

Cox Proportional Hazard Model

Chao Cheng

December 19, 2022

Contents

1 Introduction	1
2 Estimation	2
3 Inference	2

1 Introduction

In this note we will talk about the Cox's proportional hazards (Cox's PH) model. Suppose we observe some non-informatively right-censored data (U, δ) with covariate vector Z . That is, for subject i , the covariate vector is Z_i , survival time T_i and censoring time C_i . The observed data is (U_i, δ_i) where $U_i = \min(T_i, C_i)$ and $\delta_i = 1(T_i \leq C_i)$. Also $T_i \perp C_i | Z_i$.

And now we want to model the relationship between Z and T . One way to do that is to incorporate Z into the hazard function $h(\cdot)$, e.g.,

$$T \sim \text{Exp}(\lambda_Z) \implies h(t) = \lambda_Z \triangleq e^{\alpha + \beta Z} = \lambda_0 e^{\beta Z},$$

where $\lambda_0 = e^\alpha$ can be viewed as a baseline hazard. If $\beta = 0$ then Z is not associated with T .

We can generalize this idea as

$$h(t|Z) = h_0(t) \times g(Z).$$

So the hazard can be factorized and this model is sometimes called a “multiplicative intensive model” or “multiplicative hazard model” or “proportional hazard model” because this factorization implies that

$$\frac{h(t|Z = z_1)}{h(t|Z = z_2)} = \frac{g(z_1)}{g(z_2)}.$$

The hazard ratio is constant with respect to t , hence the (constant) proportional hazard. So in our previous model (the exponential survival time), the hazard ratio is

$$\frac{h(t|Z = z_1)}{h(t|Z = z_2)} = e^{\beta(z_1 - z_2)}.$$

Also this exponential form of $g(Z)$

$$h(t|Z) = h_0(t) \cdot e^{\beta Z} \tag{1}$$

is the **Cox's PH** model.

2 Estimation

(1) implies that

$$\begin{aligned}
 S(t|Z) &= \exp(-H(t|Z)) \\
 &= \exp\left(-\int_0^t h(u|Z) du\right) \\
 &= \exp\left(-\int_0^t h_0(t) du \cdot g(Z)\right) \\
 &= (S_0(t))^{g(Z)} = (S_0(t))^{\exp(\beta Z)},
 \end{aligned}$$

where $S_0(t) = \exp\left(-\int_0^t h_0(t) du\right)$, the survival function for $Z = 0$, hence $S(t|Z = 0)$. Also remember that $f(t|Z) = h(t|Z) S(t|Z)$. Thus, given n independent data (u_i, δ_i, z_i) , the likelihood ([one can refer to our previous notes about survival analysis.](#)) is

$$\begin{aligned}
 L(\beta, h_0(\cdot)) &= \prod_{i=1}^n (f(u_i|z_i))^{\delta_i} (S(u_i|z_i))^{1-\delta_i} = \prod_{i=1}^n h(u_i|z_i)^{\delta_i} S(u_i|z_i) \\
 &= \prod_{i=1}^n (h_0(u_i|z_i) e^{\beta z_i})^{\delta_i} \left(\exp\left(-\int_0^{u_i} h_0(t) dt\right) \right)^{\exp(\beta z_i)} \\
 &= \text{function}(data, h_0(\cdot), \beta).
 \end{aligned}$$

If $h_0(\cdot)$ is allowed to be “arbitrary”, then the “parameter space “ is

$$\mathcal{H} \times \mathcal{R}^p = \left\{ (h(\cdot), \beta) \mid h_0(\cdot) \geq 0, \int_0^\infty h_0(t) dt = \infty, \beta \in \mathcal{R}^p \right\},$$

where $\int_0^\infty h_0(t) dt = \infty$ ensures that $S_0(\infty) = 0$.

3 Inference

References