Exam TMA4275 May 2013

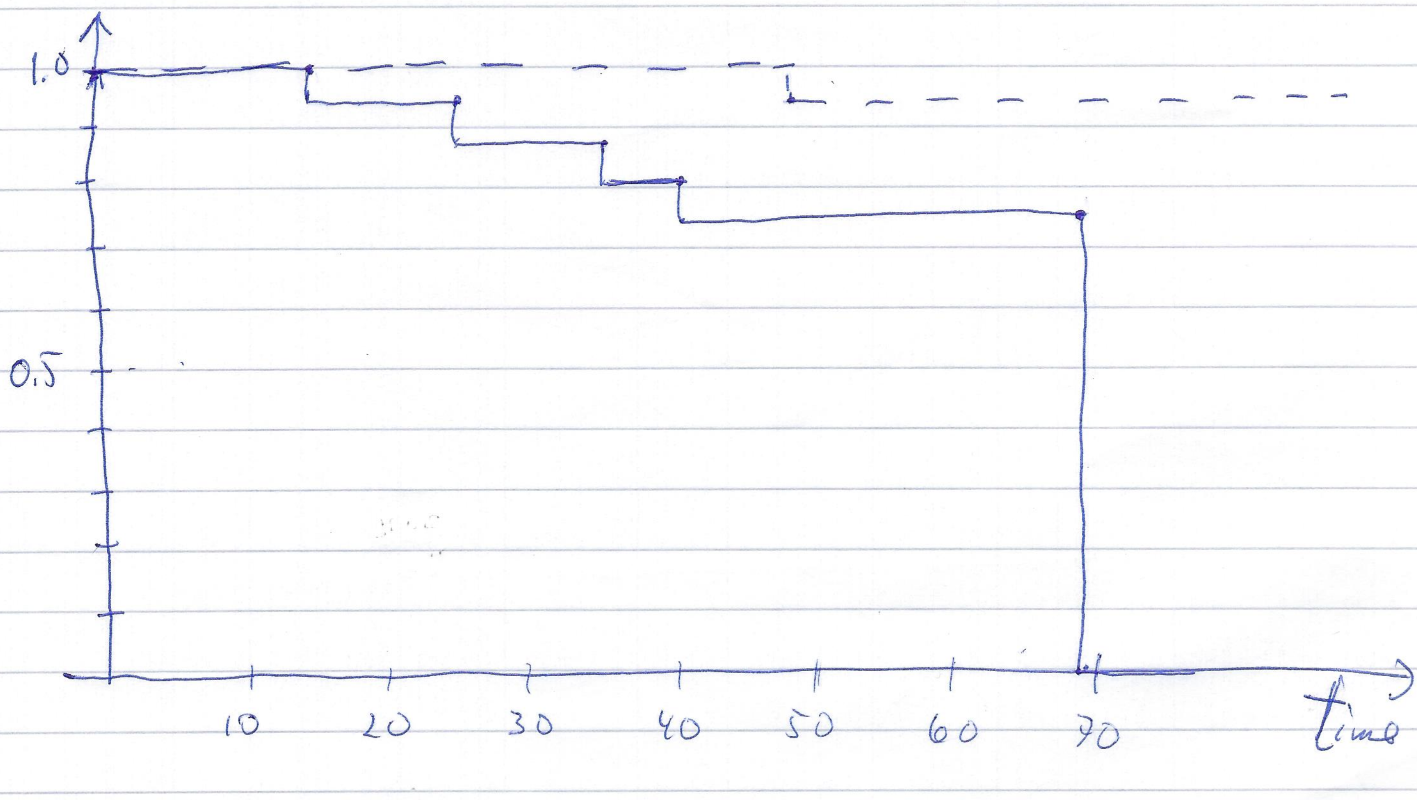
Problem 1:

a) Placebo:

Time $t_{(i)}$	n_i	d_i	$\frac{n_i - d_i}{n_i}$	KM $(\widehat{P(T > t_{(i)})})$
14	17	1	$\frac{16}{17}$	0.9412
26	14	1	$\frac{13}{14}$	0.8739
36	13	1	$\frac{12}{13}$	0.8067
42	12	1	$\frac{11}{12}$	0.7395
69	1	1	0	0

DES:

50	16	1	$\frac{15}{16}$	0.9375



Apparently, the DES treatment leads to longer survival.

Est. median lifetime Placebo: 69
Est. mean lifetime Placebo:
Area under KM curve up to 69:

$$1 \cdot 14 + 0{,}9412 \cdot 12 + 0{,}8739 \cdot 10 + 0{,}8067 \cdot 6 + 0.7395 \cdot 27 = \underline{58{,}8\cancel{4}}$$

Est. mean lifetime DES:
Area under KM curve up to 70 (largest observed time):

$$1 \cdot 50 + 0{,}9375 \cdot 20 = \underline{\underline{68.75}}$$

(Median lifetime can't be computed because estimate does not go beyond 0.5).

b) Model:

$$\ln T = \beta_0 + \beta_1 \cdot \text{Treat} + \beta_2 \cdot \text{Age} + \beta_3 \cdot \text{Shb} + \beta_4 \cdot \text{Size} + \beta_5 \cdot \text{Index} + \frac{1}{\alpha} W$$

where $\cancel{\beta_1}, \ldots, \beta_5$ are unknown regression coefficients, β_0 is the intercept; α is the shape parameter and W is Gumbel-distributed.

Median for Weibull-distribution with $R(t) = e^{-\left(\frac{t}{\theta}\right)^\alpha}$ is $\underline{\tilde{t} = \theta(\ln 2)^{1/\alpha}}$

Thus median lifetime of regression model is:

$$med = e^{\beta_0+\beta_1\cdot Treat+\dots+\beta_5\cdot Index}\cdot(\ln 2)^{1/\alpha}$$

Estimates $\hat{\beta}_i$:

Treatment: median lifetime increased by Treat = 1.
Age, Shb, Size, Index: median lifetime is reduced when these covariates are increased.

$$\frac{Med\ (DES)}{Med\ (placebo)} = e^{\beta_1}\ ;\ \text{estimated to } e^{0.425215} = \underline{1.53}$$

so increases by 53% compared to placebo.

Significant effect: Size, Index, but <u>not</u> Treat.

c) $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$.

"Coef": Test statistic $Z_1 = \frac{\hat{\beta}_1}{SD(\hat{\beta}_1)} = \frac{0.434133}{0.463267} = 0.94$

P-value: $2\cdot P(Z > 0.94)$ when $Z \sim N(0,1)$.

$= 2\cdot 0.1745 = \underline{0.349}$

"Log-Likelihood": $2(\text{loglik(full model)} - \text{loglik(reduced model)})$
is $\approx \chi^2_1$ when reduced model holds.

Here: $2(-31.434-(-32.015)) = 1.162$

so p-value is $P(\chi^2_1 > 1.162) \underset{\uparrow}{=} \begin{cases} 1.1 & 0.2943 \\ 1.2 & 0.2733 \end{cases}$ ~~0.286~~

table in
the exercise

≈ 0.286

The p-values are too large for declaring β_1 significantly different from 0.

Problem 2:

a) $N(t) \sim \text{Poisson}(\lambda t)$

~~a)~~ $E(N(t)) = \lambda t$

$$P(N(t)=1) = \frac{(\lambda t)^1}{1!} e^{-\lambda t} = \lambda t e^{-\lambda t}$$

~~b)~~ $P(S_1 > t) = P(\text{no replac. in } (0,t)) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$

So S_1 is exponentially distributed with rate λ.

b.) General likelihood, $NHPP(w(t;\theta))$:

$$\cancel{L(\theta) = \prod_{i=1}^{m} \left[\left(\prod_{j=1}^{N_i} w(t_{ij};\theta) \right) \cdot e^{\cancel{- W(\tau_i;\theta)}} \right.}$$

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

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

Here: $w(t;\theta)=\lambda$; $W(t;\theta)=\lambda t$, so

$$\underline{\underline{L(\lambda)=\lambda^{\sum_{j=1}^{m} N_j}\cdot e^{-\lambda\sum_{j=1}^{m}\tau_j}}}$$

c) $\ell(\lambda)=\left(\sum N_j\right)\ln\lambda-\lambda\sum\tau_j$

$$\ell'(\lambda)=\frac{\sum N_j}{\lambda}-\sum\tau_j$$

$$\ell'(\lambda)=0 \iff \lambda=\frac{\sum N_j}{\sum\tau_j}$$

so $$\hat{\lambda}=\frac{\sum_{j=1}^{m} N_j}{\sum_{j=1}^{m}\tau_j}=\frac{\#\text{replacements}}{\text{total time}}$$

$$\ell''(\lambda)=-\frac{\sum N_j}{\lambda^2}$$

so Observed information is $-\ell''(\hat{\lambda})=\dfrac{\sum N_j}{\hat{\lambda}^2}$

$$=\frac{\left(\sum_{j=1}^{m}\tau_j\right)^2}{\sum_{j=1}^{m} N_j}$$ and

$$\widehat{\operatorname{Var}\hat{\lambda}} = \frac{\hat{\lambda}^2}{\sum_{j=1}^{m} N_j} \quad \text{or} \quad \widehat{SD\,\hat{\lambda}} = \frac{\hat{\lambda}}{\sqrt{\sum_{j=1}^{m} N_j}} = \frac{\sqrt{\sum_{j=1}^{m} N_j}}{\sum_{j=1}^{m} \tau_j}$$

With data:

$$\hat{\lambda} = \frac{\sum N_j}{\sum \tau_j} = \frac{6}{990} = \frac{1}{165} = 0.00606$$

$$SD(\hat{\lambda}) = \frac{1}{165} \cdot \frac{1}{\sqrt{6}} = 0.00247$$

Conf. int:

Use standard interval for positive parameters:

General: $\hat{\theta} \cdot e^{\pm 1.96 \cdot \frac{SD(\hat{\theta})}{\hat{\theta}}}$

Here: $\hat{\lambda} \cdot e^{\pm 1.96 \frac{1}{\sqrt{\sum N_j}}}$

Here: $0.00606 \cdot e^{\pm 1.96 \cdot \frac{1}{\sqrt{6}}}$

$(0.00272,\ 0.01349)$

~~7~~ -

Problem 3:

a) $F(t) = P(T \le t) = P(\ln T \le \ln t)$

$= P(\mu + \sigma W \le \ln t)$

$= P\left(W \le \frac{\ln t - \mu}{\sigma}\right)$

$$= \frac{e^{\frac{\ln t - \mu}{\sigma}}}{1 + e^{\frac{\ln t - \mu}{\sigma}}}$$

$F(t_p) = p$

$\cancel{R(t_p) = 1 - p}$

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

$$e^{\frac{\ln t - \mu}{\sigma}} = p + p\,e^{\frac{\ln t - \mu}{\sigma}}$$

$$(1-p)\,e^{\frac{\ln t - \mu}{\sigma}} = p$$

$$e^{\frac{\ln t - \mu}{\sigma}} = \frac{p}{1-p}$$

$$\frac{\ln t - \mu}{\sigma} = \frac{p}{1-p}$$

$$\Rightarrow \quad \ln t_p = \mu + \sigma \ln \frac{p}{1-p}$$

$$\text{so } t_p = e^{\mu + \sigma \ln \frac{p}{1-p}} = e^{\mu} \cdot \left(\frac{p}{1-p}\right)^{\sigma}$$

Median:

$$\underline{\underline{t_{0.5} = e^{\mu}}}$$

Interquartile range:

$$t_{0.75} - t_{0.25} = \cancel{e^{\mu}\left[\left(\frac{p}{1-p}\right)\right.}$$

$$= e^{\mu}\left[\left(\frac{\frac{3}{4}}{1-\frac{3}{4}}\right)^{\sigma} - \left(\frac{\frac{1}{4}}{1-\frac{1}{4}}\right)^{\sigma}\right]$$

$$\underline{\underline{= e^{\mu}\left[3^{\sigma} - \left(\frac{1}{3}\right)^{\sigma}\right]}}$$

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

$$\cancel{-\frac{d}{dt} R(t) = \frac{d}{dt}\left(1 + e^{\frac{\ln t - \mu}{\sigma}}\right)}$$

$$F(t) + F(t)\, e^{\frac{\ln t - \mu}{\sigma}} = e^{\frac{\ln t - \mu}{\sigma}}$$

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

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

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

so $$\ln \frac{1-R(t)}{R(t)} = \frac{1}{\sigma} \ln t - \frac{\mu}{\sigma}$$

Let $\hat{R}(t)$ be a KM-estimate (or modified KM estimate) based on the censored data.

Then the points

$$\left(\ln \frac{1-\hat{R}(t_{(i)})}{\hat{R}(t_{(i)})} , \ln t_{(i)} \right)$$

will be close to a straight line if the model is correct.

The <u>slope</u> of the line then estimates $1/\sigma$, while the intercept estimates $-\frac{\mu}{\sigma}$

Thus estimates of μ and σ can be obtained.

c) $Z(t) = -\ln R(t) = -\ln\left(\frac{1}{1+e^{\frac{\ln t - \mu}{\sigma}}}\right)$

$= \ln\left(1+e^{\frac{\ln t-\mu}{\sigma}}\right)$

Known that $Z(T) \sim \text{expon}(1)$.

Cox-Snell residuals: For censored data (Y_i, δ_i); $i=1,\dots,n$, the Cox-Snell residuals are

$$(Z(Y_i), \delta_i) \; ; \; i=1,\dots,n$$

which if the model is correct are a censored sample from the standard exponential distribution.