Causal Inference with Ordinal Outcomes

Density Estimation Based Approach

Chanhyuk Park
Washington University in St. Louis

MOTIVATION: THE DISCRETIZATION PROBLEM

- Experiments to identify causal effects of treatments
- Outcomes are usually believed to be continuous in unidimensional space
  - Approval ratings (Canes-Wrone and De Marchi, 2002; Kriner and Schwartz, 2009)
  - Policy preferences (Scheve and Slaughter, 2001; Mayda and Rodrik, 2005; Wu, 2022)
- The Problem - How we measure it

MOTIVATION: THE DISCRETIZATION PROBLEM

- Experiments to identify causal effects of treatments
- Outcomes are usually believed to be continuous in unidimensional space
  - Approval ratings (Canes-Wrone and De Marchi, 2002; Kriner and Schwartz, 2009)
  - Policy preferences (Scheve and Slaughter, 2001; Mayda and Rodrik, 2005; Wu, 2022)
- The Problem - How we measure it

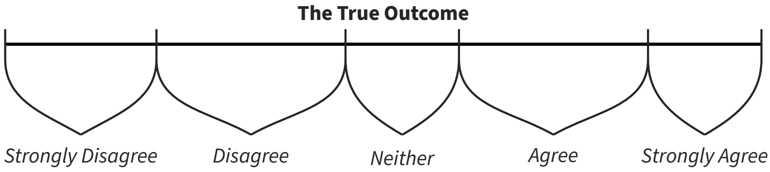
MOTIVATION: THE DISCRETIZATION PROBLEM

- Experiments to identify causal effects of treatments
- Outcomes are usually believed to be continuous in unidimensional space
  - Approval ratings (Canes-Wrone and De Marchi, 2002; Kriner and Schwartz, 2009)
  - Policy preferences (Scheve and Slaughter, 2001; Mayda and Rodrik, 2005; Wu, 2022)
- The Problem - How we measure it

The True Outcome

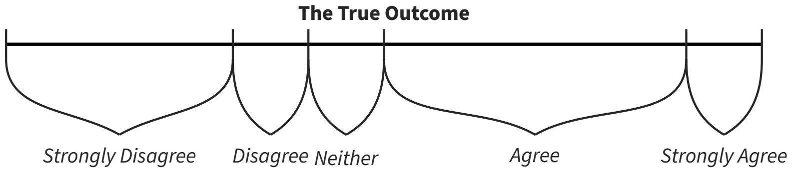
MOTIVATION: THE DISCRETIZATION PROBLEM

- Experiments to identify causal effects of treatments
- Outcomes are usually believed to be continuous in unidimensional space
  - Approval ratings (Canes-Wrone and De Marchi, 2002; Kriner and Schwartz, 2009)
  - Policy preferences (Scheve and Slaughter, 2001; Mayda and Rodrik, 2005; Wu, 2022)
- The Problem - How we measure it



MOTIVATION: THE DISCRETIZATION PROBLEM

- Experiments to identify causal effects of treatments
- Outcomes are usually believed to be continuous in unidimensional space
  - Approval ratings (Canes-Wrone and De Marchi, 2002; Kriner and Schwartz, 2009)
  - Policy preferences (Scheve and Slaughter, 2001; Mayda and Rodrik, 2005; Wu, 2022)
- The Problem - How we measure it



THE GOAL

- *Identification*
- *Estimation*

THE GOAL

- *Identification*

  Naive causal identification with these "index" may fail

  $\Longrightarrow$ Normalized Latent Treatment Effect

- *Estimation*

THE GOAL

- *Identification*

  Naive causal identification with these "index" may fail

  $\Longrightarrow$ Normalized Latent Treatment Effect

- *Estimation*

  Standard parametric regression is rigid

  $\Longrightarrow$ Flexible, density estimation based estimators

WHY NAIVE IDENTIFICATION FAILS

Preference over Political System (Y^*)

0 *10*

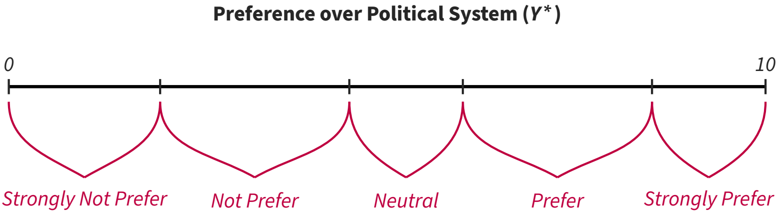
- The treatment effect we want: $\mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] = \tau^*$
- *Unknown Reporting Function g*
- *Arbitrary Numerical Index f*
- What we get: $\mathbb{E}[f(g(Y^*(1)))] - \mathbb{E}[f(g(Y^*(0)))] \neq \tau^*$

WHY NAIVE IDENTIFICATION FAILS

Preference over Political System (Y^*)

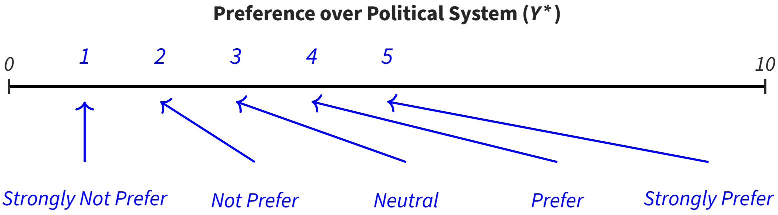
- The treatment effect we want: $\mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] = \tau^*$
- *Unknown Reporting Function g*
- *Arbitrary Numerical Index f*
- What we get: $\mathbb{E}[f(g(Y^*(1)))] - \mathbb{E}[f(g(Y^*(0)))] \neq \tau^*$

WHY NAIVE IDENTIFICATION FAILS



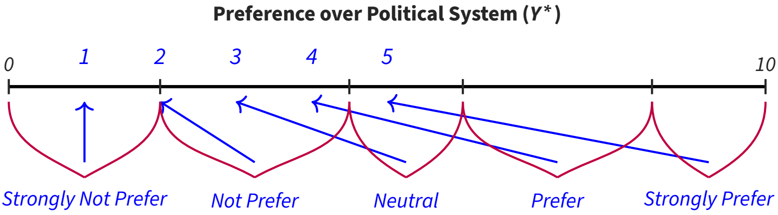
- The treatment effect we want: $\mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] = \tau^*$
- *Unknown Reporting Function g*
- *Arbitrary Numerical Index f*
- What we get: $\mathbb{E}[f(g(Y^*(1)))] - \mathbb{E}[f(g(Y^*(0)))] \neq \tau^*$

WHY NAIVE IDENTIFICATION FAILS



- The treatment effect we want: $\mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] = \tau^*$
- *Unknown Reporting Function g*
- *Arbitrary Numerical Index f*
- What we get: $\mathbb{E}[f(g(Y^*(1)))] - \mathbb{E}[f(g(Y^*(0)))] \neq \tau^*$

WHY NAIVE IDENTIFICATION FAILS



- The treatment effect we want: $\mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] = \tau^*$
- *Unknown Reporting Function g*
- *Arbitrary Numerical Index f*
- What we get: $\mathbb{E}[f(g(Y^*(1)))] - \mathbb{E}[f(g(Y^*(0)))] \neq \tau^*$

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Nonparametrically identified probability distribution of each answers
  $\Longrightarrow$ identify the Treatment Effect up to scale
  $\Longrightarrow$ treatment effect τ^* cannot be distinguished form $C\tau^*$

- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$
  3. Pure probability scale

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Nonparametrically identified probability distribution of each answers

  $\Longrightarrow$ identify the Treatment Effect up to scale

  $\Longrightarrow$ treatment effect τ^* cannot be distinguished form $C\tau^*$

- Normalization to cancel out C

  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$
  3. Pure probability scale

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Nonparametrically identified probability distribution of each answers
  $\Longrightarrow$ identify the Treatment Effect up to scale
  $\Longrightarrow$ treatment effect τ^* cannot be distinguished form $C\tau^*$
- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$
  3. Pure probability scale

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Nonparametrically identified probability distribution of each answers
  $\Longrightarrow$ identify the Treatment Effect up to scale
  $\Longrightarrow$ treatment effect τ^* cannot be distinguished form $C\tau^*$

- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$
  3. Pure probability scale

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Nonparametrically identified probability distribution of each answers
  $\implies$ identify the Treatment Effect up to scale
  $\implies$ treatment effect τ^* cannot be distinguished form $C\tau^*$

- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$
  3. Pure probability scale

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Nonparametrically identified probability distribution of each answers
  $\implies$ identify the Treatment Effect up to scale
  $\implies$ treatment effect τ^* cannot be distinguished form $C\tau^*$

- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$

     $\implies$ $\frac{\tau^*}{\sigma_{\varepsilon_i}}$ is identified
  2. If there are multiple treatments, fixing $\tau_a = 1$
  3. Pure probability scale

HOW CAN WE ESTIMATE?

- Ordered Probit, Ordered Logit, and most IRT models

  $\Longrightarrow$ Assume that ε_i follows specific distributions

- Distributional Assumptions can be violated
  - Pure misspecification
  - Unaccounted confounder

- Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

HOW CAN WE ESTIMATE?

- Ordered Probit, Ordered Logit, and most IRT models
  $\implies$ Assume that ε_i follows specific distributions

- Distributional Assumptions can be violated
  - Pure misspecification
  - Unaccounted confounder

- Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

HOW CAN WE ESTIMATE?

- Ordered Probit, Ordered Logit, and most IRT models

  $\Longrightarrow$ Assume that ε_i follows specific distributions

- Distributional Assumptions can be violated
  - Pure misspecification
  - Unaccounted confounder

- Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

HOW CAN WE ESTIMATE?

- Ordered Probit, Ordered Logit, and most IRT models

  $\Longrightarrow$ Assume that ε_i follows specific distributions

- Distributional Assumptions can be violated
  - Pure misspecification
  - Unaccounted confounder

- Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

- Nonparametric method: Kernel Density Estimation
- Parametric Generative Model: Normalizing Flows
- Plug into MLE

$$\ell(\alpha, \beta, \tau) = \sum_{i=1}^{n} \sum_{j=0}^{J} 1_{Y_i=j} \log \left[\hat{F}_{\varepsilon_i}\left(D_i, \tau_i\right) \hat{F}_{\varepsilon_i}\left(D_i, \tau_i\right) \right]$$

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

- Nonparametric method: Kernel Density Estimation
  - Smooth each observation using a kernel (usually Gaussian)

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

- Parametric Generative Model: Normalizing Flows

- Plug into MLE

$$\ell(\alpha, \beta, \tau) = \sum_{i=1}^{n}\sum_{j=0}^{J} 1_{Y_i=j} \log\left[\hat{F}_{\varepsilon_i}(D_i, \tau_i)\hat{F}_{\varepsilon_i}(D_i, \tau_i)\right]$$

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

- Nonparametric method: Kernel Density Estimation
  - Smooth each observation using a kernel (usually Gaussian)

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

- Parametric Generative Model: Normalizing Flows
  - Use the change-of-variable formula

$$f_\theta(\varepsilon) = f_Z\left(T_\theta^{-1}(\varepsilon)\right) \left|\det\left(\frac{\partial T_\theta^{-1}(\varepsilon)}{\partial \varepsilon}\right)\right|,$$

  - T_θ is a set of *invertible* transformation

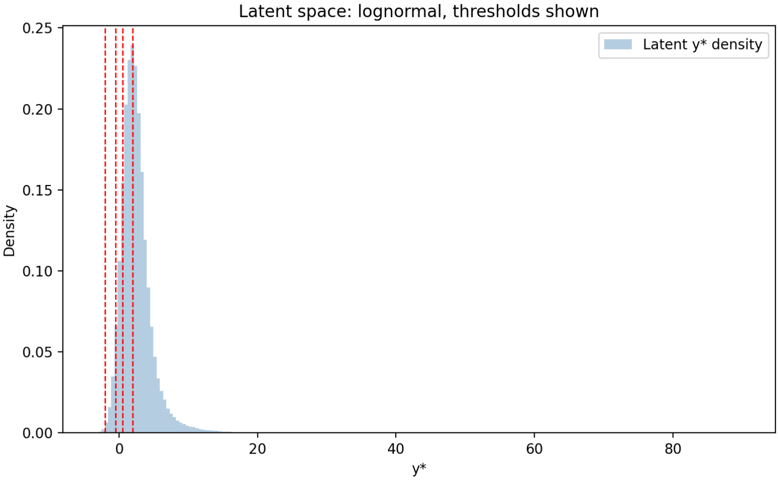
- Plug into MLE

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

- Nonparametric method: Kernel Density Estimation
- Parametric Generative Model: Normalizing Flows
- Plug into MLE

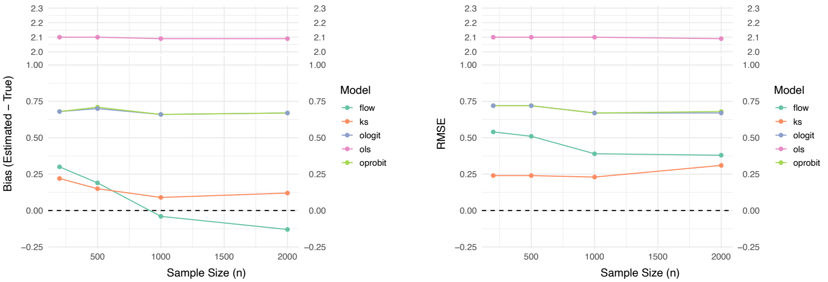
$$\ell(\alpha, \beta, \tau) = \sum_{i=1}^{n} \sum_{j=0}^{J} 1_{Y_i = j} \log \left[\hat{F}_{\varepsilon_i}\left(D_i, \tau_i\right) \hat{F}_{\varepsilon_i}\left(D_i, \tau_i\right) \right]$$

SIMULATION



Latent space: lognormal, thresholds shown
Latent y* density
Density
y*
0.25
0.20
0.15
0.10
0.05
0.00
0
20
40
60
80

SIMULATION – LOGNORMAL CASE



True treatment size: 0.46

Replication – Mattingly et al. (2025)

- Effect of promotion videos on preference over political system
- *Strongly Prefer the US* to *Strongly Prefer China*

	ATE	NLTE			
	Original – OLS	Ordered Logit	Ordered Probit	KDE-based	NF-based
China	1.04^{***}	0.73^{***}	0.76^{***}	0.36^{***}	0.37^{***}
	(0.05)	(0.04)	(0.04)	(0.07)	(0.04)
USA	-0.43^{***}	-0.35^{***}	-0.38^{***}	-0.18^{***}	-0.17^{***}
	(0.04)	(0.04)	(0.04)	(0.05)	(0.03)
Competition	0.36^{***}	0.21^{***}	0.25^{***}	0.13^{*}	0.11^{**}
	(0.05)	(0.04)	(0.04)	(0.06)	(0.04)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

CONCLUSION

- *Identification*
- *Estimation*

CONCLUSION

- *Identification*

  Naive use of ordinal "index" is not what we want

  $\Longrightarrow$ Normalized Latent Treatment Effect

- *Estimation*

CONCLUSION

- *Identification*

  Naive use of ordinal "index" is not what we want

  $\Longrightarrow$ Normalized Latent Treatment Effect

- *Estimation*

  Standard ordinal regressions risk inconsistency

  $\Longrightarrow$ KDE or NF based estimators