Binary logistic regression

Mark Andrews
Psychology Department, Nottingham Trent University

mark.andrews@ntu.ac.uk

Logistic regression's assumed model (simple case)

- For all $i \in 1 \ldots n$,

$$\begin{aligned} y_i &\sim \text{Bernoulli}(\theta_i), \\ \text{logit}(\theta_i) &= a + bx_i. \end{aligned}$$

or equivalently

$$\begin{aligned} y_i &\sim \text{Bernoulli}(\theta_i), \\ \theta_i &= \text{ilogit}(a + bx_i), \end{aligned}$$

where

$$\text{logit}(\theta_i) \triangleq \log\left(\frac{\theta}{1-\theta}\right),$$

and

$$\text{ilogit}(a + bx_i) \triangleq \frac{1}{1 + e^{-(a+bx_i)}}$$

Logistic regression's assumed model (multiple regression case)

- For all $i \in 1 \ldots n$,

$$y_i \sim \mathrm{Bernoulli}(\theta_i),$$

$$\mathrm{logit}(\theta_i) = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}$$

or equivalently

$$y_i \sim \mathrm{Bernoulli}(\theta_i),$$

$$\theta_i = \mathrm{ilogit}(\beta_0 + \sum_{k=1}^{K} \beta_k x_{ki}).$$

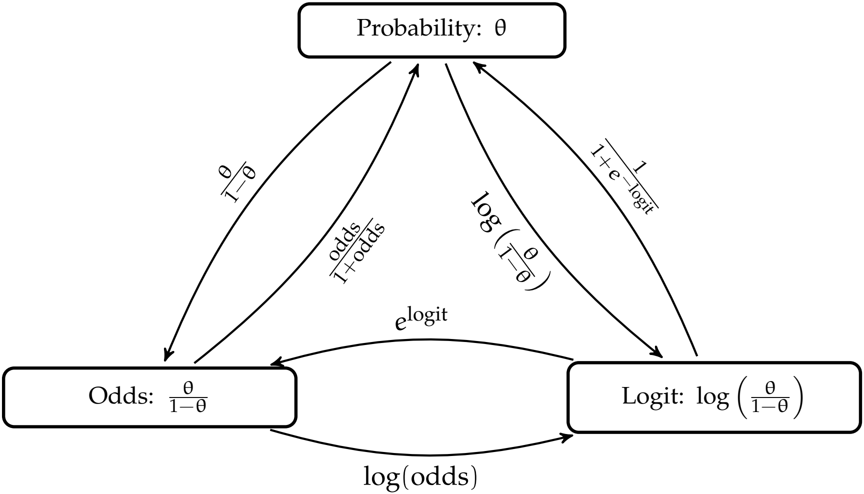
Prediction

- Given inferred values for $\beta_0, \beta_1 \ldots \beta_K$, the predicted log odds of the outcome variable taking the value of 1 if the predictor variables's values are $x_1, x_2 \ldots x_K$ is

$$\beta_0 + \sum_{k=1}^{K} \beta_k x_k$$

- Knowing the predicted log odds, the predicted probability or predicted odds is easily calculated.

From probabilities to odds to logits, and back



Understanding β coefficients

- In linear models, a coefficient for a predictor variable has a straightforward interpretation: 1 unit change for a predictor variable corresponds to β change in the outcome variable.
- As logistic regression curves are nonlinear, the change in the outcome variable is not a constant function of change in the predictor.
- This makes interpretation more challenging.
- The most common means to interpret β coefficients is in terms of odds ratios.

Odds ratios

- We have seen that an odds in favour of an event are $\frac{p}{1-p}$.
- We can compare two odds with an odds ratio.
- For example, the odds of getting a certain job for someone with a MBA might be $\frac{p}{1-p}$, while the odds of getting the same job for someone without an MBA might be $\frac{q}{1-q}$.
- The ratio of the odds for the MBA to those of the non-MBA are

$$\frac{p}{1-p} \Big/ \frac{q}{1-q}$$

- This gives the factor by which odds for the job change for someone who gains an MBA.

β *coefficients as (log) odds ratios*

- Consider a logistic regression model with a single dichotomous predictor, i.e.

$$\log\left(\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right) = \alpha + \beta x_i,$$

  where $x_i \in \{0, 1\}$.

- The log odds that $y_i = 1$ when $x_i = 1$ is $\alpha + \beta$.
- The log odds that $y_i = 1$ when $x_i = 0$ is α.
- The log odds that $y_i = 1$ when $x_i = 1$ minus the log odds that $y_i = 1$ when $x_i = 0$ is

$$(\alpha + \beta) - \alpha = \beta.$$

β *coefficients as (log) odds ratios*

- Let's denote the probability that $y_i = 1$ when $x_i = 1$ by p, and denote the probability that $y_i = 1$ when $x_i = 0$ by q.
- Subtracting the log odds is the log of the odds ratio, i.e.

$$\log\left(\frac{p}{1-p}\right) - \log\left(\frac{q}{1-q}\right) = \log\left(\frac{p}{1-p}\Big/\frac{q}{1-q}\right) = \beta$$

- As such,

$$e^{\beta} = \frac{p}{1-p}\Big/\frac{q}{1-q}.$$

- This provides a general interpretation for the β coefficients.

Model Fit with Deviance

- Once we have the maximum likelihood estimate for the parameters, we can calculate *goodness of fit*.
- The *deviance* of a model is defined

$$-2 \log L(\hat{\alpha}, \hat{\beta} | \mathcal{D}),$$

  where $\hat{\alpha}$, $\hat{\beta}$ are the mle estimates.
- This is counterpart to R^2 for generalized linear models.

Model Fit with Deviance: Model testing

- In a model with one predictor, a null model would be that $P(y_i = 1)$ is not a function of x_i.
- The difference in the deviance of the null model minus the deviance of the full model is

$$\Delta_D = D_0 - D_1 = -2 \log \frac{L(\hat{\alpha}|\mathcal{D})}{L(\hat{\alpha}, \hat{\beta}|\mathcal{D})}.$$

- Under the null hypothesis, Δ_D is distributed as χ^2 with 1df.

Deviance based model testing

- In general, we can compare any two *nested* models using χ^2 test applied to differences in deviance.
- The deviance of the subset model minus that of the full model will always be (approximately) distributed a χ^2 with df equalling the difference in the number of parameters between the two models.
- In other words, under the null hypothesis that subset and full models are identical, the difference in the deviances will be distributed as a χ^2 with df equal to the difference in the number of parameters between the two models.