Statistics for Biology and Health
Chapter 8 Refinements of the semiparametirc proportional hazards model

Qi Guo

July 24,2019

THE UNIVERSITY OF TEXAS AT DALLAS
School of Natural Sciences and Mathematics

Contents

Introduction

- When the covariate may become a time-dependent variable, in notes before we introduce a $Z(t)$ instead of Z, and for commonly used model:

$$h[t|Z(t)] = h_0(t)\exp[\beta' Z(t)] = h_0(t)\exp\left[\sum_{k=1}^{p}\beta_k Z_k(t)\right]$$

- If the proportional hazard assumption is violated for a variable, we can stratify on this variable.Stratification fits a different baseline hazard function for each stratum, so that the form of the hazard function for different levels of this variable is not constrained by their hazards being proportional.
- And the basic proportional hazards model can be extended to left-truncated survival data.

Time-Dependent Covariates

Time-Dependent Covariates

- Now our data based on a sample of size n, consists of the triple $[T_j, \delta_j, [Z_j(t), 0 \leq t \leq T_j]],\ j = 1, \cdots, n$, where T_j is the time on study for the jth patient, δ_j is the event indicator, and $Z_j(t) = [Z_{j1}(t), \cdots, Z_{jp}(t)]'$ is the vector of covariates for the jth individual.
- And we assume that the censoring time and event are independent, the distinct event time $t_1 < t_2 < \cdots < t_D$, so $Z_{(i)}(t_i)$ is the covrariate associated with the individual whose failure time is t_i and $R(t_i)$ is the risk set at time t_i, so the partial likelihood is given by:

$$L(\beta) = \prod_{i=1}^{D} \frac{\exp\left[\sum_{b=1}^{p} \beta_b Z_{(i)b}(t_i)\right]}{\sum_{j \in R(t_i)} \exp\left[\sum_{b=1}^{p} \beta_b Z_{jb}(t_i)\right]}$$

Time-Dependent Covariates

- In notes before we know there are a cut points here to coded the time-dependent variable into 0 and 1, if $t <$ time at which event occurs, let $Z(t) = 0$, and 1 o.w.
- And we skip this part as it's pretty close as before.

Stratified Proportional Hazards Models

Coding Covariates

- Here the subjects in the jth stratum have a arbitrary baseline hazards function $h_{0j}(t)$ and the effect of other explanatory variables on the hazards function can be represented by a proportional hazards model in that stratum as :

$$h_j[t|Z(t)] = h_{0j}(t)\exp[\beta' Z(t)],\ j = 1, \cdots, s$$

- The regression coefficients are assumed to be the same in each stratum although the baseline hazard functions may be different and completely unrelated.

Estimation and hypothesis testing

- The partial log likelihood function is given by:

$$LL(\beta) = [LL_1(\beta)] + [LL_2(\beta)] + \cdots + [LL_s(\beta)]$$

  where $[LL_j(\beta)]$ is the log partial likelihood using only the data for those individuals in the jth stratum.
- A key assumption in using a stratified proportional hazards model is that the covariates are acting similarly on the baseline hazard function in each stratum.
- This can be tested by using either a likelihood ratio test or a Wald test.

Estimation and hypothesis testing

- Fit the stratified model, which assumes common β's in each stratum, and obtain the log partial likelihood, $LL(b)$.
- Using only data from the jth stratum, a Cox model is fit and the estimator b_j and the log partial likelihood $LL_j(b_j)$are obtained.
- The log likelihood under the model, with distinct covariates for each of the s strata, is $\sum_{j=1}^{s} LL_j(b_j)$.
- The likelihood ratio chi square for the test that the $beta'$s are the same in each stratum is $-2[LL(b) - \sum_{j=1}^{s} LL_j(b_j)]$ which has a large-sample, chi-square distribution with $(s-1)p$ degrees of freedom under the H_0.

Left Truncation

Introduction

- Actually all of the material in this chapter is similar as before, and for the left truncated data, it arises when the event time X is the age of the subject and persons are not observed from birth but rather from some other time V corresponding to their entry into the study.
- For example, the age, X_i, at death for the ith subject in a retirement center in California was recorded. Because an individual must survive to a sufficient age V_i to enter the retirement community, and all individuals who died prior to entering the retirement community were not included in this study, the life lengths considered in this study are left-truncated.
- Another situation is when the event time X is measured from some landmark, but only subjects who experience some intermediate event at time V are to be included in the study. The times V are sometimes called delayed entry times.

Formulate a proportional model

- To formulate a proportional hazards regression model for a set of covariates Z, we model the conditional hazard rate of t, given Z and $X > V$:

$$h(t|Z, X > V) \cong \frac{P(X = t|Z, X > V)}{P(X \geq t|Z, X > V)}$$

- If the event time X and the entry time V are conditionally independent, given the covariates Z, then $h(t|Z(t), X > V) = h(t|Z(t))$

Estimate the coefficients

- The partial likelihoods are modified to account for delayed entry into the risk set, and in notes before, we only need to add a delay entry time in R codes, so for the partial likelihoods will be different in the risk set $R(t)$.
- Define the risk set $R(t)$ at time t as the set of all individuals who are still under study at a time just prior to t. Here, $R(t) = \{j | V_j < t < T_j\}$, and the rest are same.