

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Examination in: STK4080 — Survival and event history analysis

Day of examination: Wednesday December 10th, 2008.

Examination hours: 14.30 – 17.30.

This examination set consists of 4 pages.

Appendices: None.

Permitted aids: Approved calculator.

Make sure that your copy of the examination set is complete before you start solving the problems.

Problem 1.

Let T be a survival time, assumed to be continuous.

- a) Define the survival function $S(t)$ and the hazard rate $\alpha(t)$ for T . Prove the relation $S(t) = \exp\left(-\int_0^t \alpha(u)du\right)$.

Let T_1, T_2, \dots, T_n be independent and identically distributed survival times with survival function $S(t)$ and hazard rate $\alpha(t)$. We do not observe the T_i s, however, but only the right-censored survival times $\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_n$ (that are the minimum of survival times and censoring times) and the censoring indicators $D_i = I\{\tilde{T}_i = T_i\}$; $i = 1, \dots, n$.

- b) Describe what is meant by independent censoring.

Assuming censoring to be independent, we may estimate the survival function by the Kaplan-Meier estimator.

- c) Describe how you may use the Kaplan-Meier estimator to estimate the quartiles of the survival distribution. Also describe how you may obtain 95% confidence intervals for the quartiles.

(Continued on page 2.)

In a study of patients with aplastic anemia (a condition where the bone marrow is not producing enough new blood cells), 64 patients got a bone marrow transplantation and were then randomized to treatment with (i) methotrexate (MTX) and cyclosporin (CSP) or (ii) only methotrexate. For each patient one measured the time (in days) from randomization until a life threatening complication occurred (“acute graft versus host disease”). Of the 32 patients who got only MTX, 15 experienced the life threatening complication.

Below is given R-output from the `survfit` command for this group of patients:

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
9	32	1	0.969	0.0308	0.798	0.996
11	31	1	0.938	0.0428	0.773	0.984
12	30	1	0.906	0.0515	0.737	0.969
20	29	2	0.844	0.0642	0.665	0.932
22	27	1	0.813	0.0690	0.629	0.911
25	26	2	0.750	0.0765	0.562	0.866
28	23	2	0.685	0.0826	0.493	0.816
31	21	1	0.652	0.0849	0.460	0.790
35	20	2	0.587	0.0880	0.397	0.736
46	18	1	0.554	0.0890	0.366	0.707
49	17	1	0.522	0.0895	0.336	0.678

- d) Use the R-output to estimate the lower quartile in the distribution of the survival times for the patients who got only MTX. Give a 95% confidence interval for the lower quartile. (Remember that the lower quartile $\xi_{0.25}$ is given by $S(\xi_{0.25}) = 0.75$.)

Problem 2.

Assume that we have two counting process $N_1(t)$ and $N_2(t)$ with intensity processes of the multiplicative form $\lambda_1(t) = \alpha_1(t)Y_1(t)$ and $\lambda_2(t) = \alpha_2(t)Y_2(t)$, respectively. Here $\alpha_1(t)$ and $\alpha_2(t)$ are nonnegative functions, while $Y_1(t)$ and $Y_2(t)$ are predictable processes that do not depend on unknown parameters. (Usually $Y_1(t)$ and $Y_2(t)$ are numbers at risk.) We want to test the null hypothesis $H_0 : \alpha_1(t) = \alpha_2(t)$ for all $t \in [0, t_0]$.

- a) Explain why it is reasonable to base the test on a statistic of the form $Z_1(t_0) = \int_0^{t_0} L(t) \left(d\hat{A}_1(t) - d\hat{A}_2(t) \right)$, where $L(t)$ is a non-negative predictable weight process and $\hat{A}_1(t)$ and $\hat{A}_2(t)$ are the Nelson-Aalen estimators based on $N_1(t)$ and $N_2(t)$.
- b) Show that $Z_1(t_0)$ is a mean zero martingale under the null hypothesis (when considered as a process in t_0) and derive an estimator for the variance of $Z_1(t_0)$.

(Continued on page 3.)

The log-rank test corresponds to the weight process $L(t) = Y_1(t)Y_2(t)/Y_*(t)$, where $Y_*(t) = Y_1(t) + Y_2(t)$. In the study of patients with aplastic anemia described in Problem 1, the log-rank test was used to compare the two treatments. Below is given R-output from the `survdif` command (`treat=0` is the group who got both MTX and CSP, while `treat=1` is the group who got only MTX):

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
treat=0	32	5	10.2	2.65	5.49
treat=1	32	15	9.8	2.75	5.49

Chisq= 5.5 on 1 degrees of freedom, p= 0.0192

- c) Explain what the R-output tells you about the effect of treatment with both MTX and CSP compared to treatment with only MTX.

Problem 3.

Assume that we have counting processes $N_1(t), N_2(t), \dots, N_n(t)$ counting the occurrences of an event of interest for n individuals. For each individual i we have the two covariates x_{i1} and x_{i2} , and we assume that the intensity process of $N_i(t)$ takes the form $\lambda_i(t) = Y_i(t)\{\beta_0(t) + \beta_1(t)x_{i1} + \beta_2(t)x_{i2}\}$; $i = 1, 2, \dots, n$. Here $Y_i(t) = 1$ if individual i is at risk “just before” time t and $Y_i(t) = 0$ otherwise.

- a) Let $\mathbf{B}(t) = (B_0(t), B_1(t), B_2(t))^T$, where $B_j(t) = \int_0^t \beta_j(u)du$; $j = 0, 1, 2$; are the cumulative regression functions. Introduce relevant matrix- and vector-notation and derive an estimator $\hat{\mathbf{B}}(t) = (\hat{B}_0(t), \hat{B}_1(t), \hat{B}_2(t))^T$ for $\mathbf{B}(t)$. Show that the estimator is approximately unbiased.

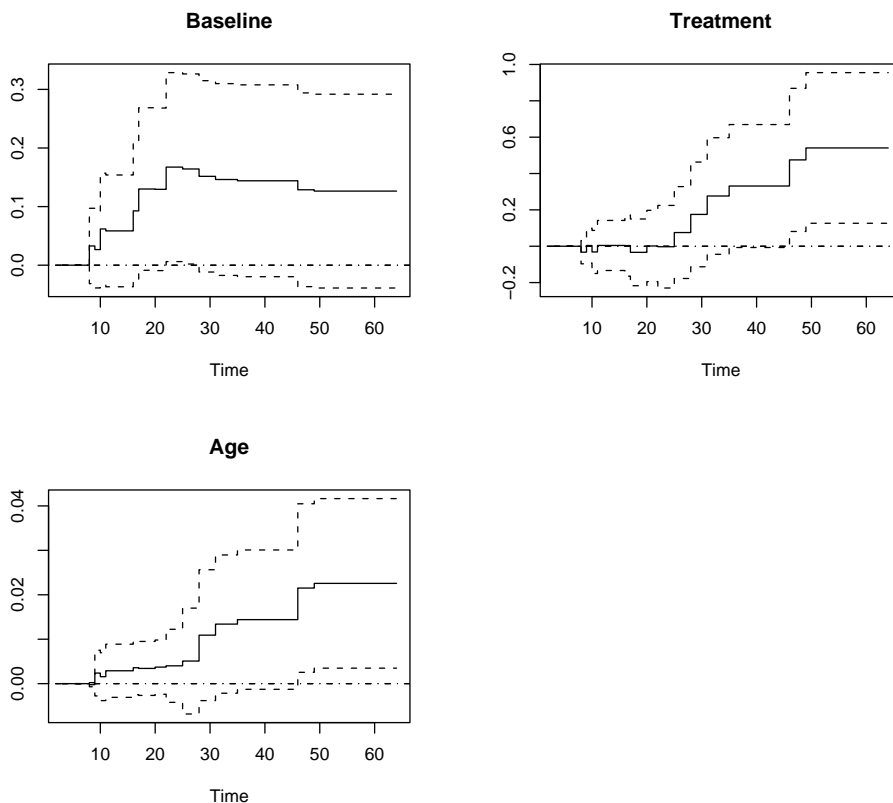
In the study of patients with aplastic anemia described in Problem 1, the ages (in years) of the patients were also recorded. (The mean age was about 20 years.) On the next page plots of $\hat{B}_0(t)$, $\hat{B}_1(t)$, and $\hat{B}_2(t)$ are given for the model with covariates:

$$x_{i1} = \begin{cases} 0 & \text{if the } i \text{ th patient got both MTX and CSP} \\ 1 & \text{if the } i \text{ th patient got only MTX} \end{cases}$$

$$x_{i2} = \text{age of the } i \text{ th patient} - 20$$

- b) Interpret the estimate $\hat{B}_0(t)$ of the cumulative baseline hazard. Also interpret the estimate $\hat{B}_1(t)$ of the cumulative regression function for treatment and the estimate $\hat{B}_2(t)$ of the cumulative regression function for age.

(Continued on page 4.)



Problem 4.

Let $M = \{M_0, M_1, M_2, \dots\}$ be a martingale in discrete time relative to the history $\{\mathcal{F}_t\}$ and assume that $M_0 = 0$.

- Give a mathematical formulation of the martingale property and show that $E(M_n) = 0$ for all n .
- Let $\langle M \rangle$ be the predictable variation process of M . Show that $M^2 - \langle M \rangle$ is a mean zero martingale. Also show that $\text{Var}(M_n) = E\langle M \rangle_n$ for all n .

Let $H = \{H_0, H_1, H_2, \dots\}$ be a predictable process.

- Define the transformation $Z = H \bullet M$ and show that Z is a mean zero martingale. Find an expression for the predictable variation process $\langle Z \rangle$.

END