STAT347: Generalized Linear Models
Lecture 3

Today's topics: Chapters 4.3-4.4

- Asymptotic distribution of the MLE estimates
- Hypothesis testing for β

1 Asymptotic distribution of the MLE estimates

- the MLE $\hat{\beta}$ is consistent when $n \to \infty$ and p is fixed.
- Asymptotic normality: when n is large

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

  where β_0 is the true value of the parameter. $(nV_{\beta_0}) = O(1))$

As an applied course, we ignore the discussions of the conditions of the above consistency and CLT results, and also skip the proofs.

1.1 Calculation of V_{β_0}

Delta method:

$$0 = \dot{L}(\hat{\beta}) \approx \dot{L}(\beta_0) + \ddot{L}(\beta_0)(\hat{\beta} - \beta_0)$$

The above approximation is a general approach and can be applied to any estimating equation $\phi(\hat{\beta}) = 0$ that results it a consistent estimate of β_0 (will discuss more in later lectures).

Thus

$$\hat{\beta} - \beta_0 \approx -\left(\ddot{L}(\beta_0)\right)^{-1} \dot{L}(\beta_0) = -\frac{1}{\sqrt{n}} \left(\frac{\ddot{L}(\beta_0)}{n}\right)^{-1} \left(\frac{\dot{L}(\beta_0)}{\sqrt{n}}\right)$$

- Under appropriate conditions, we have

  $$\ddot{L}(\beta_0)/n = \sum_i \ddot{L}_i(\beta_0)/n \to \text{Const.} \quad \text{(law of large numbers)}$$

  $$\frac{\dot{L}(\beta_0)}{\sqrt{n}} = \frac{\sum_i \dot{L}_i(\beta_0)}{\sqrt{n}} \xrightarrow{d} N(0, V) \quad \text{(central limit theorem)}$$

  Thus we have

  $$V_{\beta_0} = \left(\mathbb{E}\left(\ddot{L}(\beta_0)\right)\right)^{-1} \text{Var}\left(\dot{L}(\beta_0)\right) \left(\mathbb{E}\left(\ddot{L}(\beta_0)\right)\right)^{-1}$$

- property of the likelihood:
$$\mathrm{Var}\left(\dot{L}(\beta_0)\right) = \mathbb{E}\left(\left(\frac{\partial L}{\partial \beta}\,|_{\beta=\beta_0}\right)^2\right) = -\mathbb{E}\left(\ddot{L}(\beta_0)\right)$$
- $V_{\beta_0} = -\mathbb{E}\left(\ddot{L}(\beta_0)\right)^{-1}$
- $\hat{\beta}$ is more precise when $L(\beta)$ has larger curvature at β_0.
- See Chapter 4.2.4. $V_{\beta_0} = (X^T W X)^{-1}$ where $W = D^2 V^{-1}$

1.2 The distribution of any function $h(\hat{\beta})$

- $h(\hat{\beta})$ is a consistent estimator of $h(\beta_0)$
- Delta method:
$$h(\hat{\beta}) \approx h(\beta_0) + \dot{h}(\beta_0)^T(\hat{\beta} - \beta_0)$$
$$\sqrt{n}\left(h(\hat{\beta}) - h(\beta_0)\right) \to N\left(0, n\dot{h}(\beta_0)^T V_{\beta_0} \dot{h}(\beta_0)\right)$$
- Example: fitted values $h_i(\hat{\beta}) = g^{-1}(X_i^T \hat{\beta})$

2 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
$$H_0 : A\beta_0 = a_0 \quad V.S. \quad H_1 : A\beta_0 \neq a_0$$

2.1 Wald test

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

- $\widehat{\mathrm{Var}}(A\hat{\beta}) = A V_{\hat{\beta}} A^T$
- If $a_0 \in \mathbb{R}^1$, Wald statistic can also be written as
$$z = \frac{A\hat{\beta} - a_0}{\sqrt{\widehat{\mathrm{Var}}(A\hat{\beta})}}$$
- Under H_0, Wald statistic $z \,\dot{\sim}\, N(0,1)$
- We can also obtain a 95% CI for $A\beta_0$ as $[A\hat{\beta} - 1.96\sqrt{\widehat{\mathrm{Var}}(A\hat{\beta})}, A\hat{\beta} + 1.96\sqrt{\widehat{\mathrm{Var}}(A\hat{\beta})}]$
- When $a_0 \in \mathbb{R}^d$, then under H_0, $T \,\dot{\sim}\, \mathcal{X}_d^2$
- This is the GLM R function output for the analysis of each component β_j

2.2 A potential issue with Wald test

Let's look at an example of using Wald test for Binomial data $y_i \sim \text{Binomial}(n_i, p_i)$ where we work on the null model:

$$\log \frac{p_i}{1-p_i} = \log \frac{\mu_i}{n_i - \mu_i} = \beta_0$$

- As we use a canonical link, the asymptotic variance is $V_{\beta_0} = (X^T W X)^{-1}$ where $W = D^2 V^{-1} = D/a(\phi) = D$ (Lecture 2, section 2.2 and $a(\phi) = 1$ for Binomial data).
- $D_{ii} = \frac{1}{g'(\mu_i)} = \mu_i(n_i - \mu_i)/n_i$
- An estimate $\widehat{V}_{\beta_0} = V_{\hat{\beta}} = [(\sum_i n_i)\hat{p}(1-\hat{p})]^{-1}$ where $\hat{p}_i = \hat{p} = e^{\hat{\beta}}/(1 + e^{\hat{\beta}})$
- If we are interested in testing $H_0 : p_i \equiv 0.5$ or equivalently $H_0 : \beta_0 = 0$, the Wald statistics is

$$z = \hat{\beta}\sqrt{(\sum_i n_i)\hat{p}(1-\hat{p})}$$

- If we only have one sample with $y = 23$ and $n = 25$, then $z = 3.31$. If $y = 24$ and $n = 25$ then $z = 3.11$. Why do we have a smaller z when we have stronger evidence against the null?
- In the above specific example with only one sample, we can also obtain the CLT of $\hat{p} = y/n$, which result in another Wald statistics

$$z = \frac{\hat{p} - 0.5}{\sqrt{\hat{p}(1-\hat{p})/n}}.$$

  So the Wald statistics is not unique and depends on parameterization.
- We will discuss this more when we learn binary GLM (Chapter 5.3.3)

2.3 Score test

We only discuss the simple case

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

Last time we used the property of the likelihood that:

$$\text{Var}\left(\dot{L}(\beta_0)\right) = \mathbb{E}\left(\left(\frac{\partial L}{\partial \beta}\mid_{\beta=\beta_0}\right)^2\right) = -\mathbb{E}\left(\ddot{L}(\beta_0)\right)$$

where β_0 is the true value of the parameter. We construct the test statistics:

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

We make use of the asymptotic normal distribution of $\dot{L}(\beta_0)$. Under H_0, we have $T \to \mathcal{X}_p^2$ when $n \to \infty$.

2.4 Likelihood ratio test

We test for the null

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

where $a_0 \in \mathbb{R}^d$. 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 $n \to \infty$, under H_0

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

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

Next time: Chapters 4.4 - 4.5, deviance analysis, residuals and computation