Proportional odds models

Patrick Breheny

April 23

Introduction

- Multinomial regression requires the estimation of $(K-1)p$ parameters, and assumes nothing about the relationship between the categories
- This is very flexible of course, but has two downsides:
    - The large number of parameters can be cumbersome to interpret
    - Estimating a large number of parameters can result in high variability in the estimates
- When the categories are *ordered*, making assumptions about the relationships between them allows us to introduce some structure and estimate fewer parameters, decreasing variability and increasing interpretability

Martin county coal spill

- In October 2000, a coal slurry impoundment ruptured, emptying more than 300 million gallons of toxic coal waste into the streams of Martin County, Kentucky
- Researchers from the sociology department at the University of Kentucky carried out a study of the disaster's effect on trust in the community
- In the months following the disaster, and then again 10 years later, a survey was administered to residents of Martin County

Martin county study

- Among the survey items was the statement, "I have trust in the local government"; respondents were asked to choose their reaction to that statement from among the following options:
  - Strongly disagree
  - Disagree
  - Neutral
  - Agree
  - Strongly agree
- Such items are often referred to as being measured on a *Likert scale* after their inventor, Rensis Likert
- In addition to trust in local government and year (2011 vs. 2001), we will also consider the demographic variable Education, recorded as Less than high school/High school/Some college/College degree

Cumulative logits

- Note that, unlike the case with alligator food choices, the response categories here are ordered, which suggests a certain relationship between them

- It would be odd, for example, if comparing 2011 vs. 2001 we came to the conclusion that the number of "strongly agree" and "disagree" responses went up significantly, but that the number of "agree" and "strongly disagree" responses went down significantly

- To address this ordering, we can focus on the *cumulative logits*:

$$\log \left(\frac{\Pr(Y \leq k)}{\Pr(Y > k)} \right) = \log \left(\frac{\pi_1 + \cdots + \pi_k}{\pi_{k+1} + \cdots + \pi_K} \right)$$

The proportional odds model

- The *proportional odds model* assumes that each explanatory variable exerts the same effect on each cumulative logit regardless of the cutoff k:

$$\log\left(\frac{\Pr(Y \le k)}{\Pr(Y > k)}\right) = \alpha_k + \mathbf{x}^T\boldsymbol{\beta}$$

- Note that:
  - Each cumulative logit has its own intercept, but each explanatory variable only has a single coefficient β; thus, the model has fewer terms than a multinomial regression model
  - Writing down the proportional odds model requires us to modify the notation we've used all semester; in the above, $\mathbf{x}$ and $\boldsymbol{\beta}$ do not include a term for the intercept

Alternative parameterization

- Note that the description of the model given on the previous slide is perhaps a bit counterintuitive, in that high values of $\eta = \alpha_k + \mathbf{x}^T \boldsymbol{\beta}$ are associated with *low* values of Y

- For this reason, many people prefer to specify the model as

$$\log \left(\frac{\Pr(Y \leq k)}{\Pr(Y > k)} \right) = \alpha_k - \mathbf{x}^T \boldsymbol{\beta},$$

so that the sign of $\boldsymbol{\beta}$ has the usual meaning (*i.e.*, if positive, an increase in x is associated with an increase in Y)

- This is the formulation adopted by R; SAS uses the formulation on the previous slide

Coefficients and probabilities

- Suppose we wish to calculate $\Pr(Y = k|\mathbf{x})$ based on our model

- For the simple case where $k = 1$, the calculations are identical to logistic regression:

$$\Pr(Y = 1|\mathbf{x}) = \Pr(Y \le 1|\mathbf{x}) = \frac{\exp(\alpha_1 + \mathbf{x}^T\boldsymbol{\beta})}{1 + \exp(\alpha_1 + \mathbf{x}^T\boldsymbol{\beta})}$$

- Calculating the probabilities for other categories requires a bit more work:

$$\Pr(Y = k|\mathbf{x}) = \Pr(Y \le k|\mathbf{x}) - \Pr(Y \le k - 1|\mathbf{x})$$

Coefficients and odds ratios

- As with regular logistic regression, a key advantage of the logit link is that additive models yield constant odds ratios

- Consider comparing two arbitrary individuals, with covariates $\mathbf{x}_2$ and $\mathbf{x}_1$:

$$\frac{\Pr(Y > k|\mathbf{x}_2)/\Pr(Y \le k|\mathbf{x}_2)}{\Pr(Y > k|\mathbf{x}_1)/\Pr(Y \le k|\mathbf{x}_1)} = -(\mathbf{x}_2 - \mathbf{x}_1)^T \boldsymbol{\beta};$$

this is for the first (SAS-style) parameterization; for the second (R-style) parameterization,

$$\frac{\Pr(Y > k|\mathbf{x}_2)/\Pr(Y \le k|\mathbf{x}_2)}{\Pr(Y > k|\mathbf{x}_1)/\Pr(Y \le k|\mathbf{x}_1)} = (\mathbf{x}_2 - \mathbf{x}_1)^T \boldsymbol{\beta}$$

Remarks

- Thus, for the usual case where we consider changing only a single parameter at a time by one unit, e^β represents the odds ratio, as usual in logistic regression

- The difference, however, is that e^β now represents a cumulative odds ratio: the odds of "at least k" under two different conditions

- Note that we get exactly the same odds ratio for comparing {Neutral, Agree, Strongly agree} vs {Disagree, Strongly disagree} as when comparing {Agree, Strongly agree} vs {Neutral, Disagree, Strongly disagree}

- The odds ratio is constant across each split; hence the name "proportional odds model"

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction

Results: Odds ratio

- For the model with `Year` as the only explanatory variable, we have $\hat{\beta} = 0.618$ and

	OR	95% CI Lower	Upper	p
Year	1.9	1.3	2.6	0.0003

- One valid interpretation of this finding would be that the odds of agreeing with the statement "I trust the local government" nearly doubled between 2001 and 2011

- Or, equally valid, the odds of disagreeing with the statement fell by nearly half

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction

Descriptive statistics

Of course, this is something of an oversimplification, in that the actual increases were not perfectly proportional across all categories:

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction
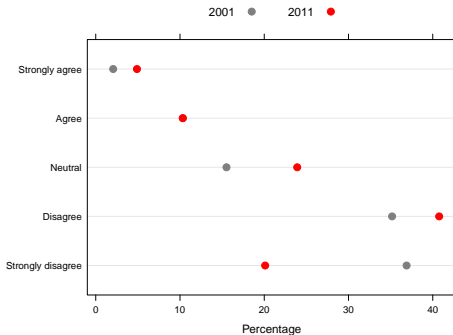
Estimated probabilities

- Estimated probability of each response, subject to model restrictions:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
2001	0.35	0.38	0.16	0.08	0.02
2011	0.23	0.37	0.22	0.14	0.04

- Observed ("raw") proportions:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
2001	0.37	0.35	0.16	0.10	0.02
2011	0.20	0.41	0.24	0.10	0.05

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction

Proportional odds vs. multinomial models

- Multinomial regression avoids the proportional odds assumption, allowing the possibility of capturing non-proportional trends in the explanatory variables

- The downsides, however, are that we don't simply get an OR for Year, we have separate ORs for Agree vs. Disagree, Strongly agree vs. neutral, neutral vs. disagree, etc.

- This is a bit cumbersome to interpret; for example, the estimated agree vs. disagree odds ratio for the multinomial model is 0.86, suggesting the opposite trend (a decline in trust over time) from most other comparisons

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction

Variability

- Furthermore, the multinomial model estimates have higher variance

- For example, $SE_{\hat{\beta}} = 0.17$ for the proportional odds model, $SE_{\hat{\beta}} = 0.33$ for the multinomial model comparing agree to disagree

- To compare the models, we could use AIC; here, the proportional odds model has AIC 1312, while the multinomial model has AIC 1313, indicating that the violation of proportional odds is not substantial enough to warrant all the extra parameters that the multinomial model introduces

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction

Some model selection

- Investigating the fit of various models:

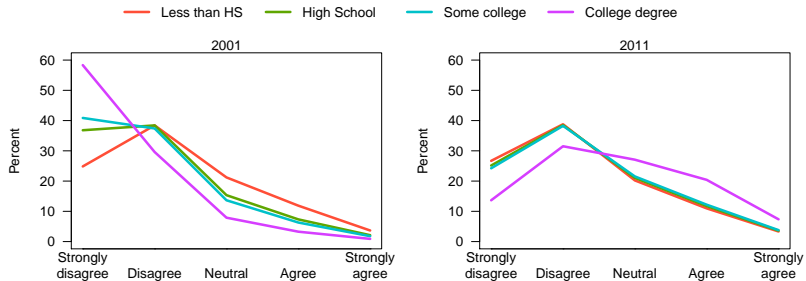|	AIC
(Intercept)	1324
Year	1312
Year + Educ	1314
Year × Educ	1309

- Interestingly, education seems to add very little as an additive effect, but has a somewhat important interaction with Year

- It is worth noting that assuming a linear trend for education has an even better fit (AIC 1303), but for the sake of illustration, I'll continue treating education as a categorical variable

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction

Education-specific odds ratios

		95% CI	
Education	OR	2.5 %	97.5 %
(No interaction)	1.9	1.3	2.6
Less than HS	0.9	0.4	1.9
High school	1.7	1.0	3.0
Some college	2.2	1.1	4.3
College degree	8.9	2.8	30.1

Including an interaction reveals that the change in trust over time
was not the same in each demographic group – trust changed very
little among those with less than a high school education and
dramatically for those with a college degree

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction

Probability estimates

Introduction
The proportional odds model
Results

Basic model
Comparison with multinomial regression
Interaction

Summary

- Ordinal responses are very common in the medical, epidemiological, and social sciences; I have been asked to analyze ordinal data on many occasions

- The proportional odds model is a rather elegant (and popular) way to handle ordinal data, respecting both its ordering as well as its categorical nature without any substantial increase in the difficulty of interpretation, as individual coefficients have odds ratio interpretations very similar to logistic regression

- Finally, there is also a rather interesting connection between the proportional odds model and nonparametric testing: with a single binary covariate, the (score) test of a proportional odds model is equivalent to the Wilcoxon rank sum test (see code for a demonstration)