GOV 2001/ 1002/ E-200 Section 11
Choice Modeling and Future Directions in Methods\textsuperscript{1}

Anton Strezhnev

Harvard University

April 20, 2016

\textsuperscript{1}These section notes are heavily indebted to past Gov 2001 TFs for slides and R code.
**LOGISTICS**

**Reading Assignment** - UPM Chapter 8, Glasgow et. al., 2012

**Final Paper** Due April 27th at 5:00 pm. But with automatic extensions to May 5th at 5:00 pm.

**Final Exam (for E-school students not doing the paper)**
Released on April 27th. You will “check out” the exam on Canvas any time during exam period. After check-out, you will have 1 week to finish. The final deadline is May 14th at 5:00 pm.

**Fill out the RSVP for the party on May 7th!** We only have 4 respondents so far!
In this section you will...
OVERVIEW

▶ In this section you will...
  ▶ learn how to model choice data
Overview

- In this section you will...
  - learn how to model choice data
  - learn how latent space models work.
Overview

- In this section you will...
  - learn how to model choice data
  - learn how latent space models work.
  - learn how to generalize from an unrepresentative sample.
In this section you will...

- learn how to model choice data
- learn how latent space models work.
- learn how to generalize from an unrepresentative sample.
- learn how to think about learning methods beyond this course!
OUTLINE

Choice Models

Ideal Point Models

Modern Survey Sampling

Learning more methods
MODELING CHOICES

We want to model some choice among a set of unordered outcomes...

...vote choice in multiparty elections.

...choices among potential coalition partners in government.

...patients choosing different types of medications.

...consumer purchasing decisions.
We want to model some choice among a set of unordered outcomes...
We want to model some choice among a set of unordered outcomes...
  - vote choice in multiparty elections.
Modeling choices

- We want to model some choice among a set of unordered outcomes...
  - ...vote choice in multiparty elections.
  - ...choices among potential coalition partners in government.
MODELING CHOICES

- We want to model some choice among a set of unordered outcomes...
  - ...vote choice in multiparty elections.
  - ...choices among potential coalition partners in government.
  - ...patients choosing different types of medications.
We want to model some choice among a set of unordered outcomes...

- vote choice in multiparty elections.
- choices among potential coalition partners in government.
- patients choosing different types of medications.
- consumer purchasing decisions.
MULTINOMIAL LOGIT
MULTINOMIAL LOGIT

- Can intuitively generalize our friendly logit model to multiple outcomes – the multinomial logit!
MULTINOMIAL LOGIT

▶ Can intuitively generalize our friendly logit model to multiple outcomes – the multinomial logit!
▶ Stochastic component: $Y_i$ is a $J$-length vector -

$$Y_i \sim \text{Multinomial}(1, \vec{\pi}_i)$$
MULTINOMIAL LOGIT

- Can intuitively generalize our friendly logit model to multiple outcomes – the multinomial logit!
- Stochastic component: $Y_i$ is a $J$-length vector -

$$Y_i \sim \text{Multinomial}(1, \vec{\pi}_i)$$

where $\vec{\pi}_i$ is a $J$-length vector of choice probabilities for each of $J$ choices: \{\(\pi_{i1}, \pi_{i2}, \ldots, \pi_{ij}\}\}
**Multinomial Logit**

- Can intuitively generalize our friendly logit model to multiple outcomes – the multinomial logit!
- Stochastic component: $Y_i$ is a $J$-length vector -

$$Y_i \sim \text{Multinomial}(1, \vec{\pi}_i)$$

where $\vec{\pi}_i$ is a $J$-length vector of choice probabilities for each of $J$ choices: \{$\pi_{i1}, \pi_{i2}, \ldots, \pi_{ij}$\}

- Systematic component:

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{K} \exp(\eta_{ik})}$$

$$\eta_{ij} = X_i \beta_j$$
Multinomial logit
MULTINOMIAL LOGIT

- Identification:
**Multinomial logit**

- Identification: Need to fix one category as “baseline”. For notation, that’s $J$. So let $\eta_{iJ} = 0$ and therefore $\exp(\eta_{iJ}) = 1$. 
MULTINOMIAL LOGIT

- Identification: Need to fix one category as “baseline”. For notation, that’s $J$. So let $\eta_{ij} = 0$ and therefore $\exp(\eta_{ij}) = 1$.
- How many parameters are we estimating?
Multinomial Logit

- Identification: Need to fix one category as “baseline”. For notation, that’s $J$. So let $\eta_{ij} = 0$ and therefore $\exp(\eta_{ij}) = 1$.
- How many parameters are we estimating? $J - 1 \times$ length of $\beta$.
- Likelihood $L(\beta|X, Y)$:

$$\propto \prod_{i=1}^{N} \prod_{j=1}^{J} \pi_{ij}^{Y_{ij}}$$

$$\propto \prod_{i=1}^{N} \prod_{j=1}^{J} \left[ \frac{\exp(\eta_{ij})}{\sum_{k=1}^{K} \exp(\eta_{ik})} \right]^{Y_{ij}}$$

$$\propto \prod_{i=1}^{N} \left[ \prod_{j=1}^{J-1} \left[ \frac{\exp(X_i\beta_j)}{1 + \sum_{k=1}^{K-1} \exp(X_i\beta_k)} \right]^{Y_{ij}} \right] \times \left[ \frac{1}{1 + \sum_{k=1}^{K-1} \exp(X_i\beta_k)} \right]^{Y_{iJ}}$$
ASSUMPTION: INDEPENDENCE OF IRRELEVANT ALTERNATIVES

▶ Likelihood massively simplified by assuming logit form for each observation.

▶ However, has implicit assumption about choice behavior: Independence of Irrelevant Alternatives (IIA).

\[
\frac{\pi_{ij}}{\pi_{ik}} = \frac{\exp(\eta_{i1})}{\exp(\eta_{ik})} = \frac{\exp(X_i \beta_1)}{\exp(X_i \beta_2)}
\]
ASSUMPTION: INDEPENDENCE OF IRRELEVANT ALTERNATIVES

- Likelihood massively simplified by assuming logit form for each observation.
ASSUMPTION: INDEPENDENCE OF IRRELEVANT ALTERNATIVES

- Likelihood massively simplified by assuming logit form for each observation.
- However, has implicit assumption about choice behavior: Independence of Irrelevant Alternatives (IIA).
ASSUMPTION: INDEPENDENCE OF IRRELEVANT ALTERNATIVES

- Likelihood massively simplified by assuming logit form for each observation.
- However, has implicit assumption about choice behavior: Independence of Irrelevant Alternatives (IIA).
- Ratio of choice probability of category 1 to 2 does not depend on any other category:

\[
\frac{\pi_{ij}}{\pi_{ik}} = \frac{\exp(\eta_{i1})}{\sum \exp(\eta_{ik})} = \frac{\exp(\eta_{i1})}{\exp(\eta_{i2})} = \frac{\exp(X_i/\beta_1)}{\exp(X_i/\beta_2)}
\]
VIOLATIONS OF IIA

Adding or removing a third option should not affect the ratio of choice probabilities between the other categories.

Commonly violated when choices are substitutes.

Red Bus/Blue Bus problem:

A person chooses between commuting by Car or a Red Bus. They're indifferent so \( \Pr(\text{Car}) = \Pr(\text{Red Bus}) = 0.5 \) and \( \Pr(\text{Car}) \Pr(\text{Red Bus}) = 1 \).

Suppose a third option is introduced - a Blue Bus. Let's assume that the color doesn't really matter to the person, so given that they take a bus, they'll take either with equal probability.

New probs: \( \Pr(\text{Car}) = 0.5 \), \( \Pr(\text{Red Bus}) = \Pr(\text{Blue Bus}) = 0.25 \). \( \Pr(\text{Car}) \Pr(\text{Red Bus}) = 0.25 \neq 1 \).
VIOLATIONS OF IIA

▶ What does it mean for IIA to be violated?

Adding or removing a third option should not affect the ratio of choice probabilities between the other categories.

Commonly violated when choices are substitutes.

Red Bus/Blue Bus problem:

A person chooses between commuting by Car or a Red Bus. They’re indifferent so $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = 0.5$. $\Pr(\text{Car}) \cdot \Pr(\text{Red Bus}) = 0.5$.

Suppose a third option is introduced - a Blue Bus. Let’s assume that the color doesn’t really matter to the person, so given that they take a bus, they’ll take either with equal probability.

New probs: $\Pr(\text{Car}) = 0.5$, $\Pr(\text{Red Bus}) = \Pr(\text{Blue Bus}) = 0.25$. $\Pr(\text{Car}) \cdot \Pr(\text{Red Bus}) = \frac{1}{2} \neq 0.5$.25 = 0.25 = 1.
Violations of IIA

- What does it mean for IIA to be violated? Adding or removing a third option should not affect the ratio of choice probabilities between the other categories.
**Violations of IIA**

- What does it mean for IIA to be violated? Adding or removing a third option should not affect the ratio of choice probabilities between the other categories.
- Commonly violated when choices are *substitutes*. 
VIOLATIONS OF IIA

- What does it mean for IIA to be violated? Adding or removing a third option should not affect the ratio of choice probabilities between the other categories.
- Commonly violated when choices are substitutes.
- Red Bus/Blue Bus problem:
VIOLATIONS OF IIA

- What does it mean for IIA to be violated? Adding or removing a third option should not affect the ratio of choice probabilities between the other categories.

- Commonly violated when choices are substitutes.

- Red Bus/Blue Bus problem:
  - A person chooses between commuting by Car or a Red Bus. They’re indifferent so $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = .5$ and $\frac{\Pr(\text{Car})}{\Pr(\text{Red Bus})} = 1$. 
VIOLATIONS OF IIA

▶ What does it mean for IIA to be violated? Adding or removing a third option should not affect the ratio of choice probabilities between the other categories.

▶ Commonly violated when choices are substitutes.

▶ Red Bus/Blue Bus problem:
  ▶ A person chooses between commuting by Car or a Red Bus. They’re indifferent so $Pr(\text{Car}) = Pr(\text{Red Bus}) = .5$ and $\frac{Pr(\text{Car})}{Pr(\text{Red Bus})} = 1$.
  ▶ Suppose a third option is introduced - a Blue Bus. Let’s assume that the color doesn’t really matter to the person, so given that they take a bus, they’ll take either with equal probability.
What does it mean for IIA to be violated? Adding or removing a third option should not affect the ratio of choice probabilities between the other categories. Commonly violated when choices are substitutes. Red Bus/Blue Bus problem:

- A person chooses between commuting by Car or a Red Bus. They’re indifferent so $Pr(\text{Car}) = Pr(\text{Red Bus}) = .5$ and $\frac{Pr(\text{Car})}{Pr(\text{Red Bus})} = 1$.
- Suppose a third option is introduced - a Blue Bus. Let’s assume that the color doesn’t really matter to the person, so given that they take a bus, they’ll take either with equal probability.
- New probs: $Pr(\text{Car}) = .5$, $Pr(\text{Red Bus}) = Pr(\text{Blue Bus}) = .25$. $\frac{Pr(\text{Car})}{Pr(\text{Red Bus})} = \frac{.5}{.25} = 2 \neq 1$
**Conditional Logit**

The multinomial model only considers attributes of individuals. But we might want to know how characteristics of alternatives/choices affect behavior?

Market research: What's the probability of buying a red car vs. a grey car?

Appointments: Given that a president picks a Supreme Court candidate, how does experience/background affect probability of appointment.

"Conditional" because we are conditioning on a choice being made among a set of alternatives.

Systematic component changes slightly - same logit form, but \( \eta_{ij} \) changes.\( \eta_{ij} = \mathbf{Z}_j \gamma \)

\( \mathbf{Z}_j \) are covariates for alternative \( j \) and \( \gamma \) are estimated coefficients.
**CONDITIONAL LOGIT**

- The multinomial model only considers attributes of individuals.
**Conditional Logit**

- The multinomial model only considers attributes of individuals. But we might want to know how characteristics of alternatives/choices affect behavior?
**Conditional Logit**

- The multinomial model only considers attributes of individuals. But we might want to know how characteristics of alternatives/choices affect behavior?
  - Market research: What’s the probability of buying a red car vs. a grey car
Conditional Logit

- The multinominal model only considers attributes of individuals. But we might want to know how characteristics of alternatives/choices affect behavior?
  - Market research: What’s the probability of buying a red car vs. a grey car
  - Appointments: Given that a president picks a Supreme Court candidate, how does experience/background affect probability of appointment.
**CONDITIONAL LOGIT**

- The multinomial model only considers attributes of individuals. But we might want to know how characteristics of alternatives/choices affect behavior?
  - Market research: What’s the probability of buying a red car vs. a grey car
  - Appointments: Given that a president picks a Supreme Court candidate, how does experience/background affect probability of appointment.
- “Conditional” because we are conditioning on a choice being made among a set of alternatives.
**Conditional Logit**

- The multinomial model only considers attributes of individuals. But we might want to know how characteristics of alternatives/choices affect behavior?
  - Market research: What’s the probability of buying a red car vs. a grey car
  - Appointments: Given that a president picks a Supreme Court candidate, how does experience/background affect probability of appointment.
- “Conditional” because we are conditioning on a choice being made among a set of alternatives.
- Systematic component changes slightly - same logit form, but $\eta_{ij}$ changes

\[ \eta_{ij} = Z_j \gamma \]
**Conditional Logit**

- The multinomial model only considers attributes of individuals. But we might want to know how characteristics of alternatives/choices affect behavior?
  - Market research: What’s the probability of buying a red car vs. a grey car
  - Appointments: Given that a president picks a Supreme Court candidate, how does experience/background affect probability of appointment.

- “Conditional” because we are conditioning on a choice being made among a set of alternatives.

- Systematic component changes slightly - same logit form, but \( \eta_{ij} \) changes

\[
\eta_{ij} = Z_j \gamma
\]

\( Z_j \) are covariates for alternative \( j \) and \( \gamma \) are estimated coefficients.
COMBINING MULTINOMIAL AND CONDITIONAL LOGIT

\[ \eta_{ij} = X_i \beta_j + Z_{ij} \gamma \]
COMBINING MULTINOMIAL AND CONDITIONAL LOGIT

- Can combine the two to estimate both individual and alternative specific attributes (and interactions!)
COMBINING MULTINOMIAL AND CONDITIONAL LOGIT

- Can combine the two to estimate both individual and alternative specific attributes (and interactions!)

\[ \eta_{ij} = X_i \beta_j + Z_{ij} \gamma \]
OUTLINE

Choice Models

Ideal Point Models

Modern Survey Sampling

Learning more methods
Latent Space Modeling

European Parliament. Photo by David Iliff. License: CC-BY-SA 3.0
Latent Space Modeling

European Parliament. Photo by David Iliff. License: CC-BY-SA 3.0

- We have high dimensional data
**Latent Space Modeling**

European Parliament. Photo by David Iliff. License: CC-BY-SA 3.0

- We have high dimensional data
  - ... $M$ votes in Congress by $N$ legislators
Latent Space Modeling

European Parliament. Photo by David Iliff. License: CC-BY-SA 3.0

- We have high dimensional data
  - ... $M$ votes in Congress by $N$ legislators
  - ... $M$ exam questions by $N$ students
- We want to summarize patterns in a meaningful way.
Latent Space Modeling

European Parliament. Photo by David Iliff. License: CC-BY-SA 3.0

- We have high dimensional data
  - ... $M$ votes in Congress by $N$ legislators
  - ... $M$ exam questions by $N$ students
- We want to summarize patterns in a meaningful way.
  - ... which legislators are the most liberal/conservative
Latent Space Modeling

We have high dimensional data

- $M$ votes in Congress by $N$ legislators
- $M$ exam questions by $N$ students

We want to summarize patterns in a meaningful way.

- which legislators are the most liberal/conservative
- which students perform the best on exams.
**Why model?**
**Why model?**

- Intuitive summary - look at % voting agreement between two legislators.
**WHY MODEL?**

- Intuitive summary - look at % voting agreement between two legislators.
- Problem!
**WHY MODEL?**

- Intuitive summary - look at % voting agreement between two legislators.
- Problem! What does 90% agreement mean?
Intuitive summary - look at % voting agreement between two legislators.

Problem! What does 90% agreement mean?

What if those 90% were unanimous votes?
**Why model?**

- Intuitive summary - look at % voting agreement between two legislators.
- Problem! What does 90% agreement mean?
  - What if those 90% were unanimous votes? What if those were votes where *only* those two legislators were on the same side?
**WHY MODEL?**

- Intuitive summary - look at % voting agreement between two legislators.
- Problem! What does 90% agreement mean?
  - What if those 90% were unanimous votes? What if those were votes where *only* those two legislators were on the same side? How should we interpret % agreement?
- Exam Analogy:
Why Model?

- Intuitive summary - look at % voting agreement between two legislators.
- Problem! What does 90% agreement mean?
  - What if those 90% were unanimous votes? What if those were votes where only those two legislators were on the same side? How should we interpret % agreement?
- Exam Analogy:
  - Year 1 - student gets 70% on an exam. Year 2 - student gets 90%.
Why model?

- Intuitive summary - look at % voting agreement between two legislators.
- Problem! What does 90% agreement mean?
  - What if those 90% were unanimous votes? What if those were votes where only those two legislators were on the same side? How should we interpret % agreement?
- Exam Analogy:
  - Year 1 - student gets 70% on an exam. Year 2 - student gets 90%. Did the student improve?
**WHY MODEL?**

- Intuitive summary - look at % voting agreement between two legislators.
- Problem! What does 90% agreement mean?
  - What if those 90% were unanimous votes? What if those were votes where *only* those two legislators were on the same side? How should we interpret % agreement?
- Exam Analogy:
  - Year 1 - student gets 70% on an exam. Year 2 - student gets 90%. Did the student improve? Or did the exam get easier?
- Simple metrics like % agreement miss important variation in agenda.
Item Response Theory (IRT)
ITEM RESPONSE THEORY (IRT)

- Developed in educational testing!
**Item Response Theory (IRT)**

- Developed in educational testing! But huge expansion into psychology, sociology and political science.
**Item Response Theory (IRT)**

- Developed in educational testing! But huge expansion into psychology, sociology and political science.
- **Goal:** Infer latent ability/preferences from observed outcomes (test questions/votes).
Item Response Theory (IRT)

- Developed in educational testing! But huge expansion into psychology, sociology and political science.
- **Goal:** Infer latent ability/preferences from observed outcomes (test questions/votes).
Simple 2-Parameter, 1-Dimensional Model

We observe a $N \times M$ matrix of roll call votes $Y$.

Assume each legislator $i$ has a single latent unobserved "ideal point" $x_i$.

For each vote $j$, the observed outcome $Y_{ij}$ is

$Y_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} \leq 0 \end{cases}$

and $z_{ij}$ is a combination of ideal point, roll call characteristics, and random error.

$z_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}$

Possible to justify this from a "utility maximization" model.
**Simple 2-Parameter, 1-Dimensional Model**

- We observe a $N$ by $M$ matrix of roll call votes $Y$. 

$\begin{align*}
\text{We observe a } N \text{ by } M \text{ matrix of roll call votes } Y. 
\end{align*}$
Simple 2-Parameter, 1-Dimensional Model

- We observe a $N$ by $M$ matrix of roll call votes $Y$. $N$ legislators.
**Simple 2-Parameter, 1-Dimensional Model**

- We observe a $N$ by $M$ matrix of roll call votes $Y$. $N$ legislators. $M$ votes.
**Simple 2-Parameter, 1-Dimensional Model**

- We observe a $N$ by $M$ matrix of roll call votes $Y$. $N$ legislators. $M$ votes.
- Assume each legislator $i$ has a single latent unobserved “ideal point” $x_i$. 

$z_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}$

Possible to justify this from a “utility maximization” model.
Simple 2-Parameter, 1-Dimensional Model

- We observe a $N$ by $M$ matrix of roll call votes $Y$. $N$ legislators. $M$ votes.
- Assume each legislator $i$ has a single latent unobserved “ideal point” $x_i$. For each vote $j$, the observed outcome $Y_{ij}$ is

$$Y_{ij} = \begin{cases} 
1 & \text{if } z_{ij} > 0 \\
0 & \text{if } z_{ij} \leq 0 
\end{cases}$$
Simple 2-Parameter, 1-Dimensional Model

- We observe a $N$ by $M$ matrix of roll call votes $Y$. $N$ legislators. $M$ votes.

- Assume each legislator $i$ has a single latent unobserved "ideal point" $x_i$. For each vote $j$, the observed outcome $Y_{ij}$ is

\[
Y_{ij} = \begin{cases} 
1 & \text{if } z_{ij} > 0 \\
0 & \text{if } z_{ij} \leq 0
\end{cases}
\]

- and $z_{ij}$ is a combination of ideal point, roll call characteristics, and random error.

\[
z_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}
\]
### Simple 2-Parameter, 1-Dimensional Model

- We observe a $N$ by $M$ matrix of roll call votes $Y$. $N$ legislators. $M$ votes.
- Assume each legislator $i$ has a single latent unobserved “ideal point” $x_i$. For each vote $j$, the observed outcome $Y_{ij}$ is

$$Y_{ij} = \begin{cases} 1 & \text{if } z_{ij} > 0 \\ 0 & \text{if } z_{ij} \leq 0 \end{cases}$$

- and $z_{ij}$ is a combination of ideal point, roll call characteristics, and random error.

$$z_{ij} = \alpha_j + \beta_j x_i + \epsilon_{ij}$$

- Possible to justify this from a “utility maximization” model.
SIMPLE 2-PARAMETER, 1-DIMENSIONAL MODEL

If we assume $\epsilon_{ij} \sim N(0,1)$, then we can write

$$\Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)$$

What does that remind us of?

A probit model!

What do the parameters mean?

$\alpha_{ij}$: "difficulty" parameter – For roll calls: if close to 0, then vote is probably evenly split. If large, then vote is probably lopsided.

$\beta_j$: "discrimination" parameter – For roll calls: How well does this vote reflect latent preferences? Positive $\beta_j$: high $x_i = \text{high } \Pr(Y_{ij} = 1)$. Negative $\beta_j$: high $x_i = \text{low } \Pr(Y_{ij} = 1)$. 
Simple 2-Parameter, 1-Dimensional Model

- If we assume $\epsilon_{ij} \sim \mathcal{N}(0, 1)$, then we can write

$$Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)$$
Simple 2-Parameter, 1-Dimensional Model

- If we assume $\epsilon_{ij} \sim \mathcal{N}(0, 1)$, then we can write
  $$Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)$$

- What does that remind us of?
**Simple 2-Parameter, 1-Dimensional Model**

- If we assume \( \epsilon_{ij} \sim \mathcal{N}(0, 1) \), then we can write

\[
Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)
\]

- What does that remind us of? A probit model!
Simple 2-Parameter, 1-Dimensional Model

- If we assume $\epsilon_{ij} \sim \mathcal{N}(0, 1)$, then we can write

$$Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)$$

- What does that remind us of? A probit model!
- What do the parameters mean?
  - $\alpha_{ij}$: “difficulty” parameter –
Simple 2-Parameter, 1-Dimensional Model

- If we assume $\epsilon_{ij} \sim \mathcal{N}(0, 1)$, then we can write

$$Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)$$

- What does that remind us of? A probit model!
- What do the parameters mean?
  - $\alpha_{ij}$: “difficulty” parameter – For roll calls: if close to 0, then vote is probably evenly split. If large, then vote is probably lopsided.
  - $\beta_j$: “discrimination” parameter –
If we assume $\epsilon_{ij} \sim \mathcal{N}(0, 1)$, then we can write

$$Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)$$

What does that remind us of? A probit model!

What do the parameters mean?

- $\alpha_{ij}$: “difficulty” parameter – For roll calls: if close to 0, then vote is probably evenly split. If large, then vote is probably lopsided.
- $\beta_j$: “discrimination” parameter – For roll calls: How well does this vote reflect latent preferences?
If we assume $\epsilon_{ij} \sim N(0, 1)$, then we can write

$$Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)$$

What does that remind us of? A probit model!

What do the parameters mean?

- $\alpha_{ij}$: “difficulty” parameter – For roll calls: if close to 0, then vote is probably evenly split. If large, then vote is probably lopsided.
- $\beta_j$: “discrimination” parameter – For roll calls: How well does this vote reflect latent preferences? Positive $\beta_j$: high $x_i = high \Pr(Y_{ij} = 1)$. 

Simple 2-Parameter, 1-Dimensional Model
Simple 2-Parameter, 1-Dimensional Model

- If we assume $\epsilon_{ij} \sim \mathcal{N}(0, 1)$, then we can write

$$Pr(Y_{ij} = 1) = \Phi(\beta_j x_i - \alpha_j)$$

- What does that remind us of? A probit model!

- What do the parameters mean?
  - $\alpha_{ij}$: “difficulty” parameter – For roll calls: if close to 0, then vote is probably evenly split. If large, then vote is probably lopsided.
  - $\beta_j$: “discrimination” parameter – For roll calls: How well does this vote reflect latent preferences? Positive $\beta_j$: high $x_i$ = high $Pr(Y_{ij} = 1)$. Negative $\beta_j$: high $x_i$ = low $Pr(Y_{ij} = 1)$. 

IRT Example

Figure: Example of latent space model with no voting error

Vote 1
\( \beta > 0 \)
2 Nay (Rep. 1, 2)
3 Yea (Rep. 3 – 5)

Vote 2
\( \beta > 0 \)
4 Nay (Rep. 1 - 4)
1 Yea (Rep. 5)

Vote 3
\( \beta < 0 \)
2 Nay (Rep. 4, 5)
3 Yea (Rep. 1-3)
IDENTIFICATION

We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).

What's the issue with ML estimates? Not identified!

Likelihood depends only on distances between ideal points. Invariant to scale or rotation!

Solutions:

- Constrain scale
- Fix some legislators' locations

Even then, ML estimates are inconsistent. As $N$ gets large, the number of parameters also grows!

More simply - it's just a hard likelihood to maximize!
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
**IDENTIFICATION**

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates?

*Likelihood depends only on distances between ideal points. Invariant to scale or rotation!*
**IDENTIFICATION**

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates? Not identified!
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates? Not identified! Likelihood depends only on distances between ideal points.
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates? Not identified! Likelihood depends only on distances between ideal points. Invariant to scale or rotation!
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates? Not identified! Likelihood depends only on distances between ideal points. Invariant to scale or rotation!
- Solutions:
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).

- What’s the issue with ML estimates? Not identified! Likelihood depends only on distances between ideal points. Invariant to scale or rotation!

- Solutions:
  - Constrain scale
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates? Not identified! Likelihood depends only on distances between ideal points. Invariant to scale or rotation!
- Solutions:
  - Constrain scale
  - Fix some legislators’ locations
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates? Not identified! Likelihood depends only on distances between ideal points. Invariant to scale or rotation!
- Solutions:
  - Constrain scale
  - Fix some legislators’ locations
- Even then, ML estimates are inconsistent.
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates? Not identified! Likelihood depends only on distances between ideal points. Invariant to scale or rotation!
- Solutions:
  - Constrain scale
  - Fix some legislators’ locations
- Even then, ML estimates are inconsistent. As $N$ gets large, the number of parameters also grows!
IDENTIFICATION

- We can write the likelihood as the product of $Y_{ij}$ over $i$ and $j$ (assuming independence between votes).
- What’s the issue with ML estimates? Not identified! Likelihood depends only on distances between ideal points. Invariant to scale or rotation!
- Solutions:
  - Constrain scale
  - Fix some legislators’ locations
- Even then, ML estimates are inconsistent. As $N$ gets large, the number of parameters also grows!
- More simply - it’s just a hard likelihood to maximize!
Bayesian Estimation

Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates). "Markov Chain Monte Carlo" (MCMC) techniques allow us to simulate draws from the posterior and obtain point estimates/credible intervals.

Intuition

\[ f(x, \alpha, \beta | Y) \]

is hard to calculate!

\[ f(\alpha, \beta | x, Y) \]

is just probit regression!

\[ f(x | \alpha, \beta, Y) \]

is also a regression problem!

MCMC methods repeatedly take draws from these conditionals. Markov chain theory tells us that this converges to drawing from the true posterior!
Bayesian Estimation

- Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates).
**Bayesian Estimation**

- Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates).
- “Markov Chain Monte Carlo” (MCMC) techniques allow us to simulate draws from the posterior and obtain point estimates/credible intervals.
Bayesian Estimation

- Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates).
- “Markov Chain Monte Carlo” (MCMC) techniques allow us to simulate draws from the posterior and obtain point estimates/credible intervals.
- Intuition
Bayesian Estimation

- Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates).
- “Markov Chain Monte Carlo” (MCMC) techniques allow us to simulate draws from the posterior and obtain point estimates/credible intervals.
- Intuition
  - Posterior $f(x, \alpha, \beta|Y)$ hard to calculate!
Bayesian Estimation

- Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates).
- “Markov Chain Monte Carlo” (MCMC) techniques allow us to simulate draws from the posterior and obtain point estimates/credible intervals.

Intuition

- Posterior $f(x, \alpha, \beta|Y)$ hard to calculate!
- But $f(\alpha, \beta|x, Y)$ is just probit regression!
Bayesian Estimation

- Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates).
- “Markov Chain Monte Carlo” (MCMC) techniques allow us to simulate draws from the posterior and obtain point estimates/credible intervals.

Intuition

- Posterior $f(x, \alpha, \beta | Y)$ hard to calculate!
- But $f(\alpha, \beta | x, Y)$ is just probit regression!
- And $f(x | \alpha, \beta, Y)$ is also a regression problem!
**Bayesian Estimation**

- Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates).

- “Markov Chain Monte Carlo” (MCMC) techniques allow us to simulate draws from the posterior and obtain point estimates/credible intervals.

- **Intuition**
  - Posterior $f(x, \alpha, \beta|Y)$ hard to calculate!
  - But $f(\alpha, \beta|x, Y)$ is just probit regression!
  - And $f(x|\alpha, \beta, Y)$ is also a regression problem!
  - MCMC methods repeatedly take draws from these conditionals.
Most modern ideal point estimation techniques rely on Bayesian approaches (with priors on the ideal point and roll call parameters to constrain the estimates).

“Markov Chain Monte Carlo” (MCMC) techniques allow us to simulate draws from the posterior and obtain point estimates/credible intervals.

**Intuition**
- Posterior $f(x, \alpha, \beta | Y)$ hard to calculate!
- But $f(\alpha, \beta | x, Y)$ is just probit regression!
- And $f(x | \alpha, \beta, Y)$ is also a regression problem!
- MCMC methods repeatedly take draws from these conditionals. Markov chain theory tells us that this converges to drawing from the true posterior!
What IRT models can show us.

Figure: Dynamic ideal point estimates of P5 countries from UNGA voting - Voeten et. al. (2015)
<table>
<thead>
<tr>
<th>Logistics</th>
<th>Overview</th>
<th>Choice Models</th>
<th>Ideal Point Models</th>
<th>Modern Survey Sampling</th>
<th>Learning more methods</th>
</tr>
</thead>
</table>

**Outline**

- Choice Models
- Ideal Point Models
- Modern Survey Sampling
- Learning more methods
The Crisis in Polling

Traditional landline-based surveys no longer get close to a representative sample of the population. Growth of cell-phone use. Abysmal response rates (5% to 15% for Pew). Pretty much all pollsters work with non-random samples. True of researchers as well. Why do we talk about random samples then? Theoretical exposition!
THE CRISIS IN POLLING

- Traditional landline-based surveys no longer get close to a representative sample of the population.
THE CRISIS IN POLLING

- Traditional landline-based surveys no longer get close to a representative sample of the population.
  - Growth of cell-phone use
The Crisis in Polling

- Traditional landline-based surveys no longer get close to a representative sample of the population.
  - Growth of cell-phone use
  - Abysmal response rates (5% to 15% for Pew)
THE CRISIS IN POLLING

▶ Traditional landline-based surveys no longer get close to a representative sample of the population.
  ▶ Growth of cell-phone use
  ▶ Abysmal response rates (5% to 15% for Pew)
▶ Pretty much all pollsters work with non-random samples.
The Crisis in Polling

- Traditional landline-based surveys no longer get close to a representative sample of the population.
  - Growth of cell-phone use
  - Abysmal response rates (5% to 15% for Pew)
- Pretty much all pollsters work with non-random samples.
- True of researchers as well.
THE CRISIS IN POLLING

- Traditional landline-based surveys no longer get close to a representative sample of the population.
  - Growth of cell-phone use
  - Abysmal response rates (5% to 15% for Pew)
- Pretty much all pollsters work with non-random samples.
- True of researchers as well. Why do we talk about random samples then?
THE CRISIS IN POLLING

- Traditional landline-based surveys no longer get close to a representative sample of the population.
  - Growth of cell-phone use
  - Abysmal response rates (5% to 15% for Pew)
- Pretty much all pollsters work with non-random samples.
- True of researchers as well. Why do we talk about random samples then? Theoretical exposition!
GENERALIZING FROM NON-RANDOM SAMPLES

How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling!

Ideally, we'd know $\text{Pr(Person Selected)}$. When we do by design, can weight by $\frac{1}{\text{Pr(Person Selected)}}$. This is rare though (especially for internet convenience samples).

One common approach – “Multilevel Regression and Post-stratification” (MRP) or Mr. P!

- **Multilevel Regression**: Make a model predicting individual response using individual and group (e.g. county/state) variables.
- **Post-stratification**: Use what we know about population-level covariate distribution to reweight data for predictions.
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample?
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling!
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling!
  Meaning: Assumptions!
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling!
  Meaning: Assumptions!
- Ideally, we’d know $Pr(\text{Person Selected})$. 
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling!
  Meaning: Assumptions!

- Ideally, we’d know $Pr($Person Selected$)$. When we do by design, can weight by $\frac{1}{Pr($Person Selected$)}$. 
GENERALIZING FROM NON-RANDOM SAMPLES

► How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling! Meaning: Assumptions!

► Ideally, we’d know $Pr(\text{Person Selected})$. When we do by design, can weight by $\frac{1}{Pr(\text{Person Selected})}$. This is rare though (especially for internet convenience samples).
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling!
  Meaning: Assumptions!
- Ideally, we’d know $Pr(\text{Person Selected})$. When we do by design, can weight by $\frac{1}{Pr(\text{Person Selected})}$. This is rare though (especially for internet convenience samples).
- One common approach – “Multilevel Regression and Post-stratification” (MRP)
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling!
  Meaning: Assumptions!
  - Ideally, we’d know $Pr(\text{Person Selected})$. When we do by design, can weight by $\frac{1}{Pr(\text{Person Selected})}$. This is rare though (especially for internet convenience samples).
  - One common approach – “Multilevel Regression and Post-stratification” (MRP) (or Mr. P!)
GENERALIZING FROM NON-RANDOM SAMPLES

► How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling! Meaning: Assumptions!

► Ideally, we’d know $Pr(\text{Person Selected})$. When we do by design, can weight by $\frac{1}{Pr(\text{Person Selected})}$. This is rare though (especially for internet convenience samples).

► One common approach – “Multilevel Regression and Post-stratification” (MRP) (or Mr. P!)
  ► Multilevel Regression:
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling!
  Meaning: Assumptions!

- Ideally, we’d know $Pr(\text{Person Selected})$. When we do by design, can weight by $\frac{1}{Pr(\text{Person Selected})}$. This is rare though (especially for internet convenience samples).

- One common approach – “Multilevel Regression and Post-stratification” (MRP) (or Mr. P!)
  - Multilevel Regression: Make a model predicting individual response using individual and group (e.g. county/state) variables.
**GENERALIZING FROM NON-RANDOM SAMPLES**

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling! Meaning: Assumptions!

- Ideally, we’d know $Pr($Person Selected$)$. When we do by design, can weight by $\frac{1}{Pr($Person Selected$)}$. This is rare though (especially for internet convenience samples).

- One common approach – “Multilevel Regression and Post-stratification” (MRP) (or Mr. P!)
  - Multilevel Regression: Make a model predicting individual response using *individual* and *group* (e.g. county/state) variables.
  - Post-stratification:
GENERALIZING FROM NON-RANDOM SAMPLES

- How do we get an unbiased estimated of the population using a non-random sample? Statistical modelling! Meaning: Assumptions!
- Ideally, we’d know $Pr(\text{Person Selected})$. When we do by design, can weight by $\frac{1}{Pr(\text{Person Selected})}$. This is rare though (especially for internet convenience samples).
- One common approach – “Multilevel Regression and Post-stratification” (MRP) (or Mr. P!)
  - Multilevel Regression: Make a model predicting individual response using individual and group (e.g. county/state) variables.
  - Post-stratification: Use what we know about population-level covariate distribution to reweight data for predictions.
Post-Stratification
Post-Stratification

- Why post-stratification?
Post-Stratification

- Why post-stratification?
  - “Stratification” because we’re trying to match our sample proportions to population proportions for certain strata (age, gender, etc...)
POST-STRATIFICATION

- Why post-stratification?
  - “Stratification” because we’re trying to match our sample proportions to population proportions for certain strata (age, gender, etc...)
  - “Post-” because we do it after sampling.
Post-Stratification

- Why post-stratification?
  - “Stratification” because we’re trying to match our sample proportions to population proportions for certain strata (age, gender, etc...)
  - “Post-” because we do it after sampling.
- Example: Using a convenient internet sample, predict % that would vote for Obama in Hennepin County, MN.
POST-STRATIFICATION

► Why post-stratification?
  ▶ “Stratification” because we’re trying to match our sample proportions to population proportions for certain strata (age, gender, etc...)
  ▶ “Post-” because we do it after sampling.
► Example: Using a convenient internet sample, predict % that would vote for Obama in Hennepin County, MN.
► How?
**Post-Stratification**

- Why post-stratification?
  - “Stratification” because we’re trying to match our sample proportions to population proportions for certain strata (age, gender, etc...)
  - “Post-” because we do it after sampling.

- Example: Using a convenient internet sample, predict % that would vote for Obama in Hennepin County, MN.

- How? Re-weight our sample to match known population characteristics of Hennepin County.
POST-STRATIFICATION

Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.

Step 2: Identify covariates that are correlated with the outcome you care about. E.g. For vote choice: Party ID, gender, race, income, etc...

Step 3: Get data on the distributions of covariates for your population of interest. Often proportions (e.g. % registered Dem, % White, etc...). Often using census data.

Step 4: Calculate weights for your observations such that the (weighted) sample distributions of covariates match the population distributions.
Post-Stratification

- Basic post-stratification weighting
**Post-Stratification**

- Basic post-stratification weighting
  - Step 1:
POST-STRATIFICATION

- Basic post-stratification weighting
  - Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.
POST-STRATIFICATION

- Basic post-stratification weighting
  - Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.
  - Step 2:
Post-Stratification

- Basic post-stratification weighting
  - Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.
  - Step 2: Identify covariates that are correlated with the outcome you care about.
**Post-Stratification**

- Basic post-stratification weighting
  - Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.
  - Step 2: Identify covariates that are correlated with the outcome you care about. E.g. For vote choice: Party ID, gender, race, income, etc...
Post-Stratification

- Basic post-stratification weighting
  - Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.
  - Step 2: Identify covariates that are correlated with the outcome you care about. E.g. For vote choice: Party ID, gender, race, income, etc...
  - Step 3: Get data on the distributions of covariates for your population of interest.
**Post-Stratification**

- Basic post-stratification weighting
  - Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.
  - Step 2: Identify covariates that are correlated with the outcome you care about. E.g. For vote choice: Party ID, gender, race, income, etc...
  - Step 3: Get data on the distributions of covariates for your population of interest. Often proportions (e.g. % registered Dem, % White, etc...).
POST-STRATIFICATION

- Basic post-stratification weighting
  - Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.
  - Step 2: Identify covariates that are correlated with the outcome you care about. E.g. For vote choice: Party ID, gender, race, income, etc...
  - Step 3: Get data on the distributions of covariates for your population of interest. Often proportions (e.g. % registered Dem, % White, etc...). Often using census data.
POST-STRATIFICATION

- Basic post-stratification weighting
  - Step 1: Identify the population of interest - e.g. whole U.S. population, state of NY, Cuyahoga County.
  - Step 2: Identify covariates that are correlated with the outcome you care about. E.g. For vote choice: Party ID, gender, race, income, etc...
  - Step 3: Get data on the distributions of covariates for your population of interest. Often proportions (e.g. % registered Dem, % White, etc...). Often using census data.
  - Step 4: Calculate weights for your observations such that the (weighted) sample distributions of covariates match the population distributions.
Calculating PS Weights

When the full distribution of strata is known, weight for observation $i$ in stratum $h$:

$$w_i = \frac{n}{n_h} \times P_h$$

$n$ is the sample size

$n_h$ is the number of sample obs. in stratum $h$

$P_h$ is the population proportion in stratum $h$

Can also think of it as $w_i = P_h p_h$ where $p_h$ is the sample proportion in stratum $h$.

Intuition: Upweight observations that are rare relative to population. Downweight observations that are common.
Calculating PS Weights

- When the full distribution of strata is known, weight for observation $i$ in stratum $h$:

$$w_i = \frac{n}{n_h} \times P_h$$
Calculating PS Weights

- When the full distribution of strata is known, weight for observation $i$ in stratum $h$:

$$w_i = \frac{n}{n_h} \times P_h$$

- $n$ is the sample size
Calculating PS Weights

- When the full distribution of strata is known, weight for observation $i$ in stratum $h$:

$$w_i = \frac{n}{n_h} \times P_h$$

- $n$ is the sample size
- $n_h$ is the number of sample obs. in stratum $h$
CALCULATING PS WEIGHS

- When the full distribution of strata is known, weight for observation $i$ in stratum $h$:

$$w_i = \frac{n}{n_h} \times P_h$$

- $n$ is the sample size
- $n_h$ is the number of sample obs. in stratum $h$
- $P_h$ is the population proportion in stratum $h$
Calculating PS Weights

- When the full distribution of strata is known, weight for observation $i$ in stratum $h$:

$$w_i = \frac{n}{n_h} \times P_h$$

- $n$ is the sample size
- $n_h$ is the number of sample obs. in stratum $h$
- $P_h$ is the population proportion in stratum $h$

- Can also think of it as

$$w_i = \frac{P_h}{p_h}$$

where $p_h$ is the sample proportion in stratum $h$.

- **Intuition:** Upweight observations that are rare relative to population. Downweight observations that are common.
Raking

However, you rarely have the full joint distribution for lots of covariates. Just the marginals.

Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.

Solution: “raking” - iteratively reweight to match the population marginals as closely as possible.

Implemented in the R package survey.

Raking procedure:

Step 1: Calculate PS weights for the first variable.
Step 2: Using those weights, calculate the new in-sample proportions of the second variable.
Step 3: Recalculate the PS weights for the second variable given the previously calculated weighting.
Step 4: Repeat across all of the variables in sequence until convergence (no change in weights).
**Raking**

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
**Raking**

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census.
RAKING

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.
Raking

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.
- Solution: “raking”
**Raking**

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.

- Solution: “raking” - iteratively reweight to match the population marginals as closely as possible
Raking

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.
- Solution: “raking” - iteratively reweight to match the population marginals as closely as possible
  - Implemented in the R package survey
Raking

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.
- Solution: “raking” - iteratively reweight to match the population marginals as closely as possible
  - Implemented in the R package `survey`
- Raking procedure:
Raking

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.
- Solution: “raking” - iteratively reweight to match the population marginals as closely as possible
  - Implemented in the R package survey
- Raking procedure:
  - Step 1: Calculate PS weights for the first variable.
Raking

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.
- Solution: “raking” - iteratively reweight to match the population marginals as closely as possible
  - Implemented in the R package survey
- Raking procedure:
  - Step 1: Calculate PS weights for the first variable.
  - Step 2: Using those weights, calculate the new in-sample proportions of the second variable.
Raking

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.
- Solution: “raking” - iteratively reweight to match the population marginals as closely as possible
  - Implemented in the R package `survey`
- Raking procedure:
  - Step 1: Calculate PS weights for the first variable.
  - Step 2: Using those weights, calculate the new in-sample proportions of the second variable.
  - Step 3: Re-calculate the PS weights for the second variable given the previously calculated weighting.
**Raking**

- However, you rarely have the full joint distribution for lots of covariates. Just the marginals.
  - Ex. We know % White, % Women, % Age 18-35 from census. But we don’t know % White Women Age 18-35.

- Solution: “raking” - iteratively reweight to match the population marginals as closely as possible
  - Implemented in the R package `survey`

- **Raking procedure:**
  - Step 1: Calculate PS weights for the first variable.
  - Step 2: Using those weights, calculate the new in-sample proportions of the second variable.
  - Step 3: Re-calculate the PS weights for the second variable given the previously calculated weighting.
  - Step 4: Repeat across all of the variables in sequence until convergence (no change in weights).
OUTLINE

Choice Models

Ideal Point Models

Modern Survey Sampling

Learning more methods
Why Methods?
**Why Methods?**

- Research is about persuasion.
Why Methods?

- Research is about persuasion.
  - Given some data, why should I believe your story about the world?
**Why Methods?**

- Research is about persuasion.
  - Given some data, why should I believe your story about the world?
- Research methods provide a *common* language for arguing from data.
Why Methods?

- Research is about persuasion.
  - Given some data, why should I believe your story about the world?
- Research methods provide a *common* language for arguing from data.
- Statistics gives us a coherent set of rules for drawing inferences given data.
Why Methods?

- Research is about persuasion.
  - Given some data, why should I believe your story about the world?
- Research methods provide a *common* language for arguing from data.
- Statistics gives us a coherent set of rules for drawing inferences given data.
  - Still have to argue for the assumptions.
**Why Methods?**

- Research is about persuasion.
  - Given some data, why should I believe your story about the world?
- Research methods provide a *common* language for arguing from data.
- Statistics gives us a coherent set of rules for drawing inferences given data.
  - Still have to argue for the assumptions.
  - But statistics gives us tools to adjudicate.
**Why Methods?**

- Research is about persuasion.
  - Given some data, why should I believe your story about the world?
- Research methods provide a *common* language for arguing from data.
- Statistics gives us a coherent set of rules for drawing inferences given data.
  - Still have to argue for the assumptions.
  - But statistics gives us tools to adjudicate. And different methods entail different arguments – they rely on different (sometimes weaker/more flexible) assumptions.
**Why Methods?**

- Research is about persuasion.
  - Given some data, why should I believe your story about the world?

- Research methods provide a *common* language for arguing from data.

- Statistics gives us a coherent set of rules for drawing inferences given data.
  - Still have to argue for the assumptions.
  - But statistics gives us tools to adjudicate. And different methods entail different arguments – they rely on different (sometimes weaker/more flexible) assumptions.
  - Alternatively, can justify assumptions via *design* (e.g. randomization/natural experiments).
Why Methods?

▶ Research is about persuasion.
  ▶ Given some data, why should I believe your story about the world?

▶ Research methods provide a *common* language for arguing from data.

▶ Statistics gives us a coherent set of rules for drawing inferences given data.
  ▶ Still have to argue for the assumptions.
  ▶ But statistics gives us tools to adjudicate. And different methods entail different arguments – they rely on different (sometimes weaker/more flexible) assumptions.
  ▶ Alternatively, can justify assumptions via design (e.g. randomization/natural experiments).

▶ Perusading researchers requires you to make arguments that make sense to both you and them.
Why Methods?

- Research is about persuasion.
  - Given some data, why should I believe your story about the world?
- Research methods provide a common language for arguing from data.
- Statistics gives us a coherent set of rules for drawing inferences given data.
  - Still have to argue for the assumptions.
  - But statistics gives us tools to adjudicate. And different methods entail different arguments – they rely on different (sometimes weaker/more flexible) assumptions.
  - Alternatively, can justify assumptions via design (e.g. randomization/natural experiments).
- Persuading researchers requires you to make arguments that make sense to both you and them. Statistical methods lay out one useful method of argumentation.
WHAT METHODS?

Don’t think only in terms of learning about specific methods. Instead, focus on how to understand those papers and how to fit those methods into your repertoire.

E.g.: Don’t just learn “text analysis,” learn how to think about high-dimensional data where $p >> n$. 
What Methods?

- Don’t think only in terms of learning about specific methods.
WHAT METHODS?

▶ Don’t think only in terms of learning about specific methods.
  ▶ You’ll be learning fancy new methods by reading papers.
WHAT METHODS?

- Don’t think only in terms of learning about specific methods.
  - You’ll be learning fancy new methods by reading papers. Instead, focus on how to understand those papers and how to fit those methods into your repertoire.
WHAT METHODS?

▶ Don’t think only in terms of learning about specific methods.
▶ You’ll be learning fancy new methods by reading papers. Instead, focus on how to understand those papers and how to fit those methods into your repertoire.
▶ E.g.: Don’t just learn “text analysis,” learn how to think about high-dimensional data where $p >> n$. 
What Methods?
**WHAT METHODS?**

- What questions does a method help you answer better?
What Methods?

▶ What questions does a method help you answer better?
  ▶ Modeling.
WHAT METHODS?

- What questions does a method help you answer better?
  - **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
What Methods?

- What questions does a method help you answer better?
  - **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    - Primary topic of this course.
WHAT METHODS?

▶ What questions does a method help you answer better?
  ▶ **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    ▶ Primary topic of this course. Choosing among models is often challenging.
WHAT METHODS?

▶ What questions does a method help you answer better?
  ▶ **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    ▶ Primary topic of this course. Choosing among models is often challenging.
    ▶ Trade-off between flexibility and parsimony.
WHAT METHODS?

▷ What questions does a method help you answer better?
  ▷ **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    ▷ Primary topic of this course. Choosing among models is often challenging.
    ▷ Trade-off between flexibility and parsimony.
    ▷ Simple example: Logit models for binary outcomes
WHAT METHODS?

▷ What questions does a method help you answer better?
  ▷ **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    ▷ Primary topic of this course. Choosing among models is often challenging.
    ▷ Trade-off between flexibility and parsimony.
    ▷ Simple example: Logit models for binary outcomes
    ▷ Complex example: Latent space models for roll call voting.
WHAT METHODS?

- What questions does a method help you answer better?
  - **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    - Primary topic of this course. Choosing among models is often challenging.
    - Trade-off between flexibility and parsimony.
    - Simple example: Logit models for binary outcomes
    - Complex example: Latent space models for roll call voting.
  - **Estimation.**
What Methods?

- **What questions does a method help you answer better?**
  - **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    - Primary topic of this course. Choosing among models is often challenging.
    - Trade-off between flexibility and parsimony.
    - Simple example: Logit models for binary outcomes
    - Complex example: Latent space models for roll call voting.
  - **Estimation.** How can I get a good estimate of a quantity using my data?
WHAT METHODS?

▶ What questions does a method help you answer better?
  ▶ **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    ▶ Primary topic of this course. Choosing among models is often challenging.
    ▶ Trade-off between flexibility and parsimony.
    ▶ Simple example: Logit models for binary outcomes
    ▶ Complex example: Latent space models for roll call voting.
  ▶ **Estimation.** How can I get a good estimate of a quantity using my data?
    ▶ Bias/Variance trade-offs. Frequentist vs. Bayesian approaches.
What Methods?

- What questions does a method help you answer better?
  - **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    - Primary topic of this course. Choosing among models is often challenging.
    - Trade-off between flexibility and parsimony.
    - Simple example: Logit models for binary outcomes
    - Complex example: Latent space models for roll call voting.
  - **Estimation.** How can I get a good estimate of a quantity using my data?
    - Bias/Variance trade-offs. Frequentist vs. Bayesian approaches.
  - **Identification.**
What Methods?

▷ What questions does a method help you answer better?
  ▷ **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    ▷ Primary topic of this course. Choosing among models is often challenging.
    ▷ Trade-off between flexibility and parsimony.
    ▷ Simple example: Logit models for binary outcomes
    ▷ Complex example: Latent space models for roll call voting.
  ▷ **Estimation.** How can I get a good estimate of a quantity using my data?
    ▷ Bias/Variance trade-offs. Frequentist vs. Bayesian approaches.
  ▷ **Identification.** How can I connect quantities I can estimate to quantities that I care about (e.g. causal effects)?
What Methods?

▶ What questions does a method help you answer better?
  ▶ **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    ▶ Primary topic of this course. Choosing among models is often challenging.
    ▶ Trade-off between flexibility and parsimony.
    ▶ Simple example: Logit models for binary outcomes
    ▶ Complex example: Latent space models for roll call voting.
  ▶ **Estimation.** How can I get a good estimate of a quantity using my data?
    ▶ Bias/Variance trade-offs. Frequentist vs. Bayesian approaches.
  ▶ **Identification.** How can I connect quantities I can estimate to quantities that I care about (e.g. causal effects)?
    ▶ Often contributions in terms of research design.
What Methods?

- What questions does a method help you answer better?
  - **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    - Primary topic of this course. Choosing among models is often challenging.
    - Trade-off between flexibility and parsimony.
    - Simple example: Logit models for binary outcomes
    - Complex example: Latent space models for roll call voting.
  - **Estimation.** How can I get a good estimate of a quantity using my data?
    - Bias/Variance trade-offs. Frequentist vs. Bayesian approaches.
  - **Identification.** How can I connect quantities I can estimate to quantities that I care about (e.g. causal effects)?
    - Often contributions in terms of research design.
What Methods?

▶ What questions does a method help you answer better?
  ▶ **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    ▶ Primary topic of this course. Choosing among models is often challenging.
    ▶ Trade-off between flexibility and parsimony.
    ▶ Simple example: Logit models for binary outcomes
    ▶ Complex example: Latent space models for roll call voting.
  ▶ **Estimation.** How can I get a good estimate of a quantity using my data?
    ▶ Bias/Variance trade-offs. Frequentist vs. Bayesian approaches.
  ▶ **Identification.** How can I connect quantities I can estimate to quantities that I care about (e.g. causal effects)?
    ▶ Often contributions in terms of research design.
    ▶ Simple example: randomization for causal effects.
What Methods?

- What questions does a method help you answer better?
  - **Modeling.** How can I parsimoniously summarize my data in a meaningful way?
    - Primary topic of this course. Choosing among models is often challenging.
    - Trade-off between flexibility and parsimony.
    - Simple example: Logit models for binary outcomes
    - Complex example: Latent space models for roll call voting.
  - **Estimation.** How can I get a good estimate of a quantity using my data?
    - Bias/Variance trade-offs. Frequentist vs. Bayesian approaches.
  - **Identification.** How can I connect quantities I can estimate to quantities that I care about (e.g. causal effects)?
    - Often contributions in terms of research design.
    - Simple example: randomization for causal effects.
    - More complex examples: instrumental variables, regression discontinuity
What Methods?

▶ All of these are elements of almost every methods paper!
  ▶ I can write down a really complicated model...
What Methods?

- All of these are elements of almost every methods paper!
  - I can write down a really complicated model... but it’s useless if I can’t estimate it!
  - I can get a really efficient estimate of some regression parameter...
What Methods?

► All of these are elements of almost every methods paper!
  ▶ I can write down a really complicated model... but it’s useless if I can’t estimate it!
  ▶ I can get a really efficient estimate of some regression parameter... but if I want to claim causality, it’s useless if I can’t also argue that it identifies a causal parameter of interest.

► Main Takeaway: Think first in terms of what you need to better argue from your data, then go out and find what you don’t know.
QUESTIONS

Questions?