UNIVERSITETET I OSLO

Matematisk Institutt

Exam in:	**STK 4080/9080 – Survival Analysis and Event History Analysis**
With:	**Nils Lid Hjort**
Auxilia:	**Calculator, plus one single page of paper with the candidate's own personal notes**
Time for exam:	**Monday 3/xii/2018, 9^{00}–13^{00}**

This exam set contains four exercises and comprises four pages.

Exercise 1: Multiple dangers

Life is dangerous at both ends and uncomfortable in the middle – well, actually, 007 Bond author Ian Fleming said this about horses, but it might be fitting nevertheless, and in this exercise we look into the probability calculus of lives threatened by multiple dangers.

(a) Suppose $A(t)$ is some cumulative hazard function, continuous and strictly increasing, with $A(0) = 0$ and $A(t) \to \infty$ as $t \to \infty$. With V a unit exponential, i.e. with constant hazard rate 1, show that the random life-time $T = A^{-1}(V)$ has this $A(t)$ as its cumulative hazard function.

(b) If the cumulative hazard function has the form $A(t) = \log(1+t)$, give a recipe for simulating random life-times T from that distribution. If you generate say a million such life-times, what can you say about their average $\bar{T}$?

(c) Suppose a new-born object lives in a world where there are ten deadly dangers, and that the random time T_j until danger no. j kills an object has continuous and positive hazard rate $\alpha_j(s)$ (supposing that the object is not already dead for other reasons), for $j = 1, \ldots, 10$. Thus the object in question dies at a time T which can be represented as the smallest of the ten T_j. Assume further that these ten mechanisms act independently of each other. Find expressions for the survival curve $S^*(t)$ and the hazard rate $\alpha^*(s)$ for T.

(d) In the situation above, find an expression for

$$\Pr\{\text{object has died from danger } j \text{ in } [t, t+\varepsilon] \,|\, \text{object has died in } [t, t+\varepsilon]\},$$

where ε is a small number. Give also the limit of this conditional probability as ε tends to zero.

(e) Suppose T is a random life-time with cumulative distribution function of the form

$$F(t) = 1 - \exp(-1.1\, t^{3/2} - 1.2\, t - 1.3\, t^{1/2}) \quad \text{for } t \geq 0.$$

Give a recipe for simulating say 1000 random T from this distribution.

Exercise 2: Frail lives

WE ALL ARE MEN, IN OUR OWN NATURES FRAIL, and capable of our flesh; few are angels. The present exercise pursues this frail Shakespearean thought in two directions.

(a) For simplicity we work with Gamma distributed frailties below. If z has a Gamma distribution with certain positive parameters (a, b), its density is

$$g(z) = \frac{b^a}{\Gamma(a)} z^{a-1} \exp(-bz) \quad \text{for } z > 0,$$

with mean a/b and variance a/b^2 (but do not prove these formulae here). Show that its Laplace transform becomes

$$L(c) = \mathrm{E}^* \exp(-cz) = \frac{b^a}{(b+c)^a} = \exp\Big\{-a \log\Big(1 + \frac{c}{b}\Big)\Big\} \quad \text{for } c \geq 0.$$

For intended clarity I use 'E^*' to mean the expectation operator with respect to the frailty distribution.

(b) Suppose individual i in a certain big population has hazard rate $\alpha(s)z_i$, where z_i is an unobserved frailty factor; thus $\alpha(s)$ is the hazard rate for individuals with $z_i = 1$. With $A(t) = \int_0^t \alpha(s)\,\mathrm{d}s$, show that the survival curve for a randomly sampled individual in the population can be expressed as

$$S^*(t) = \mathrm{E}^* \exp\{-A(t)z\}.$$

Use this to find an expression for the hazard rate $\alpha^*(s)$ for the population, when the frailty factors are distributed according to a Gamma with parameters (a, b).

(c) Assume now that the unobserved heterogeneity is additive rather than multiplicative, i.e. that individual i has hazard rate $\alpha(s) + z_i$. Find an expression for the survival function $S^*(t)$ for a randomly sampled individual in the population. Again, for the case of the frailties z_i stemming from a Gamma distribution with parameters (a, b), find a formula for the hazard rate $\alpha^*(s)$.

(d) In continuation of question (c), specialise to the case where $\alpha(s)$ is a constant α, such that the individuals have constant but different hazard rates $\alpha + z_i$. Give a formula for the hazard rate $\alpha^*(s)$ in the population, and comment briefly on the role played by the frailty for long-term survivors.

(e) A generalisation of both setups is to include both multiplicative and additive frailty mechanisms. Taking again a constant baseline hazard $\alpha(s) = \alpha$, for simplicity, suppose that individual i has constant hazard rate

$$\alpha_i(s) = \alpha z_{i,1} + z_{i,2} \quad \text{for } s \geq 0,$$

where $z_{i,1}$ and $z_{i,2}$ are independent (and not directly observed) frailties, coming from Gamma distributions with parameters respectively (b, b) and (a_2, b_2). Find a formula for the consequent hazard rate $\alpha^*(s)$ in the population.

Exercise 3: Comparing groups

TAKING AN INTEREST IN HAZARD RATES is an occupational hazard for our profession. Suppose a certain time-to-event phenomenon is studied in Denmark and Sweden, with the hazard rates in question being denoted $\alpha_1(s)$ and $\alpha_2(s)$, along with cumulatives $A_j(t) = \int_0^t \alpha_j(s)\,\mathrm{d}s$ for $j = 1, 2$. Assume that relevant survival data have been collected for n_1 Danes and n_2 Swedes, of the usual form $(t_{1,i}, \delta_{1,i})$ for the Danes and $(t_{2,i}, \delta_{2,i})$ for the Swedes.

(a) Let N_1 and N_2 be the counting processes and Y_1 and Y_2 the at-risk processes for the Danish and Swedish groups. Give expressions for the Nelson–Aalen estimators $\widehat{A}_1(t)$ and $\widehat{A}_2(t)$.

(b) For simplicity of presentation and for the convenience of some the mathematics that follow, we now assume the same sample size in each group, so $n_1 = n_2 = n$. To compare the Danes with the Swedes, consider the random process

$$Z_n(t) = \int_0^t H_n(s)\{\mathrm{d}\widehat{A}_1(s) - \mathrm{d}\widehat{A}_2(s)\} \quad \text{for } t \geq 0,$$

with the weight function

$$H_n(s) = \sqrt{\frac{Y_1(s)}{n}\frac{Y_2(s)}{n}}.$$

Give an expression for $Z_n(t)$ as a finite sum. Show that if the null hypothesis H_0 that the two hazard rates are identical holds, then $Z_n(t)$ becomes a martingale.

(c) Again assuming that H_0 holds, find an expression for the martingale variance process $\langle Z_n, Z_n\rangle(t)$, and use this to provide an estimator of the variance of $Z_n(t)$.

(d) Explain how you can construct tests for H_0, based on having computed the $Z_n(t)$ process.

(e) Somewhat more formally than you have had to argue for question (d), one may show that under null hypothesis conditions, $\sqrt{n}Z_n(t)$ converges in distribution to a well-defined Gaußian zero-mean martingale process $W(t)$, with independent increments and a certain variance function $v(t) = \mathrm{Var}\,W(t)$. Find a clear expression for this variance function. Give also an estimator for $v(t)$.

Exercise 4: Presidential survival regression

REVOLUTIONS, LIKE SATURN, EAT THEIR OWN CHILDREN, or at least some of them, we learned in the process of our revolutionary Oblig (November 2018). There we worked with models for the time from end of presidency to death, for the $n = 73$ French presidents from September 1792 to November 1795. Thanks to Céline Cunen's efforts we have a survival regression dataset of $(t_i, \delta_i, \mathrm{gironde}_i, \mathrm{vip}_i)$, for $i = 1, \ldots, n$, with $\mathrm{gironde}_i$ an indicator for having belonged to the Gironde faction and vip_i a proxy for perceived importance, namely the number of languages in which there is a wikipedia article about president i.

Consider now the hazard rate regression model

$$\alpha_i(s) = \theta\gamma s^{\gamma-1} \exp(\beta_1 \text{gironde}_i + \beta_2 \text{vip}_i) \quad \text{for } i = 1, \ldots, n,$$

wth θ and γ positive parameters.

(a) Write down a clear expression for the ensuing log-likelihood function, and explain how maximum likelihood estimates may be computed.

(b) Also explain how the standard errors (i.e. the estimated standard deviations) for the four maximum likelihoods estimators may be estimated from the data.

(c) I have carried out this analysis (supplementing the Gompertz regression modelling and analysis work carried out for the Oblig), for the French presidents dataset, and find the following maximum likelihood estimates and standard errors:

```
 0.0020  0.0015  theta
 1.8258  0.2074  gamma
-0.2577  0.3875  beta1
-0.0041  0.0117  beta2
```

Use this information for finding out (i) if the Gironde or the proxy importance covariates exert influence on the time from end of presidency to death; (ii) if the individual hazard rate functions are constant over time, i.e. if $\gamma = 1$.

– quantum satis *&* the end –