Chapter 3

Regression Models for Survival Data

Regression Models for Survival Data

- Linear and generalized linear regression models are fully parametric models that:
  - Describe the distribution of an outcome variable, Y
  - Model the relationship between Y and a vector of covariates, X
- Examples:
  - Linear regression
    - $Y \sim N(\mu, \sigma^2)$; $\mu = \beta X$
  - Logistic regression
    - $Y \sim \text{Binomial}(n, p)$; $\text{logit}(p) = \beta X$
  - Poisson regression
    - $Y \sim \text{Poisson}(\mu)$; $\log(\mu) = \beta X$

Regression Models for Survival Data

- For survival data, the outcome variable is survival time, T, and we can build a model using the hazard function, h(t)

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t \mid T \ \geq t)}{\Delta t}$$

- A simple example of a parametric regression model:

$$\log h(t|\beta, x) = \beta x$$
$$h(t|\beta, x) = e^{\beta x}$$

  - For a given set of covariates, the hazard is constant over time
  - T ~ Exponential($e^{\beta x}$)

Regression Models for Survival Data

- Often, we are much less interested in the distribution of the survival time than in the relationship between survival time and covariates
  - Example: Do subjects taking Drug A live longer than those taking Drug B?
  - For these situations, the ratio of the hazards under Drugs A and B is of interest and a fully parametric model is not needed:

$$\frac{h(t|A)}{h(t|B)} < 1 \qquad \frac{h(t|A)}{h(t|B)} > 1 \qquad \frac{h(t|A)}{h(t|B)} = 1$$

Drug A better | Drug B better | No difference

The Proportional Hazards Assumption

- Need a semi-parametric model that will allow us to model the relationship between covariates and the hazard function, without specifying the distribution of the survival time.
- **The proportional hazards assumption**: Individuals will have hazard functions that are proportional to one another. That is

$$\frac{h(t|\beta, x_1)}{h(t|\beta, x_2)},$$

the ratio of hazard functions for two individuals with covariates x_1 and x_2, does not vary with time.

The Proportional Hazards Assumption

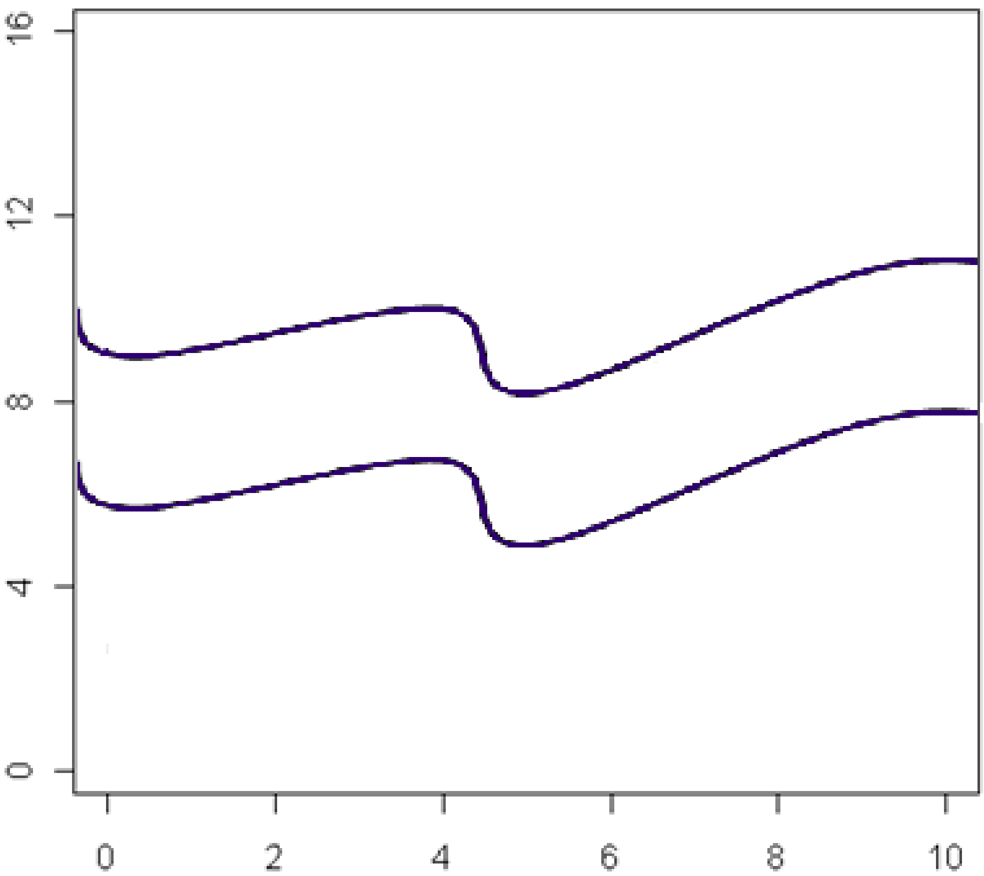
- Under the proportional hazards assumption, the hazard function given a set of covariates *x* can be expressed as

$$h(t|\beta, x) = h_0(t) \cdot r(x, \beta)$$

  - $h_0(t)$ is the baseline hazard function
  - Note that the hazard ratio, $HR(t|x_1,x_2)$, does not depend on either time or the baseline hazard:

$$HR(t|x_1, x_2) = \frac{h_0(t) \cdot r(x_1, \beta)}{h_0(t) \cdot r(x_2, \beta)} = \frac{r(x_1, \beta)}{r(x_2, \beta)}$$

The Proportional Hazards Assumption



- Generic proportional hazards

Cox Proportional Hazards Model

- The Cox proportional hazards model uses

$$r(\mathrm{x}, \beta) = e^{\beta x}$$

  - The Cox model is the most common proportional hazard model
- Under the Cox model, the hazard ratio is

$$HR(t|x_1, x_2) = e^{\beta(x_1 - x_2)}$$

- The survival function is

$$S(t|\beta, x) = [S_0(t)]^{e^{\beta x}}$$

where $S_0(t)$ is the baseline survival function

Cox Proportional Hazards Model

- The hazard function is

$$h(t|\beta, x) = h_0(t) \cdot e^{\beta x}$$

  - $\beta = 0$: x does not affect survival
  - $\beta < 0$: As x increases, the hazard decreases and survival improves
  - $\beta > 0$: As x increases, the hazard increases and survival gets worse

Estimating Cox Model Parameters

- To use a regression model, we need a way to estimate the model parameters.
- Suppose we have a sample of n observations:

$$(t_1, x_1, c_1),\ (t_2, x_2, c_2),\ \ldots\ (t_n, x_n, c_n)$$

where

  - t_j is the survival time,
  - x_j is a vector of covariates, and
  - c_j is the censoring indicator

    ($c_j = 1$ for observed, $c_j = 0$ for censored)

for the j^{th} subject.

- In order to use maximum likelihood estimation, need to think about each observation's contribution to the likelihood

Estimating Cox Model Parameters

- Usually, the maximum likelihood estimator is found by maximizing

$$L(\Omega) = \prod_{j=1}^{n} f(y_j | \Omega, x_j)$$

where f(·) is the density function for a random variable y and depends on covariates x and model parameter(s) Ω.

- However, in survival analysis, some of our observations are censored, so the true survival time is unknown.

Estimating Cox Model Parameters

- For subjects who had the event, the exact survival time T_j is known, so the contribution is
  - $f(t_j|\beta,x_j)$
- For (right-)censored subjects, all we know is that the true survival time T_j is greater than censored time t_j
  - The contribution is $S(t_j|\beta,x_j) = 1 - F(t_j|\beta,x_j)$
- So the likelihood function is

$$L(\beta) = \prod_{j=1}^{n} \left[f(t_j|\beta,x_j)\right]^{c_j} \left[S(t_j|\beta,x_j)\right]^{1-c_j}$$

Estimating Cox Model Parameters

- In order to obtain the MLE for β under the Cox model, we would need to maximize

$$L(\beta) = \sum_{j=1}^{n} \left[c_j \log h_0(t_j) + c_j \beta x_j + e^{\beta x_j} \log S_0(t_j) \right]$$

  - But the whole point of using a proportional hazards model is to avoid having to specify the baseline hazard.
  - To get around this problem, Cox proposed a partial likelihood expression that can be maximized instead.

Estimating Cox Model Parameters

- Cox partial likelihood:

$$L_p(\beta) = \prod_{j=1}^{n} \left(\frac{e^{\beta x_j}}{\sum_{k \in R(t_j)} e^{\beta x_k}} \right)^{c_j}$$

where $R(t_j)$ is the set of all subjects who are still at risk at time t_j

  - This expression assumes no tied survival times

- The maximum partial likelihood estimator, $\hat{\beta}$, is the value of β that maximizes $L_p(\beta)$

Estimating Cox Model Parameters

- The variance of $\hat{\beta}$ is estimated using the Fisher information matrix:

$$\hat{V}(\hat{\beta}) = I^{-1}(\hat{\beta})$$

where

$$I(\beta) = -\frac{\partial^2 \log L_p(\beta)}{\partial \beta^2}$$

- Software packages are needed to get $\hat{\beta}$ and $\hat{V}(\hat{\beta})$
- A 100(1-α)% CI for β is

$$\hat{\beta} \pm z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\beta})}$$

Hypothesis Tests for the Cox Model

- The Cox proportional hazards model uses the following hazard function:

$$h(t|\beta, x) = h_0(t) \cdot e^{\beta x}$$

  - If $\beta = 0$, the covariate *x* is not related to survival time.
- Often want to test the null hypothesis H_0: $\beta = 0$ against the alternative H_1: $\beta \neq 0$.

Hypothesis Tests for the Cox Model

- Partial likelihood ratio test
  - The usual likelihood ratio test takes the form

$$G = 2 \cdot log\left(\frac{likelihood\ of\ full\ model}{likelihood\ of\ reduced\ model}\right)$$

  - If H_0 is true, then G ~ χ^2_d, where *d* is the difference between the number of parameters in the two models

Hypothesis Tests for the Cox Model

- Partial likelihood ratio test
  - For Cox proportional hazards model, the partial likelihood is used to conduct the test.

$$L_p(\beta) = \prod_{j=1}^{n} \left(\frac{e^{\beta x_j}}{\sum_{k \in R(t_j)} e^{\beta x_k}} \right)^{c_j}$$

  - Reject H_0 in favor of H_1 if G $>\chi^2_{d,1-\alpha}$
  - The two-sided p-value is P($\chi^2_d \geq$ G).

Hypothesis Tests for the Cox Model

- The Wald test
  - The test statistic is

$$z = \frac{\hat{\beta}}{\sqrt{\hat{V}(\hat{\beta})}}$$

  - If H_0 is true then $z \sim N(0,1)$.
  - Reject H_0 in favor of H_1 if $z < z_{\alpha/2}$ or $z > z_{1-\alpha/2}$
  - The two-sided p-value is $2 \cdot P(Z > |z|)$, where Z follows a standard normal distribution

Hypothesis Tests for the Cox Model

- The Wald test (continued)
  - For the multivariate case, can also use the test statistic

$$z^2 = \hat{\beta}^T \left(\hat{V}(\hat{\beta})\right)^{-1} \hat{\beta}$$

    - If H_0 is true then $z^2 \sim \chi^2_d$, where d is the dimension of $\hat{\beta}$

Hypothesis Tests for the Cox Model

- The Score Test
  - The test statistic is

$$z^* = \left. \frac{\frac{\partial}{\partial \beta} \log L_p(\beta)}{\sqrt{I(\beta)}} \right|_{\beta=0}$$

  - If H_0 is true then z* ~ N(0,1).
  - One advantage of the score test is that you do not have to calculate $\hat{\beta}$ in order to perform the test
  - Square of the score test statistic is also sometimes used.

Hypothesis Tests for the Cox Model

- The three tests usually give the same result, especially when dealing with large samples
- The partial likelihood ratio test is usually chosen if the test results differ

Dealing with Tied Survival Times

- Cox's partial likelihood assumes that there are no tied survival times
  - This assumption makes sense in theory, but tied survival times are common
    - Time is continuous, but ties occur because we have imprecise measurements
  - How should the partial likelihood be modified to take ties into account?

Dealing with Tied Survival Times

- Exact partial likelihood
  - Proposed by Kalbfleisch and Prentice
  - Motivation: For any group of tied survival times, we want to know the true order in which they died.
    - Example: If three people are listed as having died on March 5, ideally we would want to know who died first, second and third on that day
  - Since we don’t know the order, we have to take into account all of the possible orderings
    - If *m* people have the same survival time, there are *m!* possible orderings.

Dealing with Tied Survival Times

- If there are large tied groups, then there will be many possible orderings.
  - In these situations, calculating the exact partial likelihood is difficult, even for SAS
- Two common approximations for the exact partial likelihood:
  - Breslow’s approximation
  - Efron’s approximation

Dealing with Tied Survival Times

- Breslow's approximation
  - More commonly used than Efron
  - Performs poorly when there are a large number of ties compared to the number at risk
- Efron's approximation
  - More accurate than Breslow
  - Takes longer to compute than Breslow

- No real consensus on which method is best
  - With no ties, the two methods give the same results.
  - With few ties, there is little difference in the results.

Estimating the Survival Function

- The hazard ratio can be used to estimate the effect of a covariate without having to specify the baseline hazard function, $h_0(t)$
- Sometimes want to estimate the survival function for a person with a specific set of covariate values.
  - Example: Fit a Cox proportional hazards model using age and gender as covariates. Want to estimate the median survival time for a 50 year-old man.

Estimating the Survival Function

- For the Cox model, the survival function is

$$S(t|\beta, x) = [S_0(t)]^{e^{\beta x}}$$

where $S_0(t)$ is the baseline survival function

  - So estimating $S(t \mid \beta, x)$ means we need an estimate of $S_0(t)$.

Estimating the Survival Function

- Can derive an estimator for $S_0(t)$ by thinking about the conditional survival probability (as in Kaplan-Meier estimation).
- Breslow's estimator:

$$\hat{S}_0(t) = \prod_{t_i \le t} \hat{p}_i$$

where

$$\hat{p}_i = \left(1 - \frac{e^{\hat{\beta} x_i}}{\sum_{j \in R(t_i)} e^{\hat{\beta} x_j}}\right)^{e^{-\hat{\beta} x_i}}$$