Cox proportional hazards models

Kunyang He, Yuang Li, Jiayi Li, Baiding Chen

biomedical Statistics (Wisdom Lake Academy of Pharmacy)
Xi'an Jiaotong-Liverpool University

5.10.2023

Outline

Outline

Review

Cox proportional hazards regression model
The Cox PH model

- is a semiparametric model
- makes no assumptions about the form of h(t) (nonparametric 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.

Model

- The survival function : the probability of being alive at time t : $S(t, X) = P(T > t, X)$
- The failure function: the probability of dying at time t: $F(t, X) = P(T \leq t, X)$
- Failure density function

$$f(t, x) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t, X)}{\Delta t} = F(t, X)'$$

Hazard function

$$h(t, X) = \lim_{\Delta t \to 0} \frac{P(t<T<t+\Delta t|T>t,X)}{\Delta t}$$
$$= \frac{1}{P(T>t,X)} \lim_{\Delta t \to 0} \frac{P(t<T<t+\Delta t,X)}{\Delta t} = \frac{f(t,X)}{S(t,X)}$$

Model

The proportional hazards regression model is given by

Cox proportional hazards models

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

- The predictors, $X_1, \ldots, X_p$ are assumed to act additively on $\log h(t \mid x)$.
- $\log h(t \mid x)$ changes linearly with the βs.
- 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 .

Model

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

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

- The point estimate for the hazard ratio is

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

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

Outline

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 .

- 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.

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

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

The log partial likelihood is given by

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

Outline

Model application

$$h(t \mid sex, age, ph.ecog) = h(t)\exp\left(\beta_1 \times sex + \beta_2 \times age + \beta_3 \times ph.ecog\right)$$

```
              coef exp(coef)  se(coef)      z Pr(>|z|)
age       0.011067  1.011128  0.009267  1.194 0.232416
sex      -0.552612  0.575445  0.167739 -3.294 0.000986 ***
ph.ecog   0.463728  1.589991  0.113577  4.083 4.45e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        exp(coef) exp(-coef) lower .95 upper .95
age        1.0111     0.9890    0.9929    1.0297
sex        0.5754     1.7378    0.4142    0.7994
ph.ecog    1.5900     0.6289    1.2727    1.9864

Concordance= 0.637  (se = 0.025 )
Likelihood ratio test= 30.5  on 3 df,   p=1e-06
Wald test            = 29.93  on 3 df,   p=1e-06
Score (logrank) test = 30.5  on 3 df,   p=1e-06
```

Figure: The result of the model

Visualizing the estimated distribution of survival time

Having fit a Cox model to the data, it's possible to visualize the predicted survival proportion at any given point in time for a particular risk group. The function survfit() estimates the survival proportion, by default at the mean values of covariates.

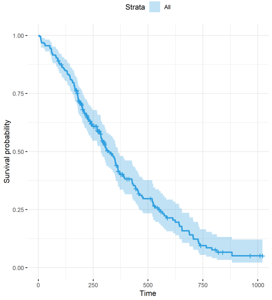

Figure: The result of the model

Visualizing the estimated distribution of survival time

Consider that, we want to assess the impact of the sex on the estimated survival probability. In this case, we construct a new data frame with two rows, one for each value of sex; the other covariates are fixed to their average values (if they are continuous variables) or to their lowest level (if they are discrete variables).

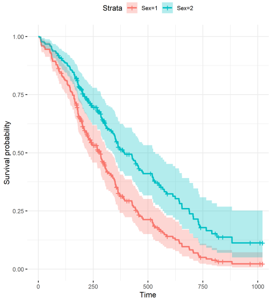

Figure: The result of the model