

# cox proportional hazard model

Jieun Shin

2023-06-11

## 1. proportional hazards regression model

비례위험모형은 다음과 같이 주어진다:

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

비례위험모형에서는 predictor  $X_1, \dots, X_p$ 이  $\log h(t|x)$ 에 가법적으로 영향을 미친다고 가정한다.  $\log h(t|x)$ 은  $\beta$ 들에 대해 선형적으로 변화한다. 그리고 predictor들의 영향은 모든 시간  $t$ 에 대해 동일하다.

$X$ 에 대한  $X^*$ 의 위험율 (HR; hazard ratio)은  $hr(X^* : X) = \frac{\exp(X^*\beta)}{\exp(X\beta)} = \exp\{(X^* - X)\beta\}$ 이다. 이에 따라 HR에 대한 point estimate은  $\hat{hr}(X^* : X) = \exp\{(X^* - X)\hat{\beta}\}$ 가 되며 여기서  $\hat{\beta}$ 는  $\beta$ 의 MLE이다. HR에 대한  $(1 - \alpha)100\%$  신뢰구간은  $\exp\{(X^* - X)\hat{\beta} \pm Z_{1-\alpha/2}\hat{se}((X^* - X)\hat{\beta})\}$ 이다.

## 2. cox proportional hazards regression model

cox PH model의 가정은 다음과 같다.

1. 다른 개체들 간의 생존시간이 독립적이다.
2. 예측 변수와 위험 사이에 곱셈 관계가 있어야 한다. (선형 관계가 아님)
3. 시간에 따라 일정한 위험비가 유지되어야 한다.

cox PH model은 기저위험률 (baseline hazard)  $h(t)$ 에 대한 가정이 없어 semi-parametric model라고도 불린다.  $h(t)$ 의 모양 (shape)에 대한 가정이 없어 non-parametric part이고, predictors의 효과에 대한 가정은 있으므로 parametric part이다. 대부분의 문제에서 hazard의 모양보다는 모수  $\beta$ 의 추정에 관심이 있는 편이다.

$\beta$ 의 추정치는 다음과 같은 partial likelihood를 최대화함으로써 얻어진다:

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

log partial likelihood는 다음과 같다:

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

모수 추정 과정에서 baseline hazard  $h(t)$ 의 형태는 추정하지 않는다. cox PH model은 baseline hazard를 추정하지 않고도 위험비 ( $\exp(\beta)$ )를 추정할 수 있도록 설계된 모형이다. 추정된  $\beta$ 는 일반 회귀분석에서 해석하듯이 하면 된다. 다만  $\exp(\beta)$ 을 HR이라고 간주하면 된다. 만약 predictor가  $X_1$ 로 한 개 뿐일 때 cox PH model은

$$h(t|X) = h(t) \exp(X_1\beta_1)$$

일 것이다.  $X_1 = a$ 와  $X_1 = b$ 의 기대위험 (expected hazard)을 비교해보자. 먼저  $X_1 = a$ 일때의 기대위험은  $h(t|X = a) = h(t) \exp(\beta_1 a)$ 이고  $X_1 = b$ 일때의 기대위험은  $h(t|X = b) = h(t) \exp(\beta_1 b)$ 이다. 두 기대위험의 비는  $\frac{h(t) \exp(\beta_1 a)}{h(t) \exp(\beta_1 b)} = \exp\{\beta_1(a - b)\}$ 으로 시간  $t$ 에 의존하지 않는 것을 알 수 있다.

### 3. predict hazard의 구현

여기까지 공부를 했는데 막상 predict hazard function은 어떻게 구현하는건지 알 길이 없었다. 이 기회에 정리해보려 한다.

먼저 데이터셋은 smoothHR라이브러리에 있는 whas500을 사용했다. time은 lenfol, reponse는 fstat이며, 독립변수를 bmi 1개로 두고 restruct cubic spline을 basis로 하였다. 이렇게 모델링을 한 뒤, predicted hazard plot을 그리기 위해 새로운 데이터 bmi=seq(0, 40, by=1)을 모델에 적합하였다.

```
# 패키지
library(smoothHR)

## Warning: 패키지 'smoothHR'는 R 버전 4.2.3에서 작성되었습니다
## 필요한 패키지를 로딩중입니다: survival
## Warning: 패키지 'survival'는 R 버전 4.2.3에서 작성되었습니다
## 필요한 패키지를 로딩중입니다: splines
library(survival)
library(rms)

## Warning: 패키지 'rms'는 R 버전 4.2.3에서 작성되었습니다
## 필요한 패키지를 로딩중입니다: Hmisc
## Warning: 패키지 'Hmisc'는 R 버전 4.2.3에서 작성되었습니다
##
## 다음의 패키지를 부착합니다: 'Hmisc'
## The following objects are masked from 'package:base':
##
##      format.pval, units
## 필요한 패키지를 로딩중입니다: lattice
## 필요한 패키지를 로딩중입니다: ggplot2
## 필요한 패키지를 로딩중입니다: SparseM
##
## 다음의 패키지를 부착합니다: 'SparseM'
## The following object is masked from 'package:base':
##
##      backsolve
head(whas500)

##   id age gender hr sysbp diasbp      bmi cvd afb sho chf av3 miord mitype year
## 1  1  83      0  89  152    78 25.54051  1  1  0  0  0    1    0    1
## 2  2  49      0  84  120    60 24.02398  1  0  0  0  0    0    1    1
## 3  3  70      1  83  147    88 22.14290  0  0  0  0  0    0    1    1
## 4  4  70      0  65  123    76 26.63187  1  0  0  1  0    0    1    1
## 5  5  70      0  63  135    85 24.41255  1  0  0  0  0    0    1    1
## 6  6  70      0  76   83    54 23.24236  1  0  0  0  1    0    0    1
##   admitdate   disdate   fdate los dstat lenfol fstat
```

```
## 1 13-01-1997 18-01-1997 31-12-2002 5 0 2178 0
## 2 19-01-1997 24-01-1997 31-12-2002 5 0 2172 0
## 3 01-01-1997 06-01-1997 31-12-2002 5 0 2190 0
## 4 17-02-1997 27-02-1997 11-12-1997 10 0 297 1
## 5 01-03-1997 07-03-1997 31-12-2002 6 0 2131 0
## 6 11-03-1997 12-03-1997 12-03-1997 1 1 1 1
```

```
fit_spline = cph(Surv(lenfol, fstat) ~ rcs(bmi,3), data=whas500)
ggplot(rms::Predict(fit_spline, bmi=seq(0, 40, by=1), fun=exp))
```

