Stratified Randomized Experiments

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

Blocking for Improved Efficiency

- How can we improve the efficiency of causal effect estimation while maintaining the unbiasedness?
- Minimize variances of potential outcomes
  $\rightsquigarrow$ conduct a randomized experiment in a group of similar units

"Block what you can and randomize what you cannot"

Box, et al. (2005). *Statistics for Experimenters*. 2nd eds. Wiley.

- Basic procedure:
  1. Blocking: create groups of similar units based on pre-treatment covariates
  2. Block randomization: completely randomize treatment assignment within each group

Stratified Design

- Setup:
  - Number of units, n
  - Number of blocks, J
  - Block size, $n_j > 2$
  - Number of treated in each block, $n_{j1} > 1$
  - Complete randomization within each block, $\Pr(T_{ij} = 1) = n_{j1}/n_j$
- Analysis:
  1. Apply Neyman's analysis to each block

$$\hat{\tau}_j = \frac{1}{n_{j1}} \sum_{i=1}^{n_j} T_{ij} Y_{ij} - \frac{1}{n_{j0}} \sum_{i=1}^{n_j} (1 - T_{ij}) Y_{ij}, \quad \widehat{\mathbb{V}(\hat{\tau}_j)} = \frac{\hat{\sigma}_{j1}^2}{n_{j1}} + \frac{\hat{\sigma}_{j0}^2}{n_{j0}}$$

  2. Aggregate block-specific estimates and variances

$$\hat{\tau}_{\text{block}} = \sum_{j=1}^{J} w_j \cdot \hat{\tau}_j, \quad \text{and} \quad \widehat{\mathbb{V}_{\text{block}}(\hat{\tau}_{\text{block}})} = \sum_{j=1}^{J} w_j^2 \cdot \widehat{\mathbb{V}(\hat{\tau}_j)}$$

where w_j is the weight for the jth block, e.g., $w_j = n_j/n$

Efficiency Gain due to Blocking

- Simple analytic framework:
  - PATE as the estimand
  - J pre-defined blocks within an infinite population
  - Stratified random sampling of $w_j \cdot n$ units within each block
  - Complete randomization of treatment assignment within each block
  - Identical treatment assignment probability across blocks $k = n_1/n$
- Key equality:

$$\underbrace{\mathbb{V}(X)}_{\text{total variance}} = \underbrace{\mathbb{E}\{\mathbb{V}(X \mid Y)\}}_{\text{within}-\text{block variance}} + \underbrace{\mathbb{V}\{\mathbb{E}(X \mid Y)\}}_{\text{across}-\text{block variance}}$$

- Difference in variance:

$$\begin{aligned} \mathbb{V}(\hat{\tau}) - \mathbb{V}_{\text{block}}(\hat{\tau}_{\text{block}}) &= \frac{1}{n}\left\{\frac{\sigma_1^2}{k} + \frac{\sigma_0^2}{1-k} - \sum_{j=1}^{J} w_j \left(\frac{\sigma_{1j}^2}{k} + \frac{\sigma_{0j}^2}{1-k}\right)\right\} \\ &\geq 0 \end{aligned}$$

The Project STAR

- Randomization was done within each school $\rightsquigarrow$ stratification!
- Effect of kindergarden class size on high school graduation:
  1. Permutation tests
     - Fisher's exact test: p-value $= 0.51$
     - Mantel-Haenszel test: p-value = 0.37
  2. Average treatment effect estimation
     - est. $= 0.018$, se. $= 0.017$
     - s.e. without stratification $\approx 7\%$ greater
- Effect of kindergarden class size on 8th grade reading score
  1. Permutation tests
     - Wilcoxon's test: p-value $= 0.121$
     - Aligned rank sum test: p-value $= 0.067$
  2. Average treatment effect estimation
     - est. $= 2.76$, se. $= 1.73$
     - s.e. without stratification $\approx 9\%$ greater

Matched-Pairs Design

- Should we keep blocking until we cannot block any further?
- Procedure:
  1. Create $J = n/2$ pairs of similar units
  2. Randomize treatment assignment within each pair
     - $W_j = 1$: first unit receives the treatment
     - $W_j = -1$: second unit receives the treatment
- Analysis:

$$\hat{\tau}_{\text{pair}} = \frac{1}{J} \sum_{j=1}^{J} W_j (Y_{1j} - Y_{2j}),$$

$$\widehat{\mathbb{V}(\hat{\tau}_{\text{pair}})} = \frac{1}{J(J-1)} \sum_{j=1}^{J} \{W_j (Y_{1j} - Y_{2j}) - \hat{\tau}_{\text{pair}}\}^2$$

Efficiency Analysis

- Neyman's stratified variance estimator is not applicable
- For SATE, $\widehat{\mathbb{V}(\hat{\tau}_{\text{pair}})}$ is conservative unless the average treatment effect is constant across pairs (Imai. 2008. *Stat. Med.*)

- For PATE, simple random sampling of pairs instead of stratified random sampling within pre-defined strata

$$\mathbb{E}(\widehat{\mathbb{V}(\hat{\tau}_{\text{pair}})}) \ = \ \frac{\sigma_1^2}{J} + \frac{\sigma_0^2}{J} - 2 \times \text{Cov}(Y_{1j}(1), Y_{2j}(0))$$

- Improved inference under stratified random sampling:
  - group similar pairs (IMBENS AND RUBIN. Chapter 10)
  - regression (Forgaty. 2018. *J. Royal Stat. Soc. B*)

Evaluation of Seguro Popular (King et al. 2009. *Lancet*)

- 50 million uninsured Mexicans $\rightsquigarrow$ catastrophic medical expenditure among poor households
- Seguro popular: delivery of health insurance, regular and preventive healthcare, medicines and health facilities
- Units: health clusters $=$ predefined health facility catchment areas
- Randomization within 74 matched pairs of "similar" health clusters
- 10 months followup survey for 50 pairs
- Outcome: proportion of households within each health cluster who experienced catastrophic medical expenditure
  - est. $= -0.013$, s.e. $= 0.007$
  - $\text{Cor}(Y_{1j}(1), Y_{2ij}(0)) = 0.482$
  - estimated s.e. under complete randomization $= 0.010$
- Wilcoxon's signed rank test:
  p-value $= 0.10$, 95% conf. int. $= [-0.026, 0.002]$

Blocking in Practice

- Univariate blocking: discrete or discretized variable
- Multivariate blocking: Mahalanobis distance

$$D(\mathbf{X}_i, \mathbf{X}_j) = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)^\top \widehat{\mathbb{V}(\mathbf{X})}^{-1} (\mathbf{X}_i - \mathbf{X}_j)}$$

- Greedy algorithms
  1. Matching: pair two units with the shortest distance, set them aside, and repeat
  2. Blocking: randomly choose one unit and choose n_j units with the shortest distances, set them aside, and repeat
- But the resulting matches may not be optimal

Optimal Matching

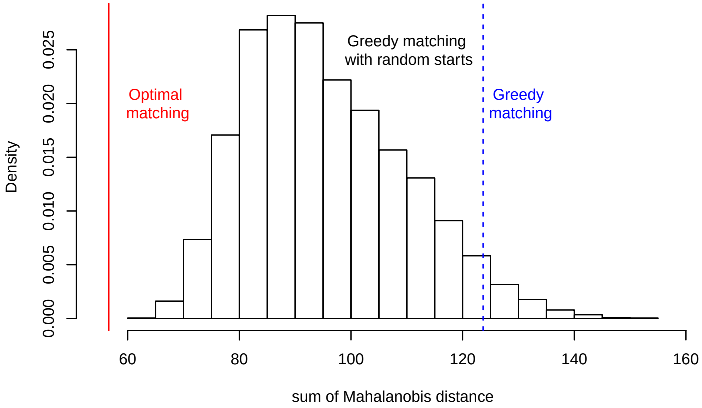
- **D**: $n \times n$ matrix of pairwise distance or a cost matrix
- Select n elements of **D** such that there is only one element in each row and one element in each column and the sum of pairwise distances is minimized
- Linear Sum Assignment Problem (LSAP)
  - Binary $n \times n$ matching matrix: **M** with $M_{ij} \in \{0, 1\}$
  - Optimization problem

$$\underset{\mathbf{M}}{\text{minimize}} \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} D_{ij} \quad \text{subject to} \quad \sum_{i=1}^{n} M_{ij} = 1, \sum_{j=1}^{n} M_{ij} = 1$$

    where we set $D_{ii} = \infty$ for all i
  - application of Hungarian algorithm etc.
  - also a special case of nonbipartite matching problem
- To be truely optimal, we must consider variable strata size

Optimal Matching for Seguro Popular

- 4 pre-treatment cluster-average covariates: age, education, household size, household assets
- 100 clusters, 50 pairs
- Minimize the sum of pairwise Mahalanobis distance



Adaptive Designs

- What happens if subjects sequentially arrive?
- Biased coin design (Efron. 1971. *Biometrika*)
  1. For the first $2m$ units, we use the Bernoulli design
  2. For a new unit, assign it to the treatment group with

$$\begin{cases} \text{probability } p & \text{if more units are in treatment group} \\ \text{probability } 1/2 & \text{if treatment and control groups have same size} \\ \text{probability } q & \text{if more units are in control group} \end{cases}$$

  - Efron suggests $p = 1/3$ and $q = 2/3$
  - Permutation tests or asymptotic approximation

- Doubly adaptive biased coin design (Eisele. 1994. *J. Stat. Plan. Inference*)
  - Modify step 2 above with the following:
    assign a new unit to the treatment group with probability
    f(proportion of the treated, $\hat{\sigma}_1/(\hat{\sigma}_1 + \hat{\sigma}_0)$)
  - optimal rule:

Summary

- Blocking improves efficiency of inference with randomized experiments while preserving the advantages of randomization
- Neyman's randomization inference allows for the efficiency analysis
  - Stratified designs
  - Matched-pair designs
  - Biased-coin designs
- Optimal matching algorithm $\rightsquigarrow$ stratification of variable size?
- Suggested reading: IMBENS AND RUBIN, Chapters 9 and 10