The Benefits of Probability Proportional to Size Sampling in Cluster-Randomized Experiments

Michael J. Higgins

Princeton University

Aug 29, 2014

Cluster-Randomized Experiments

- Cluster-randomized experiments: Randomization of treatment across clusters of units instead of the units themselves.
- Reasons for cluster-level randomization: Logistical, not to improve power.

  E.g.: Impractical or infeasible to randomize across units, avoid "interference" or "treatment contamination," etc.

Cluster-Randomized Experiments: Examples

Examples:

- Assign treatment to villages, interested in effects at individual/household level.

  National Solidarity Programme (Beath, Christia, Enikolopov)
  Clientelism and Vote Buying (Wanchekon)

- Assign treatment to households, interested in individual effects.

  Get out the vote (Gerber, Green, Larimer)

Also common in medical trials and experiments in education.

Problems with estimating treatment effects

- Unbiased and precise estimation of population average treatment effect (PATE) can be problematic.

Problems with estimating treatment effects

- Unbiased and precise estimation of population average treatment effect (PATE) can be problematic.
- With PATE, usually need to sample clusters from population (e.g. sampling a subset of villages in a country).
- Current practice: clusters sampled using SRS

Problems with estimating treatment effects

- Unbiased and precise estimation of population average treatment effect (PATE) can be problematic.
- With PATE, usually need to sample clusters from population (e.g. sampling a subset of villages in a country).
- Current practice: clusters sampled using SRS
- Horvitz-Thompson: unbiased, but not location invariant.
  Large standard errors that can increase when transforming variables.

  Difference-in-means: Location invariant but biased when treatment effects correlated with number of units within cluster (Middleton and Aronow).

Problems with estimating treatment effects

- Unbiased and precise estimation of population average treatment effect (PATE) can be problematic.
- With PATE, usually need to sample clusters from population (e.g. sampling a subset of villages in a country).
- Current practice: clusters sampled using SRS
- Horvitz-Thompson: unbiased, but not location invariant.
  Large standard errors that can increase when transforming variables.

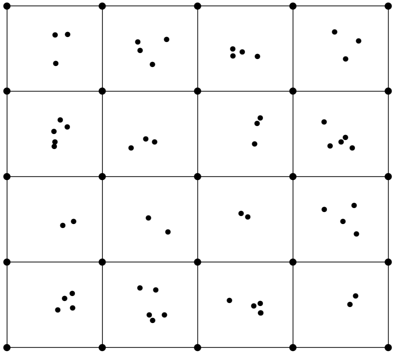
  Difference-in-means: Location invariant but biased when treatment effects correlated with number of units within cluster (Middleton and Aronow).
- Solution: Sampling clusters with probability proportional to size—the number of units within each clusters—eliminate these problems with estimation.

Cluster-Randomized Experiments: Framework

- The population of units of interest $y_1, \ldots, y_n$ partitioned into $\#c$ clusters.
- Quantity of interest: population average treatment effect (PATE).

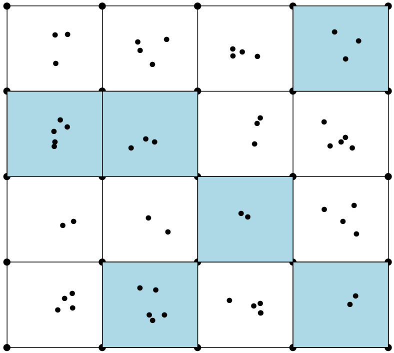
$$PATE = \frac{1}{n}\sum_{i=1}^{n} y_i(1) - \frac{1}{n}\sum_{i=1}^{n} y_i(0)$$

$y_i(1)$ is response of unit i under treatment, $y_i(0)$ is response under control.



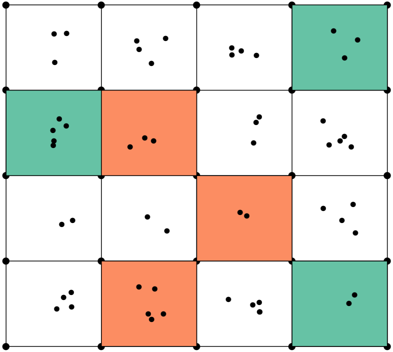
Cluster-Randomized Experiments: Framework

1. Draw a sample of $\#s$ clusters from the population of clusters.



Cluster-Randomized Experiments: Framework

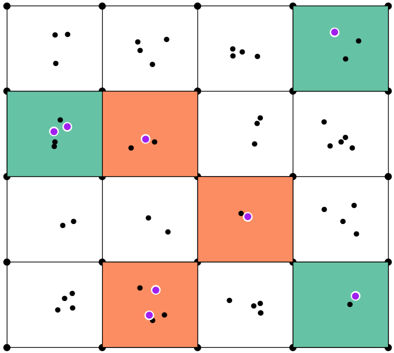
1. Draw a sample of $\#s$ clusters from the population of clusters.
2. Randomize treatment across clusters.



Cluster-Randomized Experiments: Framework

1. Draw a sample of $\#s$ clusters from the population of clusters.
2. Randomize treatment across clusters.
3. Sample $\#s_c$ units within each of the sampled clusters.

Inference on PATE made using purple sampled units.



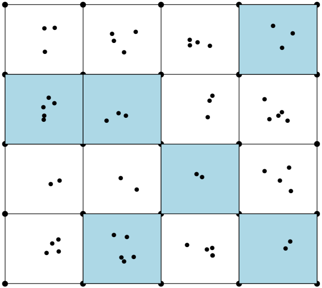
Current practice: SRS sampling of clusters

- Under Neyman-Rubin potential outcomes model:

$$Y_i = y_i(1)T_i + y_i(0)(1 - T_i)$$

Under SRS of clusters, no current estimator of PATE that is both unbiased and invariant to location shifts of potential outcomes...

- ... without the introduction of additional parameters (Des Raj) (Middleton and Aronow).



Inference on PATE: Horvitz-Thompson

- Horvitz-Thompson estimator of PATE:

$$\sum_{c \in \# s} w_c \left(\sum_{T_c = 1} \frac{1}{\# T_1} \frac{n_c}{n} \sum_{k \in c} \frac{y_{kc}(1)}{\# s_c} - \sum_{T_c = 0} \frac{1}{\# T_0} \frac{n_c}{n} \sum_{k \in c} \frac{y_{kc}(0)}{\# s_c} \right)$$

$$\begin{aligned}
\sum_{k \in c} \frac{y_{kc}(1)}{\# s_c} &= \text{Within-cluster sample mean} \\
\frac{n_c}{n} &= \text{Fraction of units contained in cluster } c \\
\# T_1 &= \text{Number of treated clusters}
\end{aligned}$$

Inference on PATE: Horvitz-Thompson

- Horvitz-Thompson estimator of PATE:

$$\sum_{c \in \# s} w_c \left(\sum_{T_c = 1} \frac{1}{\# T_1} \frac{n_c}{n} \sum_{k \in c} \frac{y_{kc}(1)}{\# s_c} - \sum_{T_c = 0} \frac{1}{\# T_0} \frac{n_c}{n} \sum_{k \in c} \frac{y_{kc}(0)}{\# s_c} \right)$$

- Unbiased, but not invariant to location shifts of potential outcomes, which inflates variance:

$$y_{kc}^* = y_{kc} + \alpha.$$

$$\widehat{PATE}^* = \widehat{PATE} + k\alpha \left(\# n_1 - \# n_0 \right).$$

- E.g.: Adjustments for inflation or currency for logged responses. Recoding binary variable from 0–1 to 1–0.

Inference on PATE: Difference-in-means

- Difference-in-means estimator of PATE:

$$\frac{\sum_{T_c=1}\sum_{k\in c} y_{kc}(1)}{\# s_1} - \frac{\sum_{T_c=0}\sum_{k\in c} y_{kc}(0)}{\# s_0}$$

- Invariant to location shifts, but biased if treatment effects are correlated with cluster sizes.

  E.g. Treatment effect smaller going from small towns to large cities.

Inference on PATE: Difference-in-means

- Difference-in-means estimator of PATE:

$$\frac{\sum_{T_c=1} \sum_{k \in c} y_{kc}(1)}{\# s_1} - \frac{\sum_{T_c=0} \sum_{k \in c} y_{kc}(0)}{\# s_0}$$

- Invariant to location shifts, but biased if treatment effects are correlated with cluster sizes.

  E.g. Treatment effect smaller going from small towns to large cities.
- Common to block or stratify on cluster size to reduce bias.

Inference on PATE: Difference-in-means

- Difference-in-means estimator of PATE:

$$\frac{\sum_{T_c=1}\sum_{k\in c} y_{kc}(1)}{\#s_1} - \frac{\sum_{T_c=0}\sum_{k\in c} y_{kc}(0)}{\#s_0}$$

- Invariant to location shifts, but biased if treatment effects are correlated with cluster sizes.

  E.g. Treatment effect smaller going from small towns to large cities.
- Common to block or stratify on cluster size to reduce bias.
- Suboptimal if other covariates predict treatment effects better.
- Still some bias in finite experiments.

Solution: Better design

- Solution: Sample clusters proportional to *size* instead of SRS.

Solution: Better design

- Solution: Sample clusters proportional to *size* instead of SRS.
- Focus attention on sampling without replacement (PPSWOR sampling):
  Fixed number of clusters sampled $\Longrightarrow$ Easier to design experiment with budget constraint.

PPSWOR Sampling

- PPSWOR: Probability of sampling cluster c in a sample of $\#s$ clusters is $\#s n_c / n$.
- Does not uniquely define a sample; joint probabilities $\pi_{cc'}$ need to be specified.

PPSWOR Sampling

- PPSWOR: Probability of sampling cluster c in a sample of $\#s$ clusters is $\#s n_c / n$.
- Does not uniquely define a sample; joint probabilities $\pi_{cc'}$ need to be specified.
- `SunterSampling` `R` package; efficiently draw approximate PPSWOR sample with nice properties.

Inference on PATE: Horvitz-Thompson

- Horvitz-Thompson estimator of PATE:

$$\sum_{c \in \# s} w_c \left(\sum_{T_c = 1} \frac{1}{\# T_1} \sum_{k \in c} \frac{y_{kc}(1)}{\# s_c} - \sum_{T_c = 0} \frac{1}{\# T_0} \sum_{k \in c} \frac{y_{kc}(0)}{\# s_c} \right)$$

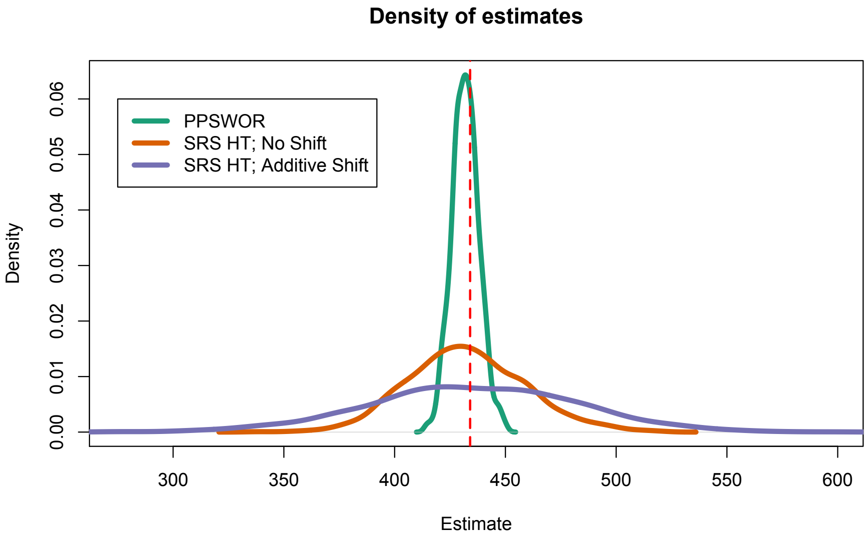
Inference on PATE: Horvitz-Thompson

- Horvitz-Thompson estimator of PATE:

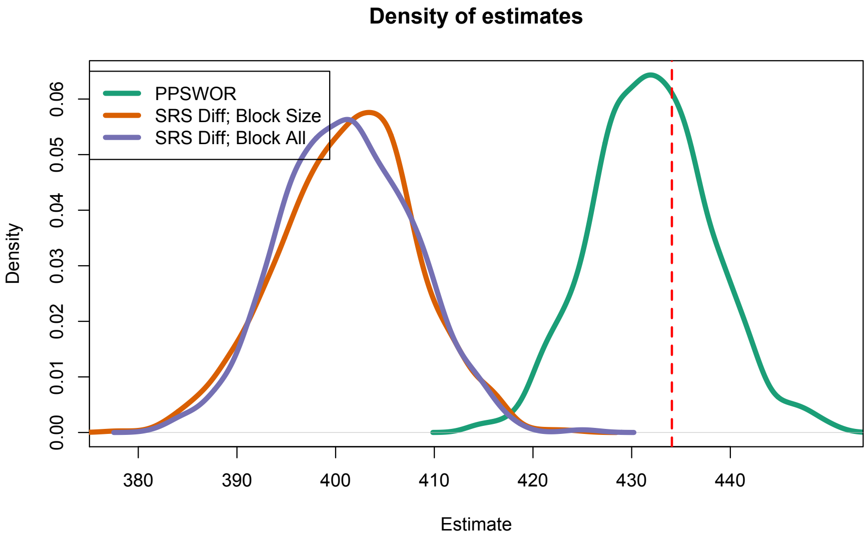
$$\sum_{c \in \# s} w_c \left(\sum_{T_c = 1} \frac{1}{\# T_1} \sum_{k \in c} \frac{y_{kc}(1)}{\# s_c} - \sum_{T_c = 0} \frac{1}{\# T_0} \sum_{k \in c} \frac{y_{kc}(0)}{\# s_c} \right)$$

- Unbiased and invariant to location shifts of potential outcomes.
- Freedom to block on covariates other than cluster size.
- Variance not dramatically different from SRS methods.
- Equivalent to difference-in-means estimator when same within-cluster sample sizes or under matched-pairs blocking.

Simulation results



Simulation results



Extensions

Results hold when:

- Stratifying clusters before sampling.
  E.g.: Sample large cities and small villages separately.
- Stratifying within-cluster sample.
  E.g.: Gender, race.
- When treatment is randomized across individual units.
  Useful when design has treatments randomized at both cluster and treatment level.

Thank you.