PoliSci 4782 Political Analysis II

Categorical Outcome Models

Seamus Wagner

The Ohio State University

Roadmap

When having categorical outcomes, ask yourself two questions:

- Is the number of categories just equal to 2?
  - If so, a binary outcome model will do (logit or probit).
  - If not, we will need a categorical outcome model (which is also based either logit or probit)
- Is there any intrinsic ordering between different categories (so that we can rank those categories in an order, i.e. low-income, middle-income, high-income)?
  - ordered outcome models (ordinal logit or ordinal probit)
  - unordered outcome models (multinomial logit or multinomial probit)

Ordinal Logit/Probit for Ordered Categorical Outcome

Ordinal Logit/Probit

- It extends from logit/probit.
- In theory, we can combine multiple adjacent categories (since they are ordered) together to form a spectrum.
- We can model the probability of moving from one category to the next (upper) category, given a changeable baseline level (which is linked to the regression intercept).

Ordinal Logit

- Consider an ordered categorical outcome Y that has k possible categories.
- For brevity, those categories are denoted as $a, b, c, \ldots k$
- We can set up the following logit models with changing cutpoint c_i:

$$
\begin{aligned}
Pr(y > a) &= logit^{-1}(X\beta - c_1) \\
Pr(y > b) &= logit^{-1}(X\beta - c_2) \\
Pr(y > c) &= logit^{-1}(X\beta - c_3) \\
&\ldots \\
Pr(y > j) &= logit^{-1}(X\beta - c_{k-1})
\end{aligned}
$$

Cutpoints

- The parameters c_k (*cutpoints*) get larger as we move to a higher category: $c_1 < c_2 < \ldots < c_{k-1}$.
- Cutpoints help locate the predicted categories by logit, but do not meddle in the cause-effect relationships between Y and $\boldsymbol{X}$.
- Thus, an ordinal logit is a logit with $k - 1$ intercepts indicating the cutpoints between categories.

Interpreting Ordinal Logit

$$Pr(Y_i \leq j) = \frac{\exp(\beta X_i - c_j)}{1 + \exp(\beta X_i - c_j)}$$

- β indicates how the corresponding variable affects the probability of changing from any given category to the next category, generally speaking (odds interpretation and divide-by-4 apply).
- The intercept c_j indicates the baseline probability that the categorical outcome is no greater than category $j + 1$ (while other explanatory variables are held at 0).
- If the outcome variable has k categories, we will have $k - 1$ intercepts.

Illustrative Example

`nes96` dataset from `faraway` package on U.S. 1996 National Election Study:

- `PID`: party identification, ordered from strong Democrat to strong Republican (outcome variable)
- `income`: income level
- `age`: age group
- `educ`: education level

Ordinal Logit in R

```r
library(MASS)
model1 <- polr(PID ~ age + as.numeric(educ) + as.numeric(income), nes96)
summary(model1)
```

33:1 lecture 9: categorical outcome

Console C:/Users/Jianzi_He/Desktop/Teaching_PS4782/Lectures/9. Models for categorical outcomes/

```
Call:
polr(formula = PID ~ age + as.numeric(educ) + as.numeric(income),
    data = nes96)

Coefficients:
                       Value Std. Error t value
age                 0.001135   0.003579   0.317
as.numeric(educ)    0.047513   0.038780   1.225
as.numeric(income)  0.057229   0.010329   5.541

Intercepts:
                Value   Std. Error t value
strDem|weakDem  -0.1507  0.2812     -0.5359
weakDem|indDem   0.8001  0.2812      2.8448
indDem|indind    1.2811  0.2836      4.5168
indind|indRep    1.4442  0.2846      5.0746
indRep|weakRep   1.8759  0.2873      6.5294
weakRep|strRep   2.7303  0.2942      9.2810

Residual Deviance: 3457.807
AIC: 3475.807
```

Interpretation

The **coefficient** of income ($\beta = 0.06$): the odds of moving from one category to the next (which applies to any two adjacent PIDs) increases by a factor of $\exp(0.06) = 1.06$ when income rises by one level.

Intercepts indicate the baseline probability for each category when independent variables are held at 0:

- `strDem|weakDem` (the first intercept) is -0.15, so the probability of being a strong is $invlogit(-0.15) = 0.46$.
- `weakDem|indDem` is 0.80, so the probability of being a strong Democrat **or** a weak Democrat is $invlogit(0.80) = 0.69$.
- ...

Interpretation

```r
library(MASS)
model1 <- polr(PID ~ age + as.numeric(educ) + as.numeric(income), nes96)
summary(model1)
```

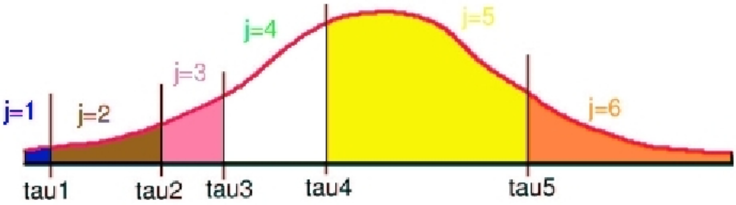
33:1 lecture 9: categorical outcome

Console C:/Users/Jianzi_He/Desktop/Teaching_PS4782/Lectures/9. Models for categorical outcomes/

```
Call:
polr(formula = PID ~ age + as.numeric(educ) + as.numeric(income),
    data = nes96)

Coefficients:
                       Value Std. Error t value
age                 0.001135   0.003579   0.317
as.numeric(educ)    0.047513   0.038780   1.225
as.numeric(income)  0.057229   0.010329   5.541

Intercepts:
                 Value   Std. Error t value
strDem|weakDem  -0.1507   0.2812    -0.5359
weakDem|indDem   0.8001   0.2812     2.8448
indDem|indind    1.2811   0.2836     4.5168
indind|indRep    1.4442   0.2846     5.0746
indRep|weakRep   1.8759   0.2873     6.5294
weakRep|strRep   2.7303   0.2942     9.2810

Residual Deviance: 3457.807
AIC: 3475.807
```

- What is the baseline probability of being a weak Democrat?
- Answer: $invlogit(0.80) - invlogit(-0.15) = 0.23$

Ordinal Probit



- Intuition: imagine you throw a ball from your own yard ($j = 1$) forward and it eventually falls into one of k possible yard divided by $k - 1$ fences ("cutpoints").
- The probit function $\mathbf{\Phi}(X\beta)$ determines the (latent) distance which we don't care; cutpoints (τ) help translate the latent value into an observed category J.
- The same coefficient interpretation approach as that for probit.

Multinomial Logit/Probit for Unordered Categorical Outcome

Unordered Categorical/Nominal Outcomes

- no superiority and inferiority between categories
- no meaningful distance between categories
- distinction between ordered and unordered categorical variable can be tricky and theory-dependent

Multiple Equation Models

- The new challenge in multinomial regression is that we are unable to put categories to a spectrum (since they are not ordered).
- The solution is to use multiple equations instead of one equation and to focus on one single category each time.
- This approach of using a system of equations is called "multiple equation model," which has other important applications in issues such as exploring complex causation and dealing with missing data.

Multinomial Logit/Probit

- It is an extension of logit/probit.
- In multinomial logit, we pick one category as the reference group and do a logit on the the probability of observations changing from that category to each of the remaining categories one at a time.
- In multinomial probit, we also pick one category, use standard normal distribution to compare the utility value of that category to each of the remaining categories one at a time, and eventually model the probability that one prefers a given category over the reference category.
- Both multinomial logit and probit yield $k-1$ sets of intercepts and coefficients, given the outcome has k categories.

Multinomial logit model

- We apply logistical regression to the odds ratio of being in one category vis-a-vis the reference category:

$$\eta_{ij} = X_i \beta_j = \log \frac{p_{ij}}{p_{i1}}, \qquad j = 2, \ldots, J$$

  Note: we have individual index i (for our units) and category index j

- The inverse logit function serves as the mean function by which we can compute the predicted probability of moving from the reference category to the corresponding one:

$$p_{ij} = \frac{\exp(X_{ij}\beta_{ij})}{1 + \sum_{j=2}^{J} \exp(X_{ij}\beta_{ij})}$$

Illustrative Example

We continue to use `nes96` dataset for demo and treat `PID` as an unordered categorical variable (this is a judgment call)

Multinomial Logit in R

```r
library(nnet)

model3 <- multinom(PID ~ as.numeric(age) + as.numeric(educ) + as.numeric(income), nes96)

summary(model3)

```

73:1 (Untitled) R Script

Console C:/Users/Jianzi_He/Desktop/Teaching_PS4782/Lectures/9. Models for categorical outcomes/

```
converged
> summary(model3)
Call:
multinom(formula = PID ~ as.numeric(age) + as.numeric(educ) +
    as.numeric(income), data = nes96)

Coefficients:
        (Intercept) as.numeric(age) as.numeric(educ) as.numeric(income)
weakDem   0.7021256   -2.209823e-02       0.04643949        0.001542891
indDem   -1.0150463   -1.918221e-02       0.12621262        0.045812593
indind   -1.6613264   -9.281564e-03      -0.10361522        0.055683638
indRep   -2.0857629    1.825893e-06       0.03347729        0.073664606
weakRep  -1.1482475   -9.190595e-03       0.03984032        0.070180544
strRep   -1.9303261   -1.849700e-03       0.10130247        0.087050320

Std. Errors:
        (Intercept) as.numeric(age) as.numeric(educ) as.numeric(income)
weakDem   0.4817966     0.006315216       0.07119764         0.01722851
indDem    0.5890586     0.007658225       0.08266329         0.02176761
indind    0.8931024     0.011048023       0.12428004         0.03316717
indRep    0.6513791     0.007700525       0.08538329         0.02422957
weakRep   0.5478319     0.006727317       0.07436736         0.02039010
strRep    0.5456234     0.006460797       0.07145470         0.02036117

Residual Deviance: 3428.426
AIC: 3476.426
```

Interpreting Intercepts

- Estimated intercepts indicate the baseline probabilities that individuals switch from the reference category to the corresponding category, while other explanatory variables are held at 0.
- "0" intercept indicates the baseline probability of being in the reference category, while other explanatory variables are held at 0.

```
> intercept <- c(0, coef(model3)[ ,1])
> intercept
              weakDem     indDem     indind     indRep    weakRep     strRep
 0.0000000  0.7021256 -1.0150463 -1.6613264 -2.0857629 -1.1482475 -1.9303261
> exp(intercept)/sum(exp(intercept))
              weakDem     indDem     indind     indRep    weakRep     strRep
0.24056876 0.48547684 0.08717867 0.04568087 0.02988160 0.07630652 0.03490674
>
```

Interpreting Coefficients

- Almost the same as to logit and ordinal logit (odds interpretation and divide-by-4 apply).
- Coefficients represent the log-odds of moving from the baseline category to a given category given a unit change in the corresponding explanatory variable.
- Exponentiated coefficients thus tell us how odds changes in the multiplicative way if the corresponding explanatory variable increases by one unit.

Independence of Irrelevant Alternatives (IIA)

- Theoretically, multinomial logit assumes IIA, which means that our preference between categories A and B is independent to any other category (the red-bus/blue-bus problem).
- This is often time an unrealistic assumption.
  - think about the fact that adding a third candidate often splits the votes
- If IIA is clearly violated, we need multinomial probit, which is theoretically less demanding but computationally more challenging

Multinomial Probit

- It can be run in `R` with the help of `mlogit` package or `MNP` package.
- It may take a long time to run and end up with unstable, weakly identified results, due to technical challenges.
- Coefficient interpretation is almost the same as to probit and ordinal probit (change in the z-score for the probability of changing from the reference category to the corresponding category).

Coming Up

- No lab or lab assignment this week (instructional break)
- Lecture and lab on duration outcomes next week