Minimax Optimal Nonparametric Estimation of Heterogeneous Treatment Effects

Zijun Gao — Stanford Statistics
Yanjun Han — Stanford EE

NeurIPS 2020

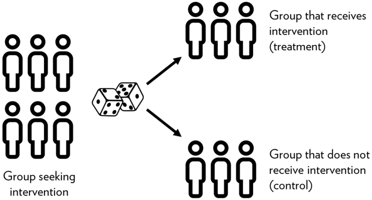
HTE Estimation

Model: Assume n treated, n control units with covariates X_i^L follow

$$Y_i^L = \mu_L(X_i^L) + \varepsilon_i^L, \quad L \in \{t, c\}.$$

Goal: Estimate the heterogeneous treatment effect (HTE)

$$\tau(x) := \mu_t(x) - \mu_c(x).$$



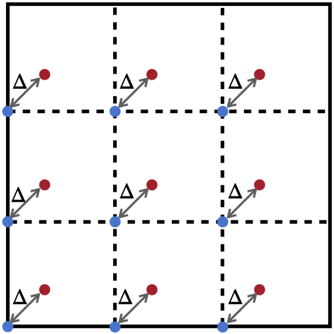
Potential outcome model (Rubin 1974)

Picture from `https://redfworkshop.org/learn/formal-evaluations`

Covariate Design

Fixed design:

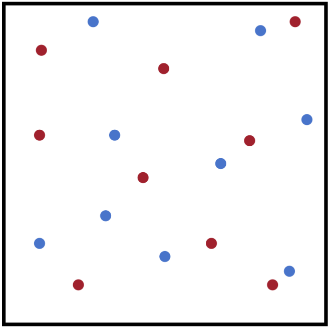
$$X_i^c = \text{Grid points};$$
$$X_i^t = X_i^c + \Delta.$$



Random design:

$$X_i^L \overset{\text{i.i.d.}}{\sim} g_L.$$

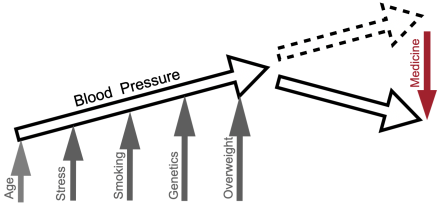
Overlap: $\kappa^{-1} \leq g_c(x)/g_t(x) \leq \kappa$.



Assumptions

- *Smoothness*: μ_t, μ_c are β_μ-smooth, τ is β_τ-smooth,

$$\color{red}{\beta_\mu \leq \beta_\tau \implies \text{Simpler HTE!}}$$



Simpler HTE (Hansen 2008, Kuënzel 2018)

- *Error*: ε_i^L are mutually independent, zero-mean, of variance σ^2.

Literature of HTE Estimation

- (Semi)parametric: μ_t, μ_c (non)parametric, τ parametric.
  - Semiparametrically efficient estimator (Robinson 1988);
  - Debiased approach handling confounders (Chernozhukov et.al. 2018)

- Nonparametric: μ_t, μ_c, τ nonparametric.
  - Optimal non-parametric estimator given crude baseline estimators (Wager et.al. 2017)
  - Optimal non-parametric estimator of modeling μ_t, μ_c not τ (Alaa et.al. 2018)

$\Downarrow$

Optimal non-parametric estimator of modeling μ_c, τ?

Minimax Formulation

Fixed design:

$$R_{\text{fixed}}(n, d, \beta_\mu, \beta_\tau, \sigma, \Delta) \triangleq \inf_{\hat{\tau}} \sup_{\substack{\mu_c \ \beta_\mu\text{-smooth} \\ \tau \ \beta_\tau\text{-smooth}}} \mathbb{E}_{\mu_c,\tau}[\|\hat{\tau} - \tau\|_1].$$

Random design:

$$R_{\text{random}}(n, d, \beta_\mu, \beta_\tau, \sigma, \kappa) \triangleq \inf_{\hat{\tau}} \sup_{\substack{\mu_c \ \beta_\mu\text{-smooth} \\ \tau \ \beta_\tau\text{-smooth} \\ 1/\kappa \le g_c/g_t \le \kappa}} \mathbb{E}_{\mu_c,\tau}[\|\hat{\tau} - \tau\|_{L_1(g_c)}].$$

Fixed Design: Main Result

Theorem (Minimax risk under fixed design)

$$R_{\mathrm{fixed}}(n, d, \beta_\mu, \beta_\tau, \sigma, \Delta) \asymp n^{-\frac{\beta_\mu}{d}} (n^{\frac{1}{d}} \|\Delta\|_\infty)^{\beta_\mu \wedge 1} + \left(\frac{\sigma^2}{n}\right)^{\frac{\beta_\tau}{2\beta_\tau + d}}.$$

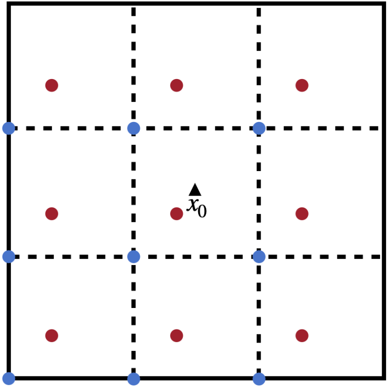
Remark:

- Tight dependence on n, σ, Δ;
- Matching bias + standard nonparametric estimation error;
- Faster than standard non-parametric rate with β_μ.

$\Uparrow$

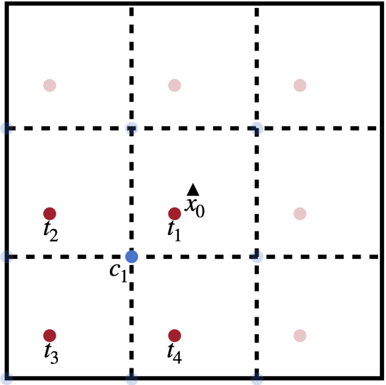
Blessing from simpler HTE!

Fixed Design: Proposed Estimator



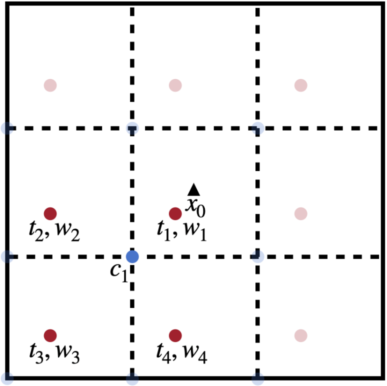
Query point x_0

Fixed Design: Proposed Estimator



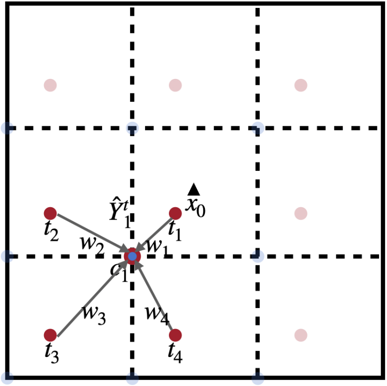
Step 1: For each control, find the nearest $\beta_\mu + 1$ treatment covariates.

Fixed Design: Proposed Estimator



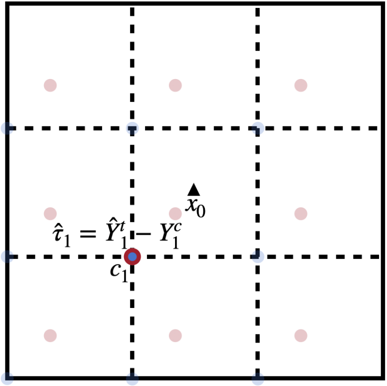
Step 2.a: Compute weights of selected treated.

Fixed Design: Proposed Estimator



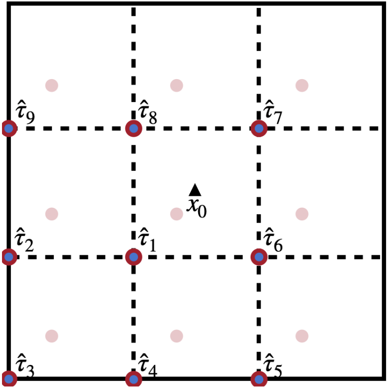
Step 2.b: Weight the responses of selected treated as pseudo-observation.

Fixed Design: Proposed Estimator



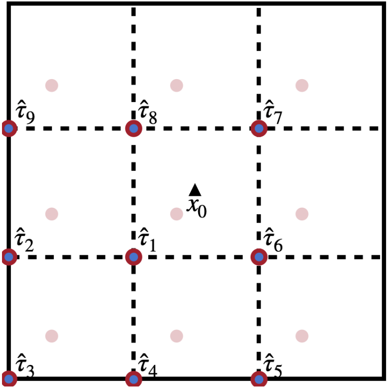
Step 3.a: Take the difference of pseudo-observation and control response.

Fixed Design: Proposed Estimator



Step 3.b: Apply kernel method to the differences with β_τ-smooth bandwidth.

Fixed Design: Proposed Estimator



Insight: Construct pseudo-observations!

Random Design: Main Result

Theorem (Minimax risk under random design)

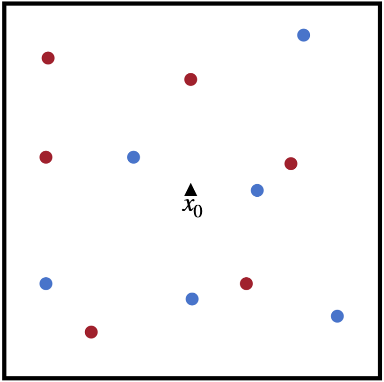
If $\beta_\tau \le 1$,

$$R_{\text{random}}(n, d, \beta_\mu, \beta_\tau, \sigma, \kappa) \asymp \left(\frac{\kappa}{n^2}\right)^{\frac{1}{d(\beta_\mu^{-1}+\beta_\tau^{-1})}} + \left(\frac{\kappa\sigma^2}{n^2}\right)^{\frac{1}{2+d(\beta_\mu^{-1}+\beta_\tau^{-1})}} + \left(\frac{\kappa\sigma^2}{n}\right)^{\frac{\beta_\tau}{2\beta_\tau+d}}.$$

Remark:

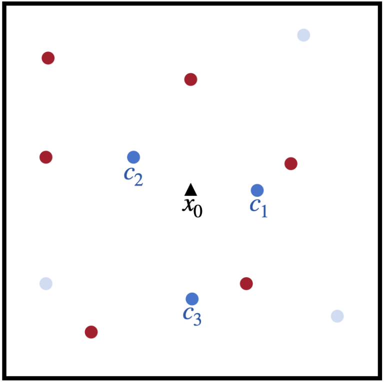
- Tight dependence on n, σ, κ;
- Three sources of errors instead of two;
- Again, faster than standard non-parametric rate with β_μ.

Random Design: Proposed Estimator



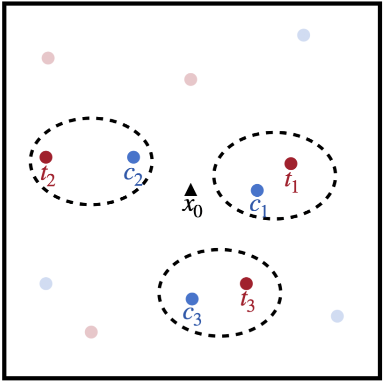
Query point x_0

Random Design: Proposed Estimator



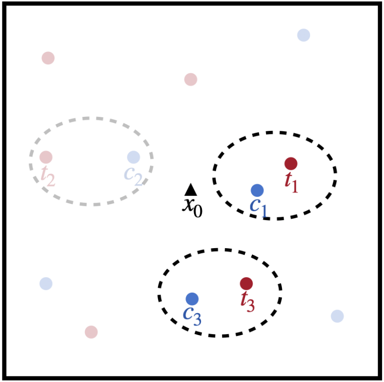
Step 1: Find m_1 nearest control covariates to x_0

Random Design: Proposed Estimator



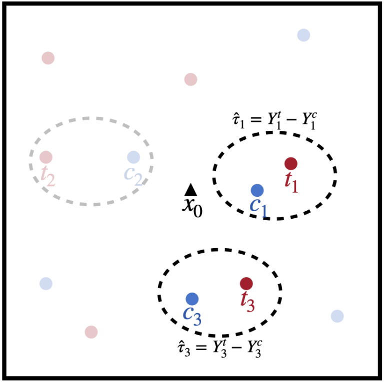
Step 2.a: Find the nearest treatment covariate of each selected control

Random Design: Proposed Estimator



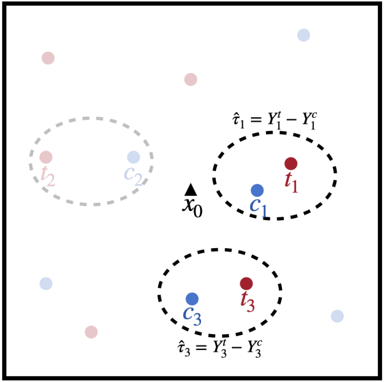
Step 2.b: Select the closest $m_2 \leq m_1$ treatment-control pairs

Random Design: Proposed Estimator



Step 3: Compute the average difference of selected pairs' responses

Random Design: Proposed Estimator



Insight: Construct pseudo-observations

+

Keep only good-quality pairs!

Optimal Parameter Choice

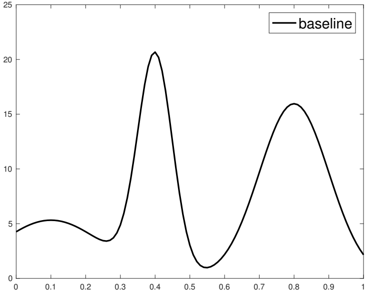
Three sources of errors:

$$R_{\text{random}} \leq \underbrace{\left(\frac{\kappa m_2}{n m_1}\right)^{\frac{\beta_\mu}{d}}}_{\substack{\text{matching bias} \\ \text{of } m_2 \text{ pairs}}} + \underbrace{\left(\frac{m_1}{n}\right)^{\frac{\beta_\tau}{d}}}_{\substack{\text{bias of } m_1 \\ \text{nearest neighbors}}} + \underbrace{\frac{\sigma}{\sqrt{m_2}}}_{\textit{noise}}.$$

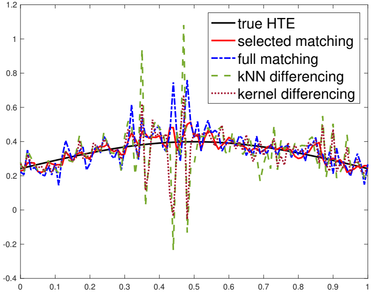
$\Uparrow$

Optimal (m_1, m_2) balance the three errors!

Simulation



Baseline



HTE estimators

Proposed method enjoys the simplicity of HTE!

---

selected matching: the minimax-optimal HTE estimator
full matching: the minimax-optimal HTE estimator keeping all m_1 treatment-control pairs
kNN differencing: difference of kNN estimators of baselines
kernel differencing: difference of kernel estimators of baselines

Thank you

Full paper: arXiv 2002.06471.

Contact: {zijungao, yjhan}@stanford.edu.