Logistic Regression

Part 2 - likelihood ratio tests

Ryan Miller

Grinnell College
Statistics

Introduction

- When fitting a logistic regression we've used the argument `family="binomial"` in the `glm()` function
  - This specifies the probability distribution involved in the model
- For a single observation, the binomial distribution leads to the following **likelihood function**:

$$Pr(y=1)^y \cdot (1 - Pr(y=1))^{1-y}$$

- In logistic regression, after applying the inverse function of log-odds, $Pr(y=1)$ is modeled by $\frac{1}{1+exp(-(\beta_0+\beta_1 X_1+...))}$
  - Thus, our modeling choices, such as which explanatory variables to include, influence the likelihood function

Likelihood

This likelihood function allows us to measure how well our model is doing:

$$Pr(y=1)^y \cdot (1 - Pr(y=1))^{1-y}$$

- If a data-point is observed to have $y = 1$, this expression reduces $Pr(y=1)$
  - A *highly effective model* should produce an estimate of $Pr(y=1)$ that is close to 1 for this data-point
  - A *non-informative model* should produce an estimate of $Pr(y=1)$ far from 1 (close to 0.5 if the data are balanced)

Likelihood (cont.)

Likelihood function (one data-point)

$$Pr(y=1)^y \cdot (1-Pr(y=1))^{1-y}$$

- Similarly, if a data-point is observed to have $y=0$, the expression reduces to $1-Pr(y=1)$
  - A *highly effective model* produces an estimate of $Pr(y=1)$ that is nearly zero for this data-point, thereby making $1-Pr(y=1)$ close to 1
  - A *non-informative model* will led to $1-Pr(y=1)$ being far from 1 ($Pr(y=1)$ might be close to 0.5 if the data are balanced)

Likelihood (cont.)

The likelihoods for every individual observation in our data set can be aggregated by multiplying them together:

$$L = \prod_{i=1}^{n} Pr(y_i = 1|x_i)^{y_i} \cdot (1 - Pr(y_i = 1|x_i))^{1-y_i}$$

- The notation $y_i = 1|x_i$ indicates the predicted probability is contingent on the values of the explanatory variables of that particular case, which aren't the same for all $i \in \{1, 2, \ldots, n\}$
- The theoretical maximum of L occurs when $Pr(y_i = 1|x_i)$ is 1 for all data-points with $y_i = 1$ and $Pr(y_i = 1|x_i)$ is 0 for all data-points with $y_i = 0$
  - The closer a model gets to this theoretical maximum, the better it fits the data

Likelihood (cont.)

For two different logistic regression models, $Pr(y_i = 1|x_i)$ can be very different:

$$\text{Model } 1: Pr(y=1) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 \text{Shot Distance}))}$$
$$\text{Model } 2: Pr(y=1) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 \text{Shot Distance} + \beta_2 \text{Touch Time}))}$$

The model whose estimates more closely resemble the observed data (ie: $Pr(y=1)$ close to 1 for $y=1$) will have a larger likelihood

Grinnell College
Statistics

Likelihood Ratio Test

- When two models are *nested*, we can compare the ratio of their likelihoods to test whether the larger model provides a significantly better fit to the data than the reduced model
  - For reasons we will not cover, -2 times the natural log of this ratio follows a Chi-squared distribution when the sample size is large
  - The degrees of freedom are the number of additional parameters present in the larger model

Likelihood Ratio Test (cont.)

All of that is to provide background into why/how we can compare nested logistic regression models using a procedure that is conceptually similar to the F-test in linear regression:

```
model1 = glm(OUTCOME ~ DRIBBLES, data = shots, family = "binomial")
model2 = glm(OUTCOME ~ DRIBBLES + SHOT_CLOCK, data = shots, family = "binomial")
lrtest(model1, model2)
```

```
## Likelihood ratio test
##
## Model 1: OUTCOME ~ DRIBBLES
## Model 2: OUTCOME ~ DRIBBLES + SHOT_CLOCK
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1   2 -82216
## 2   3 -81674  1  1084  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary

- Likelihood provides a way of numerically quantifying how well the sample data fits a particular logistic regression model
- When two logistic regression models are nested, their fits can be compared using a Likelihood Ratio Test
  - The null hypothesis of this test is that the smaller null model and the larger alternative model both fit the data equally well
- When the likelihood ratio is large (leading to a small p-value), there is evidence that the alternative model fits the data significantly better