UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

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

Day of examination: Tuesday 11 December 2012.

Examination hours: 14.30 – 18.30.

This problem set consists of 5 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

Assume that we have a counting process $N(t)$ with intensity processes of the multiplicative form $\lambda(t) = \alpha(t)Y(t)$. Here $\alpha(t)$ is a nonnegative function, while $Y(t)$ is a predictable processes that does not depend on unknown parameters. We say that the counting process satisfies the multiplicative intensity model.

a) Give two examples of situations that may be described by the multiplicative intensity model.

We will consider estimation of $A(t) = \int_0^t \alpha(u)du$. The Nelson-Aalen estimator for $A(t)$ is given by

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

where $J(u) = I\{Y(u) > 0\}$.

b) Give a motivation for the Nelson-Aalen estimator.

We introduce $A^*(t) = \int_0^t J(u)\alpha(u)du$.

c) Show that $\widehat{A}(t) - A^*(t)$ is a mean zero martingale.
Also show that the Nelson-Aalen estimator is approximately unbiased.

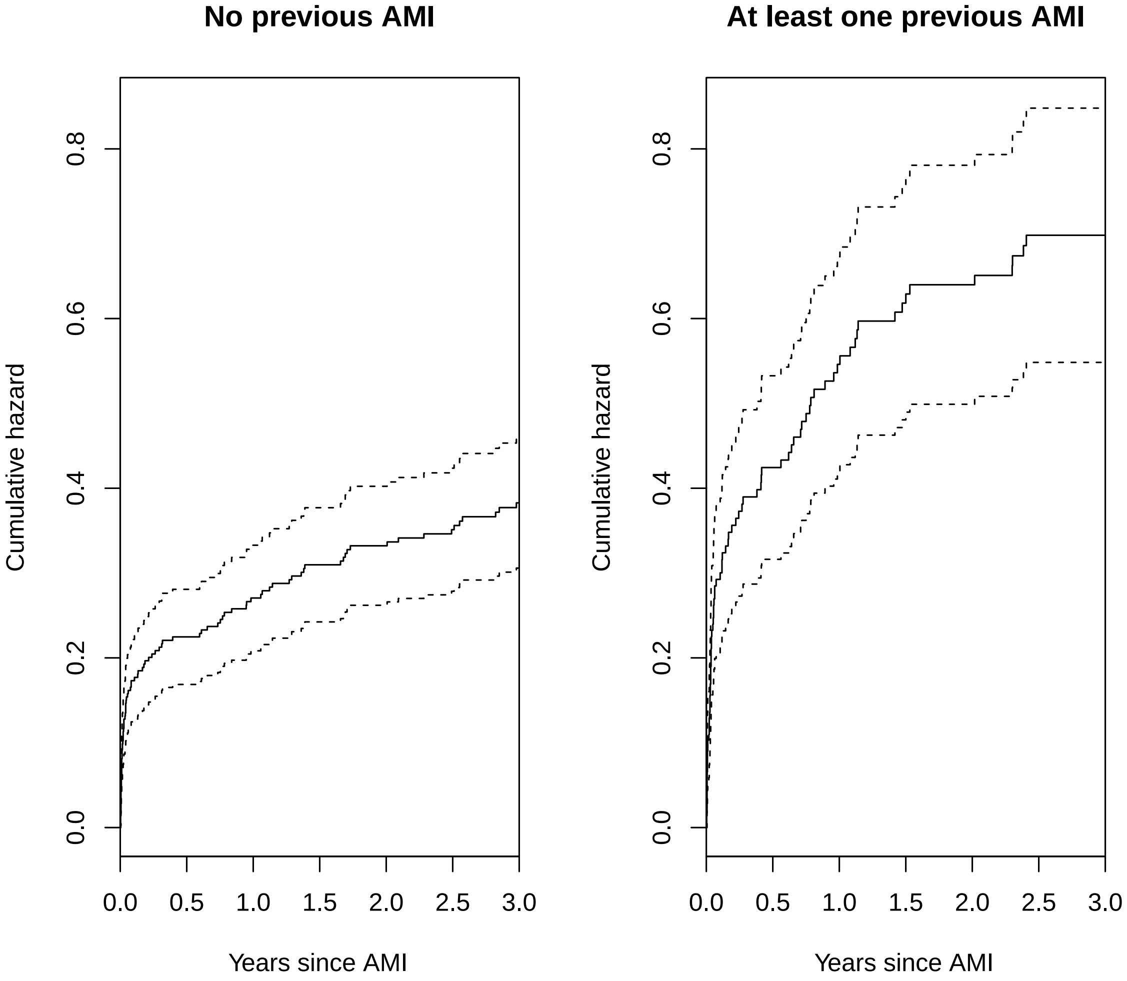
d) Derive an estimator for the variance of the Nelson-Aalen estimator.
[*Hint:* Consider the optional variation process of $\widehat{A}(t) - A^*(t)$.]

e) Explain why the Nelson-Aalen estimator $\widehat{A}(t)$ is approximately normally distributed and give a 95% confidence interval for $A(t)$.

(Continued on page 2.)

At a hospital one has studied the survival of patients admitted with an acute myocardial infarction (AMI). The figure below shows Nelson-Aalen estimates of the cumulative hazard with 95% confidence intervals for patients with no previous AMI (left) and at least one previous AMI (right).

f) Use the Nelson-Aalen plots to make rough sketches of the hazard rates for the two groups of patients. Discuss what the sketches (and the Nelson-Aalen plots) tell you about the mortality of the AMI patients.



Problem 2

In this problem we will consider the gamma frailty model with constant individual hazards. We assume that the frailty variable Z is gamma distributed with mean one and variance δ. Further given $Z = z$, we assume that the hazard rate of the survival time T is given as $\alpha(t \mid z) = z\,\alpha$, where $\alpha > 0$ is the individual hazard for an individual with unit frailty ($Z = 1$).

(Continued on page 3.)

Below you may use (without proof) that the Laplace transform $\mathcal{L}(c) = \mathrm{E}(e^{-cZ})$ of a gamma distributed random variable Z with mean one and variance δ takes the form $\mathcal{L}(c) = \{1 + \delta c\}^{-1/\delta}$.

a) Find an expression for the population survival function $S(t) = P(T > t)$.

b) Show that the population hazard $\mu(t) = -S'(t)/S(t)$ takes the form $\mu(t) = \alpha/\{1 + \delta\alpha t\}$.

c) Discuss how the result in question b) may help to give an explanation of the form of the hazard rates seen in question f) in Problem 1.

Problem 3

Assume that we have counting processes $N_1(t), N_2(t), \ldots, N_n(t)$ that record the occurrences of an event of interest for n individuals. In questions a)-c) we assume that the counting processes have intensity processes of the form $\lambda_i(t) = Y_i(t)\alpha(t, \boldsymbol{\theta})$; $i = 1, 2, \ldots, n$. Here $\alpha(t, \boldsymbol{\theta})$ is assumed to be piecewise constant, i.e. there exists a partition $0 = t_0 < t_1 < \cdots < t_K = \tau$ of the study time interval $[0, \tau]$ such that $\alpha(t, \boldsymbol{\theta}) = \sum_{k=1}^{K} \theta_k I_k(t)$, where $I_k(t) = 1$ for $t_{k-1} < t \leq t_k$, and $I_k(t) = 0$ otherwise.

For this situation one may show that the likelihood takes the form

$$L(\boldsymbol{\theta}) = \prod_{k=1}^{K} \left\{ \theta_k^{O_k} \; e^{-\theta_k R_k} \right\}.$$

Here $O_k = \int_0^\tau I_k(t) dN_{\boldsymbol{.}}(t)$ and $R_k = \int_0^\tau I_k(t) Y_{\boldsymbol{.}}(t) dt$, where $N_{\boldsymbol{.}}(t) = \sum_{i=1}^{n} N_i(t)$ and $Y_{\boldsymbol{.}}(t) = \sum_{i=1}^{n} Y_i(t)$. (You shall not prove this result.)

a) Show that the maximum likelihood estimator for θ_j becomes

$$\widehat{\theta}_j = \frac{O_j}{R_j} \quad \text{for } j = 1, \ldots, K.$$

The estimators in question a) are called occurrence/exposure rates.

b) Explain why the occurrence/exposure rates are approximately independent and normally distributed.

c) Derive an estimator for the variance of $\widehat{\theta}_j$.

We now assume that the n individuals can be divided into G groups. In the application on the following page, the groups correspond to the four cities. If individual i belongs to group g, the intensity process of $N_i(t)$ takes the form $\lambda_i(t) = Y_i(t)\alpha(t, \boldsymbol{\theta})e^{\beta_g}$, where $\alpha(t, \boldsymbol{\theta})$ is piecewise constant as described in the introduction to the problem, and $\beta_1 = 0$.

(Continued on page 4.)

For this situation the likelihood takes the form

$$L(\boldsymbol{\theta}) = \prod_{g=1}^{G}\prod_{k=1}^{K}\left\{\left(\theta_k\, e^{\beta_g}\right)^{O_{gk}}\; e^{-\theta_k\, e^{\beta_g} R_{gk}}\right\},$$

where $O_{gk} = \int_0^\tau I_k(t)dN^{(g)}(t)$ and $R_{gk} = \int_0^\tau I_k(t)Y^{(g)}(t)dt$. Here $N^{(g)}(t) = \sum N_i(t)$ and $Y^{(g)}(t) = \sum Y_i(t)$, where the sums are over all individuals i who belong to group g; $g = 1, \ldots, G$. (You shall not prove this result.)

d) Show that the likelihood is proportional to the likelihood one gets by considering the O_{gk} to be independent and Poisson distributed random variables with means $\theta_k\, e^{\beta_g} R_{gk}$.

e) Explain how we may use software for Poisson regression (like the `glm`-command in `R`) to find the maximum likelihood estimates of $\psi_k = \log\theta_k$ and β_g.

We will then consider data on lung cancer occurrence among Danish men. Table 1 gives the observed numbers of lung cancer cases in the male population in four Danish cities in the period 1968-1971 distributed over five age groups. Table 2 gives the corresponding numbers of person-years.

Table 1: Observed numbers of lung cancer cases in the male population between 1968 and 1971 in four Danish cities distributed over age groups.

Age group	Fredericia	Horsens	Kolding	Vejle	Total
40-54	11	13	4	5	33
55-59	11	6	8	7	32
60-64	11	15	7	10	43
65-69	10	10	11	14	45
70-74	11	12	9	8	40

Table 2: The numbers of person-years lived for male inhabitants in four Danish cities in the period 1968-1971, distributed over age groups.

Age group	Fredericia	Horsens	Kolding	Vejle	Total
40-54	12236	11516	12568	10080	46400
55-59	3200	4332	4200	3512	15244
60-64	2840	3692	3580	3356	13468
65-69	2324	3336	2808	2524	10992
70-74	2036	2536	2140	2156	8868

We will first consider the lung cancer rates for all cities counted together.

f) Use the totals in Tables 1 and 2 to compute the occurrence/exposure rates for lung cancer.

(Continued on page 5.)

We then use Poisson regression to find the maximum likelihood estimates of a model with piecewise constant baseline hazard over the age groups in Tables 1 and 2 and proportional effect of city. We then obtain the following (edited) output:

```
Call:
glm(formula=cancer~offset(log(personyears))+factor(agegroup)+factor(city)-1,
            family = poisson, data = lungcancer)

                          Estimate     Std. Error
factor(agegroup)40-54      -7.0318         0.2049
factor(agegroup)55-59      -5.9358         0.2126
factor(agegroup)60-64      -5.5180         0.1928
factor(agegroup)65-69      -5.2734         0.1902
factor(agegroup)70-74      -5.1832         0.1954
factor(city)Horsens        -0.1907         0.1910
factor(city)Kolding        -0.4791         0.2103
factor(city)Vejle          -0.2534         0.2033
```

f) Use the output to obtain estimates of the baseline hazard rates θ_k for the five age groups. Also obtain estimates and 95% confidence intervals for the hazard rate ratios e^{β_g} for Horsens, Kolding, and Vejle relative to Fredericicia (which is the reference city). Discuss what the results tell you on how the lung cancer risk varies with age and between the cities.

END