STAT347: Generalized Linear Models
Lecture 6

Today's topics: Chapters 5.3 - 5.5, 5.7

- Binary GLM inference
- Fitting logistic regression and the infinite estimates
- Binary GLM example

1 Binary GLM model inference

We have already learnt the inference of a general GLM model, we now look what the specific forms are for a binary GLM.

1.1 Score function

For logistic regression, as the logit link is the canonical link, the score equation is:

$$\frac{\partial L}{\partial \beta_j} = \sum_i (y_i - n_i p_i) x_{ij} = \sum_i \left(y_i - \frac{n_i e^{X_i^T \beta}}{1 + e^{X_i^T \beta}} \right) x_{ij} = 0$$

For a binary GLM with other links, when $p_i = F(\eta_i) = F(X_i^T \beta)$ as discussed in last lecture:

$$\begin{aligned} \frac{\partial L}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i) x_{ij}}{\mathrm{Var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} &= \sum_i \frac{(y_i - \mu_i) x_{ij}}{\mathrm{Var}(y_i)} n_i F'(\eta_i) \\ &= \sum_i \frac{\left[y_i - n_i F(X_i^T \beta) \right] x_{ij}}{F(X_i^T \beta) \left[1 - F(X_i^T \beta) \right]} f(X_i^T \beta) \\ &= 0 \end{aligned}$$

1.2 Variance of $\hat{\beta}$

We have derived that as $n \to \infty$

$$\mathrm{Var}(\hat{\beta}) \to (X^T W X)^{-1}$$

where $W = D^2 V^{-1}$ is a diagonal matrix

For logistic regression where the logit link is the canonical link, $W = V$

$$W_{ii} = n_i p_i (1 - p_i), \quad \widehat{W}_{ii} = n_i \frac{e^{X_i^T \hat{\beta}}}{(1 + e^{X_i^T \hat{\beta}})^2}$$

For a general link with $p_i = F(\eta_i)$

$$W_{ii} = \frac{n_i f(X_i^T \beta)^2}{F(X_i^T \beta) \left[1 - F(X_i^T \beta) \right]}$$

1.3 Hypothesis testing

Consider a simple case. Under the null model, the group data is $\sum_i y_i \sim \text{Binomial}(N, p)$. We want to test for $H_0 : \beta = \text{logit}(p_0)$ (or equivalently: $H_0 : p \equiv p_0$) where β is the constant coefficient. Define $y = \sum_i y_i/N$, then the MLE is $\hat{p} = y$. The test statistics are

Wald test:

$$\left(\frac{\hat{\beta} - \text{logit}(p_0)}{\widehat{\text{SE}}(\hat{\beta})}\right)^2 = [\text{logit}(y) - \text{logit}(p_0)]^2 Ny(1-y),$$

Or

$$\left(\frac{\hat{p} - p_0}{\widehat{\text{SE}}(\hat{p})}\right)^2 = \frac{(y-p_0)^2}{[y(1-y)/N]}$$

Likelihood ratio test:

$$-2(L_0 - L_1) = -2\log\left[\frac{p_0^{Ny}(1-p_0)^{N-Ny}}{y^{Ny}(1-y)^{N-Ny}}\right]$$

Score test:

$$T = \frac{\dot{L}(\beta_0)^T \dot{L}(\beta_0)}{-\ddot{L}(\beta_0)} = \frac{(y-p_0)^2}{[p_0(1-p_0)/N]}$$

- Wald test depends on the scale
- Wald test is less stable when y is close to 0 or 1. Read Chapter 5.3.3

1.4 Deviance

The total (residual) deviance for a binary GLM (the deviance between the saturated model and the fitted model) is

$$\begin{aligned}
D_+(y, \hat{\mu}) &= \sum_i D(y_i, n_i\hat{p}_i) \\
&= -2\sum_i \log\left[f(y_i, \hat{\theta}_i)/f(y_i, \theta_{y_i})\right] \\
&= -2\sum_i \log\left[\frac{\hat{p}_i^{y_i}(1-\hat{p}_i)^{n_i-y_i}}{(y_i/n_i)^{y_i}(1-y_i/n_i)^{n_i-y_i}}\right] \\
&= 2\sum_i y_i \log\frac{y_i}{n_i\hat{p}_i} + 2\sum_i (n_i - y_i)\log\frac{n_i - y_i}{n_i - n_i\hat{p}_i}
\end{aligned}$$

- The total deviance is different for grouped data and ungrouped data (as the saturated model is different)
- For grouped data, if represented by an $n \times 2$ table

$$D_+(y, \hat{\mu}) = 2\sum \text{observed} \ \times \log(\text{observed}/\text{fitted})$$

Chi-square goodness of fit test:

$$X^2 = \sum \frac{(\text{observed} \ - \ \text{fitted})^2}{\text{fitted}}$$

2 Binary GLM computation

For logistic regression, Newton's method = Fisher scoring = IRLS.
For IRLS, the tth iteration is

$$X^T W^{(t)}(z^{(t)} - X\beta) = 0$$

where

$$\begin{aligned} z_i^{(t)} &= X_i^T \beta^{(t)} + \left(D_{ii}^{(t)}\right)^{-1}(y_i - \mu_i^{(t)}) \\ &= \log\left(\frac{p_i^{(t)}}{1-p_i^{(t)}}\right) + \frac{y_i - n_i p_i^{(t)}}{n_i p_i^{(t)}(1-p_i^{(t)})} \end{aligned}$$

and

$$W_{ii}^{(t)} = V_{ii}^{(t)} = n_i p_i^{(t)}(1-p_i^{(t)})$$

2.1 Infinite parameter estimates

One may sometimes see this warning message using R to solve the logistic regression:
Warning message: glm.fit: fitted probabilities numerically 0 or 1 occurred
You may see very large estimates of β. What happened?

- Perfect separation: If there exists β_s such that if $X_i^T\beta_s > 0$ then $y_i = 1$ otherwise $y_i = 0$, then let $\beta = k\beta_s$.

  When $k \to \infty$, then

  $$\frac{e^{X_i^T\beta}}{1+e^{X_i^T\beta}} \to \begin{cases} 1 & \text{if } X_i^T\beta_s > 0 \\ 0 & \text{else} \end{cases}$$

  Thus, $\frac{\partial L}{\partial \beta} \to 0$ if $k \to \infty$ so the solution of the score equation is infinite

- Quasi-complete separation: If there exists β_s such that if $X_i^T\beta_s > 0$ then $y_i = 1$, if $X_i^T\beta_s < 0$ then $y_i = 0$, and if $X_i^T\beta_s = 0$ then $y_i = 0$ or 1 (allow data points on the separation hyperplane with both outcomes.

  Let $\eta_i = kX_i^T\beta_s + \beta_0$ where β_0 is any arbitrary scalar. When $k \to \infty$, then

  $$\frac{e^{kX_i^T\beta_s+\beta_0}}{1+e^{kX_i^T\beta_s+\beta_0}} \to \begin{cases} 1 & \text{if } X_i^T\beta_s > 0 \\ 0 & \text{if } X_i^T\beta_s < 0 \\ \frac{e^{\beta_0}}{1+e^{\beta_0}} & \text{if } X_i^T\beta_s = 0 \end{cases}$$

  So we still have $\frac{\partial L}{\partial \beta} \to 0$ for some β_0 and $k \to \infty$, and we still have infinite ML estimates.

- How to deal with perfect/quasi-complete separation? Read Chapter 5.4.2

3 Example: risk factors for cancer

Chapter 5.7.1

Next time: Chapter 6.1, multivariate GLM: nominal response