Political Methodology III: Model Based Inference

Justin Grimmer

Professor
Department of Political Science
Stanford University

April 23rd, 2019

Model Based Inference

1) Likelihood inference
2) Logit/Probit
3) Ordered Probit
4) Choice Models:
   - Multinomial Logit
     a) DGP
     b) IIA
     c) Optimization
     d) Quantities of Interest
     e) Interpretation
   - Multinomial Probit
     a) DGP
     b) No IIA, But No Likelihood
     c) Quantities of Interest
     d) Interpretation

Modeling Categorical Outcomes

Categorical outcome (a.k.a. discrete choice) variable:

$$Y_i \in \{1, 2, ..., J\}$$

Example: Multiparty elections

- Voters choose from more than two parties/candidates
- $1 =$ Bush; $2 =$ Clinton; $3 =$ Perot
- $1 =$ Trump; $2 =$ Cruz; $3 =$ Kasich; $4 =$ Rubio; $5 =$ Bush ...

Possible research questions:

- How are candidate characteristics associated with probability of voting? (e.g. past experience, campaign spending)
- What kind of voters tend to choose Perot instead of Bush or Clinton? (e.g. age, race, ideology)
- How would vote shares have changed if Perot had not run?

Goal: Model probability of choosing $Y_i = j$ as function of predictors

Modeling Categorical Outcomes

Categorical outcome (a.k.a. discrete choice) variable:

$$Y_i \ \in \ \{1, 2, ..., J\}$$

Example: Multiparty elections

- Voters choose from more than two parties/candidates
- $1 =$ Bush; $2 =$ Clinton; $3 =$ Perot
- $1 =$ Trump; $2 =$ Cruz; $3 =$ Kasich; $4 =$ Rubio; $5 =$ Bush ...

Possible research questions:

- How are candidate characteristics associated with probability of voting? (e.g. past experience, campaign spending)
- What kind of voters tend to choose Perot instead of Bush or Clinton? (e.g. age, race, ideology)
- How would vote shares have changed if Perot had not run?

Goal: Model probability of choosing $Y_i = j$ as function of predictors

Modeling Categorical Outcomes

Categorical outcome (a.k.a. discrete choice) variable:

$$Y_i \in \{1, 2, ..., J\}$$

Example: Multiparty elections

- Voters choose from more than two parties/candidates
- $1 =$ Bush; $2 =$ Clinton; $3 =$ Perot
- $1 =$ Trump; $2 =$ Cruz; $3 =$ Kasich; $4 =$ Rubio; $5 =$ Bush ...

Possible research questions:

- How are candidate characteristics associated with probability of voting? (e.g. past experience, campaign spending)
- What kind of voters tend to choose Perot instead of Bush or Clinton? (e.g. age, race, ideology)
- How would vote shares have changed if Perot had not run?

Goal: Model probability of choosing $Y_i = j$ as function of predictors

Modeling Categorical Outcomes

Categorical outcome (a.k.a. discrete choice) variable:

$$Y_i \in \{1, 2, ..., J\}$$

Example: Multiparty elections

- Voters choose from more than two parties/candidates
- $1 =$ Bush; $2 =$ Clinton; $3 =$ Perot
- $1 =$ Trump; $2 =$ Cruz; $3 =$ Kasich; $4 =$ Rubio; $5 =$ Bush ...

Possible research questions:

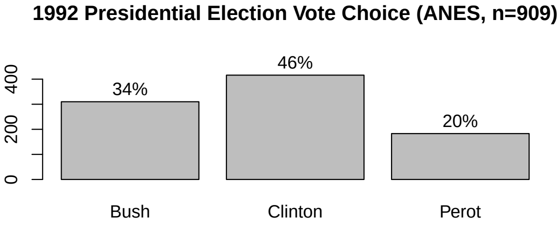
- How are candidate characteristics associated with probability of voting? (e.g. past experience, campaign spending)
- What kind of voters tend to choose Perot instead of Bush or Clinton? (e.g. age, race, ideology)
- How would vote shares have changed if Perot had not run?

Goal: Model probability of choosing $Y_i = j$ as function of predictors

Example: 1992 U.S. Presidential Election

Alvarez and Nagler (1995):

- Y_i: Vote choice in the 1992 U.S. presidential election ($1 =$ Clinton, $2 =$ Bush, $3 =$ Perot)

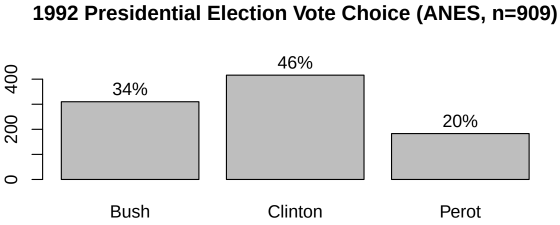


- Two types of predictors:
  - Voter-specific (V_i): age, gender, education, party, opinions, etc.
  - Candidate-varying (X_{ij}): ideological distance between voter i and candidate j

Example: 1992 U.S. Presidential Election

Alvarez and Nagler (1995):

- Y_i: Vote choice in the 1992 U.S. presidential election ($1 =$ Clinton, $2 =$ Bush, $3 =$ Perot)



- Two types of predictors:
  - Voter-specific (V_i): age, gender, education, party, opinions, etc.
  - Candidate-varying (X_{ij}): ideological distance between voter i and candidate j

Definition

Suppose we observe a trial, i, which might result in J outcomes.
And that P(outcome $= i) = \pi_i$
$Y_i = j$ if outcome j occurred and 0 otherwise.
Then Y_i follows a multinomial distribution, with

$$
\begin{aligned}
p(y_i) &= \pi_1^{I(y_i=1)}\pi_2^{I(y_i=2)}\ldots\pi_J^{I(y_i=J)} \\
p(y_i) &= \prod_{j=1}^{J}\pi_j^{I(y_j=j)}
\end{aligned}
$$

Equivalently, we'll write

$$
\begin{aligned}
Y_i &\sim \textit{Multinomial}(1, \boldsymbol{\pi}) \\
Y_i &\sim \textit{Categorical}(\boldsymbol{\pi})
\end{aligned}
$$

Multinomial Distribution

Multinomial Distribution

$$\mathsf{E}[I(Y_i = j)] = \pi_j$$

Multinomial Distribution

$$
\begin{aligned}
\mathsf{E}[I(Y_i = j)] &= \pi_j \\
\mathsf{E}[I(Y_i = J)] &= \pi_J = 1 - \pi_1 - \pi_2 - \ldots - \pi_{J-1}
\end{aligned}
$$

Multinomial Distribution

$$\begin{aligned}
\mathsf{E}[I(Y_i = j)] &= \pi_j \\
\mathsf{E}[I(Y_i = J)] &= \pi_J = 1 - \pi_1 - \pi_2 - \ldots - \pi_{J-1} \\
\mathsf{Var}[I(Y_i = j)] &= \pi_j(1 - \pi_j)
\end{aligned}$$

Multinomial Distribution

$$\begin{aligned}
\mathsf{E}[I(Y_i = j)] &= \pi_j \\
\mathsf{E}[I(Y_i = J)] &= \pi_J = 1 - \pi_1 - \pi_2 - \ldots - \pi_{J-1} \\
\mathsf{Var}[I(Y_i = j)] &= \pi_j(1 - \pi_j) \\
\mathsf{Cov}(I(Y_i = j), I(Y_i = k)] &= -\pi_j \pi_k
\end{aligned}$$

Vote Choice: Parameters of Multinomial Distribution

Suppose we make N independent draws:

$$Y_i \sim_{\text{iid}} \text{Multinomial}(1, \boldsymbol{\pi})$$

Vote Choice: Parameters of Multinomial Distribution

Suppose we make N independent draws:

$$Y_i \quad \sim_{\text{iid}} \quad \text{Multinomial}(1, \boldsymbol{\pi})$$

Then:

$$L(\boldsymbol{\pi}|\boldsymbol{Y}) \quad = \quad f(\boldsymbol{Y}|\boldsymbol{\pi})$$

Vote Choice: Parameters of Multinomial Distribution

Suppose we make N independent draws:

$$Y_i \quad \sim_{\text{iid}} \quad \text{Multinomial}(1, \boldsymbol{\pi})$$

Then:

$$\begin{aligned} L(\boldsymbol{\pi}|\boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{\pi}) \\ &= \prod_{i=1}^{N} f(Y_i|\boldsymbol{\pi}) \end{aligned}$$

Vote Choice: Parameters of Multinomial Distribution

Suppose we make N independent draws:

$$Y_i \quad \sim_{\text{iid}} \quad \text{Multinomial}(1, \boldsymbol{\pi})$$

Then:

$$\begin{aligned} L(\boldsymbol{\pi}|\boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{\pi}) \\ &= \prod_{i=1}^{N} f(Y_i|\boldsymbol{\pi}) \\ &= \prod_{i=1}^{N} \prod_{j=1}^{J} \pi_j^{I(Y_i=j)} \end{aligned}$$

Vote Choice

Obtaining maximum-likelihood estimates:

$$L(\boldsymbol{\pi}|\boldsymbol{Y}) = \prod_{i=1}^{N}\prod_{j=1}^{J} \pi_j^{I(Y_i=j)}$$

Vote Choice

Obtaining maximum-likelihood estimates:

$$
\begin{aligned}
L(\boldsymbol{\pi}|\boldsymbol{Y}) &= \prod_{i=1}^{N}\prod_{j=1}^{J} \pi_j^{I(Y_i=j)} \\
\ell(\boldsymbol{\pi}|\boldsymbol{Y}) &= \sum_{i=1}^{N}\sum_{j=1}^{J} I(Y_i=j)\log \pi_j
\end{aligned}
$$

Vote Choice

Obtaining maximum-likelihood estimates:

$$
\begin{aligned}
L(\boldsymbol{\pi}|\boldsymbol{Y}) &= \prod_{i=1}^{N}\prod_{j=1}^{J}\pi_j^{I(Y_i=j)} \\
\ell(\boldsymbol{\pi}|\boldsymbol{Y}) &= \sum_{i=1}^{N}\sum_{j=1}^{J} I(Y_i = j)\log \pi_j
\end{aligned}
$$

Include constraint that $\sum_{j=1}^{J}\pi_j = 1$

Vote Choice

Obtaining maximum-likelihood estimates:

$$
\begin{aligned}
L(\boldsymbol{\pi}|\boldsymbol{Y}) &= \prod_{i=1}^{N}\prod_{j=1}^{J}\pi_j^{I(Y_i=j)} \\
\ell(\boldsymbol{\pi}|\boldsymbol{Y}) &= \sum_{i=1}^{N}\sum_{j=1}^{J} I(Y_i = j)\log \pi_j
\end{aligned}
$$

Include constraint that $\sum_{j=1}^{J}\pi_j = 1$

$$
\ell(\boldsymbol{\pi}|\boldsymbol{Y}) = \sum_{i=1}^{N}\sum_{j=1}^{J} I(Y_i = j)\log \pi_j + \lambda(\sum_{j=1}^{J}\pi_j - 1)
$$

Vote Choice

$$
\begin{aligned}
\ell(\boldsymbol{\pi}|\boldsymbol{Y}) &= \sum_{i=1}^{N}\sum_{j=1}^{J} I(Y_i = j)\log \pi_j + \lambda(\sum_{j=1}^{J}\pi_j - 1) \\
\frac{\partial \ell(\boldsymbol{\pi}|\boldsymbol{Y})}{\partial \pi_1} &= \frac{\sum_{i=1}^{N} I(Y_i = 1)}{\pi_1} + \lambda \\
\frac{\partial \ell(\boldsymbol{\pi}|\boldsymbol{Y})}{\partial \pi_2} &= \frac{\sum_{i=1}^{N} I(Y_i = 2)}{\pi_2} + \lambda \\
\frac{\partial \ell(\boldsymbol{\pi}|\boldsymbol{Y})}{\partial \pi_3} &= \frac{\sum_{i=1}^{N} I(Y_i = 3)}{\pi_3} + \lambda \\
\vdots \quad &\vdots \quad \vdots \\
\frac{\partial \ell(\boldsymbol{\pi}|\boldsymbol{Y})}{\partial \pi_J} &= \frac{\sum_{i=1}^{N} I(Y_i = J)}{\pi_J} + \lambda \\
\frac{\partial \ell(\boldsymbol{\pi}|\boldsymbol{Y})}{\partial \lambda} &= \sum_{j=1}^{J}\pi_j - 1
\end{aligned}
$$

Set equal to zero
Solve for $\boldsymbol{\pi}^* = (\pi_1^*, \pi_2^*, \ldots, \pi_J^*)$

Vote Choice

Note that: $\sum_{i=1}^{N} \sum_{j=1}^{J} I(y_i = j) = N$

Vote Choice

Note that: $\sum_{i=1}^{N} \sum_{j=1}^{J} I(y_i = j) = N$

$$\pi_1^* = \frac{\sum_{i=1}^{N} I(Y_i = 1)}{N}$$

Vote Choice

Note that: $\sum_{i=1}^{N} \sum_{j=1}^{J} I(y_i = j) = N$

$$
\begin{aligned}
\pi_1^* &= \frac{\sum_{i=1}^{N} I(Y_i = 1)}{N} \\
\pi_2^* &= \frac{\sum_{i=1}^{N} I(Y_i = 2)}{N} \\
\pi_3^* &= \frac{\sum_{i=1}^{N} I(Y_i = 3)}{N} \\
\vdots &\ \vdots\ \ \vdots \\
\pi_J^* &= \frac{\sum_{i=1}^{N} I(Y_i = J)}{N}
\end{aligned}
$$

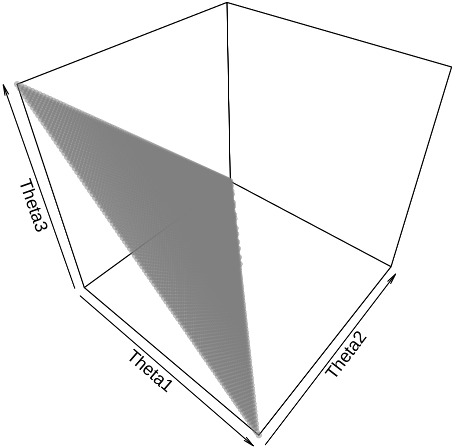
Vote Choice

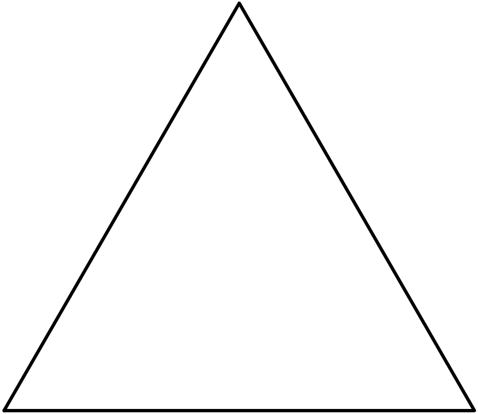
Note that: $\sum_{i=1}^{N} \sum_{j=1}^{J} I(y_i = j) = N$

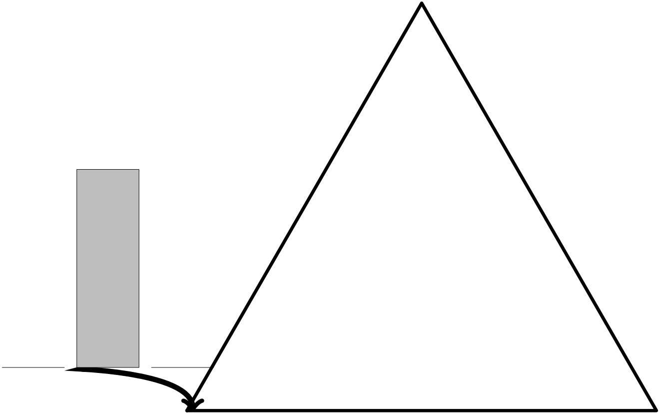
$$
\begin{aligned}
\pi_1^* &= \frac{\sum_{i=1}^{N} I(Y_i = 1)}{N} \\
\pi_2^* &= \frac{\sum_{i=1}^{N} I(Y_i = 2)}{N} \\
\pi_3^* &= \frac{\sum_{i=1}^{N} I(Y_i = 3)}{N} \\
\vdots & \quad \vdots \quad \vdots \\
\pi_J^* &= \frac{\sum_{i=1}^{N} I(Y_i = J)}{N}
\end{aligned}
$$

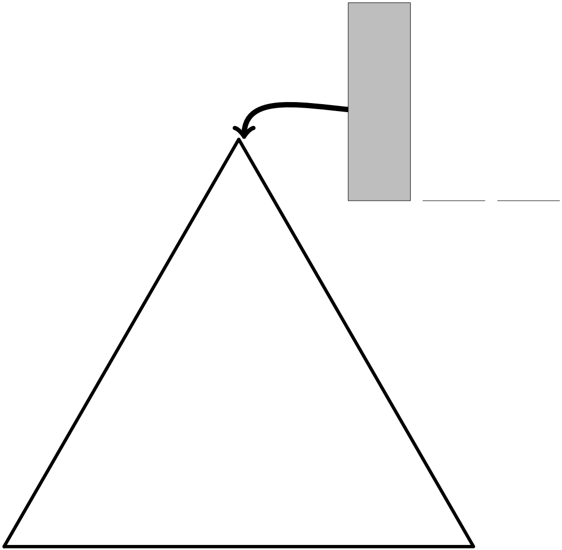
Maximum likelihood estimates$\rightsquigarrow$ Average proportion of time candidate is chosen

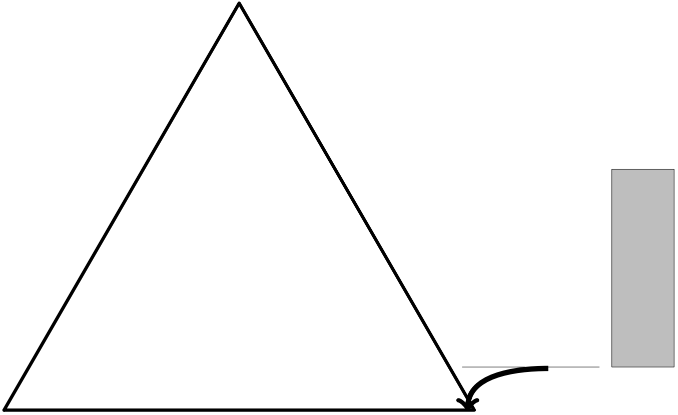
Three Candidate Choice

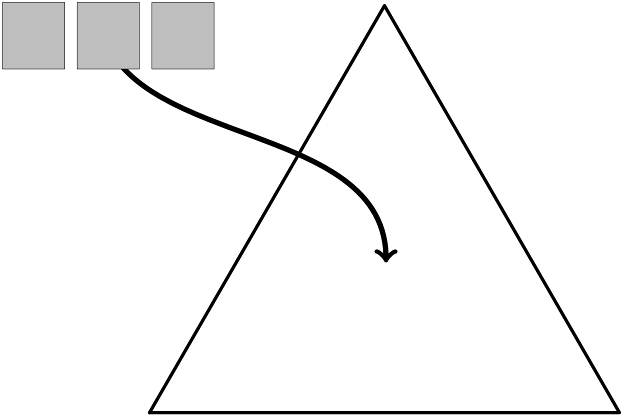

Theta3
Theta1
Theta2











Modeling Choices with Covariates

Multinomial Logit Model

- Generalize the logit model to more than two choices

- The multinomial logit model (MNL):

$$\pi_{ij} \;=\; \Pr(Y_i = j \mid V_i) \;=\; \frac{\exp(V_i' \delta_j)}{\sum_{k=1}^{J} \exp(V_i' \delta_k)},$$

  where V_i = individual-specific characteristics of unit i (and an intercept)

- Note that $\sum_{j=1}^{J} \pi_{ij} = 1$

- Need to set the base category for identifiability: $\delta_1 = 0$

- δ_j represents how characteristics of voter i is associated with probability of voting for candidate j

Multinomial Logit Model

- Generalize the logit model to more than two choices
- The multinomial logit model (MNL):

$$\pi_{ij} \;=\; \Pr(Y_i = j \mid V_i) \;=\; \frac{\exp(V_i'\delta_j)}{\sum_{k=1}^{J}\exp(V_i'\delta_k)},$$

where V_i = individual-specific characteristics of unit i (and an intercept)

- Note that $\sum_{j=1}^{J} \pi_{ij} = 1$
- Need to set the base category for identifiability: $\delta_1 = 0$
- δ_j represents how characteristics of voter i is associated with probability of voting for candidate j

Multinomial Logit Model

- Generalize the logit model to more than two choices
- The multinomial logit model (MNL):

$$\pi_{ij} \;=\; \Pr(Y_i = j \mid V_i) \;=\; \frac{\exp(V_i' \delta_j)}{\sum_{k=1}^{J} \exp(V_i' \delta_k)},$$

  where V_i = individual-specific characteristics of unit i (and an intercept)

- Note that $\sum_{j=1}^{J} \pi_{ij} = 1$
- Need to set the base category for identifiability: $\delta_1 = 0$
- δ_j represents how characteristics of voter i is associated with probability of voting for candidate j

Multinomial Logit Model

- Generalize the logit model to more than two choices
- The multinomial logit model (MNL):

$$\pi_{ij} \;=\; \Pr(Y_i = j \mid V_i) \;=\; \frac{\exp(V_i'\delta_j)}{\sum_{k=1}^{J}\exp(V_i'\delta_k)},$$

  where V_i = individual-specific characteristics of unit i (and an intercept)

- Note that $\sum_{j=1}^{J}\pi_{ij} = 1$
- Need to set the base category for identifiability: $\delta_1 = 0$
- δ_j represents how characteristics of voter i is associated with probability of voting for candidate j

Multinomial Logit Model

- Generalize the logit model to more than two choices
- The multinomial logit model (MNL):

$$\pi_{ij} \;=\; \Pr(Y_i = j \mid V_i) \;=\; \frac{\exp(V_i' \delta_j)}{\sum_{k=1}^{J} \exp(V_i' \delta_k)},$$

  where V_i = individual-specific characteristics of unit i (and an intercept)

- Note that $\sum_{j=1}^{J} \pi_{ij} = 1$
- Need to set the base category for identifiability: $\delta_1 = 0$
- δ_j represents how characteristics of voter i is associated with probability of voting for candidate j

Multinomial Logit Model: Confirming the Generalization

Suppose $J = 2$

Multinomial Logit Model: Confirming the Generalization

Suppose $J = 2$

$$
\begin{aligned}
\pi_{i1} &= \Pr(Y_i = 1 \mid V_i) \\
&= \frac{\exp(V_i^{'}\delta_1)}{\exp(V_i^{'}\delta_1) + \exp(V_i^{'}\delta_2)}
\end{aligned}
$$

Multinomial Logit Model: Confirming the Generalization

Suppose $J = 2$

$$\begin{aligned}\pi_{i1} &= \Pr(Y_i = 1 \mid V_i) \\ &= \frac{\exp(V_i^{'}\delta_1)}{\exp(V_i^{'}\delta_1) + \exp(V_i^{'}\delta_2)}\end{aligned}$$

Set $\delta_2 = 0$

Multinomial Logit Model: Confirming the Generalization

Suppose $J = 2$

$$\begin{aligned}
\pi_{i1} &= \Pr(Y_i = 1 \mid V_i) \\
&= \frac{\exp(V_i^{'}\delta_1)}{\exp(V_i^{'}\delta_1) + \exp(V_i^{'}\delta_2)}
\end{aligned}$$

Set $\delta_2 = 0$

$$\begin{aligned}
\pi_{i1} &= \frac{\exp(V_i^{'}\delta_1)}{\exp(V_i^{'}\delta_1) + 1} \\
&= \frac{1}{1 + \exp(-V_i^{'}\delta_1)}
\end{aligned}$$

Conditional Logit Model

- We can also incorporate alternative-varying predictors X_{ij}
- The conditional logit (CL) model:

$$\pi_{ij} = \Pr(Y_i = j \mid X_{ij}) = \frac{\exp(X_{ij}'\beta)}{\sum_{k=1}^{J} \exp(X_{ik}'\beta)}$$

- β represents how characteristics of candidate j for voter i are associated with voting probabilities
- X_{ij} does not have to vary across voters (e.g. whether candidate j is incumbent)
- In that case we suppress the subscript to X_j

Conditional Logit Model

- We can also incorporate alternative-varying predictors X_{ij}
- The conditional logit (CL) model:

$$\pi_{ij} \;=\; \Pr(Y_i = j \mid X_{ij}) \;=\; \frac{\exp(X_{ij}'\beta)}{\sum_{k=1}^{J}\exp(X_{ik}'\beta)}$$

- β represents how characteristics of candidate j for voter i are associated with voting probabilities
- X_{ij} does not have to vary across voters (e.g. whether candidate j is incumbent)
- In that case we suppress the subscript to X_j

Conditional Logit Model

- We can also incorporate alternative-varying predictors X_{ij}
- The conditional logit (CL) model:

$$\pi_{ij} = \Pr(Y_i = j \mid X_{ij}) = \frac{\exp(X_{ij}'\beta)}{\sum_{k=1}^{J} \exp(X_{ik}'\beta)}$$

- β represents how characteristics of candidate j for voter i are associated with voting probabilities
- X_{ij} does not have to vary across voters (e.g. whether candidate j is incumbent)
- In that case we suppress the subscript to X_j

Conditional Logit Model

- We can also incorporate alternative-varying predictors X_{ij}
- The conditional logit (CL) model:

$$\pi_{ij} \;=\; \Pr(Y_i = j \mid X_{ij}) \;=\; \frac{\exp(X_{ij}'\beta)}{\sum_{k=1}^{J} \exp(X_{ik}'\beta)}$$

- β represents how characteristics of candidate j for voter i are associated with voting probabilities
- X_{ij} does not have to vary across voters (e.g. whether candidate j is incumbent)
- In that case we suppress the subscript to X_j

MNL as a Special Case of CL

- MNL can be written using CL: Create a set of artificial alternative-varying regressors for each V_i:

$$X_{i1} = \begin{pmatrix} V_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad X_{i2} = \begin{pmatrix} 0 \\ V_i \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots, \quad X_{iJ} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ V_i \end{pmatrix}$$

- Set the element of β for X_{ij} to δ_j and you get the MNL model
- δ_1 must be set to zero for identifiability
- Thus we can write both models (and their mixture) simply as CL:

$$\pi_{ij} = \Pr(Y_i = j \mid X) = \frac{\exp(X'_{ij}\beta)}{\sum_{k=1}^{J} \exp(X'_{ik}\beta)}$$

- We use the names CL and MNL interchangeably from here on

MNL as a Special Case of CL

- MNL can be written using CL: Create a set of artificial alternative-varying regressors for each V_i:

$$X_{i1} = \begin{pmatrix} V_i \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad X_{i2} = \begin{pmatrix} 0 \\ V_i \\ \vdots \\ 0 \end{pmatrix}, \quad \cdots, \quad X_{iJ} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ V_i \end{pmatrix}$$

- Set the element of β for X_{ij} to δ_j and you get the MNL model
- δ_1 must be set to zero for identifiability

- Thus we can write both models (and their mixture) simply as CL:

$$\pi_{ij} = \Pr(Y_i = j \mid X) = \frac{\exp(X_{ij}'\beta)}{\sum_{k=1}^{J}\exp(X_{ik}'\beta)}$$

- We use the names CL and MNL interchangeably from here on

Predictor Types and Data Formats

Discrete choice data usually come in one of the two formats:

1. Wide format: N rows, $\#V + J \cdot \#X$ predictors

```
 choice women educ idist.Clinton idist.Bush idist.Perot
   Bush     1    3        4.0804     0.1024      0.2601
   Bush     1    4        4.0804     0.1024      0.2601
Clinton     1    2        1.0404     1.7424      0.2401
   Bush     0    6        0.0004     5.3824      2.2201
Clinton     1    3        0.9604    11.0220      6.2001 ...
```

2. Long format: NJ rows, $\#V + \#X$ predictors

```
chid     alt choice women educ  idist
   1    Bush   TRUE     1    3 0.1024
   1 Clinton  FALSE     1    3 4.0804
   1   Perot  FALSE     1    3 0.2601
   2    Bush   TRUE     1    4 0.1024
   2 Clinton  FALSE     1    4 4.0804
   2   Perot  FALSE     1    4 0.2601 ...
```

Predictor Types and Data Formats

- Use `reshape` to change between wide and long
- Some estimation functions (e.g. `mlogit`) can take both formats

Latent Variable Interpretation

- Recall the random utility model:

$$Y_{ij}^* = X_{ij}'\beta + \epsilon_{ij},$$

where $\begin{cases} Y_{ij}^* = \text{latent utility from choosing } j \text{ for } i \\ \epsilon_{ij} = \text{stochastic component of the utility} \end{cases}$

- Assume that voter chooses the most preferred candidate, i.e.,

$$Y_i = j \quad \text{if} \quad Y_{ij}^* \geq Y_{ij'}^* \quad \text{for any} \quad j' \in \{1, ..., J\}$$

- Assuming $\epsilon_{ij} \sim_{\text{iid}}$ type I extreme value distribution, this setup implies MNL (McFadden 1974)
- Proof for $J = 2$:

$$\begin{aligned} \Pr(Y_i = 1 \mid X) &= \Pr(Y_{i1}^* \geq Y_{i2}^* \mid X) \\ &= \Pr\left(\epsilon_{i2} - \epsilon_{i1} \leq (X_{i1} - X_{i2})'\beta\right) \\ &= \frac{\exp\left((X_{i1} - X_{i2})'\beta\right)}{1 + \exp\left((X_{i1} - X_{i2})'\beta\right)} = \frac{\exp(X_{i1}'\beta)}{\exp(X_{i1}'\beta) + \exp(X_{i2}'\beta)} \end{aligned}$$

Latent Variable Interpretation

- Recall the random utility model:

$$Y_{ij}^* = X_{ij}^{'}\beta + \epsilon_{ij},$$

where $\begin{cases} Y_{ij}^* & = & \text{latent utility from choosing } j \text{ for } i \\ \epsilon_{ij} & = & \text{stochastic component of the utility} \end{cases}$

- Assume that voter chooses the most preferred candidate, i.e.,

$$Y_i = j \quad \text{if} \quad Y_{ij}^* \geq Y_{ij'}^* \quad \text{for any} \quad j' \in \{1, ..., J\}$$

- Assuming $\epsilon_{ij} \sim_{\text{iid}}$ type I extreme value distribution, this setup implies MNL (McFadden 1974)
- Proof for $J = 2$:

$$\begin{aligned}
\Pr(Y_i = 1 \mid X) &= \Pr(Y_{i1}^* \geq Y_{i2}^* \mid X) \\
&= \Pr\left(\epsilon_{i2} - \epsilon_{i1} \leq (X_{i1} - X_{i2})^{'}\beta\right) \\
&= \frac{\exp\left((X_{i1} - X_{i2})^{'}\beta\right)}{1 + \exp\left((X_{i1} - X_{i2})^{'}\beta\right)} = \frac{\exp(X_{i1}^{'}\beta)}{\exp(X_{i1}^{'}\beta) + \exp(X_{i2}^{'}\beta)}
\end{aligned}$$

Latent Variable Interpretation

- Recall the random utility model:

$$Y_{ij}^* = X_{ij}'\beta + \epsilon_{ij},$$

where $\begin{cases} Y_{ij}^* & = \text{latent utility from choosing } j \text{ for } i \\ \epsilon_{ij} & = \text{stochastic component of the utility} \end{cases}$

- Assume that voter chooses the most preferred candidate, i.e.,

$$Y_i = j \quad \text{if} \quad Y_{ij}^* \geq Y_{ij'}^* \quad \text{for any} \quad j' \in \{1, ..., J\}$$

- Assuming $\epsilon_{ij} \sim_{\text{iid}}$ type I extreme value distribution, this setup implies MNL (McFadden 1974)

- Proof for $J = 2$:

$$\begin{aligned} \Pr(Y_i = 1 \mid X) &= \Pr(Y_{i1}^* \geq Y_{i2}^* \mid X) \\ &= \Pr\left(\epsilon_{i2} - \epsilon_{i1} \leq (X_{i1} - X_{i2})'\beta\right) \\ &= \frac{\exp\left((X_{i1} - X_{i2})'\beta\right)}{1 + \exp\left((X_{i1} - X_{i2})'\beta\right)} = \frac{\exp(X_{i1}'\beta)}{\exp(X_{i1}'\beta) + \exp(X_{i2}'\beta)} \end{aligned}$$

Latent Variable Interpretation

- Recall the random utility model:

$$Y_{ij}^* \;=\; X_{ij}^{'}\beta + \epsilon_{ij},$$

where $\begin{cases} Y_{ij}^* & = & \text{latent utility from choosing } j \text{ for } i \\ \epsilon_{ij} & = & \text{stochastic component of the utility} \end{cases}$

- Assume that voter chooses the most preferred candidate, i.e.,

$$Y_i = j \quad \text{if} \quad Y_{ij}^* \geq Y_{ij'}^* \quad \text{for any} \quad j' \in \{1, ..., J\}$$

- Assuming $\epsilon_{ij} \sim_{\text{iid}}$ type I extreme value distribution, this setup implies MNL (McFadden 1974)
- Proof for $J = 2$:

$$\begin{aligned} \Pr(Y_i = 1 \mid X) &= \Pr(Y_{i1}^* \geq Y_{i2}^* \mid X) \\ &= \Pr\left(\epsilon_{i2} - \epsilon_{i1} \leq (X_{i1} - X_{i2})^{'}\beta\right) \\ &= \frac{\exp\left((X_{i1} - X_{i2})^{'}\beta\right)}{1 + \exp\left((X_{i1} - X_{i2})^{'}\beta\right)} = \frac{\exp(X_{i1}^{'}\beta)}{\exp(X_{i1}^{'}\beta) + \exp(X_{i2}^{'}\beta)} \end{aligned}$$

Independence of Irrelevant Alternatives

$$\frac{\pi_{i1}}{\pi_{i2}}$$

Independence of Irrelevant Alternatives

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{\exp(X_{i1}'\beta)}{\sum_{k=1}^{J}\exp(X_{ik}\beta)}}{\frac{\exp(X_{i2}'\beta)}{\sum_{k=1}^{J}\exp(X_{ik}\beta)}}$$

Independence of Irrelevant Alternatives

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{\exp(X_{i1}'\beta)}{\sum_{k=1}^{J}\exp(X_{ik}\beta)}}{\frac{\exp(X_{i2}'\beta)}{\sum_{k=1}^{J}\exp(X_{ik}\beta)}} = \frac{\exp(X_{i1}'\beta)}{\exp(X_{i2}'\beta)}$$

Independence of Irrelevant Alternatives

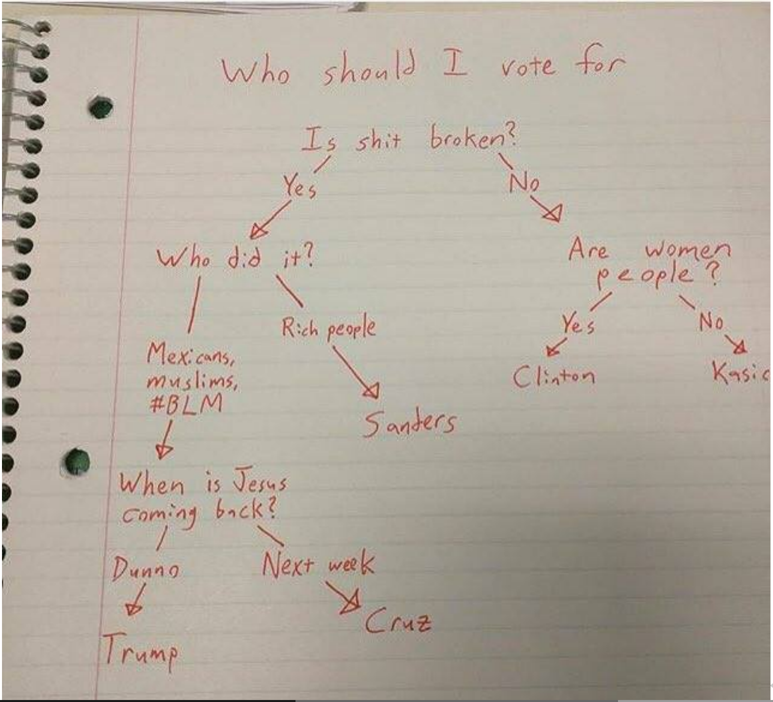
$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{\exp(X_{i1}^{'}\beta)}{\sum_{k=1}^{J}\exp(X_{ik}\beta)}}{\frac{\exp(X_{i2}^{'}\beta)}{\sum_{k=1}^{J}\exp(X_{ik}\beta)}} = \frac{\exp(X_{i1}^{'}\beta)}{\exp(X_{i2}^{'}\beta)}$$

Relative Risk: does not depend on other choices

Independence of Irrelevant Alternatives

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{\exp(X_{i1}^{'}\beta)}{\sum_{k=1}^{J}\exp(X_{ik}\beta)}}{\frac{\exp(X_{i2}^{'}\beta)}{\sum_{k=1}^{J}\exp(X_{ik}\beta)}} = \frac{\exp(X_{i1}^{'}\beta)}{\exp(X_{i2}^{'}\beta)}$$

Relative Risk: does not depend on other choices
Multinomial Choice: series of pairwise comparisons


Who should I vote for
Is shit broken?
Yes
No
Who did it?
Are women people?
Mexicans, muslims, #BLM
Rich people
Yes
No
Clinton
Sanders
When is Jesus coming back?
Dunno
Next week
Trump
Cruz

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

Trump, Cruz, and Sanders remaining candidates

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

Trump, Cruz, and Sanders remaining candidates (God, help us)

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

Trump, Cruz, and Sanders remaining candidates (God, help us)

- 50% support Sanders
- 25% support Trump
- 25% support Cruz

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

Trump, Cruz, and Sanders remaining candidates (God, help us)

- 50% support Sanders
- 25% support Trump
- 25% support Cruz

Cruz drops out, what happens to voters?

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

Trump, Cruz, and Sanders remaining candidates (God, help us)

- 50% support Sanders
- 25% support Trump
- 25% support Cruz

Cruz drops out, what happens to voters?

- Intuitively: 50% Sanders, 50% Trump

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

Trump, Cruz, and Sanders remaining candidates (God, help us)

- 50% support Sanders
- 25% support Trump
- 25% support Cruz

Cruz drops out, what happens to voters?

- Intuitively: 50% Sanders, 50% Trump
- Under model, Expected latent utilities are:

$$\begin{aligned}
\text{Sanders} &= \log 2 \\
\text{Trump} &= 0 \\
\text{Cruz} &= 0
\end{aligned}$$

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

Trump, Cruz, and Sanders remaining candidates (God, help us)

- 50% support Sanders
- 25% support Trump
- 25% support Cruz

Cruz drops out, what happens to voters?

- Intuitively: 50% Sanders, 50% Trump
- Under model, Expected latent utilities are:

$$\begin{aligned}
\text{Sanders} &= \log 2 \\
\text{Trump} &= 0 \\
\text{Cruz} &= 0 \\
\Pr(\text{Sanders}) &= \frac{\exp(\log 2)}{\exp(\log 2) + \exp(0) + \exp(0)} = 2/(2+1+1) = 0.5
\end{aligned}$$

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

But then:

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

But then:

$$\frac{\text{Pr(Sanders)}}{\text{Pr(Trump)}} = \frac{\exp(\log 2)}{\exp(0)} = 2/1$$

Independence of Irrelevant Alternatives (Red Bus, Blue Bus)

But then:

$$\frac{\mathsf{Pr(Sanders)}}{\mathsf{Pr(Trump)}} = \frac{\exp(\log 2)}{\exp(0)} = 2/1$$

Or the model predicts voters choose Sanders $2:1$ over Trump(!) resulting in (66%,33%) split

IIA and Your Work

What to do?

IIA and Your Work

What to do?

1) Alternative models: Multinomial Probit, Nested Logit

IIA and Your Work

What to do?

1) Alternative models: Multinomial Probit, Nested Logit
2) Conditional on included covariates (other information on individual voters would limit the problem)

IIA and Your Work

What to do?

1) Alternative models: Multinomial Probit, Nested Logit
2) Conditional on included covariates (other information on individual voters would limit the problem)
3) Can overstate size of problem: estimation error (standard errors) often much larger than deviation

Estimation and Inference

- Likelihood for a random sample of size N:

$$
\begin{aligned}
L(\boldsymbol{\beta} \mid Y, X) &= \prod_{i=1}^{N} \prod_{j=1}^{J} \pi_{ij}^{I(Y_i = j)} \\
&= \prod_{i=1}^{N} \prod_{j=1}^{J} \left(\frac{\exp(X_{ij}^{'} \boldsymbol{\beta})}{\sum_{k=1}^{J} \exp(X_{ik}^{'} \boldsymbol{\beta})} \right)^{I(Y_i = j)} \\
\log L(\boldsymbol{\beta} \mid Y, X) &= \sum_{i=1}^{N} \sum_{j=1}^{J} I(Y_i = j) \left(X_{ij}^{'} \boldsymbol{\beta} - \log \left(\sum_{k=1}^{J} \exp(X_{ik}^{'} \boldsymbol{\beta}) \right) \right)
\end{aligned}
$$

Estimation and Inference

- Score:

$$s(\boldsymbol{\beta}|Y_i, \boldsymbol{X}_i) \;=\; \sum_{j=1}^{J} I(Y_i = j)\left(X_{ij} - \bar{X}_i\right),$$

where $\bar{X}_i$ is the weighted average of X_i, i.e. $\bar{X}_i = \sum_{j=1}^{J} \pi_{ij} X_{ij}$

- Solve $\sum_{i=1}^{N} s(\boldsymbol{\beta}|Y_i, \boldsymbol{X}_i) = 0$ (numerically) to get $\boldsymbol{\beta}^*_{MLE}$
- It can be shown that the log-likelihood is globally concave $\Rightarrow$ guaranteed convergence to the true (not local) MLE
- Information:

$$I_N(\boldsymbol{\beta}) \;=\; -\mathsf{E}[H(\boldsymbol{\beta})] \;=\; \sum_{i=1}^{N} \sum_{j=1}^{J} \pi_{ij}\left(X_{ij} - \bar{X}_i\right)\left(X_{ij} - \bar{X}_i\right)'$$

Estimation and Inference

- Score:
$$s(\boldsymbol{\beta}|Y_i, \boldsymbol{X}_i) \;=\; \sum_{j=1}^{J} I(Y_i = j)\left(X_{ij} - \bar{X}_i\right),$$
where $\bar{X}_i$ is the weighted average of X_i, i.e. $\bar{X}_i = \sum_{j=1}^{J} \pi_{ij} X_{ij}$
- Solve $\sum_{i=1}^{N} s(\boldsymbol{\beta}|Y_i, \boldsymbol{X}_i) = 0$ (numerically) to get $\boldsymbol{\beta}^*_{MLE}$
- It can be shown that the log-likelihood is globally concave $\Rightarrow$ guaranteed convergence to the true (not local) MLE
- Information:

$$I_N(\boldsymbol{\beta}) \;=\; -\mathsf{E}[H(\boldsymbol{\beta})] \;=\; \sum_{i=1}^{N} \sum_{j=1}^{J} \pi_{ij}\left(X_{ij} - \bar{X}_i\right)\left(X_{ij} - \bar{X}_i\right)'$$

Estimation and Inference

- Score:

$$s(\boldsymbol{\beta}|Y_i, \boldsymbol{X}_i) \;=\; \sum_{j=1}^{J} I(Y_i = j)\left(X_{ij} - \bar{X}_i\right),$$

  where $\bar{X}_i$ is the weighted average of X_i, i.e. $\bar{X}_i = \sum_{j=1}^{J} \pi_{ij} X_{ij}$
- Solve $\sum_{i=1}^{N} s(\boldsymbol{\beta}|Y_i, \boldsymbol{X}_i) = 0$ (numerically) to get $\boldsymbol{\beta}^*_{MLE}$
- It can be shown that the log-likelihood is globally concave $\Rightarrow$ guaranteed convergence to the true (not local) MLE
- Information:

$$I_N(\boldsymbol{\beta}) \;=\; -\mathsf{E}[H(\boldsymbol{\beta})] \;=\; \sum_{i=1}^{N}\sum_{j=1}^{J} \pi_{ij}\left(X_{ij} - \bar{X}_i\right)\left(X_{ij} - \bar{X}_i\right)'$$

Estimation and Inference

$$\boldsymbol{\beta}^{*}_{MLE} \longrightarrow^{D} \text{Normal}\left(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1}\right)$$

Interpreting MNL/CL Coefficients

In MNL/CL, β itself is not necessarily informative about the effect of X

1. The coefficients are all with respect to the baseline category
   $\longrightarrow$ Testing $\beta_j = 0$ does not generally make sense
   (unless comparison to the baseline is the goal)

2. Changing X_{ij} has impact on $\Pr(Y_i = k \mid X)$, $k \neq j$:
   - For individual-specific characteristics (V_i), even sign of δ_j may not agree with the direction of the change in response probability for j
   - For alternative-varying characteristics (X_{ij}), sign of β does indicate the direction of the effect, but magnitude is hard to interpret

Compute a quantity that has a clear substantive interpretation!

Interpreting MNL/CL Coefficients

In MNL/CL, β itself is not necessarily informative about the effect of X

1. The coefficients are all with respect to the baseline category
   $\longrightarrow$ Testing $\beta_j = 0$ does not generally make sense
   (unless comparison to the baseline is the goal)
2. Changing X_{ij} has impact on $\Pr(Y_i = k \mid X)$, $k \neq j$:
   - For individual-specific characteristics (V_i), even sign of δ_j may not agree with the direction of the change in response probability for j
   - For alternative-varying characteristics (X_{ij}), sign of β does indicate the direction of the effect, but magnitude is hard to interpret

Compute a quantity that has a clear substantive interpretation!

Interpreting MNL/CL Coefficients

In MNL/CL, β itself is not necessarily informative about the effect of X

1. The coefficients are all with respect to the baseline category
   $\longrightarrow$ Testing $\beta_j = 0$ does not generally make sense
   (unless comparison to the baseline is the goal)
2. Changing X_{ij} has impact on $\Pr(Y_i = k \mid X)$, $k \neq j$:
   - For individual-specific characteristics (V_i), even sign of δ_j may not agree with the direction of the change in response probability for j
   - For alternative-varying characteristics (X_{ij}), sign of β does indicate the direction of the effect, but magnitude is hard to interpret

Compute a quantity that has a clear substantive interpretation!

Interpreting MNL/CL Coefficients

In MNL/CL, β itself is not necessarily informative about the effect of X

1. The coefficients are all with respect to the baseline category
   $\longrightarrow$ Testing $\beta_j = 0$ does not generally make sense
   (unless comparison to the baseline is the goal)
2. Changing X_{ij} has impact on $\Pr(Y_i = k \mid X)$, $k \neq j$:
   - For individual-specific characteristics (V_i), even sign of δ_j may not agree with the direction of the change in response probability for j
   - For alternative-varying characteristics (X_{ij}), sign of β does indicate the direction of the effect, but magnitude is hard to interpret

Compute a quantity that has a clear substantive interpretation!

Interpreting MNL/CL Coefficients

In MNL/CL, β itself is not necessarily informative about the effect of X

1. The coefficients are all with respect to the baseline category
   $\longrightarrow$ Testing $\beta_j = 0$ does not generally make sense
   (unless comparison to the baseline is the goal)

2. Changing X_{ij} has impact on $\Pr(Y_i = k \mid X)$, $k \neq j$:
   - For individual-specific characteristics (V_i), even sign of δ_j may not agree with the direction of the change in response probability for j
   - For alternative-varying characteristics (X_{ij}), sign of β does indicate the direction of the effect, but magnitude is hard to interpret

Compute a quantity that has a clear substantive interpretation!

Interpreting MNL/CL Coefficients

In MNL/CL, β itself is not necessarily informative about the effect of X

1. The coefficients are all with respect to the baseline category
   $\longrightarrow$ Testing $\beta_j = 0$ does not generally make sense
   (unless comparison to the baseline is the goal)
2. Changing X_{ij} has impact on $\Pr(Y_i = k \mid X)$, $k \neq j$:
   - For individual-specific characteristics (V_i), even sign of δ_j may not agree with the direction of the change in response probability for j
   - For alternative-varying characteristics (X_{ij}), sign of β does indicate the direction of the effect, but magnitude is hard to interpret

Compute a quantity that has a clear substantive interpretation!

Calculating Quantities of Interest

1. Choice probability:

$$\pi_j(x) \;=\; \Pr(Y_i = j \mid X = x)$$

e.g. How likely is a female college-educated conservative Republican voter to vote for Perot?

2. Predicted vote share:

$$p_j(x_1) \;\equiv\; \mathsf{E}\left[I\left(\pi_j(X_{i1} = x_1, X_{i2}) \geq \pi_k(X_{i1} = x_1, X_{i2}) \text{ for all } k\right)\right]$$

where X_{i1} is the predictor(s) of interest and X_{i2} is all other predictors

e.g. What would Perot's vote share be if all voters supported abortion?

3. Average partial (treatment) effects:

$$\tau_{jk} \;=\; \mathsf{E}\left[\pi_j(T_{ik} = 1, T_{i*}, W_i) - \pi_j(T_{ik} = 0, T_{i*}, W_i)\right]$$

where T_{ik} is treatment on candidate k, T_{i*} is treatment on others, W_i is pre-treatment covariates

- "Direct effect" if $j = k$; "indirect effect" if $j \neq k$
- If T is individual-specific, $\tau_j = \mathsf{E}[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)]$

- Estimate by plugging in sample analogues (e.g. $\pi_j \to \hat{\pi}_j$, $\mathsf{E} \to \frac{1}{n}\sum$)

Calculating Quantities of Interest

1. Choice probability:

$$\pi_j(x) \;=\; \Pr(Y_i = j \mid X = x)$$

e.g. How likely is a female college-educated conservative Republican voter to vote for Perot?

2. Predicted vote share:

$$p_j(x_1) \;\equiv\; \mathsf{E}\left[I\left(\pi_j(X_{i1} = x_1, X_{i2}) \geq \pi_k(X_{i1} = x_1, X_{i2}) \text{ for all } k\right)\right]$$

where X_{i1} is the predictor(s) of interest and X_{i2} is all other predictors

e.g. What would Perot's vote share be if all voters supported abortion?

3. Average partial (treatment) effects:

$$\tau_{jk} \;=\; \mathsf{E}\left[\pi_j(T_{ik} = 1, T_{i*}, W_i) - \pi_j(T_{ik} = 0, T_{i*}, W_i)\right]$$

where T_{ik} is treatment on candidate k, T_{i*} is treatment on others, W_i is pre-treatment covariates

- "Direct effect" if $j = k$; "indirect effect" if $j \neq k$
- If T is individual-specific, $\tau_j = \mathsf{E}[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)]$

- Estimate by plugging in sample analogues (e.g. $\pi_j \to \hat{\pi}_j$, $\mathsf{E} \to \frac{1}{n}\sum$)

Calculating Quantities of Interest

1. Choice probability:

$$\pi_j(x) \; = \; \Pr(Y_i = j \mid X = x)$$

e.g. How likely is a female college-educated conservative Republican voter to vote for Perot?

2. Predicted vote share:

$$p_j(x_1) \; \equiv \; \mathsf{E}\left[I\left(\pi_j(X_{i1} = x_1, X_{i2}) \geq \pi_k(X_{i1} = x_1, X_{i2}) \text{ for all } k\right)\right]$$

where X_{i1} is the predictor(s) of interest and X_{i2} is all other predictors
e.g. What would Perot's vote share be if all voters supported abortion?

3. Average partial (treatment) effects:

$$\tau_{jk} \; = \; \mathsf{E}\left[\pi_j(T_{ik} = 1, T_{i*}, W_i) - \pi_j(T_{ik} = 0, T_{i*}, W_i)\right]$$

where T_{ik} is treatment on candidate k, T_{i*} is treatment on others, W_i is pre-treatment covariates

- "Direct effect" if $j = k$; "indirect effect" if $j \neq k$
- If T is individual-specific, $\tau_j = \mathsf{E}[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)]$

- Estimate by plugging in sample analogues (e.g. $\pi_j \to \hat{\pi}_j$, $\mathsf{E} \to \frac{1}{n}\sum$)

Calculating Quantities of Interest

1. Choice probability:

$$\pi_j(x) \;=\; \Pr(Y_i = j \mid X = x)$$

e.g. How likely is a female college-educated conservative Republican voter to vote for Perot?

2. Predicted vote share:

$$p_j(x_1) \;\equiv\; \mathsf{E}\left[I\left(\pi_j(X_{i1} = x_1, X_{i2}) \geq \pi_k(X_{i1} = x_1, X_{i2}) \text{ for all } k\right)\right]$$

where X_{i1} is the predictor(s) of interest and X_{i2} is all other predictors

e.g. What would Perot's vote share be if all voters supported abortion?

3. Average partial (treatment) effects:

$$\tau_{jk} \;=\; \mathsf{E}\left[\pi_j(T_{ik} = 1, T_{i*}, W_i) - \pi_j(T_{ik} = 0, T_{i*}, W_i)\right]$$

where T_{ik} is treatment on candidate k, T_{i*} is treatment on others, W_i is pre-treatment covariates

- "Direct effect" if $j = k$; "indirect effect" if $j \neq k$
- If T is individual-specific, $\tau_j = \mathsf{E}[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)]$

- Estimate by plugging in sample analogues (e.g. $\pi_j \to \hat{\pi}_j$, $\mathsf{E} \to \frac{1}{n}\sum$)

Example: 1992 U.S. Presidential Election

- Model specification (Alvarez and Nagler 1995):

$$\pi_{ij} \;=\; \frac{\exp(X_{ij}'\beta + V_i'\delta_j)}{\sum_{k=1}^{J}\exp(X_{ik}'\beta + V_i'\delta_k)}$$

where

$$\begin{aligned} X_{ij} &= \{\text{ideological distance}\} \\ V_i &= \{1, \text{issue opinions, party, gender, education, age, ...}\} \end{aligned}$$

- Estimated coefficients:

$$\hat{\beta} \;=\; -0.11\ (0.02)$$

$$\hat{\delta} \;=\; \begin{bmatrix}\hat{\delta}_{\text{Bush}} & \hat{\delta}_{\text{Clinton}}\end{bmatrix} \;=\; \begin{array}{c} \begin{bmatrix} 0.67\ (0.94) & -0.41\ (0.45) \\ -0.52\ (0.11) & -0.02\ (0.12) \\ 0.54\ (0.23) & 0.30\ (0.22) \\ \vdots & \vdots \end{bmatrix} \end{array} \begin{array}{l} \text{(intercept)} \\ \text{(support abortion)} \\ \text{(female)} \\ \vdots \end{array}$$

Example: 1992 U.S. Presidential Election

- Model specification (Alvarez and Nagler 1995):

$$\pi_{ij} \;=\; \frac{\exp(X_{ij}'\beta + V_i'\delta_j)}{\sum_{k=1}^{J}\exp(X_{ik}'\beta + V_i'\delta_k)}$$

where

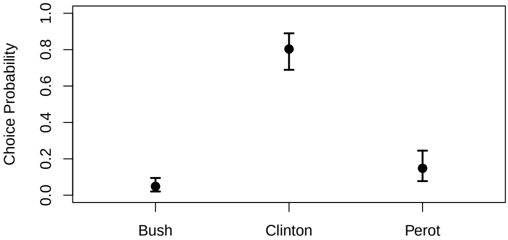
$$\begin{aligned} X_{ij} &= \{\text{ideological distance}\} \\ V_i &= \{1, \text{issue opinions, party, gender, education, age, ...}\} \end{aligned}$$

- Estimated coefficients:

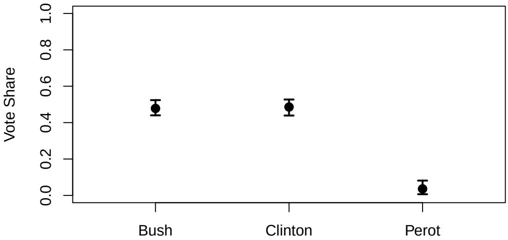
$$\begin{aligned} \hat{\beta} &= -0.11\ (0.02) \\ \hat{\delta} &= \begin{bmatrix} \hat{\delta}_{\text{Bush}} & \hat{\delta}_{\text{Clinton}} \end{bmatrix} = \begin{array}{cl} \begin{bmatrix} 0.67\ (0.94) & -0.41\ (0.45) \\ -0.52\ (0.11) & -0.02\ (0.12) \\ 0.54\ (0.23) & 0.30\ (0.22) \\ \vdots & \vdots \end{bmatrix} & \begin{array}{l} \text{(intercept)} \\ \text{(support abortion)} \\ \text{(female)} \\ \vdots \end{array} \end{array} \end{aligned}$$

Example: 1992 U.S. Presidential Election

- Estimated choice probabilities for a "typical" voter from South:

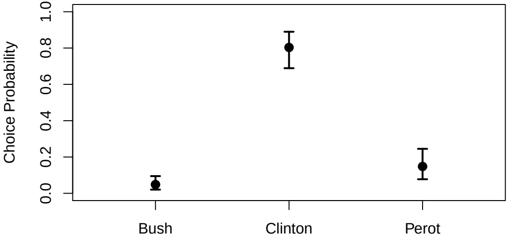


- Predicted vote shares if everyone opposed abortion:

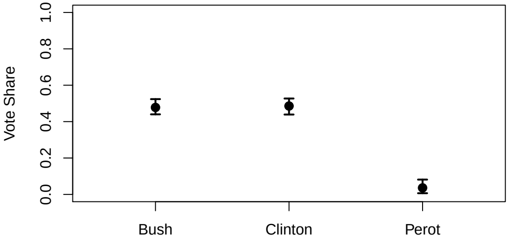


Example: 1992 U.S. Presidential Election

- Estimated choice probabilities for a "typical" voter from South:



- Predicted vote shares if everyone opposed abortion:



Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y^*_{ij} are unrelated to those affecting Y^*_{ik}
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Revisiting The IIA Assumption

- IIA (Trump, Cruz, and Sanders)
- Formally: MNL assumes ϵ_{ij} is i.i.d. $\epsilon_{ij} \perp\!\!\!\perp \epsilon_{ik}$ for $j \neq k$
- This implies that unobserved factors affecting Y_{ij}^* are unrelated to those affecting Y_{ik}^*
- When is this assumption plausible?
- Example: Multiparty election with parties R, L1 and L2.
- Do voters' unobserved ideological preferences affect Pr($Y_i =$L1) independently of their effect on Pr($Y_i =$L2)? Probably not.

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* \;=\; X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \epsilon_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \;\cdots\; Y_{iJ}^*]' \\ X_i = [X_{i1} \;\cdots\; X_{iJ}]' \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i'\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]' \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]' \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]^{'} \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]^{'} \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Multinomial Probit Model

- How can we relax the IIA assumption?
- Instead of assuming ϵ_{ij} to be i.i.d. across alternatives j, we allow ϵ_{ij} to be correlated across j within each voter i
- Multinomial probit model (MNP):

$$Y_i^* = X_i^{'}\beta + \epsilon_i \quad \text{where} \quad \begin{cases} \boldsymbol{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \Sigma_J) \\ Y_i^* = [Y_{i1}^* \ \cdots \ Y_{iJ}^*]' \\ X_i = [X_{i1} \ \cdots \ X_{iJ}]' \end{cases}$$

- Restrictions on the model for identifiability:
  - The (absolute) level of Y_i^* shouldn't matter
    $\longrightarrow$ Subtract the 1st equation from all the other equations and work with a system of $J-1$ equations with $\tilde{\epsilon}_i \sim_{\text{iid}} \text{MVN}(0, \tilde{\Sigma}_{J-1})$
  - The scale of Y_i^* also shouldn't matter
    $\longrightarrow \tilde{\Sigma}_{(1,1)} = 1$

Limitations of Multinomial Probit

- MNP has no closed-form expression for the likelihood:

$$\pi_{ij} = \int_{-\infty}^{-\ddot{X}_{1j}^\top \beta} \cdots \int_{-\infty}^{-\ddot{X}_{Jj}^\top \beta} \phi(\ddot{\epsilon}_{1j}, ..., \ddot{\epsilon}_{Jj}) d\ddot{\epsilon}_{1j} \cdots d\ddot{\epsilon}_{Jj} \text{ where } \begin{cases} \ddot{X}_{kl} & = & X_{ik} - X_{il} \\ \ddot{\epsilon}_{kl} & = & \epsilon_{ik} - \epsilon_{il} \end{cases}$$

- This makes its estimation computationally costly when J large
- Must use numerical approximation (quadratures) or simulation methods (maximum simulated likelihood or MCMC)
- Moreover, # of parameters in Σ_J increases as J gets large, but data contain little information about Σ_J:

J	3	4	5	6	7
# of elements in Σ_J	6	10	15	21	28
# of parameters identified	2	5	9	14	20

- Consequently, MNP is only feasible when J is small
- MNP in R
- Alternative solutions: Nested logit

Limitations of Multinomial Probit

- MNP has no closed-form expression for the likelihood:

$$\pi_{ij} = \int_{-\infty}^{-\ddot{X}_{1j}^{\top}\beta} \cdots \int_{-\infty}^{-\ddot{X}_{Jj}^{\top}\beta} \phi(\ddot{\epsilon}_{1j}, ..., \ddot{\epsilon}_{Jj}) d\ddot{\epsilon}_{1j} \cdots d\ddot{\epsilon}_{Jj} \text{ where } \begin{cases} \ddot{X}_{kl} & = & X_{ik} - X_{il} \\ \ddot{\epsilon}_{kl} & = & \epsilon_{ik} - \epsilon_{il} \end{cases}$$

- This makes its estimation computationally costly when J large
- Must use numerical approximation (quadratures) or simulation methods (maximum simulated likelihood or MCMC)
- Moreover, # of parameters in Σ_J increases as J gets large, but data contain little information about Σ_J:

J	3	4	5	6	7
# of elements in Σ_J	6	10	15	21	28
# of parameters identified	2	5	9	14	20

- Consequently, MNP is only feasible when J is small
- MNP in R
- Alternative solutions: Nested logit

Limitations of Multinomial Probit

- MNP has no closed-form expression for the likelihood:

$$\pi_{ij} = \int_{-\infty}^{-\ddot{X}_{1j}^\top \beta} \cdots \int_{-\infty}^{-\ddot{X}_{Jj}^\top \beta} \phi(\ddot{\epsilon}_{1j}, ..., \ddot{\epsilon}_{Jj}) d\ddot{\epsilon}_{1j} \cdots d\ddot{\epsilon}_{Jj} \text{ where } \begin{cases} \ddot{X}_{kl} & = & X_{ik} - X_{il} \\ \ddot{\epsilon}_{kl} & = & \epsilon_{ik} - \epsilon_{il} \end{cases}$$

- This makes its estimation computationally costly when J large
- Must use numerical approximation (quadratures) or simulation methods (maximum simulated likelihood or MCMC)
- Moreover, # of parameters in Σ_J increases as J gets large, but data contain little information about Σ_J:

J	3	4	5	6	7
# of elements in Σ_J	6	10	15	21	28
# of parameters identified	2	5	9	14	20

- Consequently, MNP is only feasible when J is small
- MNP in R
- Alternative solutions: Nested logit

Limitations of Multinomial Probit

- MNP has no closed-form expression for the likelihood:

$$\pi_{ij} = \int_{-\infty}^{-\ddot{X}_{1j}^\top \beta} \cdots \int_{-\infty}^{-\ddot{X}_{Jj}^\top \beta} \phi(\ddot{\epsilon}_{1j}, ..., \ddot{\epsilon}_{Jj}) d\ddot{\epsilon}_{1j} \cdots d\ddot{\epsilon}_{Jj} \text{ where } \begin{cases} \ddot{X}_{kl} & = & X_{ik} - X_{il} \\ \ddot{\epsilon}_{kl} & = & \epsilon_{ik} - \epsilon_{il} \end{cases}$$

- This makes its estimation computationally costly when J large
- Must use numerical approximation (quadratures) or simulation methods (maximum simulated likelihood or MCMC)
- Moreover, # of parameters in Σ_J increases as J gets large, but data contain little information about Σ_J:

J	3	4	5	6	7
# of elements in Σ_J	6	10	15	21	28
# of parameters identified	2	5	9	14	20

- Consequently, MNP is only feasible when J is small
- `MNP` in `R`
- Alternative solutions: Nested logit

Even When You Choose Not To Decide, You Still Have Made a Choice

Lacy and Burden (1999)
Multinomial Probit Model:
Three Choices: Bush, Perot, and Clinton
Four Choices: Bush, Perot, Clinton, and Abstention
What happens if Perot doesn't run?

	Actual	3-choice	4-choice
Bush	32.0	45.7	38.4
Clinton	48.6	54.3	61.6
Abstention	20.9	-	23.7

Perot stole from Clinton!

Model Based Inference

1) Likelihood inference
2) Logit/Probit
3) Ordered Probit
4) Choice Models:
5) Count Models (Poisson, Negative Binomial,...)