Notes for Biostatistics 140.641.01 'Survival Analysis'

First Term, 2022

Chapter 5. PHM and Beyond

Denote by $\boldsymbol{x_i}$ the $p \times 1$ covariates and $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters. The proportional hazards model (PHM) assumes

$$\begin{aligned}\lambda(t; \boldsymbol{x_i}) &= \lambda_0(t) e^{\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_p x_{ip}} \\ &= \lambda_0(t) e^{\boldsymbol{\beta}' \boldsymbol{x_i}}\end{aligned}$$

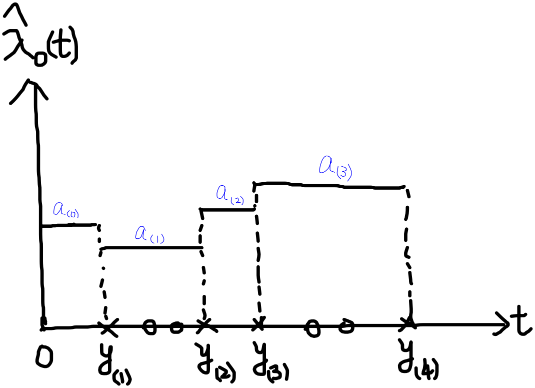
The parameters $\boldsymbol{\beta}$ can be estimated by the partial likelihood method as seen in Chapter 4. If we can estimate the baseline hazard function $\lambda_0(t)$ we then can estimate the absolute risk for developing the failure event (e.g., a certain disease)

5.1 Estimation of $\lambda_0(t)$

Suppose survival time T is continuous. Breslow (1972, *JRSS B*, discussion of Cox paper) gave a heuristic argument. He assumed $\hat{\lambda}_0(t)$ remains constant between uncensored survival times. Let $a_{(0)}, a_{(1)}, a_{(2)}, \ldots$ be constants

$$\hat{\lambda}_0(t) = \begin{cases} a_{(0)} & 0 \leq t < y_{(1)} \\ a_{(1)} & y_{(1)} \leq t < y_{(2)} \\ \ldots & \end{cases}.$$

Say, we are interested in deriving $a_{(2)}$.


$\hat{\lambda}_0(t)$
$a_{(0)}$
$a_{(1)}$
$a_{(2)}$
$a_{(3)}$
0
$y_{(1)}$
$y_{(2)}$
$y_{(3)}$
$y_{(4)}$
t

Since we observe one person fails at $y_{(2)}$, thus for given $(y_{(2)}, R_{(2)})$,

$$\begin{aligned} 1 &= \sum_{j \in R_{(2)}} \mathrm{P}\left(\boldsymbol{x_j} \text{ fails at } y_{(2)} | R_{(2)}\right) \\ &\approx \sum_{j \in R_{(2)}} (y_{(3)} - y_{(2)}) a_{(2)} e^{\boldsymbol{\beta}' \boldsymbol{x_j}} \\ &= (y_{(3)} - y_{(2)}) a_{(2)} \sum_{j \in R_{(2)}} e^{\boldsymbol{\beta}' \boldsymbol{x_j}} \end{aligned}$$

Use $\hat{\beta}$ (the mle derived from the partial likelihood) to derive

$$a_{(2)} = \frac{1}{(y_{(3)} - y_{(2)}) \sum_{j \in R_{(2)}} e^{\hat{\beta}' \boldsymbol{x_j}}}$$

Thus, the baseline hazard probability between $y_{(i)}$ and $y_{(i+1)}$ is

$$\alpha_{(i)} = a_{(i)} \cdot (y_{(i+1)} - y_{(i)}) = \frac{1}{\sum_{j \in R_{(i)}} e^{\boldsymbol{\beta}' \boldsymbol{x_j}}}$$

Remark: For continuous survival data, a more mathematical expression of this estimate is

$$\hat{\lambda}_0(t)dt = \frac{1}{\sum_{j \in R(t)} e^{\hat{\beta}' \boldsymbol{x_j}}}$$

where 'dt' means the length of a short interval (which is $y_{(i+1)} - y_{(i)}$).

To estimate the baseline cumulative hazard probability, $\Lambda_0(t)$, you just add up the estimates of the baseline hazard probabilities within the window $[0, t]$

$$\hat{\Lambda}_0(t) = \sum_{\{i:\ y_{(i)} < t\}} \frac{1}{\sum_{j \in R(y_{(i)})} e^{\hat{\beta}' \boldsymbol{x_j}}}$$

Remark: Based on the baseline cumulative hazard function $\hat{\Lambda}_0(t)$, we can obtain

$$\hat{S}(t;\boldsymbol{x_i}) = \exp\{-\hat{\Lambda}(t;\boldsymbol{x_i})\} = \exp\{-e^{\hat{\beta}'\boldsymbol{x_i}}\hat{\Lambda}_0(t)\}$$

which can be used for risk prediction.

Remark: Although the estimate of the cumulative hazard probability described above is statistically accurate when the sample size is large, the Breslow's estimate of the hazard function can be greatly improved by smoothing techniques such as kernel or spline estimation. For example, STATA uses the kernel smoothing technique to estimate the baseline hazard function:

$$\hat{\lambda}_0(y) = \frac{1}{b} \cdot \sum_{i=1}^{K} \alpha_{(i)} \cdot \phi\left(\frac{y - y_{(i)}}{b}\right)$$

where the function ϕ is a kernel function and b is a bandwidth for the kernel function.

5.2 Model fitting for PHM

Time-independent x

There are many approaches established in the literature. Here we describe only one specific approach.

Suppose we want to check on the validity of proportional hazards model. In the case that x is one-dimensional, an approach of goodness-of-fit is to partition the x-axis into K intervals, compute a separate Kaplan-Meier estimate for each interval, then apply the 2-sample goodness-of-fit procedures. When the time-independent covariate $\boldsymbol{x}$ is multi-dimensional, we consider the following approach. Define

$$\Lambda_{\boldsymbol{x_i}}(T_i) = e^{\beta' \boldsymbol{x_i}} \int_0^{T_i} \lambda_0(u) du$$

Since $\Lambda_{\boldsymbol{x_i}}(T_i)$ is monotonic in T_i, conditioning on $\boldsymbol{x_i}$, we have

$$\begin{aligned} \mathrm{P}(\Lambda_{\boldsymbol{x_i}}(T_i) > t) &= \mathrm{P}(T_i > \Lambda^{-1}_{\boldsymbol{x_i}}(t)) \\ &= \exp(-\Lambda_{\boldsymbol{x_i}}(\Lambda^{-1}_{\boldsymbol{x_i}}(t))) \\ &= e^{-t} \end{aligned}$$

Note that the distribution of the random variable $\Lambda_{\boldsymbol{x_i}}(T_i)$ is independent of $\boldsymbol{X_i} = \boldsymbol{x_i}$. Thus, unconditional on $\boldsymbol{X_i} = \boldsymbol{x_i}$, the distribution of $\Lambda_{\boldsymbol{X_i}}(T_i)$ is independent of $\boldsymbol{X_i}$ and follows Exponential($\theta = 1$) distribution. Further, the independence between $\Lambda_{\boldsymbol{X_i}}(T_i)$ and $\Lambda_{\boldsymbol{X_i}}(C_i)$ is implied by (i) independence between T_i and C_i conditioning on $\boldsymbol{X_i}$, and (ii) the fact that the distributions of $\Lambda_{\boldsymbol{X_i}}(T_i)$ and $\Lambda_{\boldsymbol{X_i}}(C_i)$ are independent of $\boldsymbol{X_i}$.

Further, $(\Lambda_{\boldsymbol{x_1}}(y_1), \delta_1)$,. . ., $(\Lambda_{\boldsymbol{x_n}}(y_n)), \delta_n)$ form a survival data sample with independent censoring. Because $\Lambda_{\boldsymbol{x_i}}(y_i)$ depends on β and $\lambda_0(t)$, substitute the corresponding estimates and define

$$\hat{\Lambda}_i = \hat{\Lambda}_{\boldsymbol{x_i}}(y_i) = e^{\hat{\beta}'\boldsymbol{x_i}} \int_0^{y_i} \hat{\lambda}_0(u)du \ .$$

Let $\hat{S}(t)$ be the Kaplan-Meier estimate based on $(\hat{\Lambda}_1, \delta_1)$,. . ., $(\hat{\Lambda}_n, \delta_n)$. Under the proportional hazards model, log$S(t) = -t$ is a linear function of t. To verify the validity of the proportional hazards model, check if

$$\frac{t}{log\hat{S}(t)} = -1$$

is approximately satisfied.

Time-dependent $\boldsymbol{x}(t)$. When the covariate $\boldsymbol{x}(t)$ is time-dependent, the above techniques no longer work for goodness-of-fit. There is a large literature regarding how to construct tests to verify the proportional hazards model assumptions. The so-called 'Martingale residuals' or 'Shoenfeld's residual' are used as the fundamental statistics for constructing the tests. For continuous survival data, define a 'residual' at $y_{(i)}$ as

$$\begin{aligned} r_{(i)} &= \boldsymbol{x_{(i)}}(y_{(i)}) - \frac{\sum_{j \in R_{(i)}} \boldsymbol{x_j}(y_{(i)}) \exp(\beta' \boldsymbol{x_j}(y_{(i)}))}{\sum_{k \in R_{(i)}} e^{\beta' \boldsymbol{x_k}(y_{(i)})}} \\ &= \boldsymbol{x_{(i)}}(y_{(i)}) - \mathrm{E}[\text{covariate at } y_{(i)} \mid R_{(i)}] \end{aligned}$$

Each residual term has 0 expectation. Thus, after replacing β by $\hat{\beta}$, the corresponding residual plot should reflect this specific feature.

4.2.7 Risk prediction

Risk prediction using baseline covariates (or biomarkers)

Risk prediction/estimation is an important topic for precision medicine.

- Risk prediction based on the PHM: With a new patient's covariates (or biomarkers) $\boldsymbol{x}$, we are interested in the individual's survival probability beyond t, that is, estimating $S(t;\boldsymbol{x})$ by

$$\hat{S}(t;\boldsymbol{x}) = \exp\{-\hat{\Lambda}(t;\boldsymbol{x})\} = \exp\{-e^{\hat{\beta}'\boldsymbol{x}} \times \hat{\Lambda}_0(t)\} \ .$$

  where $\hat{\beta}$ is the partial likelihood estimate and $\hat{\Lambda}_0(t)$ is the Breslow estimate; see slides 5 and 6.

- Prediction using the PHM is a model-based approach. When data are 'big', completely nonparametric or machine-learning methods are also available for prediction.

Landmark model prediction

For various reasons, in some studies it is desired to use biomarkers at pre-specified time points: $0 = s_0 < s_1 < \ldots < s_J = \infty$, and the **landmark hazards model**:

$$\lambda(t|\boldsymbol{x}(s_k)) = \lambda_0(t)e^{\boldsymbol{\beta}_k' \boldsymbol{x}(s_k)}, \; s_k \le t < s_{k+1}$$

- The regression parameter vector $\boldsymbol{\beta}_k$ is updated at each landmark time s_k.
- Unlike the PHM with time-dependent covariates, at pre-specified landmark times s_k, the landmark hazards model uses updated marker information $\boldsymbol{x}(s_k)$ for prediction. Methods from Cox regression model can still be used for estimation of $\boldsymbol{\beta}_k$ and $\lambda_0(t)$ as seen in the next slide.

- Compared with the PHM with time-dependent covariates, the landmark hazards model only needs marker information at pre-specified time points s_k. Methods from Cox regression model can still be used for estimation of $\boldsymbol{\beta_k}$ and $\lambda_0(t)$. For $k = 0, 1, 2, ..., J-1$, the kth partial likelihood

$$L_k = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta_k' x_i}(s_k)}}{\sum_{j \in R(y_i)} e^{\boldsymbol{\beta_k' x_j}(s_k)}} \right)^{I(\delta_i = 1, s_k \le y_i < s_{k+1})}$$

can be used to estimate $\boldsymbol{\beta_k}$. The baseline hazard function can be estimated by

$$\hat{\lambda}_0(t)dt = \frac{1}{\sum_{j \in R(t)} e^{\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \boldsymbol{x_j}(s_k)}}, \qquad s_k \le t < s_{k+1}.$$

- **Risk prediction:**
e.g., T: age onset of AD. Using marker info at age 60 to predict AD at age 60∼70; using marker info at age 70 to predict AD at age 70∼80, etc.

Using a new survivor's covariates at time s_k, $\boldsymbol{x}(s_k)$, we can predict this individual's probability to survive beyond time t, where $s_k \le t < s_{k+1}$. That is, define the conditional survival function $S^{con}(t; \boldsymbol{x}(s_k)) = P(T \ge t | T \ge s_k, \boldsymbol{x}(s_k))$, $t \in [s_k, s_{k+1})$. Then, given survivorship at time s_k, the risk probability for this individual to fail within next u time units is $1 - \hat{S}^{con}(s_k + u; \boldsymbol{x}(s_k))$. we can estimate $S^{con}(s_k + u; \boldsymbol{x}(s_k))$, $0 < u \le s_{k+1} - s_k$ by

$$\hat{S}^{con}(s_k + u; \boldsymbol{x}(s_k)) = \exp\{-e^{\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \boldsymbol{x}(s_k)} \times \int_{s_k}^{s_k+u} \hat{\lambda}_0(u) du\}$$
$$= \exp\left\{-e^{\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \boldsymbol{x}(s_k)} \times \sum_{\{i:\ s_k \le y_{(i)} < s_k + u\}} \frac{1}{\sum_{j \in R(y_{(i)})} e^{\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}' \boldsymbol{x}_{\boldsymbol{j}}(s_k)}}\right\}$$

where $\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}'$ is the partial likelihood estimate and $\hat{\lambda}_0(u)$ is the Breslow estimate derived from the data set we have in hand.

* The so-called 'survivors at t' refer to those subjects who did not exerience the failure event before t.