Logit models for multinomial responses

Today’s topics:

1. Introduction
2. Models for nominal responses
3. Models for ordered responses
4. Conditional independence

Sections skipped:
7.4.3-7.4.7

M. de Rooij 7.1
Introduction

When the dependent variable is not binary but has multiple outcome categories we use the multinomial instead of binomial for modeling.

Two types of multcategory outcome can be considered: nominal and ordinal.

We start with a logit-link type of model, but also some other links will be discussed
Nominal response variable

In logistic regression the log odds of 'success' is predicted.

When there are \( J \) outcome categories, the multinomial logit model predicts simultaneously all pairs of log odds. Not all are necessary: a good choice of \( J - 1 \) provides all information.

Let \( \pi_j(x) = P(Y = j|x) \) for given \( x \), with \( \sum_j \pi_j(x) = 1 \).

Normally a baseline category is chosen (the first or last). Logit models, then pair each response category with this baseline. The model

\[
\log \frac{\pi_j(x)}{\pi_J(x)} = \alpha_j + \beta_j x
\]

simultaneously describes the effects of \( x \) on these \( J - 1 \) logits.

Given these logits all other can be derived since:

\[
\log \frac{\pi_a(x)}{\pi_b(x)} = \log \frac{\pi_a(x)}{\pi_J(x)} - \log \frac{\pi_b(x)}{\pi_J(x)}
\]

With categorical predictors (grouped data) standard \( X^2 \) and \( G^2 \) statistics can be used for model checking.

M. de Rooij 7.3
Nominal response variable

<table>
<thead>
<tr>
<th>logit</th>
<th>intercept</th>
<th>Size</th>
<th>Hancock</th>
<th>Lake</th>
<th>Oklawaha</th>
<th>Trafford</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\log(\pi_I/\pi_F))</td>
<td>-1.55</td>
<td>1.46</td>
<td>-1.66</td>
<td>0.94</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>(\log(\pi_R/\pi_F))</td>
<td>-3.31</td>
<td>-0.35</td>
<td>1.24</td>
<td>2.46</td>
<td>2.94</td>
<td></td>
</tr>
<tr>
<td>(\log(\pi_B/\pi_F))</td>
<td>-2.09</td>
<td>-0.63</td>
<td>0.70</td>
<td>-0.65</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>(\log(\pi_O/\pi_F))</td>
<td>-1.90</td>
<td>0.33</td>
<td>0.83</td>
<td>0.01</td>
<td>1.52</td>
<td></td>
</tr>
</tbody>
</table>

1. Log odds against Fish

2. Dummy predictors where Lake George and Large have value 0

3. For each log odds a different equation

4. The \(\log(\pi_I/\pi_R)\) can be derived from \(\log(\pi_I/\pi_F) - \log(\pi_R/\pi_F)\) by subtracting each of the values: New intercept is \(-1.55 - (-3.31) = 1.76\), etc.
Nominal response variable

Response probabilities can be obtained from

$$\pi_j(x) = \frac{\exp(\alpha_j + \beta'_j x)}{1 + \sum_{h=1}^{J-1} \exp(\alpha_h + \beta'_h x)}$$

When there are continuous explanatory variables it is helpful to plot the probabilities against the values of the predictor keeping others fixed.
Fitting Nominal response variable

The procedure for fitting multinomial logit models is very similar to fitting logistic regressions. Note that it is not equal to separately fitting binomial logit models.

The multinomial response model can also be seen as a multivariate GLM, with each log odds as dependent variable. A IRLS-algorithm can be used, but the weights matrix is not diagonal anymore (but it is block-diagonal).
Nominal response variable

lEM example

* Multinomial logit model
* data; Agresti 2002
* Tables 7.3 and 7.4 (primary food choice of alligators)
* L=lake; S=size; F=food

man 3
dim 4 2 5
lab L S F
mod F|LS {FL,FS}
dat [23 4 2 2 8 ...
17 1 0 1 3]
dum 4 2 1
Ordinal response variable

Cumulative logit models define cumulative probabilities

\[ P(Y \leq j|x) = \pi_1(x) + \pi_2(x) + \ldots + \pi_j(x), \quad j = 1, \ldots, J - 1. \]

and from these the cumulative logits

\[ \text{logit}[P(Y \leq j|x)] = \log \frac{P(Y \leq j|x)}{1 - P(Y \leq j|x)} \]

A model for a single logit\([P(Y \leq j|x)]\) alone is a standard logit model. It is better to simultaneously model all cumulative logits (i.e. for different \(j\).)
**Proportional odds model**

The proportional odds model models the set of cumulative odds ratios as:

\[
\text{logit}[P(Y \leq j|x)] = \alpha_j + \beta'x
\]

that is with different intercepts (which are increasing due to the cumulative nature) and equals slopes.

The proportional odds model has the following property:

\[
\text{logit}[P(Y \leq j|x_1)] - \text{logit}[P(Y \leq j|x_2)] = \beta'(x_1 - x_2)
\]

the odds of making a response \( \leq j \) with predictor value \( x_1 \) are \( \exp(\beta'(x_1 - x_2)) \) times the odds at \( x_2 \), i.e. it is proportional to the distance between \( x_1 \) and \( x_2 \), and is independent of \( j \).
If the proportional odds model fails to fit the data

1. try another (asymmetric) link function such as the complementary log-log
2. include more predictors, i.e. quadratic or interaction predictors
3. allow for separate effects for some predictors $\beta_j$ but not all.
4. fit the baseline-category model.
Cumulative link models

The cumulative link model

\[ G^{-1}[P(Y \leq j|x)] = \alpha_j + \beta'x \]

where the link can be standard normal cdf or the complementary log-log. Just as last week.
Proportional odds model

\ell EM example

* Cumulative logit model
* data: agresti 2002, table 7.5
* cum(a) is logit link
* cum(b) is probit link
* cum(c) is complementary log-log link
man 1
con 2
dim 4
lab M x
mod M|x cum(a) \{cov(x,1) cov(x,2)\}
rec 40
dat [1 1 1
... 4 0 9]
Adjacent categories logit models

The logits of adjacent categories are modeled, i.e., the logits of ‘one step forward’. An example of such a model is:

$$\log \frac{\pi_j(x)}{\pi_{j+1}(x)} = \alpha_j + \beta' x$$

where a common effect $\beta$ is used.

When we use $\beta_j$ (instead of a single $\beta$) the adjacent category logit model is equivalent to baseline category models. The parameters from one model can be turned into the parameters of the other. And therefore, the baseline category model can be used to estimate this model.

This is the same as the difference between local odds ratios and spanning cell odds ratios.
Testing conditional independence

Like for the $2 \times 2 \times K$ case we can use likelihood ratio statistics, Wald statistics (both model based) and Cochran-Mantel-Haenszel statistics for testing conditional independence in the $J \times I \times K$ table.

Conditional independence = independence of two variables given any relations between the other variables.

For both the likelihood ratio statistic and Wald statistic a suitable model has to be defined. When $Y$ is ordinal cumulative logit models are used (although other ordinal models might be chosen).

<table>
<thead>
<tr>
<th>$Y - X$</th>
<th>Model</th>
<th>$H_0$</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ord-Ord</td>
<td>$\text{logit}[P(Y \leq j)] = \alpha_j + \beta x_i + \beta_k Z$</td>
<td>$\beta = 0$</td>
<td>1</td>
</tr>
<tr>
<td>Ord-Nom</td>
<td>$\text{logit}[P(Y \leq j)] = \alpha_j + \beta_i X + \beta_k Z$</td>
<td>$\beta_i X = 0$</td>
<td>$(I - 1)$</td>
</tr>
<tr>
<td>Nom-Ord</td>
<td>$\log \left[ \frac{P(Y=j)}{P(Y=J)} \right] = \alpha_{jk} + \beta_j x_i$</td>
<td>$\beta_j = 0$</td>
<td>$(J - 1)$</td>
</tr>
<tr>
<td>Nom-Nom</td>
<td>$\log \left[ \frac{P(Y=j)}{P(Y=J)} \right] = \alpha_{jk} + \beta_{ij}^{XY}$</td>
<td>$\beta_{ij}^{XY} = 0$</td>
<td>$(I - 1)(J - 1)$</td>
</tr>
</tbody>
</table>