Notes for Biostatistics 140.641.01 'Survival Analysis'

First Term, 2022

Chapter 4. Regression Models

Two types of regression models are frequently used to describe the association between covariates (independent variables) and a failure time outcome T (dependent variable):

- Section 4.1 AFTM: Modeling relationship between covariates and failure time T.
- Section 4.2 PHM: Modeling relationship between covariates and the hazard function of T. Chapter 4 has its focus on PHM.

The main focus of this chapter is PHM.

4.1 Accelerated Failure Time Model (AFTM)

Suppose we have a sample of n subjects from a target population. Denote by $\boldsymbol{x_i}' = (x_{i1}, \ldots, x_{ip})$ the vector of covariates for subject i.

- The AFTM is specified by the following equation

$$\begin{aligned} \log(T_i) &= \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \\ &= \beta_0 + \boldsymbol{\beta}' \boldsymbol{x_i} + \epsilon_i \end{aligned}$$

  where β_0 and $\boldsymbol{\beta}' = (\beta_1, \cdots, \beta_p)$ are the regression parameters.

- ϵ_i is the random disturbance variable, which is usually assumed to have 0 mean, variance σ^2 and pdf $f(\epsilon)$, and is independent of $\boldsymbol{x_i}$.

- Logarithm of T_i is a frequently adopted transformation. Other monotonic transformations could also be used in applications.

- An alternative way to formulate the AFTM is

$$T_i = T_{0i} \cdot e^{\boldsymbol{\beta}' \boldsymbol{x_i}} \iff \log T_i = \boldsymbol{\beta}' \boldsymbol{x_i} + \log T_{0i}, \quad \log T_{0i} \sim S_0$$

  Where the distribution of T_{0i} is independent of $\boldsymbol{x_i}$. This model is equivalent to the AFTM from last slide if $\{T_{0i}\}$ are iid, where $\mathrm{E}[\log T_{0i}] = \beta_0$. Let $\epsilon_i = \log T_{0i} - \beta_0$, then $\log T_i = \beta_0 + \boldsymbol{\beta}' \boldsymbol{x_i} + \epsilon_i$

- The larger value of $\beta_k x_{ik}$, $k = 1, \ldots, p$, is more beneficial in terms of improving overall survival time.
- if we increase the covariate value of x_{ik}, $k = 1, \ldots, p$, by one unit while holding other covariate values fixed for the ith subject, the expected value of $\log(T_i)$ is changed by β_k.

Parametric AFTMs

- For the linear model $\log T_i = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x_i} + \epsilon_i$, we can assume a parameterized distribution for the disturbance term ϵ. For example, assume $\epsilon_i \sim Normal(0, \sigma^2)$. This assumption is equivalent to assuming that T_i has log-normal distribution; that is $T_i \sim \text{Lognormal}(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x_i},\ \sigma^2)$.
- Suppose we have a sample of n subjects from a target population. For subject i, we observed covariates $x_{i1}, \cdots, x_{ip}$ and possibly censored survival time Y_i and censoring indicator Δ_i. Assume the independent censoring condition: T_i and C_i are independent conditionally on covariates $\boldsymbol{X_i} = \boldsymbol{x_i}$. The maximum likelihood estimation approach is frequently used for estimating the regression parameters and deriving inferential results. For example, if $\epsilon_i \sim N(0, \sigma^2)$, then $\log T_i \sim N(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x_i}, \sigma^2)$. Let f and S respectively be the pdf and survival function for $N(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x_i}, \sigma^2)$. Base on the observed data $(y_1, \delta_1), ..., (y_n, \delta_n)$, the likelihood function is

$$L = \prod_{i=1}^{n} f(\log y_i)^{\delta_i} S(\log y_i)^{1-\delta_i}$$

4.2 Proportional hazards model (PHM)

The proportional hazards model (PHM) is the focused interest of this chapter. Assume covariates are available on each individual

$$\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots, x_{ip})^t.$$

The PHM assumes

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

where $\boldsymbol{x_i}$ is $p \times 1$ vector of covariates and $\boldsymbol{\beta}$ is a $1 \times p$ vector of parameters. Interpretation of the model:

Hazard at t for given x_i = (baseline hazard at t) $\times$ (risk function $e^{\boldsymbol{\beta' x_i}}$)

Characteristics of PHM:

1. The PHM is a model on the basis of hazard function, in contrast with the AFTM based on survival time:

$$\text{PHM:} \quad \lambda(t; \boldsymbol{x_i}) = \lambda_0(t) e^{\boldsymbol{\beta' x_i}}$$

$$\text{AFTM:} \quad T_i = T_{0i} \cdot e^{\boldsymbol{\beta' x_i}}$$

2. The baseline hazard $\lambda_0(t)$ is left unspecified (nonparametric), thus the PHM is a semiparametric model: $\lambda_0 =$ nonparametric component, β: parametric component.

3. In most applications related to public health, the parameter β is of primary interest and $\lambda_0(t)$ is of secondary interest. However, estimation of $\lambda_0(t)$ is becoming more important in recent years, since it predicts absolute risk (hazard) for an individual with grouping covariates $\boldsymbol{x_i}$. Essentially, both β and $\lambda_0(t)$ are important for the task of risk prediction.

Interpretation of $\boldsymbol{\beta}$

Example. T: time from onset of treatment to clinically defined AIDS (using definition before Jan. 1993)

x_{i1}: CD4 count measured at baseline for subject i

x_{i2}: gender, 0 for female and 1 for male

Consider the PHM: $\lambda(t;\boldsymbol{x_i}) = \lambda_0(t)e^{\beta_1 x_{i1}+\beta_2 x_{i2}}$

Hazard ratio (or relative hazard) for subjects i and k at time t:

$$\begin{aligned} \text{H.R.}(t) &= \frac{\lambda(t;\boldsymbol{x_i})}{\lambda(t;\boldsymbol{x_k})} = \frac{\lambda_0(t)e^{\boldsymbol{\beta}'\boldsymbol{x_i}}}{\lambda_0(t)e^{\boldsymbol{\beta}'\boldsymbol{x_k}}} = e^{\boldsymbol{\beta}'(\boldsymbol{x_i}-\boldsymbol{x_k})} \\ &= e^{\beta_1(x_{i1}-x_{k1})+\beta_2(x_{i2}-x_{k2})} \end{aligned}$$

Note that the hazard ratio is time-invariant with respect to t.

With $x_{i2} = x_{k2}$ (same gender), if $\beta_1 = -0.01$, $x_{i1} = 250$, $x_{k1} = 200$, then

$$\text{H.R.}(t) = e^{-0.01 \times (250-200)} = e^{-0.5} \approx 0.61.$$

Thus, conditioning on the same gender and being AIDS-free prior to time t, the probability that subject i is diagnosed with AIDS at t is 61% of the probability that subject k is diagnosed with AIDS at t.

In general, when we change the covariate value of x_{i1} by one unit while holding x_{i2} fixed, the corresponding hazard will be changed by the proportion e^{β_1}.

4.2.1 PHM as Lehmann's Alternatives

The PHM can also be expressed as

$$S(t; \boldsymbol{x_i}) = S_0(t)^{e^{\boldsymbol{\beta' x_i}}}$$

Proof.
$$\begin{aligned} S(t; \boldsymbol{x_i}) &= e^{-\int_o^t \lambda(u;\boldsymbol{x_i})du} \\ &= e^{-\int_o^t \lambda_0(u) e^{\boldsymbol{\beta' x_i}} du} \\ &= e^{[-\int_o^t \lambda_0(u)du] \cdot e^{\boldsymbol{\beta' x_i}}} \\ &= S_0(t)^{e^{\boldsymbol{\beta' x_i}}}. \end{aligned}$$

where $S_0(t) = e^{-\int_o^t \lambda_0(u)du}$ is the baseline survival function.

We say that a class of distributions with the form $S(t) = S_0(t)^\gamma$ for some positive γ is a family of "Lehmann's alternatives". Clearly, the PHM implies that the distribution functions form a family of "Lehmann's alternatives". The PHM is a very flexible model because of its semiparametric feature, but the constant proportionality of the model is not automatic and needs to be confirmed by proper methods.

Example. A two-sample case

$$x = \begin{cases} 0 & \text{controlled} \\ 1 & \text{treated} \end{cases}$$

Under the PHM,

$$\lambda(t; x) = \lambda_0(t) e^{\beta x}.$$

That is

$$\lambda_1(t) = \lambda_0(t) e^{\beta}.$$

Using Lehmann's alternative expression, we derive

$$\begin{aligned} S_1(t) &= S_0(t)^{e^{\beta}} \\ \log S_1(t) &= e^{\beta} \times \log S_0(t) = \text{positive constant } \times \log S_0(t) \end{aligned}$$

For exploratory analysis, to examine the validity of the PHM for two-sample case, we can use the K-M estiamtes $\hat{S}_1$ and $\hat{S}_0$ to see if

$$\phi(t) = \frac{\log \hat{S}_1(t)}{\log \hat{S}_0(t)} \approx \text{ positive constant.}$$

The PHM is a valid model if $\phi(t)$ remains a positive constant over time.

4.2.2 Partial Likelihood

Assume independent censoring condition: Conditioning on $\boldsymbol{X_i} = \boldsymbol{x_i}$, T_i is independent of C_i.

Assume the PHM

$$\lambda(t; \boldsymbol{x_i}) = \lambda_0(t) e^{\beta_1 x_{i1} + \cdots + \beta_p x_{ip}} = \lambda_0(t) e^{\boldsymbol{\beta' x_i}}$$

Data : $(y_1, \delta_1, \boldsymbol{x_1}), \cdots, (y_n, \delta_n, \boldsymbol{x_n})$

$y_i =$ observed survival time

$\delta_i =$ censoring indicator

$\boldsymbol{x_i} =$ vector of covariates

$H(t^-) =$ data history up to t, but not including t

Risk set at t, $R(t) = \{(j, \boldsymbol{x_j}) : \ y_j \geq t\}$

(Risk set provides information so that you know who are at risk at t)

Assume survival times are not tied. The likelihood function is

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} f(y_i; \boldsymbol{x_i})^{\delta_i} S(y_i; \boldsymbol{x_i})^{1-\delta_i} \\
&= \prod_{y_{(i)}} p(x_{(i)} | H(y_{(i)}^-), y_{(i)}) \times \{\text{residual likelihood}\} \\
&= \left\{ \prod_{y_{(i)}} \left[\frac{e^{\boldsymbol{\beta}' \boldsymbol{x_{(i)}}}}{\sum_{j \in R(y_{(i)})} e^{\boldsymbol{\beta}' \boldsymbol{x_j}}} \right] \right\} \times \{\text{residual likelihood}\}
\end{aligned}$$

where $R(y_{(i)})$ = Risk set at the uncensored failure time $y_{(i)}$, and $\boldsymbol{x_{(i)}}$ = covariates corresponding to $y_{(i)}$. Note that this likelihood can also be expressed as

$$\mathcal{L} = \left\{ \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta}' \boldsymbol{x_i}}}{\sum_{j \in R(y_i)} e^{\boldsymbol{\beta}' \boldsymbol{x_j}}} \right)^{\delta_i} \right\} \times \{\text{residual likelihood}\}$$

The first likelihood is called the "partial likelihood". Cox (1972, JRSS-B; 1975, Biometrika) identified the above likelihood structure. Thus the partial likelihood method is also referred to as Cox's method.
The result is great!! Why?

- The result is derived under an attractive/flexible semiparametric model with desired interpretations in hazards.
- The partial likelihood only involves β!! It does not involve $\lambda_0(t)$, and thus computation of $\hat{\beta}$ is manageable and inferences can be developed.

Methodological ideas of partial likelihood?

Assume no ties in the uncensored survival times. Let L_p = The partial likelihood. Any "likelihood" must correspond to a probability (or density) of some kind. Note that

$$\begin{aligned} & \text{P}\left(\text{individual } x_{(i)} \text{ fails at } y_{(i)} \;\middle|\; \begin{array}{l}\text{a failure occurring at } y_{(i)} \text{ and} \\ \text{data history before } (<) y_{(i)}\end{array}\right) \\ = \;& \text{P}\left(x_{(i)} \text{ fails at } y_{(i)} \middle| \text{ a failure occurring at } y_{(i)} \text{ and } R_{(i)}\right) \\ = \;& \frac{\lambda_0(y_{(i)})e^{\boldsymbol{\beta}'\boldsymbol{x_{(i)}}}}{\sum_{j\in R_{(i)}} \lambda_0(y_{(i)})e^{\boldsymbol{\beta}'\boldsymbol{x_j}}} = \frac{e^{\boldsymbol{\beta}'\boldsymbol{x_{(i)}}}}{\sum_{j\in R_{(i)}} e^{\boldsymbol{\beta}'\boldsymbol{x_j}}} \end{aligned}$$

Thus, the "partial likelihood" is

$$\begin{aligned} L_p &= \prod_{y_{(i)}} \mathrm{P}(\boldsymbol{x_{(i)}} \text{ fails at } y_{(i)} | \text{a failure occurring at } y_{(i)}, R(y_{(i)})) \\ &= \prod_{y_{(i)}} \left(\frac{e^{\boldsymbol{\beta' x_{(i)}}}}{\sum_{j \in R(y_{(i)})} e^{\boldsymbol{\beta' x_j}}} \right) \\ &= \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta' x_i}}}{\sum_{j \in R(y_i)} e^{\boldsymbol{\beta' x_j}}} \right)^{\delta_i} \end{aligned}$$

Note that the partial likelihood can also be expressed as

$$L_p = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta' x_i}}}{\sum_{j \in R(y_i)} e^{\boldsymbol{\beta' x_j}}} \right)^{\delta_i}$$

Derive the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ by maximizing L_p over possible values of $\boldsymbol{\beta}$.

Example. Two-sample case

Control group: 7, 9^+, 18

Treatment group: 12, 19^+

$$x = \begin{cases} 0 & \text{in control gr.} \\ 1 & \text{in treatment gr.} \end{cases}$$

$$\underline{\text{PHM}} : \lambda(t;x) = \lambda_0(t)e^{\beta' x}$$

The partial likelihood is

$$\begin{aligned} L_p &= \left[\frac{e^{0\beta}}{e^{0\beta}+e^{0\beta}+e^{0\beta}+e^{\beta}+e^{\beta}}\right]\left[\frac{e^{\beta}}{e^{0\beta}+e^{\beta}+e^{\beta}}\right]\left[\frac{e^{0\beta}}{e^{0\beta}+e^{\beta}}\right] \\ &= \left[\frac{1}{3+2e^{\beta}}\right]\left[\frac{e^{\beta}}{1+2e^{\beta}}\right]\left[\frac{1}{1+e^{\beta}}\right] \end{aligned}$$

Obtain the mle $\hat{\beta}$ by maximizing L_p.

Real data example.

	Hazard ratio (95% CI)	p*
Univariate Cox regression models		
Treatment (nevirapine *vs* zidovudine)	0·59 (0·41–0·84)†	0·0039
Maternal HIV-1 RNA ($\log_{10}$) at pre-entry‡	2·24 (1·73–2·89)	<0·0001
Maternal CD4 at pre-entry§	1·29 (1·18–1·40)	<0·0001
Birthweight¶	1·16 (0·94–1·42)	0·16
Gender (female *vs* male)	1·09 (0·77–1·56)	0·63
Duration of labour (h)	1·00 (0·98–1·03)	0·92
Prolonged rupture of membrane (>4 h) (yes *vs* no)	1·11 (0·67–1·83)	0·69
Caesarean section (yes *vs* no)	0·79 (0·43–1·43)	0·43
Multivariate Cox regression		
Treatment (nevirapine *vs* zidovudine)	0·57 (0·40–0·83)	0·0033
Maternal HIV-1 RNA ($\log_{10}$) at pre-entry‡	1·81 (1·36–2·40)	<0·0001
Maternal CD4 at pre-entry§	1·19 (1·09–1·31)	<0·0001

*Computed from a Wald statistic. †Corresponding to a relative efficacy of 0·41 (95% CI 0·16–0·59). ‡For a unit increase of log, HIV-1 RNA copies/mL. §For a decrease of 100 cells/μL. ¶For a decrease of 500 g.

Table 3: **Prognostic factors for HIV-1 infection**

The table was published in the article, "Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: 18-month follow-up of the HIVNET 012 randomized trial" (by Jackson, et al., 2003, the Lancet, 362, pp. 859-868).

Outcome variable T: Child's age at incidence of HIV infection
What can be concluded from the table?

- Adjusted for maternal HIV-1 RNA viral loads and CD4+ counts at pre-entry, single-dose nevirapine would reduce the hazard ratio of HIV-1 infection incidence to 57% .
- Adjusted treatment effect (RH $= 0.57$) is similar to unadjusted treatment effect (RH $= 0.59$).
- Cautious: Interaction terms are not considered in analysis.
- Conceptually, in a randomized study, we do not need to adjust for confounding factors for identifying treatment effect. We do adjusted analysis here for various reason: study is not completely randomized (sampling bias), informative censoring,...

4.2.3 Generalization to Time-Dependent Covariates

There are situations that some the covariates can be time-dependent. For example,

- biomarker value changes over time
- treatment plan changes over time
- drug dosage level changes over time
- or a transformation of the above time-dependent measurements

Time-dependent covariates for the i$^{\text{th}}$ individual are

$$\boldsymbol{x_i}(t) = (x_{i1}(t), x_{i2}(t), \ldots, x_{ip}(t))$$

We shall use the general notation $\boldsymbol{x_i}(t)$ instead of $\boldsymbol{x_i}$, even though some of the covariates are time-independent. The PHM is now

$$\lambda(t; \boldsymbol{x_i}(t)) = \lambda_0(t) e^{\boldsymbol{\beta}' \boldsymbol{x_i}(t)}.$$

Remark: A more rigorous way to write the time-dependent covariates PHM is

$$\lambda(t; \boldsymbol{x_i^H}(t)) = \lambda_0(t) e^{\boldsymbol{\beta}' \boldsymbol{z_i}(t)},$$

where $\boldsymbol{x_i^H}(t) = \{\boldsymbol{x_i}(u) : \ 0 \leq u \leq t\}$ is the history of covariates prior to or at t, and $\boldsymbol{z_i}(t) = \phi(\boldsymbol{x_i^H}(t))$ is a transformed covariate vector..

With time-dependent covariates, the previous partial likelihood argument still works, and the partial likelihood becomes

$$L_p = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta}' \boldsymbol{x_i}(y_i)}}{\sum_{j \in R(y_i)} e^{\boldsymbol{\beta}' \boldsymbol{x_j}(y_i)}} \right)^{\delta_i}$$

Interpretation of $\boldsymbol{\beta}$

Example. T: time from study entry to diagnosis of AIDS (using definition before Jan. 1993)

$x_{i1}(t)$: CD4 count measured at failure time t for subject i

x_{i2}: gender, 0 for female and 1 for male

Consider the PHM: $\lambda(t; \boldsymbol{x_i}(t)) = \lambda_0(t)e^{\beta_1 x_{i1}(t)+\beta_2 x_{i2}}$.

Then, the hazard ratio for subjects i and k at time t is

$$\text{Hazard ratio at } t = \frac{\lambda(t; \boldsymbol{x_i}(t))}{\lambda(t; \boldsymbol{x_k}(t))} = \frac{\lambda_0(t)e^{\boldsymbol{\beta}' \cdot \boldsymbol{x_i}(t)}}{\lambda_0(t)e^{\boldsymbol{\beta}' \cdot \boldsymbol{x_k}(t)}} = e^{\boldsymbol{\beta}'(\boldsymbol{x_i}(t) - \boldsymbol{x_k}(t))}$$

With x_{i2} fixed (same gender) and $\beta_1 = -0.01$,
for $x_{i1}(t_1) = 250$, $x_{k1}(t_1) = 200$, then

$$\text{H.R.}(t_1) = e^{-0.01\times(250-200)} = e^{-0.5} \approx 0.61.$$

for $x_{i1}(t_2) = 225$, $x_{k1}(t_2) = 200$, then

$$\text{H.R.}(t_2) = e^{-0.01\times(225-200)} = e^{-0.25} \approx 0.78.$$

Thus, conditioning on being AIDS-free prior to t_1, the probability that subject i is diagnosed with AIDS at t_1 is 61% of the probability that subject k is diagnosed with AIDS at t_1.

Conditioning on being AIDS-free prior to t_2, the probability that subject i is diagnosed with AIDS at t_2 is 78% of the probability that subject k is diagnosed with AIDS at t_2.

How much time-dependent covariate information is needed for computation?

Example. T: time from study entry to diagnosis of AIDS

$x_{i1}(t)$: CD4 count measured at failure time t for subject i

x_{i2}: age at entry

x_{i3}: gender, 0 for female and 1 for male

Consider the PHM: $\lambda(t; \boldsymbol{x_i}(t)) = \lambda_0(t) e^{\beta_1 x_{i1}(t) + \beta_2 x_{i2} + \beta_3 x_{i3}}$.

The partial likelihood is

$$L_p = \prod_{i=1}^{n} \left(\frac{e^{\beta_1 x_{i1}(y_i) + \beta_2 x_{i2} + \beta_3 x_{i3}}}{\sum_{j \in R(y_i)} e^{\beta_1 x_{j1}(y_i) + \beta_2 x_{j2} + \beta_3 x_{j3}}} \right)^{\delta_i}$$

Suppose the observed data are

Male

I.D.	001	002
age at entry	10	12
y_i or y_i^+	12	19^+

Female

I.D.	003	004	005
age at entry	4	0	11
y_i or y_i^+	7	9^+	18

Time-dependent CD4 count

	I.D./ uncensored y_i	7	12	18
$x_{i1} = 1$	001	310	300	
	002	250	250	260
	003	420		
$x_{i1} = 0$	004	500		
	005	550	450	430

Note: Computer needs the above "time-dependent covariate" data information measured at all the uncensored failure times for time-dependent PHM analysis.

$$L_p = \left[\frac{e^{\beta_1 \cdot 420 + \beta_2 \cdot 4 + \beta_3 \cdot 0}}{e^{\beta_1 \cdot 310 + \beta_2 \cdot 10 + \beta_3 \cdot 1} + e^{\beta_1 \cdot 250 + \beta_2 \cdot 12 + \beta_3 \cdot 1} + \ldots + e^{\beta_1 \cdot 550 + \beta_2 \cdot 11 + \beta_3 \cdot 0}} \right]$$

$$\cdot \left[\frac{e^{\beta_1 \cdot 300 + \beta_2 \cdot 10 + \beta_3 \cdot 1}}{e^{\beta_1 \cdot 300 + \beta_2 \cdot 10 + \beta_3 \cdot 1} + e^{\beta_1 \cdot 250 + \beta_2 \cdot 12 + \beta_3 \cdot 1} + e^{\beta_1 \cdot 450 + \beta_2 \cdot 11 + \beta_3 \cdot 0}} \right]$$

$$\cdot \left[\frac{e^{\beta_1 \cdot 430 + \beta_2 \cdot 11 + \beta_3 \cdot 0}}{e^{\beta_1 \cdot 260 + \beta_2 \cdot 12 + \beta_3 \cdot 1} + e^{\beta_1 \cdot 430 + \beta_2 \cdot 11 + \beta_3 \cdot 0}} \right] .$$

What if we use 'time-dependent age' $x_{i2}(t)$ instead of the baseline age x_{i2} in the PHM?

Answer: **It makes no difference!** Using the baseline age x_{i2} or time-dependent age $x_{i2}(t)$ as a linear term in the proportional hazards model would end up with the same partial likelihood estimate $\hat{\beta}_2$ because

$$\begin{aligned} \lambda_0(t)e^{\beta_1 x_{i1}(t)+\beta_2 x_{i2}(t)+\beta_3 x_{i3}} &= \lambda_0(t)e^{\beta_1 x_{i1}(t)+\beta_2 (x_{i2}+t)+\beta_3 x_{i3}} \\ &= \lambda_0^*(t)e^{\beta_1 x_{i1}(t)+\beta_2 x_{i2}+\beta_3 x_{i3}} \end{aligned}$$

where $\lambda_0^*(t) = \lambda_0(t)e^{\beta_2 t}$ is also a baseline hazard function (i.e., a covariate-free function).

Practicality in real data analysis:

When time-dependent covariates are not available at ALL uncensored times, a conventional approach is to use the "carry over to the next time-point" approach; the is, use of the most recently measured covariate information. This approach is commonly used for applications but has the disadvantage from using out-of-date information.

Time-dependent treatment effect for binary x

When the PHM is used to identify the treatment effect in a clinical trial. A standard model formulation is $\lambda(t;x) = \lambda_0(t)e^{\beta x}$, where $x(=0,1)$, is the treatment indicator. The hazard ratio of the treatment group to the control group from this PHM is e^{β}.

Time-dependent covariate PHM can be used as a tool to model time-dependent treatment effect. For example, the hazard ratio of the treatment group to the control group from the PHM $\lambda(t;x) = \lambda_0(t)e^{(\beta_0+\beta_1 t)x}$ is $e^{\beta_0+\beta_1 t}$, which is time-dependent. For computation, **one can take advantage of time-dependent covariate PHM techniques** to conduct the analysis: Express the model by

$$\lambda(t;x) = \lambda_0(t)e^{\beta_0 x+\beta_1(t\times x)} = \lambda(t;x) = \lambda_0(t)e^{\boldsymbol{\beta'}\boldsymbol{x}(t)}$$

where $\boldsymbol{x}(t) = (x, t\times x)'$ and $\boldsymbol{\beta'} = (\beta_0, \beta_1)$.

In some studies it is designed to collect covariate or biomarker information at pre-specified time points: $0 = s_0 < s_1 < \ldots < s_J = \infty$, and the **landmark hazards model**:

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

- Estimation methodology for the landmark hazards model is similar to the time-dependent covariate PHM: Treat $\boldsymbol{x_i}(t) = \boldsymbol{x_i}(s_k)$ for $s_k \leq t < s_{k+1}$:

$$\lambda(t; \boldsymbol{x_i}(t)) = \lambda_0(t) e^{\boldsymbol{\beta'} \boldsymbol{x_i}(t)}.$$

  and the partial likelihood method can be applied.

- Ideally we could prefer using updated $\boldsymbol{\beta'_k}$ instead of $\boldsymbol{\beta'}$ in the above PHM. We will discuss this with more details for the topic on risk prediction in next chapter.

4.2.4 Models for Discrete Survival Data

The partial likelihood approach so far does not handle tied survival data. When we analyze discrete or grouped survival data, the problem of how to analyze such data naturally arises.

I. Proportional hazards Model

$$\lambda(t_j; \boldsymbol{x_i}(t_j)) = \lambda_0(t_j) e^{\beta' \boldsymbol{x_i}(t_j)} \quad j = 1, \ldots, J; \; i = 1, \ldots, n$$

where $t_1, \ldots, t_k$ are discrete time points of T. Note that, theoretically, the hazard probabilities $\lambda(t_j; \boldsymbol{x_i}(t_j))$ and $\lambda_0(t_j)$ should be bounded between 0 and 1. For discrete survival data, this model has nice and direct interpretation but it might encounter the problem that the estimated hazard probability exceeds upper bound 1.

Example. Consider the following simple PHM for discrete survival time: $\lambda(t; x_i) = \lambda_0(t) e^{\beta' x_i}$,

Control group	7	9^+	18	$x_1, x_2, x_3 = 0$
Treatment group	18	19^+		$x_4, x_5 = 1$

Recall the partial likelihood construction is motivated by

$\mathrm{P}(x_{(i)}$ fails at $y_{(i)} |$ a failure occurring at $y_{(i)},\ R_{(i)})$.

Now, at $y_{(2)} = 18$, the probability becomes

$$\begin{aligned} & \mathrm{P}(x_3 \text{ and } x_4 \text{ fail at } 18 \,|\text{two failures at } 18, \text{ risk set at } 18 = \{x_3, x_4, x_5\}) \\ = & \frac{\lambda_0(18)e^{\beta \cdot x_3} \cdot \lambda_0(18)e^{\beta \cdot x_4}}{\lambda_0(18)e^{\beta \cdot x_3} \cdot \lambda_0(18)e^{\beta \cdot x_4} + \lambda_0(18)e^{\beta \cdot x_4} \cdot \lambda_0(18)e^{\beta \cdot x_5} + \lambda_0(18)e^{\beta \cdot x_3} \cdot \lambda_0(18)e^{\beta' x_5}} \\ = & \frac{e^{\beta \cdot 0 + \beta \cdot 1}}{\left(e^{\beta \cdot 0 + \beta \cdot 1} + e^{\beta \cdot 1 + \beta \cdot 1} + e^{\beta \cdot 0 + \beta \cdot 1}\right)} \end{aligned}$$

The partial likelihood is

$$\begin{aligned} L_p &= \left(\frac{e^{\beta \cdot 0}}{3 \cdot e^{\beta \cdot 0} + 2 \cdot e^{\beta \cdot 1}}\right)\left(\frac{e^{\beta \cdot 0 + \beta \cdot 1}}{e^{\beta \cdot 0 + \beta \cdot 1} + e^{\beta \cdot 1 + \beta \cdot 1} + e^{\beta \cdot 0 + \beta \cdot 1}}\right) \\ &= \left(\frac{1}{3 + 2e^{\beta}}\right)\left(\frac{e^{\beta}}{2e^{\beta} + e^{2\beta}}\right) \end{aligned}$$

For the general data $(\boldsymbol{x_1}, y_1, \delta_1), (\boldsymbol{x_2}, y_2, \delta_2), \ldots, (\boldsymbol{x_n}, y_n, \delta_n)$, the partial likelihood for tied survival data is

$$L_p = \prod_{(i)} \left(\frac{e^{\sum_{j \in D_{(i)}} \boldsymbol{\beta}' \cdot \boldsymbol{x_j}(y_{(i)})}}{\sum_{\substack{\text{combinations} \\ D^*_{(i)} \subset R_{(i)}}} e^{\sum_{j \in D^*_{(i)}} \boldsymbol{\beta}' \cdot \boldsymbol{x_j}(y_{(i)})}} \right)$$

Where $D_{(i)}$ is the set of "deaths" (or failures) occurring at $y_{(i)}$, $D^*_{(i)}$ is a combination of subjects from the risk set $R_{(i)}$, with the restriction that $\#D^*_{(i)} = \#D_{(i)}$.

Remarks:

1. The hazard probability in this discrete PHM might be out of bound (> 1) – a major drawback for use of this model.

2. Computation of the mle from L_p for tied survial data might be problematic. If you have heavily tied survival data, check your computing packages to see if they handle such data.

3. Some of the computing packages use the Breslow's approach (Breslow, 1972, *Biometrics*) to handle problems with tied data. The results are reasonably accurate if you have a small proportion of ties. Here the Breslow's approach refers to: Each of a set of tied survival times is sequentially treated as though it occurred just before the others (one at a time), then use the partial likelihood approach.

II. Pooled Logistic Regression Model

Instead of the proportional hazards model, consider the pooled logistic regression model defined at different time points,

$$\frac{\lambda(t_j; \boldsymbol{x_i}(t_j))}{1-\lambda(t_j; \boldsymbol{x_i}(t_j)} = \frac{\lambda_0(t_j)}{1-\lambda_0(t_j)} e^{\beta' \cdot \boldsymbol{x_i}(t_j)} \quad j = 1, \ldots, J; \; i = 1, \ldots, n$$

where t_j, $j = 1, 2, \ldots, J$, are the discrete points of the survival time T.

There are a number of approaches developed to estimate the parameter β; see Breslow and Day (Volume 1, 1980) for details.

The pooled logistic regression model can be also expressed as

$$\frac{\lambda(t_j; \boldsymbol{x_i}(t_j))}{1 - \lambda(t_j; \boldsymbol{x_i}(t_j))} = e^{\alpha_j + \beta' \cdot \boldsymbol{x_i}(t_j)}$$

with $e^{\alpha_j} = \lambda_0(t_j)/\{1 - \lambda_0(t_j)\}$. That is, $logit\{\lambda(t_j; \boldsymbol{x_i}(t_j))\} = \alpha_j + \beta' \cdot \boldsymbol{x_i}(t_j)$ and

$$\lambda(t_j; \boldsymbol{x_i}(t_j)) = \frac{e^{\alpha_j + \beta' \cdot \boldsymbol{x_i}(t_j)}}{1 + e^{\alpha_j + \beta' \cdot \boldsymbol{x_i}(t_j)}}$$

Note that

- When $\boldsymbol{x_i}(t_j)$ is binary ($= 0$ or 1), e^{β} has the interpretation as the odds ratio of hazard probability.
- When discrete points become dense, $\lambda_0(t_j) \approx 0$ and $\lambda(t_j; \boldsymbol{x_i}(t_j)) \approx 0$, and the pooled logistic regression model approximates the proportional hazards model.
- MLE method: Create the 'local likelihood' from data in the jth risk set $\{(\boldsymbol{x_i}(t_j), I(y_i = t_j)) : y_i \in R(t_j)\}$, $j = 1, ..., J$, where a subject's binary outcome (for logistic model) is 1 if the person experiences failure at t_j, and the binary outcome is 0 otherwise. Then maximize the product of all the local likelihoods to derive estimates of α_j and β.
- If time is continuous but one only observes the data in grouped form, then the complementary log-log link model would seem more appropriate; see Appendix.

*Appendix

A few parametric AFTMs

Example. Exponential distribution model

The simplest model is the exponential model where T_{0i} (referred to as the baseline survival time) has exponential distribution with constant hazard $exp(-\beta_0)$. This is equivalent to assuming that $\sigma = 1$ and ϵ has a standard extreme value distribution with the density function $f(\epsilon) = e^{\epsilon - e^{\epsilon}}$ (So e^{ϵ} has the standard exponential distribution with constant hazard 1.).

From this specification, it is easy to see that the distribution of T_i with any covariate vector $\boldsymbol{x_i}$ is exponential with constant hazard
$\lambda(t|\boldsymbol{x_i}) = \exp\{-\beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}\}$.

Example. Weibull distribution model

The only difference between the Weibull model and the exponential model is that σ is set to be a parameter rather than being fixed to have value 1. In this case, the distribution of $\sigma\epsilon$ is an extreme value distribution with scale parameter σ.

The hazard function of T_i can be shown to be

$$\log \lambda(t|\boldsymbol{x_i}) = (\alpha - 1)\log t + \beta_0^* + x_{i1}\beta_1^* + \cdots + x_{ip}\beta_p^*,$$

where $\alpha = 1/\sigma, \beta_0^* = -\log\sigma - \beta_0/\sigma$, and $\beta_j^* = -\beta_j/\sigma$, $j = 1, \ldots, p$.

Semi-parametric transformation models (STM)

A class of semi-parametric transformation models (STM) can be viewed as a generalization of the AFTM and PHM:

$$h(T_i) = \boldsymbol{\beta}' \boldsymbol{x_i} + \epsilon_i$$

where $h(\cdot)$ is an unspecified monotone function satisfying $h(0) = -\infty$, ϵ_i is a random variable with a known distribution. Special cases of the transformation models include the proportional hazards model and the proportional odds model, where ϵ_i corresponds to the extreme-value distribution and the standard logistic distribution, respectively (Cheng et al., Biometrika 1995).

Remark: Under STM, an unknown transformation of the survival time is linearly related to the covariates with various completely specified error distributions. This class of regression models includes the proportional hazards and proportional odds models.

C-Log-Log Link model for discrete survival data

This model can be obtained by grouping time into intervals with boundaries $0 = \tau_0 < \tau_1 < ... < \tau_K = \infty$ for the continuous-time proportional-hazards model. Suppose $t_1, \ldots, t_k$ are discrete time points of T, where t_k is picked to be $\tau_{k-1} < t_k \leq \tau_k$.

- Let $\lambda(t_k; X)$ denote the hazard probability between τ_{k-1} and τ_k,

$$\begin{aligned}
\lambda(t_k; X) &= 1 - P(T > \tau_k \mid T > \tau_{k-1}, X) \\
&= 1 - \exp\{-\int_{\tau_{k-1}}^{\tau_k} h(t; X)dt\} \quad (h(t; X) \text{ is the hazard of continuous r.v. } T) \\
&= 1 - \{\exp\{-\int_{\tau_{k-1}}^{\tau_k} h_0(t; X)dt\}\}^{exp\{\beta' X\}} \\
&= 1 - \{1 - \lambda_0(t_k)\}^{exp\{\beta' X\}}
\end{aligned}$$

  This implies $1 - \lambda(t_k; X) = \{1 - \lambda_0(t_k)\}^{e^{\beta' X}}$ and $\log(-log\{1 - \lambda(t_k; X)\}) = \alpha_k + \beta' X$ where $\alpha_k = \log(-log\{1 - \lambda_0(t_k)\})$, $k = 1, \ldots, K$.

- Note that C-Log-Log Link model does not handle time-varying covariates.
- This model can be fitted to discrete survival data by generating pseudo-observations and fitting a generalized linear model with binomial error structure and complementary log-log link.