Stratification Factor or Covariate

Chao Cheng

July 17, 2025

Contents

1 Introduction

In this note we will talk about the Cox's proportional hazards (Cox's PH) model. Suppose we observe some non-informativly 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.,

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

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

In this section, we will talk about what is the objective function for Cox's model. But we will not talk about the detailed optimization algorithm. (1) implies that

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

where $S_0(t) = \exp\left(-\int_0^t h_0(u)\,\mathrm{d}u\right)$, 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

$$\begin{aligned}
L(\beta, h_0(\cdot)) =& \prod_{i=1}^{n} \left(f(u_i|z_i)\right)^{\delta_i} \left(S(u_i|z_i)\right)^{1-\delta_i} = \prod_{i=1}^{n} h(u_i|z_i)^{\delta_i}\, S(u_i|z_i) \\
=& \prod_{i=1}^{n} \left(h_0(u_i)\, e^{\beta z_i}\right)^{\delta_i} \left(\exp\left(-\int_0^{u_i} h_0(t)\,\mathrm{d}t\right)\right)^{\exp(\beta z_i)} \\
=& \text{function } (data, h_0(\cdot), \beta).
\end{aligned} \tag{2}$$

If $h_0(\cdot)$ is allowed to be "arbitary", then the "parameter space " is

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

where $\int_0^\infty h_0(t)\,\mathrm{d}t = \infty$ ensures that $S_0(\infty) = 0$.

In general this likelihood is hard to maximize. And Cox proposed this idea: to factor $L(\beta, h_0(\cdot))$ as

$$L(\beta, h_0(\cdot)) = L_1(\beta) \times L_2(\beta, h_0(\cdot)),$$

where L_1 only depends on β and its maximization $(\hat{\beta})$ enjoys nice properties such as consistency and asymptotic normality while L_2 contains relatively little information about β. And this L_1 is called a **partial likelihood**.

2.1 What is $L_1(\beta)$

In this section we introduce the L_1 proposed by Cox. First let's assume there are **NO tied** nor censoring observations. And define the distinct times of failure $\tau_1 < \tau_2 < \cdots$. Denote

$$R_j = \{i | U_i \geq \tau_j\} = \text{risk set at } \tau_j,$$

and

$$Z_{(j)} = \text{value of Z for the subject who fails at } \tau_j.$$

we can reconstruct the data from $\{\tau_j\}$, $\{R_j\}$ and $\left\{Z_{(j)}\right\}$. And L_1 is defined as

$$L_1(\beta) \triangleq \prod_j \left\{ \frac{e^{\beta Z_{(j)}}}{\sum\limits_{l \in R_j} e^{\beta Z_l}} \right\}. \tag{3}$$

3 Extensions

3.1 Stratified Cox's model

Assume we have two binary covariates: Z for treatment or control, W for male or female. We can incorporate them into the Cox's model as

$$h(t|w,z) = h_0(t)\, e^{\beta_1 z + \beta_2 w}.$$

Then test for β_1 would tell us about the treatment effect and test for β_2 would tell up whether there is difference between male and female. Like we talked in the Logrank-test notes, this model is assumeing constant hazard ratio between both treatment and gender, which means

$$HR = \frac{h(t|z=0,w=0)}{h(t|z=1,w=0)} = \frac{h(t|z=0,w=1)}{h(t|z=1,w=1)} = e^{\beta_1}$$
$$HR = \frac{h(t|z=0,w=0)}{h(t|z=0,w=1)} = \frac{h(t|z=1,w=0)}{h(t|z=1,w=1)} = e^{\beta_2}.$$

But sometimes we just want to assume the constant hazard ratio between different treatment groups and let the hazard between male and female to be "arbitrary". Then we can consider the stratified Cox's model

$$h(t|z,w) = h(t|w)\, e^{\beta z}, \tag{4}$$

where w is a categorical variable with L levels. These L levels of W can have arbitrary underlying hazard, yet within each, the treatment relative risk is e^{β}. For each stratified level, we can construct the partial likelihood as normal, denoted by $L_1^{(l)}(\beta)$, then the overall partial likelihood is

$$L_1(\beta) = \prod_{l=1}^{L} L_1^{(l)}(\beta). \tag{5}$$

Note: when Z is binary, the resulting partial likelihood score test for $\beta = 0$ reduced to stratified logrank test.

References