

Cox PH Model

Model Diagnosis

Introduction

- So far, we have learned, **for a given Cox PH model**,
 - Estimation of β , and baseline $h_0(t)$ or $S_0(t)$.
 - Inference: testing effect of predictors.
- Outstanding question:
 - **Is Cox PH model appropriate?**
- An inappropriate model can lead to biased estimation and misleading conclusions.

Introduction

- Similar to other statistical models, Cox PH model is valid under certain assumptions.
- It is important to check the validity of these assumptions and identify ways in which they might be violated.

Cox PH model

- Recall the Cox's proportional hazard model
 $h(t | \mathbf{Z}) = h_0(t) \exp(\mathbf{Z}^T \boldsymbol{\beta})$.
- Two key model assumptions
 - Proportional hazard assumption
 - The linear form $\mathbf{Z}^T \boldsymbol{\beta}$
- Also interested in diagnosis of
 - Outliers: any observation that are not well captured by the model

Residuals for Cox model

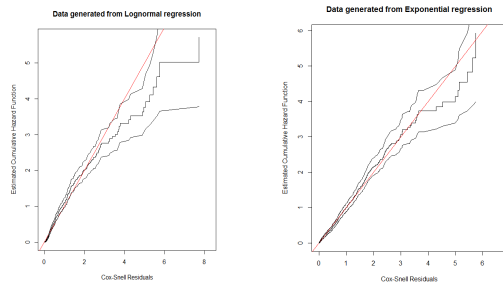
- Hypothesis testing can be used to address some of the model diagnosis questions.
- We are interested in residuals plots, a graphical way to address these questions.
- In particular, we will introduce
 - Cox-Snell residuals
 - Martingale residuals
 - Deviance residuals
 - Schoenfeld residuals

Cox-Snell residuals

- We have seen it in parametric regression model diagnosis.
- The Cox-snell residuals are defined as the estimated cumulative hazard,

$$r_{CS,i} = \hat{H}(t_i | \mathbf{Z}_i) = \hat{H}_0(t_i) \exp(\mathbf{Z}_i^T \hat{\boldsymbol{\beta}})$$
- If model is correctly specified, $H(T_i^0 | \mathbf{Z}_i) \sim \text{Exp}(1)$.
 - Obtain K-M estimator $\hat{S}_{KM}(t)$ based on $\{r_{CS,i}\}_{i=1}^n$.
 - Plot $\log \hat{S}_{KM}(t)$ against t
 - If Cox model provides a good fit, then expect a straight line.

An example



Cox-Snell residuals

- Cox-snell residuals can be used to **assess the overall fit** of the model.
- But does not provide additional information on the type of departure from the model when the plot is not linear.
- Although $H(T_i^0 | Z_i) \sim \text{Exp}(1)$ but we are using estimated $\hat{H}(T_i^0 | Z_i)$.
- Closeness of the distribution of the residuals to $\text{Exp}(1)$ depends on sample size n .

Summary

- Model diagnosis for Cox PH model
- Cox-snell residuals
- More residuals will be introduced.
