Causal Inference Workshop

Week 9 - Inference

Causal Inference Workshop

March 29, 2024

Anna Papp, ap3907@columbia.edu - SDEV 9280

Workshop outline

A. Causal inference fundamentals
   - Modeling assumptions matter too
   - Conceptual framework (potential outcomes framework)
B. Design stage: common identification strategies
   - IV + RDD [coding]
   - DiD, DiDiD, Event Studies, New TWFE Lit [coding]
   - Synthetic Control / Synthetic DiD [coding]
C. Analysis stage: strengthening inferences
   - Limitations of identification strategies, pre-estimation steps
   - Estimation [controls] and post-estimation steps [supporting assumptions]
D. Other topics in causal inference and sustainable development
   - Inference (randomization inference, bootstrapping, etc.) [coding]
   - Fixed effects, weather data regressions, remote sensing data
   - Intro to text analysis, other topics (?), and wrap up!

Outline

Workshop outline

Introduction

Randomization inference

Bootstrapping

Inference

- First focused on *identification* (overcoming selection bias)

- Last week discussed some limitation of identification strategies
  - **Pre-estimation steps**: restructuring data
  - **Estimation steps**: required/forbidden controls for *bias*; good/bad controls for *efficiency*
  - **Post-estimation steps**: diagnosis tests of *modeling assumptions*; falsification tests of *identifying assumptions*

- But we haven't (really) talked about uncertainty and inference yet
  - We talked a bit about assumptions of the CLRM (normal errors)
  - Recall that the asymptotic distribution of $\hat{\beta}_{OLS}$ is:

$$\hat{\beta}_{OLS} \overset{a}{\sim} \mathcal{N}\left(\beta_0, (X'X)^{-1}X'\Sigma X(X'X)^{-1'}\right)$$

$\rightarrow$ need a consistent estimate of the asymptotic vcov matrix in order to do *sampling-based* statistical inference

Outline

Sampling-based inference

- Our frequentist inference techniques → *sampling-based* inference
  - Consider variation in **sampling**: uncertainty about population parameters is induced by random sampling from the population
  - Properties of estimators arise from random sampling of units from a large population in combination with assumptions on the population distribution
  - *What would have occurred under a different random sample?*

Sampling-based inference

- Formulate an H_0 that represents a fact about the data we'll try to refute
  - In causal inference, generally some hypothesis of no effect
  - H_0: No *average* effect, $\mathbb{E}[Y_i^1] - \mathbb{E}[Y_i^0] = 0$
  - H_a: Some *average* effect, $\mathbb{E}[Y_i^1] - \mathbb{E}[Y_i^0] \neq 0$
- Derive a test statistic T s.t. when H_0 is true, T has a specific distribution
  - $T = \frac{\hat{\beta}-\beta_0}{SE[\hat{\beta}]} = \frac{\hat{\beta}-0}{SE[\hat{\beta}]} \rightarrow$ under H_0, the distribution of T across all random samples converges to a known distribution: Student's $\mathcal{T}$
  - When we assume that the distribution of the error term is normal, we know the distribution of T (without normality, holds asymptotically)
- Look at where our T_{obs} lies within the distribution
  - If it lies in a tail $\rightarrow \downarrow$ likelihood of data under $H_0 \rightarrow \uparrow$ confidence against H_0
  - Get p-value (prob. that the observed difference b/w groups would have been observed if they had been drawn from underlying sampling frames with no mean difference)

Sampling-based inference vs. **randomization**-based inference

- Our frequentist inference techniques → *sampling-based* inference
  - Consider variation in **sampling**: uncertainty about population parameters is induced by random sampling from the population
  - Properties of estimators arise from random sampling of units from a large population in combination with assumptions on the population distribution
  - *What would have occurred under a different random sample?*

- Is there another type of uncertainty?
  - For example, what about when the sample is not really different from population? (E.g., data for all 50 states of the US or data for all visits to a website)

Sampling-based inference vs. **randomization**-based inference

- Our frequentist inference techniques → *sampling-based* inference
  - Consider variation in **sampling**: uncertainty about population parameters is induced by random sampling from the population
  - Properties of estimators arise from random sampling of units from a large population in combination with assumptions on the population distribution
  - *What would have occurred under a different random sample?*

- In causal inference, there is also variation in *assignment of treatment* → *design-based* uncertainty
  - Consider variation in **randomization** of treatment allocation as the basis for inference
  - Lack of knowledge about the values that the regression outcome would have taken under alternative interventions
  - *What would have occurred under a different random assignment of treatment among units?*

Randomization inference

- Follow the same general approach to hypothesis testing
- Formulate an H_0 that represents a fact about the data we'll try to refute
  - Consider the **sharp** null hypothesis of no effect *for any unit*
  - $H_0 : Y_i = Y^0 = Y^1 \rightarrow$ the key to randomization inference!
  - Means that our data hold the outcomes of all possible experiments
- Derive a test statistic T s.t. when H_0 is true, T has a specific distribution
  - We can construct each possible random assignment and estimate $\hat{\beta}$ for each $\rightarrow$ this distribution **is** the reference distribution under H_0
- Look at where our T_{obs} lies within the distribution
  - Get p-value: if our $\hat{\beta}_{obs}$ is in the tails, e.g., only 2% of all random assignments produce $\hat{\beta} \geqslant \hat{\beta}_{obs}$, one-tailed p-value = 0.02
  - *Not* for constructing CIs (without additional assumptions)

Randomization inference in practice

- Simulation-based approach
- When *all* possible random assignments can be simulated, the reference distribution is known and RI produces *exact* p-values
- In practice, approximate the reference distribution, repeat many (e.g., 10,000) times:
  1. Generate fake treatment statuses (re-assign treatment randomly)
  2. Estimate the model using these fake treatments, store $\hat{\beta}$

  $\rightarrow$ obtain distribution of the $\hat{\beta}$'s

Randomization inference: a very simple simulated example

Use: `01a_rand_inf.R`

- Simulated data
- Which potential outcome do we know?
- Under strict null H_0, we actually know all!
- Generate 10,000 possible treatments
- Calculate ATE for each
- Look at the distribution of ATE

id	y	treat	y_i1	y_i0
1	8	0	NA	8
2	-3	1	-3	NA
3	-4	0	NA	-4
4	1	1	1	NA
5	23	0	NA	23
6	8	1	8	NA
7	13	1	13	NA
8	25	1	25	NA
9	12	1	12	NA
10	-4	0	NA	-4
11	-10	1	-10	NA
12	-11	0	NA	-11
13	-1	0	NA	-1
14	12	1	12	NA
15	11	1	11	NA

Randomization inference: a very simple simulated example

Use: `01a_rand_inf.R`

- Simulated data
- Which potential outcome do we know?
- Under strict null H_0, we actually know all!
- Generate 10,000 possible treatments
- Calculate ATE for each
- Look at the distribution of ATE

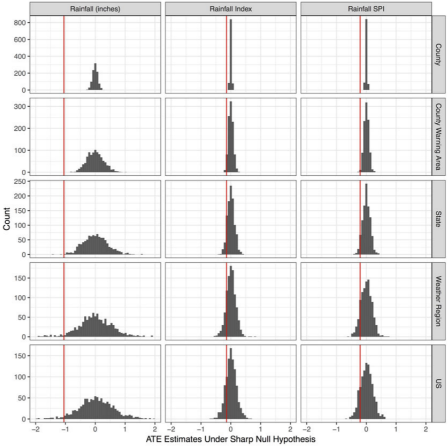
id	y	treat	y_i1	y_i0
1	8	0	8	8
2	-3	1	-3	-3
3	-4	0	-4	-4
4	1	1	1	1
5	23	0	23	23
6	8	1	8	8
7	13	1	13	13
8	25	1	25	25
9	12	1	12	12
10	-4	0	-4	-4
11	-10	1	-10	-10
12	-11	0	-11	-11
13	-1	0	-1	-1
14	12	1	12	12
15	11	1	11	11

Randomization inference: when to use RI?

- If conceptually, there is no true sampling variation to speak of (e.g., if we observe the universe of y outcomes, sampling-based p-values are meaningless)
- BUT not confined to large samples
  - RI doesn't appeal to asymptotic properties of an estimator, so can be used to make inferences about causal effects even with small data and/or few treated units
- Somewhat more robust to the presence of leverage in a few observations
  - Young 2018 reanalyzes 53 experimental articles from top econ journals → on average, randomization tests of the significance of TEs find 13% to 22% fewer significant results
- Can salvage inference with particular clustered designs (e.g., small number of assignment clusters or assignment clusters without well-defined boundaries)
  - Cooperman 2017 uses RI with rainfall data
- For more on accounting for both design-based and sampling-based uncertainty (including in observational settings), see Abadie et al. 2020

Randomization inference: when to use RI?

From Cooperman 2017:



Randomization inference: why not to use RI?

- Method for hypothesis testing, **not** for constructing CIs!

- Requires additional assumptions for constructing CIs
  - See Rosenbaum 2002, Gerber and Green 2012, and Barrios et al. 2012

Outline

Bootstrapping

- Hold on, part of the simulation approach to RI sounds a bit familiar... ? How is this different from bootstrapping?

- Bootstrapping *is* another **simulation** approach to inference, BUT it considers variation from sampling (like "traditional" sampling-based methods)
  - Bootstrap considers variation from sampling
    → resamples observations from our actual sample, with replacement
  - Simulates how *sampling* variation would affect results

- Can be very helpful when you want to do inference for something complicated when you're not sure how you should (correctly) calculate SEs
  - Simple to do in theory, though have to consider when naive bootstrap may distort correlated data (e.g., cluster bootstrap)
  - Big topic, but see Horowitz 2019 for review, see this tweet for Bayesian bootstrapping (helpful for complex models with many FEs, uses weights instead of random sampling), wild bootstrapping, etc...

Questions? Comments?

Thank you!

References

Heavily based on Claire Palandri's 2022 version of the Causal Inference Workshop.

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2020. "Sampling-Based versus Design-Based Uncertainty in Regression Analysis." *Econometrica* 88 (1): 265–296. https://doi.org/https://doi.org/10.3982/ECTA12675. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA12675. https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA12675.

Barrios, Thomas, Rebecca Diamond, Guido W. Imbens, and Michal Kolesar. 2012. "Clustering, Spatial Correlations, and Randomization Inference." *Journal of the American Statistical Association* 107 (498): 578–591. ISSN: 01621459, accessed March 28, 2024. http://www.jstor.org/stable/23239595.

Cooperman, Alicia Dailey. 2017. "Randomization Inference with Rainfall Data: Using Historical Weather Patterns for Variance Estimation." *Political Analysis* 25 (3): 277–288. ISSN: 10471987, 14764989, accessed March 28, 2024. https://www.jstor.org/stable/26563425.

Gerber, Alan S., and Donald P. Green. 2012. *Field experiments: design, analysis, and interpretation.* W.W. Norton & Company.

Horowitz, Joel L. 2019. "Bootstrap Methods in Econometrics." *Annual Review of Economics* 11 (Volume 11, 2019): 193–224. ISSN: 1941-1391. https://doi.org/https://doi.org/10.1146/annurev-economics-080218-025651. https://www.annualreviews.org/content/journals/10.1146/annurev-economics-080218-025651.

Rosenbaum, Paul. 2002. *Observational Studies.* Springer Series in Statistics. Springer Sceince & Business Media.

Young, Alwyn. 2018. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *The Quarterly Journal of Economics* 134 (2): 557–598. ISSN: 0033-5533. https://doi.org/10.1093/qje/qjy029. eprint: https://academic.oup.com/qje/article-pdf/134/2/557/28289325/qjy029.pdf. https://doi.org/10.1093/qje/qjy029.