Causal Inference under Interference

Youjin Lee

Johns Hopkins School of Public Health
Department of Biostatistics

ylee160@jhu.edu

March 13, 2018

Introduction

- I am a fourth year PhD student in the Department of Biostatistics.
- My thesis will be a (little bit random) collection of statistical inference on network data.
- What is causal inference under network setting?
- Simply speaking, violation of SUTVA assumption.

Overview

1. Potential Outcomes under Interference
   - Examples of Interference
   - Potential Outcomes
   - Causal Estimands

2. Experiment Design under Interference
   - Assumptions
   - Two-stage randomization

Examples

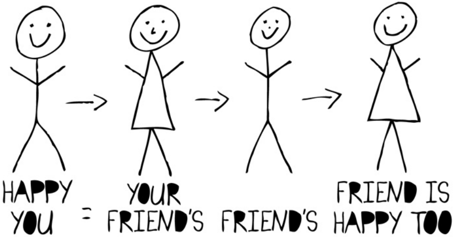

Figure: Outcome of one unit (Y_i) is often dependent on the others'. Some of them are "contagious".

Examples

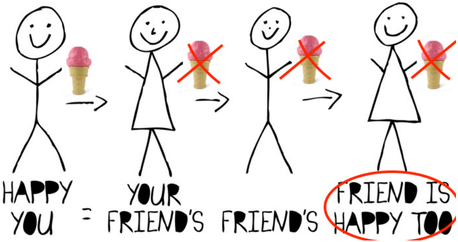

Figure: So does treatment effect. If eating icecream makes you happy, does it make your friends' friends' friend happy too? even though none of them except you do not eat ice cream?

Examples - Herd immunity

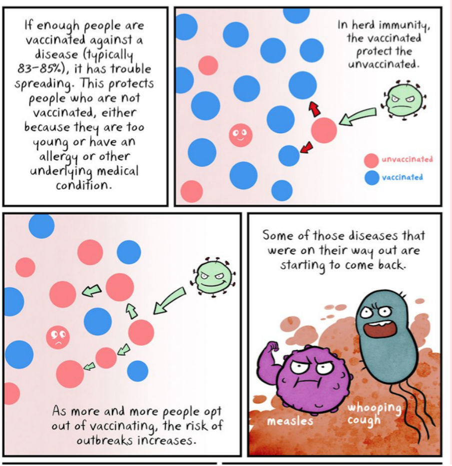

Figure: Sometimes treatment (e.g., being vaccinated) on one unit can be beneficial to others' outcomes because whether one person gets infected depends on who else in the population is vaccinated.

Examples

- Effect of housing voucher on the lease-up rate of whole villages (Sobel, 2006)
- Indirect effects of vaccination due to the level of coverage (Hudgens and Halloran, 2008)
- Educational interventions given to certain students may have an effect on other students' performance or behaviors in the same class (Rubin 1990, Rosenbaum 2007).
- Also called 'neighborhood effect', 'indirect causal effect', 'spillover effect', **interference** implies the effect of the treatment assignment of one individual on the potential outcomes of any other individuals.
- Interference is not a nuisance in causal inference in many cases!

Potential Outcomes

- Under interference, $Y_i(1) := Y_i(A_i = 1)$ can take more than one version depending on the treatment of other units ($\mathbf{A}_{-i} = \{A_j : j \neq i\}$)
  $\rightarrow$ Violation of SUTVA
- Instead, we can assume one version of $Y_i(\mathbf{A}) = Y_i(\{A_1, A_2, \ldots, A_n\})$
- That is, define the potential outcome of i under treatment assignment of all units $\mathbf{A}$.

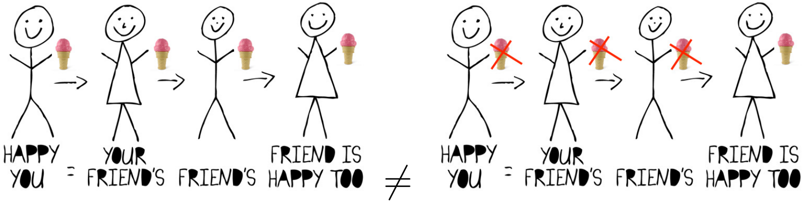

Figure: Your friends' friends' friend's potential outcomes ($Y_4(\mathbf{A})$) is not the same under these two treatment assignments even though he or she eats icecream!

Potential Outcomes

- Often, the 'level of interference' is assumed, e.g. allowing interference within "group" but no interference between the units from different groups.
- Denote treatment assignment of unit j in group i by A_{ij} $(i = 1, 2, \ldots, n_i,\ j = 1, 2, \ldots, N)$.
- Then the potential outcome of unit j in group i can be represented by $Y_{ij}(\mathbf{A}_i) = Y_{ij}(\{A_{i1}, A_{i2}, \ldots, A_{in_i}\})$.
- Under above notation, $Y_{ij}(\mathbf{A}) = Y_{ij}(\mathbf{A}')$ as long as $\mathbf{A}_i = \mathbf{A}'_i$.

Causal Estimands (Hudgens and Halloran, 2008)

Note : Depending on the literature, the exact form of the equation can be different, but what to ultimately estimate is basically the same.

1. Direct causal effects

$$\mathrm{DE}_i(\mathbf{a}_{-i}) = E\big(Y_i(\mathbf{a}_{-i}, a_i = 1)\big) - E\big(Y_i(\mathbf{a}_{-i}, a_i = 0)\big)$$

"When fixing others' treatment assignment ($\mathbf{a}_{-i}$), what is the effect from solely changing my treatment assignment?"

2. Indirect causal effects/spillover effects

$$\mathrm{SE}_i(\mathbf{a}_{-i}) = E\big(Y_i(\mathbf{a}_{-i}, a_i = 0)\big) - E\big(Y_i(\mathbf{a}_{-i} = \mathbf{0}, a_i = 0)\big)$$

"Fixing my own status as control ($a_i = 0$), what is the effect of others' treatment assignment $\mathbf{a}_{-i}$ compared to no null treatment on my potential outcomes?"

Under SUTVA, $\mathrm{SE}_i(\mathbf{a}_{-i}) = 0$ because $Y_i(\mathbf{a})$ is not affected by $\mathbf{a}_{-i}$.

Causal Estimands (Hudgens and Halloran, 2008)

- If we know the mechanism or regime of treatment assignment (say α) of $\mathbf{a}_{-i}$, we are able to calculate the allocation-specific unit average direct effects.

$$\mathrm{DE}_i(\alpha) = \sum_{a' \in \text{all possible assignments}} \mathrm{DE}_i(\mathbf{a}_{-i} = a')p(\mathbf{a}_{-i} = a')$$

- and also unit average spillover effect:

$$\mathrm{SE}_i(\alpha) = \sum_{a' \in \text{all possible assignments}} \mathrm{SE}_i(\mathbf{a}_{-i} = a')p(\mathbf{a}_{-i} = a').$$

- Network average direct and spillover effects : $\sum_{i=1}^{n} \mathrm{DE}_i(\alpha)/n$ and $\sum_{i=1}^{n} \mathrm{SE}_i(\alpha)/n$.

Causal Estimands (Hudgens and Halloran, 2008)

Now we can estimate...

- How many infections will be averted by vaccinating two-thirds of the population (treatment regime of α) compared to only vaccinating one-third of the population (treatment regime of α')?
- What proportion of households will move if two-thirds receive vouchers (treatment regime of α) compared to only one-third receiving vouchers (treatment regime of α')?

But how..?

Partial Interference

- Assumes that interference is (only) possible between units within the same group.
- This will be a reasonable assumption if the groups are sufficiently separate (e.g., in space or time).
- This assumption makes statistical inference easier (e.g. unbiased estimation for variance), but is also useful for policy perspectives.
- For example, we don't have to immunize someone in the other independent villages in order to prevent the infectious disease in one village.

Two-stage randomization

Example 1 : Housing Vouchers
Assume that households from different villages are sufficiently separated geographically so that there is no interference across the villages.

1. Some villages are randomly assigned the 'benchmark' allocation (no treatment at all) and the other villages to an allocation of α.
2. Under the 'benchmark' allocation, all units (e.g. household) in the selected group (e.g. village) do not receive a treatment.
3. Under α, a specified proportion (e.g. 30%) of units (e.g. household) are randomly chosen and receive a treatment.

Two-stage randomization

Example 2 : Vaccines

Assuming geographically separate groups, consider a 'hypothetical' two-stage randomized placebo-controlled trial.

1. Some groups are randomly assigned the allocation 1 (α) and the other groups to an allocation 2 (α').
2. Under the allocation 1 (α), 30% of individuals are randomly chosen to receive vaccine.
3. Under α',50% of individuals are randomly chosen to receive vaccine.

Comparison between 50% vaccine coverage and 30% vaccine coverage is possible.

Other Research Questions (of mine)

- If subjects are interacting each other consistently, how can we estimate causal effects of the treatment on the collective outcomes? (e.g., effect of new information on the opinion formation within the group)
- How can we identify most "causally influential" subjects on a social network? (Who we should treat to have a largest casual effect?)
- Effect of sampling dependent on social network (e.g., snowball sampling, respondent-driven sampling) on causal inference.

References

- Rosenbaum, P. R. (2007). Interference between units in randomized experiments. Journal of the American Statistical Association, 102(477), 191-200.
- Rubin, D. B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. Statistical Science, 5(4), 472-480.
- Hudgens, M. G., & Halloran, M. E. (2008). Toward causal inference with interference. Journal of the American Statistical Association, 103(482), 832-842.
- Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. Journal of the American Statistical Association, 101(476), 1398-1407.

Thank you!