Discussion of "Policy design in experiments with unknown interference" by Viviano

Timothy B. Armstrong

University of Southern California

July 2022

Nice paper! Draws on several literatures:

- Networks
- Experimental design/bandits
- Statistical treatment rules
- Diff-in-diff
- Cross-validation/sample splitting

Lots of interesting stuff here. I'll focus on one topic:

- The analysis in this paper appears to require assumptions on latent variables in a network formation model.
- But we have the ability to experiment.
- Can we guarantee (internal) validity of the procedures in this paper solely through experimental design?

Short answer: yes (in static case), by randomizing policies at cluster/network level

- Leads to similar experimental design and estimator to the paper (in static case)
- Can think of this as a reinterpretation of the identification argument in the paper

Data generating process in paper: for each network/cluster k we

1. Draw individuals $i = 1, \ldots, N$ from a (hyper)population (formally, we draw iid latent variables $U_i^{(k)}, \nu_i^{(k)}$, etc.)
2. Assign treatments to each individual, outcomes are determined by treatments, network and network formation game
3. Sample outcomes for n out of N individuals

Some critiques:

- Conceptual: what is the (hyper)population in step 1?
  - See Abadie, Athey, Imbens and Wooldridge (2020) for discussion in related setting.
- Plausibility of modeling assumptions (network formation game, model for cluster and time effects, etc.)

Alternative *design-based* approach:

- Assume only the existence of potential outcomes $Y_i^{(k)}(d) = Y_i^{(k)}(d_1, \ldots, d_{N_k})$ for individual i in cluster k under treatments $d_1, \ldots, d_{N_k}$.
- Following the paper, we are interested in an aggregate welfare measure $\bar{Y}^{(k)}(d)$ (can be $\sum_{i=1}^{N_k} Y_i^{(k)}(d)$, or something else)
- Stable unit treatment value assumption (SUTVA) only *across networks* (same as in paper).
- Clusters/networks can also have covariates $X^{(k)} = (X_1^{(k)}, \ldots, X_{N_k}^{(k)})$
- Treat covariates and potential outcomes as fixed: no need for hyperpopulation

As in the paper, define a *policy*, indexed by β, as a rule mapping covariates for cluster k to a probability distribution for treatments.

- For simplicity, consider a deterministic policy $d^{(k)}(\beta; X^{(k)})$
- Example: "treat only the first $\beta \cdot N_k$ units $i = 1, \ldots, \beta \cdot N_k$"
- Let $W_k(\beta) = \bar{Y}^{(k)}(d^{(k)}(\beta; X^{(k)}))$ denote aggregate welfare under the policy β in network/cluster k.
- Average welfare across clusters $k = 1, \ldots, K$:

$$W(\beta) = \frac{1}{K} \sum_{i=1}^{K} W_k(\beta).$$

- Welfare effect of marginal change in β:

$$V(\beta, \eta) = \frac{W(\beta + \eta) - W(\beta - \eta)}{2\eta}$$

Basic idea: we have the usual potential outcomes setting at the network/cluster level!

- $W_k(\beta)$ is the potential outcome for network/cluster k under "treatment" β
- $W(\beta)$ is the average structural/dose-response function
- $W(\beta + \eta) - W(\beta - \eta)$ is the average treatment effect for moving from $\beta - \eta$ to $\beta + \eta$.
- If we assign network/cluster k to treatment β, we get to observe $W_k^{\text{obs}} = W_k(\beta)$. sampling individuals/randomized policies

Unbiased estimate for $V(\beta, \eta) = \frac{W(\beta+\eta)-W(\beta-\eta)}{2\eta}$:

1. Draw subsets $\mathcal{K}_0$ and $\mathcal{K}_1$ of $\{1, \ldots, K\}$ "at random."
   - Assign clusters $k \in \mathcal{K}_0$ to policy $\beta - \eta$.
   - Assign clusters $k \in \mathcal{K}_1$ to policy $\beta + \eta$.
2. Use the difference-in-means estimator

$$\hat{V} = \frac{1}{2\eta} \left[\frac{1}{\#\mathcal{K}_1} \sum_{k \in \mathcal{K}_1} W_k^{\text{obs}} - \frac{1}{\#\mathcal{K}_0} \sum_{k \in \mathcal{K}_0} W_k^{\text{obs}} \right]$$

If we use matched pairs to randomize in step 1, then the estimator in step 2 is *the same as the one proposed in the paper* (in the static case), with two modifications:

- The estimator in the paper has a first differencing step
- In the paper, there is no explicit randomization at the network/cluster level

Interpretation: the modeling assumptions in the paper guarantee that assignment of network/cluster k to $\beta - \eta$ vs $\beta + \eta$ is "as good as random," after first-differencing.

- The above analysis shows that we can alternatively guarantee this by randomizing k to policy $\beta - \eta$ vs $\beta + \eta$.

Some other things we can do with design-based approach:

- Externally valid estimates for a larger population of clusters/networks, if we can sample the K clusters/networks randomly from this population
- Test the "strong null" that $W_k(\beta)$ doesn't depend on β using randomization test
  - This turns out to be equivalent to extension mentioned in footnote: using Canay, Romano and Shaikh (2017) instead of Ibragimov and Müller (2010) for inference
- Use period 0 outcomes for regression adjustment
  - Special case of Roth and Sant'Anna (2022): leads to weighted average of diff-in-diff and diff-in-means

What do the hyperpopulation and modeling assumptions in the paper buy us?

- Normality result for aggregate outcomes, guarantees on bias when η is small $\rightarrow$ inference on marginal effects
- Concentration bounds based on γ_N (could be useful for statistical power analysis)
- Framework for repeated sampling from the same network/cluster
- Many other results in the paper...

Useful to clarify when it suffices to assume (as good as) random assignment at network/cluster level, and when we need the network model.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. "Sampling-Based versus Design-Based Uncertainty in Regression Analysis." Econometrica 88, no. 1 (2020): 265–96.

Canay, Ivan A., Joseph P. Romano, and Azeem M. Shaikh. "Randomization Tests Under an Approximate Symmetry Assumption." Econometrica 85, no. 3 (2017): 1013–30.

Ibragimov, Rustam, and Ulrich K. Müller. "T-Statistic Based Correlation and Heterogeneity Robust Inference." Journal of Business Economic Statistics 28, no. 4 (October 1, 2010): 453–68.

Roth, Jonathan, and Pedro H. C. Sant'Anna. "Efficient Estimation for Staggered Rollout Designs." arXiv, January 12, 2022.

Extra slides: randomized policies and samples from clusters

To incorporate randomized treatment into the policy (as in the paper), we define a policy indexed by β as a rule mapping covariates for cluster k to a probability distribution for treatments.

- Formally, we define a probability distribution $\Pi(X^{(k)}; \beta)$ for $d_1, \ldots, d_{N_k}$ for network/cluster k with covariates $X^{(k)} = (X_1^{(k)}, \ldots, X_{N_k}^{(k)})$.
- Treat covariates $X^{(k)}$ as deterministic.
- Example: with no covariates, set $d_i \sim$ Bernoulli(β) to treat proportion β of individuals at random
- Define $W_k(\beta)$ as average welfare under policy β:

$$W_k(\beta) = E_{d \sim \Pi(X^{(k)}, \beta)} \bar{Y}^{(k)}(d)$$

Policy/treatment assignment at network/cluster level:

- If we assign network/cluster k to treatment β, we observe one draw $W_k^{\text{obs}} = \bar{Y}^{(k)}(d)$ with $d \sim \Pi(X^{(k)}, \beta)$.
- Suffices to randomly sample n_k out of N_k individuals and form W_k^{obs} as an unbiased estimate of $\bar{Y}^{(k)}(d)$
- Then $EW_k^{\text{obs}} = W_k(\beta)$
- This, along with the usual arguments, shows that diff-in-means estimate after randomization is unbiased.
- However, need to be careful when defining "strong null" for testing no policy effect. back