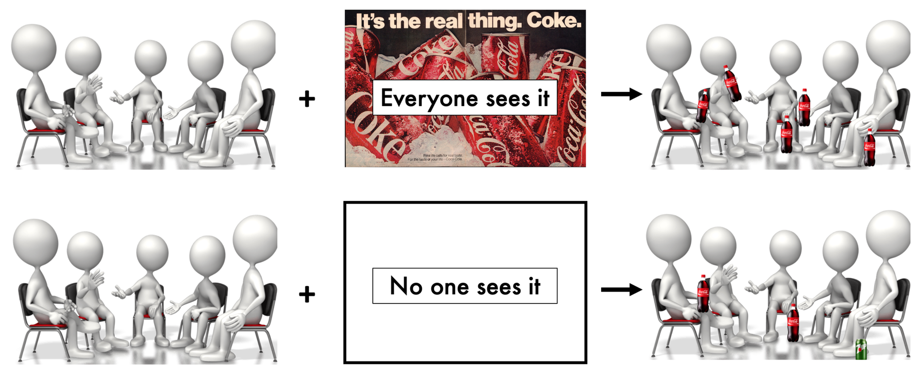
Exploiting Neighborhood Interference with Low Order Interactions under Unit Randomized Design

Mayleen Cortez-Rodriguez [1] Matthew Eichhorn [1] Christina Lee Yu [2]

[1]Center for Applied Mathematics [2]Operations Research and Information Engineering

Motivating Example

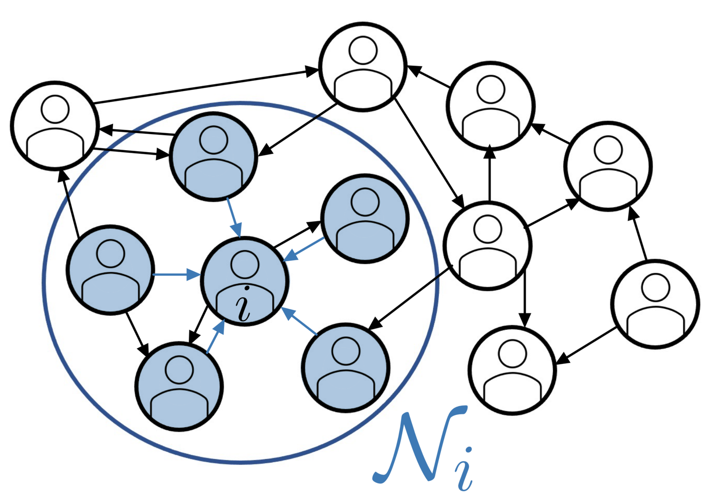
- You're deciding whether to roll out an advertising campaign
- The **Total Treatment Effect (TTE)** measures the average change in customer behavior (sales revenue) with the campaign versus without it



- The company runs a randomized experiment to estimate TTE
- **Network Interference:** Word-of-mouth spreads advertiser's message beyond direct viewers
- Traditional estimators rely on SUTVA, so are biased
- Inverse probability estimators can have high variance

Formalizing the Problem

- **Population:** n individuals
- **Network Effects:** Edge (j,i) if j's treatment affects i's outcome
  * Network structure is known *
- **Treatment:** Indicated by $\mathbf{z} \in \{0,1\}^n$
- **Outcomes:** $Y_i \colon \mathbf{z} \to \mathbb{R} \quad \forall i$



$$\text{TTE} = \frac{1}{n}\sum_{i=1}^{n}\Big(Y_i(\mathbf{1}) - Y_i(\mathbf{0})\Big)$$

Assumptions

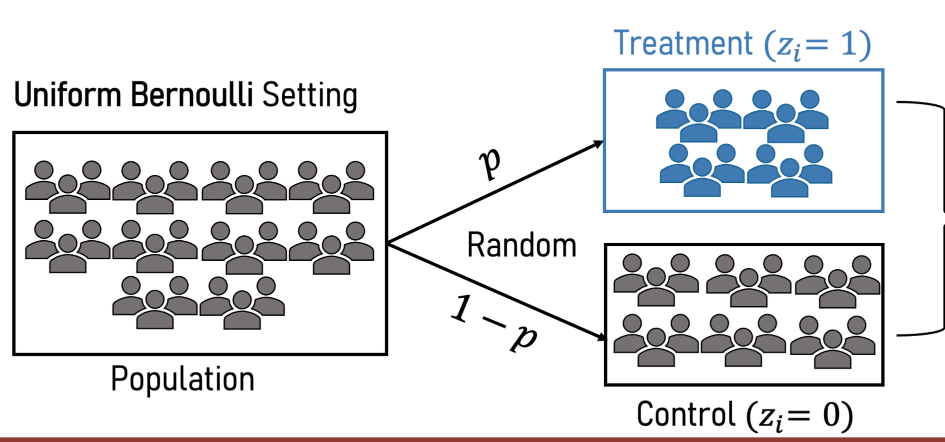
1. **Neighborhood Interference:** Y_i depends only on the treatment assignments of their in-neighbors $\{z_j\}_{j\in\mathcal{N}_i}$
2. **β-Order Interactions:** Only small subsets of *treated* neighbors affect Y_i

$$Y_i(\mathbf{z}) = \sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i\\ |\mathcal{S}|\le\beta}} c_{i,\mathcal{S}} \prod_{j\in\mathcal{S}} z_j$$

3. **Bounded Effects:** For each i, $\sum_{\mathcal{S}} |c_{i,\mathcal{S}}| = \mathcal{O}(1)$
4. **Known Network Structure:** We have knowledge of each $\mathcal{N}_i$

Bernoulli Randomized Design

Treatments sampled independently: $z_i \sim \text{Bernoulli}(p)$ with $p \in (0,1)$



Research Question

Can we design an unbiased TTE estimator under Bernoulli randomized design with a reasonable bound on its variance?

Framing Our Estimator:

We instead estimate the effect coefficient vector $\mathbf{c}_i$ for each i

Suppose we could replicate the randomized experiment R times:

$$\underbrace{\begin{bmatrix} Y_i^1(\mathbf{z}^1)\\ Y_i^2(\mathbf{z}^2)\\ \vdots \\ Y_i^R(\mathbf{z}^R)\end{bmatrix}}_{\mathbf{Y}_i} = \underbrace{\begin{bmatrix} \tilde{\mathbf{z}}_1^1 & \tilde{\mathbf{z}}_2^1 & \dots & \tilde{\mathbf{z}}_k^1\\ \tilde{\mathbf{z}}_1^2 & \tilde{\mathbf{z}}_2^2 & \dots & \tilde{\mathbf{z}}_k^2\\ \vdots & \vdots & \ddots & \vdots\\ \tilde{\mathbf{z}}_1^R & \tilde{\mathbf{z}}_2^R & \dots & \tilde{\mathbf{z}}_k^R\end{bmatrix}}_{\tilde{\mathbf{Z}}_i} \underbrace{\begin{bmatrix} c_{i,\mathcal{S}_1}\\ c_{i,\mathcal{S}_2}\\ \dots\\ c_{i,\mathcal{S}_k}\end{bmatrix}}_{\mathbf{c}_i} \qquad \tilde{\mathbf{z}}_\ell^r = \prod_{j\in\mathcal{S}_\ell} \mathbf{z}_j^r$$

If we left-multiply by $\frac{1}{R}\tilde{\mathbf{Z}}_i^{\intercal}$ and let $R\to\infty$, the LLN gives

$$\mathbb{E}[Y_i(\mathbf{z})\tilde{\mathbf{z}}] = \mathbb{E}[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^{\intercal}]\mathbf{c}_i \quad \implies \quad \mathbf{c}_i = \mathbb{E}[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^{\intercal}]^{-1}\,\mathbb{E}[Y_i(\mathbf{z})\tilde{\mathbf{z}}]$$

The SNIPE Estimator

Given realization $(\mathbf{z},\mathbf{Y})$ of our experiment, produce unbiased estimates

$$\hat{\mathbf{c}}_i = Y_i(\mathbf{z})\mathbb{E}[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^{\intercal}]^{-1}\tilde{\mathbf{z}}$$

Extend by linearity to **unbiased** estimator

$$\widehat{TTE} = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i\\ 1\le|\mathcal{S}|\le\beta}} \hat{c}_{i,\mathcal{S}}$$

$$\widehat{TTE} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}) \sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i\\ 1\le|\mathcal{S}|\le\beta}} \left(\left(\tfrac{1-p}{p}\right)^{|\mathcal{S}|} - (-1)^{|\mathcal{S}|}\right) \prod_{j\in\mathcal{S}} \tfrac{z_j - p}{1-p}$$

Special case of the *psuedoinverse estimator* of [3] and the *Riesz estimator* of [2]

Analyzing the Variance

A careful analysis shows that

$$\text{Var}\Big(\widehat{\text{TTE}}\Big) = \mathcal{O}\left(\frac{d^2}{n}\left(\frac{ed}{\beta}\cdot\max\left\{4\beta^2, \frac{1}{p(1-p)}\right\}\right)^{\beta}\right)$$

Compare to the Horvitz-Thompson estimator:

$$\widehat{\text{TTE}}_{\text{HT}} = \frac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z})\left(\prod_{j\in\mathcal{N}_i}\frac{z_j}{p} - \prod_{j\in\mathcal{N}_i}\frac{1-z_j}{1-p}\right)$$

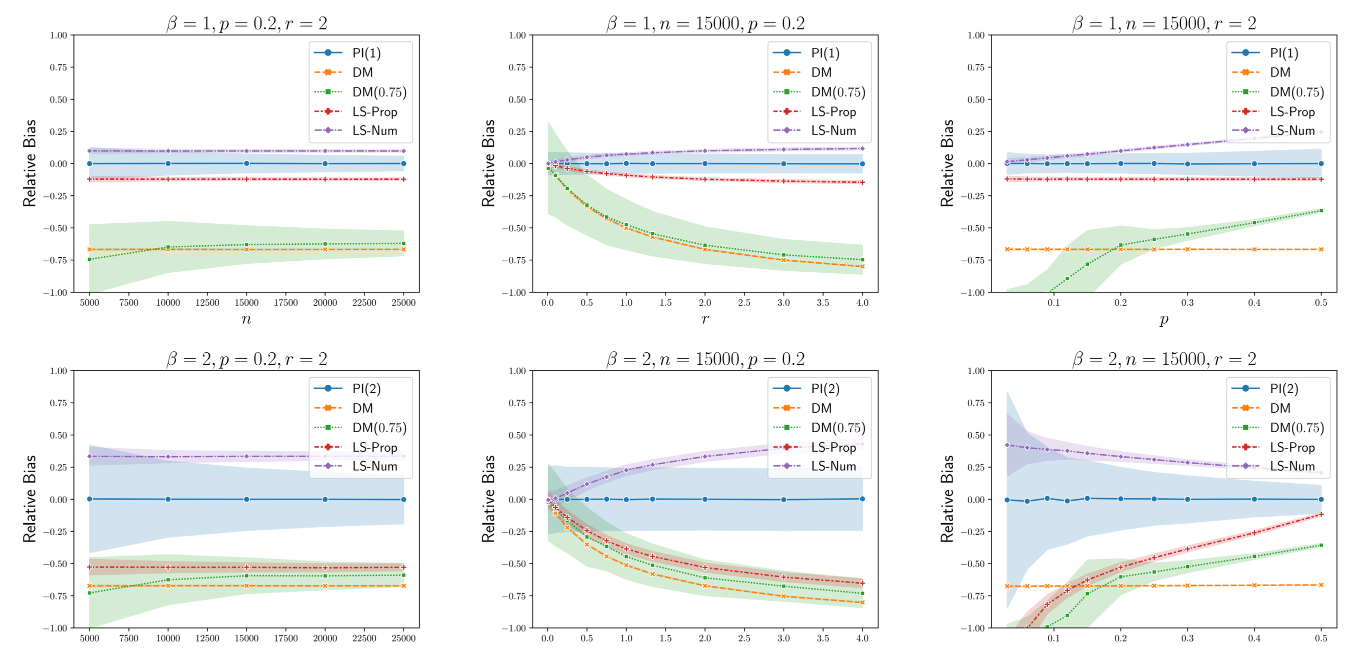
The variance of $\widehat{\text{TTE}}_{\text{HT}}$ scales as $\Theta\left(\frac{1}{p^d}\right)$ scales as [4]

The variance of SNIPE scales polynomially in d and exponentially in β, a clear improvement when $\beta \ll d$.

A minimax lower bound analysis of the MSE with Le Cam's method shows that variance $\Omega\left(\frac{1}{np^{\beta}}\right)$ is unavoidable.

Experiments

- Erdős-Rényi network of n nodes with edge probability $p_{\text{edge}} = 10/n$
- Parameter r governs the strength of interference effects
- Parameter p is the treatment budget
- Compare against difference-in-means (DM) and adjusted least-squares (LS) estimators
- **Observation**: Under a β-order outcomes model, our estimator PI(β) generally outperforms other estimators w.r.t. MSE



(a) Varying population size (b) Varying interference level (c) Varying treatment budget

Related/Ongoing Work

Unknown Networks

- The SNIPE estimator utilizes knowledge of the graph structure
- If the network is not fully known, we can use multiple rounds of experimentation to estimate TTE

Other Experimental Designs

- The framing of our estimator is not unique to Bernoulli design
- Can we select treatment assignments in a smarter way (e.g. clustering) to further reduce variance?

Model Misspecification

- It may be unreasonable to assume that β is known to the practitioner
- How can we quantify and/or minimize additional bias/variance brought about by estimating with a misspecified model?

References

[1] Mayleen Cortez, Matthew Eichhorn, and Christina Lee Yu. Exploiting neighborhood interference with low order interactions under unit randomized design. *arXiv preprint arXiv:2208.05553*, 2022.

[2] Christopher Harshaw, Fredrik Sävje, and Yitan Wang. A design-based riesz representation framework for randomized experiments. *arXiv preprint arXiv:2210.08698*, 2022.

[3] Adith Swaminathan, Akshay Krishnamurthy, Alekh Agarwal, Miro Dudik, John Langford, Damien Jose, and Imed Zitouni. Off-policy evaluation for slate recommendation. *Advances in Neural Information Processing Systems*, 30, 2017.

[4] Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337.