Multinomial Choice

Tyler Ransom

Duke University

June 15, 2012

Multinomial Choice

- In Applied Micro, a common and straightforward way to model an agent's behavior is through a discrete choice model
- Agent faces J mutually-exclusive choices
  - E.g. get to work via bicycle, bus, car, or other
- We want to see how the agent's dependent variable (taking on values $1, 2, \ldots, J$) varies with some covariates (X)

Binary logit

- Recall the choice probabilities from the binary logit model, where $y = X\beta + \varepsilon$ and $y \in \{0, 1\}$:
  $P_1 = \frac{\exp(X\beta)}{1+\exp(X\beta)}$
  $P_0 = 1 - P_1 = \frac{1}{1+\exp(X\beta)}$
- β_0 is normalized to 0 in this model to identify β

Multinomial Logit

- Given $y = X\beta + \varepsilon$, the probability that choice j is selected is $P_j = \frac{\exp(X\beta_j)}{\sum_k \exp(X\beta_k)}$
- Like in the binary case, normalize one of the β_j to 0 (for identification)
- Then the choice probability formula becomes $P_j = \frac{\exp(X\beta_j)}{1+\sum_h \exp(X\beta_h)}$
- Multinomial logit just adds more terms to the denominator and has J separate probability formulas (instead of two)

Applications of Multinomial Choice

IO

- Consumer decides which brand of a product to purchase (Demand estimation)
- Firm decides whether or not to enter a market (entry game)

Labor

- Individual determines whether to attend college, enter the workforce, or exit the labor force
- Individual chooses which occupation (or industry) to work in

Health

- Worker chooses a health insurance plan from a menu
- Patient decides which hospital to attend for treatments

Urban/Environmental

- Individual chooses which city (or neighborhood within a city) to live in
- Parents choose which school to enroll children in

Random Utility Models

- Useful to think of a multinomial choice model as a utility model
- Each agent gets utility from each of the choices, but the researcher only observes the choice of the agent (not the utility)
- Researcher assumes that utility-maximizing choice is the one observed
- This class of models is referred to as random utility models ("random" because the choice has a random aspect to it—ε—and "utility" because utility is being modeled)

Identification in Random Utility Models

- Normalizations:
- Scale: In the random utility model, one can multiply each of the alternatives by a constant without affecting which alternative is the max
- Level: Similarly, one can add a positive constant to all alternatives without affecting the arg max
- In practice, scale is set by normalizing the variance of ε_{ij} and level is set by normalizing one of the β_j's to be zero

Distribution of Error Term

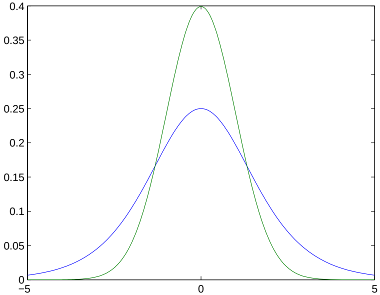
- Our random utility model is $u_{ij} = x_i\beta_j + \varepsilon_{ij}$ where i indexes individuals and j indexes choices
- For a multinomial logit, assume ε_{ij} is iid standard Type I Extreme Value $[F(x) = e^{-e^{-x}}]$
- Normalize the variance of ε_{ij} to $\pi^2/6$
- The familiar logistic distribution is nothing more than a difference in two T1EV variables with the appropriate normalizations already imposed
- This yields the choice probability formulas described earlier
- Standard T1EV (Gumbel) distribution looks similar to a normal, but with fatter tails

Distribution of Error Term

- If errors are normal, assume $\varepsilon_{ij} \sim N(0, \Omega)$
- Where Ω is some covariance matrix
- Note: can't estimate all covariance matrix elements, because scale needs to be set
- No one way to normalize this, but most common way is to set one of the variances to 1

Logit vs. Probit

Figure: Logit vs. Probit PDF



Note: Logit in blue; Normal in green

Independence of Irrelevant Alternatives (IIA)

- The major drawback of the logit model is its vulnerability to the IIA assumption, which imposes too many restrictions on substitution patterns (see next slide)
- Mathematically, this is manifest by comparing the ratio of two P_{ij}'s:
- $P_{i1} = \exp(x_i\beta_1)/(\text{denominator})$
- $P_{i3} = \exp(x_i\beta_3)/(\text{denominator})$
- $\frac{P_{i1}}{P_{i3}} = \exp(x_i\beta_1)/\exp(x_i\beta_3) = \exp(x_i(\beta_1 - \beta_3))$
- So the ratio of the probabilities is the same regardless of how many other choices there are

Red Bus / Blue Bus Example

- McFadden (1974) provides a famous example of the drawbacks to the IIA assumption
- Suppose a commuter is initially faced with two options: red bus and car, and that each choice probability is 1/2
- Now suppose a new choice is introduced: blue bus
- The IIA assumption requires that the new choice probabilities are 1/3 for each choice
- But given that commuters don't care what color the bus is, the actual change in probabilities is unlikely to be so drastic

Getting around IIA

- Nested Logit
  - Allows correlation among alternatives within a "nest" of choices
- Other GEV forms
  - Assume ε_{ij} is distributed Generalized Extreme Value (instead of iid Type 1 EV)
- Note: GEV forms still have closed-form choice probabilities, which is a huge advantage
- You'll cover these topics in more detail in Peter Arcidiacono's 2nd-year class

Choice of Error: Logit vs. Probit

- Logit Pros:
  - Closed-form choice probabilities
- Logit Cons:
  - Have to have iid assumption
  - This leads to the IIA problem
- Probit Pros:
  - Flexibly accommodates covariance among choice errors, so no issues with IIA
- Probit Cons:
  - Choice probabilities don't have closed form
  - Requires computation (and simulation) of a $J - 1$ dimensional integral

Logit Drawbacks

- Three main drawbacks to the logit
  - It can't handle individual-specific betas
  - Restrictive substitution patterns (IIA)
  - Can't handle temporally-correlated ε_{ij}'s in panel data very well
- Normal can get around all three of these, but at significant computational cost
- Nested Logit and other GEV models can only correct for the IIA assumption
- At the end of the day, most researchers still use logit because its computational ease is worth more than its drawbacks