Norwegian University of Science and Technology
Department of Mathematical Sciences

Contact during Exam:
Rupali Akerkar (40 33 20 37)

EXAM IN TMA4275 LIFETIME ANALYSIS

Wednesday 30 May 2012
Tid: 09:00 – 13:00

Permitted Aids:
Statistiske tabeller og formler (Tapir forlag)
Approved Calculator (HP30S)
K. Rottmann: Matematisk formelsamling
One yellow sheet (A4 with stamp) with your own formulae and notes

Grading: 20 June 2011

Problem 1

A random sample of $n = 100$ items were put on test at a particular point in time and the test was run until $t_c = 1000$ hours, at which point $r = 20$ units had been reported as failing. The reported failure times were noted. We can denote these by $t_1, \cdots, t_{20}$. Engineers believe that, based on the physical knowledge of the failure mechanism, the one-parameter exponential distribution can be used to describe time to failure. The exponential cumulative distribution function (CDF) is given by
$F(t, \theta) = 1 - \exp(-t/\theta)$.

a) For the given data, write down an expression for the likelihood function of θ.

b) Derive an expression for the maximum likelihood estimate of θ based on the given data.

Problem 2

A clinical trial to evaluate the efficacy of chemotherapy for a specific cancer was conducted. After reaching a state of remission (disappearance of cancer) through treatment, the patients who entered the study were randomized into two groups. The first group received maintenance chemotherapy, the second (or control) group did not. For a preliminary analysis during the course of the trial the data were as follows:
Length of complete remission (in weeks).

Maintenance group: 9, 13, 13^+, 23, 24^+, 34, 45^+, 55, 161^+
Control group: 5, 13, 13, 16^+, 20, 21, 43, 45
+ indicates censored observation.

a) What does it (in general) mean that an observation is censored?
What are typical causes (in general) leading to a dataset containing censored observations?

b) Let $R_M(t)$ and $R_C(t)$ be the survival functions for the length of remission for the maintenance group and the control group, respectively. Compute the Kaplan - Meier (KM) estimators for $R_M(t)$ and $R_C(t)$ and plot them on the same graph. Compare the KM curves. Which group seems to have better length of remission (cancer free survival)?

c) Explain how one may use the logrank test to test the null hypothesis that patients in the maintenance group and the control group have the same survival distribution. The total number of expected events (end of remission) for the maintenance group is 7.73 and for the control group is 4.27. Use these results to compute the test statistic and perform the test with significance level 5%. What do you conclude about whether or not the two groups have the same survival distribution?

One wants to use a Cox model for the data.

Let the only covariate be treatment, represented by the variable x with values 1 for the maintenance group and 0 for the control group.

d) Write down a simple Cox-model which leads to a test for this situation. Formulate a null hypothesis and an alternative hypothesis to test that the cancer free survival times are the same for the two groups.
(You are not asked to perform the testing nor writing down the partial likelihood).

This test was performed in R and the result is given on the next page:

```
coxph(formula = Surv(time, status) ~ trt, data = data)

     coef exp(coef) se(coef)     x      p
trt -1.06       0.347    0.636  -1.67   0.096

Likelihood ratio test = 2.92  on 1 df, p=0.0873  n= 17, number of events= 12
```

Use this information to conclude whether the cancer free survival times are the same for the two groups.

Problem 3

Consider a component where the time to failure T has the survival (reliability) function

$$R(t;\alpha,\theta) = \frac{1}{1+(\frac{t}{\theta})^{\alpha}} \quad \text{for} \quad t > 0 \tag{1}$$

where $\alpha > 0$ and $\theta > 0$ are parameters.

a) Let the pth quantile $t_p(\alpha,\theta)$ for T be given by $P(T \leq t_p(\alpha,\theta)) = p$ for $0 < p < 1$.
Show that

$$\ln t_p(\alpha,\theta) = \ln\theta + \frac{1}{\alpha}\ln\frac{p}{1-p}$$

where ln is the natural logarithm.
Derive an expression for the median in the distribution of T.
Also find the expression for the first and the third quantiles, ($Q1 = t_{0.25}$ and $Q3 = t_{0.75}$).

b) Show that the hazard function (failure rate function) for T is given by

$$z(t;\alpha,\theta) = \frac{\alpha t^{\alpha-1}}{\theta^{\alpha}+t^{\alpha}} \quad \text{for} \quad t > 0.$$

Problem 4 is on the next page.

Problem 4

One has registered failures of a special valve until time $\tau = 900$ days after start. The following data were observed (failure times in days after start):

$270, 520, 700, 810, 860$

It is assumed that the failure times follow a non-homogeneous Poisson process (NHPP).

a) How would you estimate the cumulative intensity $W(t)$ when no parametric assumptions are made? Draw the estimated curve for $W(t)$.
Does the plot indicate a trend in the failure intensity?
Perform the Laplace test to investigate whether there is a trend in the failure times.
What is your conclusion if you use 5% level of significance ?

b) Assume that the ROCOF $w(t)$ of the failure process is

$$w(t) = e^{\alpha+\beta t} \quad \text{for} \quad t > 0 \tag{2}$$

where $-\infty < \alpha < \infty$ and $-\infty < \beta < \infty$ are unknown parameters.
What is this model called? What is the cumulative ROCOF for this model?
What are the answers to the two previous questions if $\beta = 0$.

c) Write down the log likelihood for the model when $w(t)$ is given by equation (2). Use the log likelihood to write down the equations that will determine the maximum likelihood estimates (MLE) of α and β. (You do not need to solve these equations.)

d) Let $N(\tau)$ be the number of failures up to time $\tau = 900$ days. Show that the MLE of α satisfies

$$\hat{\alpha} = \ln\left(\frac{N(\tau)\cdot\hat{\beta}}{\exp(\hat{\beta}\cdot\tau) - 1}\right)$$

where $\hat{\beta}$ is the MLE for β. Calculate $\hat{\alpha}$ for the given data set when you get to know that $\hat{\beta} = 0.003$.
Draw the estimated curve for the expected number of failures as a function of time on the same graph as the plot in (a).
Does it seem that the model in (2) gives a reasonable fit to the data?