STAT347: Generalized Linear Models
Lecture 2

Today's topics: Chapter 4.2

- maximum likelihood estimation for GLM
- Asymptotic distribution of $\hat{\beta}$

1 Likelihood score equations

Assume each observation y_i follows a one-parameter exponential distribution

$$f(y_i; \theta_i) = e^{y_i\theta_i - b(\theta_i)} f_0(y_i)$$

and the link function $g(\mu_i) = X_i^T\beta$. Then for n independent observations, the log likelihood is

$$L = \sum_i L_i = \sum_i [y_i\theta_i - b(\theta_i)] + \sum_i \log f_0(y_i)$$

1.1 For the canonical link

If $\theta_i = X_i^T\beta$, then

$$L = \sum_j (\sum_i y_i x_{ij})\beta_j - \sum_i b(X_i^T\beta) + \sum_i \log f_0(y_i)$$

- Score equation for β_j

$$\frac{\partial L}{\partial \beta_j} = \sum_i y_i x_{ij} - \sum_i b'(X_i^T\beta)x_{ij} = \sum_i (y_i - \mu_i)x_{ij} = 0$$

- score equation for a Poisson and linear model (Section 4.2.2)
- L is a concave function of β.

1.2 For a general link

Let $\eta_i = g(\mu_i)$ Then

$$\frac{\partial L_i}{\partial \beta_j} = \frac{\partial L_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}$$

We have

- $\frac{\partial L_i}{\partial \theta_i} = y_i - b'(\theta_i) = y_i - \mu_i$
- $\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta_i)} = \frac{1}{\mathrm{Var}(y_i)}$ (last lecture)
- $\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \mu_i}{\partial g(\mu_i)} = \frac{1}{g'(\mu_i)}$
- $\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}$

Thus, the score equation

$$\frac{\partial L}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{\mathrm{Var}(y_i)} \frac{1}{g'(\mu_i)} = 0$$

- The score equation only depends on the mean and variance of y
- Matrix form of the score equation:

$$\dot{L}(\beta) = X^T D V^{-1}(y - \mu) = 0$$

  where $V = \mathrm{diag}(\mathrm{Var}(y_1), \cdots, \mathrm{Var}(y_n))$ and $D = \mathrm{diag}(g'(\mu_1), \cdots, g'(\mu_n))^{-1}$, $y = (y_1, \cdots, y_n)$ and $\mu = (\mu_1, \cdots, \mu_n)$.

2 Asymptotic distribution of GLM

- 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. $(nV_{\beta_0}) = O(1))$

2.1 Calculation of V_{β_0}

Delta method:

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

Thus

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

- property of the likelihood: $\mathrm{Var}\left(\dot{L}(\beta_0)\right) = \mathbb{E}\left(\left(\frac{\partial L}{\partial \beta} \mid_{\beta=\beta_0}\right)^2\right) = -\mathbb{E}\left(\ddot{L}(\beta_0)\right)$
- $V_{\beta_0} = -\mathbb{E}\left(\ddot{L}(\beta_0)\right)^{-1}$
- $\hat{\beta}$ 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 $W = D^2 V^{-1}$

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

- $h(\hat{\beta})$ is a consistent estimator of $h(\beta_0)$
- Delta method:

$$h(\hat{\beta}) = h(\beta_0) + \dot{h}(\beta_0)^T(\hat{\beta} - \beta_0)$$

$$\sqrt{n}\left(h(\hat{\beta}) - h(\beta_0)\right) \to N\left(0, \dot{h}(\beta_0)^T V_{\beta_0} \dot{h}(\beta_0)\right)$$

- Example: fitted values $h_i(\hat{\beta}) = g^{-1}(X_i^T \hat{\beta})$

Next time: Chapter 4.3-4.4, Hypothesis testing, deviance