Reliability Data and Survival Analysis
Homework 5

Chenghua Liu
liuch18@mails.tsinghua.edu.cn
Department of Computer Science
Tsinghua University

1

Suppose $\lambda(t \mid Z) = \lambda_0(t) e^{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3}$ where

$$Z_1 = \begin{cases} 0 & \text{tr. } 0 \\ 1 & \text{tr. } 1 \end{cases} \quad Z_2 = \begin{cases} 0 & \text{female} \\ 1 & \text{male} \end{cases}$$
$$\text{and } Z_3 = Z_1 \cdot Z_2$$

What values of $\beta_1, \beta_2, \beta_3$ correspond to

(a) treatment hazard ratio same in males as in females

(b) no treatment effect in males, but an effect in females

(c) no treatment effect

solution:

(a)

$$\frac{\lambda(t \mid Z_1 = 1, Z_2 = 1)}{\lambda(t \mid Z_1 = 0, Z_2 = 1)} = \frac{\lambda(t \mid Z_1 = 1, Z_2 = 0)}{\lambda(t \mid Z_1 = 0, Z_2 = 0)} \Rightarrow \frac{e^{\beta_1 + \beta_2 + \beta_3}}{e^{\beta_2}} = \frac{e^{\beta_1}}{e^0} \Rightarrow \beta_3 = 0$$

(b)

$$\begin{aligned} & \frac{\lambda(t \mid Z_1 = 1, Z_2 = 1)}{\lambda(t \mid Z_1 = 0, Z_2 = 1)} = 1, \quad \frac{\lambda(t \mid Z_1 = 1, Z_2 = 0)}{\lambda(t \mid Z_1 = 0, Z_2 = 0)} \neq 1 \\ & \Rightarrow \frac{e^{\beta_1 + \beta_2 + \beta_3}}{e^{\beta_2}} = 1, \quad \frac{e^{\beta_1}}{e^0} \neq 1 \\ & \Rightarrow \beta_1 \neq 0, \quad \beta_1 + \beta_3 = 0 \end{aligned}$$

(c)

$$
\begin{aligned}
&\frac{\lambda\left(t \mid Z_1=1, Z_2=1\right)}{\lambda\left(t \mid Z_1=0, Z_2=1\right)}=1, \quad \frac{\lambda\left(t \mid Z_1=1, Z_2=0\right)}{\lambda\left(t \mid Z_1=0, Z_2=0\right)}=1 \\
&\Rightarrow \frac{e^{\beta_1+\beta_2+\beta_3}}{e^{\beta_2}}=1, \quad \frac{e^{\beta_1}}{e^0}=1 \\
&\Rightarrow \beta_1=0, \quad \beta_3=0
\end{aligned}
$$

2

We will complete the following questions (Q2-Q6) using the dataset 'anderson'. As we show in the class, the data set consists of survival times (in weeks) on 42 leukemia patients, half of whom receive a new therapy and the other half of whom get a standard therapy. The exposure variable of interest is treatment status ($Rx = 0$ if new treatment, $Rx = 1$ if standard treatment). Two other variables for control are log white blood cell count (i.e., $\log WBC$) and sex. Failure status is defined by the status variable (0 if censored, 1 if failure).

Suppose you want to use an extended Cox model to assess the PH assumption for the Sex variable, by using a heaviside function approach designed to yield a constant hazard ratio for less than 15 weeks of follow-up and a constant hazard ratio for 15 weeks or more of follow-up. State two equivalent alternative extended Cox models that will carry out this approach, one model containing one heaviside function and the other model containing two heaviside functions.

solution:

We define heaviside functions

$$I_1(t) = \mathbf{1}_{\{0 \leq t < 15\}}, \quad I_2(t) = \mathbf{1}_{\{15 \leq t\}}$$

1. Model containing one heaviside function.

$$\lambda(t \mid Z(t)) = \lambda_0(t) \exp\left\{\beta_1(Sex) + \beta_2(\log WBC) + \beta_3(Rx) + \gamma(Sex)I_1(t)\right\}$$

   and we have null hypothesis $H_0 : \gamma = 0$

2. Model containing two heaviside functions.

$$\lambda(t \mid Z(t)) = \lambda_0(t) \exp\left\{\beta_2(\log WBC) + \beta_3(Rx) + \gamma_1(Sex)I_1(t) + \gamma_2(Sex)I_2(t)\right\}$$

   and we have null hypothesis $H_0 : \gamma_1 = \gamma_2$

3

Consider an alternative approach to controlling for Sex using an extended Cox model. Define an interaction term between sex and time that allows for diverging survival curves over time.

(a) Please write down the corresponding extended Cox model, which contains Rx, $log\ WBC$, and Sex as main effects plus the product term $Sex \times time$.

(b) Further, please express the hazard ratio for the effect of Sex adjusted for Rx and $log\ WBC$ at 8 and 16 weeks.

(c) Run the model in R.

solution:

(a)

$$\lambda(t \mid Z(t)) = \lambda_0(t) \exp\left\{\beta_1(Sex) + \beta_2(\log WBC) + \beta_3(Rx) + \gamma(Sex \times t)\right\}$$

(b)

the hazard ratio for the effect of Sex adjusted for Rx and $log\ WBC$ at 8 weeks is $\exp\left(\beta_1 + 8\gamma\right)$

the hazard ratio for the effect of Sex adjusted for Rx and $log\ WBC$ at 16 weeks is $\exp\left(\beta_1 + 16\gamma\right)$

(c)

```
library(survival)
ads<-read.csv("/Users/liuchenghua/Downloads/anderson.txt",sep="")
fit<-coxph(Surv(Time, Status) ~ Sex + logWBC + Rx+tt(Sex),data = ads,tt=function(x,t,...)x*t)
summary(fit)
#####output#####
Call:
coxph(formula = Surv(Time, Status) ~ Sex + logWBC + Rx + tt(Sex),
    data = ads, tt = function(x, t, ...) x * t)

  n= 42, number of events= 30

            coef exp(coef) se(coef)      z Pr(>|z|)
Sex       1.7277    5.6278   0.8966  1.927   0.0540 .
logWBC    1.5352    4.6423   0.3406  4.508 6.56e-06 ***
Rx        1.1864    3.2752   0.4828  2.457   0.0140 *
tt(Sex) -0.1655    0.8474   0.0905 -1.829   0.0674 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        exp(coef) exp(-coef) lower .95 upper .95
Sex        5.6278     0.1777    0.9707    32.627
logWBC     4.6423     0.2154    2.3814     9.050
Rx         3.2752     0.3053    1.2713     8.438
tt(Sex)    0.8474     1.1800    0.7097     1.012

Concordance= 0.874  (se = 0.04 )
Likelihood ratio test= 51.09  on 4 df,   p=2e-10
Wald test            = 36.73  on 4 df,   p=2e-07
Score (logrank) test = 59.05  on 4 df,   p=5e-12
```

4

We have fitted a Cox PH model containing the three predictives Rx, $log\ WBC$, and Sex.

(a) Which of the variables in the model fitted above are time-independent and which are time-dependent?

(b) Based on the results, is the PH assumption satisfied for the model being fit? Explain briefly.

(c) Suppose you want to use an extended Cox model to assess the PH assumption for all three variables in the above model. State the general form of an extended Cox model that will allow for this assessment.

solution:

(a)

```
fit_t<-coxph(Surv(Time, Status) ~ Rx + logWBC + Sex + tt(Rx) + tt(logWBC) + tt(Sex),
                    ads, tt = function(x, t, ...)x*t)
summary(fit_t)
#####output#####
Call:
coxph(formula = Surv(Time, Status) ~ Rx + logWBC + Sex + tt(Rx) +
    tt(logWBC) + tt(Sex), data = ads, tt = function(x, t, ...) x *
    t)

  n= 42, number of events= 30

                coef exp(coef) se(coef)      z Pr(>|z|)
Rx           1.63665   5.13795  0.92942  1.761   0.0782 .
logWBC       1.43612   4.20435  0.58631  2.449   0.0143 *
Sex          1.88721   6.60095  0.95712  1.972   0.0486 *
tt(Rx)      -0.05103   0.95025  0.09121 -0.559   0.5758
tt(logWBC)   0.01205   1.01212  0.06572  0.183   0.8545
tt(Sex)     -0.19073   0.82636  0.10487 -1.819   0.0689 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
Rx             5.1379     0.1946    0.8311    31.762
logWBC         4.2044     0.2378    1.3324    13.267
Sex            6.6010     0.1515    1.0114    43.083
tt(Rx)         0.9503     1.0524    0.7947     1.136
tt(logWBC)     1.0121     0.9880    0.8898     1.151
tt(Sex)        0.8264     1.2101    0.6728     1.015

Concordance= 0.87  (se = 0.041 )
Likelihood ratio test= 51.46  on 6 df,   p=2e-09
Wald test            = 36.52  on 6 df,   p=2e-06
Score (logrank) test = 59.14  on 6 df,   p=7e-11
```

By checking p-value (row 16 17 18),we think that Sex is time-dependent , Rx and $logWBC$ is time-independent.

(b)

No, because Sex is time-dependent.

(c)

h_i is function of t, the model:

$$\begin{aligned}\lambda(t \mid Z(t)) =& \lambda_0(t)\exp\{\beta_1(Sex) + \beta_2(\log WBC) + \beta_3(Rx) \\ &+ \gamma_1(Sex)h_1(t) + \gamma_2(logWBC)h_2(t) + \gamma_3(lRx)h_3(t)\}\end{aligned}$$

and we have null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0$

5

Compare between the extended Cox models and a stratified Cox procedure that stratifies on the Sex variable but keeps Rx and log WBC in the model.

solution:

Extended Cox:

3(c)

Stratified Cox:

```
fit_stf <- coxph(Surv(Time, Status) ~ strata(Sex) + logWBC + Rx,data = ads)
summary(fit_stf)
#####output#####
Call:
coxph(formula = Surv(Time, Status) ~ strata(Sex) + logWBC + Rx,
    data = ads)

  n= 42, number of events= 30

          coef exp(coef) se(coef)     z Pr(>|z|)
logWBC 1.4537    4.2787   0.3441 4.225 2.39e-05 ***
Rx     0.9981    2.7131   0.4736 2.108   0.0351 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

       exp(coef) exp(-coef) lower .95 upper .95
logWBC     4.279     0.2337     2.180     8.398
Rx         2.713     0.3686     1.072     6.864

Concordance= 0.812  (se = 0.059 )
Likelihood ratio test= 32.06  on 2 df,   p=1e-07
Wald test            = 22.75  on 2 df,   p=1e-05
Score (logrank) test = 30.8  on 2 df,   p=2e-07
```

Compared output of two models ,we can found that both $logWBC$ and Rx are similar, and $coef$ in stratified Cox model are smaller.

6

The following table gives a small data set of survival times and a covariate z :

patient id	survival time (in years)	z
1	6^+	2
2	5	3
3	7^+	2
4	9^+	1
5	8	4
6	3	2

where + means a right censored observation. Assuming a proportional hazards model

$$\lambda(t \mid z) = \lambda_0(t)\mathrm{e}^{z\beta}$$

(a) Write down the partial likelihood of β.

(b) Plot the log partial likelihood of β in $[-2, 2]$, and convince yourself that this function is concave (Hint: using the griding method to make the plot).

(c) Find $\hat{\beta}$ that maximize this log partial likelihood function, calculate the second derivative of the log partial likelihood function at $\hat{\beta}$.

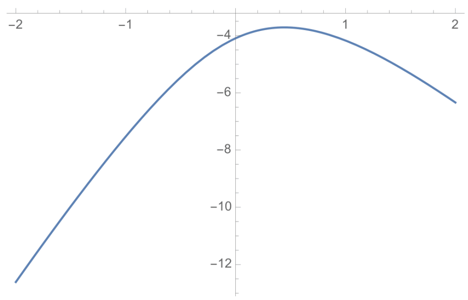
solution:

(a)

$$PL(\beta) = \frac{e^{2\beta}}{e^{2\beta}+e^{3\beta}+e^{2\beta}+e^{2\beta}+e^{4\beta}+e^{\beta}} \times \frac{e^{3\beta}}{e^{3\beta}+e^{2\beta}+e^{2\beta}+e^{4\beta}+e^{\beta}} \times \frac{e^{4\beta}}{e^{4\beta}+e^{\beta}}$$

(b)

We plot the log partial likelihood of β in $[-2, 2]$ by mathematica.Intuitively, this is indeed a concave function.Moreover,through the concave-preserving transformation, we can also quickly determine that this is a concave function.



(c) From the figure we can see that the maximum value is between 0 and 1.We get the value $\beta = 0.443395$ when PL is maximum -3.71412 by mathematica. And the second derivative of the log partial likelihood function at $\hat{\beta}$ is -3.51626 .