Staggered Rollout Designs Enable Causal Inference Under Interference Without Network Knowledge

Changhao Shi

Renmin University of China, Statistics

November 8, 2023

- Cortez, M., Eichhorn, M. and Yu, C., 2022. Staggered rollout designs enable causal inference under interference without network knowledge. *Advances in Neural Information Processing Systems, 35*, pp.7437-7449.

Classic causal inference

- Binary treatment $Z_i \in \{0, 1\}$
- Potential outcomes $Y_i(1)$ and $Y_i(0)$
- Causal effects: comparisons of potential outcomes
- Common choice: average causal effect (ACE)

$$
\begin{aligned}
\mathrm{ACE} &\overset{\text{def}}{=} E\{Y(1) - Y(0)\} \\
&= E\{Y(1)\} - E\{Y(0)\} \\
&\overset{(1)}{=} E\{Y(1) \mid Z = 1\} - E\{Y(0) \mid Z = 0\} \\
&\overset{(2)}{=} E\{Y^{\text{obs}} \mid Z = 1\} - E\{Y^{\text{obs}} \mid Z = 0\}
\end{aligned}
$$

  - (1) holds when $\{Y(1), Y(0)\} \perp\!\!\!\perp Z$
  - (2) holds when $Y^{\text{obs}} = Y(1)Z + Y(0)(1 - Z)$
- For many policy makers, ACE is the quantity of interest

Causal inference under interference

- Violation of SUTVA
- Common in advertising, epidemiology and educational studies
- Potential outcomes $Y_i(z)$, where $z \in \{0,1\}^n$
- Causal effects of interest
  - total treatment effect (TTE)
    $$\mathrm{TTE} = \frac{1}{n}\sum_{i=1}^{n}\{Y(\mathbf{1}) - Y(\mathbf{0})\}$$
  - average direct effect (ADE)
    $$\mathrm{ADE} = \frac{1}{n}\sum_{i=1}^{n} E\big\{Y_i(z_i = 1, Z_{-i}) - Y_i(z_i = 0, Z_{-i})\big\}$$
  - average indirect effect (AIE)
    $$\mathrm{AIE} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} E\{Y_j(z_i = 1, Z_{-i}) - Y_j(z_i = 0, Z_{-i})\}$$
- Standard methods for ACE cannot be applied naively

General framework for interference

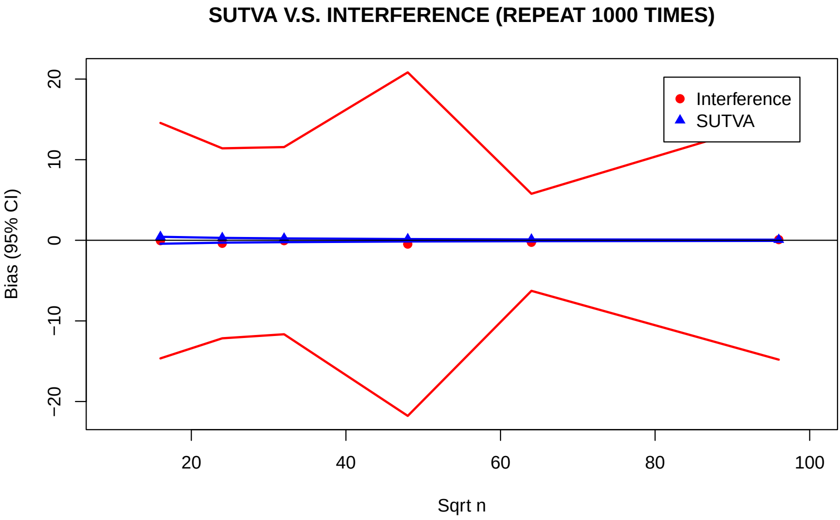
- A social network
  - through which individuals interfere each other
  - observable and correctly measured
- An exposure mapping
  - determines the extent and intensity of the interference
  - technically reduces the number of potential outcomes
  - canonical examples (minor notation abuse)
    - (no interference) $Y_i(z) = Y_i(z_i)$
    - (neighborhood interference) $Y_i(z) = Y_i(z_{\mathcal{N}_i})$
    - (arbitrary interference) $Y_i(z) = Y_i(z)$
    - ("individualized" interference) $Y_i(z) = Y_i(?)$
- Estimators: ht, hajek, difference-in-means, etc
- Experimental designs $Z \sim P(z)$: complete randomization, Bernoulli randomization, cluster randomization, etc

What's the difficulty? theoretically

- Explosion of the number of potential outcomes
  - unidentifiability
  - inconsistency
  - hard to interpret
  - problems in design and estimation

	Treatment					
Units	1	2	$\dots$	j	$\dots$	m
1	$Y_1(\mathbf{z_1})$	$Y_1(\mathbf{z_2})$	$\dots$	$Y_1(\mathbf{z}_j)$	$\dots$	$Y_1(\mathbf{z}_m)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
i	$Y_i(\mathbf{z_1})$	$Y_i(\mathbf{z_2})$	$\dots$	$Y_i(\mathbf{z}_j)$	$\dots$	$Y_i(\mathbf{z}_m)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
n	$Y_n(\mathbf{z_1})$	$Y_n(\mathbf{z_2})$	$\dots$	$Y_n(\mathbf{z}_j)$	$\dots$	$Y_n(\mathbf{z}_m)$

What's the difficulty? empirically



More difficult situation: unobservable networks

- Structure of social network may be unavailable, costly to collect or even “ill-defined” (say, time-varying network)
- Corte et al. (2022) says one can still get unbiased estimator for TTE and bound its variance under this situation: polynomial interpolation is all you need (and some additional assumptions, of course)
- More specifically, if you
  - get a sample including n individuals
  - care about the TTE of some policy
  - have no idea about the underlying social network

  then you can
  - implement staggered rollout design
  - calculate graph agnostic estimators
  - draw your conclusions

Notation and framework for unobservable networks

- An unknown directed graph with edge set $E \subset [n] \times [n]$
- An edge $(j, i) \in E$ means i is affected by j's treatment
- In-neighborhood of i: $\mathcal{N}_i = \{j \in [n] : (j, i) \in E\}$
- Potential outcomes function: $Y_i : \{0, 1\}^n \to \mathbb{R}$
- Under assumption of consistency, one may see

$$Y_i(\mathbf{z}) = \sum_{\mathcal{S} \subseteq [n]} a_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j \prod_{j' \in [n] \setminus \mathcal{S}} (1 - z_{j'}) = \sum_{\mathcal{S} \subseteq [n]} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j \quad (1)$$

- Equation (1) means $Y_i(z)$ is a polynomial in z of degree at most n
- Estimand of interest: $\mathrm{TTE} := \frac{1}{n} \sum_{i=1}^{n} (Y_i(\mathbf{1}) - Y_i(\mathbf{0}))$

Assumptions

- (Neighborhood Interference) $Y_i(\mathbf{z})$ only depends on the treatment of individuals in $\mathcal{N}_i$ (including i). Equivalently, $Y_i(\mathbf{z}) = Y_i(\mathbf{z}')$ for any $\mathbf{z}$ and $\mathbf{z}'$ such that $\mathbf{z}_j = \mathbf{z}'_j$ for all $j \in \mathcal{N}_i$.
- (Bounded Potential Outcomes)

$$Y_{\max} := \max_{i \in [n]} \textstyle\sum_{\mathcal{S} \subseteq \mathcal{N}_i, |\mathcal{S}| \leq \beta} |c_{i,\mathcal{S}}|.$$

- (Low Polynomial Degree) The potential outcomes model has polynomial degree at most β, i.e. there exist coefficients $\{c_{i,\mathcal{S}}\}_{i \in [n], \mathcal{S} \subseteq [n]}$ such that for all i and $\mathbf{z}$,

$$Y_i(\mathbf{z}) = \textstyle\sum_{\mathcal{S} \subseteq \mathcal{N}_i, |\mathcal{S}| \leq \beta} c_{i,\mathcal{S}} \cdot \mathrm{I}\big(\mathcal{S} \text{ treated}\big) = \sum_{\mathcal{S} \subseteq \mathcal{N}_i, |\mathcal{S}| \leq \beta} c_{i,\mathcal{S}} \prod_{j \in \mathcal{S}} z_j.$$

- ("Time-Invariant" Potential Outcomes)

$$Y_{i,t}^{\mathbf{obs}} = Y_i(\mathbf{z}^t) + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \overset{iid}{\sim} N(0, \sigma^2).$$

Staggered Rollout Design

- Treatment is incrementally given to random subsets of individuals
  - treatment is assigned to individuals in T stages
  - individuals' outcomes are measured $T+1$ times
    - a baseline measurement before treatment
    - a measurement after each treatment round
- Treatment assignment in round t: $\mathbf{z}^t$
  - each entry z_i^t is monotone increasing with t
- Staggered rollout bernoulli design (BRD($\mathbf{p}$))
  - cumulative treatment probabilities $0 < p_1 < \cdots < p_T \leq b \ll 1$
  - $u_i \overset{iid}{\sim} U(0,1)$, for each $i \in [n]$
  - $z_i^t = \mathbb{1}(u_i \leq p_t)$, for each $t \in [T]$
- Staggered rollout completely randomized design (CRD($\mathbf{k}$))
  - cumulative treatment numbers $0 = k_0 < k_1 < \cdots < k_T \leq c \ll n$
  - $\mathbf{z}^t \sim \text{CRD}(k_t - k_{t-1})$ out of the remaining untreated individuals

Graph Agnostic Estimators

- (Lagrange Interpolation) Given a dataset $\{(x_t, y_t)\}_{t=0}^{T}$ with distinct x-coordinates, the unique polynomial F of degree at most T with $F(x_t) = y_t$ for each t is given by

$$F(x) = \sum_{t=0}^{T} \ell_{t,\mathbf{x}}(x) \cdot y_t, \qquad \ell_{t,\mathbf{x}}(x) = \prod_{\substack{s=0 \\ s \neq t}}^{T} \frac{x - x_s}{x_t - x_s}.$$

- Polynomial interpolation (PI) estimator

$$\widehat{\mathrm{TTE}}_{\mathrm{PI}}(\mathbf{x}) := \begin{cases} \sum_{t=0}^{T} \left(\ell_{t,\mathbf{x}}(1) - \ell_{t,\mathbf{x}}(0)\right) \left(\frac{1}{n} \sum_{i=1}^{n} Y_{i,t}^{\mathbf{obs}}\right) & x_0 < x_1 < \ldots < x_T, \\ 0 & x_t = x_{t-1} \text{ for some } t. \end{cases}$$

  - i.e. implement PI for $\{(x_t, \overline{y_t^{\mathrm{obs}}})\}_{t=0}^{T}$, where $\overline{y_t^{\mathrm{obs}}} = \frac{1}{n} \sum_{i=1}^{n} y_{i,t}^{\mathrm{obs}}$

Theoretical Results

- (**Theorem 1**) Consider a potential outcomes model with degree β. Under a BRD(p) with $p_0 = 0$, the estimator $\widehat{\mathrm{TTE}}_{\mathrm{PI}}(\mathbf{p})$ is unbiased with variance

$$O\Big(\beta^2 Y_{\max}^2 \frac{d^2}{n} \Delta_{\mathbf{p}}^{-2\beta} + \frac{\sigma^2 \beta}{n} \Delta_{\mathbf{p}}^{-2\beta}\Big).$$

- (**Theorem 2**) Consider a potential outcomes model with degree β. Under a CRD(k) with $k_0 = 0$, the estimator $\widehat{\mathrm{TTE}}_{\mathrm{PI}}(\mathbf{k}/n)$ is unbiased with variance

$$O\Big(\beta^2 Y_{\max}^2 \Big(\frac{d^2}{n} + \frac{\beta^2}{k_1}\Big) \cdot \Big(\frac{n}{\Delta_{\mathbf{k}}}\Big)^{2\beta} + \frac{\sigma^2 \beta}{n} \Big(\frac{n}{\Delta_{\mathbf{k}}}\Big)^{2\beta}\Big).$$

Intuition-Why does it work?

- In general, when $F(x)$ is a polynomial in x of degree T
  - $F(x) = a_T x^T + \cdots + a_1 x + a_0$ (definition)
    - equivalently, $F(x) = \sum_{t=0}^{T} \ell_{t,\mathrm{x}}(x) F(x_t)$ (linear w.r.t $F(x_t)$)
  - $\widehat{F}(x) = \sum_{t=0}^{T} \ell_{t,\mathrm{x}}(x) \widehat{F}(x_t)$
    - $E(\widehat{F}(x)) = F(x)$ if $E(\widehat{F}(x_t)) = F(x_t)$
    - $\ell_{t,\mathrm{x}}(x)$ is nonrandom
- Assume $Z \sim \mathcal{D}_x$, where $\mathcal{D}_x$ is a parameterized class of distributions
  - let $P_{\mathbf{Z} \sim \mathcal{D}_0}(\mathbf{Z} = \mathbf{0}) = 1$ and $P_{\mathbf{Z} \sim \mathcal{D}_1}(\mathbf{Z} = \mathbf{1}) = 1$
  - define $F_{\mathcal{D}}(x) = \mathbb{E}_{\mathbf{Z} \sim \mathcal{D}_x} \left[\frac{1}{n} \sum_{i=1}^{n} Y_i(\mathbf{Z}) \right]$
  - then $\mathrm{TTE} = F_{\mathcal{D}}(1) - F_{\mathcal{D}}(0)$
  - under "suitable" designs, $F_{\mathcal{D}}(x)$ will be a polynomial in x
    - (BRD($\mathbf{p}$)) $F_B(p) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{S} \subseteq \mathcal{N}_i, |\mathcal{S}| \leq \beta} c_{i,\mathcal{S}} \cdot p^{|\mathcal{S}|}$
    - (CRD($\mathbf{k}$)) $F_C(\frac{k}{n}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{\mathcal{S} \subseteq \mathcal{N}_i, |\mathcal{S}| \leq \beta} c_{i,\mathcal{S}} \cdot \left[\frac{k}{n} \right]^{|\mathcal{S}|}$

Simulation Settings

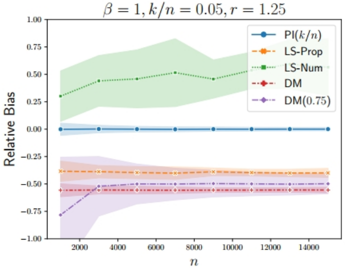
- Random networks generated from a configuration model (say, SBM, SW, etc) with degrees distributed as a power law with exponent 2.5
- For degree β, consider potential outcomes model

$$Y_i(\mathbf{z}) = c_{i,\emptyset} + \sum_{j \in \mathcal{N}_i} \tilde{c}_{ij} z_j + \sum_{\ell=2}^{\beta} \left(\frac{\sum_{j \in \mathcal{N}_i} \tilde{c}_{ij} z_j}{\sum_{j \in \mathcal{N}_i} \tilde{c}_{ij}} \right)^{\ell},$$

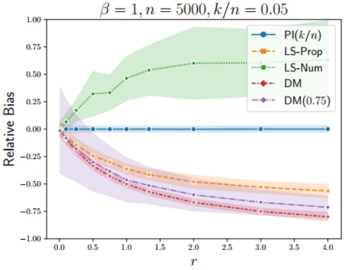
  where

  - $c_{i,\emptyset} \sim U[0,1]$, $\tilde{c}_{ii} \sim U[0,1]$
  - for $i \neq j$, $\tilde{c}_{ij} = v_j |\mathcal{N}_i| / \sum_{k:(k,j) \in E} |\mathcal{N}_k|$ for $v_j \sim U[0, r]$
  - hyperparameter r governs the relative magnitude of the network effects
- Set $\sigma = 0$, which is in $\varepsilon_{i,t} \overset{iid}{\sim} N(0, \sigma^2)$

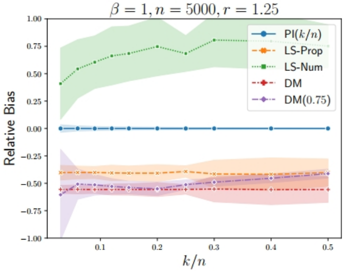
Simulation results



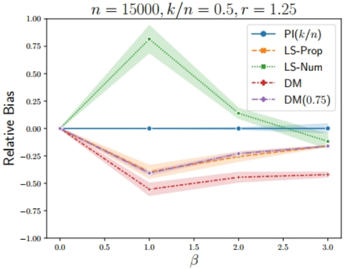
(a) Varying size of the population



(b) Varying ratio of direct:indirect effects



(c) Varying treatment budget



(d) Varying the model degree

Discussion

- A novel idea
  - staggered rollout design
  - polynomial interpolation
- Future work
  - model selection when β is unknown
  - dynamic setting
    - time-dependent noise
    - time-varying effects
    - time-varying networks
  - beyond polynomial
    - sublinear functions
    - monotone functions

Thank you!