

Lecture 17

Cox proportional hazards models

BIOST 515

March 4, 2004

Review

The proportional hazards regression model is given by

$$h(t|X) = h(t) \exp(X_1\beta_1 + \cdots + X_p\beta_p).$$

- The predictors, X_1, \dots, X_p are assumed to act additively on $\log h(t|x)$.
- $\log h(t|x)$ changes linearly with the β s.
- The effect of the predictors is the same at all times t .
- In parametric survival models, we make an assumption on the shape of the underlying hazard, $h(t)$, and therefore we are making assumptions about S , H and f .

- The hazard ratio for a subject with a set of predictors X^* compared to a subject with a set of predictors X is

$$hr(X^* : X) = \frac{\exp(X^* \beta)}{\exp(X \beta)} = \exp\{(X^* - X)\beta\}.$$

- The point estimate for the hazard ratio is

$$\hat{hr}(X^* : X) = \frac{\exp(X^* \hat{\beta})}{\exp(X \hat{\beta})} = \exp\{(X^* - X)\hat{\beta}\},$$

where $\hat{\beta}$ is the maximum likelihood estimate of β .

- We can construct $(1 - \alpha)100\%$ confidence intervals for the hazard ratio as

$$\exp\{(X^* - X)\hat{\beta} \pm Z_{1-\alpha/2} \hat{se}((X^* - X)\hat{\beta})\}.$$

Cox proportional hazards regression model

The Cox PH model

- is a semiparametric model
- makes no assumptions about the form of $h(t)$ (non-parametric part of model)
- assumes parametric form for the effect of the predictors on the hazard

In most situations, we are more interested in the parameter estimates than the shape of the hazard. The Cox PH model is well-suited to this goal.

Brief overview of estimation of β

Parameter estimates in the Cox PH model are obtained by maximizing the partial likelihood as opposed to the likelihood. The partial likelihood is given by

$$L(\beta) = \prod_{Y_i \text{ uncensored}} \frac{\exp(X_i\beta)}{\sum_{Y_j \geq Y_i} \exp(X_j\beta)}$$

The log partial likelihood is given by

$$l(\beta) = \log L(\beta) = \sum_{Y_i \text{ uncensored}} \{X_i\beta - \log[\sum_{Y_j \geq Y_i} \exp(X_j\beta)]\}$$

Partial likelihood

- Cox and others have shown that this partial log-likelihood can be treated as an ordinary log-likelihood to derive valid (partial) MLEs of β .
- Therefore we can estimate hazard ratios and confidence intervals using maximum likelihood techniques discussed previously. The only difference is that these estimates are based on the partial as opposed to the full likelihood.
- The partial likelihood is valid when there are no ties in the data set. That is no two subjects have the same event time.
- If there are ties in the data set, the true partial log-likelihood function involves permutations and can be time-consuming to compute. In this case, either the Breslow or Efron approximations to the partial log-likelihood can be used.

Model assumptions and interpretations of parameters

- Same model assumptions as parametric model - except no assumption on the shape of the underlying hazard.
- Parameter estimates are interpreted the same way as in parametric models, except no shape parameter is estimated because we are not making assumptions about the shape of the hazard.

Example

$$h(t|rx, age) = h(t) \exp(\beta_1 \times rx + \beta_2 \times age)$$

```
> cph1=coxph(Surv(futime, fustat)~rx+age , ovarian)
> summary(cph1)
```

	coef	exp(coef)	se(coef)	z	p
rx	-0.804	0.448	0.6320	-1.27	0.2000
age	0.147	1.159	0.0461	3.19	0.0014

	exp(coef)	exp(-coef)	lower .95	upper .95
rx	0.448	2.234	0.130	1.54
age	1.159	0.863	1.059	1.27

```
Rsquare= 0.457 (max possible= 0.932 )
Likelihood ratio test= 15.9 on 2 df, p=0.000355
Wald test = 13.5 on 2 df, p=0.00119
Score (logrank) test = 18.6 on 2 df, p=9.34e-05
```


Interpreting the output from R

This is actually quite easy. The $coxph()$ function gives you the hazard ratio for a one unit change in the predictor as well as the 95% confidence interval. Also given is the Wald statistic for each parameter as well as overall likelihood ratio, wald and score tests.

What if we wanted to estimate $hr(rx = 1, age = 50 : rx = 2, age = 60)$?

The point estimate is easily obtained as

$$\exp\{-0.804(1 - 2) + 0.147(50 - 60)\} = 0.514.$$

How do we interpret this quantity?

How do we obtain the confidence interval?

- We can get the variance-covariance matrix for the parameter estimates from the fitted object.

```
> cph1$var
              [,1]      [,2]
[1,] 0.399486408 0.004696922
[2,] 0.004696922 0.002129550
> sqrt(diag(cph1$var))
[1] 0.63204937 0.04614705
```

- Estimate the variance of $\hat{\beta}_1(1 - 2) + \hat{\beta}_2(50 - 60)$.

$$\begin{aligned}\widehat{\text{var}}(-\hat{\beta}_1 - 10\hat{\beta}_2) &= \widehat{\text{var}}(\hat{\beta}_1) + 100\widehat{\text{var}}(\hat{\beta}_2) + \\ &\quad 2(-1)(-10)\widehat{\text{cov}}(\hat{\beta}_1, \hat{\beta}_2) \\ &= 0.399 + 100 \times 0.00213 + 20 \times 0.00470 \\ &= 0.706\end{aligned}$$

- The 95% CI for $hr(rx = 1, age = 50 : rx = 2, age = 60)$ is

$$0.514 \times \exp(\pm 1.96 \times \sqrt{0.706}) = (0.099, 2.67)$$

What is the appropriate interpretation for this confidence interval?

Partial likelihood ratio tests

What if we want to do a likelihood ratio test for $H_0 : \beta_2 = 0$ in the previous model?

Reduced model:

```
>cph1r=coxph(Surv(futime, fustat)~rx, ovarian)
> cph1$loglik
[1] -34.98494 -27.04190
> cph1r$loglik
[1] -34.98494 -34.45921
```

$$-2 \log L (\text{full model}) = -2 \times -27.04190$$

$$-2 \log L (\text{reduced model}) = -2 \times -34.45921$$

$$X^2 = -2 \log L (\text{reduced model}) - (-2 \log L (\text{full model})) = 68.92 - 54.08 = 14.83 > \chi_1^2 = 3.84$$

Stratification

The stratified Cox model allows the form of the underlying hazard function to vary across levels of stratification variables.

What does this mean?

- Suppose we have a predictor of interest, X , this could be treatment, for example.
- Z is a secondary categorical predictor that we want to adjust for when making inferences about X 's relationship to the time-to-event endpoint.
- Note: While the risk factor or exposure that Z represents does not have to be categorical, Z does. Z could be the

result of breaking a continuous predictor into groups.

To adjust for Z without estimating its effect on the outcome, we can perform a stratified analysis by fitting

$$h(t|X, Z = j) = h_j(t) \exp(X\beta), j = 1, \dots, C \text{ (number of levels in } Z)$$

There is only one parameter fit for X 's effect on the hazard, but it is adjusted for Z . This is different than an interaction.

Why stratify?

- Allows a factor to be adjusted for without estimating its effect.
- Predictor may not satisfy proportional hazards assumption, and it may be too complicated to model the hazard ratio for that predictor as a function of time.
- Can be used to make graphical checks of the proportional hazards assumption. We will look at this more later.

Stratification example

Suppose in the previous ovarian cancer example, we stratify by age instead of including it as a predictor in the model. We will create two strata, women over the age of 60 ($n=7$) and women younger than 60 ($n=19$). There are 6 events in each group. We fit the following regression model,

$$h(t|rx, agegrp) = h_{agegrp}(t) \exp(rx \times \beta), \quad agegrp = 1, 2.$$

```
> cph2=coxph(Surv(futime, fustat)~rx+strata(age>60), ovarian)
> cph2
```

	coef	exp(coef)	se(coef)	z	p
rx	-0.43	0.65	0.6	-0.716	0.47

```
Likelihood ratio test=0.52 on 1 df, p=0.471 n= 26
```


How does this compare to an additive model with *agegrp* as a predictor?

```
> cph3=coxph(Surv(futime, fustat)~rx+(age>60), ovarian)
```

```
> summary(cph3)
```

Call:

```
coxph(formula = Surv(futime, fustat) ~ rx + (age > 60), data = ovarian)
```

n= 26

	coef	exp(coef)	se(coef)	z	p
rx	-0.273	0.761	0.613	-0.444	0.66000
age > 60TRUE	2.258	9.560	0.685	3.294	0.00099

	exp(coef)	exp(-coef)	lower .95	upper .95
rx	0.761	1.313	0.229	2.53
age > 60TRUE	9.560	0.105	2.495	36.63

Rsquare= 0.364 (max possible= 0.932)

Likelihood ratio test= 11.8 on 2 df, p=0.00281

Wald test = 12.2 on 2 df, p=0.00221

Score (logrank) test = 17.3 on 2 df, p=0.000172

What about an interaction?

```
> coxph(Surv(futime, fustat)~rx*(age>60), ovarian)
```

```
Call:
```

```
coxph(formula = Surv(futime, fustat) ~ rx * (age > 60), data = ovarian)
```

	coef	exp(coef)	se(coef)	z	p
rx	-0.350	0.705	0.818	-0.428	0.67
age > 60TRUE	2.012	7.479	1.847	1.089	0.28
rx:age > 60TRUE	0.176	1.192	1.230	0.143	0.89

```
Likelihood ratio test=11.8 on 3 df, p=0.0082 n= 26
```