aligned}
L(\theta) &= P(D_1 = d_1, D_2 = d_2, \cdots, D_r = d_r) \\
&= \prod_{i=1}^{r} P(D_i = d_i) = \prod_{i=1}^{r} \frac{W(h_{i-1}, h_i; \theta)^{d_i}}{d_i!} e^{-W(h_{i-1}, h_i; \theta)} \\
&= \left\{ \prod_{i=1}^{r} \frac{W(h_{i-1}, h_i; \theta)^{d_i}}{d_i!} \right\} \cdot e^{-\sum_{i=1}^{r} W(h_{i-1}, h_i; \theta)} \\
&= \left\{ \prod_{i=1}^{r} \frac{W(h_{i-1}, h_i; \theta)^{d_i}}{d_i!} \right\} \cdot e^{-W(\tau; \theta)}
\end{aligned}
$$

since
$\sum_{i=1}^{r} W(h_{i-1}, h_i; \theta) = \sum_{i=1}^{r} \int_{h_{i-1}}^{h_i} w(u; \theta) du = \int_0^{\tau} w(u; \theta) du = W(\tau; \theta)$

Recall $L(\theta) = \left\{\prod_{i=1}^{r} \frac{W(h_{i-1},h_i;\theta)^{d_i}}{d_i!}\right\} \cdot e^{-W(\tau;\theta)}$.

If data are given by counts D_i in intervals $(h_{i-1}, h_i]$, then we maximize this $L(\theta)$ to find MLE.

If times are instead given by exact times $s_1, s_2, \cdots, s_N$, then we let the grid of h_i be more and more dense and get in the limit 0 or 1 event in each interval $(h_{i-1}, h_i]$.

Now, when $d_i = 0$, the contribution to the product is $\frac{W(h_{i-1},h_i;\theta)^0}{0!} = 1$, which can be ignored.

When $d_i = 1$, the contribution is

$$W(h_{i-1}, h_i; \theta) = \int_{h_{i-1}}^{h_i} w(u;\theta)du \approx w(s_k;\theta)(h_i - h_{i-1})$$

Since the $h_i - h_{i-1}$ are fixed by us, we let the contribution to the likelihood be $w(s_k, \theta)$ for $k = 1, 2, \ldots, N$.

Hence we get the likelihood for exactly observed failure times from a single repairable system:

$$L(\theta) = \left\{ \prod_{i=1}^{N} w(s_i, \theta) \right\} e^{-W(\tau;\theta)}$$

If there are data from m systems with the same $w(t;\theta)$, then the likelihood is the product of the likelihoods for each system:

$$L(\theta) = \prod_{j=1}^{m} \left(\left\{ \prod_{i=1}^{N_j} w(s_{ij};\theta) \right\} e^{-W(\tau_j;\theta)} \right)$$

Here system j ($j = 1, \ldots, m$) is observed on the time interval $(0, \tau_j]$, with N_j observations $s_{1j}, s_{2j}, \ldots, s_{N_j j}$.

The log-likelihood function for m systems is hence

$$\ell(\theta) = \sum_{j=1}^{m} \sum_{i=1}^{N_j} \ln w(s_{ij}; \theta) - \sum_{j=1}^{m} W(\tau_j; \theta)$$

Example - power law NHPP: $w(t; \lambda, \beta) = \lambda\beta t^{\beta-1}$, $W(t; \lambda, \beta) = \lambda t^{\beta}$

$$\ell(\lambda, \beta) = \sum_{j=1}^{m} \sum_{i=1}^{N_j} (\ln \lambda + \ln \beta + (\beta - 1) \ln s_{ij}) - \sum_{j=1}^{m} \lambda \tau_j^{\beta}))$$

$$= N \ln \lambda + N \ln \beta + (\beta - 1) S - \lambda \sum_{j=1}^{m} \tau_j^{\beta}$$

where $N = \sum_{j=1}^{m} N_j$ = total number of observations,
and $S = \sum_{j=1}^{m} \sum_{i=1}^{N_j} \ln s_{ij}$ = sum of logs of all observations.

Recall $\ell(\lambda, \beta)) = N \ln \lambda + N \ln \beta + (\beta - 1)S - \lambda \sum_{j=1}^{m} \tau_j^{\beta}$

$$\frac{\partial \ell}{\partial \lambda} = \frac{N}{\lambda} - \sum_{j=1}^{m} \tau_j^{\beta} = 0 \qquad (1)$$

$$\frac{\partial \ell}{\partial \beta} = \frac{N}{\beta} + S - \lambda \sum_{j=1}^{m} \tau_j^{\beta} \ln \tau_j = 0 \qquad (2)$$

(1) gives $\lambda = \frac{N}{\sum_{j=1}^{m} \tau_j^{\beta}}$

put this into (2)

$$\frac{N}{\beta} + S - \frac{N \sum_{j=1}^{m} \tau_j^{\beta} \ln \tau_j}{\sum_{j=1}^{m} \tau_j^{\beta}} = 0$$

Can solve this numerically for β to get $\hat{\beta}$ and then put $\hat{\lambda} = \frac{N}{\sum_{j=1}^{m} \tau_j^{\hat{\beta}}}$

Special case: If all $\tau_j \equiv \tau$ are equal (including the case m=1)

$$\frac{N}{\beta} + S - \frac{Nm\tau^{\beta} \ln \tau}{m\tau^{\beta}} = 0$$

$$\frac{N}{\beta} + S - N \ln \tau = 0$$

$$\beta = \frac{N}{N \ln \tau - S}$$

So in this case:

$$\hat{\beta} = \frac{N}{N \ln \tau - S}$$

FINDING MLE BY PROFILE LOG LIKELIHOOD

If β is *known*, then we found by solving $\partial\ell/\partial\lambda = 0$,

$$\hat{\lambda}(\beta) = \frac{N}{\sum_{j=1}^{m} \tau_j^{\beta}}$$

The profile log likelihood for β is therefore

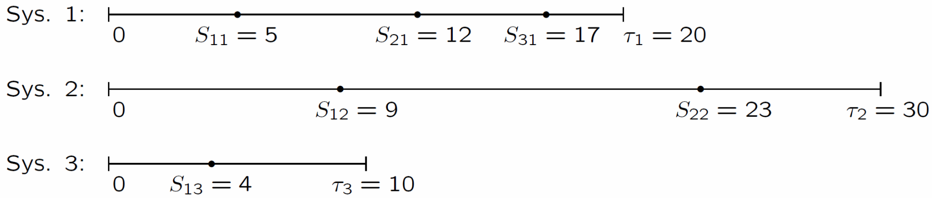
$$\begin{aligned} \tilde{\ell}(\beta) &= \ell(\hat{\lambda}(\beta), \beta) \\ &= N \ln \hat{\lambda}(\beta) + N \ln \beta + (\beta - 1)S - \hat{\lambda}(\beta) \cdot \sum_{j=1}^{m} \tau_j^{\beta} \\ &= N \ln N - N \ln(\sum_{j=1}^{m} \tau_j^{\beta}) + N \ln \beta + (\beta - 1)S - N \end{aligned}$$

Then to find the MLE of λ, β we can

1. Maximize $\tilde{\ell}(\beta)$ to find $\hat{\beta}$
2. Put $\hat{\lambda} = \hat{\lambda}(\hat{\beta})$

SIMPLE EXAMPLE: DATA

SIMPLE EXAMPLE WITH THREE SYSTEMS

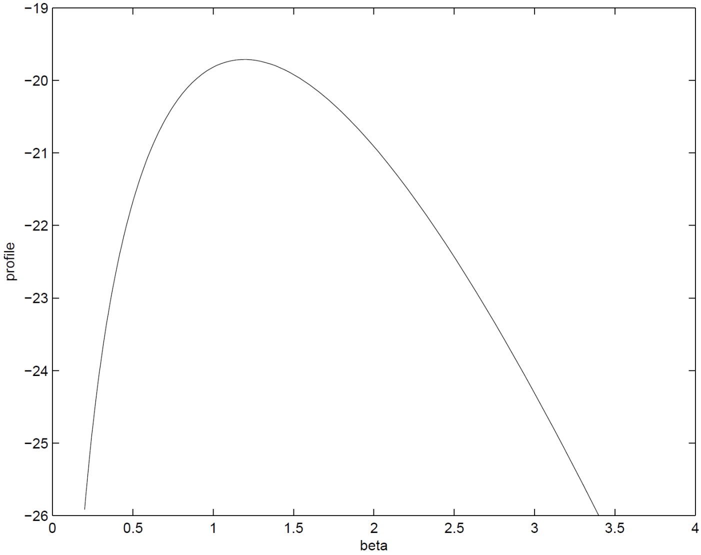


$N = 6,\ S = \ln 5 + \ln 12 + \ln 17 + \ln 9 + \ln 23 + \ln 4 = 13.6466$, so

$$
\begin{aligned}
\tilde{\ell}(\beta) &= N \ln N - N \ln\Big(\sum_{j=1}^{m} \tau_j^{\beta}\Big) + N \ln \beta + (\beta - 1)S - N \\
&= 6 \ln 6 - 6 \ln(20^{\beta} + 30^{\beta} + 10^{\beta}) + 6 \ln \beta + (\beta - 1)13.6466 - 6
\end{aligned}
$$

See graph next slide!

SIMPLE EXAMPLE: PROFILE LOG LIKELIHOOD OF β



$\tilde{\ell}(\beta)$, maximum at $\hat{\beta} = 1.194$, $\hat{\lambda}(\hat{\beta}) = \frac{6}{20^{\hat{\beta}} + 30^{\hat{\beta}} + 10^{\hat{\beta}}} = 0.0548$

Confidence interval for β using profile log likelihood:

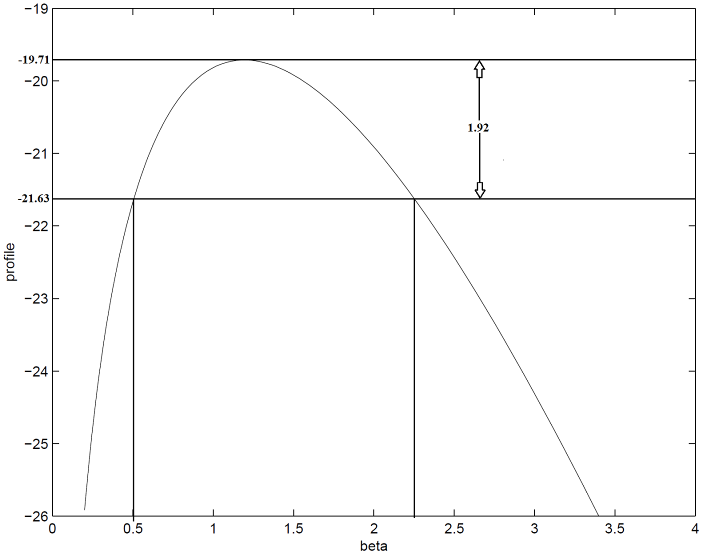
$$W(\beta) = 2(\underbrace{\ell(\hat{\lambda}(\hat{\beta}), \hat{\beta})}_{=\ell(\hat{\lambda},\hat{\beta})} - \underbrace{\ell(\hat{\lambda}(\beta), \beta)}_{=\tilde{\ell}(\beta)}) \approx \chi_1^2$$

$\Longrightarrow$ (as before) that a 95% (approx) confidence interval for β is obtained by cutting off profile log-likelihood at maximum value minus 1.92, i.e., at

$$-19.71 - 1.92 = -21.63$$

(see next slide).

THE 1.92 CI FOR β



The approximate 95% CI for β is hence $(0.50, 2.25)$

Recall $\ell(\lambda, \beta) = N \ln \lambda + N \ln \beta + (\beta - 1)S - \lambda \sum_{j=1}^{m} \tau_j^{\beta}$

$$
\begin{aligned}
\frac{\partial \ell}{\partial \lambda} &= \frac{N}{\lambda} - \sum_{j=1}^{m} \tau_j^{\beta} \\
\frac{\partial^2 \ell}{\partial \lambda^2} &= -\frac{N}{\lambda^2} \\
\frac{\partial^2 \ell}{\partial \lambda \partial \beta} &= -\sum_{j=1}^{m} (\ln \tau_j) \tau_j^{\beta} \\
\frac{\partial \ell}{\partial \beta} &= \frac{N}{\beta} + S - \lambda \sum_{j=1}^{m} (\ln \tau_j) \tau_j^{\beta} \\
\frac{\partial^2 \ell}{\partial \beta^2} &= \frac{N}{\beta^2} - \lambda \sum_{j=1}^{m} (\ln \tau_j)^2 \tau_j^{\beta}
\end{aligned}
$$

$$\begin{bmatrix} -\frac{\partial^2 l}{\partial \lambda^2} & -\frac{\partial^2 l}{\partial \lambda \partial \beta} \\ -\frac{\partial^2 l}{\partial \lambda \partial \beta} & -\frac{\partial^2 l}{\partial \beta^2} \end{bmatrix}_{\lambda=\hat{\lambda} \quad \beta=\hat{\beta}} = \begin{bmatrix} \frac{N}{\hat{\lambda}^2} & \sum_{j=1}^m (\ln \tau_j)\tau_j^{\hat{\beta}} \\ \sum_{j=1}^m (\ln \tau_j)\tau_j^{\hat{\beta}} & \frac{N}{\hat{\beta}^2} + \hat{\lambda}\sum_{j=1}^m (\ln \tau_j)^2 \tau_j^{\hat{\beta}} \end{bmatrix}$$

In the simple example the observed information matrix is

$$\begin{bmatrix} \frac{6}{0.0538^2} = 2072.9 & 30^{\hat{\beta}} \ln 30 + 20^{\hat{\beta}} \ln 20 + 10^{\hat{\beta}} \ln 10 = 347.03 \\ 347.03 & \frac{6}{1.2^2} + 0.0538 \cdot 1096 = 63.1315 \end{bmatrix}$$

with inverse given by

$$\begin{bmatrix} 0.006050 & -0.03325 \\ -0.03325 & 0.1986 \end{bmatrix}$$

We can read off from the inverse observed information matrix: $\widehat{Var}\hat{\beta} = 0.1986$, so the standard error is $\widehat{SD}\hat{\beta} = 0.4457$.

The standard 95% CI for β is $\underbrace{\hat{\beta}}_{1.194} \pm 1.96 \cdot 0.4457$, i.e. $[0.320, 2.067]$

This is the interval that MINITAB gives

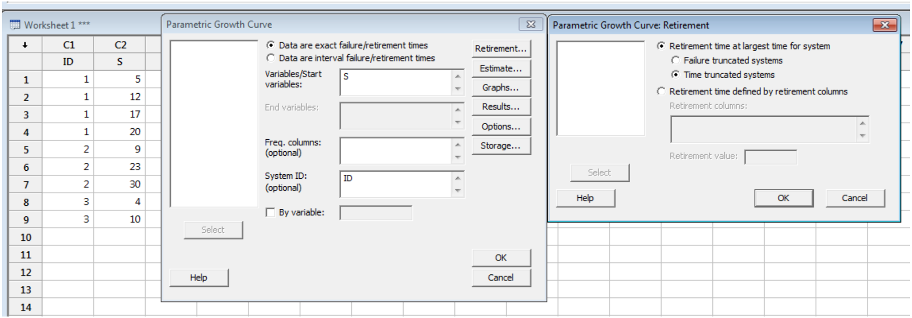
BUT MINITAB ought to use the "standard interval for positive parameters", $1.94 \cdot e^{\pm 1.96 \frac{0.4457}{1.194}}$, i.e. [0.574, 2.482]
(which is also closer to the 1.92 interval).

Finally, $\widehat{Var}\hat{\lambda} = 0.006050$, i.e. $\widehat{SD}\hat{\lambda} = 0.0775$.

Note that MINITAB uses another parameterization, with $W(t) = \left(\frac{t}{\theta}\right)^{\alpha}$.

Standard 95% CI for λ: $\underbrace{\hat{\lambda}}_{0.0548} \pm 1.96 \cdot 0.0775 = [0, 0.2067]$, while standard CI for positive: $\hat{\lambda} e^{\pm 1.96 \frac{SD(\hat{\lambda})}{\hat{\lambda}}} = 0.0548 \cdot e^{\pm 1.96 \frac{0.0775}{0.0548}}$ gives [0.0034,0.8761].

Stat > Reliability/Survival > Repairable System Analysis > Parametric Growth Curve



Simple Example With 3 Systems

Power Law NHPP Model: $W(t; \alpha, \theta) = (t/\theta)^\alpha$

Results for: SimpleNHPP.MTW

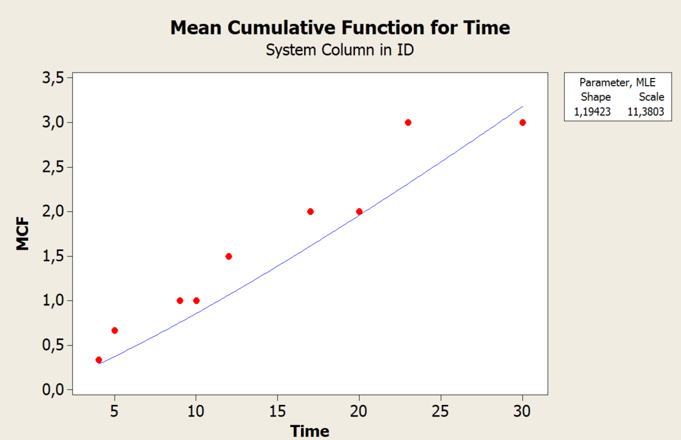
Parametric Growth Curve: Time

```
System:   ID

Model: Power-Law Process
Estimation Method: Maximum Likelihood


Parameter Estimates
```

Parameter	Estimate	Standard Error	95% Normal CI Lower	95% Normal CI Upper
Shape	1,19423	0,445	0,323015	2,06545
Scale	11,3803	4,840	1,89335	20,8672


Mean Cumulative Function for Time
System Column in ID
Parameter, MLE
Shape Scale
1,19423 11,3803
MCF
3,5
3,0
2,5
2,0
1,5
1,0
0,5
0,0
5
10
15
20
25
30
Time

MINITAB's POWER LAW: $w(t) = \frac{\text{Shape}}{\text{Scale}} \left(\frac{t}{scale} \right)^{shape-1}$

POWER LAW IN SLIDES: $w(t) = \lambda \beta t^{\beta - 1}$.

Thus: Shape $= \beta$, so the same value should be obtained.

But $\lambda = \frac{1}{scale^{shape}} =_{\text{from MINITAB}} \frac{1}{11.3803^{1.19423}} = 0.0548$ (which we already obtained)

EXAMPLE: GRAMPUS DATA

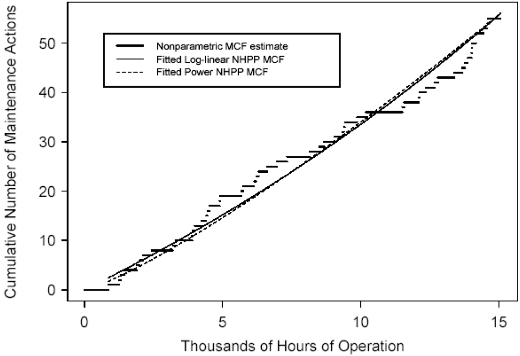


Figure shows estimates of $W(t)$ by:

- Nelson-Aalen estimator (here just $\hat{W}(t) = N(t)$)
- Power-law estimate $\hat{W}(t) = \hat{\lambda} t^{\hat{\beta}} = 2.06 \cdot t^{1.22}$
- Log-linear estimate $\hat{W}(t) = \frac{e^{\hat{\alpha}}}{\hat{\beta}}(e^{\hat{\beta} t} - 1) = \frac{e^{1.01}}{0.0377}(e^{0.0377 \cdot t} - 1)$