Ordered and Multinomial Logit

Nick Waterbury

November 9, 2017

Motivation

Motivation

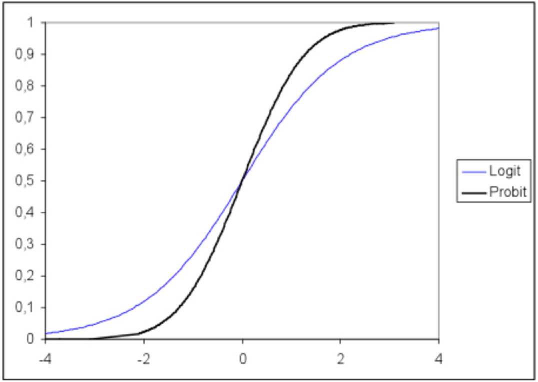
LIBRARY OF CONGRESS

Introduction

- For the standard logistic regression model there are two possible outcomes for the dependent variable.
  - Yes/No Votes
  - Democrat/Republican
- For many research questions, however, we are interested in discrete dependent variables that have more than two potential outcomes.
- Today, I will discuss two models that can handle these questions: the *ordinal* and the *multinomial* logistic regression.

Logit vs. Probit

- The difference in estimation is in the link functions (log vs. CDF normal).
- Regardless of the model used, results will be similar.



Ordered Logit

Ordinal Variables

- The dependent variable is an ordinal variable.
- The possible categories in the dependent variable have a natural order.
  - Agreement Scores- Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, Strongly Agree
  - Movie Ratings- 1 to 5 stars
  - Nearly any continuous variable (e.g. income)
- The measure is represented numerically, but these values are not cardinal numbers.

Proportional Odds Assumption

The model only applies to data that meets the proportional odds assumption.
Consider the following probabilities of the outcomes:

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
P1	*P2*	*P3*	*P4*	*P5*

Then the logarithm of the odds (not the probabilities) form an arithmetic series.

Proportional Odds Assumption

In more detail this means:

	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
Probability	$P1$	$P2$	$P3$	$P4$	$P5$
Formula	$\text{Log}(\frac{P1}{(P2+P3+P4+P5)})$	$\text{Log}(\frac{(P1+P2)}{(P3+P4+P5)})$	$\text{Log}(\frac{(P1+P2+P3)}{(P4+P5)})$	$\text{Log}(\frac{(P1+P2+P3+P4)}{P5})$	$\text{Log}(P1+P2+P3+P4+P5)$
Value	A	A+D	A+2D	A+3D	A+4D
Alternative Value	B-4D	B-3D	B-2D	B-D	B

Notice this means the ordered logit model assumes that the distance between each category of the outcome is proportional.

Proportional Odds Assumption

What to do when the assumption is violated?

1. Do nothing. Use ordered logistic regression because the practical implications of violating this assumption are debated in the literature.
2. Use a multinomial logit model. This frees you of the proportionality assumption, but it is less parsimonious and often dubious on substantive grounds.
3. Dichotomize the outcome and use binary logistic regression. This is a common solution, but you lose information and it could alter your substantive conclusions.
4. Use a model that does not assume proportionality.
    - Generalized Ordered Logit Models
    - Heterogeneous Choice Models

Estimating the Model

- Estimate the cumulative probability of being in one category versus all lower or higher categories.
- First recall the logistic regression for binary data:

$$\text{Log}(p/(1-p)) \;=\; \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

- Now the ordinal logistic regression:

$$\text{Log}(p_i/(1-p_i)) \;=\; \alpha_i + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

  where p_i is the probability of an outcome less than or equal to i, k is the total number of covariates, and α_i is the intercept for an outcome less than or equal to i.
- Each logit has its own α term but the same β coefficient. This means that the effect of the independent variables is the same for the different logit functions.

Interpreting the Results

- The coefficients of the ordinal logistic model cannot be interpreted like OLS coefficients.
- We are not modeling Y_i, but rather the probability of obtaining a value of Y_i.
- A one unit increase in a covariate will yield differing changes in the probability of some outcome Y_i depending on the values of the other covariates.
- Especially when there are many covariates, we can only produce predicted probabilities for specific, individual observations.

Interpreting the Results

- The most straightforward way to interpret results may be the odds ratio.
- The odds ratio can be found by exponentiating the coefficients of the ordered logit.
- When we view the change in levels in a cumulative sense and interpret the coefficients in odds, we are comparing the people who are in groups greater than i versus those who are in groups less than or equal to i (when i is the level of the outcome variable).
- The interpretation would be that for a one unit change in the predictor variable, the odds for an observation in a group that is greater than i versus less than or equal to i are the proportional odds times larger.
- This may make more sense in an example (coming very soon).

Potential Problems or Concerns

1. Heteroskedasticity
   - Non-constant error variance is much more concerning for logistic regression in comparison to OLS.
   - The standard errors are incorrect and (unlike OLS regression) the parameter estimates are biased.
   - Standardization is inherent in the model; unless the residual variability is identical across populations, the standardization of coefficients for each group will also differ.
2. Violation of the Proportional Odds Assumption
   - The assumption is often not satisfied and can alter the interpretations drawn from the model .
   - We can perform a Brant test to see if the assumption is being violated.

Example

- This example uses a dataset from Princeton's Data and Statistical Services webpage.
- There are 70 observations of individuals from 7 different countries, with a dependent variable of their opinion on a political assertion (4 point agreement scale).
- The variables of interest are $x1$, $x2$, and $x3$. They are all continuous variables.
- I will make use of the `MASS` package in `R`.

Example

```
# Getting sample data
library(foreign)
mydata <- read.dta("http://dss.princeton.edu/training/Panel101.dta")

# Loading library -MASS-
library(MASS)

# Running the ordered logit model
m1 <- polr(opinion ~ x1 + x2 + x3, data=mydata, Hess=TRUE)
```

Store results | Outcome | Predictors | Data source | Required for SE

```
summary(m1)
Call:
polr(formula = opinion ~ x1 + x2 + x3, data = mydata, Hess = TRUE)

Coefficients:
     Value Std. Error t value
x1 0.98140     0.5641  1.7397
x2 0.24936     0.2086  1.1954
x3 0.09089     0.1549  0.5867

Intercepts:
                Value   Std. Error t value
Str agree|Agree -0.2054  0.4682    -0.4388
Agree|Disag      0.7370  0.4697     1.5690
Disag|Str disag  1.9951  0.5204     3.8335

Residual Deviance: 189.6382
AIC: 201.6382
```

Example

```
# Getting coefficients and p-values

m1.coef <- data.frame(coef(summary(m1)))

m1.coef$pval = round((pnorm(abs(m1.coef$t.value), lower.tail = FALSE) * 2),2)

m1.coef

                      Value Std..Error    t.value pval
x1               0.98139603  0.5641136  1.7397134 0.08
x2               0.24935530  0.2086027  1.1953599 0.23
x3               0.09089175  0.1549254  0.5866807 0.56
Str agree|Agree -0.20542664  0.4682027 -0.4387558 0.66
Agree|Disag      0.73696754  0.4696907  1.5690486 0.12
Disag|Str disag  1.99507902  0.5204282  3.8335334 0.00
```

Example

- However, for ease in interpretation we should look at the odds ratios and predicted probabilities.
- The odds ratio is the exponentiated value of the coefficients.
- Keeping all other variables constant, when an observation's value of $x1$ increases one unit, it is 2.668 times more likely to be in a higher category. The odds of moving to a higher category in the outcome variable are very high when $x1$ is increased by one unit.

```
m1.or=exp(coef(m1))
m1.or
      x1       x2       x3
2.668179 1.283198 1.095150
```

Example

- Remember, it is only possible to interpret predicted probabilities at specific values.
- Below we compare values of $x3$ at 1 and 2 while keeping all other covariates at their means.
- Both observations have the highest probability of being in the *Disagree* category of the DV.

```
setup1
         x1        x2 x3
1 0.6480006 0.1338694  1
2 0.6480006 0.1338694  2
```

```
# Use "class" for the predicted category
setup1[, c("pred.prob")] <- predict(m1, newdata=setup1, type="class")
setup1

         x1        x2 x3 pred.prob
1 0.6480006 0.1338694  1     Disag
2 0.6480006 0.1338694  2     Disag
```

Example

- Lastly, a Brant test reveals this model does not violate the Proportional Odds Assumption.
- A violation would be denoted by a statistically significant result.

```
library(brant)
brm1<-brant(m1)
```

```
--------------------------------------------
Test for        X2        df        probability
--------------------------------------------
Omnibus                -1.19        6          1
x1               2.07        2        0.36
x2               0.54        2        0.76
x3               -17.9        2         1
--------------------------------------------

H0: Parallel Regression Assumption holds
```

Multinomial Logit

Nominal Variables

- The dependent variable is a nominal variable with more than two possible categories.
- The possible categories in the dependent variable are mutually exclusive, but have no natural order.
  - Blood Type
  - Parties in the UK House of Commons
  - Level of Legal Change in a SCOTUS decision (Wahlbeck 1997)
- Some sort of order can be arbitrarily created, but most descriptive calculations for this variable are meaningless.

Independence of Irrelevant Alternatives Assumption

- Independence of Irrelevant Alternatives states the odds of preferring one class over another do not depend on the presence or absence of other unconsidered alternatives.
- For example, the relative probabilities of taking a car or bus to work do not change if a bicycle is added as an additional possibility.
- This allows the choice of k alternatives to be modeled as a set of $k - 1$ independent binary choices, in which one alternative is chosen as a base and the other $k - 1$ compared against it, one at a time.
- Consider the relative probabilities of taking a car or a red bus or a blue bus.
- Violations of this assumption can lead to incorrect interpretations of results.
- A nested mutinomial logit is a potential remedy.

Estimating the Model

- Estimate a series of binary logit models.
- One group is chosen to be the base (reference) category for the other groups.
- For example if we want to compare never voters, former voters, and current voters, we make never voters the base category and two models are estimated:
  1. Current voters vs. Never voters
  2. Former voters vs. Never voters

Estimating the Model

- We are interested in the likelihood that an actor will choose a particular alternative, compared to an alternative that serves as a referent.
- According to Greene (1990), the functional form of the multinomial logit is:

$$P(Y = j) = \frac{\exp(\beta_j x_i)}{1 + \sum_{k=1}^{m} \exp(\beta_k x_i)} \qquad \text{for } j = 1, 2, 3, ..., J.$$

$$P(Y = 0) = \frac{1}{1 + \sum_{k=1}^{m} \exp(\beta_k x_i)}$$

- In these equations j indicates a specific choice in the set of all $J + 1$ alternatives, βx is the matrix of estimated coefficients and independent variables, k is one of the J choices, and m equals the number of J alternatives.

Interpreting the Results

- Again we cannot think of the coefficients produced in our model as similar to OLS coefficients.
- The results are estimates of our predicted variables' affect on the category of dependent variable as compared to the base category.
- Coefficients are valuable for direction, but not for magnitude.
- As with the ordered logit, predicted probabilities can only be performed for specific observations.
- The odds ratio allows for an easier interpretation of results.

Odds Ratio

- Separate odds ratios are determined for all independent variables for each category of the dependent variable with the exception of the reference category, which is omitted from the analysis.
- The $\exp(\beta)$ coefficient represents the change in the odds of the dependent variable being in a particular category vis-a-vis the reference category.
- This change in odds is associated with a one unit change of the corresponding independent variable.

Example

- This example uses a dataset from Wahlbeck's "The Life of the Law: Judicial Politics and Legal Change" (1997)
- There are 147 observations with 10 covariates that attempt to capture the facts of a case.
- Wahlbeck is interested in the probability that the Supreme Court will produce restrictive or expansive change, vis-a-vis retaining the status quo or effecting no legal change.
- Consequently, multinomial logit generates two sets of coefficients, one for restrictive change and another for expansive change.
- I will make use of the `nnet` package in `R`.

Example

- The code below show the tabulation of the DV.
  - 0 = the reference category; retaining the status quo
  - 1 = implementing restrictive legal change
  - 2 = implementing expansive legal change
- By default the reference is the first category.

```
library(foreign)
library(nnet)
library(stargazer)
wahlbeck<-read.dta("LegalChangeJOP.dta")
table(wahlbeck$trandir)

 0  1  2
83 39 26
```

Example

```
attach(wahlbeck)
multi1 <- multinom(trandir~defend+govt+stunion+
                govtatt2+defatty2+govtamic+defamic+law+congcoal
                +compmed2)
```

MULTINOMIAL LOGIT COEFFICIENTS FOR THE DIRECTION OF LEGAL CHANGE

Variable	Restrictive Change		Expansive Change	
Constant	3.128*	(1.838)	2.314	(1.947)
Policy Views				
Supreme Court Values	−2.501*	(1.102)	2.804**	(1.142)
Legal Constraint				
Number of Consecutive Decisions	−0.543**	(0.219)	0.075	(0.157)
Litigation Factors				
Searched Party Status	−0.564	(0.370)	0.100	(0.342)
Government Party Status	−0.800	(0.434)	−1.050*	(0.503)
Searched Attorney Experience	0.009	(0.010)	−0.002	(0.027)
Government Attorney Experience	0.001	(0.002)	−0.028*	(0.016)
Searched Amicus Support	0.933	(0.347)	0.741*	(0.377)
Government Amicus Support	−0.417	(0.221)	−1.179**	(0.497)
Political Factors				
President	13.733*	(7.660)	−7.063	(9.919)
Congress	−0.014	(0.018)	−0.013	(0.021)
Number of cases	147			
-2 X log-likelihood	208.665***			
Pseudo-R^2	27.5			
Correctly categorized	61.2%			
Proportional reduction of error	33.7%			

Note: Numbers in parentheses are the standard errors.
* Significant at ≤ .05 one-tailed test
** Significant at ≤ .01 one-tailed test
*** Significant at ≤ .001 one-tailed test

Example

- For ease of interpretation we should consider the odds ratios.
- Keeping all other variables constant, if the government's status increases one unit (it is binary), SCOTUS is 0.45 times more likely to produce restrictive legal change as opposed to retaining the status quo.
- Keeping all other variables constant, if the government's status increases one unit, SCOTUS is 0.35 times more likely to produce expansive legal change as opposed to retaining the status quo.

```
multi1_or<-exp(coef(multi1))
multi1_or
```

```
  (Intercept)    defend      govt      stunion  govtatt2  defatty2  govtamic
1    22.69573 0.5691274 0.4498717 9.328502e+05 1.0014312 1.0086717 0.6594596
2    10.08660 1.1055438 0.3505535 9.163304e-04 0.9727016 0.9973696 0.3074418
   defamic       law  congcoal    compmed2
1 2.541981 0.5814676 0.9865138  0.08201073
2 2.094579 1.0783170 0.9875301 16.46882446
```

Example

```
allmean <- data.frame(defend=rep(mean(defend, na.rm = TRUE),3),
                      stunion=rep(mean(stunion, na.rm = TRUE),3),
                      govtatt2=rep(mean(govtatt2, na.rm = TRUE)
                                   ,3),
                      defatty2=rep(mean(defatty2, na.rm = TRUE),3),
                      govtamic=rep(mean(govtamic, na.rm = TRUE),3),
                      defamic=rep(mean(defamic, na.rm = TRUE),3),
                      law=rep(mean(law, na.rm = TRUE),3),
                      congcoal=rep(mean(congcoal, na.rm = TRUE),3),
                      compmed2=rep(mean(compmed2, na.rm = TRUE),3),
                      govt = c(1, 2, 3))
allmean[, c("pred.prob")] <- predict(multi1, newdata=allmean,
                                     type="class")
```

```
    defend    stunion govtatt2 defatty2 govtamic   defamic      law congcoal
1 1.993827 0.03414815 75.66013 34.94771 0.808642 0.5740741 2.272109 70.49383
2 1.993827 0.03414815 75.66013 34.94771 0.808642 0.5740741 2.272109 70.49383
3 1.993827 0.03414815 75.66013 34.94771 0.808642 0.5740741 2.272109 70.49383
    compmed2 govt pred.prob
1 0.00621118    1         1
2 0.00621118    2         0
3 0.00621118    3         0
```

Example

- There are limited diagnostic tests available to test the IIA asumption.
- The Hausman-McFadden Test estimates the multinomial logit twice. First on the full set of possible alternatives and then on a subset of the possible alternatives.
- If IIA holds, the two sets of estimates should not be statistically distinguishable.
- The command in `R` is `hmftest()` in the `mnlogit` library.
- However, some have questioned the reliability of the test and suggest a theory driven approach.

Thank You!