CSS MS bootcamp: Data-based Approaches to Ideological Conflict

Professor Seth Hill
Department of Political Science
UC San Diego
sjhill@ucsd.edu

August 21, 2023

About me

- Professor of Political Science. Twelfth year at UCSD.
- PhD UCLA, postdoc Yale.
- Interested in how citizens motivate politician behavior.
- Vote choice, candidate behavior, representation, elections, learning about politics.
- `www.sethjhill.com`

You?

- Name and goal for program.

Political coalitions

- Start with a little political science.
- Representative democracies := electorate empowers small group of citizens to make law.
- Voters' challenge: state must make decisions on many different issues. Yet voters mostly get single choices in elections.
- Political coalitions form to bundle multiple issues and offer simplified choice to voters.
- How do issues go together? Don't really know. But they do.

Political coalitions

- Ideology used to describe political coalitions that represent many issues.
- Ideological types Left and Right go back to French Revolution.
- Estates General met at Versailles, May 5 1789: clergy (First Estate), nobility (Second Estate), commoners (Third Estate).
- Supporters of King Louis XVI sat on right of chamber, opponents on left.
- Third Estate breaks away to form National Assembly. Eventually guillotine, Napoleon, etc.
- Names Left and Right stuck.

Political coalitions

- What do we mean by *ideology*, *left*, and *right* today?
- Joe Biden left. Donald Trump right. Why?
- Not agreed upon, but common definition:
- Ideology a general orientation toward government intervention.
- Left-progressives more favorable towards intervention.
- Right-conservatives less favorable.

Ideology and elections

- If candidates roughly summarized by an ideology, voters can choose by the more simple metric ideology.
- Alternative: having to learn every issue position of every candidate, anticipate what might happen in the future, etc.
- If voters across electorate make choices based on candidate ideology, election determines ideology of those who control government.
- If voters want more intervention, favor and elect more left-leaning politicians.
- If voters want less intervention, favor and elect more right-leaning politicians.

Ideology and elections

- This ideological theory of elections generates representation without voters needing to know everything.
- Task of political science: how well do voters get what they want out of government?
- Ideological theory of elections can be used to develop research questions.
- For example: When the legislature full of right-leaning politicians, what happens to policy?
- For example: Are voters who favor more government intervention more likely to vote for left-leaning candidates?

Towards computational social science

- Computer revolution allows political scientists to use data to estimate left-right ideology.

1. Legislators cast thousands of roll call votes in legislature. Do votes look well summarized by latent dimension(s) of ideology?
2. Voters cast multiple votes in elections. Do vote choices look well summarized by ideology?
3. Campaign contributors give different amounts of money to different candidates. Do contributions look well summarized by ideology?
4. Survey respondents answer many questions. Do responses look well summarized by ideology?

Latent variable models

- Ideology a latent variable. Something we want for social scientific purposes but cannot measure directly.
- Actually, many problems across social sciences latent variables:
- Measuring intelligence (latent) by responses to test questions (observed).
- Measuring utility (latent) by choices made in a decision situation (observed).
- Measuring attitudes (latent) by responses to survey questions (observed).
- Measuring career prospects (latent) by course work, written work, capstone project (observed).

Latent variables models

- Statistical problem:
- Inference about *unobserved* quantities given related *observed* quantities.
- Solution:
- Assume observed quantities a function, manifestation, or indicator of latent variable(s).

Latent variable models

All latent variables models have common features:

1. Multiple indicators $\boldsymbol{y}$.
2. Parametric connection from latent variable ψ to observed indicators $\boldsymbol{y}$ via parameters $\boldsymbol{\theta}$, e.g., coefficients.

- Similar goal but distinct grounding from machine learning nonparametric classification methods, e.g., k-nearest neighbors, support vector machines, kernel, etc.
- Political science methods developed before machine learning.
- Also connect to a theory of elections.

Simple example

- Students in a class take a multiple choice test, goal to estimate knowledge of material.
- Score each response correct/incorrect 1/0.

Simple example

- Students in a class take a multiple choice test, goal to estimate knowledge of material.
- Score each response correct/incorrect 1/0.
- Assume ψ_i latent knowledge of student *i*, and $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})$ a vector of 1/0 response for student *i* to each question *j*.

Simple example

- Students in a class take a multiple choice test, goal to estimate knowledge of material.
- Score each response correct/incorrect 1/0.
- Assume ψ_i latent knowledge of student i, and $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})$ a vector of 1/0 response for student i to each question j.
- Probability model? Response binary: probit.
  $\rightarrow$ Probability student i responds correctly to question j, $\Phi(\beta_j \psi_i - \alpha_j)$.

Simple example

- Students in a class take a multiple choice test, goal to estimate knowledge of material.
- Score each response correct/incorrect 1/0.
- Assume ψ_i latent knowledge of student i, and $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})$ a vector of 1/0 response for student i to each question j.
- Probability model? Response binary: probit.
  $\rightarrow$ Probability student i responds correctly to question j, $\Phi(\beta_j \psi_i - \alpha_j)$.
- What we did: collected observed indicators related to latent variable of interest, identified parametric model relating observed to unobserved through parameters, here $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

Latent variable models

- Different variable types of observed responses lead to different models, but all have similar goals.
- Factor analysis or principal components for continuous responses.
- Item-response theory (IRT) models usually categorical responses.
- IRT and its variants commonly used in analysis of voting because model parameters can be mapped back to theories of spatial voting.

Categorical responses

- Item-response theory developed to model binary correct/incorrect responses in test-settings to measure latent intelligence.
- Why not simply sum correct answers?
- Because each test question not equally easy to answer so sum can be inaccurate.
- $\rightarrow$ IRT models re-weight responses to different questions based on difficulty to get more accurate measure of latent intelligence.

Imaginary test

1. Is America more a democracy or a representative democracy?
2. What is derivative of natural log?
3. How many letters are in English alphabet?
4. Is chocolate ice cream better than vanilla ice cream?

Imaginary test

1. Is America more a democracy or a representative democracy?
2. What is derivative of natural log?
3. How many letters are in English alphabet?
4. Is chocolate ice cream better than vanilla ice cream?

- We might grade this test simply out of 4, and our measure of intelligence/knowledge would be fraction for each student {0/4, 1/4, 2/4, 3/4, 4/4}.

Imaginary test

1. Is America more a democracy or a representative democracy?
2. What is derivative of natural log?
3. How many letters are in English alphabet?
4. Is chocolate ice cream better than vanilla ice cream?

- However, we might not think each question equally representative.
- Should we really rate a student who answers only questions 1 and 2 correct equally able as a student who gets only questions 3 and 4 correct?

Item-response models

- *Single-parameter item-response model* allows questions of different *difficulty*.
- Define π_{ij} probability student i answers question j correctly. (unobserved)
- Assume π_{ij} a parametric function of student's ability and question difficulty
$$\pi_{ij} = \Pr(y_{ij} = 1 | \psi_i, \alpha_j) = F(\psi_i - \alpha_j),$$
where F is some function that translates real values to [0,1], e.g., probit or logit.

Item-response models

- Probability a function of student's ability and question difficulty

$$\pi_{ij} = \Pr(y_{ij} = 1 | \psi_i, \alpha_j) = F(\psi_i - \alpha_j),$$

- ψ_i ability of subject i: As ψ_i increases, so does π_{ij}.
- α_j *item difficulty* for question j: As α_j increases, π_{ij} decreases.
- From our previous set of questions we might guess that $\alpha_2 > \alpha_1 > \alpha_3$ with $\alpha_4 =$?.

Item-response models

$$\pi_{ij} = \Pr(y_{ij} = 1 | \psi_i, \alpha_j) = F(\psi_i - \alpha_j),$$

- Likelihood for single-parameter item response model similar to probit or logit, except across $J > 1$ items:

$$\mathcal{L} = \prod_{i=1}^{N} \prod_{j=1}^{J} (\pi_{ij})^{y_{ij}} (1 - \pi_{ij})^{1 - y_{ij}}.$$

- Estimate parameters $\boldsymbol{\theta} = [\boldsymbol{\psi}, \boldsymbol{\alpha}]$.
- Either or both might be of interest.

Item-response models

- With single-parameter item-response model, unclear what would/should occur with question 4 about ice cream ...
- $\rightarrow$ Two-parameter item-response model.
- Adds second parameter (β here) to capture extent to which item captures variation in ability, so-called discrimination:

$$\pi_{ij} = \Pr(y_{ij} = 1 | \psi_i, \beta_j, \alpha_j) = F(\psi_i \beta_j - \alpha_j).$$

Item-response models

$$\pi_{ij} = \Pr(y_{ij} = 1 | \psi_i, \beta_j, \alpha_j) = F(\psi_i \beta_j - \alpha_j).$$

- As β_j increases from zero, probability of correct response ...?
- As β_j decreases from zero, probability of *incorrect* response ...?
- When β_j close to zero, correct response to question j not much related to subject ability. Might imagine β_4 for ice cream question to be ≈ 0.
- Note single-parameter IRT model is two-parameter model with $\beta_k = 1 \; \forall \; k \; \in \; [1, \ldots, J]$.

Item-response models

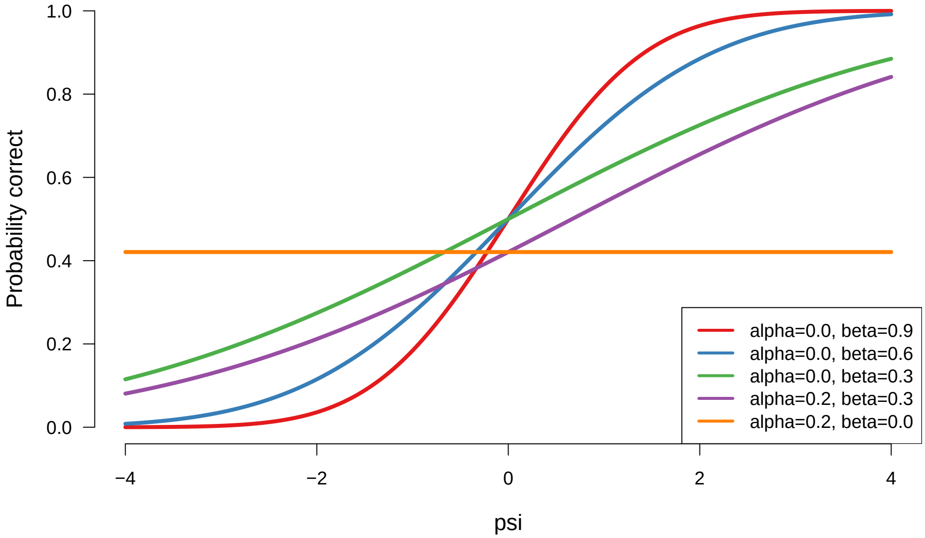
$$\pi_{ij} = \Pr(y_{ij} = 1 | \psi_i, \beta_j, \alpha_j) = F(\psi_i \beta_j - \alpha_j).$$

- Likelihood extends single-parameter model,

$$\mathcal{L} = \prod_{i=1}^{N} \prod_{j=1}^{J} (\pi_{ij})^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}},$$

  estimate parameters $\boldsymbol{\theta} = [\boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\alpha}]$.

Item-response curves



Item-response models

- Q: What do item-response models have to do with political ideology?
- Operationalizing what is called the *Euclidean spatial voting model* to an IRT statistical model.
- Latent traits/abilities = legislator/voter's ideology, called *ideal point*.
- Item difficulty and discrimination parameters = functions of features of what legislators vote upon.

Example voting model

- Imagine we observe a set of roll call votes from legislators in Congress:

1. Endorsing apple pie as an important American food.
2. Increasing estate tax rate.
3. Cutting Medicaid.
4. Prohibiting drones in United States.

- As with test questions, we might imagine these questions have varying difficulty (how many yes votes they receive), as well as varying discrimination (how related to latent dimension of ideology).

Procedure

- Normally, for IRT estimation we create a matrix representation of item data.
- Legislators in rows, votes in columns (*roll call matrix*).
- Response for person *i* to item *j* in corresponding cell:

	Vote			
	Pie	Tax	Medicaid	Drones
Legislator				
A	Yes	Yes	No	Yes
B	Yes	Yes	No	No
C	Yes	No	Yes	No
D	Yes	No	Yes	Yes
E	Yes	No	Yes	Yes

Procedure

- Full likelihood:

$$\mathcal{L} = \prod_{i=1}^{N} \prod_{j=1}^{J} \Phi(\psi_i \beta_j - \alpha_j)^{y_{ij}} (1 - \Phi(\psi_i \beta_j - \alpha_j))^{1-y_{ij}}.$$

	Vote			
	Pie	Tax	Medicaid	Drones
Legislator				
A	1	1	0	1
B	1	1	0	0
C	1	0	1	0
D	1	0	1	1
E	1	0	1	1

Procedure

- Likelihood, Legislator A:

$$\mathcal{L}_A = \Phi(\psi_A \beta_{Pie} - \alpha_{Pie}) \times \Phi(\psi_A \beta_{Tax} - \alpha_{Tax}) \times$$
$$(1 - \Phi(\psi_A \beta_{Medicaid} - \alpha_{Medicaid})) \times \Phi(\psi_A \beta_{Drones} - \alpha_{Drones}).$$

	Vote			
	Pie	Tax	Medicaid	Drones
Legislator				
A	1	1	0	1
B	1	1	0	0
C	1	0	1	0
D	1	0	1	1
E	1	0	1	1

Workshop: Your turn

- Suppose we knew $\boldsymbol{\alpha} = (0.9, 0.2, 0.5, 0.4)$ and $\beta = (0.05, 0.7, -0.2, -0.1)$ for $j =$ (Pie, Tax, Medicaid, Drones).
- Use five OLS regressions to estimate $\psi_A, \psi_B, \psi_C, \psi_D, \psi_E$.
- Remember, $\mathcal{L}_{ij} = \Phi(\psi_i \beta_j - \alpha_j)^{y_{ij}} (1 - \Phi(\psi_i \beta_j - \alpha_j))^{1-y_{ij}}$.
- Hint: With a linear model, if $y = a + bx + e$, it is also the case that $y - a = bx + e$.

	Vote			
	Pie	Tax	Medicaid	Drones
Legislator				
A	1	1	0	1
B	1	1	0	0
C	1	0	1	0
D	1	0	1	1
E	1	0	1	1

Workshop

`R` code:

Workshop

R code:

```
est_ideo = function(y, a=c(0.9,0.2,0.5,0.4),
                       b=c(0.05,0.7,-0.2,-0.1)) {
  # Estimate ideology by OLS for vote vector y.
  # Arguments.
  #  y - vector of votes.
  #  a,b - vectors of vote parameters.
  # Set up data frame.
  dat = data.frame(y=y,a=a,b=b)
  dat[,"yLessA"] = dat$y - dat$a
  # Regression with no intercept.
  lm_out = lm(yLessA ~ -1 + b, data=dat)
  # Return estimate of ideology (coefficient on b).
  summary(lm_out)$coefficients['b',]
}
est_ideo(y=c(1, 1, 0, 1)) # Legislator A.
est_ideo(y=c(1, 1, 0, 0)) # Legislator B.
est_ideo(y=c(1, 0, 1, 0)) # Legislator C.
est_ideo(y=c(1, 0, 1, 1)) # Legislators D and E.
```

Break

- Let's take 15 minutes, then return to statistical estimation in `R`.

Statistical estimation

- Problem: We observe $\boldsymbol{y}$ but not $\boldsymbol{\psi}$, $\boldsymbol{\alpha}$, or $\boldsymbol{\beta}$.
- If we knew $\boldsymbol{\psi}$, could regress (logit/probit) $\boldsymbol{y}$ on $\boldsymbol{\psi}$ to estimate α_j and β_j, vote by vote.
- If we knew $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, could regress (logit/probit) $\boldsymbol{y}$ on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to estimateψ_i, legislator by legislator.

Statistical estimation

- Problem: We observe $\boldsymbol{y}$ but not $\boldsymbol{\psi}$, $\boldsymbol{\alpha}$, or $\boldsymbol{\beta}$.
- If we knew $\boldsymbol{\psi}$, could regress (logit/probit) $\boldsymbol{y}$ on $\boldsymbol{\psi}$ to estimate α_j and β_j, vote by vote.
- If we knew $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, could regress (logit/probit) $\boldsymbol{y}$ on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to estimateψ_i, legislator by legislator.
- Approach: pool all i and all j into one estimation. Leverage information in both directions, votes and legislators.

Statistical estimation

- Options: Bayesian or maximum likelihood inference.
- Bayesian has many nice properties, but a bit more setup.
- Today: maximum likelihood.

Workshop with me

1. Toy example.
2. Legislator ideology by party, 103rd House (1993-1994).

- `www.sethjhill.com/computational-social-science/`

Workshop if time

1. More complicated example.
2. Legislator ideology by party, 89th House (1965-1966).
3. Legislator ideology by party, 117th House (2021-2022).
4. Comparisons.

Spatial voting model

Spatial voting model

- Each vote/bill represented by points in Euclidean space. Taking one-dimensional case, each voter assumed to have a most preferred policy position, θ.
- Voter's utility for various alternatives defined by function of distance between position of alternative and voter's ideal point. Function might be quadratic, Gaussian, etc.
- Assuming quadratic utility, voter's utility for alternative A is

$$U(\theta, A) = -(\theta - A)^2 + \epsilon,$$

where ϵ an idiosyncratic shock, often assumed $\sim N(0, \sigma^2)$ and i.i.d. across alternatives.

Relationship to spatial model

- Difference in utility between two alternatives, A and B is

$$\begin{aligned} U(\theta, A) - U(\theta, B) &= -(\theta - A)^2 + \epsilon_A - [-(\theta - B)^2 + \epsilon_B] \\ &= (B^2 - A^2) + 2(A - B)\theta - (\epsilon_B - \epsilon_A). \end{aligned}$$

- If voters (sincerely) choose alternative offering higher utility, probability of choice A over choice B is

$$\begin{aligned} \Pr(A|\theta) &= \Pr\left[(B^2 - A^2) + 2(A - B)\theta - (\epsilon_B - \epsilon_A) > 0\right] \\ &= \Pr\left[(B^2 - A^2) + 2(A - B)\theta > (\epsilon_B - \epsilon_A)\right]. \end{aligned}$$

Relationship to spatial model

$$\Pr(A|\theta) = \Pr\big[(B^2 - A^2) + 2(A - B)\theta > (\epsilon_B - \epsilon_A)\big].$$

- If $\epsilon \sim N(0, \sigma^2)$, then $\epsilon_B - \epsilon_A \sim N(0, \sqrt{2}\sigma^2)$ (by i.i.d.).
- Define $\alpha = (A^2 - B^2)/\sqrt{2}\sigma^2$.
- Define $\beta = 2(A - B)/\sqrt{2}\sigma^2$, and
- Let V_i indicate voter i's choice between A and B, where $V_i = 1$ when A chosen and $V_i = 0$ when B chosen. Then, we have following probabilistic model of choice:

$$\Pr(V_i = v_i|\theta, \alpha, \beta) = \Phi(\beta\theta_i - \alpha)^{v_i}\big(1 - \Phi(\beta\theta_i - \alpha)\big)^{1-v_i}.$$

- Two-parameter probit IRT model! (Assuming logistic errors instead of normal errors would yield logit IRT)