Optimal Stratification of Survey Experiments

Max Cytrynbaum

April 1, 2026

Introduction

This paper introduces a new family of **fine stratification** methods to design representative and balanced experiments.

These designs increase the **precision** of treatment effect estimation, making the most efficient use of limited experimental resources.

Experimental Design

Large design literature on how to **assign treatments**.

This paper jointly models two different design decisions.

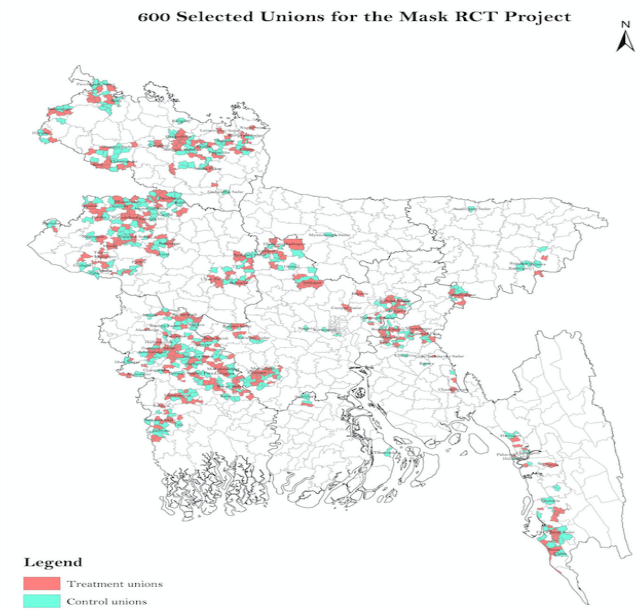
1. **Sampling** of experimental participants.
2. **Assignment** of treatments to the sampled units.

Motivating Example

Abaluck et al. (2021): Effect of mask distribution on COVID-19 infection levels, mask-wearing behavior in villages in Bangladesh.

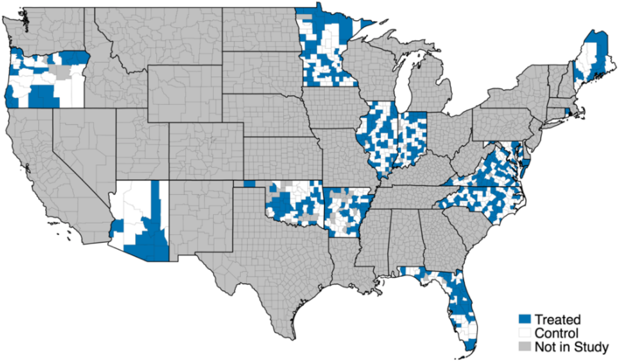
1. Sampling: which villages should be in the experiment?
2. Assignment: how should treatments be assigned to these villages?

Abaluck et al. (2021)



Breza et al. (2021)

Effect of Facebook ads telling people to stay home.



New Design Tools

1. **Sampling** of participants.
2. **Assignment** of treatments.

Local Randomization

1. Stratified randomization within fine strata optimized to reduce estimator variance.
2. Matched **k-tuples**, generalizing matched pairs to assignment propensity $p(x) \neq 1/2$ and sampling propensity $q(x)$.
3. Allows representative sampling, e.g. out of 500 villages, which 100 should you select?

Related Literature

1. Covariate-Adaptive Sampling and Assignment

- Sampling - Cochran (1977), Chen and Rao (2007), Yang et al. (2021).
- Stratification and matched pairs - Bugni et al. (2018), Bai et al. (2021), Bai (2022).
- Rerandomization - Morgan and Rubin (2012), Li et al. (2018).

Contribution: Finely stratified sampling into an experiment with arbitrary sampling propensity $q(x)$. Matched pairs randomization for $p(x) \neq 1/2$, general dimension.

Related Literature

2. Optimal Design / Design with a Pilot.

- Hahn et al. (2012), Bai (2022), Tabord-Meehan (2022), Cai and Shah (2023).

Contribution: Formulate and solve budget-constrained optimal stratification problem, feasible implementation using pilot data.

3. Inference for Stratified Sampling and Assignment.

- "Pairs-of-pairs": Abadie and Imbens (2008), Bugni et al. (2018), Bai et al. (2021)
- de Chaisemartin and Ramirez-Cuellar (2021), Fogarty (2018), Higgins et al. (2015), Imai et al. (2009).

Contribution: Asymptotically exact inference for joint finely stratified sampling and assignment.

Outline

Outline

Design Model

Model.

1. Observe N units with covariates $(X_i)_{i=1}^N$.
2. Select $n = qN$ units with **sampling propensity** $q = a/k$
3. Denote $T_i = 1$ if sampled and $T_i = 0$ otherwise.
4. Assigned to $D_i = 1$ with assignment propensity $p = a'/k'$.
5. Then $Y_i = T_i\big[D_i Y_i(1) + (1 - D_i) Y_i(0)\big]$.

Goal. Estimation and inference on ATE $= E[Y(1) - Y(0)]$ or SATE $= E_N[Y_i(1) - Y_i(0)]$.

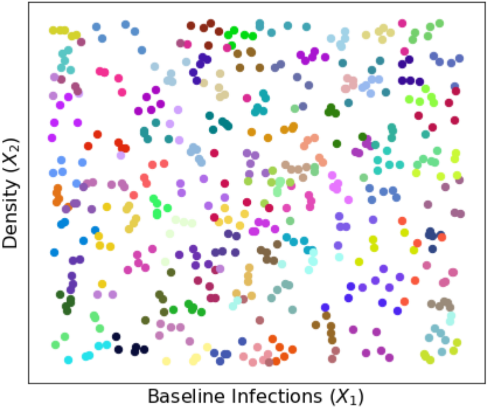
Local Randomization. Match units into groups of size $|g| = k$ using stratification variables $\psi(X) \in \mathbb{R}^d$.

Group Homogeneity. For $\psi \in \mathbb{R}^d$ groups must satisfy

$$\frac{1}{N} \sum_g \sum_{i,j \in g} |\psi_i - \psi_j|_2^2 = o_p(1)$$

Visualizing Homogeneity Condition

$$\frac{1}{N}\sum_{g}\sum_{i,j\in g}|\psi_i - \psi_j|_2^2 = o_p(1) \qquad \psi_i = (\text{Infections}, \text{Educ.})$$



Local Randomization

$T_{1:N}$ are **completely randomized** with probability q if $T_i = 1$ for exactly qN units, uniformly at random

$$T_{1:N} \sim \mathrm{CR}(q)$$

Local Randomization. Complete randomization <u>within</u> groups satisfying homogeneity condition

$$(T_i)_{i \in g} \sim \mathrm{CR}(q) \tag{2.1}$$

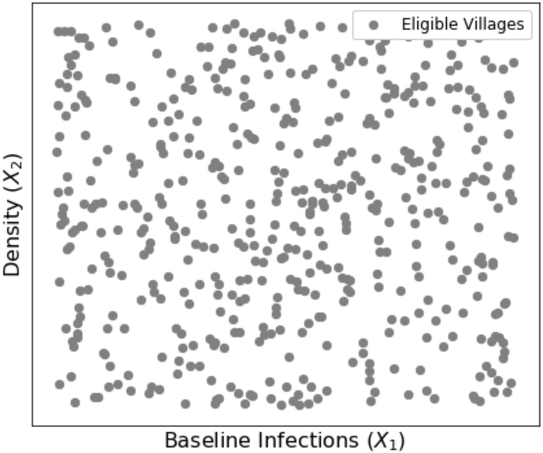
$$\frac{1}{N} \sum_{g} \sum_{i,j \in g} |\psi_i - \psi_j|_2^2 = o_p(1) \tag{2.2}$$

We denote sampling and assignment by

$$T_{1:N} \sim \mathrm{Loc}(\psi, q) \qquad D_{1:N} \sim \mathrm{Loc}(\psi, p)$$

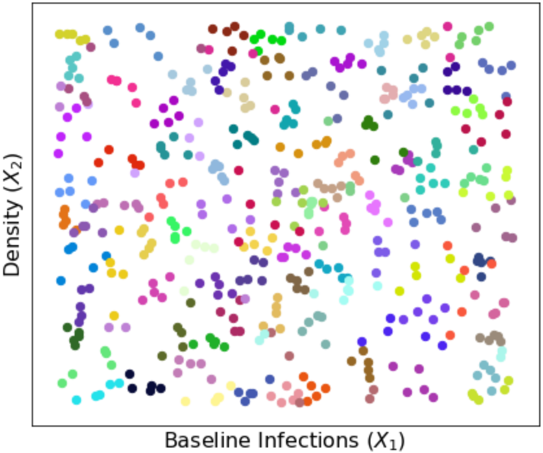
Local Randomization

$$T_{1:N} \sim \text{Loc}(\psi, 1/5) \qquad D_{1:N} \sim \text{Loc}(\psi, 2/3)$$



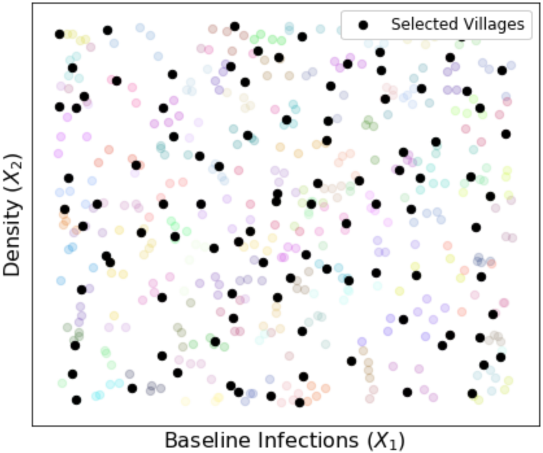
Local Randomization

$$T_{1:N} \sim \text{Loc}(\psi, 1/5) \qquad D_{1:N} \sim \text{Loc}(\psi, 2/3)$$



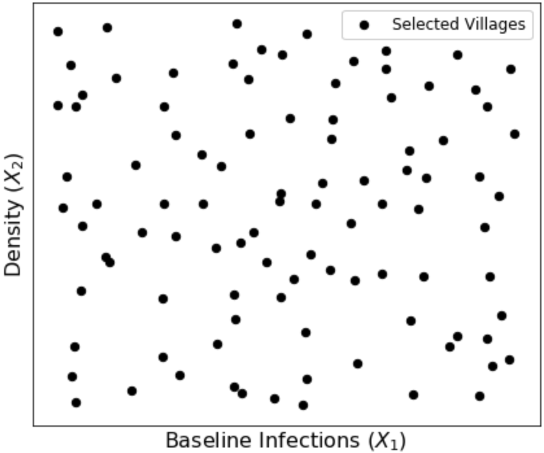
Local Randomization

$$T_{1:N} \sim \text{Loc}(\psi, 1/5) \qquad D_{1:N} \sim \text{Loc}(\psi, 2/3)$$



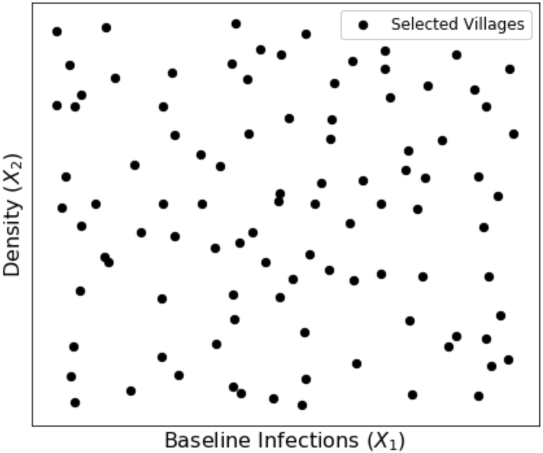
Local Randomization

$$T_{1:N} \sim \text{Loc}(\psi, 1/5) \qquad D_{1:N} \sim \text{Loc}(\psi, 2/3)$$



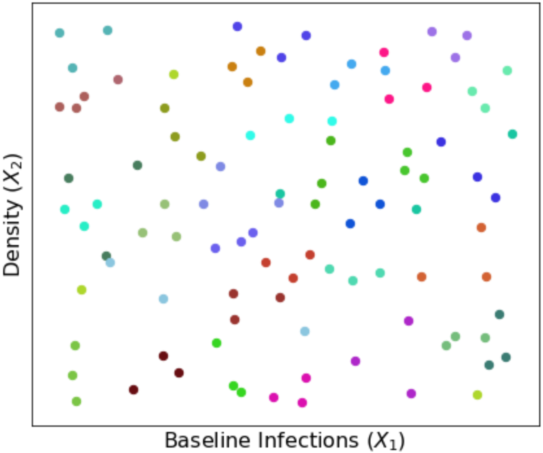
Local Randomization

$$T_{1:N} \sim \text{Loc}(\psi, 1/5) \qquad D_{1:N} \sim \text{Loc}(\psi, 2/3)$$



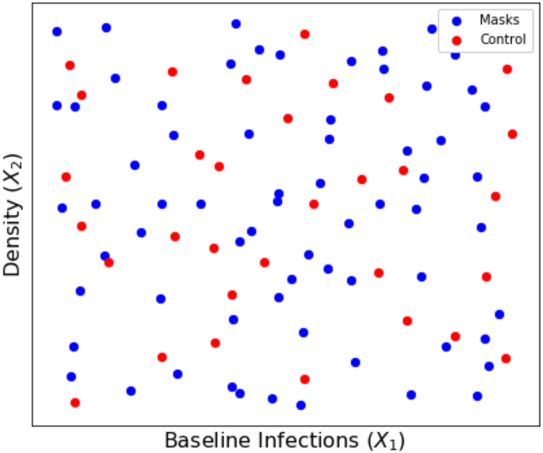
Local Randomization

$T_{1:N} \sim \text{Loc}(\psi, 1/5) \qquad D_{1:N} \sim \text{Loc}(\psi, 2/3)$



Local Randomization

$$T_{1:N} \sim \text{Loc}(\psi, 1/5) \qquad D_{1:N} \sim \text{Loc}(\psi, 2/3)$$



Unified Design Framework

Classical Designs:

- Complete Randomization: $D_{1:N} \sim \text{Loc}(1, p)$
- Coarse Stratification: $D_{1:N} \sim \text{Loc}(\psi, p)$ for $\psi(X) \in \{1, \ldots, K\}$
- Matched Pairs: $D_{1:N} \sim \text{Loc}(\psi, 1/2)$ for $\psi(X) \in \mathbb{R}^d$

Example. $D = 1$ costs more than $D = 0$. Can afford $p = 2/7$

- Let $D_{1:N} \sim \text{Loc}(\psi, 2/7)$ "matched 7-tuples"
- Welfare: e.g. JTPA has $p = 2/3$ "matched triples"

Example. $N = 1000$ villages, can afford $n = 250$.

- Let $T_{1:N} \sim \text{Loc}(\psi, 1/4)$.
- Matched 4-tuples sampling.

Outline

Plan for Theory Discussion

Steps

1. Algorithms for computing homogeneous groups.
2. CLT for ATE / SATE estimation.
3. Optimal designs / pilot design.
4. Asymptotically exact inference.

Assumptions

Assumptions. Data $W_{1:n} = (X_i, Y_i(0), Y_i(1))_{i=1}^N \overset{\text{iid}}{\sim} F$ such that $E[Y(d)^2] < \infty$ for $d = 0, 1$ and $E[|\psi(X)|_2^\alpha] < \infty$ for some $\alpha > \dim(\psi) + 1$.

Remove Lipschitz condition on $E[Y(d)|\psi(X) = \psi]$ and $\text{Var}(Y(d)|\psi(X) = \psi)$, as well as boundedness of $\psi(X)$.

Algorithms

$$\frac{1}{N}\sum_{g}\sum_{i,j\in g}|\psi_i-\psi_j|_2^2 = o_p(1) \qquad \text{(homogeneity)}$$

Theorem. Optimal groups of size $|g| = k$ satisfy

$$\min_{(g)} \frac{1}{N}\sum_{g}\sum_{i,j\in g}|\psi_i-\psi_j|_2^2 = O_p\left(N^{2/\alpha-2/(d(\psi)+1)}\right) = o_p(1)$$

if $E[|\psi(X)|_2^\alpha] < \infty$. If $\psi(X)$ bounded, becomes $O\left(N^{-2/(d(\psi)+1)}\right)$, sharpening $N^{-1/d(\psi)}$ rate for $k = 2$ in Bai et al. (2021).

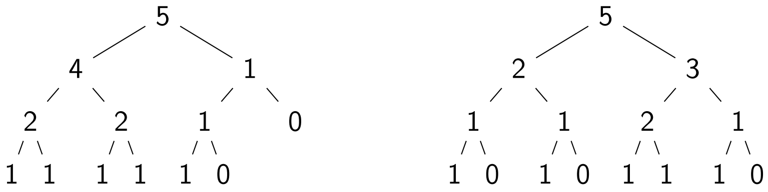
Computation. For $k = 2$, optimal pairs computable in $O(N^3)$ time by Derigs (1988) algorithm. Expected NP-hard for $k > 2$.

Algorithms

Iteratively apply optimal pairing ($k = 2$) algorithm to make $|g| = k > 2$.

At each step, match centroids $|g|^{-1} \sum_{i \in g} \psi_i$ of groups formed in previous step into pairs.

Many ways to implement for each k, e.g. $5 = 3 + 2$ or $5 = 4 + 1$.



Large Experiments

Complexity. Runs in $O(N^3)$ time. For e.g. matched 5-tuples, algorithm takes 23 seconds if $n = 500$, about 20 minutes if $n = 2000$.

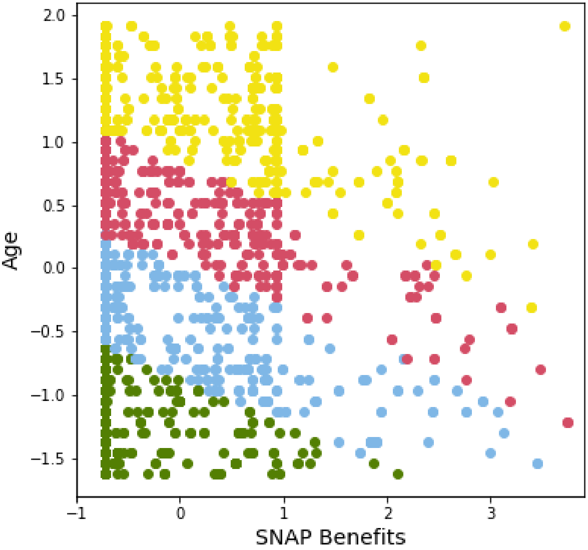
Large experiments - e.g. Domurat (2021), $n = 87394$, would take over 3.8 years.

PCA Folds. Let Z_1 first principal component of $(\psi_i)_{i=1}^n$:

1. Partition $(\psi_i)_{i=1}^n$ into K folds by $(1/K)$th quantile of sorted projections $Z_1'\psi_i$.
2. Separately in each fold, run the procedure above.

For $K = 40$, would run in 20 minutes in Domurat (2021) data with parallelization.

Large Experiments



ATE Estimation

Estimate $\text{ATE} = E[Y(1) - Y(0)]$ using $Y \sim 1 + D$. Let $\widehat{\theta}$ denote coefficient on D. Let $c(\psi) = E[Y(1) - Y(0)|\psi]$ and define the **balance function**

$$b(\psi; p) = \left(\frac{1-p}{p}\right)^{1/2} E[Y(1)|\psi] + \left(\frac{p}{1-p}\right)^{1/2} E[Y(0)|\psi]$$

Variance Decomposition. Suppose $T_{1:N} \sim \text{CR}(q)$ and $D_{1:N} \sim \text{CR}(p)$. Then $\sqrt{n}(\widehat{\theta} - \text{ATE}) \Rightarrow \mathcal{N}(0, V)$

$$V = \underbrace{\text{Var}(c(\psi))}_{\text{Heterogeneity}} + \underbrace{\text{Var}(b(\psi))}_{\text{Assignment}} + E\left[\frac{\sigma_1^2(\psi)}{p} + \frac{\sigma_0^2(\psi)}{1-p}\right]$$

Residual variance $\sigma_d^2(\psi) = \text{Var}(Y(d)|\psi)$.

ATE Estimation

$$V = \underbrace{\text{Var}(c(\psi))}_{\text{Heterogeneity}} + \underbrace{\text{Var}(b(\psi))}_{\text{Assignment}} + E\left[\frac{\sigma_1^2(\psi)}{p} + \frac{\sigma_0^2(\psi)}{1-p}\right]$$

Theorem. Suppose $T_{1:N} \sim \text{Loc}(\psi, q)$ and $D_{1:N} \sim \text{Loc}(\psi, p)$. Then $\sqrt{n}(\widehat{\theta} - \text{ATE}) \Rightarrow \mathcal{N}(0, V)$

$$V = q\,\text{Var}(c(\psi)) + E\left[\frac{\sigma_1^2(\psi)}{p} + \frac{\sigma_0^2(\psi)}{1-p}\right]$$

If $N = 1000$ and $n = 100$ then $q = 1/10$. First term behaves like infeasible larger experiment.

If $q = 1$, this is Hahn (1998) variance bound for ATE. Nonparametric regression adjustment "by design."

SATE Estimation

Define SATE $= E_N[Y_i(1) - Y_i(0)]$ in the eligible population.

Theorem. Suppose $T_{1:N} \sim \text{Loc}(\psi, q)$ and $D_{1:N} \sim \text{Loc}(\psi, p)$. Then $\sqrt{n}(\widehat{\theta} - \text{SATE}) \Rightarrow \mathcal{N}(0, V_{\text{SATE}})$

$$V_{\text{SATE}} = E\left[\frac{\sigma_1^2(\psi)}{p} + \frac{\sigma_0^2(\psi)}{1-p} - q\sigma_\tau^2(\psi)\right]$$

with $\sigma_\tau^2(\psi) = \text{Var}(Y(1) - Y(0)|\psi)$ not identified.

Inference can be based on lower bounds $\sigma_\tau^2(\psi) \geq (\sigma_1(\psi) - \sigma_0(\psi))^2 \geq 0$.

Fine Stratification with Varying Propensities

Consider a sampling propensity $q(\psi) \in \{a_j/k_j : j \in J\}$, e.g. $q(\psi) \in \{1/2, 1/5\}$.

Procedure.

1. (1) Form propensity strata $\{i : q(\psi_i) = a_j/k_j\}$.
2. (2) In each propensity stratum $\{i : q(\psi_i) = a_j/k_j\}$, draw $T_{1:N} \sim \mathrm{Loc}(\psi, a_j/k_j)$.

Example. $N = 1000$ villages in urban (U) and rural (R) areas

- Sampling proportions $q(U) = 1/2$ and $q(R) = 1/5$.
- Do matched pairs in U with $q = 1/2$ and matched 5-tuples in R with $q = 1/5$, still finely balancing ψ.
- Expect $D = 1$ is best, prioritize $p = 2/3$ for vulnerable villages.
- Matched triples with varying propensity $p(\psi) \in \{1/3, 2/3\}$.

ATE Estimation with Varying Propensities

Define the double IPW estimator

$$\widehat{\theta} = E_N\left[\frac{T_i D_i Y_i}{q(\psi_i)p(\psi_i)}\right] - E_N\left[\frac{T_i(1-D_i)Y_i}{q(\psi_i)(1-p(\psi_i))}\right]$$

Theorem. Suppose $T_{1:N} \sim \text{Loc}(\psi, q(\psi))$ and $D_{1:N} \sim \text{Loc}(\psi, p(\psi))$. Then $\sqrt{n}(\widehat{\theta} - \text{ATE}) \Rightarrow \mathcal{N}(0, \bar{q}V)$ for $\bar{q} = E[q(\psi)]$ and

$$V = \text{Var}(c(\psi)) + E\left[\frac{1}{q(\psi)}\left(\frac{\sigma_1^2(\psi)}{p(\psi)} + \frac{\sigma_0^2(\psi)}{(1-p(\psi))}\right)\right]$$

For $q = 1$, ATE variance bound for propensity $p(\psi)$.

For $q \neq 1$, variance of "double" AIPW estimator adjusting nonparametrically for imbalances in both sampling and assignment variables.

ATE Estimation with Varying Propensities

Suppose $|\widehat{m}_d - E[Y(d)|\psi]|_{2,\psi} = o_p(1)$. Define the estimator

$$\widehat{\theta}_{AIPW} = E_n[\widehat{m}_1(\psi_i) - \widehat{m}_0(\psi_i)] + E_n\left[\frac{T_i D_i(Y_i - \widehat{m}_1(\psi_i))}{q(\psi_i)p(\psi_i)} - \frac{T_i(1-D_i)(Y_i - \widehat{m}_0(\psi_i))}{q(\psi_i)(1-p(\psi_i))}\right]$$

Theorem. Suppose $T_{1:N} \overset{\text{iid}}{\sim} \text{Bernoulli}(q(\psi_i))$ and $D_{1:N} \overset{\text{iid}}{\sim} \text{Bernoulli}(p(\psi_i))$. Then $\sqrt{n}(\widehat{\theta}_{AIPW} - \text{ATE}) \Rightarrow \mathcal{N}(0, \bar{q}V)$

$$V = \text{Var}(c(\psi)) + E\left[\frac{1}{q(\psi)}\left(\frac{\sigma_1^2(\psi)}{p(\psi)} + \frac{\sigma_0^2(\psi)}{(1-p(\psi))}\right)\right]$$

Same variance as IPW estimator with $T_{1:N} \sim \text{Loc}(\psi, q(\psi))$ and $D_{1:N} \sim \text{Loc}(\psi, p(\psi))$.

Nonparametric regression adjustment "by design."

Intuition

For a set $A \subseteq \mathbb{R}^{\dim(\psi)}$ define the **realized propensity**

$$\widehat{p}(A) \equiv E_n[D_i | \psi_i \in A].$$

Local Randomization. For $D_{1:n} \sim \text{Loc}(\psi, p)$ can show

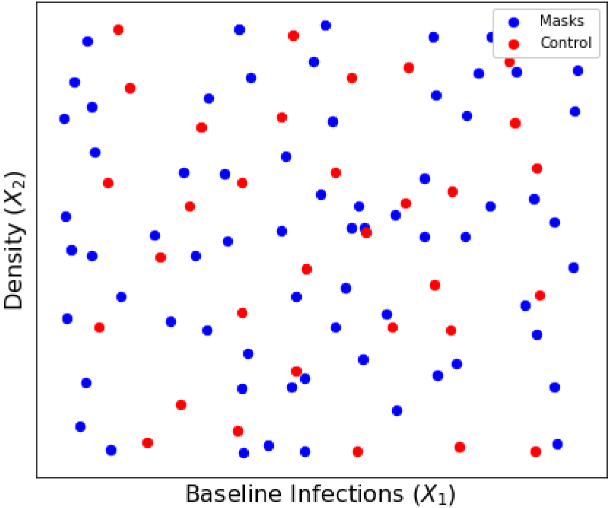
$$\widehat{p}(A) = p + o_p(n^{-1/2}) \qquad A \subseteq \mathbb{R}^{\dim(\psi)}$$

Even holds for $P(\psi \in A_n) \to 0$, giving a "local" implementation of p with respect to ψ.

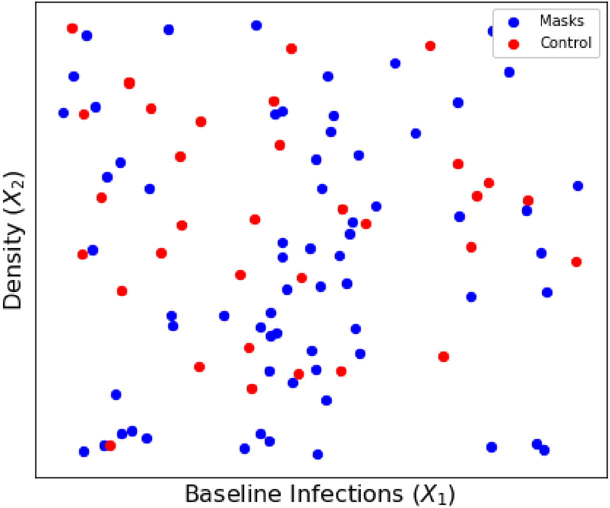
By contrast, if $D_{1:n} \sim \text{CR}(p)$ then $\widehat{p}(A) = p + O_p(n^{-1/2})$.

Efficiency. Hirano et al. (2003) - Nonparametrically re-estimate known p with $\widehat{p}(\psi)$ and do IPW. Here, p is implemented correctly at design-time.

Local Randomization



Complete Randomization



Optimal Sampling

Goal. Solve for optimal sampling propensity $q^*(\psi)$ for sampling costs $C(\psi)$ and budget constraint $\bar{B}$. For example,

$$C(\psi; p) = C_s(\psi) + p(\psi)C_1(\psi) + (1 - p(\psi))C_0(\psi).$$

We have $\sqrt{N}(\widehat{\theta} - \text{ATE}) \Rightarrow \mathcal{N}(0, V(q))$, with objective

$$V(q) = \text{Var}(c(\psi)) + E\left[\frac{1}{q(\psi)}\left(\frac{\sigma_1^2(\psi)}{p(\psi)} + \frac{\sigma_0^2(\psi)}{(1 - p(\psi))}\right)\right]$$

Define the **ex-ante** residual variance

$$\bar{\sigma}^2(\psi) = \frac{\sigma_1^2(\psi)}{p(\psi)} + \frac{\sigma_0^2(\psi)}{1 - p(\psi)}$$

Leads to optimal sampling problem

$$\min_{0 \le q \le 1} E\left[\frac{\bar{\sigma}^2(\psi)}{q(\psi)}\right] \quad \text{s.t.} \quad E[C(\psi)q(\psi)] \le \bar{B}$$

Optimal Sampling

$$\min_{0\leq q\leq 1} E\left[\frac{\bar{\sigma}^2(\psi)}{q(\psi)}\right] \quad \text{s.t.} \quad E[C(\psi)q(\psi)] \leq \bar{B}$$

Theorem. Define the candidate solution

$$q^*(\psi) = \bar{B} \cdot \frac{\bar{\sigma}(\psi)C(\psi)^{-1/2}}{E[\bar{\sigma}(\psi)C(\psi)^{1/2}]}$$

If $\sup_\psi q^*(\psi) \leq 1$, then $q^*(\psi)$ is optimal.

Homogeneous Costs. If $C(\psi) = 1$ then $q^*(\psi) \propto \bar{\sigma}(\psi)$.

Homoskedasticity. $q^*(\psi) \propto C(\psi)^{-1/2}$ if $p(\psi) = p$.

Optimal Spending at ψ is

$$S(\psi) \propto \bar{\sigma}(\psi)C(\psi)^{1/2}dP(\psi)$$

Optimal Stratification

Optimal Stratification. Implement jointly optimal $q^*(\psi)$ and $p^*(\psi)$ for stratification variables ψ.

$$p^*(\psi) = \frac{\sigma_1(\psi)}{\sigma_1(\psi) + \sigma_0(\psi)} \quad q^*(\psi) = \bar{B}\frac{(\sigma_1(\psi) + \sigma_0(\psi))C(\psi)^{-1/2}}{E[(\sigma_1(\psi) + \sigma_0(\psi))C(\psi)^{1/2}]}$$

Let $\widehat{p}(\psi)$ and $\widehat{q}(\psi)$ consistent pilot estimates of $p^*(\psi)$ and $q^*(\psi)$. Requires $\|\widehat{\sigma}_d - \sigma_d\|_{2,\mathcal{X}} \to 0$.

Let $\widehat{p}_n(\psi)$ and $\widehat{q}_n(\psi)$ discretizations, rounding to nearest a/k_n with $k_n \to \infty$ and $k_n = o(\sqrt{n})$.

Theorem. Suppose $T_{1:n} \sim \text{Loc}(\psi, \widehat{q}_n(\psi))$ and $D_{1:n} \sim \text{Loc}(\psi, \widehat{p}_n(\psi))$. Then $\sqrt{n}(\widehat{\theta} - \text{ATE}) \Rightarrow \mathcal{N}(0, \bar{q}V^*)$.

$$\text{Var}(c(\psi)) + \min_{\substack{0 \le q,p \le 1 \\ E[C(\psi)q(\psi)] = \bar{B}}} E\left[\frac{1}{q(\psi)}\left(\frac{\sigma_1^2(\psi)}{p(\psi)} + \frac{\sigma_0^2(\psi)}{(1-p(\psi))}\right)\right]$$

Global Optimality

Bai (2022) that matching units into strata by sorted $b(X_i)$ values is the optimal stratification, for fixed p.

Result. If $b(x)$ known, stratified designs are suboptimal, derive globally optimal design for $p = 1/2$.

Define Max-Cut problem with edge weights $w_{ij} = b(X_i)b(X_j)$

$$\max_{E_0,E_1} \sum_{i,j} b(X_i)b(X_j)\mathbb{1}(i \in E_1, j \in E_0) \quad \text{s.t.} \quad E_0 \sqcup E_1 = \{1, \ldots, n\}$$

Let E_0^*, E_1^* optimal and $d_i^* = \mathbb{1}(i \in E_1^*)$ and define

$$P^*(D_{1:n} = d_{1:n}^* | X_{1:n}) = P^*(D_{1:n} = 1 - d_{1:n}^* | X_{1:n}) = 1/2$$

Theorem. The design P^* has

$$\text{MSE}_{P^*}(\widehat{\theta} | X_{1:n}) \leq \text{MSE}_P(\widehat{\theta} | X_{1:n})$$

for all P with $P(D_i = 1) = 1/2$ and $D_{1:n} \perp\!\!\!\perp W_{1:n} | X_{1:n}$.

Estimating What to Stratify On

Feasible version of globally optimal alternating design or Bai (2022) design: plug in pilot estimate $\widehat{b}(x)$.

Feasible version of Bai (2022) design: $D_{1:n} \sim \mathrm{Loc}(\widehat{b}, p)$.

Equivalent to AIPW estimation with regressions estimated in pilot, up to $O_p(n^{-1})$.

Robustified Version. Draw $T_{1:n} \sim \mathrm{Loc}(\psi, q)$ and $D_{1:n} \sim \mathrm{Loc}(\psi, p)$ with

$$\psi = (\widehat{c}, \widehat{b}, \psi')$$

for some ψ' chosen ex-ante. Can find optimal propensities $q^*(\psi)$ and $p^*(\psi)$ for this ψ as previously discussed.

Outline

Inference

Suppose $T_{1:N} \sim \text{Loc}(\psi, q(\psi))$ and $D_{1:N} \sim \text{Loc}(\psi, p(\psi))$. By earlier theorem, $\sqrt{n}(\widehat{\theta} - \text{ATE}) \Rightarrow \mathcal{N}(0, V)$.

Theorem. (Inference) Define variance estimator

$$\widehat{V} = \text{Var}_n\left(\frac{T_i(D_i - p(\psi_i))Y_i}{q(\psi_i)(p - p^2)(\psi_i)}\right) - \widehat{v}_1 - \widehat{v}_0 - 2\widehat{v}_{10}.$$

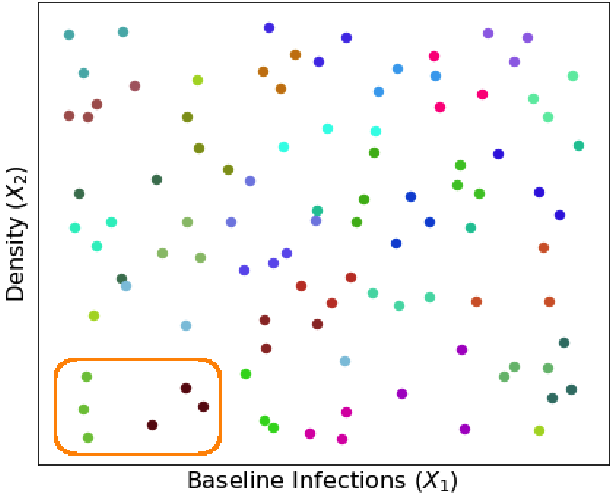
Then $\widehat{V} \xrightarrow{p} V$.

The matching corrections $\widehat{v}_1$, $\widehat{v}_0$, $\widehat{v}_{10}$ estimate how well ψ predicts outcomes. For example, $\widehat{v}_1 \approx E[E[Y(1)|\psi]^2]$.

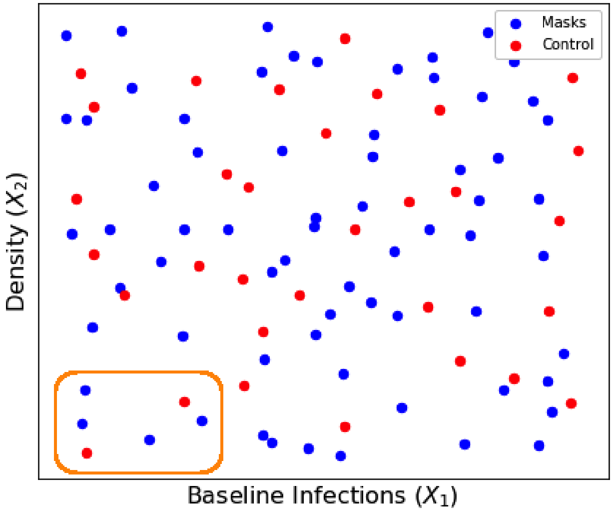
In particular, for $w_i^1 = (1 - p_i q_i)/(p_i q_i)^2$

$$\widehat{v}_1 = (2n)^{-1} \sum_{g \in \mathcal{G}_n^\nu} \sum_{\substack{i,j \in g \\ i \neq j}} D_i D_j (w_i^1 w_j^1)^{1/2} Y_i Y_j \approx E[E[Y(1)|\psi]^2]$$

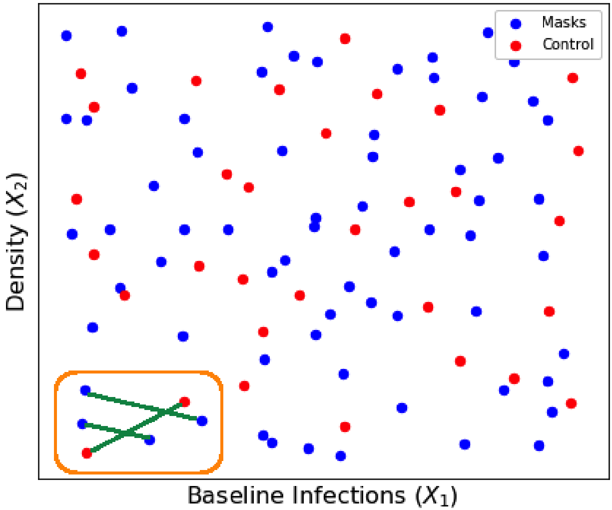
Measuring Similarity of Outcomes



Measuring Similarity of Outcomes



Measuring Similarity of Outcomes



Outline

Empirical Application

Goal. Quantify variance reduction from finely stratified sampling, assignment, optimal propensities for DGP's from experimental economics.

Use $N = 9$ empirical papers, 7 from AER 2021-2022 plus Banerjee et al. (2021) and Finklestein et al. (2012).

Impute missing potential outcomes $\widehat{Y}_i(d) = Y_{j(i)}(d)$ with $j(i) = \text{argmin}_{j:D_j=d} |\psi_i - \psi_j|_2$.

Draw $(\widehat{Y}_i(0), \widehat{Y}_i(1), \psi_i)_{i=1}^n$ with replacement, make design, reveal outcomes $\widehat{Y}_i = T_i(D_i \widehat{Y}_i(1) + (1 - D_i)\widehat{Y}_i(0))$, estimate and form confidence interval.

Empirical Application

Costs. Define costs

$$C(\psi) = \mathbb{1}(|\psi|_2 \leq \kappa) + 5\mathbb{1}(|\psi|_2 > \kappa)$$

with $\kappa = \text{Median}_{i=1}^{n}|\psi_i|_2$ and $\bar{B} = 1.5$. Results in sampling proportion $\bar{q} = E[q^*(\psi)] \approx 0.7$.

Designs.

1. **CR**. Draw $T_{1:n} \sim \text{CR}(q_n^*)$ and $D_{1:n} \sim \text{CR}(p)$ for budget-exhausting q_n^*.
2. **CR, Loc**. Same but $D_{1:n} \sim \text{Loc}(\psi, p)$.
3. **Loc**. Same but $T_{1:n} \sim \text{Loc}(\psi, q_n^*)$.
4. **Hom**. $T_{1:n} \sim \text{Loc}(\psi, q_n^*(\psi))$ assuming $\sigma_d^2(\psi)$ constant.
5. **Pilot S/L**. $T_{1:n} \sim \text{Loc}(\psi, \widehat{q}_n(\psi))$ and $D_{1:n} \sim \text{Loc}(\psi, \widehat{p}_n)$ with $n_{pilot} = 100, 400$.

Empirical Application

	Design, Paper	A.	Ban.	Bay.	C.	De.	Do.	F.	H.	L.
SD	CR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	CR, Loc	0.81	0.48	0.51	0.88	0.55	0.93	0.79	0.73	0.87
	Loc	0.75	0.45	0.39	0.84	0.47	0.88	0.74	0.70	0.76
	Hom.	0.72	0.46	0.44	0.79	0.58	0.86	0.85	0.68	0.90
	Pilot S	0.72	0.45	0.55	0.81	0.57	1.01	0.75	0.71	0.75
	Pilot L	0.70	0.42	0.52	0.81	0.52	0.87	0.65	0.68	0.71
%ΔCI	CR	0	0	0	0	0	0	0	0	0
	CR, Loc	-11	-49	-41	-6	-36	-4	-14	-10	-11
	Loc	-12	-48	-38	-5	-33	-4	-14	-15	-11
	Hom.	-21	-44	-47	-6	-40	-8	-8	-16	-10
	Pilot S	-19	-47	-23	-5	-23	3	-18	-15	-21
	Pilot L	-21	-52	-31	-5	-30	-9	-27	-18	-24
Cover	CR	0.95	0.95	0.95	0.94	0.96	0.95	0.94	0.94	0.96
	CR, Loc	0.96	0.96	0.99	0.96	0.98	0.96	0.96	0.98	0.96
	Loc	0.98	0.97	1.00	0.96	1.00	0.97	0.97	0.97	0.98
	Hom.	0.97	0.98	0.98	0.97	0.97	0.96	0.96	0.98	0.95
	Pilot S	0.97	0.98	0.99	0.96	0.99	0.95	0.97	0.98	0.97
	Pilot L	0.97	0.97	0.99	0.97	0.99	0.96	0.97	0.97	0.96
	n	1451	903	550	91	446	1000	1903	116	770
	$\dim(\psi)$	8	6	3	3	3	4	6	4	3

Empirical Application

Implications.

1. Clear case for finely stratified sampling and assignment.
2. Variance estimation reflects precision from stratified sampling and assignment, though conservative.
3. Use $q^*_{hom}(\psi)$ under homoskedasticity or very heterogeneous costs.
4. Use $\widehat{q}(\psi)$ or $\widehat{p}(\psi)$ if have large pilot, previous observational data, large heteroskedasticity.

Summary and Conclusion

Local Randomization. Generic data-adaptive fine stratification $D_{1:n} \sim \text{Loc}(\psi, p(\psi))$ for $\psi \in \mathbb{R}^d$.

Application: finely stratified sampling into an experiment.

1. Sampling reduces variance due to treatment effect heterogeneity.
2. Assignment reduces variance due to outcome heterogeneity.

Optimal Stratification. Formalized and solved budget-constrained optimal stratification problem, with feasible pilot versions.

- Globally optimal covariate-adaptive design for $p = 1/2$.

Asymptotically exact inference accounting for stratified sampling and assignment.