Cox Proportional Hazard Model

Estimation

Introduction

- Cox proportional hazard model
$$h(t|\boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{Z}^T\boldsymbol{\beta})$$
  - The coefficients β quantify the effects of covariates on hazard function.
- Therefore $\boldsymbol{\beta}$ is the focus of our inference, while $h_0(t)$, the baseline hazard, is of less interest. Thus it is treated as "nuisance parameter".
- In the following, we will focus on
  - Estimate $\boldsymbol{\beta}$ and derive its statistical properties
  - Hypothesis testing for $\boldsymbol{\beta}$

Parameter estimation

- Cox (1972) constructed a partial likelihood to estimate $\boldsymbol{\beta}$.
- Suppose there is no ties in the event times.
- Let $t_1 < t_2 < \cdots < t_D$ be the ordered event times.
- At each event time t_i, define
  - $Z_{(i)}$ denote the covariate vector associated with the individual whose failure time i t_i.
  - R_i be the risk set at time t_i with $R_i = \{j\colon\ t_j \geq t_i\}$.

Partial likelihood

- Cox (1972) proposed the partial likelihood

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{D} \frac{\exp\left(\boldsymbol{Z}_{(i)}^T \boldsymbol{\beta}\right)}{\sum_{j\in R_i} \exp\left(\boldsymbol{Z}_j^T \boldsymbol{\beta}\right)}.$$

  - It only involves $\boldsymbol{\beta}$,not the baseline hazard $h_0(t)$.
  - Does not use information on the actual censoring and event times.
  - Only need the rankings to determine the at-risk sets $\{R_i\}_{i=1}^n$.

Partial likelihood

- The partial likelihood is treated as usual likelihood. Thus estimation and inference of $\boldsymbol{\beta}$ can be carried out by the usual means.
- The maximum partial likelihood estimator (MPLE)

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\beta} PL(\boldsymbol{\beta}) = \arg\max_{\beta} \prod_{i=1}^{D} \frac{\exp\left(\boldsymbol{Z}_{(i)}^T \boldsymbol{\beta}\right)}{\sum_{j\in R_{(t_i)}} \exp\left(\boldsymbol{Z}_j^T \boldsymbol{\beta}\right)}.$$

- The likelihood based testing methods can also be used to make inference about $\widehat{\boldsymbol{\beta}}$ (more later).

An example

- Consider a simple data example

ID	T	δ	Z
1	9	1	4
2	8	0	5
3	6	1	7
4	10	1	3

- What is the partial likelihood for this data?

An example

t_i	Risk set	$Z_{(i)}$	Contribution to the partial likelihood
6	{1, 2, 3, 4}	7	$\varphi_1(\beta) = \frac{\exp(7\beta)}{\exp(4\beta)+\exp(5\beta)+\exp(7\beta)+\exp(3\beta)}$
9	{1, 4}	4	$\varphi_2(\beta) = \frac{\exp(4\beta)}{\exp(4\beta)+\exp(3\beta)}$
10	{4}	3	$\varphi_3(\beta) = \frac{\exp(3\beta)}{\exp(3\beta)}$

The partial likelihood is the product of the three terms in the last column.

$$PL(\beta) = \varphi_1(\beta)\varphi_2(\beta)\varphi_3(\beta).$$

Partial likelihood function

- How to understand the partial likelihood function?
- We justify the partial likelihood using a simple example.
- Suppose data has only two observations

Individual	Observed time	Corvariates	Hazard
1	T_1	$\boldsymbol{Z}_1$	$h_1(t)$=$h_0(t)\exp(\boldsymbol{Z}_1{}^T\boldsymbol{\beta})$
2	T_2	$\boldsymbol{Z}_2$	$h_2(t)$=$h_0(t)\exp(\boldsymbol{Z}_2{}^T\boldsymbol{\beta})$

Partial likelihood function

Suppose the first event occurred at time t. How likely is it that the subject who had event was subject 1?

$$P(\text{ Subject 1 had event at } t \mid \text{one event at } t\text{ })$$
$$= \frac{P(\text{Sub 1 failed at t and Sub 2 didn't} \mid \text{survival to } t)}{P(\text{One event at } t)}$$
$$= \frac{P(A_1 \cap A_2^c)}{P(\{A_1 \cap A_2^c\} \cup \{A_2 \cap A_1^c\})}$$

- Let A_j = the event that subject j had event at $[t, t+\Delta t)$, given that he/she still alive at t, for j=1,2.
- $P(A_1 \cap A_2^c) = P(A_1)P(A_2^c) \approx h_1(\text{t})\Delta t(1 - h_2(\text{t})\Delta t) \approx h_1(\text{t})\Delta t$
- $P(\{A_1 \cap A_2^c\} \cup \{A_2 \cap A_1^c\})$=$P(A_1 \cap A_2^c\ ) + P(A_2\ \cap A_1^c) \approx h_1(\text{t})\Delta t + h_2(\text{t})\Delta t$

Partial likelihood

$$P(\text{ Subject 1 had event at } t \mid \text{One event at } t\,)$$
$$\approx \frac{h_1(\mathrm{t})}{h_1(\mathrm{t}) + h_2(\mathrm{t})}$$
$$= \frac{h_0(t)\exp(Z_1^T\beta)}{h_0(t)\exp(Z_1^T\beta) + h_0(t)\exp(Z_2^T\beta)}$$
$$= \frac{\exp(Z_1^T\beta)}{\exp(Z_1^T\beta) + \exp(Z_2^T\beta)}$$

- The baseline $h_0(t)$ is canceled out.

Conditional probability

- The partial likelihood
$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{D} \varphi_i = \prod_{i=1}^{D} \frac{\exp\left(\boldsymbol{Z}_{(i)}^T\boldsymbol{\beta}\right)}{\sum_{j\in R_i}\exp\left(\boldsymbol{Z}_j^T\boldsymbol{\beta}\right)}.$$
- The term φ_i characterizes the conditional probability that the event occurred to subject (i), given the information at time t_i.

Conditional probability

- The partial likelihood is obtained by taking products of these conditional probabilities over all failure times.
- Note that the arbitrary baseline hazard has been eliminated and resulting likelihood only involves $\boldsymbol{\beta}$.
- The exact values for censoring times and events times are not important (nonparametric).
- We only need the risk sets $R(t_i)$ at each event time.

Estimation using partial likelihood

- We estimate $\boldsymbol{\beta}$ by maximizing the partial likelihood

$$\mathrm{PL}(\beta) = \prod_{i=1}^{D} \frac{\exp\left(Z_{(i)}^T \beta\right)}{\sum_{j \in R(t_i)} \exp\left(Z_j^T \beta\right)},$$

or equivalently maximizing the log partial likelihood

$$\log PL(\beta) = \sum_{i=1}^{D} \left\{ Z_{(i)}^T \beta - \log \sum_{j \in R(t_i)} \exp\left(Z_j^T \beta\right) \right\}.$$

- Often numerical procedures such as the Newton-Raphson algorithm can be used to solve for the maximum partial likelihood estimate (MPLE).
- In SAS, PROC PHREG does it for you.