Chapter 7

(AST405) Lifetime data analysis

Md Rasel Biswas

Lecture Outline

1. 7. Semiparametric Multiplicative Hazards Regression Models
   - 7.1 Methods for continuous multiplicative hazards model
   - 7.2 Comparison of two or more lifetime distributions

Section 1

7. Semiparametric Multiplicative Hazards Regression Models

Subsection 1

7.1 Methods for continuous multiplicative hazards model

7.1 Methods for continuous multiplicative hazards model

- Models in which covariates have a multiplicative effect on the hazard function play an important role in the analysis of lifetime data
- Proportional hazard (PH) model is one of such models
- Depending on whether baseline hazard function is left arbitrary or not, PH model could be either semiparametric or parametric
- In this section, semiparametric PH models are discussed, where baseline hazard function is left arbitrary

7.1 Methods for continuous multiplicative hazards model

- The hazard function is modeled as

$$\begin{aligned} h(t \mid \mathbf{x}) &= h_0(t) \exp(\mathbf{x}'\beta) \\ &= h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p) \end{aligned} \tag{1}$$

  - $h(t|x)$ = hazard at time t for a person with covariates x
  - $h_0(t)$ = **baseline hazard** (unspecified)
  - $\beta = (\beta_1, \ldots, \beta_p)'$ $\rightarrow$ vector of regression coefficients
  - Covariate vector $\mathbf{x}$ could include time-varying covariate
  - No intercept term is included in $\mathbf{x}'\beta$

- Model (Equation 1) is known as "Cox's proportional hazards model" or simply "Cox model"

- No distributional assumption is required for estimating the parameters of the Model (Equation 1)

7.1 Methods for continuous multiplicative hazards model

- The cumulative baseline hazard function is defined as

$$H_0(t) = \int_0^t h_0(u)\, du \tag{2}$$

- The baseline survivor function

$$S_0(t) = \exp\big[-H_0(t)\big] \tag{3}$$

- The survivor function of T given covariate vector $\mathbf{x}$

$$S(t \,|\, \mathbf{x}) = \big[S_0(t)\big]^{\exp(\mathbf{x}'\beta)} \tag{4}$$

Estimation of model parameters

- Data

$$\left\{(t_i, \delta_i, \mathbf{x}_i), i = 1, \ldots, n\right\}$$

- Parameters of interest are $h_0(t)$ and β

Estimation of model parameters

Log-likelihood function

$$\begin{aligned}
\ell(h_0(t), \beta) &= \log \prod_{i=1}^{n} \left[f(t_i; \mathbf{x}_i)\right]^{\delta_i} \left[S(t_i; \mathbf{x}_i)\right]^{1-\delta_i} \\
&= \sum_i \left\{ \delta_i \log \left[h_0(t_i) \exp(\mathbf{x}_i' \beta)\right] + \exp(\mathbf{x}_i' \beta) \log S_0(t_i) \right\} \\
&= \sum_i \left\{ \delta_i \left[\log h_0(t_i) + \mathbf{x}_i' \beta \right] + \exp(\mathbf{x}_i' \beta) \log S_0(t_i) \right\} \qquad (5)
\end{aligned}$$

- No unique solutions of the parameters because the number of parameters to be estimated is greater than the number of observations

Estimation of model parameters

- Complete likelihood function is not useful for estimating parameters of Cox's proportional hazards model
- There are a number of different likelihood functions defined for estimating parameters, of which Cox's "partial likelihood function" is widely used for PH models
- Log-partial-likelihood function is defined as

$$\ell_1(\beta) = \log \prod_{i=1}^{n} \left(\frac{\exp(\mathbf{x}_i'\beta)}{\sum_{k=1}^{n} Y_k(t_i) \exp(\mathbf{x}_k'\beta)} \right)^{\delta_i} \tag{6}$$

  - $Y_k(t) = I(t_k \geq t) \rightarrow$ indicates whether the kth subject is still in the risk set at time t or not

Estimation of model parameters

- Partial likelihood function can be treated as a regular likelihood function for making statistical inference
- For partial likelihood function, the parameters of interest is β and the estimated parameters

$$\hat{\beta} = \arg\max_{\beta \in \Theta} \ell_1(\beta)$$

  follow asymptotically normal distribution, similar to MLEs
- The baseline hazard functions are estimated from the full likelihood function with regression parameters are assumed to be known, i.e. $\ell(h_0(t), \hat{\beta})$

Estimation of model parameters

- Obtain the expression of partial likelihood function for the following censored sample

time	x
3	1
5	0
8	1
4+	1
10	0

Subsection 2

7.2 Comparison of two or more lifetime distributions

7.2 Comparison of two or more lifetime distributions

- Let $S_j(t)$ be the survivor function of lifetime T_j, $j = 1, 2$
- Data available

$$\{(t_i, \delta_i, x_i),\ i = 1, \ldots, n\}$$

  - $x_i = I(i\text{th subject is from group } 1)$

- Null hypothesis

$$H_0 : S_1(t) = S_2(t)$$

7.2 Comparison of two or more lifetime distributions

- Consider PH model

$$h(t \mid x) = h_0(t) \exp(\beta x) \quad \Rightarrow \quad S(t \mid x) = [S_0(t)]^{\exp(\beta x)}$$

- We can obtain

$$S_2(t) = S(t \mid x = 0) = S_0(t)$$

$$S_1(t) = S(t \mid x = 1) = [S_0(t)]^{\exp(\beta)} = [S_2(t)]^{\exp(\beta)}$$

7.2 Comparison of two or more lifetime distributions

- The null hypothesis under proportional model assumption

$$H_0 : S_1(t) = S_2(t) \quad \Rightarrow \quad H_0 : \beta = 0$$

- Large sample-based property of MLE $\hat{\beta}$ can be used to test the null hypothesis

7.2 Comparison of two or more lifetime distributions

- Log-likelihood function

$$\ell(\beta) = \log \prod_{i=1}^{n} \left(\frac{e^{\beta x_i}}{\sum_{k=1}^{n} Y_k(t_i)\, e^{\beta x_k}} \right)^{\delta_i}$$

$$= \sum_{i=1}^{n} \left(\delta_i\, x_i\, \beta - \delta_i \log \sum_{k=1}^{n} Y_k(t_i)\, e^{\beta x_k} \right)$$

7.2 Comparison of two or more lifetime distributions

- Score function

$$U(\beta) = \sum_{i=1}^{n} \left(\delta_i \, x_i - \frac{\delta_i \sum_{k=1}^{n} Y_k(t_i) \, e^{\beta x_k} \, x_k}{\sum_{k=1}^{n} Y_k(t_i) \, e^{\beta x_k}} \right)$$

$$= \sum_{i=1}^{n} \left(d_{1i} - \frac{d_i \, n_{1i} \, e^{\beta}}{n_{1i} \, e^{\beta} + n_{2i}} \right)$$

  - $d_i = \delta_i$
  - $d_{1i} = \delta_i \, x_i = I(ith$ subject from group 1$)$
  - $n_{1i} = \sum_{k=1}^{n} Y_k(t_i) x_k \rightarrow$ number of group 1 subjects at risk at time t_i
  - $n_{2i} = \sum_{k=1}^{n} Y_k(t_i)(1 - x_k) \rightarrow$ number of group 2 subjects at risk at time t_i

7.2 Comparison of two or more lifetime distributions

- Information matrix

$$I(\beta) = -\frac{d_i\, n_{1i}\, e^{\beta} n_{1i}\, e^{\beta} - d_i (n_{1i}\, e^{\beta} + n_{2i}) n_{1i}\, e^{\beta}}{\left(n_{1i}\, e^{\beta} + n_{2i}\right)^2}$$

$$= \frac{d_i\, n_{1i} n_{2i} e^{\beta}}{\left(n_{1i}\, e^{\beta} + n_{2i}\right)^2}$$

7.2 Comparison of two or more lifetime distributions

- Confidence interval for β can be obtained from the following pivotal quantity

$$Z(\beta) = \frac{U(\beta)}{[I(\beta)]^{1/2}}$$

  which follows an asymptotic standard normal distribution

- $100(1-\alpha)\%$ confidence interval for β can be obtained from the set of values of β that satisfy

$$Z(\beta) \leq z_{1-\alpha}$$

7.2 Comparison of two or more lifetime distributions

- Under $H_0 : \beta = 0$

$$U(0) = \sum_{i=1}^{n} \left(d_{1i} - \frac{d_i \, n_{1i}}{n_{1i} + n_{2i}} \right)$$

$$I(0) = \sum_{i=1}^{n} \frac{d_i \, n_{1i} n_{2i}}{\left(n_{1i} + n_{2i} \right)^2}$$

- Test statistic

$$Z = \frac{U(0)}{[I(0)]^{1/2}} \sim \mathcal{N}(0, 1)$$

  - MLE of β does not require to test $H_0 : \beta = 0$ using the statistic Z

7.2 Comparison of two or more lifetime distributions

- The expression of $U(0)$ can be considered as the difference between observed number of deaths from group 1, (d_{1i}), at time t_i and the corresponding expected number of deaths

$$d_i \times \frac{n_{1i}}{n_{1i} + n_{2i}}$$

- At time t_i, there are $n_i = n_{1i} + n_{2i}$ subjects are at risk and d_i is either 0 or 1 (i.e. there is no ties in the lifetime)

group	event	alive	at risk
1	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
2	d_{2i}	$n_{2i} - d_{2i}$	n_{2i}
	d_i	$n_i - d_i$	n_i

7.2 Comparison of two or more lifetime distributions

- This score test for the Cox model to compare two groups is also known as **log-rank** test.

Example 7.1.1

Data below show remission times (in weeks) for 40 leukemia patients who were randomly assigned either treatment A or B

```
tab7_1_1
```

```
# A tibble: 40 x 3
    time status group
   <dbl>  <dbl> <chr>
 1     1      1 A
 2     3      1 A
 3     3      1 A
 4     6      1 A
 5     7      1 A
 6     7      1 A
 7    10      1 A
 8    12      1 A
 9    14      1 A
10    15      1 A
# i 30 more rows
```

Example 7.1.1

```
survdiff(Surv(time, status) ~ group,
         data = tab7_1_1)
```

```
Call:
survdiff(formula = Surv(time, status) ~ group, data = tab7_1_1)

          N Observed Expected (O-E)^2/E (O-E)^2/V
group=A 20       17     21.5     0.951      2.36
group=B 20       20     15.5     1.322      2.36

 Chisq= 2.4  on 1 degrees of freedom, p= 0.1
```

Example 7.1.1

```
coxph(Surv(time, status) ~ group, data = tab7_1_1) %>%
    tidy()
```

```
# A tibble: 1 x 5
  term   estimate std.error statistic p.value
  <chr>     <dbl>     <dbl>     <dbl>   <dbl>
1 groupB    0.503     0.332      1.51   0.130
```

Example 7.2.1

- Patients with *cystic fibrosis* are susceptible to an accumulation of mucus in lungs, which leads to pulmonary exacerbation and deterioration of lung function
- A clinical trial was conducted to investigate the efficacy of the new drug DNase-1
  - Subjects are randomly assigned to a new treatment or a placebo
- Time of interest is the time to first exacerbation after randomization and data on fev (forced expiatory volume at the time of randomization) are also measured

Example 7.2.1

Table 1.4. Times to First Pulmonary Exacerbation for 10 Subjects

t (days)[a]	trt	fev[b]
168*	1	28.8
169*	1	64.0
65	0	67.2
168*	1	57.6
171*	0	57.6
166*	1	25.6
168*	0	86.4
90	0	32.0
169*	1	86.4
8	0	28.8

[a] Starred values are censoring times.
[b] fev measure is percent of predicted normal fev, based on sex, age, and height.

Example 7.2.1

Creating the data from the R object `rhDNase`

```
tab1_4 <- as_tibble(rhDNase) %>%
  filter(is.na(ivstart) | ivstart > 0) %>%
  mutate(time0 = as.numeric(end.dt - entry.dt),
         status = as.numeric(!is.na(ivstart)),
         time = if_else(status == 1, ivstart, time0),
         fevm = fev - mean(fev)) %>%
  group_by(id) %>%
  mutate(visit = n()) %>%
  ungroup()
```

Example 7.2.1

Cox's PH model

$$h(t \mid \text{trt, fevm}) = h_0(t) \exp(\beta_1 \text{ trt} + \beta_2 \text{ fevm})$$

R code for fitting the model

```
mod1 <- coxph(Surv(time,  status) ~ trt + fevm,
              data = tab1_4)
```

Example 7.2.1

Estimates of regression coefficients

```
tidy(mod1)
```

```
# A tibble: 2 x 5
  term  estimate std.error statistic  p.value
  <chr>    <dbl>     <dbl>     <dbl>    <dbl>
1 trt    -0.352    0.106       -3.31 9.47e- 4
2 fevm   -0.0188   0.00226     -8.31 9.63e-17
```

- Treatment group patients have lower hazard for time to first exacerbation
- As FEV value increases the hazard of first exacerbation decreases
- Effects of treatment and FEV are significant on the hazard of first exacerbation decreases

Example 7.2.1

Table 3: Estimates and corresponding confidence intervals of the parameters of Cox's PH model

term	estimate	p.value	HR	2.5 %	97.5 %
trt	-0.352	0.001	0.703	0.571	0.867
fevm	-0.019	0.000	0.981	0.977	0.986

- Treatment group patients have about 30% lower hazard of first exacerbation than that of the placebo group patients provided FEV value remains constant
- For 1-unit increase of FEV value, hazard of first exacerbation decreases about 2% provided treatment group remains constant

Example 7.2.1

`survfit()` provides estimate of survivor function and corresponding standard errors

```
tidy(survfit(mod1)) %>%
    as_tibble()
```

```
# A tibble: 161 x 8
    time n.risk n.event n.censor estimate std.error conf.high conf.l
   <dbl>  <dbl>   <dbl>    <dbl>    <dbl>     <dbl>     <dbl>    <dl
 1     1    761       1        0    0.999   0.00138      1         0.9
 2     5    760       3        0    0.994   0.00277      1.00      0.9
 3     6    757       1        0    0.993   0.00311      0.999     0.9
 4     8    756       4        0    0.988   0.00420      0.996     0.9
 5     9    752       3        0    0.983   0.00489      0.993     0.9
 6    11    749       2        0    0.981   0.00530      0.991     0.9
 7    13    747       2        0    0.978   0.00569      0.989     0.9
 8    14    745       2        0    0.975   0.00606      0.987     0.9
 9    15    743       4        0    0.970   0.00675      0.983     0.9
10    16    739       2        0    0.967   0.00708      0.980     0.9
# i 151 more rows
```