Cox Proportional Hazard Model

Chao Cheng

December 28, 2022

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

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

Also this exponential form of $g\left(Z\right)$

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

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

2 Estimation

(1) implies that

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

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

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

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

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

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

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

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

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

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

(Cox model assumes the time measurement to be continuous, but here we think about discrete time point for some intuition.)

2.1.1 Intuition: profile likelihood perspective

Note that under this setting (no tie, no censor), the full likelihood (2) becomes

$$L(\beta, h_0(\cdot)) = \prod_{i=1}^{n} h_0(u_i) e^{\beta z_i} \left(\exp\left(-\int_0^{u_i} h_0(t)\, \mathrm{d}t \right) \right)^{\exp(\beta z_i)}.$$

Furthermore, we can assume $u_i = \tau_i$, i.e. the data has been sorted based on survival time. And use the KM idea, i.e. assume the survival function is **discrete** with baseline hazard value h_j at u_j. Then this likelihood becomes

$$L(\beta, h_1, \cdots, h_n) = \prod_{i=1}^{n} h_i e^{\beta z_i} \exp\left(-\sum_{j=1}^{i} h_j \right)^{\exp(\beta z_i)}. \tag{4}$$

Note that, in previous notes we have deduct that in discrete case, for any $t \in [v_j, v_{j+1})$:

$$H(t) = \sum_{i=1}^{j} h_i \qquad S(t) = \prod_{i=1}^{j} (1 - h_i).$$

Here in (4) we use the approximation that $e^{-h_j} \approx 1 - h_j$ when h_j is close to 0.

We can use the method of profile likelihood: That is, for any given β, we maximize L (or equivalently, $\log L$) over h_js so the result is a function of β. Taking derivative, we have

$$\frac{\partial \log L}{\partial h_j} = \frac{1}{h_j} - \sum_{i \leq j} \exp(\beta z_i), \qquad j = 1, \cdots, n.$$

Set them to 0 we have $\hat{h}_j = 1/\sum_{i\leq j}\exp(\beta z_i)$. And the log profile likelihood of β is

$$\begin{aligned}
\log L_{profile}(\beta) =& \log\left\{\prod_{i=1}^{n}\frac{\exp(\beta z_i)}{\sum_{k\leq i}\exp(\beta z_k)}\exp\left(-\sum_{j=1}^{i}\frac{1}{\sum_{k\leq j}\exp(\beta z_k)}\right)^{\exp(\beta z_i)}\right\} \\
=& \sum_{i=1}^{n}\left\{\log\left\{\frac{\exp(\beta z_i)}{\sum_{k\leq i}\exp(\beta z_k)}\exp\left(-\sum_{j=1}^{i}\frac{1}{\sum_{k\leq j}\exp(\beta z_k)}\right)^{\exp(\beta z_i)}\right\}\right\} \\
=& \sum_{i=1}^{n}\left\{\log\left(\frac{\exp(\beta z_i)}{\sum_{k\leq i}\exp(\beta z_k)}\right) - \exp(\beta z_i)\cdot\left(\sum_{j=1}^{i}\frac{1}{\sum_{k\leq j}\exp(\beta z_k)}\right)\right\} \\
=& \sum_{i=1}^{n}\left\{\log\left(\frac{\exp(\beta z_i)}{\sum_{k\leq i}\exp(\beta z_k)}\right)\right\} - \sum_{i=1}^{n}\sum_{j=1}^{i}\frac{\exp(\beta z_i)}{\sum_{k\leq j}\exp(\beta z_k)},
\end{aligned} \tag{5}$$

where the second part of last equation can be reduced to $-n$?, which means

$$L_{profile}(\beta) \propto \prod_{i=1}^{n}\frac{\exp(\beta z_i)}{\sum_{k\leq i}\exp(\beta z_k)}.$$

And this is what Cox uses as $L_1(\beta)$.

2.1.2 Intuition: conditional distribution perspective

Given the fact that someone survies up to just prior to τ_j, hence in the risk set R_j, the hazard of someone with covariate value z failing at $t = \tau_j$ is

$$h_0(\tau_j)\exp(\beta z).$$

In discrete case, this is the conditional probability (in continuous case, this hazard value can go beyond 1.) of someone fails at τ_j given the fact that subject survives past τ_{j-1}.

To be added.

2.2 What if there is censoring?

Then (3) is still used.

2.3 What if there are tied event times?

Through the probability of tie existance is 0 in the continuous time case, in real life it is pretty comman.

2.3.1 Exact partial likelihood

The exact partial likelihood considers all the possible rankings for the tied observations. Specifically

$$L_1(\beta) = \prod_{j=1}^{K} \left\{ \sum_{\left(k_1,\cdots,k_{d_j}\right)=(1,2,\cdots,d_j)} \prod_{i=1}^{d_j} \left\{ \frac{\exp\left(\beta z_{(j)}^{k_i}\right)}{\sum_{l\in R_j} \exp(\beta z_l) - \sum_{s=i}^{d_j} \exp\left(\beta z_{(j)}^{(s)}\right)} \right\} \right\}.$$

The computation of the exact partial likelihood gets out of hand pretty quickly as d_j increases. So some modification/approximation methods are proposed.

2.3.2 Breslow's approximation

2.3.3 Efron's approximation

3 Inference

References