Exercise 4.1: Cox Score function and Information

Assume that $N_i(t)$ are counting processes with intensity processes $\lambda_i(t) = Y_i(t)\alpha_0(t)\exp(\beta' x_i)$. The log-partial likelihood can then be written

$$\log(L(\beta)) = \int_0^\tau \sum_{i=1}^n \left[\beta' x_i - \log(S^{(0)}(\beta,t))\right] dN_i(t)$$

where

$$S^{(0)}(\beta,t) = \sum_{i=1}^n Y_i(t)\exp(\beta' x_i)$$

Let also

- $S^{(1)}(\beta,t) = \sum_{k=1}^n x_k Y_k(t)\exp(\beta' x_k) = \frac{\partial S^{(0)}(\beta,t)}{\partial \beta}$
- $S^{(2)}(\beta,t) = \sum_{k=1}^n x_k x_k^\top Y_k(t)\exp(\beta' x_k) = \frac{\partial^2 S^{(0)}(\beta,t)}{\partial \beta^2}$

4.1a: Cox score function

$$U(\beta) = \frac{\partial \log(L(\beta))}{\partial \beta} = \sum_{i=1}^{n} \int_0^{\tau} \left[x_i - \frac{\partial \log(S^{(0)}(\beta,t))}{\partial \beta} \right] dN_i(t)$$

$$= \sum_{i=1}^{n} \int_0^{\tau} \left[x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)} \right] dN_i(t)$$

since

$$\frac{\partial \log(S^{(0)}(\beta, t))}{\partial \beta} = \frac{1}{S^{(0)}(\beta, t))} \frac{\partial S^{(0)}(\beta, t))}{\partial \beta} = \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t))}$$

4.1b: Cox information matrix

$$I(\beta) = -\frac{\partial U(\beta)}{\partial \beta^\top} = -\sum_{i=1}^n \int_0^\tau \left[-\frac{\partial(S^{(1)}(\beta,t)/S^{(0)}(\beta,t))}{\partial \beta^\top} \right] dN_i(t)$$

$$= \sum_{i=1}^n \int_0^\tau \left[\frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)} - \frac{S^{(1)}(\beta,t)S^{(1)}(\beta,t)^\top}{S^{(0)}(\beta,t)^2} \right] dN_i(t)$$

$$= \int_0^\tau \left[\frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)} - \frac{S^{(1)}(\beta,t)S^{(1)}(\beta,t)^\top}{S^{(0)}(\beta,t)^2} \right] dN_\bullet(t)$$

since $\frac{\partial(S^{(1)}(\beta,t)/S^{(0)}(\beta,t))}{\partial \beta^\top} =$

$$\frac{1}{S^{(0)}(\beta,t)} \frac{\partial S^{(1)}(\beta,t)}{\partial \beta^\top} - S^{(1)}(\beta,t) \frac{\partial(1/S^{(0)}(\beta,t))}{\partial \beta^\top}$$

$$= \frac{S^{(2)}(\beta,t)}{S^{(0)}(\beta,t)} - \frac{S^{(1)}(\beta,t)S^{(1)}(\beta,t)^\top}{S^{(0)}(\beta,t)^2}$$

Exercise 4.2: Cox Score test and Log-rank test

Assume that x_i is a binary variable. Then,

$$S^{(1)}(0,t) = \sum_{i=1}^{n} x_i Y_i(t) \exp(0x_i) = Y_{\bullet 1}(t),$$

i.e. the no. at risk with $x_i = 1$. Correspondingly

$$S^{(0)}(0,t) = \sum_{i=1}^{n} Y_i(t) \exp(0x_i) = Y_{\bullet 0}(t) + Y_{\bullet 1}(t) = Y_{\bullet}(t),$$

i.e. the total no. at risk. Thus

$$U(0) = \sum_{i=1}^{n} \int [x_i - \frac{S^{(1)}(0,t)}{S^{(0)}(0,t)}] dN_i(t) = \mathrm{O}_1 - \mathrm{E}_1$$

where O_1 is no. deaths with $x_i = 1$ and $\mathrm{E}_1 = \int Y_{\bullet 1}(t) \frac{dN_{\bullet}(t)}{Y_{\bullet}(t)}$, thus $U(0)$ is equal to the log-rank statistic!

Comparison Cox-regression and Log-rank

```
> summary(coxph(Surv(lifetime,dead)~ulcer,data=mel))

       coef exp(coef) se(coef)     z       p
ulcer -1.47      0.23    0.295 -4.98 6.3e-07

      exp(coef) exp(-coef) lower .95 upper .95
ulcer      0.23       4.36     0.129      0.41

Rsquare= 0.13   (max possible= 0.937 )
Likelihood ratio test= 28.4  on 1 df,   p=9.68e-08
Wald test            = 24.8  on 1 df,   p=6.3e-07
Score (logrank) test = 29.6  on 1 df,   p=5.41e-08

> survdiff(Surv(lifetime,dead)~ulcer,data=mel)

          N Observed Expected (O-E)^2/E (O-E)^2/V
ulcer=1  90       41     21.2      18.5      29.6
ulcer=2 115       16     35.8      10.9      29.6

 Chisq= 29.6  on 1 degrees of freedom, p= 5.41e-08
```

Variance of $U(0) = \mathbf{O}_1 - \mathbf{E}_1$

is estimated by

$$I(\beta) = \int \{ \frac{S^{(2)}(0,t)}{S^{(0)}(0,t)} - [\frac{S^{(1)}(0,t)}{S^{(0)}(0,t)}]^2 \} dN_\bullet(t)$$

But since x_i is binary we get

$$S^{(2)}(0,t) = \sum_{i=1}^{n} x_i^2 Y_i(t) = \sum_{i=1}^{n} x_i Y_i(t) = S^{(1)}(0,t) = Y_{\bullet 1}(t)$$

Thus

$$I(0) = \int \{ \frac{Y_{\bullet 1}(t)}{Y_\bullet(t)} - [\frac{Y_{\bullet 1}(t)}{Y_\bullet(t)}]^2 \} dN_\bullet(t) = \int \frac{Y_{\bullet 0}(t) Y_{\bullet 1}(t)}{Y_\bullet(t)^2} dN_\bullet(t)$$

which equals the previously derived variance formula of $\mathrm{O}_1 - \mathrm{E}_1$ (and ignoring ties).

Generalized log-rank test

For general x_i (not necessarily binary) we have

$$U(0) = \sum_{i=1}^{n} \int [x_i - \frac{\sum_{k=1}^{n} x_k Y_k(t)}{Y(t)}] dN_i(t)$$

which has variance under $\mathrm{H}_0 : \beta = 0$ given by

$$I(0) = \int \left\{ \frac{\sum_{i=1}^{n} x_i^2 Y_i(t)}{Y(t)} - \left[\frac{\sum_{i=1}^{n} x_i^2 Y_i(t)}{Y(t)} \right] \right\} dN_{\bullet}(t)$$

The generalized log-rank test is the given by that under H_0:

$$\frac{U(0)^2}{I(0)} \sim \chi_1^2$$