Identification and Estimation of Treatment and Interference Effects in Observational Studies on Networks[1]

Nankai University

October 9, 2022

[1] Source : Forastiere L, Airoldi E M, Mealli F. Identification and estimation of treatment and interference effects in observational studies on networks[J]. Journal of the American Statistical Association, 2021, 116(534): 901-918.

Interference

- Interference is said to be present when a treatment, exposure, or intervention, on one unit has an effect on the response of another unit [Cox 1958]

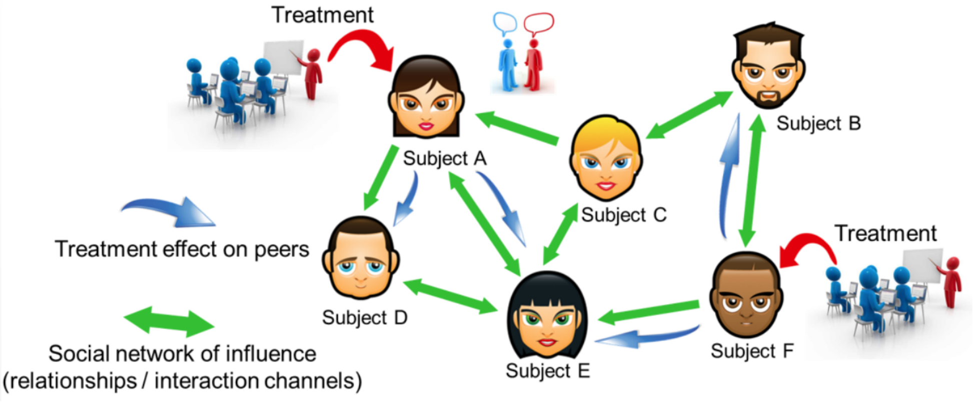

Figure 1: from Kao(2017)

Related literature

- assume no interference effect [contained in SUTVA,Rubin 1980] ,
- unreasonable and risky (Sobel 2006)
- Experimental studies
  - cluster randomized trials , two stage randomization (Hudgens and Halloran 2008; Sinclair 2011) , Insulated Neighbors Randomization (Toulis and Kao 2013)
  - test and estimate the total treatment effect (Rosenbaum 2007), separately test for treatment and spillover effects (Bowers, Fredrickson, and Panagopoulos 2013) , Horvitz–Thompson estimator (Aronow and Samii 2017)

Related literature

- spillover effect for observational studies
  - parametric multilevel approach (Raudenbush 2006) , three-level generalized hierarchical linear model (Verbitsky-Savitz and Raudenbush 2012) , IPW estimators (Tchetgen Tchetgen and VanderWeele 2012)
- causal inference for observational studies
  - generalized IPW estimator (Liu et al. 2016) , targeted maximum likelihood estimator (TMLE) (Sofrygin and van der Laan 2017) , extended this TMLE estimator to allow for dependence (Ogburn et al 2017)

Contents

I. Causal problem and identification

Notation

- Undirected network $G = (\mathcal{N}, \mathbb{E})$
- Treatment $Z_i \in \{0, 1\}$, outcome $Y_i \in \mathcal{Y}$
- covariates $\mathbf{X}_i \in \mathcal{X}$,individual covariates $\mathbf{X}_i^{ind} \in \mathcal{X}^{ind}$,neighborhood covariates $\mathbf{X}_i^{neigh} \in \mathcal{X}^{neigh}$
- For unit i, consider partition $(i, \mathcal{N}_i, \mathcal{N}_{-i})$, $(Z_i, \mathbf{Z}_{\mathcal{N}_i}, \mathbf{Z}_{\mathcal{N}_{-i}})$, $(Y_i, \mathbf{Y}_{\mathcal{N}_i}, \mathbf{Y}_{\mathcal{N}_{-i}})$

Potential Outcomes and Neighborhood Interference

- ***Stable unit treatment on neighborhood value assumption (SUTNVA)***

Assumption 1 (No multiple version of treatment(consistency))

If $\mathbf{Z}$=$\mathbf{z}$ then $Y_i = Y_i(\mathbf{z})$

Assumption 2 (Neighborhood interference)

Given a function g_i: $\{0,1\}^{\mathcal{N}_i} \to \mathcal{G}_i$,$\forall i \in \mathcal{N}$, $\forall \mathbf{Z}_{\mathcal{N}_{-i}}$, $\mathbf{Z}'_{\mathcal{N}_{-i}}$ and $\forall \mathbf{Z}_{\mathcal{N}_i}$, $\mathbf{Z}'_{\mathcal{N}_i}$: $g_i(\mathbf{Z}_{\mathcal{N}_i}) = g_i(\mathbf{Z}'_{\mathcal{N}_i})$, the following equality holds :

$$Y_i(Z_i, \mathbf{Z}_{\mathcal{N}_i}, \mathbf{Z}_{\mathcal{N}_{-i}}) = Y_i(Z_i, \mathbf{Z}'_{\mathcal{N}_i}, \mathbf{Z}'_{\mathcal{N}_{-i}})$$

Each unit's potential outcome is subject to two treatments : the *individual treatment* Z_i , and *the neighborhood treatment* G_i $[G_i = g_i(\mathbf{Z}_{\mathcal{N}_i}) \in \mathcal{G}_i]$

Individual and Neighborhood Treatments

- The assignment mechanism is the ***probability distribution of the joint treatment*** in the whole sample, given all covariates and potential outcomes.

$$P(\mathbf{Z}, \mathbf{G} | \mathbf{X}, \{\mathbf{Y}(z, g), z = 0, 1; g \in \mathcal{G}\}) =$$

$$\begin{cases} P(\mathbf{Z} | \mathbf{X}, \{\mathbf{Y}(z, g), z = 0, 1; g \in \mathcal{G}\}) & \text{if} \quad \mathbf{G} = \mathbf{g}(\mathbf{Z}) \\ 0 & \textit{otherwise} \end{cases}$$

Causal Estimands : Main Effects and Spillover effects

Average dose-response function (ADRF)

$$\mu(z, g; V) = E[Y_i(z, g) | i \in V] \qquad \forall z \in \{0, 1\}, g \in \mathcal{G}$$

Main effect $\tau(g)$ and overall main effect τ

$$\tau(g) = \mu(1, g; V_g) - \mu(0, g; V_g)$$
$$\tau = \sum_{g \in \mathcal{G}} \tau(g) P(G_i = g)$$

Spillover effect $\delta(g; z)$ and overall spillover effect $\Delta(z)$

$$\delta(g; z) = \mu(z, g; V_g) - \mu(z, 0; V_g)$$
$$\Delta(z) = \sum_{g \in \mathcal{G}} \delta(g; z) P(G_i = g)$$

$$TE = \sum_{g \in \mathcal{G}} E[Y_i(Z_i = 1, G_i = g) - Y_i(Z_i = 0, G_i = 0) | i \in V_g] P(G_i = g) = \tau + \Delta(0)$$

Unconfoundedness of the Joint Treatment

Assumption 3 (Unconfoundedness of individual and neighborhood treatment)

$$Y_i(z, g) \perp\!\!\!\perp Z_i, G_i | \mathbf{X}_i \qquad \forall z \in \{0, 1\}, g \in \mathcal{G}_i, \forall i$$

Theorem 1 (Identification of ADRF)

Under Assumption 1 (no multiple versions of treatment), Assumption2 (neighborhoood interference) , and Assumption 3 (unconfoundedness) , we have

$$E[Y_i(z, g)|i \in V_g] = \sum_{\mathbf{x} \in \mathcal{X}} E[Y_i|Z_i = z, G_i = g, \mathbf{X_i} = \mathbf{x}, i \in V_g]P(\mathbf{X_i} = \mathbf{x}|i \in V_g)$$

II. Bias analysis

Bias When SUTVA Is Wrongly Assumed

Under SUTVA ,average (individual) treatment effect : $\tau_{sutva} = E[Y_i(Z_i = 1) - Y_i(Z_i = 0)]$

Naive estimator

$$\tau^{obs}_{\mathbf{X}^\star} = \sum_{\mathbf{x} \in \mathcal{X}^\star} (E[Y_i | Z_i = 1, \mathbf{X}_i^\star] - E[Y_i | Z_i = 1, \mathbf{X}_i^\star]) P(\mathbf{X}_i^\star = \mathbf{x})$$

The difference between τ and the quantity $\tau^{obs}_{\mathbf{X}^\star}$ represents the bias for τ of a naive approach that neglects interference

Bias of Naive Estimator When Unconfoudedness Holds

Theorem 2.A

Let $G = (\mathcal{N}, \mathbb{E}$ be a known social network and lei $\mathcal{N}_i$ be the neighborhood of unit i asa defined by the presence of edges . Let $Z_i \in \{0, 1\}$ be a binary treatment assigned to unit i and let G_i be a deterministic function of the subset of the treatment vector $\mathbf{Z}$ in the neighborhood $\mathcal{N}_i$, that is , $G_i = g_i(\mathbf{Z}_{\mathcal{N}_i})$, with $g_i : \{0, 1\}^{N_i} \to \mathcal{G}_i$. If

1. Assumption 1 holds
2. Assumption 2 holds , given function $g_i(\cdot)$ for each unit $i \in \mathcal{N}$
3. Assumption 3 holds conditional on $\mathbf{X}_i^\star$, that is , $Y_i(z, g) \perp\!\!\!\perp Z_i, G_i|\mathbf{X}_i^\star$, $\forall z \in \{0, 1\}, g \in \mathcal{G}_i$

then the following equality holds

$$\tau_{\mathbf{X}^\star}^{obs} = \sum_{\mathbf{x}\in\mathcal{X}^\star} \Big(\sum_{g\in\mathcal{G}} E[Y_i(1, g)|\mathbf{X}_i^\star = \mathbf{x}, i \in V_g]P(G_i = g|Z_i = 1, , i \in X_i^\star = \mathbf{x})$$

$$- E[Y_i(0, g)|, \mathbf{X}_i^\star = \mathbf{x}, i \in V_g]P(G_i = g|Z_i = 0, \mathbf{X}_i^\star = \mathbf{x})\Big) P(\mathbf{X}_i^\star = \mathbf{x})$$

Bias of Naive Estimator When Unconfoudedness Holds

Corollary 1

Under the three conditions of Theorem 2.A and the additional condition

4. Z_i and G_i are independent conditional on $\mathbf{X}_i^\star$, that is , $Z_i \perp\!\!\!\perp G_i | \mathbf{X}_i^\star$

the following equality holds :

$$\tau_{\mathbf{X}^\star}^{obs} = \tau$$

Corollary 2

Under the three conditions of Theorem 2.A and $Z_i \not\perp\!\!\!\perp G_i | \mathbf{X}_i^\star$, an unbiased estimator of $\tau_{\mathbf{X}^\star}^{obs}$ would be biased for the overall main effect τ , with bisas given by

Bias of Naive Estimator When Unconfoudedness Holds

Corollary 2 continued

$$
\begin{aligned}
\tau^{obs}_{\mathbf{X}^\star} - \tau = \sum_{\mathbf{x}\in\mathcal{X}^\star}\sum_{g\in\mathcal{G}} \Big(& E[Y_i|Z_i = 1, G_i = g, \mathbf{X}_i^\star = \mathbf{x}, i \in V_g] \\
& - E[Y_i|Z_i = 1, G_i = g', \mathbf{X}_i^\star = \mathbf{x}, i \in V_g]\Big) \\
& \Big(P(G_i = g|Z_i = 1, \mathbf{X}_i^\star = \mathbf{x}) - P(G_i = g|\mathbf{X}_i^\star = \mathbf{x})\Big) P(\mathbf{X}_i^\star = \mathbf{x}) \\
& - \sum_{\mathbf{x}\in\mathcal{X}^\star}\sum_{g\in\mathcal{G}} \Big(E[Y_i|Z_i = 0, G_i = g, \mathbf{X}_i^\star = \mathbf{x}, i \in V_g] \\
& - E[Y_i|Z_i = 0, G_i = g', \mathbf{X}_i^\star = \mathbf{x}, i \in V_g]\Big) \\
& \Big(P(G_i = g|Z_i = 0, \mathbf{X}_i^\star = \mathbf{x}) - P(G_i = g|\mathbf{X}_i^\star = \mathbf{x})\Big) P(\mathbf{X}_i^\star = \mathbf{x})
\end{aligned}
$$

Bias of Naive Estimator When Unconfoudedness Does Not Hold

Theorem 2.B

Under Assumption 1 and 2 , if Assumption 3 does not hold conditional on $\mathbf{X}_i^\star$,that is $Y_i(z,g) \not\perp\!\!\!\perp Z_i, G_i | \mathbf{X}_i^\star$, but holds conditioanl on $\mathbf{X}_i^\star$ and an additional vector of covariates $\mathbf{U}_i \in \mathcal{U}$, that is , $Y_i(z,g) \perp\!\!\!\perp Z_i, G_i | \mathbf{X}_i^\star, \mathbf{U}_i$, an unbiased estimator of $\tau^{obs}_{\mathbf{X}^\star}$ would be unbiased for the overall main effect τ , with bias $\tau^{obs}_{\mathbf{X}^\star} - \tau$ given by

$$
\begin{aligned}
&= \sum_{\mathbf{x}\in\mathcal{X}^\star} \sum_{g\in\mathcal{G}} \sum_{u\in\mathcal{U}} \Big(E[Y_i | Z_i = 1, G_i = g, \mathbf{U}_i = \mathbf{u}, \mathbf{X}_i^\star = \mathbf{x}, i \in V_g] \\
&- E[Y_i | Z_i = 1, G_i = g', \mathbf{U}_i = \mathbf{u}', \mathbf{X}_i^\star = \mathbf{x}, i \in V_g] \Big) \\
&\Big(P(\mathbf{U}_i = \mathbf{u} |, Z_i = 1, G_i = g, \mathbf{X}_i^\star = \mathbf{x}) P(G_i = g | Z_i = 1, \mathbf{X}_i^\star = \mathbf{x}) \\
&- P(\mathbf{U}_i = \mathbf{u} | G_i = g, \mathbf{X}_i^\star = \mathbf{x}) P(G_i = g | \mathbf{X}_i^\star = \mathbf{x}) \Big) P(\mathbf{X}_i^\star = \mathbf{x})
\end{aligned}
$$

Bias of Naive Estimator When Unconfoudedness Does Not Hold

Theorem 2.B continued

$$
\begin{aligned}
&= \sum_{\mathbf{x} \in \mathcal{X}^\star} \sum_{g \in \mathcal{G}} \sum_{u \in \mathcal{U}} \Big(E[Y_i | Z_i = 0, G_i = g, \mathbf{U}_i = \mathbf{u}, \mathbf{X}_i^\star = \mathbf{x}, i \in V_g] \\
&- E[Y_i | Z_i = 0, G_i = g', \mathbf{U}_i = \mathbf{u}', \mathbf{X}_i^\star = \mathbf{x}, i \in V_g] \Big) \\
&\Big(P(\mathbf{U}_i = \mathbf{u} |, Z_i = 0, G_i = g, \mathbf{X}_i^\star = \mathbf{x}) P(G_i = g | Z_i = 0, \mathbf{X}_i^\star = \mathbf{x}) \\
&- P(\mathbf{U}_i = \mathbf{u} | G_i = g, \mathbf{X}_i^\star = \mathbf{x}) P(G_i = g | \mathbf{X}_i^\star = \mathbf{x}) \Big) P(\mathbf{X}_i^\star = \mathbf{x})
\end{aligned}
$$

Bias of Naive Estimator When Unconfoudedness Does Not Hold

Corollary 3

Under the following conditions

1. Assumption 1 holds and , if
2. Assumption 2 holds given the function $g_i(\cdot)$ for each unit $i \in \mathcal{N}$
3. Assumption 3 holds conditional on $\mathbf{X}_i^\star$ and $\mathbf{U}_i$, that is , $Y_i(z, g) \perp\!\!\!\perp Z_i, G_i | \mathbf{X}_i^\star, \mathbf{U}_i$, $\forall z \in \{0, 1\}, g \in \mathcal{G}_i$
4. Z_i and G_i are independent conditioanl on $\mathbf{X}_i^\star$,that is , $Z_i \perp\!\!\!\perp G_i | \mathbf{X}_i^\star$

$$\begin{aligned}\tau_{\mathbf{X}^\star}^{obs} - \tau = \sum_{\mathbf{x} \in \mathcal{X}^\star} \sum_{u \in \mathcal{U}} \Big(& E[Y_i | Z_i = z, \mathbf{U}_i = \mathbf{u}, \mathbf{X}_i^\star = \mathbf{x}] \\ & - E[Y_i | Z_i = z, \mathbf{U}_i = \mathbf{u}', \mathbf{X}_i^\star = \mathbf{x}] \Big) \\ & \Big(P(\mathbf{U}_i = \mathbf{u} | Z_i = 1, \mathbf{X}_i^\star = \mathbf{x}) - P(\mathbf{U}_i = \mathbf{u} | Z_i = 1, \mathbf{X}_i^\star = \mathbf{x}) \Big) P(\mathbf{X}_i^\star = \mathbf{x})\end{aligned}$$

where $z \in \{0, 1\}$

Bias of Naive Estimator When Unconfoudedness Does Not Hold

Corollary 4

Under the following conditions

1. SUTVA holds
2. Assumption 3 holds conditional on $\mathbf{X}_i^\star$ and $\mathbf{U}_i$, that is , $Y_i(z,g) \perp\!\!\!\perp Z_i, G_i | \mathbf{X}_i^\star, \mathbf{U}_i$, $\forall z \in \{0,1\}, g \in \mathcal{G}_i$

Bias $\tau^{obs}_{\mathbf{X}^\star} - \tau$ given only by the unmeasured confounder $\mathbf{U}_i$ as in equation from corollary 3

III.Proposed method

Generalized Propensity Score

- *joint propensity score* $\psi(z; g; x) = P(Z_i = z, G_i = g | \mathbf{X}_i = \mathbf{x})$
  - Balancing Property $P(Z_i = z, G_i = g | \mathbf{X}_i, \psi(z; g; \mathbf{X}_i)) = P(Z_i = z, G_i = g | \psi(z; g; \mathbf{X}_i))$
  - Conditional unconfoundedness $Y_i(z, g) \perp\!\!\!\perp Z_i, G_i | \psi(z; g; \mathbf{X}_i) \; \forall z \in \{0, 1\}, g \in \mathcal{G}_i$
- *individual propensity score* $\phi(z; x^z) = P(Z_i = z | \mathbf{X}_i^z = \mathbf{x}^z)$
- *neighborhood propensity score* $\lambda(g; z; x^g) = P(G_i = g | Z_i = z, \mathbf{X}_i^g = \mathbf{x}^g)$
- *factorization of joint propensity score* $\psi(z; g; x) = \phi(z; x^z)\lambda(g; z; x^g)$
  - *conditional unconfoundedness*

$$Y_i(z, g) \perp\!\!\!\perp Z_i, G_i | \lambda(g; z; \mathbf{X}_i^g), \phi(1; \mathbf{X}_i^z) \quad \forall z \in \{0, 1\}, g \in \mathcal{G}_i$$

Propensity Score Based Estimator for Main Effects and Spillover Effects

unbiased estimator of $\mu(z, g; V)$

- $E\Big[E[Y_i|Z_i = z, G_i = g, \psi(z; g; \mathbf{X}_i)]|Z_i = z, G_i = g\Big]$
- $E\Big[E[Y_i|Z_i = z, G_i = g, \lambda(g; z; \mathbf{X}_i^g), \phi(1; \mathbf{X}_i^z)]|Z_i = z, G_i = g\Big]$ (separate joint propensity score)

Estimation Procedure : Subclassification and GPS

- Derive a subcalssification on the individual propensity score $\phi(1; \mathbf{X}_i^z)$
  (a) estimate $\phi(1; \mathbf{X}_i^z)$ by logistic regression : $Z_i \sim \mathbf{X}_i^z$
  (b) predict $\phi(1; \mathbf{X}_i^z)$
  (c) identify J subclasses B_j by $\phi(1; \mathbf{X}_i^z)$, where $\mathbf{X}_i^z \perp\!\!\!\perp \mathbf{Z}_i | i \in B_j$

- Within B_j ,estimate $\mu_j(z, g; V_g) = E[Y_i(z, g) | i \in B_j^g]$, where $B_j^g = V_g \cap B_j$
  (a) estimate model for $\lambda(g; z; x^g)$ by $f^G(g, z, \mathbf{X}_i^g)$
  (b) estimate model for $Y_i(z, g)$ by $f^Y(z, g, \lambda(g; z; \mathbf{X}_i^g))$
  (c) predict $\lambda(g; z; \mathbf{X}_i^g)$ and $Y_i(z, g)$
  (d) estimate $\hat{\mu}_j(z, g; V_g) = \sum_{i \in B_j^g} \hat{Y}_i(z, g) / |B_j^g|$

- Derive the ADRF $\hat{\mu}(z, g; V_g) = \sum_{j=1}^{J} \hat{\mu}_j(z, g; V_g) \pi_j^g$, where $\pi_j^g = \frac{|B_j^g|}{\nu_g}$

IV.Simulation studies

Takeaways for simulation

- Aims
  - validate the analytical derivation of the bias for the main effect when interference is wrongly ruled out
  - show the performance of the proposed estimators in a realistic sample
- Data : We use friendship network data collected through the National Longitudinal Study of Adolescent Health . (29 schools for a total of 16410 students)
- Variables
  - Z_i : denote whether student i was covered or not by some health insurance
  - Y_i : denote the number of days student i missed school because of illness in one given year
  - $\mathbf{X}_i^{ind} = (race_i, grade_i)$ and $\mathbf{X}_i^{neig} = (\frac{\sum_{k \in \mathcal{N}_i} race_k}{N_i}, \frac{\sum_{k \in \mathcal{N}_i} grade_k}{N_i}, N_i)$

Four scenarios of dependence between Z_i and G_i [2]

- Scenario 1 : $Z_i \perp\!\!\!\perp G_i|\mathbf{X}_i^{ind}$.
  (treatment generating process : $logit(P(Z_i = 1)) = -18 + 2grade_i + 3race_i$)

- Scenario 2 : $Z_i \perp\!\!\!\perp G_i|\mathbf{X}_i^{ind}, \mathbf{X}_i^{neig}$
  (treatment generating process :
  $logit(P(Z_i = 1)) = -47 + 2grade_i + 4race_i + 3friends.grade_i + 5friends.race_i$)

- Scenario 3 : $Z_i \perp\!\!\!\perp G_i|\mathbf{X}_i^{ind}, \mathbf{N}_i$
  (treatment generating process : $logit(P(Z_i = 1)) = -49 + 3grade_i + 4race_i + 4N_i$)

- Scenario 4 : $Z_i \not\perp\!\!\!\perp G_i|\mathbf{X}_i^{ind}, \mathbf{X}_i^{neig}$
  (treatment generating process : $logit(P(Z_i = 1)) = -20 + 2grade_i + 3race_i + 4G_i$)

[2] In all scenarios but the third, G_i is the proportion of friends with health insurance among the first five best friends. In the third scenario, G_i is the number of "treated" friends among all friends

Outcome Models

- <u>Outcome Models for Main Effects Simulations</u>

$$Y_i(z,g)|\mathbf{X}_i^{ind} \sim \mathcal{N}(\mu(z,g,\mathbf{X}_i^{ind}),1)$$

$$\mu(z,g,\mathbf{X}_i^{ind}) = 15 - 7\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7) - 15z + 3z\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7) + \delta g$$

- $X_i^{ind} = [race_i + grade_i]$, $X_i^g = [race_i, grade_i, friends.race_i, friends.grade_i, N_i]$, $\delta \in (-5,-8,-10)$ corresponding to a low medium and high level of interference

- main effects : $\tau(g) = -15 + 3\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7)\ \forall g \in \mathcal{G} \implies \tau = -15 + 3\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7)$

- spillover effects : $\delta(g;z) = \delta g \implies \Delta(z) = \delta E[G_i]\ \forall z = 0,1$

Outcome Models

- Outcome Models for Spillover Effects Simulations

$$Y_i(z,g)|\mathbf{X}_i^{ind},\mathbf{X}_i^{g} \sim \mathcal{N}(\mu(z,g,\mathbf{X}_i^{ind},\mathbf{X}_i^{g}),1)$$

$$\mu(z,g,\mathbf{X}_i^{ind},\mathbf{X}_i^{g}) = 15 + friends.grade_i + 7 friends.race_i - 10\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7) - 10z + \delta g - 10\lambda(g;z,\mathbf{X}_i^{g} + 5g\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7) + 3zg$$

- $X_i^{ind} = [race_i + grade_i]$, $X_i^{g} = [race_i, grade_i, friends.race_i, friends.grade_i, N_i]$,
  $\delta \in (-5,-8,-10)$ corresponding to a low medium and high level of interference

- main effects : $\tau(g) = -10 + 3g \implies \tau = -10 + 3E[G_i]$
  $\tau(g) = -15 + 3\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7)\ \forall g \in \mathcal{G} \implies \tau = -15 + 3\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7)$

- spillover effects : $\delta(g;z) = \delta g 5gE[\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7)] - 10\lambda(g;z,\mathbf{X}_i^{g})$

$\implies \Delta(z) = \delta E[G_i] + 5E[G_i]E[\mathbf{I}(\phi(1;\mathbf{X}_i^{ind}) \geq 0.7)] - 10E[\lambda(G_i;z,\mathbf{X}_i^{g})] + 3zE[G_i]$

Main Effects : Bias of Naive Estimator and GPS-Based Estimator

- Table 1 shows the bias computed using formulas in Theorem 2.A and Theorem 2.B, bias resulting from neglecting interference and from adjusting for different sets of covariates : $\mathbf{X}_i^\star = \{\varnothing, \mathbf{X}_i^{ind}, \mathbf{X}_i^z\}$

Scenario	Interference	Bias($\emptyset$)	Bias($\mathbf{X}_i^{\text{ind}}$)	Bias($\mathbf{X}_i^z$)
1	Low	−5.977	−0.045	−0.045
$(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}})$	Medium	−6.200	−0.072	−0.072
	High	−6.323	−0.090	−0.090
2	Low	−5.749	−1.636	−0.034
$(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, \mathbf{X}_i^{\text{neigh}})$	Medium	−6.498	−2.618	−0.054
	High	−6.998	−3.273	−0.068
3	Low	−4.158	−1.247	−0.047
$(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, N_i)$	Medium	−4.792	−2.079	−0.075
	High	−5.744	−3.327	−0.095
4	Low	−9.504	−1.414	−1.415
$(Z_i \not\!\perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, \mathbf{X}_i^{\text{neigh}})$	Medium	−11.681	−2.263	−2.263
	High	−13.132	−2.829	−2.825

Table 1: Computed bias for τ

Main Effects : Bias of Naive Estimators and GPS-Based Estimator

- Table 2 reports the mean bias and root mean squared error of six estimators in all scenarios.

Scenario	Interference	Unadjusted Bias	Unadjusted RMSE	Regression $\sim Z_i, \mathbf{X}_i^{\text{ind}}$ Bias	Regression $\sim Z_i, \mathbf{X}_i^{\text{ind}}$ RMSE	Subclass $\hat{\phi}(1, \mathbf{X}_i^{\text{ind}})$ Bias	Subclass $\hat{\phi}(1, \mathbf{X}_i^{\text{ind}})$ RMSE	Regression $\sim Z_i, \mathbf{X}_i^{Z}$ Bias	Regression $\sim Z_i, \mathbf{X}_i^{Z}$ RMSE	Subclass $\hat{\phi}(1; \mathbf{X}_i^{Z})$ Bias	Subclass $\hat{\phi}(1; \mathbf{X}_i^{Z})$ RMSE	Subclass $\hat{\phi}(1, \mathbf{X}_i^{Z})$ and GPS Bias	Subclass $\hat{\phi}(1, \mathbf{X}_i^{Z})$ and GPS RMSE
1 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}})$	Low	−5.104	5.104	−3.429	3.431	−0.016	0.130	−3.254	3.256	0.006	0.074	−0.002	0.064
	Medium	−5.278	5.278	−3.529	3.532	−0.021	0.149	−3.256	3.258	0.007	0.090	0.002	0.067
	High	−5.396	5.396	−3.598	3.601	−0.021	0.167	−3.258	3.261	0.015	0.108	0.003	0.067
2 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, \mathbf{X}_i^{\text{neigh}})$	Low	−5.136	5.136	−3.477	3.477	−1.652	1.668	−2.156	2.157	−0.066	0.331	−0.000	0.036
	Medium	−5.884	5.884	−4.521	4.521	−2.620	2.631	−2.318	2.319	−0.097	0.511	0.003	0.039
	High	−6.384	6.384	−5.222	5.223	−3.272	3.281	−2.430	2.432	−0.123	0.635	−0.002	0.036
3 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, N_i)$	Low	−4.106	4.107	−2.079	2.089	−1.146	1.149	−1.033	1.049	0.078	0.553	−0.002	0.230
	Medium	−4.725	4.726	−2.886	2.893	−1.911	1.913	−1.041	1.056	0.082	0.556	0.019	0.233
	High	−5.655	5.656	−4.098	4.103	−3.058	3.059	−1.054	1.068	0.084	0.559	0.012	0.234
4 $(Z_i \not\perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, \mathbf{X}_i^{\text{neigh}})$	Low	−8.670	8.670	−7.104	7.104	−1.443	1.445	−7.076	7.076	−1.738	1.740	0.000	0.105
	Medium	−10.761	10.761	−8.861	8.862	−2.309	2.309	−8.824	8.823	−2.784	2.785	−0.003	0.095
	High	−12.154	12.154	−10.030	10.030	−2.883	2.884	−9.985	9.985	−3.477	3.478	0.010	0.090

Table 2: Estimation of main effect τ

Spillover Effects

- Table 3 and Table 4 report the mean bias and root mean squared error of all estimators for spillover effects $\Delta(0)$ and $\Delta(1)$.

Scenario	Interference	Unadjusted regression $\sim Z_i, G_i$ Bias	Unadjusted regression $\sim Z_i, G_i$ RMSE	Subclass $\hat{\phi}(1, \mathbf{X}_i^z)$ Bias	Subclass $\hat{\phi}(1, \mathbf{X}_i^z)$ RMSE	GPS Bias	GPS RMSE	Subclass $\hat{\phi}(1, \mathbf{X}_i^z)$ and GPS Bias	Subclass $\hat{\phi}(1, \mathbf{X}_i^z)$ and GPS RMSE
1 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}})$	Low	1.576	1.581	2.871	2.901	−1.457	1.465	0.003	0.091
	Medium	1.568	1.574	2.838	2.872	−1.462	1.472	0.005	0.114
	High	1.571	1.580	2.878	2.910	−1.467	1.479	0.000	0.148
2 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, \mathbf{X}_i^{\text{neigh}})$	Low	3.503	3.506	4.132	5.259	−0.574	0.590	0.033	0.098
	Medium	3.506	3.510	4.037	4.933	−0.579	0.599	0.041	0.118
	High	3.485	3.489	4.204	5.232	−0.592	0.613	0.034	0.129
3 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, N_i)$	Low	5.445	5.446	7.005	7.048	0.380	0.399	−0.035	0.091
	Medium	5.455	5.456	7.009	7.050	0.381	0.401	−0.033	0.091
	High	5.441	5.442	7.054	7.096	0.381	0.400	−0.033	0.099
3 $(Z_i \not\!\perp\!\!\!\perp G_i \mid \mathbf{X}_i^{z}, \mathbf{X}_i^{\text{neigh}})$	Low	3.002	3.002	2.577	2.584	−1.201	1.202	0.071	0.111
	Medium	3.002	3.002	2.556	2.563	−1.202	1.203	0.068	0.111
	High	3.005	3.005	2.548	2.555	−1.200	1.201	0.066	0.111

Table 3: Estimation of $\Delta(0)$

Spillover Effects

Scenario	Interference	Unadjusted regression $\sim Z_i, G_i$		Subclass $\hat{\phi}(1, \mathbf{X}_i^z)$		GPS		Subclass $\hat{\phi}(1, \mathbf{X}_i^z)$ and GPS	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}})$	Low	3.195	3.196	2.870	2.878	0.432	0.441	0.003	0.055
	Medium	3.197	3.198	2.863	2.871	0.437	0.449	0.002	0.083
	High	3.198	3.200	2.869	2.877	0.434	0.452	0.004	0.103
2 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, \mathbf{X}_i^{\text{neigh}})$	Low	2.483	2.485	2.842	2.862	0.386	0.393	0.001	0.059
	Medium	2.481	2.485	2.885	2.907	0.385	0.394	−0.000	0.072
	High	2.473	2.477	2.882	2.919	0.379	0.391	−0.005	0.088
3 $(Z_i \perp\!\!\!\perp G_i \mid \mathbf{X}_i^{\text{ind}}, N_i)$	Low	3.668	3.669	5.921	5.951	0.559	0.562	−0.038	0.045
	Medium	3.668	3.668	5.944	5.975	0.558	0.561	−0.040	0.046
	High	3.669	3.669	5.943	5.982	0.558	0.561	−0.038	0.048
3 $(Z_i \not\perp\!\!\!\perp G_i \mid \mathbf{X}_i^z, \mathbf{X}_i^{\text{neigh}})$	Low	−3.280	3.280	−5.182	5.182	0.367	0.370	−0.019	0.056
	Medium	−3.277	−3.277	−5.182	5.182	0.368	0.371	−0.021	0.057
	High	−3.281	3.281	−5.178	5.178	0.364	0.367	−0.026	0.059

Table 4: Estimation of $\Delta(1)$

Spillover Effects

- Figure 2 depicts the scatterplot of the observed outcomes and the estimated ADRFs $\mu(0, g)$ and $\mu(1, g)$, $g \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$,for Scenario 2

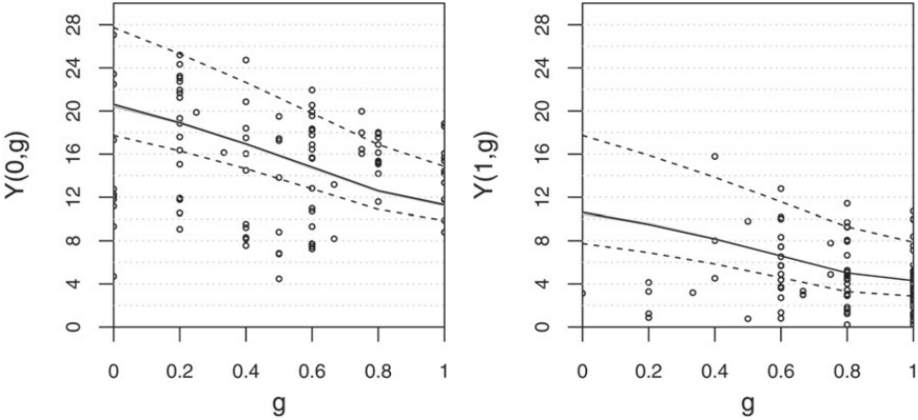

Figure 2: Estimated dose-response function $\mu(0, g)$ and $\mu(1, g)$

Spillover Effects

- Figure 3 depicts spillover effects $\delta(0, g)$ i.e $\mu(0, g) - \mu(0, 0)$ and spillover effects $\delta(1, g)$ i.e $\mu(1, g) - \mu(1, 0)$

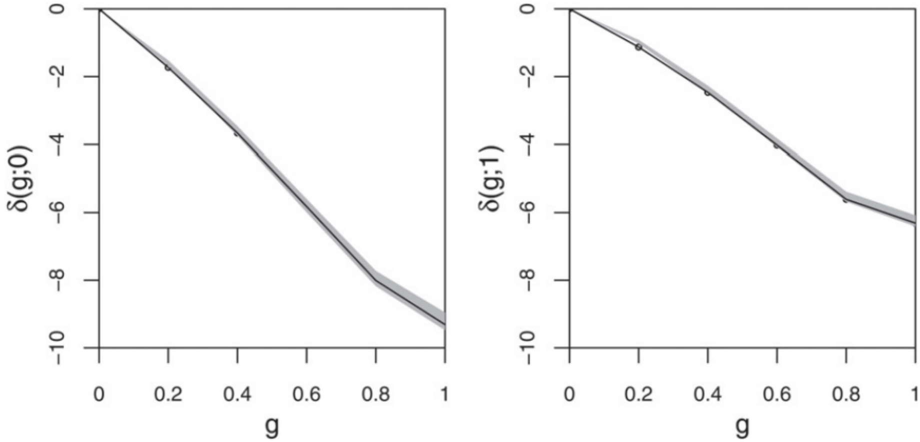

Figure 3: Estimated spillover effects $\delta(g; z)$

Spillover Effects

- Figure 4 depicts estimated main effects $\tau(g)$. i.e $\mu(1, g) - \mu(0, g)$.

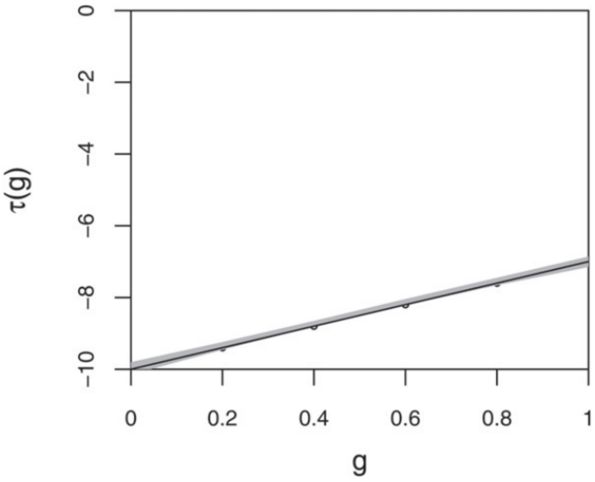

Figure 4: Estimated main effects $\tau(g)$

V . Conclusion

Concluding Remarks

- SUTVA $\longrightarrow$ SUTNVA , potential outcome $Y_i(\mathbf{Z}_i, \mathbf{G}_i)$
- simple assignment mechanism $\longrightarrow$ compound assignment mechanism
- naive estimator $\longrightarrow$ Subclassfication and GPS estimator
- Bias sources : $\mathbf{Z}_{\mathcal{N}_i}$, $P(Z_{i,res}, G_{i,res}|\mathbf{X}_i^\star)$
- limitations : network fully known and fixed , model dependent
- future : sensitivity analysis , Bayesian semi-parametric approaches , account for network uncertainty

Software

- networkinference - Python package by Michael P. Leung
- CausalModel - Python package by Qu, Zhaonan and Xiong, Ruoxuan and Liu, Jizhou and Imbens, Guido
- inferference - R package by Bradley Saul
- clusteredinterference - R package by Brian G. Barkley
- Interference - R package by Georgia Papadogeorgou

Thank You !