

Cox Proportional Hazard Model

Chao Cheng

December 29, 2022

Contents

1	Introduction	1
2	Estimation	2
2.1	What is $L_1(\beta)$	3
2.1.1	Intuition: profile likelihood perspective	3
2.1.2	Intuition: conditional distribution perspective	4
2.2	What if there is censoring?	5
2.3	What if there are tied event times?	5
2.3.1	Exact partial likelihood	5
2.3.2	Breslow's approximation	5
2.3.3	Efron's approximation	5
3	Inference	5

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.,

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

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 $g(Z)$

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

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

2 Estimation

(1) implies that

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

where $S_0(t) = \exp\left(-\int_0^t h_0(u) du\right)$, 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

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

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

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

where $\int_0^\infty h_0(t) dt = \infty$ 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 $\tau_1 < \tau_2 < \dots$. Denote

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

and

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

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

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

(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

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

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

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

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

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

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_j s so the result is a function of β . Taking derivative, we have

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

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 survives 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} .

Now, given the risk set R_j and the fact that the subject with z_* fails at τ_j , this conditional so-called “probability”/“likelihood” is

$$\frac{h_0(\tau_j) \exp(\beta z_*)}{\sum_{l \in R_j} h_0(\tau_j) \exp(\beta z_l)} = \frac{\exp(\beta z_*)}{\sum_{l \in R_j} \exp(\beta z_l)}.$$

Then the cumproduct over all τ_j leads to L_1 in Cox’s model.

Note: strictly speaking, this is not a conditional probability. In discrete case, let $h(z_1, \tau), \dots, h(z_n, \tau)$ denotes the hazard of subject i at τ , which is the conditional probability of the subject fails at τ_j given the fact that the subject survives past τ_{j-1} . Then given the risk set R_j and the fact that exactly one subject fails at τ_j . The probability of that subject being $z_i, i \in R_j$ is

$$\frac{h(z_i, \tau) \prod_{j \neq i} (1 - h(z_j, \tau))}{\sum_{i \in R_j} \left(h(z_i, \tau) \prod_{l \in R_j, l \neq i} (1 - h(z_l, \tau)) \right)}$$

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 existence is 0 in the continuous time case, in real life it is pretty common.

2.3.1 Exact partial likelihood

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

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

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