TMA 4275 - 31 May 2014

Problem 1

a) It follows that $Z_U(u) = \left(\frac{u}{\theta}\right)^\alpha$

so $z_U(u) = Z_U'(u) = \frac{\alpha u^{\alpha-1}}{\theta^\alpha}$ for $u > 0$.

Thus for a patient with covariate x is

$$z(t;x) = \alpha e^{-\alpha(\beta_0+\beta_1 x)} t^{\alpha-1}$$

So: <u>Without</u> pneum.: $x=0$: $z(t;0) = \alpha e^{-\alpha\beta_0} t^{\alpha-1}$

<u>With</u> " : $x=1$: $z(t;1) = \alpha e^{-\alpha(\beta_0+\beta_1)} t^{\alpha-1}$

Relative risk <u>without</u> vs. with:

$$\frac{z(t;0)}{z(t;1)} = e^{\alpha\beta_1}$$

b) Estimated model:

$$\ln T = 2.53917 + 0.741551 x + \frac{1}{1.91178} W$$

Significant difference?

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

Test statistic: $Z = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{0.741551}{0.344049} = 2.16$

For a 5% test we reject if $|\hat{\beta}_1| \geq 1.96$

so we reject, and conclude that there is

a difference between the two groups.

Relative risk:

$$e^{\hat{\alpha}\hat{\beta}_1} = e^{1.91178 \cdot 0.741551} = e^{1.417682}$$

$$= \underline{4.127}$$

This means that the rate of discharge for

a non-peumonia patient is 4.127 times

the one for a pneumonia patient.

The shape $\alpha = 1.91$ means that there is

an increasing discharge rate with time

(so "the longer you stay, the larger is the

probability of discharge the next day").

c) 95% CI for β_1: (Standard interval since

β_1 can be both negative and positive)

$$\hat{\beta}_1 \pm 1.96\, SE(\hat{\beta}_1)$$

$0.741551 \pm 1.96 \cdot 0.344049$

0.742 ± 0.674 , or $\underline{(0.068, 1.416)}$

95% CI for α: (Stand. interval for positive parameters)

$$\hat{\alpha} \cdot e^{\pm 1.96 \frac{SE(\hat{\alpha})}{\hat{\alpha}}}$$

$$1.91178 \cdot e^{\pm 1.96 \cdot \frac{0.461130}{1.91178}}$$

$\underline{(1.192, 3.067)}$

Would like to test H_0: exp. distr. vs H_1: Weibull

i.e. $H_0: \alpha = 1$ vs. $H_1: \alpha \neq 1$

A 5% test is obtained by rejecting H_0 if $1 \notin$ confidence interval. Since this is the case (lower bound is > 1), we reject the null hypothesis and conclude that an exponential model is not appropriate.

Alternatively we can test $H_0: \alpha = 1$ vs. $H_1: \alpha \neq 1$ by the log-likelihood test, which uses the fact that

$$2(\text{loglik full model} - \text{loglik restricted model})$$

$$\approx \chi^2_{\underset{\text{difference in \# parameters}}{1}} \text{ when } H_0 \text{ holds.}$$

Then reject if 2·differens is ≥ 3.84

We get from MINITAB output

$$2(\text{loglik Weibull} - \text{loglik Expon}) = 5.366$$

so we reject H_0 (as we did with the first test).

d) $H_0: \quad R_0(t) = R_1(t)$ for all t
vs. $H_1: R_0(t) \neq R_1(t)$ for some t.

Test statistic:

$$V = \underbrace{\frac{(4-6.80)^2}{6.80}}_{\substack{\text{with} \\ \text{pneum}}} + \underbrace{\frac{(6-3.20)^2}{3.20}}_{\substack{\text{without} \\ \text{pneum}}} = 3.60$$

so we do not reject at 5% (requires 3.84), since test statistic is $\approx \chi^2_1$ when H_0 holds.

The numbers 6.80 and 3.20 are obtained as follows:

Time	Risk 0	Risk 1	Risk	Fail 0	Fail 1	Fail	E0	E1
2	8	7	15	1	0	1	$8 \cdot \frac{1}{15}$	$7 \cdot \frac{1}{15}$
6	6	6	12	2	0	2	$6 \cdot \frac{2}{12}$	$6 \cdot \frac{2}{12}$
9	4	6	10	0	1	1	$4 \cdot \frac{1}{10}$	$6 \cdot \frac{1}{10}$
10	4	5	9	1	0	1	$4 \cdot \frac{1}{9}$	$5 \cdot \frac{1}{9}$
11	3	5	8	1	0	1	$3 \cdot \frac{1}{8}$	$5 \cdot \frac{1}{8}$
17	1	4	5	0	1	1	$1 \cdot \frac{1}{5}$	$4 \cdot \frac{1}{5}$
23	1	3	4	1	0	1	$1 \cdot \frac{1}{4}$	$3 \cdot \frac{1}{4}$
24	0	3	3	0	1	1	0	$3 \cdot \frac{1}{3}$
32	0	1	1	0	1	1	0	$1 \cdot 1$
							3.20 ~~6.80~~	6.80

So - we do not reject at 5% with this test, which essentially tests the same as the test for $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$, which was rejected in (b)

A possible reason is that we in this subpoint make fewer assumptions on the underlying distributions, which usually leads to lower testing power.

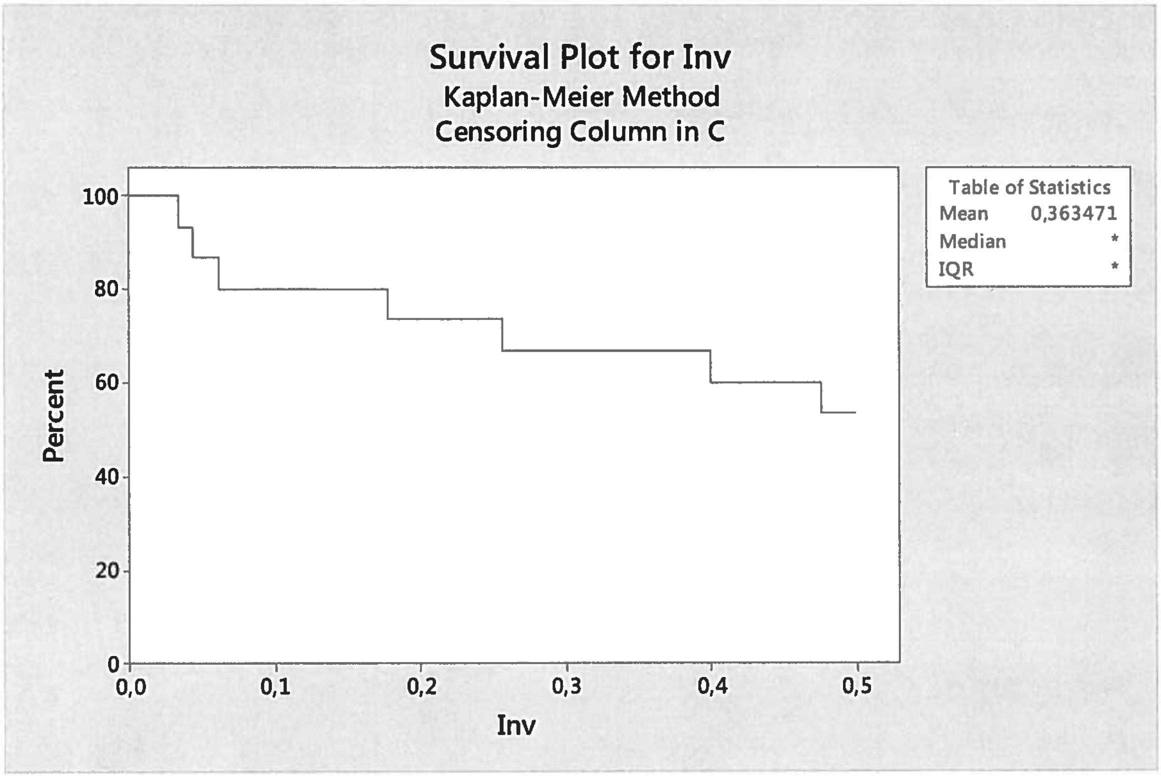
Problem 2

a) See lectures/slides for def. of censoring.

The failure times during the 2 first hours are left-censored, since all we know is that they are ≤ 2.0.

b) The exact observed T-times are clearly transformed to exact observed V-times. For the censored T-times, knowing that $T \leq 2.0$ is equivalent to knowing that $V \geq 1/2.0 = 0.5$, which means right censoring at 0.5. (In fact, the V-data is a type I censored data set).

The KM-plot/estimate for V has been done in MINITAB (next page)



Distribution Analysis: Inv

```
Variable: Inv

Censoring Information  Count
Uncensored value           7
Right censored value       8

Censoring value: C = 0


Nonparametric Estimates


Characteristics of Variable

             Standard     95,0% Normal CI
Mean(MTTF)      Error     Lower      Upper
  0,363471  0,0508824  0,263743  0,463199

Median = *
IQR = *  Q1 = 0,178571  Q3 = *


Kaplan-Meier Estimates

            Number  Number     Survival  Standard   95,0% Normal CI
    Time   at Risk  Failed  Probability     Error     Lower    Upper
0,034722        15       1     0,933333  0,064406  0,807100  1,00000
0,044444        14       1     0,866667  0,087771  0,694639  1,00000
0,061728        13       1     0,800000  0,103280  0,597576  1,00000
0,178571        12       1     0,733333  0,114180  0,509545  0,95712
0,256410        11       1     0,666667  0,121716  0,428107  0,90523
0,400000        10       1     0,600000  0,126491  0,352082  0,84792
0,476190         9       1     0,533333  0,128812  0,280866  0,78580
```

c) $R_T(t) = P(T > t) = P\left(\frac{1}{V} > t\right) = P\left(V < \frac{1}{t}\right)$

$= 1 - P\left(V > \frac{1}{t}\right) = 1 - R_V\left(\frac{1}{t}\right)$

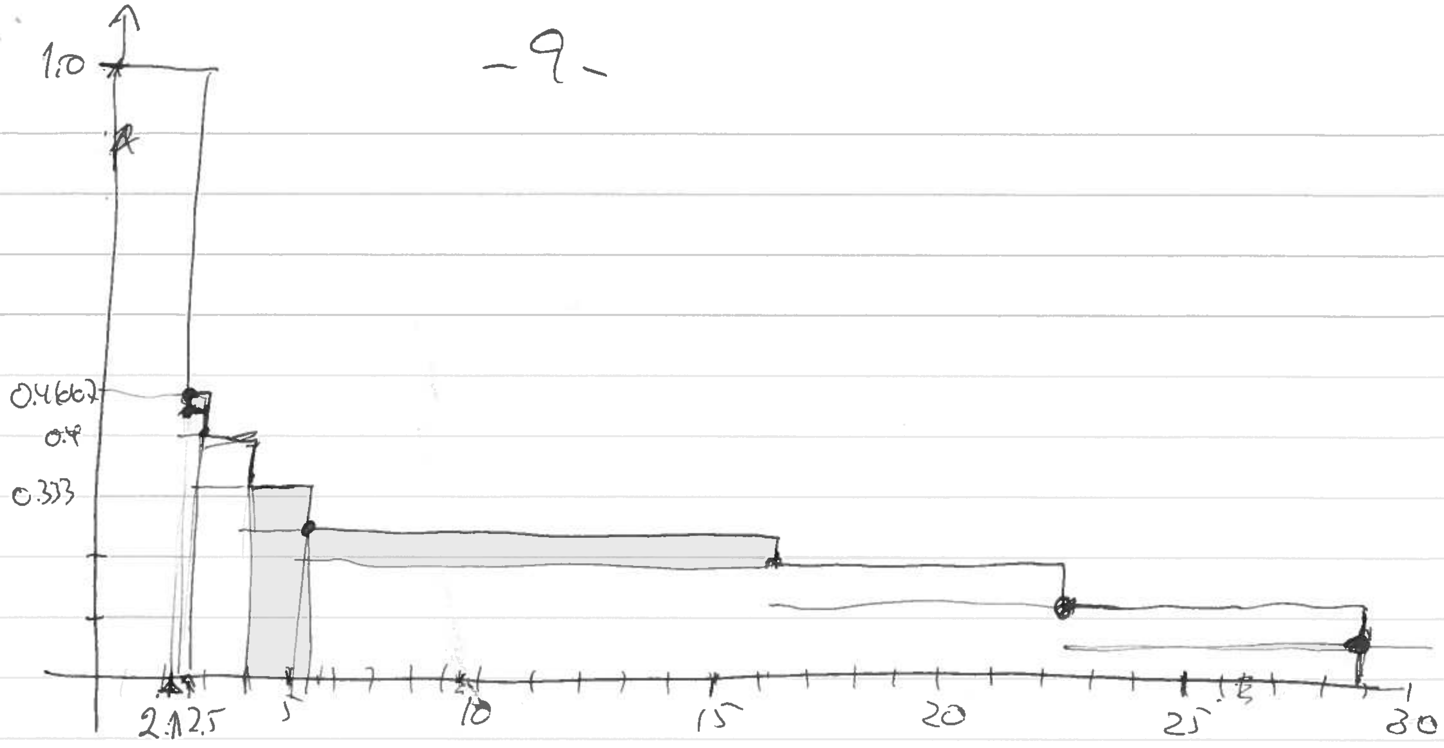
~~since~~ using that V has a continuous distribution.

Thus we will use

$$\hat{R}_T(t) = 1 - \hat{R}_V\left(\frac{1}{t}\right)$$

and compute it at the points t where $\frac{1}{t}$ is a failure time of ~~R~~ V.

v	$t = 1/v$	$\hat{R}_V(v)$	$\hat{R}_T(t)$
0.035	28.8	0.9333	0.0667
0.044	22.5	0.8667	0.1333
0.062	16.2	0.8	0.2
0.179	5.6	0.7333	0.2667
0.256	3.9	0.6667	0.3333
0.4	2.5	0.6	0.4
0.476	2.1	0.5333	0.4667



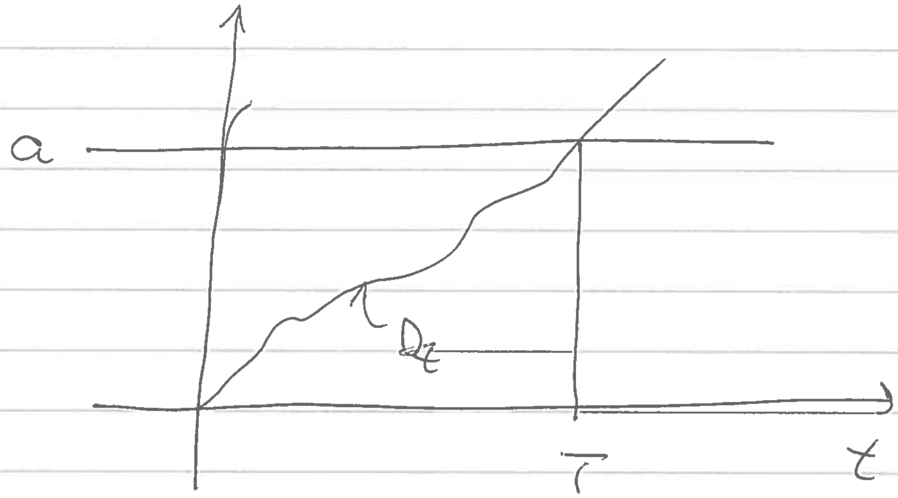
Can estimate MTTF by the area under the curve until 28.8 (last observation)

$$\widehat{MTTF} = 2.1 \cdot 1 + (2.5-2.1)\,0.4667$$
$$+ (3.9-2.1)\cdot 0.4 + (5.6-3.9)\cdot 0.3333$$
$$+ (16.2-5.6)\cdot 0.2667$$
$$+ (22.5-16.2)\cdot 0.2$$
$$+ (28.8-22.5)\cdot 0.1333$$
$$= 8.5001$$

~~(som en kan~~
(which is far from the median

In fact - the data are simulated from the distribution (4), for which the expected values does not exist (i.e., is ∞).

Problem 3



a) $P(T > t) = P(D_t < a)$: (*)

If ~~we~~ $T > t$, then we know that by time t the D_t has not yet reached the level a.

Now assume

$$D_t = Bt \quad \text{where } B \sim \text{expon}(\theta).$$

Then by (*)

$$P(T > t) = P(D_t < a) = P(Bt < a)$$

$$= P\left(B < \frac{a}{t}\right) = 1 - e^{-\frac{a\theta}{t}}$$

so $F_T(t) = e^{-\frac{a\theta}{t}}$

b) Solve

$$F(t_p; \theta) = p$$

$$e^{-\frac{\theta}{t_p}} = p$$

$$-\frac{\theta}{t_p} = \ln p$$

$$\underline{t_p = -\frac{\theta}{\ln p}}$$

Median: $p = \frac{1}{2} = 0.5$

$$\underline{\underline{t_{0.5}}} = -\frac{\theta}{\ln(\frac{1}{2})} = \underline{\underline{\frac{\theta}{\ln 2}}}$$

Interquartile range:

$$\underline{\underline{t_{0.75} - t_{0.25}}} = -\frac{\theta}{\ln \frac{3}{4}} + \frac{\theta}{\ln \frac{1}{4}} = \underline{\underline{\theta\left(\frac{1}{\ln \frac{4}{3}} - \frac{1}{\ln 4}\right)}}$$

c) Likelihood contribution of a left-censored observation is $F(t;\theta)$, and for an exact observation $f(t;\theta)$.

Thus

$$L(\theta) = \left(e^{-\frac{\theta}{2.0}}\right)^8 \cdot \prod_{i=1}^{7}\left(\frac{\theta}{t_i^2} e^{-\frac{\theta}{t_i}}\right)$$

where $t_1, \ldots, t_7$ are the exact observations.

(Density is $f(t;\theta) = F'(t;\theta) = \frac{\theta}{t^2} e^{-\frac{t}{\theta}}$

Log likelihood is therefore

$$\ell(\theta) = -8 \cdot \frac{\theta}{2.0} + 7\ln\theta - \sum_{i=1}^{7} \ln t_i^2 - \theta\sum_{i=1}^{7}\frac{1}{t_i}$$

$$\ell'(\theta) = -4 + \frac{7}{\theta} - \sum_{i=1}^{7}\frac{1}{t_i} = 0$$

$$\Updownarrow$$

$$\hat{\theta} = \frac{7}{4 + \sum_{i=1}^{7}\frac{1}{t_i}} = \frac{7}{5.452} = \underline{\underline{1.28}}$$

$$\widehat{\text{median}} = \frac{\hat{\theta}}{\ln 2} = \underline{\underline{1.85}}$$

(This is not unreasonable, since 8 out of 15 ~~observation~~ lifetimes are ≤ 2.0, so the empirical median - which for a complete dataset is ~~the~~ the 8th biggest, is probably slightly less than 2.0.)