Summary Notes Cox Model

Monday, October 9, 2023 17:28

Regression in general:

$$Z = \frac{\hat{\beta}_j - 0}{\widehat{SE}(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)}$$

Dependent variable = function of independent variables

Linear regression: mean of continuous var

$$E[Y_i] = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

Intercept that characterizes the dependent variable in the reference group

Logistic regression: log odds

$$\log\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}$$

Coefficients that characterize the effect on the dependent variable of a one-unit change in the independent variable

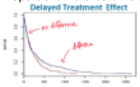
We fit the model using maximum likelihood and obtain estimates of these parameters $\hat{\beta}_0, \ldots, \hat{\beta}_k$. If we wish to obtain a predicted probability of positive response $\hat{p}_i$ for person i with covariates $X_{i1}, \ldots, X_{ik}$, we can enter their values into the following equation:

$$\hat{p}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik})}$$

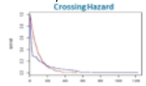
Cox Model:

Under the proportional hazards assumption, the survival curves diverge over time (until survival approaches 0).
Proportional hazards is violated if delayed treatment effect or crossing hazards. Aka, if relationship between the hazards changes over time.

Delayed Treatment Effect



Crossing Hazard



Dependent variable = function of independent variables

Cox model:

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

Baseline hazard function that characterizes the dependent variable in the reference group

Coefficients that characterize the effect on the dependent variable of a one-unit change in the independent variable

hazard is a function of t; no longer just mean or log odds

hazard in group w/ all covariates 0

covariates are multiplicative

βj is the log hazard ratio for the jth covariate
$\exp(\beta j)$ is the hazard ratio for the jth covariate

- For a binary covariate, this is the hazard ratio comparing the group with Xij = 1 vs. the group with Xij = 0 (holding all other covariates constant)
  - For a binary covariate, β_j is the log hazard ratio of Y_i for the group with X_{ij} = 1 compared to the group with X_{ij} = 0, holding all other covariates constant.
    - If $\beta_j > 0$, the hazard ratio $\exp(\beta_j) > 1$; this means that a higher value of X_{ij} is associated with a higher hazard and worse survival.
    - If $\beta_j < 0$, the hazard ratio $\exp(\beta_j) < 1$; this means that a higher value of X_{ij} is associated with a lower hazard and better survival.
    - If $\beta_j = 0$, the hazard ratio $\exp(\beta_j) = 1$; this means that different values of X_{ij} will have the same hazard and same survival.
- For a continuous covariate, this is the hazard ratio for a one-unit increase in Xij, e.g. Xij = 10 vs. Xij = 9 (holding all other covariates constant)

$$\frac{h_i' t}{h_i t} = \exp(\beta_j)$$

$$\log\left[\frac{h_i'(t)}{h_i(t)}\right] = \beta_j$$

$$H_1(t) = \exp(\beta)H_0(t)$$

$$\log(H_1(t)) = \beta + \log(H_0(t))$$

$$S_1(t) = [S_0(t)]^{\exp(\beta)}$$

Prediction

- The proportional hazards model gives us a framework for making predictions

hazard function
$$\hat{h}_i(t) = \hat{h}_0(t)\exp(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik})$$

cumulative hazard function
$$\hat{H}_i(t) = \hat{H}_0(t)\exp(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik})$$

$$\hat{S}_i(t) = \exp\left(-\hat{H}_i(t)\right)$$
Survival

- Using partial likelihood estimation, we estimate $\hat{\beta}_1, \ldots, \hat{\beta}_k$ but not the flexible baseline hazard or cumulative hazard functions

Baseline cumulative hazard

- **Breslow estimator** of the baseline cumulative hazard function:

$$\hat{H}_0(t) = \sum_{j:t_j \le t} \frac{d_j}{\sum_{i=1}^{n_j} \exp(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik})}$$

instead, sum each individual's hazard ratio

- Related to the Nelson-Aalen cumulative hazard estimator:

$$\hat{H}(t) = \sum_{j:t_j \le t} \frac{d_j}{n_j}$$

Breslow estimator reduces to Nelson-Aalen at baseline because all covariates are 0

Likelihood:

Semiparametric regression

The Cox proportional hazards model has the following form:

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

Cox's model is known as a **semiparametric regression model** because it has parametric component and a nonparametric component. The nonparametric component of the model is the flexible baseline hazard $h_0(\cdot)$. The parametric component of the model is the term $\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})$, where each covariate X_{ij} has a multiplicative effect on the hazard function.

Likelihood estimates:

We are able to estimate the $\beta_1, \ldots, \beta_k$ without estimating the entire baseline hazard $h_0(\cdot)$. This is because the full likelihood can be split into two terms:

$$L(\beta_1, \ldots, \beta_k)L(h_0(\cdot), \beta_1, \ldots, \beta_k)$$

The first term $L(\beta_1, \ldots, \beta_k)$ depends on $\beta_1, \ldots, \beta_k$ but does not depend on $h_0(\cdot)$. It is known as the **partial likelihood** because it is only one part of the full likelihood. The second term depends on $h_0(\cdot)$, as well as containing a little bit of information on the log hazard ratios.

In Cox proportional hazards regression, instead of calculating the full likelihood for the data, we calculate the partial likelihood. We find the values of $\beta_1, \ldots, \beta_k$ where the partial likelihood maximizes (as if it were a full likelihood). These are our **maximum partial likelihood estimates** $\hat{\beta}_1, \ldots, \hat{\beta}_k$. The maximum partial likelihood estimator is consistent and asymptomatically normal.

Cox partial likelihood → aka $L(\beta_1, \ldots, \beta_k)$ from previous slide

- Consider unique failure times $T_1, \ldots, T_m$
- At time T_i, individual i fails; the risk set at that time is $R(T_i)$
- The partial likelihood is the product over all m unique failure times:

$$\prod_{i=1}^{m} \frac{\exp(\beta_1 X_{i1} + \cdots + \beta_k X_{ik})}{\sum_{j \in R(T_i)} \exp(\beta_1 X_{j1} + \cdots + \beta_k X_{jk})}$$

hazard rate for person who did fail vs. hazard of everyone still at-risk

- The contribution at each failure time is the conditional probability that individual i was the one to fail out of risk set $R(T_i)$

- It is possible to estimate $\beta_1, \ldots, \beta_k$ without estimating $h_0(\cdot)$

$$L(\beta_1, \ldots, \beta_k)L(h_0(\cdot), \beta_1, \ldots, \beta_k)$$

- The left-hand term is known as the **Cox partial likelihood**

Partial likelihood contribution:

For simplicity, assume that there is one unknown β and one covariate X_j for each person j. The partial likelihood contribution is:

$$\frac{h_i(T_i)}{\sum_{j \in R(T_i)} h_j(T_i)} = \frac{h_0(T_i)\exp(\beta X_i)}{\sum_{j \in R(T_i)} h_0(T_i)\exp(\beta X_j)} = \frac{\exp(\beta X_i)}{\sum_{j \in R(T_i)} \exp(\beta X_j)}$$

Handling ties:

Method	Approach	Advantages/Disadvantages
Exact method	Considers all $d_i!$ possible combinations of breaking ties	(++) Most accurate (−) Computationally complex
Efron approximation	Approximates the mean denominator of the Exact method	(+) Accurate (+) Computationally efficient
Breslow approximation	Ignores the fact that failures are removed from the risk set	(+) Simplest (−) Least accurate

See more in Reading 5

Prediction

- We can predict survival at time t for a person with covariates $X_{i1}, \dots, X_{ik}$ as:

$$\tilde{S}_i(t) = \exp\left(-\tilde{H}_0(t)\exp\left(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}\right)\right)$$

$$= \left[\tilde{S}_0(t)\right]^{\exp\left(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}\right)}$$

$$= \left[e^{-\tilde{H}_0(t)}\right]^{\exp\left(\hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}\right)}$$