Department of Mathematical Sciences

Examination paper for TMA4275 Lifetime analysis

Examination date: 21.05.2024

Examination time (from-to): 15.00 - 19.00

Permitted examination support material: C

- Approved calculator
- Tabeller og formler i statistikk, Akademika
- A yellow sheet of paper (A5 with a stamp) with personal handwritten formulas and notes

Academic contact during examination: Håkon Tjelmeland
Phone: 4822 1896

Academic contact present at the exam location: Yes, approximately 17.00

OTHER INFORMATION

Language: Your answers may be given in Norwegian or English.

Get an overview of the question set before you start answering the questions.

Read the questions carefully and make your own assumptions. If a question is unclear/vague, make your own assumptions and specify them in your answer. The academic person is only contacted in case of errors or insufficiencies in the question set. Address an invigilator if you suspect errors or insufficiencies. Write down the question in advance.

Hand drawings: For all the questions you should give your answers on handwritten sheets. At the bottom of the question you will find a seven-digit code. Fill in this code in the top left corner of the sheets you wish to submit. We recommend that you do this during the exam. If you require access to the codes after the examination time ends, click “Show submission”.

Weighting: The exam consists of eight parts. These eight parts are given equal weight in the evaluation of your solution.

Notifications: If there is a need to send a message to the candidates during the exam (e.g. if there is an error in the question set), this will be done by sending a notification in Inspera. A dialogue box will appear. You can re-read the notification by clicking the bell icon in the top right-hand corner of the screen.

Withdrawing from the exam: If you become ill or wish to submit a blank test/withdraw from the exam for another reason, go to the menu in the top right-hand corner and click “Submit blank”. This cannot be undone, even if the test is still open.

Access to your answers: After the exam, you can find your answers in the archive in Inspera. Be aware that it may take a working day until any hand-written material is available in the archive.

1 Hazard rate and lifetime

Introduction: Let T be a lifetime for which the hazard rate is given by

$$\alpha(t) = 2\sqrt{t}.$$

Problem: Find the two probabilities

$$P(T > 1.5) \quad \textbf{and} \quad P(T \leq 2|T > 1.5).$$

Find also the median value for T.

Maximum marks: 10

2 Cox model in R

Introduction: The R output included below shows the result when estimating a Cox model to a survival time data set. The fitted model includes three time invariant covariates; x1, x2 and x3. All three covariates are continuous and have values ranging approximately between -3.0 and 3.0.

R-output:

```
summary(res.cox)
Call:
coxph(formula = Surv(time, death) ~ x1 + x2 + x3, data = dataset)

  n= 250, number of events= 170

       coef    exp(coef)  se(coef)     z      Pr(>|z|)
x1  0.04681   1.04793    0.09089   0.515   0.60651
x2 -0.29279   0.74618    0.09810  -2.985   0.00284 **
x3  0.15813   1.17131    0.09290   1.702   0.08873 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

   exp(coef) exp(-coef) lower .95 upper .95
x1    1.0479    0.9543    0.8769    1.2523
x2    0.7462    1.3402    0.6157    0.9044
x3    1.1713    0.8537    0.9763    1.4052

Concordance= 0.547  (se = 0.027 )
Likelihood ratio test = 9.6  on 3 df,   p=0.02
Wald test            = 9.48  on 3 df,   p=0.02
Score (logrank) test = 9.52  on 3 df,   p=0.02
```

Problem: Write down the estimated relative risk function.

Find a 95% confidence interval for the ratio of the hazard rate of an individual with x2$= -2.0$ over the hazard rate of an individual with x2$= 2.0$, when the x1 and x3 values of the two individuals are the same.

If you were to analyse this data set further, discuss what your next steps in the analysis would be.

Maximum marks: 10

3 Estimation in a parametric model

Introduction: Consider a situation where we observe failures occurring in a number of units. The units are of two types, type A and type B. We observe failures for n units of type A og for m units of type B. Assume we have no censoring and that whenever a unit is failing it is immediately repaired. We assume that a unit that has been repaired may fail again and that the intensity for failure is not influenced by the previous failures/repairs. We assume there is no limit on how many times a unit can fail, so the number of units at risk is the same at all times.

Letting t denote time from a unit is set in use for the first time, assume that the units of type A has an intensity for failure given by

$$\lambda_A(t) = \alpha t^{\gamma-1},$$

whereas the units of type B has an intensity for failure given by

$$\lambda_B(t) = \beta t^{\gamma-1}.$$

Note that the value for γ is the same for both type A and type B units, so the model has three unknown parameters, namely α, β and γ.

Problem A: Assuming we have observed the $n+m$ units up to a time τ, find a formula for the likelihood function. Introduce and define necessary notation in addition to what is given above so that you can make your answer precise.

Show that the log-likelihood function can be expressed as

$$\ell(\alpha,\beta,\gamma) = s_0 + s_1 \ln \alpha + s_2 \ln \beta + s_3 \gamma - \frac{n\alpha}{\gamma}\tau^{\gamma} - \frac{m\beta}{\gamma}\tau^{\gamma},$$

where s_0, s_1, s_2 and s_3 are functions of the observations. Find thereby also expressions for s_0, s_1, s_2 and s_3.

Problem B: If possible, find explicit formulas for the maximum likelihood estimators for α, β and γ. If this is not possible, optimise analytically with respect to as many of the three parameters as possible and find an expression for the profile likelihood for the remaining parameter(s). *Note: You may do your derivations and specify your answers in terms of the observed statistics s_0, s_1, s_2 and s_3.*

Maximum marks: 20

4 One sample nonparametric test

Introduction: Let $N(t)$ be a counting process with a multiplicative intensity process $\lambda(t) = Y(t)\alpha(t)$, where $\alpha(t)$ is the hazard rate and $Y(t)$ is a non-negative predictable process with respect to the history $\{\mathcal{F}_t\}$ generated by $N(t)$. As usual we let $A(t) = \int_0^t \alpha(s)ds$ denote the corresponding integrated hazed rate, and define the Nelson-Aalen estimator by

$$\hat{A}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s),$$

where $J(t) = I(Y(t) > 0)$ is equal to one when $Y(t) > 0$ and zero otherwise.

In this problem we will consider a hypothesis test situation where the null hypothesis is

$$H_0 : \alpha(t) = \alpha_0(t) \text{ for } t \in [0, t_0],$$

where $\alpha_0(t)$ is a given (known) hazard rate. To define when to reject H_0 we will use the statistic

$$Z(t_0) = \int_0^{t_0} Y(s)\left(d\hat{A}(s) - dA_0^\star(s)\right),$$

where $A_0^\star(t) = \int_0^t J(s)\alpha_0(s)ds$.

Note: In this problem you can use without proof that the Doob-Meyer decomposition of a counting process is

$$N(t) = \int_0^t \lambda(s)ds + M(t),$$

where $M(t)$ is a mean zero martingale with respect to the history generated by $N(t)$,, and that the predictable variation process of a counting process martingale is

$$\langle M \rangle(t) = \int_0^t \lambda(s)ds.$$

Problem A: Use the Doob-Meyer decomposition of $N(t)$ to show that

$$\hat{A}(t) - A_0^\star(t) = \int_0^t J(s)(\alpha(s) - \alpha_0(s))ds + \int_0^t \frac{J(s)}{Y(s)} dM(s).$$

Thereafter use this to find an expression for $E[Z(t_0)]$, and in particular show that this expectation is equal to zero when H_0 is true.

Problem B: In the following you should assume H_0 to be true.

Find the predictable variation process of $Z(t_0)$.

Use the derived expression for $\langle Z \rangle(t_0)$ to define an unbiased estimator for the variance of $Z(t_0)$.

Explain how the above can be used to decide when to reject H_0.

Maximum marks: 20

5 Nelson-Aalen for a Markov chain

Introduction: Assume we have $n = 5$ units that we number from 1 to $5.$ At any time a unit can be in three possible states which we denote by $0, 1$ and $2,$ respectively. We let $X_i(t)$ denote the state of unit i at time $t,$ and assume transitions between the three possible states occur according to a Markov process where we denote by $\alpha_{gh}(t)$ the transition intensity from state g to state h for $g \neq h,$ at time $t.$ We assume it to be known that $\alpha_{12}(t) = \alpha_{20}(t) = \alpha_{21}(t) = 0,$ whereas the remaining transition intensities $\alpha_{01}(t), \alpha_{02}(t)$ and $\alpha_{10}(t)$ are unknown. Note that this in particular means that state 2 is an absorbing state.

All the $n = 5$ units start in state 0 at time $t = 0,$ and we observe them until they reach the absorbing state. In the following we specify the observations for each of the units by giving the times at which they change state and (in parentheses) the state it is entering at this time.

Unit $1:$ 0.53 (1), 1.05 (0), 1.39 (2)
Unit $2:$ 1.02 (2)
Unit $3:$ 1.54 (1), 2.01 (0), 2.79 (1), 2.84 (0), 2.94 (2)
Unit $4:$ 0.11 (1), 0.44 (0), 0.92 (1), 1.53 (0), 2.39 (2)
Unit $5:$ 0.57 (2)

So unit $1,$ for example, is in state 0 in the time interval $[0, 0.53),$ is thereafter in state 1 in the time interval $[0.53, 1.05),$ is back in state 0 in the time interval $[1.05, 1.39),$ and finally enter the absorbing state at time $1.39.$

Hint: To solve this problem you may use the following expressions that we have discussed in class, where the notation is as defined in class.

The Nelson-Aalen estimator for the non-diagonal elements in the matrix $\mathbf{A}(t) = [A_{gh}(t)],$ where $A_{gh}(t) = \int_0^t \alpha_{gh}(s)ds, g \neq h,$ is given by

$$\hat{A}_{gh}(t) = \int_0^t \frac{J_g(s)}{Y_g(s)} dN_{gh}(s),\ g \neq h,$$

For $g = h$ we have $\hat{A}_{gg}(t) = -\sum_{h \neq g} \hat{A}_{gh}(t).$

The Kaplan-Meier estimator for the matrix $\mathbf{P}(s,t) = [P_{gh}(s,t)],$ where $P_{gh}(s,t) = P(X(t) = h | X(s) = g),\ t > s$ is given by

$$\hat{\mathbf{P}}(s,t) = \prod_{j: s < T_j \leq t} \left(\mathbf{I} + \Delta\hat{\mathbf{A}}(T_j) \right),$$

where $\Delta\hat{\mathbf{A}}(T_j) = \hat{\mathbf{A}}(T_j) - \hat{\mathbf{A}}(T_j-).$

Problem A: Use the above data to evaluate by hand the Nelson-Aalen estimator $\hat{A}_{01}(t)$ for $t \in [0, 1.0].$ Make a sketch of your result.

Problem B: Use the above data to evaluate by hand the Kaplan-Meier estimator $\hat{\mathbf{P}}(0.50, 0.75).$

Maximum marks: 20