STAT 410 - Linear Regression

Lecture 15

Meng Li

Department of Statistics

Nov 9, 2017

- Thus far we have been concerned with developing models where the response variable is numeric and ideally follows a normal distribution.
- In this lecture, we consider the situation in which the response variable is based on a series of “yes"/“no" responses.
- Ideally such responses follow a Bernoulli distribution or in general binomial distribution in which case the appropriate model is a logistic regression model.
- Data: $(y_i, \mathbf{x}_i)$ where y_i takes values either 1 or 0.
- It makes sense to assume $y_i \sim$ Bernoulli$(\mathbf{x}_i\boldsymbol{\beta})$ but $\mathbf{x}_i\boldsymbol{\beta}$ may violate the constraint that a success probability $\pi_i = E(y_i)$ is between 0 and 1.

Logistic regression

- We introduce the logistic transformation to guarantee the success probability takes values in the unit interval:

$$\pi_i = E(y_i) = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})} = \frac{1}{1+\exp(-\mathbf{x}_i'\boldsymbol{\beta})}, \tag{1}$$

  where the function $f(t) = 1/(1+\exp(-t))$ is the logistic function.
- In other words, we model the transformed mean using linear regression:

$$\log \frac{\pi_i}{1-\pi_i} = \mathbf{x}_i'\boldsymbol{\beta},$$

  where the function $g(\pi) = \log(\pi/(1-\pi))$ for $\pi \in (0,1)$ is the logit function.
- The logit function is the inverse of the logistic function.
- The ratio $\pi/(1-\pi)$ is called the odds.

Estimation

- Model: $y_i \sim \text{Bernoulli}(\pi_i)$ and $\pi_i = \frac{1}{1+\exp(-\mathbf{x}_i'\boldsymbol{\beta})}$.
  - Alternative form: $P(y_i = 1) = \frac{1}{1+\exp(-\mathbf{x}_i'\boldsymbol{\beta})}$.
- The density of each y_i is

$$\pi_i^{y_i}(1-\pi_i)^{1-y_i} = \left(\frac{\pi_i}{1-\pi_i}\right)^{y_i}(1-\pi_i).$$

- Therefore, the log-likelihood function is

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left\{y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log(1-\pi_i)\right\} \tag{2}$$

$$= \sum_{i=1}^{n} y_i \mathbf{x}_i'\boldsymbol{\beta} - \sum_{i=1}^{n}\log(1+\exp(\mathbf{x}_i'\boldsymbol{\beta})) \tag{3}$$

- The MLE is obtained by numerically maximizing the log-likelihood function though the Newton-Raphson algorithm or iteratively reweighted least squares.

Inference

- The asymptotic distributions of $\hat{\boldsymbol{\beta}}$ is available due to the properties of MLE.
  - “Asymptotic" means the sample size is sufficiently large.
  - Under regular conditions, any MLE is asymptotically normal with mean $\boldsymbol{\beta}$ and covariance matrix $I(\boldsymbol{\beta})^{-1}$, where $I(\boldsymbol{\beta})$ is the Fisher information contained in the full data, i.e., the expectation of negative second derivatives of $\log L(\boldsymbol{\beta})$.
- Consequently, we have $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{XVX}')^{-1}$, where $\mathbf{V} = \text{diag}(\pi_1(1-\pi_1), \ldots, \pi_n(1-\pi_n))$, asymptotically.
- This enables us to conduct hypothesis testing on $\boldsymbol{\beta}$ and construct confidence intervals.

Comments

- Similarly to linear regression, we have diagnostic checking, model comparison etc. for logistic regression.
- The logistic function fits into the framework of generalized linear models (GLM), which allows for Binomial, Poisson, and general exponential family distributions.
- Other link functions are available, for example, the Probit model uses
$$\pi_i = \Phi(\mathbf{x}_i'\boldsymbol{\beta}),$$
where $\Phi(\cdot)$ is the CDF of standard normal. The function $\Phi^{-1}(\cdot)$ is called the probit function.
- In general, the selection of link functions are not important.