Proportional Hazards Regression[1]

STA312 Spring 2019

[1] See last slide for copyright information.

Background Reading

Chapter 5 in *Applied Survival Analysis Using R* by Dirk Moore

Overview

1. Model
2. Estimation

Proportional Hazards

- Suppose two individuals have different $\mathbf{x}$ vectors of explanatory variable values.
- They have different hazard functions because their λ values are different.
- But the *hazard ratio* $\frac{h_1(t)}{h_2(t)}$ does not depend on time t.
- Exponential regression and Weibull regression fit this pattern.
- Proportional hazards regression is a generalization.

Proportional Hazards Regression

Also called Cox regression after Sir David Cox

Write the hazard function

$$\begin{aligned} h_i(t|\boldsymbol{\beta}) &= h_0(t)\,\psi_i(\boldsymbol{\beta}) \\ &= h_0(t)\,e^{\mathbf{x}_i^\top \boldsymbol{\beta}} \end{aligned}$$

- $h_0(t)$ is called the *baseline hazard function.*
- Baseline because it's the hazard function when $\psi(\boldsymbol{\beta}) = 1$.
- Maybe the patient is in the reference category, and the quantitative explanatory variables are centered.
- In theory $\psi(\boldsymbol{\beta})$ could be almost anything as long as the resulting hazard function is positive.
- But in practice it's almost always $e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$, Cox's original suggestion.

Exponential and Weibull Regression

$h_i(t|\boldsymbol{\beta}) = h_0(t)\,\psi_i(\boldsymbol{\beta}) = h_0(t)\,e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$

- Exponential regression: $h_i(t|\boldsymbol{\beta}) = \lambda = e^{-\mathbf{x}_i^\top \boldsymbol{\beta}}$
  - $h_0(t) = 1$
  - $\psi_i(\boldsymbol{\beta}) = e^{-\mathbf{x}_i^\top \boldsymbol{\beta}}$
- Weibull regression: $h_i(t|\boldsymbol{\beta}) = \frac{1}{\sigma} \exp\{-\frac{1}{\sigma}\mathbf{x}_i^\top \boldsymbol{\beta}\} t^{\frac{1}{\sigma}-1}$
  - $h_0(t) = \frac{1}{\sigma} t^{\frac{1}{\sigma}-1}$
  - $\psi_i(\boldsymbol{\beta}) = \exp\{-\frac{1}{\sigma}\mathbf{x}_i^\top \boldsymbol{\beta}\}$
- Are these really special cases of the proportional hazards model, with $\psi_i(\boldsymbol{\beta}) = e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$?
- Yes, by a re-parameterization. β_j of proportional hazards $= -\beta_j$ of exponential regression.
- β_j of proportional hazards $= -\beta_j/\sigma$ of Weibull regression.
- The main implication is that in proportional hazards regression, the coefficients mean the opposite of what you are used to.
- Anything that makes $\mathbf{x}_i^\top \boldsymbol{\beta}$ bigger will increase the hazard, and make the chances of survival *smaller*.

The Hazard Ratio

Form a ratio of hazard functions. In the numerator, increase $x_{i,k}$ by one unit while holding all other $x_{i,j}$ values constant.

$$\begin{aligned}\frac{h_1(t)}{h_2(t)} &= \frac{h_0(t)\exp\{\beta_0+\beta_1 x_{i,1}+\cdots+\beta_k(x_{i,k}+1)+\cdots+\beta_{p-1}x_{i,p-1}\}}{h_0(t)\exp\{\beta_0+\beta_1 x_{i,1}+\cdots+\beta_k x_{i,k}+\cdots+\beta_{p-1}x_{i,p-1}\}}\\ &= e^{\beta_k}\end{aligned}$$

- Holding the other $x_{i,j}$ values constant is the meaning of "controlling" for explanatory variables.
- If $\beta_k > 0$, increasing $x_{i,k}$ increases the hazard.
- If $\beta_k < 0$, increasing $x_{i,k}$ decreases the hazard.

Semi-parametric

$h_i(t|\boldsymbol{\beta}) = h_0(t)\, e^{\mathbf{x}_i^\top \boldsymbol{\beta}}$

- The unknown quantities in the model are the vector of regression parameters $\boldsymbol{\beta}$, and the unknown baseline hazard function $h_0(t)$.
- We can avoid making any assumptions about $h_0(t)$.
- But because of $\boldsymbol{\beta}$, it's at least partly parametric.

Estimation: Using Ideas From Kaplan-Meier

- As in the Kaplan-Meier estimate, we focus on the uncensored observations, for which the failure time is known.
- The censored observations will have their influence by disappearing from the set of individuals at risk.
- There are $D = \sum_{i=1}^{n} \delta_i$ uncensored observations.
- Denote the ordered times at which failures occur by $t_1, \ldots t_D$.
- This notation can be confusing, because the entire set of times, including censoring times, is usually denoted $t_1, \ldots t_D$.
- Some books (for example Chapter 3 in *Applied Survival Analysis* by Hosmer and Lemeshow, available from https://b-ok.org) use the notation $t_{(1)}, \ldots t_{(D)}$.
- The index set of individuals at risk at failure time t_j is R_j.
- One of them fails.

Hazard

- The hazard function $h(t_j) = \lim_{\Delta \to 0} \frac{P(t_j \leq T \leq t_j + \Delta | T \geq t_j)}{\Delta}$ is roughly proportional to the probability of failure at time t_j, conditionally on survival to that point.
- Make it an actual probability. Normalize it, dividing by the total hazards of all the individuals at risk:

$$q_i = 1 - p_i = \frac{h_0(t) e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{\sum_{j \in R_i} h_0(t) e^{\mathbf{x}_j^\top \boldsymbol{\beta}}} = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{\sum_{j \in R_i} e^{\mathbf{x}_j^\top \boldsymbol{\beta}}}$$

- First, notice that the baseline hazard cancels.
- These really are like the p_i and q_i in Kaplan-Meier estimation.
- Except, instead of dividing by the *number* of individuals at risk, they are weighted by their hazards.
- And those hazards depend on the explanatory variable values through $\boldsymbol{\beta}$.

Estimating $\boldsymbol{\beta}$

Now we have failure probabilities $q_i = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{\sum_{j \in R_i} e^{\mathbf{x}_j^\top \boldsymbol{\beta}}}$.

How can these be used to estimate $\boldsymbol{\beta}$? Cox suggested

- Multiply them together and treat them as a likelihood.
- Take the minus log, and minimize.
- He suggested that all the usual likelihood theory should hold.
- Fisher information, asymptotic normality, likelihood ratio tests: everything.
- He called it *partial* likelihood.
- **Why?!**

Partial Likelihood

Using $h(t) = \frac{f(t)}{S(t)}$,

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f(t_i|\theta)^{\delta_i}\, S(t_i|\theta)^{1-\delta_i} \\
&= \prod_{i=1}^{n} (h(t_i|\theta)S(t_i|\theta))^{\delta_i}\;\, S(t_i|\theta)^{1-\delta_i} \\
&= \prod_{i=1}^{n} h(t_i|\theta)^{\delta_i}\, S(t_i|\theta)^{\delta_i+1-\delta_i} \\
&= \prod_{i=1}^{n} h(t_i|\theta)^{\delta_i}\, S(t_i|\theta) \\
&= \prod_{i=1}^{D} h(t_{(i)}|\theta) \prod_{i=1}^{n} S(t_i|\theta)
\end{aligned}
$$

Continuing the likelihood calculation

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{D} h(t_{(i)}|\theta) \prod_{i=1}^{n} S(t_i|\theta) \\
&= \prod_{i=1}^{D} h_0(t_{(i)}) e^{\mathbf{x}_{(i)}^{\top}\boldsymbol{\beta}} \prod_{i=1}^{n} S(t_i|\boldsymbol{\beta}, h_0) \\
&= \frac{\prod_{i=1}^{D} h_0(t_{(i)}) e^{\mathbf{x}_{(i)}^{\top}\boldsymbol{\beta}}}{\prod_{i=1}^{D} \sum_{j \in R_{(i)}} h_0(t_{(i)}) e^{\mathbf{x}_{j}^{\top}\boldsymbol{\beta}}} \left(\prod_{i=1}^{D} \sum_{j \in R_{(i)}} h_0(t_{(i)}) e^{\mathbf{x}_{j}^{\top}\boldsymbol{\beta}} \right) \prod_{i=1}^{n} S(t_i|\boldsymbol{\beta}, h_0) \\
&= \prod_{i=1}^{D} \frac{h_0(t_{(i)}) e^{\mathbf{x}_{(i)}^{\top}\boldsymbol{\beta}}}{\sum_{j \in R_{(i)}} h_0(t_{(i)}) e^{\mathbf{x}_{j}^{\top}\boldsymbol{\beta}}} \left(\prod_{i=1}^{D} \sum_{j \in R_{(i)}} h_0(t_{(i)}) e^{\mathbf{x}_{j}^{\top}\boldsymbol{\beta}} \right) \prod_{i=1}^{n} S(t_i|\boldsymbol{\beta}, h_0)
\end{aligned}
$$

Partial Likelihood

$$L(\boldsymbol{\beta}, h_0) = \prod_{i=1}^{D} \left(\frac{e^{\mathbf{x}_{(i)}^\top \boldsymbol{\beta}}}{\sum_{j \in R_{(i)}} e^{\mathbf{x}_j^\top \boldsymbol{\beta}}} \right) \left(\prod_{i=1}^{D} \sum_{j \in R_{(i)}} h_0(t_{(i)}) e^{\mathbf{x}_j^\top \boldsymbol{\beta}} \right) \prod_{i=1}^{n} S(t_i | \boldsymbol{\beta}, h_0)$$

- The red product is Cox's partial likelihood.
- Properties similar to ordinary likelihood were proved years later.
- There are fairly convincing arguments that the black product is negligible for large samples.
- Lack of dependence on the baseline hazard is a good feature.
- This is the state of the art.

Hypothesis Tests

As Cox hypothesized, all the usual likelihood theory applies to partial likelihood.

- Consistency (i.e., large-sample accuracy)
- Asymptotic normality.
- Fisher information
- Z-tests
- Wald tests
- Score tests
- Likelihood ratio tests
- Call them *partial* likelihood ratio tests.

Estimating the Survival Function: Background

Using $H(t) = \int_0^t h(x)\,dx$ and $S(t) = e^{-H(t)}$

- Proportional hazards says $h(t|\boldsymbol{\beta}) = h_0(t)\,e^{\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}}$
- This makes it clear that $h_0(t)\,e^{\beta_0}$ cancels in numerator and denominator of the partial likelihood.
- $h_0(t)$ is the hazard function when all explanatory variable values are zero *and* $\beta_0 = 0$.
- $H_0(t) = \int_0^t h_0(x)\,dx$ is the baseline cumulative hazard function.
- $S_0(t) = e^{-H_0(t)}$ is the baseline survival function.
- With a little work we can show $S(t) = S_0(t)^{\exp\{\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}\}}$.
- This could be written $S(t|\mathbf{x}_i)$.

Estimation (Cox and Oakes, 1982, p. 108)

Using $S_0(t) = e^{-H_0(t)}$ and $S(t|\mathbf{x}_i) = S_0(t)^{\exp\{\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}\}}$

Cox suggested $H_0(t) \approx \sum_{t_{(i)}<t} \frac{d_{(i)}}{\sum_{j \in R_{(i)}} e^{\beta_0 + \mathbf{x}_j^\top \boldsymbol{\beta}}}$. Multiplying both sides by e^{β_0}, which is invisible in Cox's argument, arrive at

$$e^{\widehat{\beta}_0} \widehat{H}_0(t) = \sum_{t_{(i)}<t} \frac{d_{(i)}}{\sum_{j \in R_{(i)}} e^{\mathbf{x}_j^\top \widehat{\boldsymbol{\beta}}}}$$

Then, $e^{-\widehat{H}_0(t) e^{\widehat{\beta}_0}} = \widehat{S}_0(t)^{e^{\widehat{\beta}_0}}$. Raise that to the power $\mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}$, and get

$$\widehat{S}_0(t)^{e^{\widehat{\beta}_0 + \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}}} = \widehat{S}(t|\mathbf{x}_i)$$

It works

- As usual, later work clarified matters and eliminated most of the guesswork.
- Cox's estimate of $S(t)$ is show to arise from Breslow's method of approximating the partial likelihood when there are ties.
- There are several other estimates, all yielding results that are pretty close.
- In every case, β_0 is there, but usually it's invisible.

Copyright Information