STAT 6390: Analysis of Survival Data
Textbook coverage: Chapter 3

Steven Chiou

Department of Mathematical Sciences,
University of Texas at Dallas

Parametric regression models

- Previously, we learned to fit parametric regression model with the form

$$Y = \log(T) = \alpha + X'\beta + \epsilon, \tag{1}$$

  with survreg.

- The following table gives some of the common distributions in survreg.

dist in survreg	Distribution of T	Distribution of ϵ
exponential	exponential	extreme values
weibull	Weibull	extreme values
loglogistic	log-logistic	logistic
lognormal	log-normal	normal

- Here, models are named for the distribution of T, not ϵ.
- Different distributions assume different shapes for the hazard function.
- See Chapter 2.2 of Kalbfleisch and Prentice (2011) for more comprehensive discussion.

The proportional hazards model

- In the previous note, we note that the hazard at time t for an individual can be written as

$$\lambda(t; x) = \lambda \cdot r(X'\beta)$$

under the exponential assumption.

- To generalize this, we could consider modeling the hazard function

$$h(t; x) = h_0(t)r(X'\beta), \tag{2}$$

where $h_0(t)$ is an unspecified function.

- We assume X is time independent.
- In (2) is the product of two functions:
  - $h_0(t)$ characterizes how the hazard function changes as a function of time
  - $r(X'\beta)$ characterizes how the hazard function changes as a function of subject covariates.
- $h_0(t)$ is referred to as the *baseline hazard function* when $r(X'\beta)$ is parameterized such that $r(0) = 1$.

Hazard ratio

- Under the model (2), the ratio of the hazard function for two subject with covariate x_1 and x_0 is

$$\mathrm{HR}(t; x_1, x_0) = \frac{h(t; x_1)}{h(t; x_0)} = \frac{r(x_1' \beta)}{r(x_0' \beta)}. \tag{3}$$

- This implies that the hazard ratio (HR) depends only on the function $r(X' \beta)$ and does not require the actual form of $h_0(t)$.

- This gives the *proportional hazard* assumption.

The Cox model

- Cox (1992) is the first to propose model (2) with $r(X'\beta) = e^{X'\beta}$.
- With $r(X'\beta) = e^{X'\beta}$, (2) reduces to

$$h(t; x) = h_0(t)e^{X'\beta}, \tag{4}$$

  and the hazard ratio $\mathrm{HR}(t; x_1, x_0)$ becomes $e^{(x_1 - x_0)'\beta}$.
- Equivalently, the log hazard ratio is $(x_1 - x_0)'\beta$.

The Cox model

- Suppose we have a "continuous" covariate x_k, which corresponds to β_k.
- Holding other covariate values at constant, then the log hazard ratio

$$\log\left\{\mathrm{HR}(t, x_k + 1, x_k)\right\} = \log\{h(t, x_k + 1)\} - \log\{h(t, x_k)\} = \beta_k.$$

- β_k is the increase in log-hazard with one unit increase in x_k at any time.
- e^{β_k} is the hazard ratio associated with one unit increase in x_k.
- If x_k is a "categorical" covariate and can only takes on two possible values, 0 and 1, then β_k is the difference in log-hazard between the two groups.

The Cox model

- From the definition of hazard function, we have

$$P(t \leq T < t + dt | T \geq t, x) \approx h(t|x)dt.$$

- This implies

$$\frac{P(t \leq T < t + dt | T \geq t, x_k + 1)}{P(t \leq T < t + dt | T \geq t, x_k)} \approx e^{\beta_k}.$$

- Thus, e^{β_k} can be loosely interpreted in terms of conditional probabilities of dying.

The Cox model

- Since β characterizes the covariate effect, the main focus of the inference procedure is to estimate β and test $H_0 : \beta = 0$.
- As in many other regression model, we will also diagnostics procedure.
- The baseline hazard function, $h_0(t)$, can be treated as a nuisance parameter.

Likelihood approach

- Recall the likelihood we derived in note 3:

$$L = \prod_{i=1}^{n} \{f_T(t_i)\}^{\Delta_i} \cdot \{S_T(t_i)\}^{1-\Delta_i} = \prod_{i=1}^{n} \{h_T(t_i)\}^{\Delta_i} \cdot S_T(t_i). \qquad (5)$$

- If T has hazard function (4), then

$$h_T(t) = h_0(t)e^{X'\beta} \text{ and } S_T(t) = \exp\left\{-H_0(t)e^{X'\beta}\right\},$$

where $H_0(t) = \int_0^t h_0(u)\,\mathrm{d}u$.

- The complete likelihood can be obtained by plugging the above into (5).
- The maximization of the likelihood would requires solving for the unknown parameter β and unspecified baseline hazard at t_i's.

Partial likelihood approach

- Consider the *conditional probability* the ith individual fails at t_i, given the risk set at t_i:

$$\text{P(the } i\text{th individual dies|one death at } t_i)$$

$$= \frac{\text{P(the } i\text{th individual dies |survival to } t_i)}{\text{P(one death at } t_i|\text{survival to } t_i)}$$

$$= \frac{h_i(t; x)}{\sum_{j: t_j \geq t_i}^{n} h_j(t; x)} = \frac{e^{X_i' \beta}}{\sum_{j: t_j \geq t_i}^{n} e^{X_j' \beta}},$$

where the last equality follows from the Cox model assumption.

- The *partial likelihood* is formed by multiplying these conditional probabilities:

$$L_p(\beta) = \prod_{i=1}^{n} \left(\frac{e^{X_i' \beta}}{\sum_{j: t_j \geq t_i}^{n} e^{X_j' \beta}} \right)^{\Delta_i} \tag{6}$$

Partial likelihood approach

- The expression in (6) assumes no ties and excludes terms with $\Delta_i = 0$.
- A different approach to derive (6) is to consider the complete likelihood (5), and decompose it to

$$
\begin{aligned}
L &= \prod_{i=1}^{n} \{h_T(t_i)\}^{\Delta_i} \cdot S_T(t_i) \\
&= \prod_{i=1}^{n} \left(\frac{h_i(t_i)}{\sum_{j:t_j \geq t_i}^{n} h_j(t_i)} \right)^{\Delta_i} \cdot \left(\sum_{j:t_j \geq t_i}^{n} h_j(t_i) \right)^{\Delta_i} \cdot S_T(t_i) \\
&:= \prod_{i=1}^{n} L_1 \cdot L_2 \cdot L_3,
\end{aligned}
$$

where L_1 reduces to the partial likelihood in (6).

Partial likelihood approach

- Cox (1975) argues that L_1 carries "most" of the information about β.
- Cox (1975) also argues that L_2 and L_3 carry information about $\lambda_0(t)$.
- Cox (1975) suggested treating $L_p(\beta)$ as a regular likelihood function and making inference on β accordingly.
- The maximum partial likelihood estimator (MPLE) of β gives an unbiased estimator for β.
- The information matrix based on $L_p(\beta)$ can be used to derive standard error for the MPLE.

Partial likelihood approach

- Suppose (for now) that we only have one type of covariate, e.g, $p = 1$.
- For the ease of notation, we will right $j \in R(t_i)$ instead of $j : t_j \geq t_i$ to denote the index j is from the risk set.
- The log of the partial likelihood, $\log\{L_p(\beta)\}$ has the form

$$\ell_p(\beta) = \sum_{i=1}^{n} \Delta_i \left[X_i \beta - \log \left\{ \sum_{j \in R(t_i)} e^{X_j \beta} \right\} \right].$$

- The first derivative gives the *score function*:

$$\frac{\mathrm{d}\ell_p}{\mathrm{d}\beta} = \sum_{i=1}^{n} \Delta_i \left[X_i - \left\{ \frac{\sum_{j \in R(t_i)} X_j e^{X_j \beta}}{\sum_{j \in R(t_i)} e^{X_j \beta}} \right\} \right]. \tag{7}$$

Partial likelihood approach

- Notice that the fraction in the summation can be expressed as

$$\frac{\sum_{j \in R(t_i)} X_j e^{X_j \beta}}{\sum_{j \in R(t_i)} e^{X_j \beta}} = \sum_{j \in R(t_i)} \frac{X_j e^{X_j \beta}}{\sum_{k \in R(t_i)} e^{X_k \beta}} = \sum_{j \in R(t_i)} X_j \cdot \omega_{ij}(\beta),$$

where $\omega_{ij}(\beta) = \frac{e^{X_j \beta}}{\sum_{k \in R(t_i)} e^{X_k \beta}}$.

- $\omega_{ij}(\beta)$ can be seem as the conditional probability of death at t_j.
- This implies $\bar{X} = \sum_{j \in R(t_i)} X_j \cdot \omega_{ij}(\beta)$ is like a weighted average of X over all the individuals in the risk set at t_i.
- This further reduced (7) to

$$\sum_{i=1}^{n} \Delta_i (X_i - \bar{X}),$$

a familiar $\sum (O_i - E_i)$ form!

Partial likelihood approach

- The second derivative is

$$\frac{\mathrm{d}^2 \ell_p}{\mathrm{d}\beta^2} = -\sum_{i=1}^{n} \Delta_i \sum_{j \in R(t_i)} w_{ij}(\beta) \cdot \left(X_j - \bar{X}\right)^2.$$

- Then $\hat{\mathrm{Var}}(\hat{\beta}) = I(\hat{\beta})^{-1}$, where $I(\beta) = -\frac{\mathrm{d}^2 \ell_p}{\mathrm{d}\beta^2}$.

Reference

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Cox, D. R. (1992). Regression models and life-tables. In *Breakthroughs in statistics*, pages 527–541. Springer.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.