STOR 590: ADVANCED LINEAR MODELS

Instructor: Richard L. Smith

Class Notes:

August 26, 2020

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Model Selection: Nested Case

- Suppose we want to compare two models ω and Ω, where ω is a subset of Ω, $p_\omega < p_\Omega$ parameters

- Let $\widehat{\theta}_\omega, \widehat{\theta}_\Omega$ be the parameter estimates under both models

- $D = 2\{\ell(\theta_\omega) - \ell(\theta_\Omega)\} > 0$ is called the *deviance*

- If $H_0: \ \omega$ is true then the distribution of D is approximately $\chi^2_{p_\Omega - p_\omega}$ — analogous to the F-test for ANOVA.

- This is the *likelihood ratio test* (LRT). The text (Appendix A2, page 378) discusses two other tests, the *Wald test* and the *score test*, but the LRT is the one most used.

Model Selection: Comparing Many Models

- In practice, not all models are nested, and even if they were, doing many hypothesis tests is not usually a good idea (multiple testing or "data snooping" problem)

- Alternatives use automated selection criteria. Example are:
  - AIC: minimize $2\ell(\widehat{\theta}) + 2p$
  - BIC: minimize $2\ell(\widehat{\theta}) + p \log n$
  - DIC: minimize $D(\bar{\theta}) + 2p_D$ where D is deviance, $p_D = \overline{D(\theta)} - D(\bar{\theta})$ and $\bar{\cdot}$ denotes the mean

- *Note*: Faraway uses ℓ to denote the log likelihood, whereas I have used it for the negative log likelihood.

Example 1: Linear Regression with known σ^2

- $f(y_i;\beta) \propto \exp\left\{-\frac{1}{2\sigma^2}(y_i - \sum_j x_{ij}\beta_j)^2\right\}$

- Ignoring constants, $\ell(\beta) = \frac{1}{2\sigma^2}\sum_i (y_i - \sum_j x_{ij}\beta_j)^2$

- $\frac{\partial \ell}{\partial \beta_k} = \frac{1}{\sigma^2}\sum_i x_{ik}(y_i - \sum_j x_{ij}\beta_j)$

- $\frac{\partial^2 \ell}{\partial \beta_k \partial \beta_m} = \frac{1}{\sigma^2}\sum_i x_{ik}x_{im}$

- Setting $\frac{\partial \ell}{\partial \beta_k} = 0$ for all k gives the standard normal equations

- $H(\beta)$ or $I(\beta)$ are both $\frac{1}{\sigma^2}X^T X$ so they lead to the standard formula $\sigma^2\left(X^T X\right)^{-1}$ for the variance-covariance matrix of $\widehat{\beta}$.

Example 2:

Logistic Regression With One Covariate

- $f_i(y_i\ ;\ \theta)\ =\ \frac{\exp\{y_i(\beta_0+\beta_1 x_i)\}}{1+\exp(\beta_0+\beta_1 x_i)}$, $\theta = (\beta_0, \beta_1)$
- $\ell\ =\ \sum_i \log\{1 + \exp(\beta_0 + \beta_1 x_i)\} - \sum_i y_i(\beta_0 + \beta_1 x_i)$