Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Contact during exam:
Bo Lindqvist 975 89 418

EXAM IN TMA4275 LIFETIME ANALYSIS

Saturday 4 June 2005

Time: 09:00–13:00

Aids:
All printed and handwritten aids. All calculators permitted.

Grading: 25 June 2005.

Problem 1

A study has been conducted to find out how the lifetime of a particular brand of brake pads depends on the type of driving. Six cars of the same make and year were selected, and the type of driving for each of them was classified by the variable w with values $0 =$ *mainly city driving*, $1 =$ *mixed driving*, or $2 =$ *mainly highway driving*.

The following data were obtained:

y_i	d_i	w_i	x_{i1}	x_{i2}
22.7	1	0	0	0
41.0	0	1	1	0
54.2	1	2	0	1
59.8	0	0	0	0
62.4	1	1	1	0
73.1	0	2	0	1

Here y_i (with unit 1000 km) is the observed brake pad lifetime and d_i is the censoring status for car number i ($d_i = 1$ means non-censored, $d_i = 0$ means right censored). The covariate w_i is the value of the variable w for type of driving. In the analysis we will represent w by the covariates x_1 and x_2, where $x_1 = 1$ if $w = 1$ and $x_1 = 0$ otherwise; and $x_2 = 1$ if $w = 2$ and $x_2 = 0$ otherwise. The values of x_1 and x_2 for each car are also given in the table for simplicity.

In the study one uses a Cox model for the data.

a) Write down an expression for the hazard rate (failure rate) of the lifetime of a brake pad as a function of time t and the covariates x_1 and x_2. What are the underlying assumptions when a Cox model is used for this situation?

Find an expression for the reliability function (survival function) of a brake pad of a car which is mainly used on highways.

b) Calculate an expression for Cox' partial likelihood for the given data.

c) It turns out that the maximum value of the log partial likelihood is -3.67. Use this to compute the value of a test statistic for testing the null hypothesis that type of driving has no influence on the lifetime of the brake pads.

What is the conclusion of this test? Use significance level 5%.

Problem 2

One has recorded times of critical failures of a single compressor until $\tau = 650$ days after start. It is assumed that repairs are minimal and take a negligible time. Thus we assume that failure times follow a non-homogeneous Poisson process with intensity function $w(t)$ and cumulative intensity function $W(t)$.

The following data were observed (in days after start):

i	s_i
1	120
2	347
3	420
4	512
5	595

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?

Find an approximate 95% confidence interval for the value of $W(595)$.

b) Perform the Laplace test to investigate whether there is a trend in the failure times. What is the conclusion if you use significance level 5%?

What can the sign of this test statistic tell us about possible increasing or decreasing trend?

In the rest of the problem we assume that the ROCOF $w(t)$ of the failure process is of so called *log linear* type, i.e.

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

where $\alpha > 0$ and $-\infty < \beta < \infty$ are unknown parameters.

c) Show that

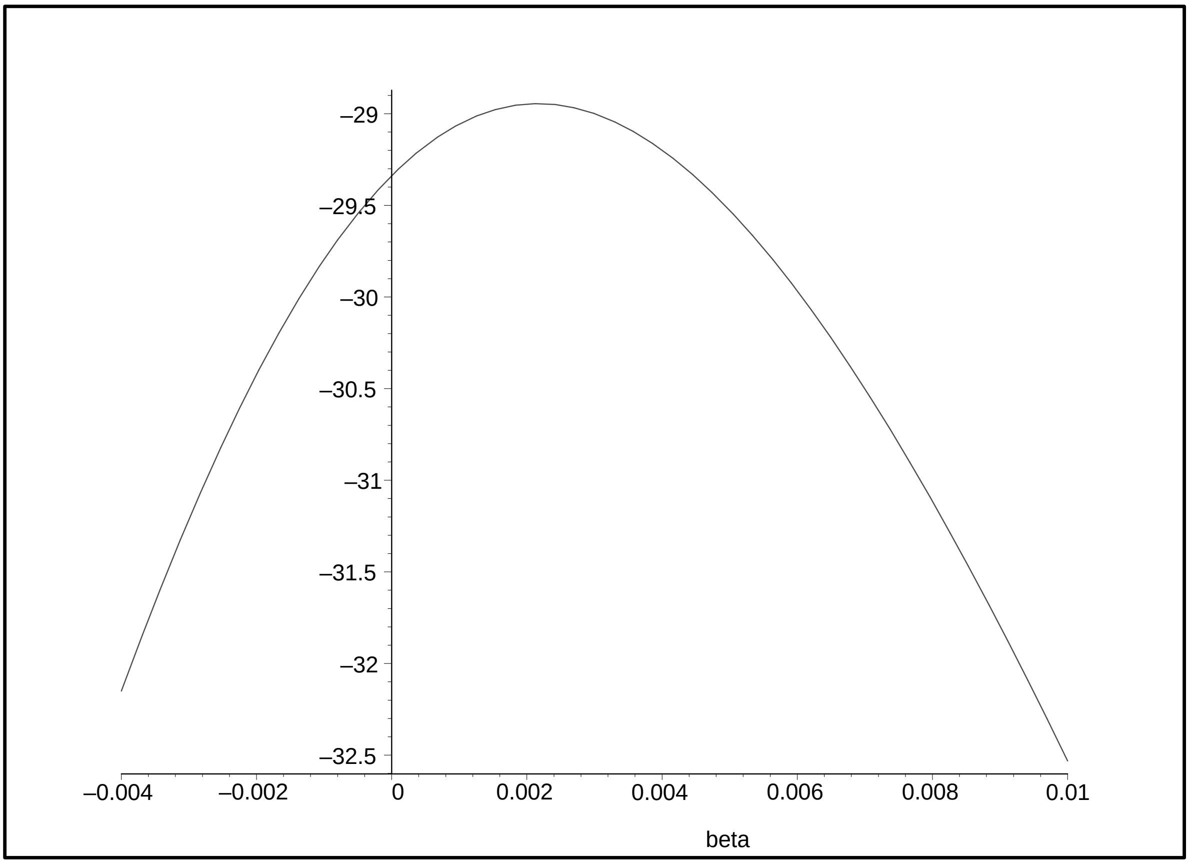
$$W(t) = \frac{e^{\alpha}(e^{\beta t} - 1)}{\beta}.$$

Show that the log likelihood function of the data can be written

$$l(\alpha, \beta) = 5\alpha + 1994\beta - \frac{e^{\alpha}(e^{650\beta} - 1)}{\beta}.$$

d) Calculate the profile log likelihood function $\tilde{l}(\beta)$ of β.

A graph of $\tilde{l}(\beta)$ is given below.



Use the graph (roughly)

1. to find the maximum likelihood estimate of β,
2. to find an approximate 95% confidence interval of β,
3. to test, with significance level 5%, the null hypothesis of no trend in the failure process.

e) Suppose that the data were instead given in intervals only, i.e. as follows:

Interval	Number of failures
$(0, 100]$	0
$(100, 200]$	1
$(200, 300]$	0
$(300, 400]$	1
$(400, 500]$	1
$(500, 600]$	2
$(600, 650]$	0

Find an expression for the log likelihood function in this case.

Problem 3

Let $\Phi_0(w)$ and $\phi_0(w)$ be, respectively, the cumulative distribution function and the probability density function of a random variable W with possible values in $(-\infty, \infty)$. Assume furthermore that $\phi_0(w) > 0$ for $w \in (-\infty, \infty)$.

a) Explain how $\Phi_0(w)$ and $\phi_0(w)$ are used to define a *log-location-scale* family of distributions for a lifetime T, indexed by parameters $-\infty < \mu < \infty$ and $\sigma > 0$.

In particular write down the cumulative distribution function and the probability density function for T, expressed in terms of $\Phi_0(w)$, $\phi_0(w)$, μ and σ.

Let W have the standard *logistic* distribution, i.e. let

$$\Phi_0(w) = \frac{e^w}{1 + e^w} \text{ for } -\infty < w < \infty. \tag{1}$$

Recall that the corresponding family of distributions for T is then called the *log-logistic* family.

b) Show that the log-logistic family of distributions for T can be reparameterized with $\alpha > 0$, $\gamma > 0$ so that the reliability function (survival function) can be written

$$R_T(t) = \frac{1}{1 + \gamma t^\alpha} \text{ for } t > 0.$$

Explain how this can be used to construct a probability plot for checking whether a right censored data sample can be assumed to come from the log-logistic family.

c) Explain how W can be used to define a parametric regression model for lifetimes T with corresponding covariate vector $x = (x_1, \dots, x_k)$.

Give an expression for the reliability function of T with covariate vector x when $\Phi_0(w)$ is the standard logistic distribution function given by (1).