Logistic Regression Models for Discrete Time Hazards

Previously . . .

We saw that, under the assumptions of **conditionally independent** and **non-informative** censoring, the (partial) likelihood expression becomes

$$L(\theta) = \prod_{i=1}^{n} \prod_{s=1}^{\infty} P(Y(s)|\mathcal{H}^Y(s)) = \prod_{s=1}^{C_i} h_i(s;\theta)^{Y_i(s)}(1 - h_i(s;\theta))^{1-Y_i(s)}.$$

Here $h_i(s;\theta)$ is **some parametric model** for the hazard!

This looks familiar . . .

This is a **binomial likelihood** with probabilities given by $h_i(s; \theta)$.

Discrete Time Survival Data with GLMs

Basic Model

Suppose that D_j is a categorical time period variable, then we can take

$$\text{logit}\{h(s;\alpha)\} = \alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{C_i} D_{C_i}.$$

This can be fit using **logistic regression**.

Fitting the Model

Suppose that `my_df` contains the relevant data in **person-period** format.

```
glm(Y ~ -1 + factor(time),
    family = binomial,
    data = my_df)
```

Inference and Confidence

- This technique allows **standard GLM inference**.

Inference and Confidence

- This technique allows **standard GLM inference**.
- We can get $h(t) = \text{expit}(\alpha_t)$, with **corresponding confidence intervals**.

Inference and Confidence

- This technique allows **standard GLM inference**.
- We can get $h(t) = \text{expit}(\alpha_t)$, with **corresponding confidence intervals**.
- Can test for **equality of hazards**, for instance,

$$H_0 : \alpha_t = \alpha_{t+1} \implies H_0 : h(t) = h(t+1).$$

Inference and Confidence

- This technique allows **standard GLM inference**.
- We can get $h(t) = \text{expit}(\alpha_t)$, with **corresponding confidence intervals**.
- Can test for **equality of hazards**, for instance,

$$H_0 : \alpha_t = \alpha_{t+1} \implies H_0 : h(t) = h(t+1).$$

- This is a **maximum likelihood estimator**, allowing for MLE-type inference.

Inference and Confidence

- This technique allows **standard GLM inference**.
- We can get $h(t) = \text{expit}(\alpha_t)$, with **corresponding confidence intervals**.
- Can test for **equality of hazards**, for instance,

$$H_0 : \alpha_t = \alpha_{t+1} \implies H_0 : h(t) = h(t+1).$$

- This is a **maximum likelihood estimator**, allowing for MLE-type inference.
- Gives **survivor function** through

$$S(t) = \prod_{j=1}^{t} \{1 - h(j)\}.$$

Survivor Function Inference

Confidence Intervals

In general we need to use the **multivariate delta method** to get confidence intervals for the **survivor function**.

This is based on intervals around the **log-transform**.

$$\text{var}\left\{\log \widehat{S}(t)\right\} \approx G \,\text{var}\left(\widehat{\alpha}\right) G',$$

which is estimated as

$$\widehat{\text{var}}\left\{\log \widehat{S}(t)\right\} \approx \widehat{G}\widehat{\text{var}}\left(\widehat{\alpha}\right) \widehat{G}'.$$

Confidence Intervals (Cont.)

In this case, we estimate $\text{var}(\widehat{\alpha})$ using the **GLM fit**.

Confidence Intervals (Cont.)

In this case, we estimate $\text{var}\left(\widehat{\alpha}\right)$ using the **GLM fit**.

We estimate G, as

$$\widehat{G} = \begin{bmatrix} -\widehat{h}(1) & 0 & 0 & \cdots & 0 \\ -\widehat{h}(1) & -\widehat{h}(2) & 0 & \cdots & 0 \\ -\widehat{h}(1) & -\widehat{h}(2) & -\widehat{h}(3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\widehat{h}(1) & -\widehat{h}(2) & -\widehat{h}(3) & \cdots & -\widehat{h}(C) \end{bmatrix}.$$

Combining these Results

A **confidence interval** for $\widehat{S}(j)$ is then given by,

$$\exp\left(\log\left\{\widehat{S}(j)\right\} \pm Z_{\alpha/2} \times \sqrt{\widehat{\text{var}}\left\{\log \widehat{S}(t)\right\}_{(j,j)}} \right).$$

The Proportional Odds Model

What if we want **the hazard** to differ based on **covariates**?

Extending the Logistic Regression

Suppose that we fit

$$\text{logit}\{h(s;\alpha)\} = \alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{C_i} D_{C_i} + X_i'\beta.$$

Here we let **all relevent covariates** be contained in X_i.

The Proportional Odds Assumption

Consider if $X_i = \text{Sex}_i$, giving the model

$$\text{logit}\{h(s;\alpha)\} = \alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{C_i} D_{C_i} + \beta \text{Sex}_i.$$

For **every time** we would find that $\text{logit}\{h(s;\alpha,\text{F})\} - \text{logit}\{h(s;\alpha,\text{M})\} = \beta$, and so the **odds ratio** is given by $\exp(\beta)$.

The Proportional Odds Assumption

Consider if $X_i = \text{Sex}_i$, giving the model

$$\text{logit}\left\{h(s;\alpha)\right\} = \alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{C_i} D_{C_i} + \beta\text{Sex}_i.$$

For **every time** we would find that $\text{logit}\left\{h(s;\alpha,\text{F})\right\} - \text{logit}\left\{h(s;\alpha,\text{M})\right\} = \beta$, and so the **odds ratio** is given by $\exp(\beta)$.

We call this the **proportional odds model** since the odds differ by a **constant, multiplicative constant**. That is, they are proportional.

Testing the Proportional Odds Assumption

We can test the validity of **the proportional odds assumption** by fitting

$$\text{logit}\{h(s;\alpha)\} = \alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{C_i} D_{C_i} + \beta_1 \text{Sex}_i + \beta_2 \text{Sex}_i D_2 + \cdots + \beta_{C_i} D_{C_i} \text{Sex}_i.$$

Testing the Proportional Odds Assumption

We can test the validity of **the proportional odds assumption** by fitting

$$\text{logit}\{h(s;\alpha)\} = \alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{C_i} D_{C_i} + \beta_1 \text{Sex}_i + \beta_2 \text{Sex}_i D_2 + \cdots + \beta_{C_i} D_{C_i} \text{Sex}_i.$$

Then we test

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_{C_i} = 0,$$

using a simple **nested likelihood ratio test**.

Summary

- The partial likelihood for discrete time survival analysis admits a **binomial representation**.

Summary

- The partial likelihood for discrete time survival analysis admits a **binomial representation**.
- We can use **logistic regression** to estimate the **hazards** based on a factor variable for time.

Summary

- The partial likelihood for discrete time survival analysis admits a **binomial representation**.
- We can use **logistic regression** to estimate the **hazards** based on a factor variable for time.
- The **survivor function** is estimable through the cumulative product.

Summary

- The partial likelihood for discrete time survival analysis admits a **binomial representation**.
- We can use **logistic regression** to estimate the **hazards** based on a factor variable for time.
- The **survivor function** is estimable through the cumulative product.
- **Standard inference** exists for the hazard function and we can use the **multivariate delta** for the survivor function.

Summary

- The partial likelihood for discrete time survival analysis admits a **binomial representation**.
- We can use **logistic regression** to estimate the **hazards** based on a factor variable for time.
- The **survivor function** is estimable through the cumulative product.
- **Standard inference** exists for the hazard function and we can use the **multivariate delta** for the survivor function.
- Can add in variates making the **proportional odds** assumption, which can be tested using **deviance tests**.