

# Cox Proportional Hazard Model

Chao Cheng

July 18, 2025

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Estimation</b>	<b>2</b>
2.1	What is $L_1(\beta)$	3
2.1.1	Intuition: parametric survival distributions	3
2.1.2	Intuition: profile likelihood perspective	5
2.1.3	Intuition: conditional distribution perspective	6
2.2	What if there is censoring?	7
2.3	What if there are tied event times?	7
2.3.1	Exact partial likelihood	7
2.3.2	Breslow's approximation	7
2.3.3	Efron's approximation	8
2.3.4	One example for comparison	8
2.3.5	A fast algorithm for Breslow's and Efron's approximation	8
<b>3</b>	<b>Inference</b>	<b>8</b>
<b>4</b>	<b>Extensions</b>	<b>9</b>
4.1	Stratified Cox's model	9
4.2	Cox's model for comparing multiple groups and trend test	10
4.2.1	Comparing multiple groups	10
4.2.2	Trend test	10
4.3	Time varying covariates	10
4.4	Estimating $H_0(t)$	10

## 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  $(U, \delta)$  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  $(U_i, \delta_i)$  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

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

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

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  $(u_i, \delta_i, z_i)$ , the likelihood (one can refer to our previous notes about survival analysis.) is

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

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  $\beta$  and its maximization ( $\hat{\beta}$ ) enjoys nice properties such as consistency and asymptotic normality while  $L_2$  contains relatively little information about  $\beta$ . 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

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

### 2.1.1 Intuition: parametric survival distributions

Suppose that base hazard  $h_0(t)$  is from Weibul distribution  $W(p, \lambda)$ , that is  $h_0(t) = p\lambda^p t^{p-1}$ , then (2) becomes

$$\begin{aligned} L(\beta, h_0(\cdot)) &= \prod_{i=1}^n (p\lambda^p u_i^{p-1} e^{\beta z_i})^{\delta_i} \left( \exp \left( - \int_0^{u_i} p\lambda^p u_i^{p-1} dt \right) \right)^{\exp(\beta z_i)} \\ &= \prod_{i=1}^n (p\lambda^p u_i^{p-1} e^{\beta z_i})^{\delta_i} \exp(-\lambda^p u_i^p \exp(\beta z_i)) \end{aligned}$$

And the log-likelihood is

$$\log L = \sum_{i=1}^n \{ \delta_i \log(p\lambda^p u_i^{p-1} e^{\beta z_i}) - \lambda^p u_i^p \exp(\beta z_i) \}$$

Here we utilize profile-likelihood perspective:

1. Fix  $\beta$  and  $p$ , we have

$$\begin{aligned} \frac{\partial \log L}{\partial \lambda} &= \sum_{i=1}^n \left\{ \frac{\delta_i p}{\lambda} - p\lambda^{p-1} u_i^p \exp(\beta z_i) \right\} \\ &= \frac{pn_{evt}}{\lambda} - p\lambda^{p-1} \sum_{i=1}^n u_i^p \exp(\beta z_i) \end{aligned}$$

Set this to 0, we have

$$\hat{\lambda} = \left( \frac{n_{evt}}{\sum_{i=1}^n u_i^p \exp(\beta z_i)} \right)^{1/p}$$

2. Plugin  $\hat{\lambda}$  and fix  $\beta$ , we have

$$\begin{aligned} & \frac{\partial \log L}{\partial p} \\ &= \frac{\partial}{\partial p} \left\{ \sum_{i=1}^n \left\{ \delta_i \left\{ \log p + \log \left( \frac{n_{evt}}{\sum_{i=1}^n u_i^p \exp(\beta z_i)} \right) + \log u_i^{p-1} + \beta z_i \right\} - \frac{n_{evt}}{\sum_{i=1}^n u_i^p \exp(\beta z_i)} u_i^p \exp(\beta z_i) \right\} \right\} \\ &= \frac{\partial}{\partial p} \left\{ n_{evt} \log p + n_{evt} \log \left( \frac{n_{evt}}{\sum_{i=1}^n u_i^p \exp(\beta z_i)} \right) + \sum_{i=1}^n \delta_i (\log u_i^{p-1} + \beta z_i) - \frac{n_{evt} \sum_{i=1}^n u_i^p \exp(\beta z_i)}{\sum_{i=1}^n u_i^p \exp(\beta z_i)} \right\} \\ &= \frac{n_{evt}}{p} - n_{evt} \frac{\partial}{\partial p} \log \left( \sum_{i=1}^n u_i^p \exp(\beta z_i) \right) + \sum_{i=1}^n \delta_i \log u_i \\ &= \frac{n_{evt}}{p} - \frac{n_{evt}}{\sum_{i=1}^n u_i^p \exp(\beta z_i)} \cdot \sum_{i=1}^n (\log u_i \cdot u_i^p \exp(\beta z_i)) + \sum_{i=1}^n \delta_i \log u_i \end{aligned}$$

Set this to 0 to get  $\hat{p}$  but **unfortunately** there seems no analytical solution to this.

Now let's assume a more specil case  $p = 1$ , that means the survival function is exponential distribution. Then we will have log-likelihood as

$$\log L = \sum_{i=1}^n \{ \delta_i \log(\lambda e^{\beta z_i}) - \lambda u_i \exp(\beta z_i) \} = n_{evt} \log \lambda + \sum_{i=1}^n \delta_i \beta z_i - \lambda \sum_{i=1}^n u_i \exp(\beta z_i)$$

Again, fix  $\beta$  and take derivative will give us

$$\frac{\partial \log L}{\partial \lambda} = \frac{n_{evt}}{\lambda} - \sum_{i=1}^n u_i \exp(\beta z_i)$$

hence

$$\hat{\lambda} = \frac{n_{evt}}{\sum_{i=1}^n u_i \exp(\beta z_i)}$$

Plug in this value will give us the profile log-likelihood of  $\beta$  as

$$\begin{aligned}
\log L_{profile}(\beta) &= n_{evt} \log \left( \frac{n_{evt}}{\sum_{i=1}^n u_i \exp(\beta z_i)} \right) + \sum_{i=1}^n \delta_i \beta z_i - \frac{n_{evt}}{\sum_{i=1}^n u_i \exp(\beta z_i)} \sum_{i=1}^n u_i \exp(\beta z_i) \\
&\stackrel{\text{w.r.t. } \beta}{\propto} -n_{evt} \log \left( \sum_{i=1}^n u_i \exp(\beta z_i) \right) + \sum_{i=1}^n \delta_i \beta z_i \\
&= \sum_{i=1}^n \delta_i \left( \beta z_i - \log \left( \sum_{i=1}^n u_i \exp(\beta z_i) \right) \right) \\
&= \sum_{i=1}^n \delta_i \log \left( \frac{\exp(\beta z_i)}{\sum_{i=1}^n u_i \exp(\beta z_i)} \right)
\end{aligned}$$

From the previous equation we can see that to maximize  $L_{profile}(\beta)$ , which is equivalent to maximize  $\log L_{profile}(\beta)$ , is equivalent to maximize

$$\begin{aligned}
&\exp \left( \sum_{i=1}^n \delta_i \log \left( \frac{\exp(\beta z_i)}{\sum_{i=1}^n u_i \exp(\beta z_i)} \right) \right) \\
&= \prod_{i=1}^n \left( \frac{\exp(\beta z_i)}{\sum_{i=1}^n u_i \exp(\beta z_i)} \right)^{\delta_i}
\end{aligned}$$

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

### 2.1.2 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  $\beta$ , we maximize  $L$  (or equivalently,  $\log L$ ) over  $h_j$ s so the result is a function of  $\beta$ . 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  $\beta$  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  $L_1(\beta)$ .

### 2.1.3 Intuition: conditional distribution perspective

Given the fact that someone survives up to just prior to  $\tau_j$ , 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  $\tau_j$  given the fact that subject survives past  $\tau_{j-1}$ .

Now, given the risk set  $R_j$  and the fact that the subject with  $z_*$  fails at  $\tau_j$ , 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  $\tau_j$  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  $\tau$ , which is the conditional probability of the subject fails at  $\tau_j$  given the fact that the subject survives past  $\tau_{j-1}$ . Then given the risk set  $R_j$  and the fact that exactly one subject fails at  $\tau_j$ . 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 existance is 0 in the continuous time case, in real life it is pretty common. Let

$\tau_1 < \tau_2 < \dots < \tau_k$	distinct failure times
$d_j$	number of failures at $\tau_j$
$z_{(j)}^{(1)}, z_{(j)}^{(2)}, \dots, z_{(j)}^{(d_j)}$	values of $z$ for the $d_j$ subjects who fail at $\tau_j$
$R_j$	as before

Then the exact, and two approximation of the partial likelihood are shown below.

### 2.3.1 Exact partial likelihood

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

$$\begin{aligned} L_1(\beta) &= \prod_{j=1}^K \left\{ \sum_{(k_1, \dots, k_{d_j})=(1,2,\dots,d_j)} \prod_{i=1}^{d_j} \left\{ \frac{\exp(\beta z_{(j)}^{(k_i)})}{\sum_{l \in R_j} \exp(\beta z_l) - \sum_{s=1}^{i-1} \exp(\beta z_{(j)}^{(k_s)})} \right\} \right\} \\ &= \prod_{j=1}^K \left\{ \left[ \frac{\prod_{i=1}^{d_j} \exp(\beta z_{(j)}^{(i)})}{\sum_{l \in R_j} \exp(\beta z_l)} \right] \cdot \left[ \sum_{(k_1, \dots, k_{d_j})=(1,2,\dots,d_j)} \prod_{i=1}^{d_j} \frac{1}{\sum_{l \in R_j} \exp(\beta z_l) - \sum_{s=1}^{i-1} \exp(\beta z_{(j)}^{(k_s)})} \right] \right\}. \end{aligned} \quad (6)$$

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

$$L_1(\beta) = \prod_{j=1}^K \left\{ \prod_{i=1}^{d_j} \left\{ \frac{\exp(\beta z_{(j)}^{(i)})}{\sum_{l \in R_j} \exp(\beta z_l)} \right\} \right\} = \prod_{j=1}^K \left\{ \frac{\prod_{i=1}^{d_j} \exp(\beta z_{(j)}^{(i)})}{\left( \sum_{l \in R_j} \exp(\beta z_l) \right)^{d_j}} \right\}. \quad (7)$$

The idea is to treat these  $d_j$  subjects separately as that in (3), the **same** risk set is used for each and product the results together.

### 2.3.3 Efron's approximation

$$L_1(\beta) = \prod_{j=1}^K \left\{ \frac{\prod_{i=1}^{d_j} \exp(\beta z_{(j)}^{(i)})}{\prod_{i=1}^{d_j} \left\{ \sum_{l \in R_j} \exp(\beta z_l) - \frac{i-1}{d_j} \sum_{s=1}^{d_j} \exp(\beta z_{(j)}^{(s)}) \right\}} \right\}, \quad (8)$$

which is quite accurate for moderate  $d_j$ .

### 2.3.4 One example for comparison

Let's assume  $R_j = \{1, 2, 3\}$  and death set at  $\tau_j$  is  $D_j = \{1, 2\}$ , which means two tied event at  $\tau_j$ . Then at this time point,

1. the exact partial likelihood, from (6) is

$$\begin{aligned} & \frac{e^{\beta z_1}}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3}} \times \frac{e^{\beta z_2}}{e^{\beta z_2} + e^{\beta z_3}} + \frac{e^{\beta z_2}}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3}} \times \frac{e^{\beta z_1}}{e^{\beta z_1} + e^{\beta z_3}} \\ &= \frac{e^{\beta(z_1+z_2)}}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3}} \times \left( \frac{1}{e^{\beta z_2} + e^{\beta z_3}} + \frac{1}{e^{\beta z_1} + e^{\beta z_3}} \right) \end{aligned}$$

2. the breslow's approximation, from (7) is

$$\begin{aligned} & \frac{e^{\beta(z_1+z_2)}}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3}} \times \frac{1}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3}} \\ &= \frac{e^{\beta(z_1+z_2)}}{(e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3})^2} \end{aligned}$$

3. the efron's approximation, from (8) is

$$\frac{e^{\beta(z_1+z_2)}}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3}} \times \frac{1}{e^{\beta z_1} + e^{\beta z_2} + e^{\beta z_3} - \frac{1}{2}(e^{\beta z_1} + e^{\beta z_2})}$$

So normally speaking, efron's approximation is more accurate than breslow's, but breslow's is more easire to cmpute. In real application, when  $d_j$  is big, it may be appropriate to consider analysis for discrete survival function.

### 2.3.5 A fast algorithm for Breslow's and Efron's approximation

To be added.

## 3 Inference

The idea is to just proceed with partial likelihood as if it is the full likelihood.



- Maximize  $L_1$  over  $\beta$ . The maximization of full likelihood is sometimes called the “semiparametric” MLE. Usually the estimate  $\hat{\beta} = \operatorname{argmax} L_1$  can not be obtained in closed form.
- Approximate the variance of  $\hat{\beta}$  by inverse of observed information from  $L_1$ .
- Use Wald test, score test, LRT as in ordinary ML settings.

Here we consider a scalar value  $z$ . And we can see how it is easy to compute using the breslow’s approximation. It is easy to verify from (7) that

$$U(\beta) = \frac{\partial \log L_1(\beta)}{\partial \beta} = \sum_{j=1}^K \left\{ \sum_{i=1}^{d_j} z_{(j)}^{(i)} - d_j \cdot \sum_{l \in R_j} w_l^{(j)} z_l \right\}$$

$$\hat{I}(\beta) = -\frac{\partial^2 \log L_1(\beta)}{\partial \beta^2} = \text{to be added.},$$

where

$$w_l^{(j)} = \frac{e^{\beta z_l}}{\sum_{m \in R_j} e^{\beta z_m}}.$$

For Efron’s approximation, the results is **to be added**.

**Wald test:** based on

$$\hat{\beta} \stackrel{apx}{\sim} N\left(\beta, \hat{I}^{-1}(\hat{\beta})\right).$$

**Score test:** The null hypothesis is  $H_0 : \beta = 0$ , based on this assuming  $U(0) / \sqrt{I(0)} \stackrel{apx}{\sim} N(0, 1)$ .

## 4 Extensions

### 4.1 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  $\beta_1$  would tell us about the treatment effect and test for  $\beta_2$  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

$$h(t|z, w) = h(t|w) e^{\beta z}, \tag{9}$$

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  $e^\beta$ . For each stratified level, we can construct the partial likelihood as normal, denoted by  $L_1^{(l)}(\beta)$ , then the overall partial likelihood is

$$L_1(\beta) = \prod_{l=1}^L L_1^{(l)}(\beta). \quad (10)$$

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

## 4.2 Cox's model for comparing multiple groups and trend test

### 4.2.1 Comparing multiple groups

Suppose  $Z$  is a categorical variable with  $p+1$  levels:  $0, 1, \dots, p$ . One way to compare the  $p+1$  survival distributions is a logrank test as introduced in previous notes. One we can use a Cox model with dummy variable

$$Z_i = 1(Z = i), \quad i = 1, \dots, p.$$

And testing the hypothesis  $\beta_1 = \beta_2 = \dots = \beta_p = 0$ . This test is equivalent to the logrank test for comparing  $p+1$  groups. So again we can see that logrank test is oriented towards PH alternatives. And logrank test can be adjusted for other covariates by using the Cox's model form.

### 4.2.2 Trend test

Next let's assume the  $p+1$  levels of  $Z$  is ordinal, such as increasing doses of the same drug. We can construct a new scalar variable

$$Z^* = c_i, \quad \text{if } Z = i, \quad i = 0, 1, \dots, p,$$

where  $c_0 = 0$  and  $c_1, \dots, c_p$  are some constants. Then build the Cox's model with  $Z^*$  and test the coefficient  $\beta^* = 0$  is equivalent to the logrank trend test with weight  $c_1, \dots, c_p$ . Again, logrank test is oriented towards PH alternatives.

## 4.3 Time varying covariates

### 4.4 Estimating $H_0(t)$

Most of the time when using Cox's model, we are making inference about coefficient  $\beta$ . But sometimes we would also want to estimate  $H_0(t)$ . Maybe we want to check the shape the baseline hazard, or make prediction about survival probability, etc. There is NO unique correct way to estimate  $H_0(t)$ . One popular method is proposed by Breslow:

$$\widehat{H}_0(t) = \sum_{\tau_j \leq t} \widehat{\Delta H}_0(\tau_j),$$

where

$$\widehat{\Delta H}_0(\tau_j) = \frac{d_j}{\sum_{l \in R_j} e^{\hat{\beta} z_l}}.$$

This is a discrete estimator and when  $\hat{\beta} = 0$ , this reduced to  $d_j/Y(\tau_j)$ , as in the Nelson-Aalen estimator of  $H_0(\cdot)$ .

## References