Average direct and indirect causal effects under interference

Authors: Yuchen Hu, Shuangning Li And Stefan Wager

Speaker: Changhao Shi

June 12, 2023

Outline

Outline

Neyman-Rubin Framework

- The classical way of analysing randomized trials, following Neyman (1923) and Rubin (1974), centres on the average treatment effect (ATE) defined using potential outcomes.
- Given a sample of $i = 1, \ldots, n$ units used to study the effect of a binary treatment $W_i \in \{0, 1\}$, we posit potential outcomes $Y_i(0), Y_i(1) \in \mathbb{R}$ corresponding to the outcomes we **would have measured** had we assigned the ith unit to control and to treatment, respectively, i.e., we observe $Y_i = Y_i(W_i)$.
- We then proceed by arguing that the sample average treatment effect $\tau_{\mathrm{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$ admits a simple unbiased estimator under random assignment of treatment.

Causal Inference under Interference

- One limitation of this classical approach is that it rules out interference and instead introduces an assumption that the observed outcome for any given unit does not depend on the treatments assigned to other units, i.e., Y_i is not affected by W_j for any $j \neq i$.
- However, in a wide variety of applied settings, such interference effects not only exist, but are often of considerable scientific interest (Sacerdote, 2001; Miguel & Kremer, 2004; Bakshy et al., 2012; Bond et al., 2012; Cai et al., 2015; Rogers & Feller, 2018).
- For example, in an education setting, it may be of interest to understand how a pedagogical innovation affects not just certain targeted students, but also their peers.

Challenges

A major difficulty in working under interference is that one no longer has a single obvious average effect parameter to target like ATE. In the general setting, each unit now has 2^n potential outcomes corresponding to every possible treatment combination assigned to the n units.

	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	$Y_n(\mathbf{z_1})$	$Y_n(\mathbf{z_2})$	$\dots$	$Y_n(\mathbf{z}_j)$	$\dots$	$Y_n(\mathbf{z}_m)$

图: Potential outcomes under interference.

Motivations

The existing literature has mostly side-stepped this issue by framing the estimand in terms of specific policy interventions. However, this paradigm does not provide researchers with simple, nonparametric and agnostic average causal estimands that can be studied without spelling out a specific policy intervention of interest.

Main Results

- In this paper authors study a pair of averaging causal estimands, the average direct and indirect effects, that are valid under interference and yet, unlike existing targets, can be defined and estimated using a single experiment and do not need to be defined in terms of hypothetical policy interventions.
- Qualitatively, the average direct effect measures the extent to which, in a given experiment and on average, the outcome Y_i of a unit is affected by its own treatment W_i; meanwhile, the average indirect effect measures the responsiveness of Y_i to treatments W_j given to other units $j \neq i$.

Outline

Notations and Assumptions

For convenience, we use the shorthand $Y_i(w_j = x; W_{-j})$ to denote the potential outcome we would observe for the ith unit if we were to assign the jth unit to treatment status $x \in \{0, 1\}$ and maintain all units, but the jth at their realized treatments $W_{-j} \in \{0, 1\}^{n-1}$. Expectations E are over the treatment assignment only; potential outcomes are held fixed.

Assumption 1

For units $i = 1, \ldots, n$, there are potential outcomes $Y_i(w) \in \mathbb{R}$, $w \in \{0, 1\}^n$, such that given a treatment vector $W \in \{0, 1\}^n$ we observe outcomes $Y_i = Y_i(W)$.

Definitions

Definition 1

Under Assumption 1, the average direct effect of a binary treatment is

$$\tau_{\mathrm{ADE}} = \frac{1}{n}\sum_{i=1}^{n} E\big\{Y_i(w_i = 1; W_{-i}) - Y_i(w_i = 0; W_{-i})\big\}.$$

Definition 2

Under Assumption 1, the average indirect effect of a binary treatment is

$$\tau_{\mathrm{AIE}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} E\{Y_j(w_i = 1; W_{-i}) - Y_j(w_i = 0; W_{-i})\}.$$

Interpretations

- The definition of the direct effect τ_{ADE} is standard. In a study without interference, τ_{ADE} matches the usual ATE.
- The definition of the indirect effect is an immediate formal generalization of τ_{ADE} to cross-unit treatment effects. In the no-interference case we clearly have $\tau_{AIE} = 0$.

Definitions

Definition 3

Under Assumption 1, the average overall effect of a binary treatment is

$$\tau_{A0E} = \tau_{\mathrm{ADE}} + \tau_{\mathrm{AIE}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\big\{ Y_j(w_i = 1; W_{-i}) - Y_j(w_i = 0; W_{-i})\big\}.$$

Definition 4

Under Assumption 1 and in a Bernoulli design, the infinitesimal policy effect is

$$\tau_{INT}(\pi) = 1 \cdot \nabla_{\pi} E_{\pi}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \sum_{k=1}^{n} \frac{\partial}{\partial \pi_k} E_{\pi}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right).$$

Interpretations

Theorem 1

Under Assumption 1 and in a Bernoulli design, $\tau_{AOE}(\pi) = \tau_{INF}(\pi)$.

Theorem 1 provides an alternative lens on our definition of the indirect effect. Suppose, for example, that a researcher knew they wanted to study nudge interventions, the total effect of which is $\tau_{INF}(\pi)$, and was also committed to the standard definition of the average direct effect given in Definition 1. Then it would be natural to define an indirect effect as $\tau_{INF}(\pi) - \tau_{ADE}(\pi)$, i.e., to characterize as indirect any effect of the nudge intervention that is not captured by the direct effect; this is, for example, the approach implicitly taken in Heckman et al. (1998).

Outline

The Direct Effects

Hudgens & Halloran (2008) proposed a definition that relies on conditional expectations:

$$\tau_{\mathrm{HH,DE}} = \frac{1}{n}\sum_{i=1}^{n}\{E(Y_i \mid W_i = 1) - E(Y_i \mid W_i = 0)\}.$$

- In a Bernoulli design, $\tau_{HH,DE} = \tau_{ADE}$.
- In other designs, such as completely randomized designs or stratified designs, $\tau_{HH,DE}$ and τ_{ADE} do not match.
- A major drawback of $\tau_{HH,DE}$ is that it conflates the effect of setting $w_i = x$ on the ith unit's outcome and the effect of setting $w_i = x$ on the distribution of W_{-i}.

The Indirect Effects -1

In the case of comparing two Bernoulli trials with randomization probabilities π and π', Hudgens & Halloran (2008) defined

$$\tau_{\mathrm{IE}}(\pi, \pi') = \frac{1}{n} \sum_{i=1}^{n} \big[E_{\pi}\{ Y_i(w_i = 0; W_{-i})\} - E_{\pi'}\{ Y_i(w_i = 0; W_{-i})\} \big].$$

- It's a interesting quantity to consider if we can run many independent experiments that test different overall treatment levels.
- It does not enable us to describe indirect effects in a single randomized study.

The Indirect Effects -2

Assume existence of functions $h_i : \{0,1\}^n \to \{1,\dots,K\}$ such that potential outcomes $Y_i(w)$ depend on w only via the compressed representation $h_i(w)$, i.e., $Y_i(w) = Y_i(w')$ whenever $h_i(w) = h_i(w')$. Then we can define

$$\mu(k) = \frac{1}{n}\sum_{i=1}^{n} E\{Y_i \mid h_i(W') = k\}, \quad \tau(k,k') = \mu(k') - \mu(k), \quad 1 \leq k \neq k' \leq K.$$

- Again, it is conceptually attractive and sometimes enable us to very clearly express the answer to a natural policy question.
- Again, it can not describe indirect effects in a single randomized study.
- Not widely to use as the number K of possible exposure types gets large.

The Indirect Effects -3

Aronow et al. (2021) considered a setting where treatments are assigned to points in a geographic space and sought to estimate the average effect of treatment at an intervention point on outcomes at points that are a distance d away,

$$\tau_{\mathrm{AMR}}(d;\pi) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in S_i(d)} \frac{E_\pi\{Y_j(w_i=1;W_{-j}) - Y_j(w_i=0;W_{-j})\}}{|\mathcal{S}_i(d)|},$$

where $\mathcal{S}_i(d) = \{j : \Delta(i,j) = d\}$, $\Delta(i,j)$ measures the distance between points i and j.

- Similar to τ_{AIE}, but the normalization factor $|\mathcal{S}_i(d)|^{-1}$ used in τ_{AMR} would invalidate Theorem 1.

Outline

Parametric Models

The purpose of this section is to examine our abstract, nonparametric definition of the indirect effect in Definition 2, and to confirm that it still corresponds to an estimand we would want to interpret as an indirect effect when we restrict our attention to simpler parametric models.

Example 1 (Cai et al. (2015))

Consider a network model: there is an edge matrix $E_{ij} \in \{0, 1\}$ such that W_j can affect Y_i only if the corresponding units are connected by an edge, i.e., $E_{ij} = 1$. Then considered a linear-in-means model

$$Y_i = \beta_1 + \beta_2 W_i + \beta_3 \frac{\sum_{j+i} E_j W_j}{\sum_{j \neq i} E_{ij}} + \varepsilon_i, \quad E(\varepsilon_i \mid W) = 0.$$

- We can show that $\tau_{ADE} = \beta_2$, $\tau_{AIE} = \beta_3$.

Example 2 (Bakshy et al., 2012)

The model in Example 1 assumes that the ith unit responds in the same way to treatment assigned to any of its neighbours. However, this restriction may be implausible in many situations. A natural generalization of Example 1 is that

$$Y_i = \alpha_i + \beta_i W_i + \sum_{j \neq i} v_j W_j + \varepsilon_i, \quad E(\varepsilon_i \mid W) = 0.$$

- We can show that $\tau_{ADE} = n^{-1} \sum_{i=1}^{n} \beta_i$ and $\tau_{AIE} = n^{-1} \sum_{i=1}^{n} \sum_{j \neq i} v_{ij}$.

Example 3 (Sinclair et al. (2012))

In studying the effects of persuasion campaigns or other types of messaging, one might assume that people respond most strongly if they receive a communication directly addressed to them, but may also respond if a member of their neighbourhood or household gets a communication. This assumption can be formalized in terms of a model where each unit has four potential outcomes:

$$Y_i = \begin{cases} Y_i(\text{treated\&exposed}), & W_i = 1 \text{ and i has a treated neighbour,} \\ Y_i(\text{treated}), & W_i = 1 \text{ but i has no treated neighbours,} \\ Y_i(\text{exposed}), & W_i = 0 \text{ but i has a treated neighbours,} \\ Y_i(\text{none}), & W_i = 0 \text{ and i has no treated neighbours.} \end{cases}$$

Example 3 (Sinclair et al. (2012))

The estimands of interest:

$$\tau_{\mathrm{SELF},1} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(\text{treated\&exposed}) - Y_i(\text{exposed})\},$$

$$\tau_{\mathrm{SELF},0} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(\text{treated}) - Y_i(\text{none})\}$$

$$\tau_{\mathrm{SPILL},1} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(\text{treated\&exposed}) - Y_i(\text{treated})\},$$

$$\tau_{\mathrm{SPILL},0} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(\text{exposed}) - Y_i(\text{none})\}.$$

We can show that

$$\tau_{\mathrm{ADE}} = \left(\rho - \frac{\rho}{m}\right)\tau_{\mathrm{SELF},1} + \left(1 - \rho + \frac{\rho}{m}\right)\tau_{\mathrm{SELF},0},$$

$$\tau_{\mathrm{AIE}} = (m-1)\left\{\frac{\rho}{m}\tau_{\mathrm{SPILL},1} + \left(1 - \rho + \frac{\rho}{m}\right)\tau_{\mathrm{SPILL},0}\right\}.$$

Outline

Contributions

$$\tau_{\mathrm{AIE}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} E\{Y_j(w_i = 1; W_{-i}) - Y_j(w_i = 0; W_{-i})\}.$$

Limitations

The estimands defined here will in general depend on the design. Figure 2 illustrates this phenomenon in a Bernoulli experiment by plotting $\tau_{ADE}(\pi)$ and $\tau_{AIE}(\pi)$ as functions of π in the following three structural models with constant treatment probability π :

- $Y_i = \frac{\sum_{i \neq j} E_{ij} W_j}{300} + \frac{2 W_i}{3} + \varepsilon_i,$
- $Y_i = 1 - \left(1 - \frac{\sum_{i \neq j} E_{ij} W_j}{\sum_{i \neq j} E_{ij}}\right)^2 \left(1 - \frac{W_i}{2}\right) + \varepsilon_i,$
- $Y_i = W_i \left\{ e_i - 3\left(e_i - \frac{1}{2}\right)^3 \right\} + \varepsilon_i, \quad e_i = \frac{\sum_{i \neq j} E_{ij} W_j}{\sum_{i \neq j} E_{ij}}.$

Remark 1

Authors believe such dependence to be largely unavoidable when seeking to define nonparametric estimands in the generality considered here.

Limitations

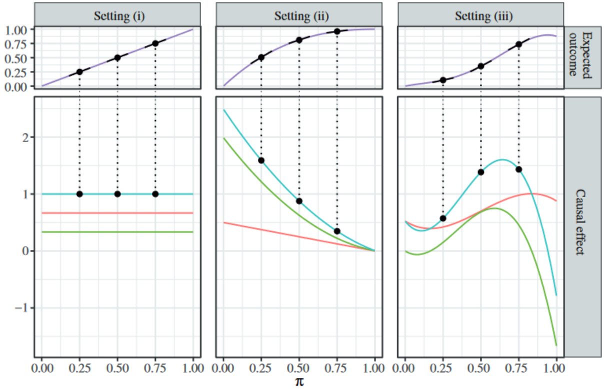

Fig. 1. Plots of τ_{ADE} (red), τ_{AIE} (green), τ_{INF} (blue) and the expected potential outcome $E_\pi(Y_i)$ (purple) as functions of the treatment probability π. The slopes of tangent line segments on the purple curve, which represent the derivative of $E_\pi(Y_i)$ at those points, are the same as the values on the blue curve, τ_{INF}. Theorem 1 establishes that $\tau_{\mathrm{INF}} = \tau_{\mathrm{ADE}} + \tau_{\mathrm{AIE}}$. In the plots, the blue curve corresponds to the sum of the red curve and the green curve. We consider the three settings in (4), where in all cases we assume constant treatment assignment probability $\pi_i = \pi_0$ and take the number of neighbours to be $\sum_{i \neq j} E_{ij} = 100$.

图: Estimands depend on the design.

Further Discussion

- It seems that the contribution of this paper is just a definition.
- Li and Wager (2022, AOS) proposed consistent estimators for their estimands. Under some standard assumptions, they established a center limit theorem for the direct effect.

Thank you!