Department of Mathematical Sciences

Examination paper for TMA4275 Lifetime Analysis

Examination date: June 2nd 2021

Examination time (from-to): 09.00 - 13.30

Permitted examination support material: A / All support material is allowed

Academic contact during examination: Håkon Tjelmeland
Phone: 4822 1896

Technical support during examination: Orakel support services
Phone: 73 59 16 00

If you experience technical problems during the exam, contact Orakel support services as soon as possible before the examination time expires. If you don't get through immediately, hold the line until your call is answered.

OTHER INFORMATION

Make your own assumptions: If a question is unclear/vague, make your own assumptions and specify them in your answer. Only contact academic contact in case of errors or insufficiencies in the question set.

Cheating/Plagiarism: The exam is an individual, independent work. Examination aids are permitted, but make sure you follow any instructions regarding citations. During the exam it is not permitted to communicate with others about the exam questions, or distribute drafts for solutions. Such communication is regarded as cheating. All submitted answers will be subject to plagiarism control. *Read more about cheating and plagiarism here.*

Citations: If using expressions from the course textbook you should refer to the equation numbers in the textbook, explain why this expression is valid in the situation considered in your problem, and start by copying the expression in 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. All candidates will also receive an SMS to ensure that nobody misses out on important information. Please keep your phone available during the exam.

Weighting: The exam set consists of nine items, Problems 1, 2, 3, 4, 5, 6, 7a, 7b and 7c. These nine items are given equal weight in the evaluation.

ABOUT SUBMISSION

How to answer questions: In problems that ask for a file upload, the uploaded file should be a **pdf file**. The uploaded files should only contain R input/output and your **handwritten** derivations and discussions. **It is strongly advised that you upload a solution as soon as you are finished**

with an exercise. You should not upload all the solutions at the end of the exam period as this may cause technical problems.

When correcting the exam emphasis will be placed on the answers being logical and that they contain all natural intermediate calculation. Except in the multiple choice questions in Problems 2 and 5, you should give reasons for all your answers.

File upload: When working in other programs because parts of/the entire answer should be uploaded as a file attachment – make sure to save your work regularly.

All files must be uploaded before the examination time expires.

The file types allowed are specified in the upload assignment(s).

30 minutes are added to the examination time to manage the sketches/calculations/files. The additional time is included in the remaining examination time shown in the top left-hand corner.

NB! You are responsible to ensure that the file(s) are correct and not corrupt/damaged. Check the file(s) you have uploaded by clicking "Download" when viewing the question. All files can be removed or replaced as long as the test is open.

How to digitize your sketches/calculations
How to create PDF documents
Remove personal information from the file(s) you want to upload

Automatic submission: Your answer will be submitted automatically when the examination time expires and the test closes, if you have answered at least one question. This will happen even if you do not click "Submit and return to dashboard" on the last page of the question set. You can reopen and edit your answer as long as the test is open. If no questions are answered by the time the examination time expires, your answer will not be submitted. This is considered as "did not attend the exam".

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.

Accessing your answer post-submission: You will find your answer in Archive when the examination time has expired.

1 1A

Introduction: Let T be a lifetime for which the survival function is given as

$$S(t) = P(T > t) = \frac{1+\theta+2\theta t}{1+\theta} \cdot e^{-2\theta t} \quad \text{for} \quad t \geq 0,$$

where $\theta > 0$ is a parameter.

Problem: Derive formulas for the density function for T, $f(t)$, and for the hazard rate $\alpha(t)$.

Note: You should upload a pdf file that contains image(s) of your **handwritten** solution to the problem. When correcting the exam emphasis will be placed on the answers being logical and that they contain all natural intermediate calculation.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

2 1B

Introduction: Let T be a lifetime for which the survival function is given as

$$S(t) = P(T > t) = \frac{1 + \phi + 4\phi t}{1 + \phi} \cdot e^{-4\phi t} \quad \text{for} \quad t \geq 0,$$

where $\phi > 0$ is a parameter.

Problem: Derive formulas for the density function for T, $f(t)$, and for the hazard rate $\alpha(t)$.

Note: You should upload a pdf file that contains image(s) of your **handwritten** solution to the problem. When correcting the exam emphasis will be placed on the answers being logical and that they contain all natural intermediate calculation.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

3 1C

Introduction: Let T be a lifetime for which the survival function is given as

$$S(t) = P(T > t) = \frac{1 + \beta + 5\beta t}{1 + \beta} \cdot e^{-5\beta t} \quad \text{for} \quad t \geq 0,$$

where $\beta > 0$ is a parameter.

Problem: Derive formulas for the density function for T, $f(t)$, and for the hazard rate $\alpha(t)$.

Note: You should upload a pdf file that contains image(s) of your **handwritten** solution to the problem. When correcting the exam emphasis will be placed on the answers being logical and that they contain all natural intermediate calculation.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

4 2

Introduction: Let $\{N(t); t \geq 0\}$ be a counting process with intensity process $\lambda(t)$, i.e.

$\lambda(t)dt = P(dN(t) = 1|\mathcal{F}_{t-})$,

where $\{\mathcal{F}_t\}$ is the history containing all information about $N(t)$ up to time and including time t.

Problem: Adopting the standard notation used in the course, specify which of the following statements you from the above information can know for sure is true.

Select the statements you can know are true.

- ☐ $P(dN(t) > 1) = 0$
- ☐ $N(t)$ is Poisson distributed with mean value $\Lambda(t) = \int_0^t \lambda(s)ds$
- ☐ $N(t)$ is a sub-martingale
- ☐ $2N(t) + 5$ is a counting process
- ☐ For all $0 < t_1 < t_2 < t_3$, $N(t_2) - N(t_1)$ and $N(t_3) - N(t_2)$ are independent
- ☐ $\lambda(t) \in [0, 1]$ for all $t \geq 0$

Maximum marks: 10

5 3

Introduction: The table below shows the results of a study where children with leukemia are treated with a drug (6-MP) to prevent relapse. Some of the children were treated with the drug 6-MP and others got placebo. In the table, right censored observations are marked with an asterix (*).

Placebo: 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23
6-MP: 6, 6, 6, 6*, 7, 9*, 10, 10*, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35*

These data are also given in Exercise 3.1 in page 126 in the course textbook, and can be downloaded from https://www.mn.uio.no/math/english/people/aca/borgan/data/leukemia.txt.

Problem: Use commands in R to compute and plot Kaplan-Meier estimates for the placebo group and for the 6-MP group. Include in your plots also confidence intervals.

Explain how you from such a plot can find a confidence interval for the median survival time. What are the confidence intervals for the median survival times in each of the two groups?

Note: You should upload a pdf file that contains image(s) of your R commands (handwritten or copied from your R session), the plots produced by R and a **handwritten** explanation of how to find a confidence interval from an R plot.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

6 4

Introduction: The R output included below shows the result when estimating a Cox model to a set of survival times for patients with lung cancer. The fitted model includes four time invariant covariates; sex, age, wt.loss and ph.karno. The covariate sex is binary (1 for males and 2 for females), age is a continuous covariate with values ranging from 39 to 82, wt.loss is also continuous and has values ranging from -24 to 68, and ph.karno is continuous having values ranging from 50 to 100.

R-output:

```
summary(res.cox)
Call:
coxph(formula = Surv(time, status) ~ sex + age + wt.loss + ph.karno, data = lung)

  n= 214, number of events= 152 (14 observations deleted due to missingness)

              coef     exp(coef)  se(coef)      z     Pr(>|z|)
sex        -0.513955   0.598125   0.174410   -2.947   0.00321 **
age         0.015140   1.015255   0.009837    1.539   0.12379
wt.loss    -0.002246   0.997757   0.006357   -0.353   0.72389
ph.karno   -0.012871   0.987211   0.006184   -2.081   0.03741 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Problems: Write down the estimated relative risk function.

What covariates are significant at the 5% level?

Find a 95% confidence interval for the ratio of the hazard rate of a male over the hazard rate of a female of the same age and with the same values for the covariates wt.loss and ph.karno.

In the estimated model, which of the four covariates have the largest effect on the survival probability?

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

Note: You should upload a pdf file that contains image(s) of your **handwritten** solution to the problem. When correcting the exam emphasis will be placed on the answers being logical and that they contain all natural intermediate calculation.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

7 5A

Introduction: Let $\{M_n; n = 0, 1, 2, \ldots\}$ be a discrete time mean zero martingale with respect to a history $\{\mathcal{F}_n\}$, and let H_n be a predictable discrete time process with respect to the same history $\{\mathcal{F}_n\}$.

Problem: Based on the information about the processes M_n and H_n given above, which of the following processes can you conclude is a mean zero martingale?

Select the processes that are mean zero martingales.

- [ ] $8M_n$
- [ ] $2M_n + 4$
- [ ] $M_n^2 + M_n$
- [ ] $(H \bullet M)_n = \sum_{s=1}^{n} H_s(M_s - M_{s-1})$
- [ ] $(H^2 \bullet M)_n = \sum_{s=1}^{n} H_s^2(M_s - M_{s-1})$
- [ ] $\sum_{s=1}^{n} (2s+1)H_s(M_s - M_{s-1})$

Maximum marks: 10

8 5B

Introduction: Let $\{M_n; n = 0, 1, 2, \ldots\}$ be a discrete time mean zero martingale with respect to a history $\{\mathcal{F}_n\}$, and let H_n be a predictable discrete time process with respect to the same history $\{\mathcal{F}_n\}$.

Problem: Based on the information about the processes M_n and H_n given above, which of the following processes can you conclude is a mean zero martingale?

Select the processes that are mean zero martingales.

- ☐ $(H \bullet M)_n = \sum_{s=1}^{n} H_s(M_s - M_{s-1})$
- ☐ $\sum_{s=1}^{n}(2s+3)H_s(M_s - M_{s-1})$
- ☐ $4M_n$
- ☐ $(H^2 \bullet M)_n = \sum_{s=1}^{n} H_s^2(M_s - M_{s-1})$
- ☐ $2M_n + 8$
- ☐ $M_n^2 + M_n$

Maximum marks: 10

9 6

Introduction: Let $\{N(t), t \geq 0\}$ be a counting process with intensity process $\lambda(t)$. Define a process $X(t)$ by

$$X(t) = \sum_{i=1}^{N(t)} Z_i,$$

where $Z_1, Z_2, \dots$ are independent and identically distributed random variables, independent of $N(t)$. Moreover we assume the mean value of the Z_i's to be positive, i.e. $\mu = \mathrm{E}[Z_i] > 0$. Let $\{\mathcal{F}_t\}$ be the history that at time t contains all information about the process $N(t)$ up to and including time t and all information about $Z_i, i = 1, \dots, N(t)$.

Problem: Show that $X(t)$ is a sub-martingale with respect to the history $\{\mathcal{F}_t\}$. *Hint: you may use the law of double expectation by conditioning on the value of* $N(t)$.

Find the compensator $X^\star(t)$ and the martingale process $M(t)$ in the Doob-Meyer decomposition of $X(t)$,

$$X(t) = X^\star(t) + M(t).$$

Note: You should upload a pdf file that contains image(s) of your **handwritten** solution to the problem. When correcting the exam emphasis will be placed on the answers being logical and that they contain all natural intermediate calculation.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

10 7A

Introduction: Consider the following parametric model for possibly right censored survival data. Assume we have n individuals and that we for individual number i observe $(\widetilde{T}_i, D_i)$, where D_i equals one if we observe the survival time for this individual and equals zero if we observe a censoring time, and $\widetilde{T}_i$ is the survival time for individual number i if $D_i = 1$ and is otherwise the censoring time for this individual.

For each individual we have two (time invariant) covariates, x_{i1} and x_{i2} for individual number i, where $x_{i1} \in \{0, 1\}$ and $x_{i2} \in \mathbb{R}$. The hazard rate for the survival time for individual number i we assume to be given by

$$\alpha_i(t) = \nu \exp\{\beta_1 x_{i1} + \beta_2 x_{i2}\},$$

where ν, β_1 and β_2 are parameters that we want to estimate.

Problems:

a) Starting from the general formula for the likelihood function for counting process models given in (5.4) in our textbook,

$$L(\theta) = \left\{ \prod_{i=1}^{n} \prod_{0<t\leq\tau} \lambda_i(t,\theta)^{\Delta N_i(t)} \right\} \exp\left\{ -\int_0^\tau \lambda_\bullet(t;\theta)dt \right\},$$

derive a formula for the log-likelihood function for the survival data situation described above and show that it can be expressed as

$$\ell(\nu, \beta_1, \beta_2) = D_\bullet \ln \nu + \beta_1 D_\bullet^{(1)} + \beta_2 \sum_{i=1}^{n} D_i x_{i2} - \nu \left(\sum_{i:x_{i1}=0} e^{\beta_2 x_{i2}} \widetilde{T}_i + e^{\beta_1} \sum_{i:x_{i1}=1} e^{\beta_2 x_{i2}} \widetilde{T}_i \right)$$

where

$$D_\bullet = \sum_{i=1}^{n} D_i \quad \text{and} \quad D_\bullet^{(1)} = \sum_{i:x_{i1}=1} D_i$$

b) If possible, find explicit formulas for the maximum likelihood estimators for ν, β_1 and β_2. If this is not possible, optimise analytically with respect the parameter(s) where this is possible and explain how the profile likelihood for the remaining parameter(s) can be found.

c) Assume that you have found values for the maximum likelihood estimates $\hat{\nu}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ by maximising the above log-likelihood. Now we want to test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. Find a test statistic you can use to decide whether or not H_0 should be rejected, i.e. explain what type of test you are using and develop the necessary formulas needed to perform the test for the model specified above.

Note: You should upload a pdf file that contains image(s) of your **handwritten** solution to the problem. When correcting the exam emphasis will be placed on the answers being logical and that they contain all natural intermediate calculation.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 30

11 7B

Introduction: Consider the following parametric model for possibly right censored survival data. Assume we have n individuals and that we for individual number i observe $(\widetilde{T}_i, D_i)$, where D_i equals one if we observe the survival time for this individual and equals zero if we observe a censoring time, and $\widetilde{T}_i$ is the survival time for individual number i if $D_i = 1$ and is otherwise the censoring time for this individual.

For each individual we have two (time invariant) covariates, x_{i1} and x_{i2} for individual number i, where $x_{i1} \in \{0, 1\}$ and $x_{i2} \in \mathbb{R}$. The hazard rate for the survival time for individual number i we assume to be given by

$$\alpha_i(t) = \kappa \exp\{\beta_1 x_{i1} + \beta_2 x_{i2}\},$$

where κ, β_1 and β_2 are parameters that we want to estimate.

Problems:

a) Starting from the general formula for the likelihood function for counting process models given in (5.4) in our textbook,

$$L(\theta) = \left\{ \prod_{i=1}^{n} \prod_{0<t\leq\tau} \lambda_i(t,\theta)^{\Delta N_i(t)} \right\} \exp\left\{ -\int_0^\tau \lambda_\bullet(t;\theta)dt \right\},$$

derive a formula for the log-likelihood function for the survival data situation described above and show that it can be expressed as

$$\ell(\kappa, \beta_1, \beta_2) = D_\bullet \ln\kappa + \beta_1 D_\bullet^{(1)} + \beta_2 \sum_{i=1}^{n} D_i x_{i2} - \kappa \left(\sum_{i:x_{i1}=0} e^{\beta_2 x_{i2}} \widetilde{T}_i + e^{\beta_1} \sum_{i:x_{i1}=1} e^{\beta_2 x_{i2}} \widetilde{T}_i \right)$$

where

$$D_\bullet = \sum_{i=1}^{n} D_i \quad \text{and} \quad D_\bullet^{(1)} = \sum_{i:x_{i1}=1} D_i$$

b) If possible, find explicit formulas for the maximum likelihood estimators for κ, β_1 and β_2. If this is not possible, optimise analytically with respect the parameter(s) where this is possible and explain how the profile likelihood for the remaining parameter(s) can be found.

c) Assume that you have found values for the maximum likelihood estimates $\widehat{\kappa}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ by maximising the above log-likelihood. Now we want to test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. Find a test statistic you can use to decide whether or not H_0 should be rejected, i.e. explain what type of test you are using and develop the necessary formulas needed to perform the test for the model specified above.

Note: You should upload a pdf file that contains image(s) of your **handwritten** solution to the problem. When correcting the exam emphasis will be placed on the answers being logical and that they contain all natural intermediate calculation.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 30

12 7C

Introduction: Consider the following parametric model for possibly right censored survival data. Assume we have n individuals and that we for individual number i observe $(\widetilde{T}_i, D_i)$, where D_i equals one if we observe the survival time for this individual and equals zero if we observe a censoring time, and $\widetilde{T}_i$ is the survival time for individual number i if $D_i = 1$ and is otherwise the censoring time for this individual.

For each individual we have two (time invariant) covariates, x_{i1} and x_{i2} for individual number i, where $x_{i1} \in \{0, 1\}$ and $x_{i2} \in \mathbb{R}$. The hazard rate for the survival time for individual number i we assume to be given by

$$\alpha_i(t) = \rho \exp\{\beta_1 x_{i1} + \beta_2 x_{i2}\},$$

where ρ, β_1 and β_2 are parameters that we want to estimate.

Problems:

a) Starting from the general formula for the likelihood function for counting process models given in (5.4) in our textbook,

$$L(\theta) = \left\{ \prod_{i=1}^{n} \prod_{0<t\leq\tau} \lambda_i(t,\theta)^{\Delta N_i(t)} \right\} \exp\left\{ -\int_0^\tau \lambda_\bullet(t;\theta)dt \right\},$$

derive a formula for the log-likelihood function for the survival data situation described above and show that it can be expressed as

$$\ell(\rho,\beta_1,\beta_2) = D_\bullet \ln \rho + \beta_1 D_\bullet^{(1)} + \beta_2 \sum_{i=1}^{n} D_i x_{i2} - \rho \left(\sum_{i:x_{i1}=0} e^{\beta_2 x_{i2}} \widetilde{T}_i + e^{\beta_1} \sum_{i:x_{i1}=1} e^{\beta_2 x_{i2}} \widetilde{T}_i \right)$$

where

$$D_\bullet = \sum_{i=1}^{n} D_i \quad \text{and} \quad D_\bullet^{(1)} = \sum_{i:x_{i1}=1} D_i$$

b) If possible, find explicit formulas for the maximum likelihood estimators for ρ, β_1 and β_2. If this is not possible, optimise analytically with respect the parameter(s) where this is possible and explain how the profile likelihood for the remaining parameter(s) can be found.

c) Assume that you have found values for the maximum likelihood estimates $\hat{\rho}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ by maximising the above log-likelihood. Now we want to test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. Find a test statistic you can use to decide whether or not H_0 should be rejected, i.e. explain what type of test you are using and develop the necessary formulas needed to perform the test for the model specified above.

Note: You should upload a pdf file that contains image(s) of your **handwritten** solution to the problem. When correcting the exam emphasis will be placed on the answers being logical and that they contains all natural intermediate calculation.

Upload your file here. Maximum one file.

All file types are allowed. Maximum file size is **50 GB**

Select file to upload

Maximum marks: 30