Efficient Nonparametric Estimation of Stochastic Policy Effects with Clustered Interference

Chanhwa Lee and Michael Hudgens

Dept of Bios, U of North Carolina @ Chapel Hill

Join work with Prof. Donglin Zeng, Dept of Bios, U of Michigan @ Ann Arbor

Interference

- Interference: Treatment of one individual may affect another's outcome
- Examples: economics, education, infectious diseases, political science, social networks, spatial analyses, ...
- Partial or clustered interference: Individuals may be partitioned into groups (clusters) s.t. there is no interference between individuals in different groups allowing for, but not assuming the existence of, interference within clusters
- Goal: Draw inference about treatment effects that allow for / quantify interference within clusters, if present

Estimands

- Interesting aspect of causal inference with interference; many choices
- Policy Q: counterfactual setting where conditional distribution of cluster's treatment $\mathbf{A}$ given cluster's covariates $\mathbf{X}$ is Q
- $\mu(Q)$ expected potential outcome (PO) under policy Q
- $\mu_a(Q)$ expected PO when individual receives a under policy Q for $a = 0, 1$
- Effects defined by contrasts such as indirect / spillover effect $\mu_0(Q) - \mu_0(Q')$

Two-stage randomization

- What would be ideal randomized experiment when partial/cluster interference might be present?
- Two-stage randomization (Hayes et al. 2000, Longini et al. 2002, Borm et al. 2005, Sinclair et al. 2012, Ichino and Schündeln 2012, Basse and Feller 2017, ...)
    1. Groups to policies (allocation strategies), $\{Q_0, Q_1\}$
    2. Given 1, randomize individuals to treatment/control, $\{0, 1\}$

Cash Transfer Program (Baird et al. 2014)

- Groups: enumerations areas (EAs) in Zomba district of Malawi
- Individuals: never married females ages 13-22
- Assignment mechanism
  1. Randomized EAs to 0%, 33%, 66% or 100% saturation
  2. Randomized participants to cash transfer conditional on EA assignment from step 1

Two-staged Randomized Experiments

- Inference in two-staged randomized studies: H. and Halloran (2008); Tchetgen Tchetgen and VanderWeele (2012); Liu and H. (2014); Baird et al. (2014); Rigdon and H. (2015); Basse and Feller (2018); and many others
- No interference and independence btwn clusters; no confounding

Partial Interference and Observational Data

- Methods for observational studies to allow partial interference: assume (i) random sample from super-population of groups and (ii) no unmeasured confounders
- Tchetgen Tchetgen and VanderWeele (2012), Perez-Heydrich et al. (2014) and Liu et al. (2016) consider IPW estimators, with weights based on inverse of group-level propensity score
- Liu et al. (2019) DR estimators; Park and Kang (2022) sp efficient estimators
- Chakladar et al. (2022) right censoring, inverse probability of censoring weights

Policy-relevant Estimands

- These methods generally target same estimands as two-stage randomized experiments, i.e., scenarios where individuals independently select treatment w/ same probability, aka, type B policy
- In observational setting, other counterfactual policies may be more relevant, e.g., if interference within clusters, might expect within cluster treatment selection dependence

Policy-relevant Estimands

- Papadogeorgou et al (*Biometrics* 2019) and Barkley et al (*Ann of App Stat* 2020) estimands shift/modify treatment distribution: (i) permits treatment selection dependence within clusters, (ii) preserves ranking of individuals within clusters by probability of treatment, (iii) based on assumed parametric model of $\mathbf{A} \mid \mathbf{X}$
- Lee, Zeng, Hudgens, 2023. Efficient Nonparametric Estimation of Stochastic Policy Effects with Clustered Interference. *arXiv*
  General class of estimands that do not require parametric modeling of $\mathbf{A} \mid \mathbf{X}$
  Nonparametric sample splitting estimators, flexible data-adaptive estimation of nuisance functions, CAN at the usual parametric rate

Definitions

Observed data

- m clusters, N_i individuals in cluster $i \in \{1, \ldots, m\}$
- Unit j in cluster i,
  $Y_{ij} \in \mathbb{R}$: outcome, $A_{ij} \in \{0, 1\}$: treatment, $\mathbf{X}_{ij} \in \mathbb{R}^p$: covariates
- $\mathbf{O}_i = (\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, N_i)$: observed data for cluster i
- $\mathcal{A}(N_i) = \{0, 1\}^{N_i}$: set of all length N_i binary vectors

Potential outcome

- $Y_{ij}(\mathbf{a}_i)$: potential outcome for unit j in cluster i when individuals in the cluster receives treatment assignment according to $\mathbf{a}_i \in \mathcal{A}(N_i)$
- $Y_{ij}(\mathbf{a}_i) = Y_{ij}(a_{ij}, \mathbf{a}_{i(-j)})$, $\mathbf{a}_{i(-j)} = (a_{i1}, \ldots, a_{i(j-1)}, a_{i(j+1)}, \ldots, a_{iN_i})$
- **No interference**: $Y_{ij}(a_{ij}, \mathbf{a}_{i(-j)}) = Y_{ij}(a_{ij}, \mathbf{a}'_{i(-j)})$

Treatment allocation policy

- Counterfactual scenario that a cluster of size N_i with cluster-level covariate $\mathbf{X}_i$ receives treatment $\mathbf{a}_i \in \mathcal{A}(N_i)$ with probability $Q(\mathbf{a}_i|\mathbf{X}_i, N_i)$.
- $Q(\cdot|\mathbf{X}_i, N_i)$: probability dist'n on $\mathcal{A}(N_i)$
- Deterministic policy

$$Q_{\text{All}}(\mathbf{a}_i|\mathbf{X}_i, N_i) = \prod_{j=1}^{N_i} \mathbb{1}(a_{ij} = 1)$$

- Type B policy (Tchetgen Tchetgen & VanderWeele 2012)

$$Q_{\text{B}}(\mathbf{a}_i|\mathbf{X}_i, N_i; \alpha) = \prod_{j=1}^{N_i} \alpha^{a_{ij}}(1-\alpha)^{1-a_{ij}}$$

- GLMM shift policy (Papadogeorgou et al. 2019, Barkley et al. 2020)

$$P(\mathbf{A}_i|\mathbf{X}_i, N_i) = \int \prod_{j=1}^{N_i}\{g(\mathbf{X}_{ij}^\top\beta + u)\}^{A_{ij}}\{1 - g(\mathbf{X}_{ij}^\top\beta + u)\}^{1-A_{ij}} dF(u)$$

$$Q_{\text{GLMM}}(\mathbf{a}_i|\mathbf{X}_i, N_i; \gamma) = \int \prod_{j=1}^{N_i}\{g(\mathbf{X}_{ij}^\top\gamma + u)\}^{a_{ij}}\{1 - g(\mathbf{X}_{ij}^\top\gamma + u)\}^{1-a_{ij}} dF(u)$$

- CIPS policy (Kennedy 2019, Lee et al. 2023)

$$Q_{\text{CIPS}}(\mathbf{a}_i|\mathbf{X}_i, N_i; \delta) = \prod_{j=1}^{N_i}(\pi_{ij,\delta})^{a_{ij}}(1-\pi_{ij,\delta})^{1-a_{ij}}$$

  - ▷ $\pi_{ij} = \mathbb{P}(A_{ij} = 1|\mathbf{X}_i, N_i)$ propensity score
  - ▷ $\pi_{ij,\delta} = \delta(\mathbf{X}_i, N_i)\pi_{ij}/\{\delta(\mathbf{X}_i, N_i)\pi_{ij} + 1 - \pi_{ij}\}$ shifted propensity score
  - ▷ $\delta(\mathbf{X}_i, N_i)$ user-specified known function
    e.g., $\delta(\mathbf{X}_i, N_i) = \delta_0$, $\delta(\mathbf{X}_i, N_i) = \delta_0(1 + 1/N_i)$
  - ▷ Shifting the propensity score distribution s.t

$$\frac{\pi_{ij,\delta}}{1-\pi_{ij,\delta}} \Big/ \frac{\pi_{ij}}{1-\pi_{ij}} = \delta(\mathbf{X}_i, N_i)$$

  - ▷ Risk of COVID19 when the odds of vaccination were 2 times the observed

- TPB policy (Lee et al. 2023)

$$Q_{\text{TPB}}(\mathbf{a}_i|\mathbf{X}_i, N_i; \rho) = \mathbb{1}(\overline{\mathbf{a}}_i \geq \rho)\frac{\mathbb{P}(\mathbf{a}_i|\mathbf{X}_i, N_i)}{\sum_{\overline{\mathbf{a}}'_i \geq \rho} \mathbb{P}(\mathbf{a}'_i|\mathbf{X}_i, N_i)}$$

  - $\overline{\mathbf{a}}_i = N_i^{-1} \sum_{j=1}^{N_i} a_{ij}$,
  - $\mathbb{P}(\mathbf{a}_i|\mathbf{X}_i, N_i) = \mathbb{P}(\mathbf{A}_i = \mathbf{a}_i|\mathbf{X}_i, N_i)$ observed joint probability of treatment
  - Counterfactual scenario that the proportion of treated individuals in each cluster is at least $\rho \in [0, 1]$
  - Risk of COVID19 when at least 50% of individuals in each city are vaccinated

Estimands

- Expected average potential outcome under policy Q

$$\mu(Q) = \mathbb{E}\left\{N_i^{-1}\sum_{j=1}^{N_i}\sum_{\mathbf{a}_i\in\mathcal{A}(N_i)} Y_{ij}(\mathbf{a}_i)Q(\mathbf{a}_i|\mathbf{X}_i, N_i)\right\}$$

- Expected average potential outcome **when treated** under policy Q

$$\mu_1(Q) = \mathbb{E}\left\{N_i^{-1}\sum_{j=1}^{N_i}\sum_{\mathbf{a}_{i(-j)}\in\mathcal{A}(N_i-1)} Y_{ij}(1, \mathbf{a}_{i(-j)})Q(\mathbf{a}_{i(-j)}|\mathbf{X}_i, N_i)\right\}$$

  - ▷ $Q(\mathbf{a}_{i(-j)}|\mathbf{X}_i, N_i) = Q(1, \mathbf{a}_{i(-j)}|\mathbf{X}_i, N_i) + Q(0, \mathbf{a}_{i(-j)}|\mathbf{X}_i, N_i)$ probability of all units in cluster i other than j receiving treatment $\mathbf{a}_{i(-j)}$ under policy Q.
  - ▷ No interference: $Y_{ij}(1, \mathbf{a}_{i(-j)}) \equiv Y_{ij}(1) \implies \mu_1(Q) \equiv \mathbb{E}\left\{N_i^{-1}\sum_{j=1}^{N_i} Y_{ij}(1)\right\}$

- Expected average potential outcome **when untreated** under policy Q

$$\mu_0(Q) = \mathbb{E}\left\{N_i^{-1}\sum_{j=1}^{N_i}\sum_{\mathbf{a}_{i(-j)}\in\mathcal{A}(N_i-1)} Y_{ij}(0, \mathbf{a}_{i(-j)})Q(\mathbf{a}_{i(-j)}|\mathbf{X}_i, N_i)\right\}$$

Estimands: Causal effects

- $DE(Q) = \mu_1(Q) - \mu_0(Q)$: effect of treatment under policy Q.
  - Vaccine effect on COVID19 when 50% of neighbors vaccinated
- $OE(Q, Q') = \mu(Q) - \mu(Q')$: compares two policies Q and Q' overall
  - Difference between overall COVID19 risk when 50% versus 30% of neighbors vaccinated
- $SE_1(Q, Q') = \mu_1(Q) - \mu_1(Q')$: average potential outcomes **when treated** under policy Q vs. Q'
  - Difference between a **vaccinated** individual's risk of COVID19 when 50% versus 30% of neighbors vaccinated

Assumptions and Identifiability

(A1) *Consistency*: $Y_{ij} = \sum_{\mathbf{a}_i \in \mathcal{A}(N_i)} Y_{ij}(\mathbf{a}_i)\mathbb{1}(\mathbf{A}_i = \mathbf{a}_i)$

(A2) *Conditional Exchangeability*: $Y_{ij}(\mathbf{a}_i) \perp\!\!\!\perp \mathbf{A}_i | \mathbf{X}_i, N_i$ for all $\mathbf{a}_i \in \mathcal{A}(N_i)$

(A3) *Positivity*: $\mathbb{P}(A_{ij} = 1 | \mathbf{X}_i, N_i) \in (c, 1-c)$ for some $c \in (0,1)$

(A4) *Finite moments*: $\left|\mathbb{E}(Y_{ij}^p | \mathbf{A}_i, \mathbf{X}_i, N_i)\right| \leq C$ for all $p \leq 4$ and some $C < \infty$

(A5) *Finite cluster size*: $\mathbb{P}(N_i \leq n_{\max}) = 1$ for some $n_{\max} \in \mathbb{N}$

Lemma (Identifiability of Causal Estimands)

$$\Psi(w) = \mathbb{E}\left\{\sum_{\mathbf{a}_i \in \mathcal{A}(N_i)} w(\mathbf{a}_i, \mathbf{X}_i, N_i)^\top \mathbb{E}\left(\mathbf{Y}_i | \mathbf{A}_i = \mathbf{a}_i, \mathbf{X}_i, N_i\right)\right\}$$

Nonparametric EIF

Theorem

Assume the EIF of $w(\mathbf{a}, \mathbf{x}, n)$ *is* $\varphi^*_{w(\mathbf{a},\mathbf{x},n)}(\mathbf{O}_i) = \{\mathbb{1}(\mathbf{X}_i = \mathbf{x}, N_i = n)/d\mathbb{P}(\mathbf{x}, n)\}$ $\phi(\mathbf{A}_i, \mathbf{X}_i, N_i; \mathbf{a})$ *for fixed* $(\mathbf{a}, \mathbf{x}, n) \in \mathcal{A}(n) \times \mathcal{X}(n) \times \mathbb{N}$*. Then, the EIF of* $\Psi(w)$ *is*

$$\varphi^*(\mathbf{O}_i) = \sum\nolimits_{\mathbf{a}_i \in \mathcal{A}(N_i)} \{w(\mathbf{a}_i, \mathbf{X}_i, N_i) + \phi(\mathbf{A}_i, \mathbf{X}_i, N_i; \mathbf{a}_i)\}^\top \mathbb{E}(\mathbf{Y}_i | \mathbf{A}_i = \mathbf{a}_i, \mathbf{X}_i, N_i) + \mathbb{P}(\mathbf{A}_i | \mathbf{X}_i, N_i)^{-1} w(\mathbf{A}_i, \mathbf{X}_i, N_i)^\top \{\mathbf{Y}_i - \mathbb{E}(\mathbf{Y}_i | \mathbf{A}_i, \mathbf{X}_i, N_i)\} - \Psi(w).$$

Corollary

$$\varphi^*_{\mu(Q)}(\mathbf{O}_i) = \sum\nolimits_{\mathbf{a}_i \in \mathcal{A}(N_i)} \{Q(\mathbf{a}_i | \mathbf{X}_i, N_i) + \phi_Q(\mathbf{A}_i, \mathbf{X}_i, N_i; \mathbf{a}_i)\} \mathbb{E}(\overline{\mathbf{Y}}_i | \mathbf{A}_i = \mathbf{a}_i, \mathbf{X}_i, N_i) + \frac{Q(\mathbf{A}_i | \mathbf{X}_i, N_i)}{\mathbb{P}(\mathbf{A}_i | \mathbf{X}_i, N_i)} \{\overline{\mathbf{Y}}_i - \mathbb{E}(\overline{\mathbf{Y}}_i | \mathbf{A}_i, \mathbf{X}_i, N_i)\} - \mu(Q)$$

Nuisance functions

Nuisance functions $\boldsymbol{\eta} = (G, H, w, \phi)$:

1. Cluster outcome regression $G(\mathbf{a}_i, \mathbf{x}_i, n_i) = \mathbb{E}\big(\mathbf{Y}_i | \mathbf{A}_i = \mathbf{a}_i, \mathbf{X}_i = \mathbf{x}_i, N_i = n_i\big)$
2. Cluster treatment probability $H(\mathbf{a}_i, \mathbf{x}_i, n_i) = \mathbb{P}\big(\mathbf{A}_i = \mathbf{a}_i | \mathbf{X}_i = \mathbf{x}_i, N_i = n_i\big)$
3. Weight function $w(\mathbf{a}_i, \mathbf{x}_i, n_i)$
4. EIF of the weight function $\phi(\mathbf{a}'_i, \mathbf{x}_i, n_i; \mathbf{a}_i)$

Parametric: GLMM

Data-adaptive: Mixed effect ML (Ngufor et al. 2019), Smoothed kernel regression for dependent data (Park & Kang 2022), etc.

Nuisance functions

If $Y_{ij} \perp\!\!\!\perp Y_{ik} | \mathbf{A}_i, \mathbf{X}_i, N_i$ and $A_{ij} \perp\!\!\!\perp A_{ik} | \mathbf{X}_i, N_i$ (not necessarily marginal indep.),

1. Individual outcome regression

$$\begin{aligned} g(j, \mathbf{a}_i, \mathbf{x}_i, n_i) =& \mathbb{E}\big(Y_{ij} | \mathbf{A}_i = \mathbf{a}_i, \mathbf{X}_i = \mathbf{x}_i, N_i = n_i\big) \\ =& \mathbb{E}\big(Y_{ij} | A_{ij} = a_{ij}, \overline{\mathbf{A}}_{i(-j)} = \overline{\mathbf{a}}_{i(-j)}, \mathbf{X}_{ij} = \mathbf{x}_{ij}\big) =: g^*(a_{ij}, \overline{\mathbf{a}}_{i(-j)}, \mathbf{x}_{ij}) \end{aligned}$$

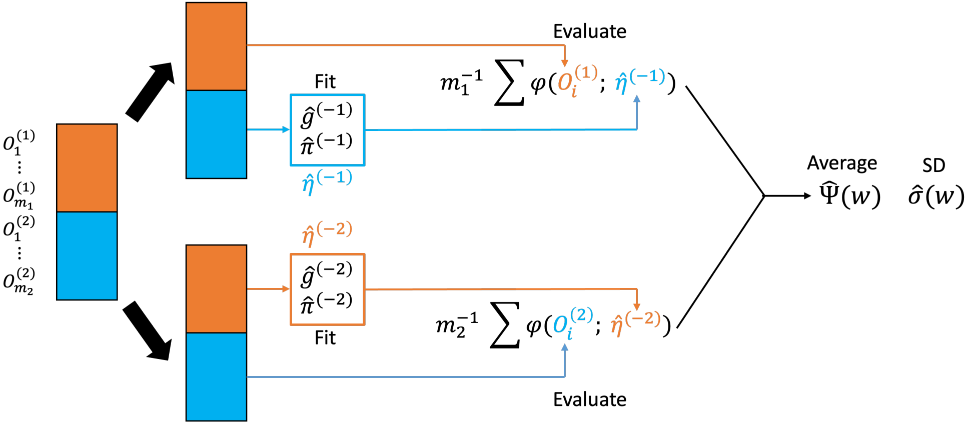
2. Individual propensity score

$$\begin{aligned} \pi(j, \mathbf{x}_i, n_i) =& \mathbb{P}(A_{ij} = 1 | \mathbf{X}_i = \mathbf{x}_i, N_i = n_i) \\ =& \mathbb{P}(A_{ij} = 1 | \mathbf{X}_{ij} = \mathbf{x}_{ij}) =: \pi^*(\mathbf{x}_{ij}) \end{aligned}$$

Parametric: GLM

Data-adaptive: SVM, RF, SuperLearner (van der Laan et al. 2007), etc.

Sample Splitting Estimator



Sample Splitting Estimator

$$\widehat{\Psi}(w) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{m_k}\sum_{i:S_i=k}\varphi(\mathbf{O}_i;\widehat{\boldsymbol{\eta}}^{(-k)})$$
$$\widehat{\sigma}^2(w) = \frac{1}{K}\sum_{k=1}^{K}\left\{\frac{1}{m_k}\sum_{i:S_i=k}\varphi(\mathbf{O}_i;\widehat{\boldsymbol{\eta}}^{(-k)}) - \widehat{\Psi}(w)\right\}^2$$

- $\sum_{\mathbf{a}_i\in\mathcal{A}(N_i)}\widehat{h}^{(-k)}(\mathbf{O}_i,\mathbf{a}_i)$ can be computationally intensive. Approximate this by $2^{N_i}\sum_{q=1}^{r}\widehat{h}^{(-k)}(\mathbf{O}_i,\mathbf{a}_i^{(q)})/r$ where $\mathbf{a}_i^{(q)}$ $(q=1,\ldots,r)$ is randomly sampled from $\mathcal{A}(N_i)$.
- Specific sample split introduces finite sample variability. Repeat sample splitting S times and then take the median of S estimators to get a split-robust estimator (Chernozhukov et al. 2018).

Theoretical results

$$\widehat{\Psi}(w) - \Psi(w) = \frac{1}{K}\sum_{k=1}^{K}\Big[(\mathbb{P}_m^k - \mathbb{P})\varphi(\mathbf{O};\boldsymbol{\eta}) + (\mathbb{P}_m^k - \mathbb{P})\big\{\varphi(\mathbf{O};\widehat{\boldsymbol{\eta}}^{(-k)}) - \varphi(\mathbf{O};\boldsymbol{\eta})\big\}$$
$$+ \mathbb{P}\big\{\varphi(\mathbf{O};\widehat{\boldsymbol{\eta}}^{(-k)}) - \varphi(\mathbf{O};\boldsymbol{\eta})\big\}\Big]$$

- $(\mathbb{P}_m^k - \mathbb{P})\varphi(\mathbf{O};\boldsymbol{\eta}) \sim N\big(0, \sigma^2(w)\big)$: CLT
- $(\mathbb{P}_m^k - \mathbb{P})\big\{\varphi(\mathbf{O};\widehat{\boldsymbol{\eta}}^{(-k)}) - \varphi(\mathbf{O};\boldsymbol{\eta})\big\} = O_{\mathbb{P}}\left(||\varphi(\mathbf{O};\widehat{\boldsymbol{\eta}}^{(-k)}) - \varphi(\mathbf{O};\boldsymbol{\eta})||/m_k^{1/2}\right)$
- $\mathbb{P}\big\{\varphi(\mathbf{O};\widehat{\boldsymbol{\eta}}^{(-k)}) - \varphi(\mathbf{O};\boldsymbol{\eta})\big\} = r_w^2 + r_G r_H + r_G r_\phi$

Conditions

(B1) *Bounded* H and $\widehat{H}^{(-k)}$: $H(\mathbf{a}, \mathbf{x}, n) \in (c, 1-c)$ and $\widehat{H}^{(-k)}(\mathbf{a}, \mathbf{x}, n) \in (c, 1-c)$

(B2) *Bounded* G and $\widehat{G}^{(-k)}$: $\left\| G(\mathbf{a}, \mathbf{x}, n) \right\|_2 \leq C$ and $\left\| \widehat{G}^{(-k)}(\mathbf{a}, \mathbf{x}, n) \right\|_2 \leq C$

(B3) *Bounded* ϕ and $\widehat{\phi}^{(-k)}$: $\left\| \phi(\mathbf{a}', \mathbf{x}, n; \mathbf{a}) \right\|_2 \leq C$ and $\left\| \widehat{\phi}^{(-k)}(\mathbf{a}', \mathbf{x}, n; \mathbf{a}) \right\|_2 \leq C$

(B4) *Bounded* w and $\widehat{w}^{(-k)}$: $\left\| w(\mathbf{a}, \mathbf{x}, n) \right\|_2 \leq C$ and $\left\| \widehat{w}^{(-k)}(\mathbf{a}, \mathbf{x}, n) \right\|_2 \leq C$

(B5) *Convergence rate of* $\widehat{H}^{(-k)}$: $\left\| \sum_{\mathbf{a} \in \mathcal{A}(N)} \left| \left(\widehat{H}^{(-k)} - H \right)(\mathbf{a}, \mathbf{X}, N) \right| \right\|_{L_2(\mathbb{P})} = O_{\mathbb{P}}(r_H)$

(B6) *Convergence rate of* $\widehat{G}^{(-k)}$: $\left\| \sum_{\mathbf{a} \in \mathcal{A}(N)} \left\| \left(\widehat{G}^{(-k)} - G \right)(\mathbf{a}, \mathbf{X}, N) \right\|_2 \right\|_{L_2(\mathbb{P})} = O_{\mathbb{P}}(r_G)$

(B7) *Convergence rate of* $\widehat{\phi}^{(-k)}$: $\left\| \sum_{\mathbf{a} \in \mathcal{A}(N)} \left\| \left(\widehat{\phi}^{(-k)} - \phi \right)(\mathbf{A}, \mathbf{X}, N; \mathbf{a}) \right\|_2 \right\|_{L_2(\mathbb{P})} = O_{\mathbb{P}}(r_\phi)$

(B8) *Second order convergence rate of* $\widehat{w}^{(-k)}$:

$$\left\| \sum_{\mathbf{a} \in \mathcal{A}(N)} \left\| \left(\widehat{w}^{(-k)} - w \right)(\mathbf{a}, \mathbf{X}, N) + \sum_{\mathbf{a}' \in \mathcal{A}(N)} \widehat{\phi}^{(-k)}(\mathbf{a}', \mathbf{X}, N; \mathbf{a}) H(\mathbf{a}', \mathbf{X}, N) \right\|_2 \right\|_{L_2(\mathbb{P})} = O_{\mathbb{P}}(r_w^2)$$

Theoretical results

Theorem

Under the mild conditions s.t nuisance function estimators have convergence rate of $m^{-1/4}$, *then* $\sqrt{m}\{\widehat{\Psi}(w) - \Psi(w)\}/\widehat{\sigma}(w) \xrightarrow{d} N(0,1)$. *Also,* $\widehat{\sigma}^2(w) \xrightarrow{p} \sigma^2(w)$, *where* $\sigma^2(w) = \mathbb{E}\left[\left\{\varphi^*(\mathbf{O};\boldsymbol{\eta})\right\}^2\right]$ *is the nonparametric efficiency bound of* $\Psi(w)$. *Thus, the proposed estimators are consistent, asymptotically normal, and nonparametric efficient.*

Examples

Policy	Consistency	Asymptotic Normality	Consistent Variance Estimator	Notes
Type B	$r_G = o(1)$ or $r_H = o(1)$	$r_G \cdot r_H = o(m^{-1/2})$	$r_G = r_H = o(1)$	$\phi = 0$, DR
CIPS	$r_\pi = o(1)$	$r_\pi = o(m^{-1/4})$, $r_\pi \cdot r_g = o(m^{-1/2})$	$r_\pi = r_g = o(1)$	Individual level nuisance functions
TPB	$r_H = o(1)$	$r_H = o(m^{-1/4})$, $r_H \cdot r_G = o(m^{-1/2})$	$r_H = r_G = o(1)$	

Examples

Theorem

Consider the collection of type B policies indexed by $\alpha \in \mathbb{A} = [\alpha_l, \alpha_u]$, *where* $0 < \alpha_l < \alpha_u < 1$. *Then,* $\sqrt{m}\{\widehat{\mu}_{\text{B}}(\cdot) - \mu_{\text{B}}(\cdot)\} \rightsquigarrow \mathbb{G}(\cdot)$ *in* $\ell^\infty(\mathbb{A})$ *as* $m \to \infty$, *where* $\mathbb{G}(\cdot)$ *is a mean zero Gaussian process with covariance* $\mathbb{E}\{\mathbb{G}(\alpha)\mathbb{G}(\alpha')\} = \mathbb{E}\{\varphi^*_{\mu_{\text{B}}(\alpha)}(\mathbf{O})\varphi^*_{\mu_{\text{B}}(\alpha')}(\mathbf{O})\}$ *where* $\varphi^*_{\mu_{\text{B}}(\alpha)}(\mathbf{O})$ *is the EIF of* $\mu_{\text{B}}(\alpha)$.

Theorem

Consider the collection of CIPS policies with constant $\delta(\mathbf{X}_i, N_i) = \delta_0$ *indexed by* $\delta_0 \in \mathbb{D} = [\delta_l, \delta_u]$, *where* $0 < \delta_l < \delta_u < \infty$. *Then,* $\sqrt{m}\{\widehat{\mu}_{\text{CIPS}}(\cdot) - \mu_{\text{CIPS}}(\cdot)\} \rightsquigarrow \mathbb{G}(\cdot)$ *in* $\ell^\infty(\mathbb{D})$ *as* $m \to \infty$, *where* $\mathbb{G}(\cdot)$ *is a mean zero Gaussian process with covariance* $\mathbb{E}\{\mathbb{G}(\delta_0)\mathbb{G}(\delta_0')\} = \mathbb{E}\{\varphi^*_{\mu_{\text{CIPS}}(\delta_0)}(\mathbf{O})\varphi^*_{\mu_{\text{CIPS}}(\delta_0')}(\mathbf{O})\}$ *where* $\varphi^*_{\mu_{\text{CIPS}}(\delta_0)}(\mathbf{O})$ *is the EIF of* $\mu_{\text{CIPS}}(\delta_0)$.

Simulations

- $D = 1000$ simulations, each consisted of $m = 500$ clusters
- $N_i \overset{iid}{\sim} Unif\{5, 6, \ldots, 20\}, i = 1, \ldots, m$
- $C_i \sim \text{N}(0, 1)$: one cluster-level covariate
- $X_{ij1} \sim \text{N}(0, 1)$, $X_{ij2} \sim \text{Bernoulli}(0.5)$: individual-level covariates
- $A_{ij} \sim \text{Bernoulli}(\pi_{ij})$: treatment status, $Y_{ij} \sim \text{Bernoulli}(g_{ij})$: outcome
  $\pi_{ij} = \text{expit}(0.1 + 0.2|X_{ij1}| + 0.2|X_{ij1}|X_{ij2} + 0.1\mathbb{1}(C_i > 0))$
  $g_{ij} = \text{expit}(3 - 2A_{ij} - \overline{\mathbf{A}}_{i(-j)} - 1.5|X_{ij1}| + 2X_{ij2} - 3|X_{ij1}|X_{ij2} - 2\mathbb{1}(C_i > 0))$
- CIPS policy with constant $\delta(\mathbf{X}_i, N_i) = \delta_0 \in \{0.5, 1, 2\}$
- Sample splitting with $K = 2$, $r = 100$, $S = 1$
- Nuisance functions estimation
  - Nonparametric: SuperLearner (logistic reg, RF, GAM, single-layer NN)
  - Parametric: logistic reg

Simulations

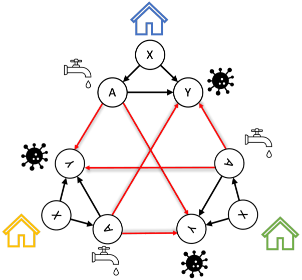
Table 1: Simulation results for nonparametric and parametric sample splitting estimators for CIPS policy with constant δ

		Nonparametric					Parametric				
Estimand	Truth	Bias	RMSE	ASE	ESE	Cov %	Bias	RMSE	ASE	ESE	Cov %
$\mu(2)$	0.300	-0.003	0.013	0.013	0.013	94.9%	-0.010	0.017	0.013	0.014	87.1%
$\mu_1(2)$	0.224	-0.004	0.014	0.014	0.014	93.7%	-0.017	0.022	0.014	0.015	75.9%
$\mu_0(2)$	0.507	0.003	0.018	0.017	0.017	94.5%	0.025	0.031	0.019	0.019	75.6%
$DE(2)$	-0.283	-0.007	0.019	0.019	0.018	94.2%	-0.042	0.046	0.021	0.020	45.6%
$SE_1(2,1)$	-0.018	-0.002	0.010	0.010	0.010	94.3%	-0.004	0.012	0.011	0.012	93.4%
$SE_0(2,1)$	-0.022	-0.002	0.012	0.012	0.012	94.4%	-0.006	0.015	0.014	0.014	92.5%
$OE(2,1)$	-0.063	-0.003	0.009	0.009	0.009	93.9%	-0.010	0.015	0.010	0.010	77.7%
$TE(2,1)$	-0.306	-0.009	0.017	0.015	0.015	92.1%	-0.047	0.050	0.017	0.017	22.7%

RMSE: root mean squared error, ASE: average standard error estimates, ESE: standard deviation of estimates, Cov %: 95% CI coverage, RMSE Ratio: RMSE ratio of nonparametric and parametric estimators

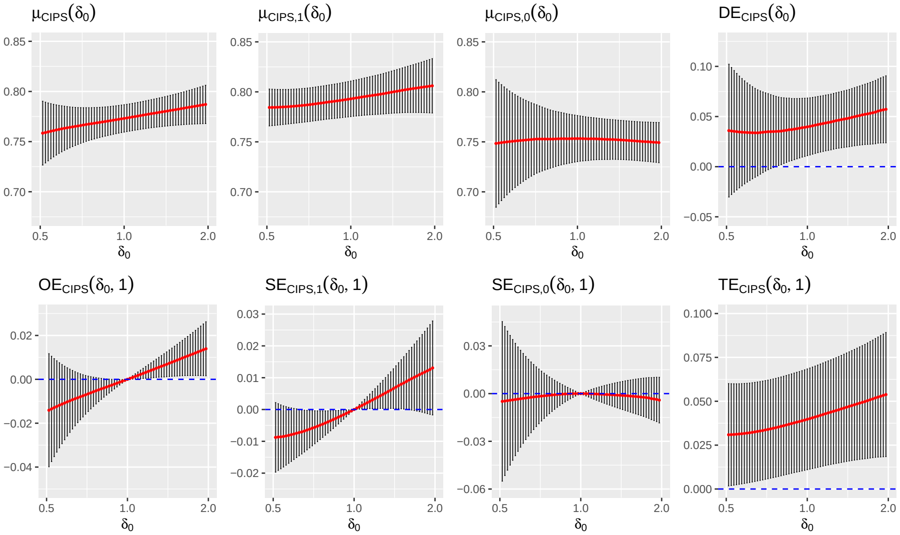
Application to Senegal Demographic and Health Survey

- Whether water, sanitation, and hygiene (WASH) facilities decrease diarrhea incidence among children under clustered interference? (Benjamin-Chung et al. 2018)
- How does diarrhea incidence change if the odds of having WASH facility change? [CIPS policy]
- How does diarrhea incidence change if at least 50% of households have WASH facility? [TPB policy]

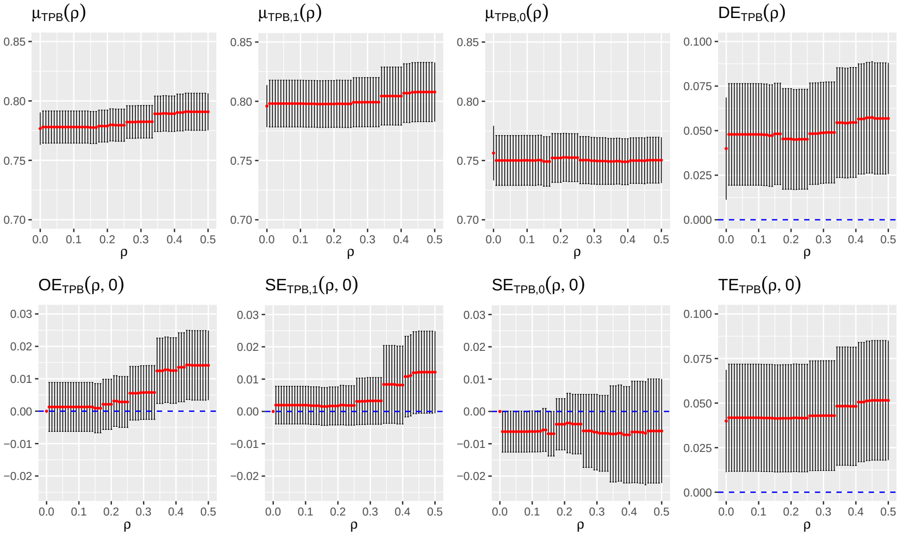


- ▷ Cluster: Census block ($i = 1, \ldots, 1074$)
- ▷ Unit: Household ($j = 2, \ldots, 12$)
- ▷ $Y_{ij} = \mathbb{1}$(All children diarrhea-free)
- ▷ $A_{ij} = \mathbb{1}$(WASH facility)
- ▷ X_{ij} = Demographic, Socioeconomic status
- ▷ Sample splitting estimator with Super Learner estimator including penalized logistic regression, spline regression, GAM, GBM, RF, Neural net

Application to Senegal DHS (cont'd)



Application to Senegal DHS (cont'd)



Discussion

- Nonparametric methods are developed which can be used to draw inference about treatment effects in the presence of clustered interference, can be applied to any treatment allocation policy, allowing for units' propensity to vary by their covariates and are not based on parametric model
- Proposed nonparametric efficient sample splitting estimators exploit a variety of data-adaptive methods, and therefore are robust to model mis-specification compared to parametric estimators
- Application to the Senegal DHS data suggested that having a private water source or flushable toilet decreases the risk of diarrhea among children, and that children from WASH households may receive an additional protective spillover effect from neighboring WASH households

Acknowledgements

Taylor J. Krajewski, Michael Jesupphasuk, Chan Park

This work was supported by NIH grant R01 AI085073.

References

Barkley, B. G., Hudgens, M. G., Clemens, J. D., Ali, M. & Emch, M. E. (2020), 'Causal inference from observational studies with clustered interference, with application to a cholera vaccine study', *The Annals of Applied Statistics* **14**(3), 1432–1448.

Benjamin-Chung, J., Amin, N., Ercumen, A., Arnold, B. F., Hubbard, A. E., Unicomb, L., Rahman, M., Luby, S. P. & Colford Jr, J. M. (2018), 'A randomized controlled trial to measure spillover effects of a combined water, sanitation, and handwashing intervention in rural bangladesh', *American Journal of Epidemiology* **187**(8), 1733–1744.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018), 'Double/debiased machine learning for treatment and structural parameters', *The Econometrics Journal* **21**(1), C1–C68.

Kennedy, E. H. (2019), 'Nonparametric causal effects based on incremental propensity score interventions', *Journal of the American Statistical Association* **114**(526), 645–656.

Lee, C., Zeng, D. & Hudgens, M. G. (2023), 'Efficient nonparametric estimation of stochastic policy effects with clustered interference', *arXiv preprint arXiv:2212.10959v2* .

References

Ngufor, C., Van Houten, H., Caffo, B. S., Shah, N. D. & McCoy, R. G. (2019), 'Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin a1c', *Journal of Biomedical Informatics* **89**, 56–67.

Papadogeorgou, G., Mealli, F. & Zigler, C. M. (2019), 'Causal inference with interfering units for cluster and population level treatment allocation programs', *Biometrics* **75**(3), 778–787.

Park, C. & Kang, H. (2022), 'Efficient semiparametric estimation of network treatment effects under partial interference', *Biometrika* **109**(4), 1015–1031.

Tchetgen Tchetgen, E. J. & VanderWeele, T. J. (2012), 'On causal inference in the presence of interference', *Statistical Methods in Medical Research* **21**(1), 55–75.

van der Laan, M. J., Polley, E. C. & Hubbard, A. E. (2007), 'Super learner', *Statistical Applications in Genetics and Molecular Biology* **6**, 1–21.