STK 4080/9080: Survival Analysis and Event History Analysis

Notes to the Exam, December 3rd, 2018

Nils Lid Hjort

Exercise 1: Multiple dangers

(a) That $T = A^{-1}(V) \le t$ is the same as $V \le A(t)$, so the cumulative distribution of such a T becomes $1 - \exp\{-A(t)\}$.

(b) Setting $A(t) = \log(1+t)$ equal to v leads to $t = \exp(v)-1$, so the variable $T = \exp(V) - 1$ has the distribution in question. I can simulate such lifetimes via $T_j = \exp(V_j) - 1$, with the V_j from the unit exponential. Its mean is

$$\int_0^\infty \{\exp(v) - 1\} \exp(-v)\,\mathrm{d}v,$$

which is infinite. This is also clear from the cdf of T directly, which is $t/(t+1)$, with density $1/(t+1)^2$, with a too fat tail for finiteness of its mean.

(c) Consider $T = \min(T_1, \ldots, T_k)$. Its survival distribution is

$$\Pr\{T \ge t\} = \Pr\{T_1 \ge t, \ldots, T_k \ge t\} = \exp\{-A_1(t) - \cdots - A_k(t)\},$$

with A_j the cumulative of α_j. Hence $A^*(t) = A_1(t) + \cdots + A_k(t)$ and $\alpha^*(s) = \alpha_1(s) + \cdots + \alpha_k(s)$.

(d) The probability in question is the ratio between

$$\Pr\{T_j \in [t, t+\varepsilon], \text{the others bigger than } t\} = \alpha_j(t)\varepsilon \exp\{-A_j(t)\} \prod_{\ell \ne j} \exp\{-A_\ell(t)\}$$

and

$$\Pr\{T \in [t, t+\varepsilon]\} = \sum_{\ell=1}^k \alpha_\ell(t)\varepsilon \exp\Big\{-\sum_{\ell=1}^k A_\ell(t)\Big\}.$$

The ratio becomes

$$q_j(t) = \frac{\alpha_j(t)}{\alpha_1(t) + \cdots + \alpha_k(t)}.$$

(e) The cumulative hazard rate takes the form $A_1(t) + A_2(t) + A_3(t)$, which means T can be represented as $\min(T_1, T_2, T_3)$. Here $1.1\,t_1^{3/2} = v_1$ leads to $T_1 = (V_1/1.1)^{2/3}$; $1.2\,T_2 = V_2$ leads to $T_2 = V_2/1.2$; and $1.3\,T_3^{1/2} = V_3$ leads to $T_3 = (V_3/1.3)^2$, with V_1, V_2, V_3 independent unit exponentials.

Exercise 2: Frail lives

(a) We have

$$L(c) = \int_0^\infty \exp(-cz) g(z)\,\mathrm{d}z = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a)}{(b+c)^a} = \exp\{-a \log(1 + c/b)\}.$$

(b) The survival curve for a randomly sampled individual becomes

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

With the Gamma (a, b), the survival curve becomes

$$S^*(t) = \exp[-a \log\{1 + A(t)/b)\}],$$

which means cumulative hazard

$$A^*(t) = a \log\{1 + A(t)/b\},$$

which means hazard rate

$$\alpha^*(s) = a \frac{1}{1 + A(t)/b} \frac{\alpha(t)}{b} = \frac{a}{b} \frac{\alpha(t)}{1 + A(t)/b}.$$

(c) An individual with frailty z has cumulative hazard $A(t) + zt$. Hence

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

For the Gamma (a, b) frailty distribution, this becomes

$$S^*(t) = \exp\{-A(t) - a \log(1 + t/b)\},$$

with cumulative hazard rate

$$A^*(t) = A(t) + a \log(1 + t/b),$$

and hazard rate

$$\alpha^*(s) = \alpha(s) + \frac{a}{b} \frac{1}{1 + s/b}.$$

(d) For a constant $\alpha(s) = \alpha$, the above gives

$$\alpha^*(s) = \alpha + \frac{a}{b} \frac{1}{1 + s/b},$$

and for high s the second terms vanishes. Thus long-time survivors end up having the same hazard as those with zero frailty.

(e) The cumulative hazard rate for an individual with frailties z_1 and z_2 is $\alpha z_1 t + z_2 t$, and

$$S^*(t) = \mathrm{E}^*\, S(t \,|\, z) = \mathrm{E}^*\, \exp(-\alpha z_1 t - z_2 t) = L_1(\alpha t) L_2(t),$$

in terms of the two Laplace transforms at work. With the two gammas, this yields

$$S^*(t) = \exp\{-b \log(1 + \alpha t/b)\} \exp\{-a_2 \log(1 + t/b_2)\},$$

with cumulative hazard rate

$$A^*(t) = b \log(1 + \alpha t/b) + a_2 \log(1 + t/b_2),$$

and hazard rate

$$\alpha^*(s) = \frac{\alpha}{1 + \alpha s/b} + \frac{a_2}{b_2}\frac{1}{1 + s/b_2}.$$

Exercise 3: Comparing groups

(a) We have

$$\widehat{A}_1(t) = \int_0^t \frac{\mathrm{d}N_1(s)}{Y_1(s)} \quad \text{and} \quad \widehat{A}_2(t) = \int_0^t \frac{\mathrm{d}N_2(s)}{Y_2(s)}.$$

(b) With the usual martingales,

$$\mathrm{d}M_1(s) = \mathrm{d}N_1(s) - Y_1(s)\alpha_1(s)\,\mathrm{d}s \quad \text{and} \quad \mathrm{d}M_2(s) = \mathrm{d}N_2(s) - Y_2(s)\alpha_2(s)\,\mathrm{d}s,$$

which leads to

$$\mathrm{d}\widehat{A}_1(s) = \frac{\mathrm{d}M_1(s)}{Y_1(s)} + J_1(s)\,\mathrm{d}A_1(s) \quad \text{and} \quad \mathrm{d}\widehat{A}_2(s) = \frac{\mathrm{d}M_2(s)}{Y_2(s)} + J_2(s)\,\mathrm{d}A_2(s),$$

with the usual $J_1(s) = I\{Y_1(s) \ge 1\}$ and $J_2(s) = I\{Y_2(s) \ge 1\}$. Hence, under the null hypothesis where A_1 and A_2 are identical to a common A,

$$H_n(s)\{\mathrm{d}\widehat{A}_1(s) - \mathrm{d}\widehat{A}_2(s)\} = H_n(s)\Big\{\frac{\mathrm{d}M_1(s)}{Y_1(s)} - \frac{\mathrm{d}M_2(s)}{Y_2(s)}\Big\},$$

since $H_n(s) = \{Y_1(s)Y_2(s)\}^{1/2}/n$ is nonzero precisely when $J_1(s)$ and $J_2(s)$ are equal to 1. Hence Z_n is a difference of two independent martingales, and therefore a martingale.

(c) Under the null,

$$\begin{aligned}
\langle Z_n, Z_n\rangle(t) &= \int_0^t H_n(s)^2\Big\{\frac{1}{Y_1(s)^2}\,\mathrm{d}\langle M_1, M_1\rangle(s) + \frac{1}{Y_2(s)^2}\,\mathrm{d}\langle M_2, M_2\rangle(s)\Big\} \\
&= \int_0^t H_n(s)^2\Big\{\frac{\alpha(s)\,\mathrm{d}s}{Y_1(s)} + \frac{\alpha(s)\,\mathrm{d}s}{Y_2(s)}\Big\} \\
&= \frac{1}{n^2}\int_0^t J_1 J_2 Y_1 Y_2\Big(\frac{1}{Y_1} + \frac{1}{Y_2}\Big)\alpha(s)\,\mathrm{d}s. \\
&= \frac{1}{n^2}\int_0^t (Y_1 + Y_2)\alpha(s)\,\mathrm{d}s.
\end{aligned}$$

An estimator of the variance of $Z_n(t)$ is

$$\widehat{\sigma}(t)^2 = \frac{1}{n^2}\int_0^t \{Y_1(s) + Y_2(s)\}\,\mathrm{d}\widehat{A}(s),$$

with $\widehat{A}(t)$ an estimator of the common $A(t)$. When using the natural

$$\widehat{A}(t) = \int_0^t \frac{\mathrm{d}N_1 + \mathrm{d}N_2}{Y_1 + Y_2},$$

the Nelson–Aalen estimator based on the combined sample, we have

$$\widehat{\sigma}(t)^2 = \frac{N_1(t) + N_2(t)}{n^2}.$$

Other versions are available.

(d) We may plot $Z_n(t)/\widehat{\sigma}(t)$, which is approximately a standard normal, under the null, for each t. We may also form tests based on $\max_{c\le t\le d} |Z_n(t)|/\widehat{\sigma}(t)$, etc. Its limit distribution is the absolute maximum of a normalised Brownian motion over an interval.

(e) Under the null, it follows from central limit theory for martingales that $\sqrt{n}Z_n(t) \to_d W(t)$, a Gaußian martingale, with variance function $v(t)$, the limit in probability of $n\langle Z_n, Z_n\rangle(t)$, namely

$$v(t) = \int_0^t \{y_1(s) + y_2(s)\}\alpha(s)\,\mathrm{d}s,$$

with y_1 and y_2 the limit functions of Y_1/n and Y_2/n. It can be estimated using $n\widehat{\sigma}(t)^2$, which becomes as simple as $\{N_1(t) + N_2(t)\}/n$.

Exercise 4: Presidential survival regression

(a) The log-likelihood function becomes

$$\begin{aligned}\ell_n(\theta,\gamma,\beta_1,\beta_2) &= \sum_{i=1}^n \int_0^\tau \{\log\alpha_i(s)\,\mathrm{d}N_i(s) - Y_i(s)\alpha_i(s)\,\mathrm{d}s\} \\ &= \sum_{i=1}^n \{\log\theta + \log\gamma + (\gamma-1)\log t_i + \mathrm{gironde}_i\beta_1 + \mathrm{vip}_i\beta_2\}\delta_i \\ &\qquad - \sum_{i=1}^n \theta t_i^\gamma \exp(\beta_1\mathrm{gironde}_i + \beta_2\mathrm{vip}_i).\end{aligned}$$

I'd throw this to an optimiser.

(b) Standard errors are estimated as the square roots of the diagonal elements of the inverse of the 4×4 Fisher information matrix,

$$\widehat{J}_{\mathrm{obs}} = -\frac{\partial^2 \ell_n(\widehat{\eta})}{\partial\eta\partial\eta^{\mathrm{t}}},$$

writing η for the four parameters.

(c) The Wald ratios $\widehat{\beta}_j/\widehat{\kappa}_j$ are too small in size to signal that any of the two coefficients are different from zero. But

$$(\widehat{\gamma} - 1)/\widehat{\kappa} = (1.8258 - 1)/0.2074 = 3.9816,$$

which convincingly shows that γ is bigger than 1. Det skulle bare mangle.

– quantum satis *&* the end –