PH 716 Applied Survival Analysis

Part V: Cox Proportional Hazards Model

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2024/Feb/18 22:53:17

Assumptions for Cox proportional hazards (PH) model

- Observed $\widetilde{T}_i = \tilde{t}_i$ and $\Delta_i = \delta_i$ (event indicator)
- T_i are independent across i, given $x_{i1}, \ldots, x_{ip}$
- The independent and non-informative censoring
- $\lambda_{T_i}(t) = h(t \mid x_{i1}, \ldots, x_{ip}) = h_0(t)\exp(\sum_{j=1}^p x_{ij}\beta_j)$, or equiv. $\ln \lambda_{T_i}(t) = \ln h_0(t) + \sum_{j=1}^p x_{ij}\beta_j$
  - $h_0(t)$ (the baseline hazard): obtained when all covariates are zeros and left unspecified
    - A semi-parametric generalized linear model: nonparmetric baseline hazard + paramatric proportion
  - Proportional hazards: the HR between any two individuals, say $\lambda_{T_{i_1}}(t)/\lambda_{T_{i_2}}(t) = \exp(\sum_{j=1}^p x_{i_1 j}\beta_j - \sum_{j=1}^p x_{i_2 j}\beta_j)$, is constant over time
  - No intercept β_0
  - Interpretation of β_j: $\exp(\beta_j)$ is the HR associated with one-unit change in the jth covariate, fixing everything else

Weibull regression as a special case of Cox PH model

- Recall the Weibull regression: $\ln T_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \sigma\varepsilon_i$ with $\varepsilon_i \overset{\text{iid}}{\sim} F_{\varepsilon_i}(\epsilon) = 1 - \exp(-\exp\epsilon)$
  - $S_{T_i}(t) = \exp[-\{t/\exp(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)\}^{1/\sigma}] \Rightarrow \lambda_{T_i}(t) = (1/\sigma)t^{1/\sigma - 1}\exp\{(-\beta_0 - \sum_{j=1}^p x_{ij}\beta_j)/\sigma\}$
- $\lambda_{T_i}(t) = h_0(t)\exp(\sum_{j=1}^p x_{ij}\beta_j^*)$ if $h_0(t) = (1/\sigma)t^{1/\sigma-1}\exp(-\beta_0/\sigma)$ and $\beta_j^* = -\beta_j/\sigma$, $j = 1, \ldots, p$
- The only continuous-time model that is both a Cox PH and an AFT model

Partial likelihood (assuming no tied failure time)

- The observed-data likelihood $L(\boldsymbol{\beta}, h_0) = \prod_i \lambda_{T_i}(\tilde{t}_i)^{\delta_i} S_{T_i}(\tilde{t}_i)$ relying on both $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_j]^\top$ and unspecified $h_0(\cdot)$
- Further assumptions
  - K and only K distinct, ordered failure times, say $t_1 < \cdots < t_K$
  - No tied failure time: for each k, there is one and only one individual, say subject i_k, who fails at t_k
  - Risk set $\mathcal{R}(t) = \{i : \widetilde{T}_i \geq t\}$: the set of individuals who are known to survive just prior to time t
- Rephrase $L(\boldsymbol{\beta}, h_0)$:

$$L(\boldsymbol{\beta}, h_0) = \prod_{i=1}^n \lambda_{T_i}(\tilde{t}_i)^{\delta_i} S_{T_i}(\tilde{t}_i) = \prod_{i=1}^n \left\{\frac{\lambda_{T_i}(\tilde{t}_i)}{\sum_{\ell\in\mathcal{R}(\tilde{t}_i)} \lambda_{T_\ell}(\tilde{t}_i)}\right\}^{\delta_i} \times \left\{\sum_{\ell\in\mathcal{R}(\tilde{t}_i)} \lambda_{T_\ell}(\tilde{t}_i)\right\}^{\delta_i} \times S_{T_i}(\tilde{t}_i)$$

- Take the partial likelihood (i.e., the first term of the above $L(\boldsymbol{\beta}, h_0)$)

$$pL(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{\lambda_{T_i}(\tilde{t}_i)}{\sum_{k\in\mathcal{R}(\tilde{t}_i)} \lambda_{T_k}(\tilde{t}_i)} \right\}^{\delta_i} = \prod_{i=1}^{n} \left\{ \frac{\exp(\sum_{j=1}^{p} x_{ij}\beta_j)}{\sum_{\ell\in\mathcal{R}(\tilde{t}_i)} \exp(\sum_{j=1}^{p} x_{\ell j}\beta_j)} \right\}^{\delta_i} = \prod_{k=1}^{K} \frac{\exp(\sum_{j=1}^{p} x_{kj}\beta_j)}{\sum_{\ell\in\mathcal{R}(t_k)} \exp(\sum_{j=1}^{p} x_{\ell j}\beta_j)}$$

  as a surrogate of $L(\boldsymbol{\beta}, h_0)$ in estimating $\boldsymbol{\beta}$

  - Cox (1972) argued that $L_{\text{partial}}(\boldsymbol{\beta})$ contained almost all the information about $\boldsymbol{\beta}$
  - Extensive evidence, both theoretical and numerical, supported this argument in the past few decades
  - Log-partial likelihood

$$p\ell(\boldsymbol{\beta}) = \ln pL(\boldsymbol{\beta}) = \sum_{k=1}^{K} \left\{ \sum_{j=1}^{p} x_{kj}\beta_j - \ln \sum_{\ell\in\mathcal{R}(\tilde{t}_k)} \exp\left(\sum_{j=1}^{p} x_{\ell j}\beta_j \right) \right\}$$

- Another look at $pL(\boldsymbol{\beta})$:

$$pL(\boldsymbol{\beta}) = \prod_{k=1}^{K} \frac{\Pr(\text{subject } i_k \text{ fails at time } t_k \mid \text{it is at risk at } t_k)}{\Pr(\text{there is one and only one failure at time } t_k \mid \text{it is at risk at } t_k)}$$

Ex. 5.1 The calculation of partial likelihood

i	$\tilde{t}_i$	δ_i	x_i
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

- Key point: follow the following definition (no need to reorder failure times) and fill in the table

$$pL(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left\{ \frac{\exp(\sum_{j=1}^{p} x_{ij}\beta_j)}{\sum_{\ell\in\mathcal{R}(\tilde{t}_i)} \exp(\sum_{j=1}^{p} x_{\ell j}\beta_j)} \right\}^{\delta_i}$$

i	$\tilde{t}_i$	δ_i	x_i	$\mathcal{R}(\tilde{t}_i)$	$\left\{\frac{\exp(x_k\beta)}{\sum_{\ell\in\mathcal{R}(t_k)}\exp(x_\ell\beta)}\right\}^{\delta_i}$
1	9	1	4		
2	8	0	5		
3	6	1	7		
4	10	1	3		

Ex. 5.2 The calculation of partial likelihood: comparison of two groups

- Covariate x_i indicating the group label

i	$\tilde{t}_i$	δ_i	x_i	$\mathcal{R}(\tilde{t}_i)$	$\frac{\exp(x_k\beta)}{\sum_{\ell\in\mathcal{R}(t_k)}\exp(x_\ell\beta)}$
1	4	0	0		
2	7	1	0		

i	$\tilde{t}_i$	δ_i	x_i	$\mathcal{R}(\tilde{t}_i)$	$\frac{\exp(x_k\beta)}{\sum_{\ell\in\mathcal{R}(t_k)}\exp(x_\ell\beta)}$
3	8	0	0		
4	9	1	0		
5	10	0	0		
6	3	1	1		
7	5	1	1		
8	5	0	1		
9	6	1	1		
10	8	0	1		

```r
library(survival)
data = data.frame(
  tte = c(4,7,8,9,10,3,5,5,6,8),
  delta = c(0,1,0,1,0,1,1,0,1,0),
  x = c(0,0,0,0,0,1,1,1,1,1)
)
fit = coxph(Surv(tte,delta)~x+x2, data = data)
fit1 = coxph(Surv(tte,delta)~x2, data = data)
anova(fit, fit1)
summary(fit)
```

- $\exp(\beta)$ is the HR of $group = 1$ against $group = 0$, fixing everything else (if any). It implies that the hazard in group 1 is $\exp(\beta) \times 100\%$ that in group 0.
- Is there any difference between the survival of the two groups? There are at least four p-values. Which one shall we refer to?
- What are meanings of other digits in the output?
- What if there are more covariates?

Ex. 5.3. Leukemia data (with tied event/failure times)

```r
survival::leukemia
```

Partial likelihood (Cox's modification)

- Assumptions
  - K and only K distinct, ordered failure times, say $t_1 < \cdots < t_K$
  - d_k failures at time t_k: there are d_k individuals, say subject $i_{k,1}, \ldots, i_{k,d_k}$, who fail at t_k
  - Risk set $\mathcal{R}(t) = \{i : \widetilde{T}_i \geq t\}$: the set of individuals who are known to survive just prior to time t
- Accordingly

$$pL(\boldsymbol{\beta}) = \prod_{k=1}^{K} \frac{\Pr(\text{subjects } i_{k,1},\ldots,i_{k,d_k} \text{ fail at time } t_k \mid \text{they are at risk at } t_k)}{\Pr(\text{there are } d_k \text{ failures at time } t_k \mid \text{they are at risk at } t_k)} = \prod_{k=1}^{K} \frac{\exp(\sum_j \sum_{i\in R_0} x_{ij}\beta_j)}{\sum_{R\in\mathcal{S}(k)} \exp(\sum_j \sum_{i\in R} x_{ij}\beta_j)}$$

  - $\mathcal{S}(k)$: the set of all possible combinations of d_k individuals that can be drawn from $\mathcal{R}(\tilde{t}_k)$
    - If $R \in \mathcal{S}(k)$, then R is a set of d_k individuals who are at risk at t_k.
      - Specifically, $D(t_k) = \{i_{k,1}, \ldots, i_{k,d_k}\} \in \mathcal{S}(k)$ denotes the set of all the d_k individuals who fail at time t_k
- Labeled as `exact` by `survival::coxph`

Partial likelihood (Breslow's approximation)

- Keeping the assumptions for the Cox's modification
- Substitute $\{\sum_{\ell\in\mathcal{R}(t_k)}\exp(\sum_{j=1}^p x_{\ell j}\beta_j)\}^{d_k}$ for the denominator of Cox's modification

$$pL(\boldsymbol{\beta}) = \prod_{k=1}^{K} \frac{\exp(\sum_j \sum_{i\in D(t_k)} x_{ij}\beta_j)}{\{\sum_{\ell\in\mathcal{R}(t_k)}\exp(\sum_{j=1}^p x_{\ell j}\beta_j)\}^{d_k}}$$

Partial likelihood (Efron's approximation)

- Keeping the assumptions for the Cox's modification
- Substitute $\{\sum_{\ell\in\mathcal{R}(t_k)}\exp(\sum_{j=1}^p x_{\ell j}\beta_j)\}^{d_k}$ for the denominator of Cox's modification

$$pL(\boldsymbol{\beta}) = \prod_{k=1}^{K} \frac{\exp(\sum_j \sum_{i\in D(t_k)} x_{ij}\beta_j)}{\prod_{m=1}^{d_k}\{\sum_{\ell\in\mathcal{R}(t_k)}\exp(\sum_{j=1}^p x_{\ell j}\beta_j) - \frac{m-1}{d_k}\sum_{i\in D(t_k)}\exp(\sum_j x_{ij}\beta_j)\}}$$

- Default tie-handling method by `survival::coxph`

Summary of handling ties

- With no ties, all approximation options give exactly the same results
- With only a few ties, all approximations yield pretty much the same results
- With many ties (relative to the number at risk), both of Breslow's and Efron's approximations yield coefficients $\boldsymbol{\beta}$ that are biased toward 0.
- Computing time of Cox's method is substantially longer than that of approximate methods. But it is not a big issue with today's hardwares.
- The Efron's approximation almost always works better than the Breslow's method, without consuming more time.

Revisit Ex. 5.3. Leukemia data (with tied event/failure times)

```r
library(survival)
data = survival::leukemia
fit1 = coxph(Surv(time,status)~x, data = data)
fit2 = coxph(Surv(time,status)~x, data = data, ties = 'efron')
fit3 = coxph(Surv(time,status)~x, data = data, ties = 'breslow')
fit4 = coxph(Surv(time,status)~x, data = data, ties = 'exact')
c(coef(fit1), coef(fit2), coef(fit3), coef(fit4))
```

CIs and hypothesis tests for HRs

- Suppose the HR of interest is the one associated with the one-unit increase of the jth covairate, i.e., $\exp(\beta_j)$
- $\text{var}\{\exp(\hat\beta_j)\} \approx \exp(2\hat\beta_j)\text{var}(\hat\beta_j)$ (delta method)
  - Hence $\text{se}(\exp(\hat\beta_j)) \approx \exp(\hat\beta_j)\text{se}(\hat\beta_j)$
- 95% CI for $\exp(\beta_j)$
  - $\exp(\hat\beta_j) \pm \Phi^{-1}(.975) \times \text{se}(\exp(\hat\beta_j))$

    - $\Phi^{-1}(.975)$ (≈ 1.96): the .975 quantile of $N(0,1)$
  - $\exp(\hat{\beta}_j \pm \Phi^{-1}(.975) \times \text{se}(\hat{\beta}_j))$ (preferred; why?)

- Hypothesis test for $H_0 : \exp(\beta_j) = 1$ (i.e., $\beta_j = 0$) vs. H_1 : otherwise.
  - Wald test statistic: $\hat{\beta}_j/\text{se}(\hat{\beta}_j) \approx N(0,1)$ under H_0
    - Equivalent to checking whether $\exp(\hat{\beta}_j \pm \Phi^{-1}(.975) \times \text{se}(\hat{\beta}_j))$ covers 1

- LRT to compare two nested models
  - Model 1 nested to Model 2
    - Model 1: $h(t \mid x_{i1}, \ldots, x_{ip}) = h_0(t)\exp(\sum_{j=1}^{p} x_{ij}\beta_j)$
    - Model 2: $h(t \mid x_{i1}, \ldots, x_{ip}, x_{i,q+1}, \ldots, x_{i,p+q}) = h_0(t)\exp(\sum_{j=1}^{p+q} x_{ij}\beta_j)$
  - H_0 : Model 1 is correct (i.e., $\beta_{p+1} = \cdots = \beta_q = 0$) vs. H_1 : Model 2 is correct
  - Test statistic: $2(\ln L_{\text{Model2}} - \ln L_{\text{Model1}}) \approx \chi^2(q)$ under H_0

Ex. 5.4. Nursing home data

- Variables:
  - ID: Patient ID
  - lstay: Length of stay of a resident (in days)
  - age: Age of a resident
  - trt: Nursing home assignment (1: receive treatment, 0: control)
  - gender: Gender (1:male, 0:female)
  - marstat: Marital status (1: married, 0: not married)
  - hlstat: Health status (2: second best, 5: worst)
  - cens: Censoring indicator (1:censored, 0: discharged)

```r
options(digits=4)
library(survival)
data = read.csv("NursingHome.csv")
data$event <- 1-data$cens
fit1 <- coxph(Surv(lstay,event) ~ trt + age + gender + marstat + hlstat, data=data)
summary(fit1)

# Testing if trt is necessary against the full model
fit2 <- coxph(Surv(lstay,event) ~ age + gender + marstat + hlstat, data=data)
anova(fit1, fit2)
summary(fit2)

# Testing if trt, age and marstat are necessary against the full model
fit3 <- coxph(Surv(lstay,event) ~ gender + hlstat, data=data)
anova(fit1, fit3)
summary(fit3)
```