Multinomial Logistic Regression

Jeffrey Woo

MSDS, University of Virginia

Response with More than Two Classes

Previously, we learned the (binary) logistic regression model where the response is categorical with two classes. The multinomial logistic regression model is used to handle response variables with more than 2 classes.

1. Setup

2. Interpreting Coefficients

Setup

In binary logistic regression model, we modeled the log odds as a linear combination of the predictors, i.e.,

$$\operatorname{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X}\boldsymbol{\beta}, \tag{1}$$

where π denotes the probability of "success", so $1 - \pi$ denotes the probability of "failure".

Setup

Another way to express (1) is to denote the probabilities of "success" and "failure" as π_1 and π_2 respectively. So the logistic regression model in (1) can be written as

$$\log\left(\frac{\pi_1}{\pi_2}\right) = \boldsymbol{X}\boldsymbol{\beta}. \tag{2}$$

The formulation above in (2) is used for multinomial logistic regression.

Notation

Suppose there are $m + 1$ classes for the response variable. Let $\pi_{i,c}$ denote the probability that observation i belongs to class c for $c = 1, 2, \cdots, m + 1$. Using the notation for binary logistic regression in (2), we have

$$\pi'_{i,1,2} = \log\left(\frac{\pi_{i,1}}{\pi_{i,2}}\right) = \boldsymbol{X_i'}\boldsymbol{\beta_{1,2}}.$$

We use $\pi'_{i,1,2}$ and $\boldsymbol{\beta_{1,2}}$ to emphasize that we are modeling the log of the ratio of the probabilities for classes 1 and 2. $\log\left(\frac{\pi_{i,1}}{\pi_{i,2}}\right)$ is called a multinomial **logit**.

Logits

Suppose we have a response variable with 3 classes. There are 3 pairs of classes that we can model, so we have 3 logits:

$$\begin{aligned}
\pi'_{i,1,2} &= \log\left(\frac{\pi_{i,1}}{\pi_{i,2}}\right) = \boldsymbol{X_i'}\beta_{\mathbf{1,2}}, \\
\pi'_{i,1,3} &= \log\left(\frac{\pi_{i,1}}{\pi_{i,3}}\right) = \boldsymbol{X_i'}\beta_{\mathbf{1,3}}, \\
\pi'_{i,2,3} &= \log\left(\frac{\pi_{i,2}}{\pi_{i,3}}\right) = \boldsymbol{X_i'}\beta_{\mathbf{2,3}}.
\end{aligned}$$

Logits

It turns out we only need 2, or 1 less than the number of classes, logits. One class will be chosen as the **reference class**, and all the other classes will be compared to the reference class. So, suppose class $m+1$ is the reference class, we need to consider only m comparisons to the reference class. The logit for the cth comparison is

$$\pi'_{i,c,m+1} = \log\left(\frac{\pi_{i,c}}{\pi_{i,m+1}}\right) = \boldsymbol{X_i'}\boldsymbol{\beta_{c,m+1}}$$

for $c = 1, \cdots, m$. Note that $\frac{\pi_{i,c}}{\pi_{i,m+1}}$ is called the **relative risk** of belonging to class c versus belonging to the reference class.

Logits

Since we consider class $m+1$ to be the reference, all comparisons are made with class $m+1$, so we let $\pi'_{i,c} = \pi'_{i,c,m+1}$ and $\boldsymbol{\beta_c} = \boldsymbol{\beta_{c,m+1}}$, so the logits are

$$\pi'_{i,c} = \log\left(\frac{\pi_{i,c}}{\pi_{i,m+1}}\right) = \boldsymbol{X_i'}\boldsymbol{\beta_c} \tag{3}$$

for $c = 1, \cdots, m$. We can view (3) as modeling the log relative risk as a linear combination of the predictors.

Logits

We only need m logits because any other comparisons can be obtained from them. For example, suppose $m + 1 = 4$, and we wish to compare classes 1 and 2. We have:

$$\begin{aligned}
\log\left(\frac{\pi_{i,1}}{\pi_{i,2}}\right) &= \log\left(\frac{\pi_{i,1}}{\pi_{i,4}} \times \frac{\pi_{i,4}}{\pi_{i,2}}\right) \\
&= \log\left(\frac{\pi_{i,1}}{\pi_{i,4}}\right) - \log\left(\frac{\pi_{i,2}}{\pi_{i,4}}\right) \\
&= \boldsymbol{X_i'}\beta_{\mathbf{1}} - \boldsymbol{X_i'}\beta_{\mathbf{2}}.
\end{aligned}$$

Logits

In general, to compare classes k and l, we have

$$\log\left(\frac{\pi_{i,k}}{\pi_{i,l}}\right) = \boldsymbol{X_i'}(\beta_{\boldsymbol{k}} - \beta_{\boldsymbol{l}}).$$

Probabilities for Classes

The probabilities for each class can be found by

$$\pi_{i,c} = \frac{\exp\left(\boldsymbol{X_i'}\beta_{\boldsymbol{c}}\right)}{1 + \sum_{k=1}^{m} \exp\left(\boldsymbol{X_i'}\beta_{\boldsymbol{k}}\right)} \tag{4}$$

for $c = 1, \cdots, m$. For the reference class, the probability is

$$\pi_{i,m+1} = \frac{1}{1 + \sum_{k=1}^{m} \exp\left(\boldsymbol{X_i'}\beta_{\boldsymbol{k}}\right)} \tag{5}$$

Interpreting Coefficients

The coefficients can be interpreted in one of the following ways:

- The $(k+1)$th element of $\boldsymbol{\beta_c}$ can be interpreted as the **difference in log relative risk** of belonging to class c versus the reference class with a one-unit increase in the kth predictor, given the other predictors are held constant; OR
- The **relative risk ratio** of belonging to class c versus the reference class with a one-unit increase in the kth predictor, given the other predictors are held constant, is the $(k+1)$th element of $\exp(\boldsymbol{\beta_c})$; OR
- The relative risk of belonging to class c versus belonging to the reference class is **multiplied** by the $(k+1)$th element of $\exp(\boldsymbol{\beta_c})$ with a one-unit increase in the kth predictor, given the other predictors are held constant

Worked Example: Contraceptives in Indonesia

Data from national survey of married women in Indonesia.

- y: contraceptive method use (no use, long-term, short-term)
- x_1: number of children
- x_2: religion (coded 1 for Islam, 0 for other)

Worked Example: Contraceptives in Indonesia

```
Coefficients:
           (Intercept)   children wife_relIslam
long-term   -0.4334275 0.15751853    -0.8739442
short-term  -0.3080111 0.08266053    -0.1822090

Std. Errors:
           (Intercept)   children wife_relIslam
long-term    0.1835733 0.02914340     0.1819265
short-term   0.1818469 0.02639178     0.1793848
```

Worked Example: Contraceptives in Indonesia

Coefficient for children for first logit: The relative risk of using long-term contraceptives versus not using contraceptives is multiplied by $\exp(0.1575) = 1.171$ for each additional child, while controlling for religion.

Worked Example: Contraceptives in Indonesia

For a woman who has 2 children and is not muslim, what is her predicted probability of using long-term contraceptives? Short-term contraceptives? No contraceptives?
Using (4) and (5), we have

$$\pi_{i,c} = \frac{\exp\left(\boldsymbol{X_i'}\beta_{\boldsymbol{c}}\right)}{1 + \sum_{k=1}^{m} \exp\left(\boldsymbol{X_i'}\beta_{\boldsymbol{k}}\right)}$$

for $c = 1, \cdots, m$. For the reference class, the probability is

$$\pi_{i,m+1} = \frac{1}{1 + \sum_{k=1}^{m} \exp\left(\boldsymbol{X_i'}\beta_{\boldsymbol{k}}\right)}$$