UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Examination in: STK4080/STK9080 — Survival and event history analysis

Day of examination: Tuesday 9 December 2014.

Examination hours: 09.00 – 13.00.

This problem set consists of 6 pages.

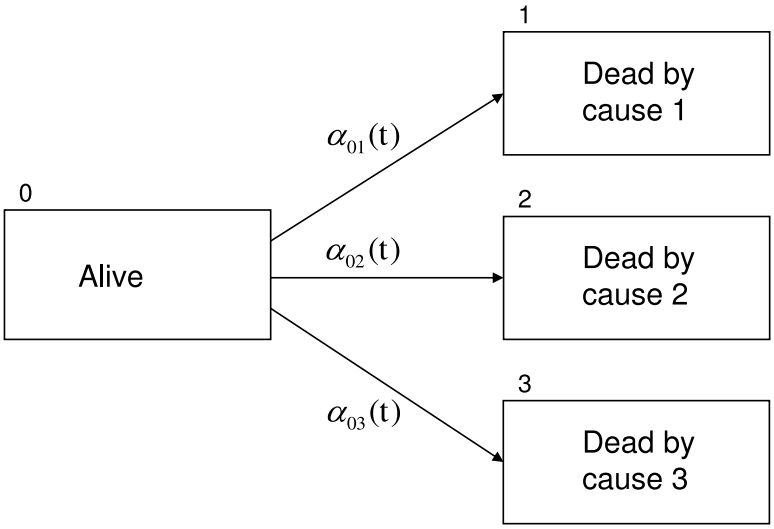
Appendices: None.

Permitted aids: Approved calculator.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

In this problem we will consider the competing risks model with three causes of death. The model is illustrated in the figure below.



a) Explain how the transition probabilities $P_{0h}(s,t)$; $h = 0,1,2,3$; and the cause-specific hazards (or transition intensities) $\alpha_{0h}(t)$; $h = 1,2,3$; are defined. Give formulas that show how the transition probabilities may be expressed by the cause-specific hazards. (You shall *not* prove these formulas.)

Assume that we have a sample of n individuals who are followed from time 0 until death of one of the three causes or to censoring. We denote by $T_1 < T_2 < \cdots$ the times when deaths from any of the three causes are observed, and let $N_{0h}(t)$ be the process that counts the number of individuals who are observed to die from cause h in the time

(Continued on page 2.)

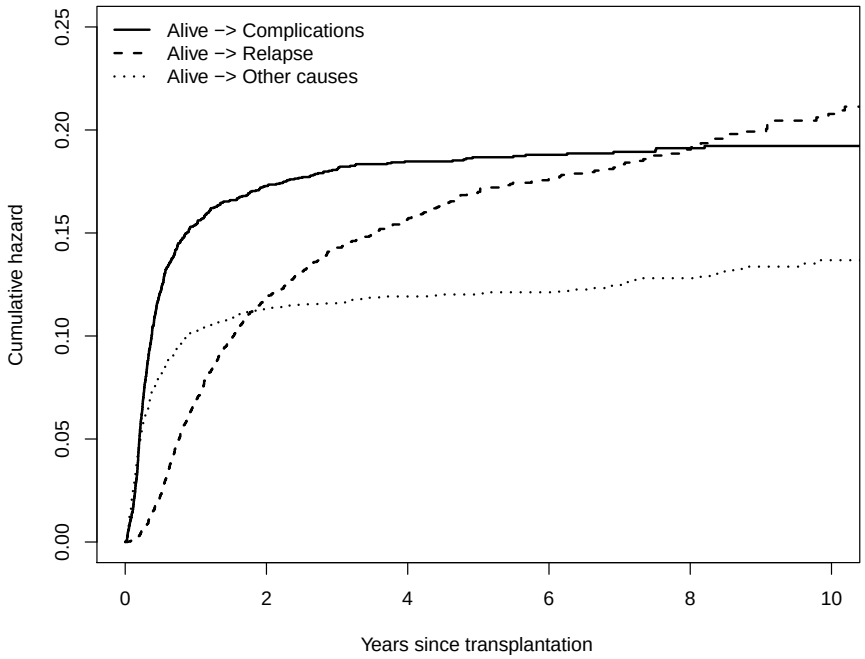
interval $[0, t]$; $h = 1, 2, 3$. Further we let $Y_0(t)$ denote the number of individuals at risk (i.e. in state 0) just before time t.

b) Explain that we may estimate the cumulative cause-specific hazards $A_{0h}(t) = \int_0^t \alpha_{0h}(u)du$ by the Nelson-Aalen estimator. Also give formulas for the empirical transition probabilities $\widehat{P}_{00}(s,t)$ and $\widehat{P}_{0h}(s,t)$; $h = 1, 2, 3$.

From the European Society for Blood and Marrow Transplantation we have data for 4800 leukemia patients who have had a bone marrow transplantation between 20 years and 40 years of age. The patients have been followed after the transplantation, and it is recorded if and when they die

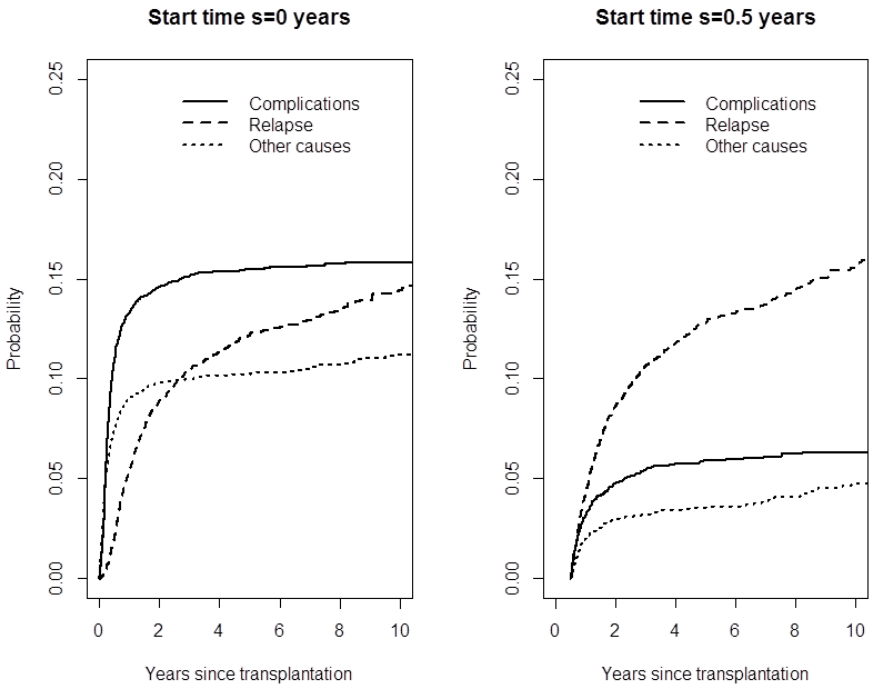
- from a complication of the transplantation (infections, immunological reactions)
- after a relapse of the disease
- from other causes.

c) The figure below gives Nelson-Aalen estimates of the three cumulative cause-specific hazards. Explain what the estimates tell you about the hazards for the three causes of death.



d) The figure on the next page gives plots of the empirical transition probabilities $\widehat{P}_{0h}(s,t)$; $h = 1, 2, 3$ for $s = 0$ years (to the left) and $s = 0.5$ years (to the right). Give an interpretation of these plots.

(Continued on page 3.)



Problem 2

Let $T_1, T_2, \ldots, T_n$ be independent survival times for n individuals. In this problem we will assume that the hazard rates for the individuals are constant, but that they vary between the individuals. More specifically, we assume that we have independent and identically distributed frailties $Z_1, Z_2, \ldots, Z_n$, and that given Z_i the survival time T_i has hazard rate $\alpha(t|Z_i) = Z_i\alpha$; $i = 1, \ldots, n$. Here $\alpha > 0$ is the hazard rate for an individual with unit frailty (i.e. $Z_i = 1$).

We do not observe the T_i's; only the right-censored survival times $\widetilde{T}_1, \widetilde{T}_2, \ldots, \widetilde{T}_n$ and the indicators $D_i = I\{\widetilde{T}_i = T_i\}$; $i = 1, \ldots, n$.

We introduce $H_i = (\widetilde{T}_i, D_i)$; $i = 1, \ldots, n$. Then the conditional likelihood, i.e. the likelihood given the frailties $Z_1, \ldots, Z_n$, is given by

$$L_{\text{cond}} = \prod_{i=1}^{n} P(H_i \,|\, Z_i), \tag{1}$$

where $P(H_i \,|\, Z_i)$ is the contribution to the conditional likelihood for the ith individual.

a) Explain that $P(H_i \,|\, Z_i) = \{Z_i\,\alpha\}^{D_i} \exp\{-Z_i\,\alpha\,\widetilde{T}_i\}$.

The frailties $Z_1, \ldots, Z_n$ are not observed. So the appropriate likelihood to use for statistical inference is the marginal likelihood, i.e. the likelihood for the observed data

(Continued on page 4.)

$(\widetilde{T}_i, D_i)$; $i = 1, 2, \ldots, n$. This is given by

$$L_{\text{marg}} = \prod_{i=1}^{n} P(H_i), \tag{2}$$

where $P(H_i) = E\{P(H_i \,|\, Z_i)\}$ and expectation is with respect to the frailty distribution.

b) Let $\mathcal{L}(c)$ be the Laplace transform of the frailty distribution. Show that

$$P(H_i) = \alpha^{D_i}(-1)^{D_i}\mathcal{L}^{(D_i)}(\alpha\,\widetilde{T}_i),$$

where $\mathcal{L}^{(0)}(c) = \mathcal{L}(c)$ and $\mathcal{L}^{(1)}(c) = \mathcal{L}'(c)$.

We now assume that the frailties are gamma distributed with mean one and variance δ. We then know that $\mathcal{L}(c) = \{1 + \delta c\}^{-1/\delta}$ (and you shall *not* prove this).

c) Give an expression for the logarithm of the marginal likelihood (i.e. $\log L_{\text{marg}}$).

d) Describe briefly how one may test the null hypothesis $H_0 : \delta = 0$ versus the alternative hypothesis $H_A : \delta > 0$.

Problem 3

Assume that we have counting processes $N_1(t), N_2(t), \ldots, N_n(t)$ with no simultaneous jumps that register the occurrences of an event of interest for n individuals. For each individual i we have a covariate x_i, and we assume that the intensity process of $N_i(t)$ takes the form

$$\lambda_i(t) = Y_i(t)\{\beta_0(t) + \beta_1(t)\,x_i\} \tag{3}$$

for $i = 1, 2, \ldots, n$. Here $Y_i(t) = 1$ if individual i is at risk "just before" time t and $Y_i(t) = 0$ otherwise. We introduce the vector

$$\mathbf{N}(t) = (N_1(t), N_2(t), \ldots, N_n(t))^T$$

and the matrix

$$\mathbf{X}(t) = \begin{pmatrix} Y_1(t) & Y_1(t)\,x_1 \\ Y_2(t) & Y_2(t)\,x_2 \\ \vdots & \vdots \\ Y_n(t) & Y_n(t)\,x_n \end{pmatrix}.$$

We further introduce $\mathbf{B}(t) = (B_0(t), B_1(t))^T$, where $B_j(t) = \int_0^t \beta_j(u)du$ for $j = 0, 1$. From the lectures we know that we may estimate $\mathbf{B}(t)$ by

$$\widehat{\mathbf{B}}(t) = \int_0^t J(u)\left\{\mathbf{X}(u)^T\mathbf{X}(u)\right\}^{-1}\mathbf{X}(u)^T d\mathbf{N}(u), \tag{4}$$

where $J(u) = I\{\mathbf{X}(u) \text{ has full rank}\}$. We also introduce $\mathbf{B}^*(t) = \int_0^t J(u)\,d\mathbf{B}(u)$.

(Continued on page 5.)

a) Show that $\widehat{\mathbf{B}}(t) - \mathbf{B}^*(t)$ equals a vector-valued stochastic integral, and explain why this implies that $\widehat{\mathbf{B}}(t)$ is almost an unbiased estimator of $\mathbf{B}(t)$.

From (4) it follows by straightforward matrix multiplication that

$$\widehat{B}_1(t) = \sum_{i=1}^{n} \int_0^t J(u) \frac{Y_i(u)\,\{x_i - \overline{x}(u)\}}{S_{xx}(u)}\, dN_i(u), \tag{5}$$

where

$$\overline{x}(u) = \frac{1}{Y_\cdot(u)} \sum_{i=1}^{n} Y_i(u)\, x_i$$

with $Y_\cdot(u) = \sum_{i=1}^{n} Y_i(u)$, and

$$S_{xx}(u) = \sum_{i=1}^{n} Y_i(u)\,\{x_i - \overline{x}(u)\}^2\,. \tag{6}$$

You shall *not* prove (5).

We then consider testing of the null hypothesis

$$H_0 : \beta_1(t) = 0 \quad \text{for all } t \in [0, t_0] \tag{7}$$

for a suitably chosen t_0 (where usually we will choose t_0 as the upper time limit of the study). We will base a test on the statistic

$$Z_1(t_0) = \int_0^{t_0} L(u)\, d\widehat{B}_1(u), \tag{8}$$

where $L(u)$ is a nonnegative predictable "weight process" that is assumed to be zero whenever $J(u) = 0$. One possible choice of the weight process is $S_{xx}(u)$ given by (6).

b) Show that $Z_1(t_0)$ is a mean zero martingale (when considered as a process in t_0) when the null hypothesis holds true, and explain why it is reasonable to use $Z_1(t_0)$ as a test statistic.

c) Show that when the null hypothesis holds true we have

$$\langle Z_1 \rangle(t_0) = \int_0^{t_0} \frac{L(u)^2}{S_{xx}(u)}\, \beta_0(u) du, \tag{9}$$

where $S_{xx}(u)$ is given by (6).

Hint: Insert (5) in (8) and note that when the null hypothesis holds true the intensity processes take the form $\lambda_i(t) = Y_i(t)\beta_0(t)$. Also remember that if $M_1(t), \ldots, M_n(t)$ are martingales derived from counting processes $N_1(t), \ldots, N_n(t)$ with intensity processes $\lambda_1(t), \ldots, \lambda_n(t)$, and $H_1(t), \ldots, H_n(t)$ are predictable processes, then

$$\Big\langle \sum_{i=1}^{n} \int H_i\, dM_i \Big\rangle(t) = \sum_{i=1}^{n} \int_0^t H_i(u)^2 \lambda_i(u) du.$$

(Continued on page 6.)

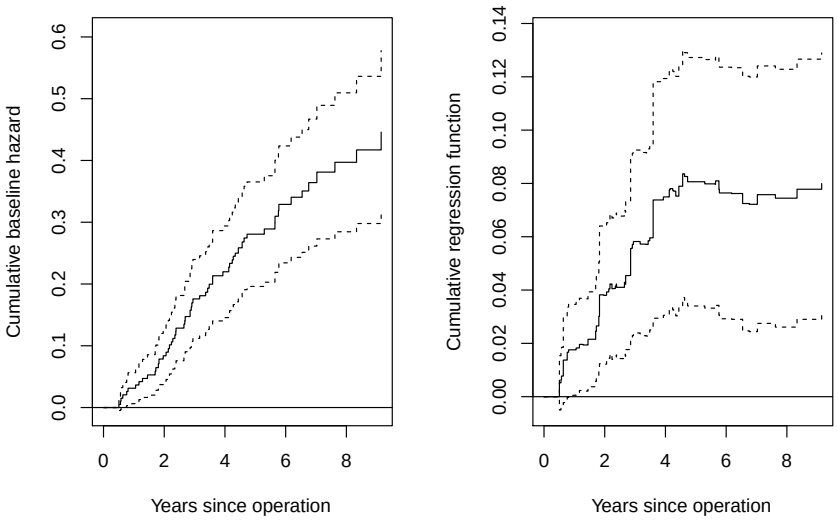
When the null hypothesis holds true, we may estimate $B_0(t)$ by the Nelson-Aalen estimator. By replacing $\beta_0(u)du$ in (9) by the increment of the Nelson-Aalen estimator, we obtain the following estimator for the variance of (8)

$$V_{11}(t_0) = \int_0^{t_0} \frac{L(u)^2}{S_{xx}(u)} \frac{dN_{\cdot}(u)}{Y_{\cdot}(u)}, \tag{10}$$

where $N_{\cdot}(u) = \sum_{i=1}^{n} N_i(u)$. One may show that when the null hypothesis holds true (10) is an unbiased estimator for the variance of (8) and the standardized test statistic $Z_1(t_0)/\sqrt{V_{11}(t_0)}$ is approximately standard normally distributed. (You shall *not* show these results.)

In the period 1962-77 a total of 205 patients with malignant melanoma (cancer of the skin) were operated at Odense University hospital in Denmark. A number of covariates were recorded at operation, and the patients were followed up until death from malignant melanoma or censoring by death from other causes or by closure of the study at December 31, 1977. Here we will only consider the covariate tumor thickness (in mm), which we centre by subtracting the mean tumor thickness 2.92 mm.

d) The additive model (3) has been fitted to melanoma data, and the figure below gives estimates of the cumulative baseline hazard (to the left) and the cumulative regression function for centred tumor thickness (to the right). Give an interpretation of the two estimates.



e) For the weight process (6), the statistic (8) and its variance estimator (10) take the values $Z_1(t_0) = 102.2$ and $V_1(t_0) = 364.1$ when t_0 is chosen as the upper time-point of the study. Use this information to test the null hypothesis (7).

END