UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Exam in:	STK4080 — Survival and event history analysis.
Day of exam:	Friday December 10th 2010.
Exam hours:	14.30 – 18.30.

This examination set consists of 4 pages.

Appendices:	None.
Permitted aids:	Approved calculator.

Make sure that your copy of this examination set is complete before answering.

FASIT

Problem 1.

(a) Since $\alpha(s) = \frac{f(s)}{S(s)}$ where the density $f(s) = -S'(s)$ we have

$$A(t) = \int_0^t \alpha(s)ds = \int_0^t \frac{f(s)}{S(s)}ds = -\int_0^t \frac{dS(s)}{S(s)} = -\int_1^{S(t)} \frac{dS}{S} = -\log(S(t)).$$

(b) For an exact observed lifetime $\tilde{T}_i$ with $D_i = 1$ we have likelihood contribution $f(\tilde{T}_i) = \alpha(\tilde{T}_i)\exp(-A(\tilde{T}_i))$. For a right censored survival time $\tilde{T}_i$ with $D_i = 0$ we get a contribution $S(\tilde{T}_i) = \exp(-A(\tilde{T}_i))$. Put together we get

$$L = \prod_{i=1}^n (\alpha(\tilde{T}_i)\exp(-A(\tilde{T}_i))^{D_i} \exp(-A(\tilde{T}_i))^{1-D_i} = \prod_{i=1}^n \alpha(\tilde{T}_i)^{D_i} \exp(-A(\tilde{T}_i)).$$

(Continued on page 2.)

(c) Since $M(t)$ is a martingale with expectation zero and $H(s) = \frac{\hat{S}(s-)}{S^\star(s)}\frac{J(s)}{Y(s)}$ is a predictable function we get that $\frac{\hat{S}(t)}{S^\star(t)} - 1 = \int_0^t H(s)dM(s)$ is also a martingale with expectation zero. Thus $\mathrm{E}[\frac{\hat{S}(t)}{S^\star(t)}] = 1$ and since $S^\star(t) = S(t)$ if $J(t) = 1$ we have a property closely related to unbiasedness.

Furthermore the predictable variation process of $\frac{\hat{S}(t)}{S^\star(t)} - 1$ becomes

$$\langle \int_0^t H(s)dM(s)\rangle = \int_0^t H(s)^2 d\langle M\rangle(s) = \int_0^t \left(\frac{\hat{S}(s-)}{S^\star(s)Y(s)}\right)^2 Y(s)\alpha(s)ds$$

Since $(\int_0^t H(s)dM(t))^2 - \langle \int_0^t H(s)dM(s)\rangle$ is a martingale with expectation zero we have

$$\mathrm{Var}\left[\frac{\hat{S}(t)}{S^\star(t)}\right] = \mathrm{E}[\langle \int_0^t H(s)dM(s)\rangle] = \mathrm{E}\left[\int_0^t \left(\frac{\hat{S}(s-)}{S^\star(s)}\right)^2 \frac{J(s)\alpha(s)ds}{Y(s)}\right]$$

An estimator of the variance of $\hat{S}(t) \approx S(t)\frac{\hat{S}(t)}{S^\star(t)}$ is given by Greenwoods formula

$$\hat{S}(t)^2 \int_0^t \frac{dN(s)}{Y(s)(Y(s)-1)}$$

where we "estimate" $\alpha(s)ds$ by $dN(s)/Y(s)$ (Other variance estimators are possible).

(d) Read off vertical lines at $1-p$. The estimate of the p 100%-percentile is the value along the x-axis where the vertical line crosses the Kaplan-Meier estimate. The 95% CI is from the value where the vertical line crosses the lower confidence limit to where it crosses the upper confidence limit of the survival function.

Problem 2.

(a) We have

$$\frac{\exp(\beta x_i)}{\sum_{k\in\mathcal{R}(t_i)}\exp(\beta x_k)} = \frac{\exp(\beta x_i)\alpha_0(t_i)}{\sum_{k\in\mathcal{R}(t_i)}\exp(\beta x_k)\alpha_0(t_i)} = \frac{\alpha_i(t_i)}{\sum_{k\in\mathcal{R}(t_i)}\alpha_k(t_i)}$$

and so the expression has the interpretation as the probability that individual i experienced the event given that there was an event among the $\mathcal{R}(t_i)$ individuals at risk.

A product of such conditional probabilities is a sensible objective function for estimating parameters of the model. Since the baseline hazard $\alpha_0(t)$ cancels out in the expression we may be estimate β without specifying the baseline.

(Continued on page 3.)

(b) We have

$$\sum_{i=1}^{n}[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]Y_i(t)\alpha_i(t)=\{\sum_{i=1}^{n}[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]Y_i(t)\exp(\beta x_i)\}\alpha_0(t)=0$$

since

$$\sum_{i=1}^{n}[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]Y_i(t)\exp(\beta x_i)=S^{(1)}(\beta,t)-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}S^{(0)}(\beta,t)=0.$$

Thus

$$\begin{aligned}\textstyle\sum_{i=1}^n\int[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]dM_i(t) &= \textstyle\sum_{i=1}^n\int[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}][dN_i(t)-Y_i(t)\alpha_i(t)dt]\\ =\textstyle\sum_{i=1}^n\int[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]dN_i(t) &-\textstyle\sum_{i=1}^n\int[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]Y_i(t)\alpha_i(t)dt\\ =\textstyle\sum_{i=1}^n\int[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]dN_i(t)-0 &=U(\beta)\end{aligned}$$

(c) Since the $M_i(t)$ are orthogonal martingales with expectation zero and $U(\beta)$ a sum of integrals of predictable functions with respect to these martingales it follows that $U(\beta)$ has expectation zero.

The variance of the score follows from

$$\mathrm{Var}(U(\beta))=\mathrm{E}[\langle U(\beta)\rangle]=\mathrm{E}\{\sum_{i=1}^{n}\int[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]^2d\langle M_i\rangle(t)\}$$

which in turn equals, since $d\langle M_i\rangle(t)=Y_i(t)\alpha_i(t)dt=Y_i(t)\exp(\beta x_i)\alpha_0(t)dt$,

$$\mathrm{E}\{\sum_{i=1}^{n}\int[x_i-\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]^2Y_i(t)\exp(\beta x_i)\alpha_0(t)dt\}=\cdots=\mathrm{E}\{\int[\frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)}-(\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)})^2]S^{(0)}(\beta,t)\alpha_0(t)dt\}$$

where $S^{(2)}(\beta,t)=\sum_{i=1}^n x_i^2Y_i(t)\exp(\beta x_i)$ (If you get as far as the left-hand side of the above formula we consider your answer complete).

(d) With covariates x_i that can only take two values, 0 and 1 we get $S^{(1)}(0,s)=\sum_{i=1}^n Y_i(s)x_i\exp(0x_i)=\sum_{i=1}^n Y_i(s)x_i=Y_{\bullet1}(s)=$ the number at risk with $x_i=1$ and and $S^{(0)}(0,s)=\sum_{i=1}^n Y_i(s)\exp(0x_i)=\sum_{i=1}^n Y_i(s)=Y_{\bullet1}(s)+Y_{\bullet1}(s)=$ the total number at risk. But then

$$\begin{aligned}U(0)&=\textstyle\sum_{i=1}^n\int[x_i-\frac{S^{(1)}(0,t)}{S^{(0)}(0,t)}]dN_i(t)\\ &=\textstyle\int\sum_{i=1}^n x_idN_i(t)-\int\sum_{i=1}^n\frac{Y_{\bullet1}(t)}{Y_{\bullet0}(t)+Y_{\bullet1}(t)}dN_i(t)\\ &=\textstyle N_{\bullet1}(\tau)-\int Y_{\bullet1}(s)\frac{dN_{\bullet1}(s)+dN_{\bullet0}(s)}{Y_{\bullet1}(s)+Y_{\bullet0}(s)}\end{aligned}$$

which is the log-rank statistic.

(Continued on page 4.)

Problem 3.

(a)

Covariate	$\hat{\beta}_j$	se_j	$\hat{HR}_j = \exp(\widehat{\beta}_j)$	95% CI $= \widehat{HR}_j \exp(\pm 1.96 se_j)$
HIV-positive (x_{i1})	0.79	0.22	2.20	[1.43 , 3.39]
Women (x_{i2})	0.02	0.22	1.02	[0.66 , 1.57]
Age > 29 (x_{i3})	0.74	0.23	2.10	[1.34 , 3.29]

We see that being HIV-infected and being 30+ years roughly doubles the mortality rate, whereas men and women appear to have roughly the same mortality.

The confidence intervals for HIV-infection and age does not include the value 1 (no difference) and so are significant at the 5% level, whereas the interval for sex includes 1 and the difference is not significant.

(b) Other regression methods

- Proportional hazards models with general risk function $\psi(\beta, x) \neq \exp(\beta' x)$
- Additive hazards models (Aalen-regression)
- Accelerated failure time models
- Poisson-regression models - assuming that the baseline is piecewise constant

END