SOC 7717 Event History Analysis and Sequence Analysis

Week 6: Semi-Parametric Methods

Wen Fan

Spring 2019

Outline

Semi-Parametric Models

Partial Likelihood Method

Interpretation of Coefficients

The Proportional Hazards Assumption

Estimating the Survivor and Hazard Functions

Stratification

Time-Varying Explanatory Variables

Semi-Parametric Models

Problems with parametric models

- Must choose one distribution, and often no good basis for choice
- Conventional models are somewhat restricted in shape of hazard function
- Many parametric models don't allow for time-varying explanatory variables

Semi-Parametric models

Problems resolved in 1972 by David Cox in his *Journal of the Royal Statistical Society B* paper, "Regression Models and Life Tables"

Proposed:

- A more general model: proportional hazards model
- An estimation method: partial likelihood

Cox model as a proportional hazards model

A simple generalization of Weibull and Gompertz models

For time-constant covariates, we can express the model as:

$$h_i(t) = h_0(t)\exp(\beta_1 X_{i1} + ... + \beta_k X_{ik}).$$

Therefore, the hazard for individual i at time t is the product of

- A function $h_0(t)$ that is left unspecified, except that it cannot be negative (baseline hazard)
- A linear function of a set of k covariates (log relative-hazard, risk score), which is then exponentiated (relative hazard)

Cox model as a proportional hazards model

Why is it called the proportional hazards model? Take the ratio of the hazards for any two individuals

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t)e^{\beta_1 X_{i1}+...+\beta_k X_{ik}}}{h_0(t)e^{\beta_1 X_{j1}+...+\beta_k X_{jk}}} = \frac{e^{\beta_1 X_{i1}+...+\beta_k X_{ik}}}{e^{\beta_1 X_{j1}+...+\beta_k X_{jk}}}$$

which is constant over time, i.e., it does not depend on time. The arbitrary function $h_0(t)$ cancels out

Cox model as a roportional hazards model

The Cox proportional hazards model does not make any assumption about the shape of the underlying hazards, but makes the assumption that the hazards for subgroups are proportional over follow-up time, although it can easily be extended to allow for non-proportional hazards

Implicit assumptions

- No covariates that affect the hazard are omitted from the model
- All covariates are measured without error
- Observations are independent
- Censoring is noninformative

Partial Likelihood Method

Partial likelihood

One of Cox's major contribution was to invent a new method of estimation without specifying exactly what $h_0(t)$ is

Similar to maximum likelihood, but maximizes a "partial" likelihood rather than the usual likelihood

Basic ideas about partial likelihood

Likelihood function for data from PH models can be factored into two parts:

$$L = A(\beta) \times B(\beta, h_0(t))$$

- Part A contains information about the β's only
- Part B contains information about both the β's and $h_0(t)$

In partial likelihood estimation, we throw away part B and treat part A as an ordinary likelihood function

Part A depends only on the order in which events occurred, not on the exact times of occurrence

Properties of partial likelihood estimators

- Consistent
- Asymptotically normal
- Not fully efficient because some information is lost (usually not worth worrying about)

Tied data

- Marginal calculation (exact-marginal calculation, continuous-time calculation)
  - Tied data as a result of our limited ability to measure time
  - `exactm`
  - Approximation
    - Breslow's approximation: `breslow` (default)
    - Efron approximation: `efron`
- Partial calculation (exact-partial calculation, discrete-time calculation)
  - Events really occur at the same time
  - `exactp`

Tied data

- When ties are relatively few, it makes little difference which method is used
- When the number of ties is large, the approximation methods tend to yield coefficients that are biased toward 0
- Exact methods need a substantial amount of computer time for large data sets containing many ties
- If the exact methods are too time-consuming, use the Efron approximation, at least for model exploration

Interpretation of Coefficients

Interpreting the estimated coefficients

In PH regression, the baseline hazard component, $h_0(t)$, vanishes from the partial likelihood; we only obtain estimates of the regression coefficients associated with the explanatory variables $X_1, ..., X_k$

Consider the simplest possible setup, one involving only a single binary variable, X:

$$\log h(t) = \log h_0(t) + \beta X \Rightarrow$$

- When $X = 0$, $\log h(t; X = 0) = \log h_0(t) + 0 = \log h_0(t)$
- When $X = 1$, $\log h(t; X = 1) = \log h_0(t) + \beta$

Therefore $\beta = \log h(t; X = 1) - \log h(t; X = 0) = \log \frac{h(t;X=1)}{h(t;X=0)}$

Interpreting the estimated coefficients

- That is, β is the logarithm of the ratio of the hazard rate for subjects belonging to the group denoted by $X = 1$ to the hazard rate for subjects belonging to the group indicated by $X = 0$
- The parameter β represents log relative risk and $\exp(\beta)$ represents relative risk
- A confidence interval for β, given by $\hat{\beta} \pm 1.96$SE, represents a range of plausible values for the log relative risk associated with the corresponding explanatory variable

Interpreting the estimated coefficients

Corresponding confidence intervals for the relative risk associated with the same covariate are obtained by transforming the confidence interval for β, i.e.,

$$(\beta_l, \beta_u) \Rightarrow (e^{\beta_l}, e^{\beta_u})$$

Since the estimates $\beta_1, ..., \beta_k$ are obtained simultaneously, these estimated relative risks adjust for the effect of all the remaining covariates included in the fitted model

The Proportional Hazards Assumption

The proportional hazards assumption

- The proportional hazards assumption is a strong assumption and its appropriateness should always be assessed
- The assumption says that the ratio of the hazard functions for any two subgroups (i.e., two groups with different values of the explanatory variable X) is constant over follow-up time
  - Note that it is the hazard ratio which is assumed to be constant. The hazard can vary freely with time

The proportional hazards assumption

- When comparing an aggressive therapy vs. a conservative therapy, for example, it is possible that the patients receiving the aggressive therapy do worse earlier, but then have a lower hazard than those receiving the conservative therapy
- $\longrightarrow$ The ratio of the hazard functions will not be constant over time
- Generally, if the hazard functions cross, it is possible that the effect of the variable will not be statistically significant despite the presence of a potentially interesting effect
- As such, it is important to plot survival curves before fitting the model and to assess the appropriateness of the proportional hazards assumption after the model has been fitted

Commonly used methods for assessing the assumption

- Plotting the cumulative survivor functions and checking that they do not cross (although the survivor functions do not have to cross for the hazards to be non-proportional)
- Plotting the log cumulative hazard functions over time and checking for parallelism
- Including time-by-covariate interaction terms in the model and testing statistical significance
- Plotting Schoenfeld's residuals against time to identify patterns

1. Plots of the log cumulative hazard function

Another way to express the PH model is in terms of the survivor function:

$$S(t;X) = S_0(t)^{\exp(\beta_1 X_1 + \ldots + \beta_k X_k)},$$

where $S_0(t)$ is called the "baseline" survivor function. It's the estimated survivor function for someone whose X values are all 0

Suppose we have only a single binary variable X, then:

$$S(t;X=1) = S(t;X=0)^{\exp(\beta)} (\because S(t;X=0) = S_0(t))$$

Taking natural logarithms of both sides gives:

$$\log S(t;X=1) = \exp(\beta) \log S(t;X=0)$$

Taking natural logarithms of the negatives of both sides gives:

$$\log[-\log S(t;X=1)] = \beta + \log[-\log S(t;X=0)]$$

1. Plots of the log cumulative hazard function

- Consequently, if the proportional hazards assumption holds, plots of $\log[-\log S(t)]$ vs. t for each group will be parallel, with the constant difference between them equal to the coefficient β
- Plots of $\log[-\log S(t)]$ are often called log cumulative hazard plots, because $-\log S(t)$ is equivalent to the cumulative hazard function
- `stphplot, by(var)`
  - » Note that the lines do not have to be straight, it is only necessary for there to be a constant difference between the lines

2. Including time-by-covariate interaction terms

Consider again a PH model with one single binary variable, X_1, and an interaction term between X_1 and t, which takes the value 1 if an exposure is present and 0 if it is absent

$$h(t) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_1 t)$$

- A significant estimate for β_2 indicates that the hazard ratio is non-constant over time
- Note: this is not a general test of the proportional hazards assumption. It merely tests whether the hazard ratio changes *monotonically* with time

2. Including time-by-covariate interaction terms

In practice, it is possible to fit any model of the form

$$h(t) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_1 f(t)),$$

where $f(t)$ is a function of time

- A similar approach to testing the PH assumption is to partition the time axis and fit separate models for different time periods
- If the PH assumption is appropriate, the parameter estimates will be similar for each model
- The problem is in how to partition the time axis, i.e., the choice of cutpoints

3. Schoenfeld residuals

Suppose individual i with covariates X_i has an event at time t_i. Suppose that 20 individuals (including person i) are at risk at the same time t_i. Denote these 20 individuals by $j = 1, \ldots, 20$

For any individual k within this set of 20, based on the estimated Cox model, we can calculate the probability that k had an event at time t_i

$$p_k = \frac{e^{\beta X_k}}{\sum_j e^{\beta X_j}}$$

The "expected" values of the covariates at time t_i can then be written as

$$\bar{X}_i = \sum_j X_j p_j$$

The vector of Schoenfeld residuals for individual i is defined as $X_i - \bar{X}_i$

3. Schoenfeld residuals

If the proportional hazards assumption is satisfied, there should be no relationship between Schoenfeld residuals and time

In other words, we can model the Schoenfeld residuals as a function of time and test the hypothesis of a zero slope

To get a detailed test for each predictor, you must use the "scaled" Schoenfeld residuals

There are options for `stphtest` that look at the relationship between these residuals and other function of time: `km`, `log`, `rank`

Adequacy of Proportional Hazards Assumption

Many people worry about whether their data satisfy the proportional hazards assumption

- A legitimate concern, but probably overemphasized
- Even when violated, likely to be a pretty good approximation
- More serious problems are likely to be omitted explanatory variables, measurement error, or informative censoring

Other residuals for Cox regression

- Cox-Snell residuals: useful in assessing overall model fit
- Martingale residuals: a transformation of Cox-Snell residuals, useful in determining the functional form of the covariates
- Deviance residuals: a further transformation of martingale residuals
  - Behave much like standardized residuals from OLS regression: symmetrically distributed around 0 with an approximate standard deviation of 1

Martingale residuals

Martingale residuals are useful for examining functional form

- Fit a model with no covariates, and request the martingale residuals
- Plot the residuals against each quantitative covariate, using a smoothing operator to estimate the functional form

Sensitivity analysis for informative censoring

Are the results sensitive to the possibility that the censoring may be informative rather than noninformative? Redo the analysis in two different ways:

- Treat all randomly censored cases as though they experienced events immediately after being censored
  - » `stset` the data without specifying a censoring indicator. Then estimate the model
- Treat all randomly censored cases as though the censoring occurred after the largest event time that is observed in the sample
  - » Replace the duration time for the censored cases and then `stset` the data with the censoring indicator

Sensitivity analysis for informative censoring

Logic: If you don't know when events occurred, consider extreme possibilities: (1) immediately after censoring, (2) later than any other observed events

If the results don't change, then you can feel more confident that censoring is not an issue

Estimating the Survivor and Hazard Functions

Estimating the survivor and hazard functions

Partial likelihood treats the hazard (and therefore the survival) function as a "nuisance function" that cancels out of the estimating equations

However, once the regression parameters are estimated, it's possible to get a semi-parametric estimate of the survivor function for specified values of the covariates

Recall that another way to express the proportional hazards model is in terms of the survivor function

$$S_i(t) = S_0(t)^{\exp(\beta X_i)}$$

Estimating the survivor and hazard functions

The problem with the baseline survivor or cumulative hazard functions is that $X = 0$ may be a very rare or impossible condition

It may be more informative to calculate the survivor function evaluated at the means of the covariates:

$$\bar{S(t)} = [S_0(t)]^{\exp(\beta\bar{X})}$$

Stratification

Stratification

If explanatory variable is categorical (e.g., race or sex), we can accommodate nonproportionality by stratification, e.g.,

Males:

$$\log h(t) = h_M(t) + \beta_1 X_1 + \beta_2 X_2 + ...$$

Females:

$$\log h(t) = h_F(t) + \beta_1 X_1 + \beta_2 X_2 + ...$$

Thus, the regression coefficients are the same, but the arbitrary functions of time are allowed to differ ($h_M(t)$ and $h_F(t)$). These two equations can be estimated simultaneously

Stratification in Stata

```
stcox varlist, strata(sex)
```

These estimates control for sex, but without imposing the assumption that the effect of sex is constant over time

Advantages and disadvantages of stratification

Advantages

- Stratification allows for any change in the effect of a variable over time (whereas interaction method requires that you choose a particular form for the interaction)
- It takes less computing time

Disadvantages

- No estimates are obtained for the effect of the stratifying variable
- It can be less efficient than the interaction method if the form of the interaction with time is correctly specified

Time-Varying Explanatory Variables

Time-varying Explanatory Variables

So far we have assumed that explanatory variables do not change over time

But this is not appropriate for many important cases

- E.g., if we want to know how income affects residential mobility, it is crucial to measure income at multiple time points

Time-varying Explanatory Variables

$$\log h(t) = \log h_0(t) + \beta_1 X_1 + \beta_2 X_2(t) + \ldots$$

May want to lag variables. E.g., if job performance affects the hazard of promotion, may want to lag the performance measure by some appropriate amount, say, a year

$$\log h(t) = \log h_0(t) + \beta_1 X_1 + \beta_2 X_2(t-1) + \ldots$$

Could also have more than one lagged version of the same variable

Observational requirements

At the time that each event occurs, must know the values of the explanatory variables for all the individuals at risk at that time

- "At risk" means that they have not already experienced an event or been censored
- E.g., If an event occurs at time 8, and 20 persons were at risk at time 8, one must know the values of the time-varying variables at time 8 for all 20 persons

Some typical cases

1. We know the exact time of any changes, and the values of the variable before and after the change

- E.g., $X(t)$ is a dummy for marital status and we know the date of any marriage or dissolution $\longrightarrow$ the observational requirements met

2. We observe a variable at fixed intervals of time, but we don't know whether any changes occurred between intervals

- E.g., blood pressure measured daily, employment status measured annually $\longrightarrow$ need some ad hoc method for assigning values for any point in time

Some typical cases

2.1. Suppose event times are measured less precisely than the intervals for the X variables

E.g., X variables are measured weekly and we know only the month of the event time. Then

- Take an average of the values of the X variables in the month at which the event occurs, or
- Take the value of the X variable at the beginning of the relevant interval (avoids possible endogeneity of the X variable)

Some typical cases

2.2. Suppose event times are measured more precisely than the intervals for the X variables

E.g., exact date of marriage and annual measurements of income (in the preceding month). Then we can

- Take closest value. Thus, if an event occurs at time 9.7 and X-variables are measured at times 9 and 10, use $X(t = 10)$
- Use linear interpolation. Since 9.7 is 70% of the distance between 9 and 10, use $X(t = 9.7) = .7X(t = 10) + .3X(t = 9)$
- Take closest preceding value, e.g., $X(t = 9)$ (avoids possible endogeneity)

Model setup with time-dependent covariates

Two different ways of setting up models with time-dependent covariates

- Wide form: One record for each individual, multiple covariate values are separate variables on the data record
- Long form: One record for each interval of time in which covariates are constant. Time-dependent variables treated just like other variables

We will focus on long form by using episode splitting

Episode splitting

Create a separate record for each interval of time in which the covariates remain constant

This record should include the beginning time and ending time of the interval

If the interval did not end in an event, code it as censored

`stsplit`