Lecture14-Handout

Elizabeth Colantuoni

5/11/2021

I. Objectives:

Upon completion of this session, you will be able to do the following:

- Understand and explain the motivation for modeling the hazard/survival rather than density/distribution functions for censored time-to-event data
- Describe and implement the Cox proportional hazards regression model for time-to-event data
- Understand the idea behind a profile likelihood

II. Survival outcomes

A. Survival outcome Definition

Let D be the time to an event of interest and Let C be the time to censoring, $D > 0$ and $C > 0$.

Define δ as the indicator that the event occurred ($\delta = 0$ if the event was censored).

Then, we get to observe $T_i = min(D_i, C_i)$ and δ_i for each subject i.

We assume D and C are independent and (T_i, δ_i) is independent of (T_j, δ_j) for all i and j.

B. Key quantities and relationships

We can define a series of quantities that can be used to describe the distribution of T; these quantities relate to one another!

Function	$F(t)$ Distribution	$S(t)$ Survival	$f(t)$ Density	$h(t)$ Hazard
Definition	$Pr(T \leq t)$	$Pr(T > t)$	$\lim_{dt \to 0} \frac{Pr(t < T \leq t + dt)}{dt}$	$\lim_{dt \to 0} \frac{Pr(t < T \leq t + dt \mid T > t)}{dt}$
Relationship to: $F(t)$		$1 - F(t)$	$\frac{d}{dt} F(T)$	$\frac{d}{dt} log\left(1 - F(T)\right)$
$h(t)$	$1 - exp\left(-\int_0^t h(u)du\right)$	$exp\left(-\int_0^t h(u)du\right)$	$h(t)exp\left(-\int_0^t h(u)du\right)$	
Common Models				
Exponential	$1 - e^{-\lambda t}$	$e^{-\lambda t}$	$\lambda e^{-\lambda t}$	λ
Weibull	$1 - e^{-(\lambda t)^p}$	$e^{-(\lambda t)^p}$	$p\lambda t^{p-1} e^{-(\lambda t)^p}$	$p\lambda^p t^{p-1)}$
Gamma	num.integration	num.integration	$\frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}$	num.integration

C. Targets for inference

We are primarily interested in making inference about:

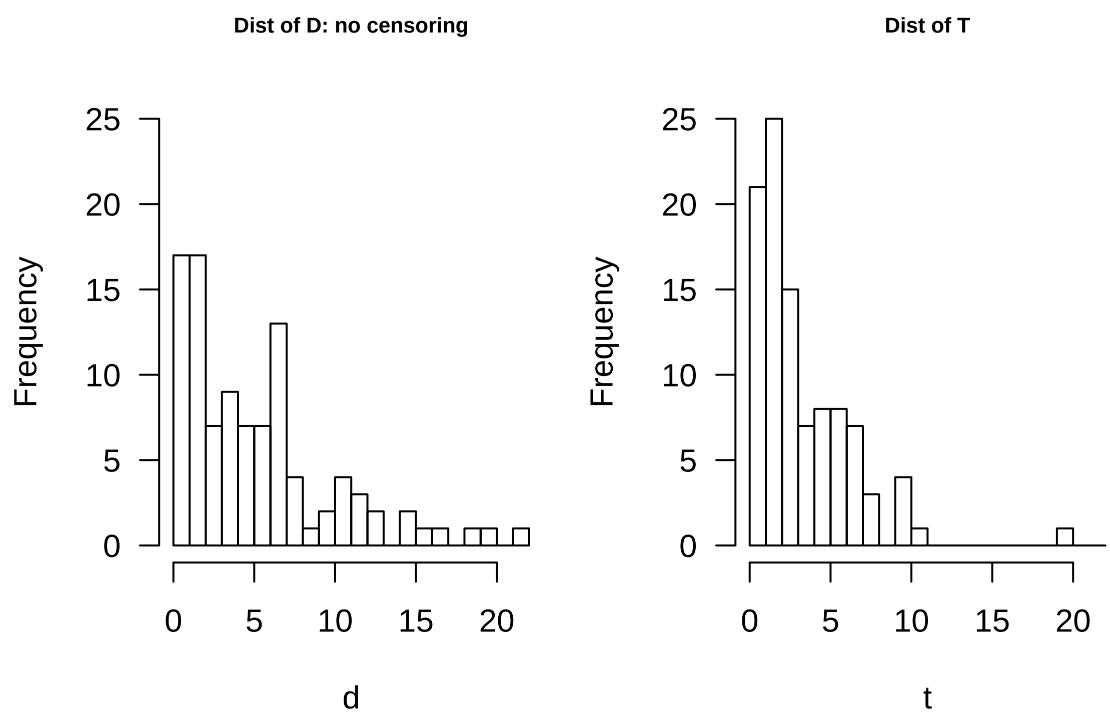
- $S(t) = Pr(T > t)$
- $h(t) = \lim_{dt \to 0} \frac{Pr(t < T \leq t + dt | T > t)}{dt} = \frac{f(t)}{S(t)} = -\frac{d}{dt} log S(t)$

D. Why do we target $S(t)$ and $h(t)$?

Why in survival analysis do we focus on making inference about $S(t)$ and $h(t)$ compared to $f(t)$ and $F(t)$?

The main reason is that censoring complicates our ability to view $f(t)$ since the times we observe are truncated for individuals who experience censoring. See the figure below for an example where $D \sim Exp(0.2)$ and $C \sim Exp(0.1)$. We have plotted the histogram of D, the approximation of $f(t)$ and the histogram for T. The distribution of T is shifted to the left as expected as individuals are censored prior to observing D.

```
set.seed(7361)
d = rexp(100,rate=0.2)
c = rexp(100,rate=0.1)
t = apply(cbind(d,c),1,min)
par(mfrow=c(1,2),mar=c(4,4,4,1))
hist(d,breaks=seq(0,22),freq=TRUE,main="Dist of D: no censoring",cex.main=0.65,cex=0.5,ylim=c(0,25),las=
hist(t,breaks=seq(0,22),freq=TRUE,main="Dist of T",cex.main=0.65,cex=0.5,ylim=c(0,25),las=1)
```



Similar arguments hold with respect to directly modeling $F(t) = Pr(T \leq t)$ since for individuals who are censored we don't know if the event occurs prior to t or not.

E. Likelihood of the data

We have for each subject: $(T_i, \delta_i), X_i$ for $i = 1, ..., n$

We can express the likelihood as:

$$Pr(T_i, \delta_i | X_i, \theta) = \prod_{i=1}^{n} \left[f(t_i|\delta_i, X_i)^{\delta_i} \left(1 - F(t_i|\delta_i, X_i)\right)^{1-\delta_i} \right] Pr(\delta_i | X_i)$$

III. Proportional Hazards Model

The proportional hazards model was proposed by Sir David Cox, 1972 (JRSS-B). This model allows for the regression analysis of survival endpoints under the assumption of proportional hazards. NOTE: There are tweaks to this model that allow for non-proportional hazards assumptions that we won't have time to cover but that you can explore/discover on your own.

The model is:

$$\lambda(t|X) = \lambda_0(t) e^{X\beta}$$

$$log(\lambda(t|X)) = log(\lambda_0(t)) + X\beta$$

where

- $X = (X_1, X_2, ..., X_p)$, no intercept!
- $log(\lambda_0(t))$ is the "baseline hazard" and is the intercept which depends on t
- $\beta_j = log\left(\frac{\lambda(t|X_1,...,X_j=x_j+1,...,X_p)}{\lambda(t|X_1,...,X_j=x_j,...,X_p)}\right)$, the log relative hazard.

A. Checking the proportional hazards assumption

Recall that:

$$\begin{aligned} S(t|X) &= exp\left(-\int_0^t \lambda_0(u) e^{X\beta} du\right) \\ &= exp\left(-e^{X\beta} H_0(t)\right), \ H_0(t) \text{ is the baseline cumulative hazard} \end{aligned}$$

$$log(-log(S(t|X))) = log(H_0(t)) + X\beta$$

So plotting the $log(-log(S(t|X = x)))$ vs. $log(t)$ for values of x will allow you to visually inspect the proportional hazards assumption. If $log(-log(S(t|X = x)))$ for different values of x are parallel, then this supports the proportional hazards assumption.

B. Estimation of β with arbitrary $\lambda_0(t)$

The estimation procedures maximizes the partial likelihood function for event times $t_1 < t_2 < ... < t_n$ with risk sets (i.e. subjects who are still at risk of experiencing the events) $R_1 \supset R_2 \supset .... \supset R_n$.

$$\begin{aligned}
L(\beta) &= \prod_{i=1}^{n} Pr(\text{person } i \text{ has the event at } t_i | 1 \text{ person in risk set } R_i \text{ has the event}) \\
&= \prod_{i=1}^{n} \left[\frac{\lambda_0(t) e^{X_i^{'}\beta}}{\sum_{j \in R_i} \lambda_0(t) e^{X_j^{'}\beta}} \right] \\
&= \prod_{i=1}^{n} \left[\frac{e^{X_i^{'}\beta}}{\sum_{j \in R_i} e^{X_j^{'}\beta}} \right]
\end{aligned}$$

1. Note on the partial likelihood

The Cox model partial likelihood function is a profile likelihood.

What is a profile likelihood? And how does it work?

Assume you have the following model: $Pr(Y = y|\theta = (\theta_1, \theta_2))$ where θ_1 is of interest, and θ_2 is a nuisance (it is an unknown but you are not interested in making inference about it).

- You observe $y_1, ..., y_n$ and the likelihood function is: $L(y|\theta) = \prod_{i=1}^{n} f(y_i|\theta)$.
- Define $\hat{\theta}_2(\theta_1, y)$ to be the value for $\hat{\theta}_2$ that maximizes the likelihood (solves the score equation) when θ_1 is fixed.
- The profile likelihood is then defined as $PL(y|\theta) = \prod_{i=1}^{n} f(y_i|\theta_1, \hat{\theta}_2)$.
- If $\hat{\theta}_1$ maximizes the profile likelihood, then it is the maximum likelihood estimate.

C. Example

Going back to the example of time to death from hospitalization among a group of persons hospitalized for a severe mental disorder.

We will consider two Cox Proportional Hazards models:

- Model A: $log(\lambda(t|male)) = log(\lambda_0(t)) + \beta_1 male$
- Model B: $log(\lambda(t|male, age)) = log(\lambda_0(t)) + \beta_1 male + \beta_2 age$

```
library(survival)
```

```
## Warning: package 'survival' was built under R version 3.6.3
```

```
d = read.table("./survival.csv",sep=",",header=T)
d$event = 1 - d$censor

fitA = coxph(Surv(survive,event)~male,data=d)
summary(fitA)
```

```
## Call:
## coxph(formula = Surv(survive, event) ~ male, data = d)
##
##   n= 26, number of events= 14
##
##         coef exp(coef) se(coef)      z Pr(>|z|)
## male -0.7511    0.4718   0.6055 -1.241    0.215
##
##      exp(coef) exp(-coef) lower .95 upper .95
## male    0.4718      2.119     0.144     1.546
##
## Concordance= 0.628  (se = 0.062 )
## Likelihood ratio test= 1.66  on 1 df,   p=0.2
## Wald test            = 1.54  on 1 df,   p=0.2
## Score (logrank) test = 1.61  on 1 df,   p=0.2
```

We estimate that at any point during the hospitalization, the hazard of death for males is 0.47 times the hazard of death for females (95% CI for the hazard ratio: 0.14 to 1.55).

```r
fitB = coxph(Surv(survive,event)~male+age,data=d)
summary(fitB)
```

```
## Call:
## coxph(formula = Surv(survive, event) ~ male + age, data = d)
##
##   n= 26, number of events= 14
##
##         coef exp(coef) se(coef)     z Pr(>|z|)
## male 0.52374   1.68833  0.73753 0.710  0.47762
## age  0.20753   1.23063  0.05828 3.561  0.00037 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## male     1.688     0.5923    0.3978     7.165
## age      1.231     0.8126    1.0978     1.380
##
## Concordance= 0.816  (se = 0.081 )
## Likelihood ratio test= 20.91  on 2 df,   p=3e-05
## Wald test            = 14.3  on 2 df,   p=8e-04
## Score (logrank) test = 21.27  on 2 df,   p=2e-05
```

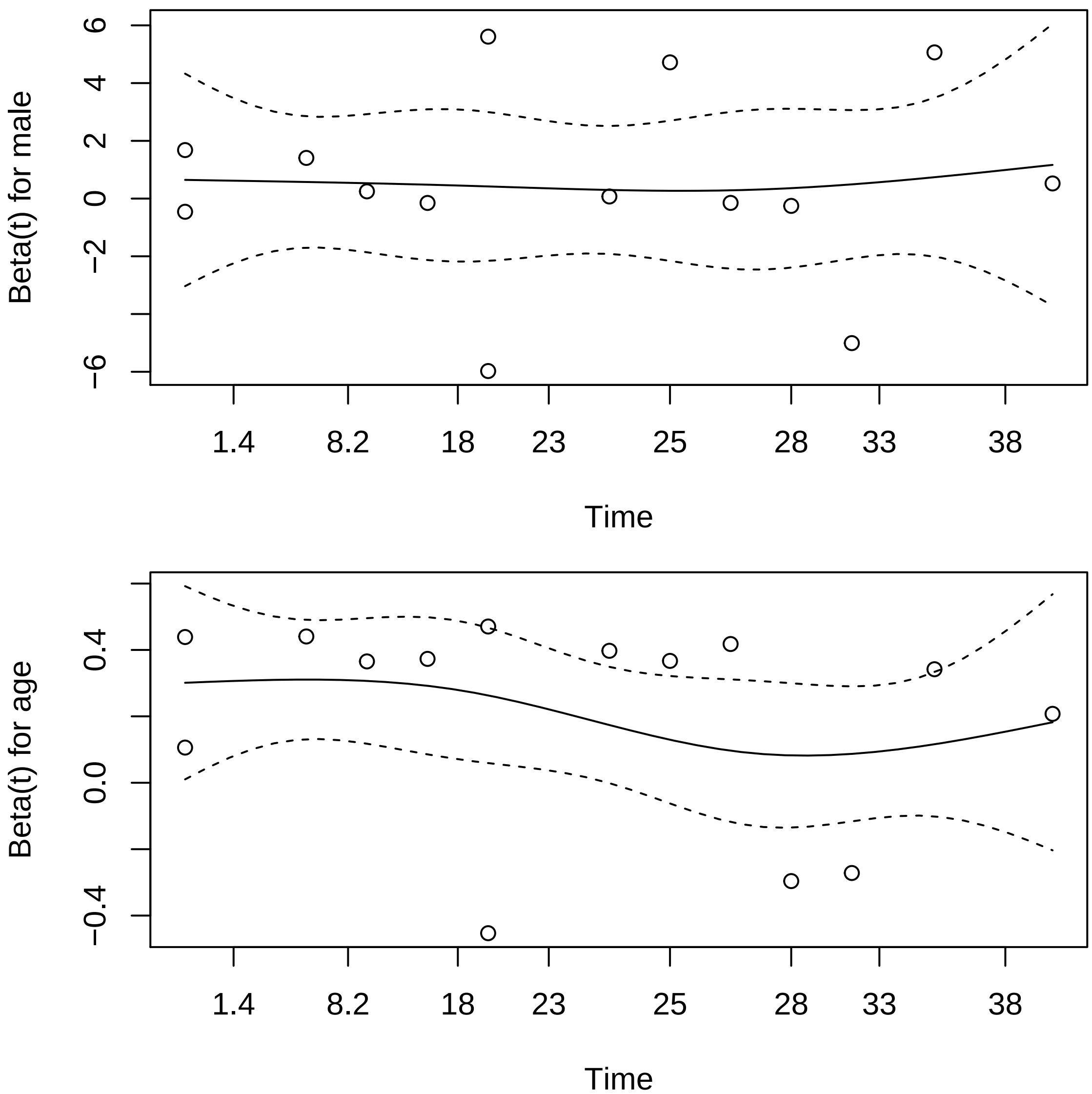
We estimate that at any time during the hospitalization, the hazard of death for males is 69% greater than that of females who are of the same age (95% CI for the relative hazard: 0.40 to 7.12).

We estimate that at any time during the hospitalization, patients who are the same gender but whom are older by 1 year of age have an increased hazard of death by 23% (95% CI for the relative hazard: 1.10 to 1.38).

We can also evaluate the proportional hazards assumption by assessing the Schoenfeld residuals which can be computed separately for each exposure variable in the model. If the mean of the Schoenfeld residuals are 0 for all time, then the proportional hazards assumption holds.

In the figure below, the unique event times have been transformed so that the values plotted along the x-axis are the estimates of $S(t)$ (Kaplan-Meier estimator) at the unique event times. The unique event times label where the survival function changes.

```
temp <- cox.zph(fitB)
par(mfrow=c(2,1),mar=c(4,4,1,1))
plot(temp)
```



There is evidence that the age at hospitalization may be violating the proportional hazards assumption. Does this make sense to you?

There are ways to account for non-proportional hazards, e.g. estimate a time specific effect of age. Here is one vignette that is a good starting place: https://cran.r-project.org/web/packages/Greg/vignettes/timeSplitter.html.

Alternatively, we could label the x-axis using the unique event times. See below.

```
temp <- cox.zph(fitB,transform="identity")
par(mfrow=c(2,1),mar=c(4,4,1,1))
plot(temp)
```