Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Contact during exam:
Bo Lindqvist 975 89 418

EXAM IN TMA4275 LIFETIME ANALYSIS

Monday 18 May 2009

Time: 09:00–13:00

Aids:
All printed and handwritten aids permitted, approved simple calculator.

Grading: 9 June 2009

ENGLISH

Problem 1 *Lung cancer study*

The table on the next page contains data on the survival of 10 patients with inoperative lung cancer, where the lifetimes are measured in days from time of entering the study and until death. Complete follow-up was obtained on all patients so that the exact lifetimes T_i are observed (see table).

As is common in studies of this kind, an interim analysis was conducted on a particular day, by which only 5 patients had died. The censored data obtained on that date are the Y_i given in the table, and the D_i are the statuses at that day, with $D_i = 1$ if patient is dead and $D_i = 0$ if patient is alive at the day of the interim analysis.

i	Y_i	D_i	T_i
1	27	1	27
2	108	0	441
3	124	1	124
4	136	1	136
5	170	1	170
6	186	0	191
7	210	1	210
8	231	0	308
9	306	0	511
10	410	0	559

a) Compute the Kaplan-Meier estimator $\hat{R}(t)$ of the survival function $R(t)$ for the lifetime T of a patient with inoperative lung cancer, based on the censored data

$$(Y_i, D_i);\ i = 1, 2, \ldots, 10.$$

Draw the estimated curve.

The probability of surviving 6 months, i.e. 180 days, is of interest. What is the estimate $\hat{R}(180)$?

Compute the estimated standard deviation of $\hat{R}(180)$ and use this to find an approximate 95% confidence interval for $R(180)$ based on the interim data.

b) Draw the empirical survival function $R^*(t)$ for T based on the complete survival data T_i; $i = 1, 2, \ldots, 10$ in the same diagram as the Kaplan-Meier estimate in point (a).

What is the estimate $R^*(180)$? Compute the estimated standard deviation of $R^*(180)$. (*Hint:* One way of doing this is by using that the random variable

$$Z(180) = \text{ number of death times that are} > 180$$

is binomially distributed $(10, R(180))$, and expressing $R^*(180)$ by $Z(180)$).

Compare the results of this point with the results of point (a) and give a comment.

c) The mean time to death, $E(T)$, is of interest in the study. A commonly used method to estimate $E(T)$ from censored data is by computing the area under the Kaplan-Meier curve until the last observed time (censored or non-censored).

Explain the reasoning behind this estimator and do the computation using the censored data (Y_i, D_i) and results from point (a).

How would you estimate $E(T)$ from the complete data T_i? Compute also this estimate of $E(T)$ and compare it to the result based on the censored data.

Problem 2 *Copy machines I*

A company has installed copy machines in several of its divisions. Two machines of the same type are installed in two such divisons. The machines are serviced every 30 days and the company records all failure times between two services, with time measured from 0 to 30 (days). It is assumed that repairs are minimal and take a negligible time, so that failure times of the two machines are assumed to follow non-homogeneous Poisson processes (NHPPs). Further, the failure processes corresponding to the two machines are assumed independent.

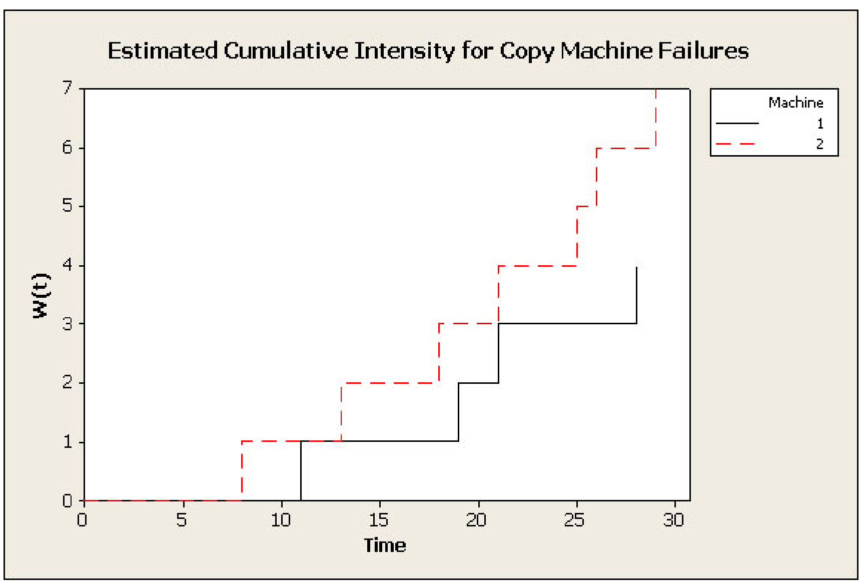
Let the two machines under study have possibly different intensity functions (ROCOF functions) $w_1(t)$ and $w_2(t)$, respectively, and let $W_1(t)$ and $W_2(t)$ denote the corresponding cumulative intensity functions (CROCOF functions).

The following failure times (in days) were observed in a particular 30 days period:

Copy machine 1: 11, 19, 21, 28.

Copy machine 2: 8, 13, 18, 21, 25, 26, 29.

a) The figure below shows estimated curves $\hat{W}_1(t)$ and $\hat{W}_2(t)$.



How are the estimated curves calculated?

Do the curves indicate a trend in the ROCOF functions of the copy machines?

What are the practical interpretations of $W_1(30)$ and $W_2(30)$?

Write down the estimates $\hat{W}_1(30)$ and $\hat{W}_2(30)$ together with their estimated standard deviations.

b) To investigate whether there is a trend in the ROCOF functions $w_1(t)$ and $w_2(t)$ one performed Military Handbook trend tests separately for each of the two machines.

What are the null hypotheses and alternative hypotheses of these tests, and what are the conclusions if you use significance level 5% for each test?

Finally a pooled Military Handbook test was performed for the data from the two machines.

Write down the null hypothesis and alternative hypothesis in this case.

What is the conclusion if 5% significance level is used for the pooled test?

Problem 3 *Copy machines II*

Consider the same situation as in Problem 2.

One now assumes that the failure times of the copy machines follow NHPPs with parametric ROCOF functions of the power-law type. More precisely it is assumed that the ROCOF functions $w_1(t)$ and $w_2(t)$ are given by, respectively,

$$w_1(t) = \lambda\beta t^{\beta-1} \quad (1)$$
$$w_2(t) = \delta\lambda\beta t^{\beta-1} \quad (2)$$

for $t > 0$, where $\lambda > 0$, $\beta > 0$, $\delta > 0$ are unknown parameters. Here δ describes the possible divergence in ROCOF between the two machines, with $\delta = 1$ corresponding to the two machines having the same ROCOF function. Note also that the power β is assumed to be the same for the two machines.

Let, respectively, N_1 and N_2 be the number of failures observed for the two machines in a 30 days period.

a) What are the probability distributions of N_1 and N_2 under the given assumptions?

Express $E(N_1)$ and $E(N_2)$ in terms of the parameters λ, β, δ.

Can you give a simple interpretation of the parameter δ in terms of these expectations?

b) Write down the general expression for the log-likelihood function for failure times $S_1, S_2, \ldots, S_{N(\tau)}$ for a single NHPP with ROCOF function $w(t)$ and CROCOF function $W(t)$, observed on a time interval $[0, \tau]$.

Use this to show that the log-likelihood function for all the failure times of the two copy machines, $S_{11}, S_{21}, \ldots, S_{N_1 1}$ for machine 1, and $S_{12}, S_{22}, \ldots, S_{N_2 2}$ for machine 2, can be written

$$l(\lambda, \beta, \delta) = (N_1 + N_2)\ln\lambda + (N_1 + N_2)\ln\beta + N_2 \ln\delta + (\beta - 1)U - \lambda(1 + \delta)\cdot 30^{\beta}$$

where

$$U = \sum_{i=1}^{N_1} \ln S_{i1} + \sum_{i=1}^{N_2} \ln S_{i2}$$

and ln is the natural logarithm.

It can be shown (you should not do this) that the maximum likelihood estimators of the parameters are given by

$$\begin{aligned} \hat{\delta} &= \frac{N_2}{N_1} \\ \hat{\beta} &= \frac{N_1 + N_2}{(N_1 + N_2)\ln 30 - U} \\ \hat{\lambda} &= \frac{N_1 + N_2}{(1 + \hat{\delta}) \cdot 30^{\hat{\beta}}} \end{aligned}$$

c) Calculate the estimates for λ, β and δ for the data given in Problem 2. You can use that $U = 32.1426$ without computing this yourself.

The inverse of the observed information matrix of (λ, β, δ) (in this order) is given by the following:

$$\begin{bmatrix} 5.278 \cdot 10^{-5} & -4.451 \cdot 10^{-3} & -1.446 \cdot 10^{-3} \\ -4.451 \cdot 10^{-3} & 0.3960 & 0 \\ -1.446 \cdot 10^{-3} & 0 & 1.203 \end{bmatrix}$$

Use this to compute an approximate 95% confidence interval for β.

How can you use this confidence interval to test the null hypothesis $H_0 : \beta = 1$ against the alternative hypothesis $H_1 : \beta \neq 1$? What is then the conclusion? (You need not give detailed reasons for the method used here).

Explain why the above null hypothesis $H_0 : \beta = 1$ exactly corresponds to the null hypothesis tested with the pooled Military Handbook test in Problem 2(b).

d) The hypotheses in point (c) can also be tested by using the log-likelihood function $l(\lambda, \beta, \delta)$.

Explain how this can be done and perform the test with approximate significance level 5%.

To simplify the computations you can use without computing it yourself that $l(\hat{\lambda}, \hat{\beta}, \hat{\delta}) = -26.88$.