STAT347: Generalized Linear Models
Lecture 3

Today's topics: Chapters 4.3-4.4

- Hypothesis testing for β
- Deviance analysis of a GLM

1 Wald, likelihood-ratio and score tests

In last lecture, we have mentioned that when n is large

$$\hat{\beta} - \beta_0 \dot{\sim} N(0, V_{\beta_0})$$

How to test

$$H_0 : A\beta_0 = a_0 \quad V.S. \quad H_1 : A\beta_0 \neq a_0$$

1.1 Wald test

Test statistics:

$$T = (A\hat{\beta} - a_0)^T \left[\widehat{\text{Var}}(A\hat{\beta})\right]^{-1} (A\hat{\beta} - a_0)$$

- $\widehat{\text{Var}}(A\hat{\beta}) = AV_{\hat{\beta}}A^T$
- If $a_0 \in \mathbb{R}^1$, Wald statistic can also be written as

$$z = \frac{A\hat{\beta} - a_0}{\sqrt{\widehat{\text{Var}}(A\hat{\beta})}}$$

- Under H_0, Wald statistic $z \dot{\sim} N(0, 1)$ and $T = z^2 \dot{\sim} \mathcal{X}_1^2$
- When $a_0 \in \mathbb{R}^d$, then under H_0, $T \dot{\sim} \mathcal{X}_d^2$
- This is the GLM R package output for the analysis of each component β_j

1.2 Likelihood ratio test

The likelihood ratio test statistics is

$$-2 \log \Lambda = -2 \left(L(\tilde{\beta}) - L(\hat{\beta})\right)$$

where $\tilde{\beta}$ is the MLE of β under the constraint $A\beta = a_0$, and $\hat{\beta}$ is our original MLE of β without any constraint. As $n \to \infty$,

$$-2 \log \Lambda \to \mathcal{X}_d^2$$

1.3 Score test

We consider a simpler case

$$H_0 : \beta = \beta_0 \in \mathbb{R}^p \quad V.S. \quad H_1 : \beta \neq \beta_0$$

Last time:

$$\text{Var}\left(\dot{L}(\beta)\right) = \mathbb{E}\left(\left(\frac{\partial L}{\partial \beta}\right)^2\right) = -\ddot{L}(\beta)$$

where β is the true value of the parameter. Thus, under H_0,

$$\text{Var}\left(\dot{L}(\beta_0)\right) = -\ddot{L}(\beta_0)$$

Test statistics:

$$T = -\dot{L}(\beta_0)^T \left(\ddot{L}(\beta_0)\right)^{-1} \dot{L}(\beta_0)$$

Under H_0, $T \to \mathcal{X}_p^2$ when $n \to \infty$.

- Relationship among the three tests: Section 4.3.4
- Construct CI: invert tests (illustrate more in later lectures)

2 Deviance analysis

2.1 Definition (more general than the textbook)

Consider density function $f(y;\theta) = e^{y\theta - b(\theta)} f_0(y)$ at two values θ_1 and θ_2. Measure the “distance” between two distributions:

$$\begin{aligned} D(\theta_1, \theta_2) = 2\mathbb{E}_{\theta_1}\left\{\log \frac{f(y;\theta_1)}{f(y;\theta_2)}\right\} &= 2\mathbb{E}_{\theta_1}\left\{y(\theta_1 - \theta_2) - b(\theta_1) + b(\theta_2)\right\} \\ &= 2\left[\mu_1(\theta_1 - \theta_2) - b(\theta_1) + b(\theta_2)\right] \end{aligned}$$

Remember the 1-to-1 mapping between μ and θ, we also write $D(\mu_1, \mu_2) = D(\theta_{\mu_1}, \theta_{\mu_2})$

- Generally, $D(\mu_1, \mu_2) \neq D(\mu_2, \mu_1)$
- KL divergence: $D(\mu_1, \mu_2)/2$
- If f is the normal density with $\sigma = 1$, then $D(\mu_1, \mu_2) = (\mu_1 - \mu_2)^2$

Deviance between the saturated model: $\hat{\mu} = y$ and another model with μ:

$$D(y, \mu) = 2\left[y(\theta_y - \theta) - b(\theta_y) + b(\theta)\right] = -2\log\left[f(y, \theta)/f(y, \theta_y)\right]$$

With samples $(X_1, y_1), (X_2, y_2), \cdots, (X_n, y_n)$, the total deviance in GLM (residual deviance, the deviance definition in the text book)

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

Null deviance:

$$\sum_i D(y_i, \bar{y})$$

where $\bar{y} = \sum_i y_i/n$

2.2 Deviance analysis for nested models

Consider the canonical link $\theta = X\beta$ where $\beta \in \mathbb{R}^p$. Let $\beta = \begin{pmatrix} \beta^{(1)} \\ \beta^{(2)} \end{pmatrix}$ where $\beta^{(1)} \in \mathbb{R}^{p_1}$ and $X = \begin{pmatrix} X^{(1)} & X^{(2)} \end{pmatrix}$.

We call $\mathcal{M}^{(1)}$ with

$$\theta = X^{(1)}\beta^{(1)}$$

a nested model of the full model $\mathcal{M}$. Let $\hat{\beta}^{(1)}$ be the MLE solution of the model $\mathcal{M}^{(\infty)}$ and $\hat{\mu}^{(1)}$ be the corresponding estimated expectation of y in the fitted model.

Then,

$$D_+(y, \hat{\mu}^{(1)}) - D_+(y, \hat{\mu}) = -2\left[L(\hat{\beta}^{(1)}) - L(\hat{\beta})\right]$$

- Additivity theorem (Efron Annals 1978)

$$D_+(\hat{\mu}, \hat{\mu}^{(1)}) = D_+(y, \hat{\mu}^{(1)}) - D_+(y, \hat{\mu})$$

  - Need to prove that $\sum_i (y_i - \hat{\mu}_i)\left(\theta_{\hat{\mu}_i} - \theta_{\hat{\mu}_i^{(1)}}\right) = 0$
  - Geometric interpretation for linear OLS

- Test for $H_0 : \beta^{(2)} = 0$. Under H_0,

$$D_+(\hat{\mu}, \hat{\mu}^{(1)}) = D_+(y, \hat{\mu}^{(1)}) - D_+(y, \hat{\mu}) \to \mathcal{X}^2_{p-p_1}$$

- R^2 in GLM:

$$1 - \frac{D_+(y, \hat{\mu})}{\sum_i D(y_i, \bar{y})}$$

- Deviance analysis table

Next time: Chapter 4.5 and 4.7, computation, building GLM example