STAT/BIOSTAT 534 Statistical Computing
Univariate Logistic Regression

Adrian Dobra
adobra@uw.edu

We assume to have observed the n independent samples

$$\mathcal{D} = \{(y_1, x_1), \ldots, (y_n, x_n)\}.$$

The response variable Y is binary, i.e. $y_i \in \{0, 1\}$ for $i = 1, 2, \ldots, n$. The explanatory variable X can be continuous or discrete. We consider the univariate logistic regression model

$$\log \frac{P(y = 1|x)}{P(y = 0|x)} = \beta_0 + \beta_1 x. \tag{1}$$

Our model assumptions say that each y_i follows a Bernoulli distribution with probability of success $\pi_i = P(y_i = 1|x_i)$:

$$y_i \sim Ber(\pi_i).$$

Since the samples are assumed to be independent, the likelihood is:

$$L(\beta_0, \beta_1|\mathcal{D}) = \prod_{i=1}^{n} [P(y_i = 1|x_i)]^{y_i} [1 - P(y_i = 1|x_i)]^{1-y_i},$$

where

$$\pi_i = P(y_i = 1|x_i) = logit^{-1}(\beta_0 + \beta_1 x_i) \in (0, 1).$$

We define the *logit* function:

$$logit : (0, 1) \longrightarrow (-\infty, +\infty), \quad logit(p) = \log \frac{p}{1 - p}.$$

Its inverse is:

$$logit^{-1} : (-\infty, +\infty) \longrightarrow (0, 1), \quad logit^{-1}(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

The log-likelihood is

$$\begin{aligned} l(\beta_0, \beta_1|\mathcal{D}) &= \log L(\beta_0, \beta_1|\mathcal{D}), & (2) \\ &= \sum_{i=1}^{n} (y_i \log \pi_i + (1 - y_i) \log[1 - \pi_i]). & (3) \end{aligned}$$

Simple calculations show that

$$
\begin{aligned}
\frac{\partial l(\beta_0,\beta_1|\mathcal{D})}{\partial \beta_0} &= \sum_{i=1}^{n}[y_i - \pi_i],\\
\frac{\partial l(\beta_0,\beta_1|\mathcal{D})}{\partial \beta_1} &= \sum_{i=1}^{n}[y_i x_i - \pi_i x_i].
\end{aligned}
$$

The gradient of $l(\beta_0,\beta_1)$ is

$$
\nabla l(\beta_0,\beta_1) = \left(\begin{array}{c} \frac{\partial l(\beta_0,\beta_1)}{\partial \beta_0} \\ \frac{\partial l(\beta_0,\beta_1)}{\partial \beta_1} \end{array} \right).
$$

The Hessian matrix associated with $l(\beta_0,\beta_1)$ is

$$
D^2 l(\beta_0,\beta_1) = \left[\begin{array}{cc} \frac{\partial^2 l(\beta_0,\beta_1)}{\partial \beta_0^2} & \frac{\partial^2 l(\beta_0,\beta_1)}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 l(\beta_0,\beta_1)}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 l(\beta_0,\beta_1)}{\partial \beta_1^2} \end{array} \right].
$$

We determine the mode log-likelihood (2), i.e.

$$
\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = \mathrm{argmax}_{(\beta_0,\beta_1)\in\Re^2} l(\beta_0,\beta_1),
$$

by employing the Newton-Raphson algorithm. The procedure starts with the initial values $\left(\beta_0^{(0)}, \beta_1^{(0)}\right) = (0,0)$. At iteration k, we update our current estimate $\left(\beta_0^{(k-1)}, \beta_1^{(k-1)}\right)$ of the mode $\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$ to a new estimate $\left(\beta_0^{(k)}, \beta_1^{(k)}\right)$ as follows:

$$
\left(\begin{array}{c} \beta_0^{(k)} \\ \beta_1^{(k)} \end{array} \right) = \left(\begin{array}{c} \beta_0^{(k-1)} \\ \beta_1^{(k-1)} \end{array} \right) - \left[D^2 l(\beta_0^{(k-1)}, \beta_1^{(k-1)}) \right]^{-1} \nabla l(\beta_0^{(k-1)}, \beta_1^{(k-1)}).
$$

The procedure stops when the estimates of the mode do not change after performing a new update, i.e. $|\beta_0^{(k)} - \beta_0^{(k-1)}| < \epsilon$ and $|\beta_1^{(k)} - \beta_1^{(k-1)}| < \epsilon$. Here ϵ is some small positive number, e.g. 0.0001.