Cox Proportional Hazard Model

Introduction

Road map

- **Kaplan-Meier Estimator**
  - A nonparametric method to estimate survival function.
  - A useful descriptive tool.
- **Logrank Test**
  - A nonparametric method to test for differences in S(t) for a categorical covariate,
  - Can't quantify the difference.
- **Accelerated Failure Time model**
  - A parametric regression model to quantify the effects of covariates (both continuous and categorical) on survival,
  - Need to specify a distribution on survival time.
- **Cox Proportional Hazard model**
  - A semiparametric regression model,
  - More flexible than the AFT model.

Introduction

- First proposed by David Cox in 1972 JRSSB paper

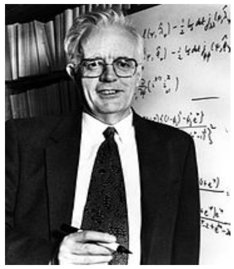
Cox, David R (1972). "*Regression Models and Life-Tables*". *Journal of the Royal Statistical Society, Series B 34 (2): 187-220.*

- The most cited paper in JRSSB, and one of the most cited statistical papers in the literature.

Paper	Total Citation	Citation per year
Kaplan-Meier (1958)	46443	800
Cox model (1972)	40310	916
EM (Dempster 1977)	43524	1116
Bootstrap (Efron 1979)	12674	342
FDR (Benjamini & Hockberg 1995)	30345	1445

Introduction (Wikipedia)

- David Cox: a British statistician
- His contributions:
  - Box-Cox transformation,
  - Logrank test (Mantel-Cox test)
  - Proportional hazard model
  - Cox process.



Introduction

- Let $h(t|\boldsymbol{Z})$ be the hazard rate at time t for an individual with covariates $\boldsymbol{Z} = (Z_1, \dots, Z_p)^T$.
- The hazard rate $h(t|\boldsymbol{Z})$ characterizes the instantaneous failure rate given the patient is still alive prior to t.
- How to model $h(t|\boldsymbol{Z})$?

Introduction

- Try a linear model
  $$h(t|\boldsymbol{Z}) = \beta_0 + \beta_1 Z_1 + \cdots + \beta_p Z_p$$
  - Covariates have an additive effect on hazard function
  - β_0 is the baseline hazard rate for individuals with $\boldsymbol{Z}$=0
- A unreasonable model
  - Hazard function is non-negative with $h(t|\boldsymbol{Z}) \geq 0$.
  - The linear model can not guarantee non-negativity.

Introduction

- Try an exponential linear model
  $h(t|\boldsymbol{Z}) = \exp(\beta_0 + \beta_1 Z_1 + \cdots + \beta_p Z_p)$
  - Covariates have a multiplicative effect on hazard function
  - $\exp(\beta_0)$ is the baseline hazard rate for individuals with $\boldsymbol{Z}$=0
- An unreasonable model
  - It assumes a constant hazard or an exponential distribution.
  - Unrealistic for most applications.

Proportional hazard model

- The basic proportional hazard regression model (Cox 1972) assumes
  $$h(t|\boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{Z}^T\beta)$$
  - $h_0(t)$ an arbitrary baseline hazard (nonparametric)
  - $h_0(t)$ depends on time t, but not on covariate
  - $\exp(\boldsymbol{Z}^T\beta)$ depends on covariates, but not on time t
  - it separates the shape of the hazard function from the effects of covariates.
- It is *semiparametric* because a parametric form assumed for covariate effect and a nonparametric form for the baseline hazard rate.

Proportional Hazard Model

- A good model
  $$h(t|\boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{Z}^T\beta)$$
  - $\exp(\boldsymbol{Z}^T\beta)$
    - guarantees positivity of hazard function
    - Easy interpretation of covariate effects on hazard
  - Nonparametric $h_0(t)$
    - No distributional assumption on survival time
  - Product form
    - Separate the estimation of β and $h_0(t)$
    - Can quantify covariate effects β, without knowing $h_0(t)$.

Proportional hazard model

- Cox proportional hazard model
$$h(t|\boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{Z}^T\beta)$$
- Baseline hazard $h_0(t)$ can be any nonnegative function of t.
- If $h_0(t)$ has a parametric form, then the distribution of $T_0|Z$ will be fully specified (parametric model).
- For example, $h_0(t) = \exp(\beta_0)$ gives exponential regression model.
- In Cox model, we do not want to fully specify the distribution of $T_0|\boldsymbol{Z}$ and leave $h_0(t)$ unspecified (nonparametric).

Proportional Hazard Model

- In Cox proportional hazard model
$$h(t|\boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{Z}^T\beta)$$
- No intercept β_0 term is allowed in $\boldsymbol{Z}^T\beta$, since it is confounded with the baseline hazard.

Proportional hazard

- Cox proportional hazard model
$$h(t|\boldsymbol{Z}) = h_0(t)\exp(\boldsymbol{Z}^T\beta)$$
- An important property: **a proportional hazard model** $\frac{h(t|\boldsymbol{Z}=\boldsymbol{z})}{h(t|\boldsymbol{Z}=\boldsymbol{z}^*)} = \frac{h_0(t)\exp(\boldsymbol{z}^T\beta)}{h_0(t)\exp(\boldsymbol{z}^{*T}\beta)} = \exp\left((\boldsymbol{z}-\boldsymbol{z}^*)^T\beta\right)$,
- It is a constant over time.

Interpretation of hazard

- The hazard ratio $\frac{h(t|\boldsymbol{Z}=\boldsymbol{z})}{h(t|\boldsymbol{Z}=\boldsymbol{z}^*)}$ can be understood as the relative risk of having the event for an individual with risk factor $\boldsymbol{z}$ as compared to an individual with risk factor $\boldsymbol{z}^*$.
- For binary Z, $\frac{h(t|Z=1)}{h(t|Z=0)} = \frac{h_0(t)\exp(\beta)}{h_0(t)} = \exp(\beta)$.
  - Z= 1 if treatment; Z= 0 if placebo
  - $\exp(\beta)$ is the risk of having the event for an individual received treatment relative to the risk for an individual received the placebo.
- The relative risk does not change in time for Cox PH model.

Interpretation of hazard

For example,

- If $\frac{h(t|Z=1)}{h(t|Z=0)} = \exp(\beta_1) = 2$, then the risk of having the event of an individual received treatment is always twice the risk for an individual received the placebo, no matter how long the patient has survived.
- If $\frac{h(t|Z=1)}{h(t|Z=0)} = \exp(\beta_1) = 1/2$, then the risk of having the event of an individual received treatment is always half of an individual received the placebo, no matter how long the patient has survived.

Proportional Hazard Model

- Warning: Proportional hazard is an important assumption of the Cox model.
- Need to check if such assumption is appropriate in real data analysis.
- Later, we will introduce some extensions to the Cox model to deal with the non-proportional hazard scenarios.