Randomization tests of causal effects with interference between units

Panagiotis (Panos) Toulis
panos.toulis@chicagobooth.edu

Econometrics and Statistics
University of Chicago, Booth School of Business

Joint work with Guillaume Basse (Harvard), Avi Feller (UC Berkeley)
Motivation

- Rogers and Feller (2018) ran a two-stage randomized experiment to engage parents of students who were frequently absent.
- Interested in primary effect for targeted student, and spillovers to siblings.
Assign fixed number $N_1$ of households to treatment
Our focus: are there spillovers to “exposed” units?
Typically, we consider only two potential outcomes of unit $i$ (=student) for treatment and control, $Y_i(1), Y_i(0)$.
Typically, we consider only two potential outcomes of unit $i$ (=student) for treatment and control, $Y_i(1), Y_i(0)$.

In absenteeism study, we assume three potential outcomes:

1. $Y_i(\text{treated}) = \text{unit treated, siblings in control}$,
2. $Y_i(\text{control}) = \text{unit in control, siblings in control}$,
3. $Y_i(\text{exposed}) = \text{unit in control, one sibling treated}$.

Outcome of some unit $i$ depends on treatment of others—this is called interference.

We want a randomization test of “no spillovers” hypothesis:

$$H_0 : Y_i(\text{exposed}) = Y_i(\text{control}) .$$
Quick overview of classical randomization tests without interference.

Extension of classical tests to settings with interference.

Application to absenteeism data.
Randomization test without interference (Fisher, 1930)

- Cross-fertilization has effect over self-fertilization?
- Test no treatment effect hypothesis:

\[ H_0 : Y_i(\text{self}) = Y_i(\text{cross}). \]
Randomization test without interference (Fisher, 1930)

- Cross-fertilization has effect over self-fertilization?
- Test no treatment effect hypothesis:

\[ H_0 : Y_i(\text{self}) = Y_i(\text{cross}). \]

<table>
<thead>
<tr>
<th>unit</th>
<th>( Z_i )</th>
<th>self-fertilized</th>
<th>cross-fertilized</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>?</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>20</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>?</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>?</td>
</tr>
</tbody>
</table>

- Observed diff in means: \( t^{\text{obs}} = 13 - 11 = +2 \) (test statistic).
- Standard error?
Sample counterfactual $Z'$ according to design.

Observed diff in means for $Z'$ is $+12$. We imputed unknown values based on null hypothesis.

All possible values of test statistic: $\pm 12, \pm 6, \pm 2$. So, $t^{\text{obs}} = 2$ is not significant ($p$-value = 0.5).

Benefits of this test: valid in finite samples; no model assumptions; invariance to data transformations.
Problem with interference

Under $Z$:

- treated (1,1)
- exposed (0,1)
- exposed (0,1)
Problem with interference

Under $Z$: 

Under $Z'$:

Under $H_0$ of no spillovers, we can impute the outcomes of the exposed units (2 and 3).

But we can't impute the outcomes for unit 1 ($H_0$ is not sharp enough). => That unit should not be used in the test statistic.
Aronow (2012) and Athey, Eckles, Imbens (2017) proposed conditioning the randomization test:

1. Select (at random, or based on covariates) a set $F$ of focal units.
2. Define test statistic only on units in $F$.
3. Calculate assignments for which $H_0$ is sharp for all focal units:
4. Run Fisher test uniformly on set from (3).

Test is straightforward (conceptually) and valid.
Aronow (2012) and Athey, Eckles, Imbens (2017) proposed conditioning the randomization test:

1. Select (at random, or based on covariates) a set $F$ of focal units.
2. Define test statistic only on units in $F$.
3. Calculate assignments for which $H_0$ is sharp for all focal units:
4. Run Fisher test uniformly on set from (3).

---

- Test is straightforward (conceptually) and valid.
- But there are some subtle problems:
  1. focal units are dropped from study if $H_0$ is not sharp for them;
  2. test is not a permutation test, in general.
Avoid having to drop units.

**Test 1:**
1. In a treated household, choose one exposed unit at random.
2. In a control household, choose any one unit at random.
3. Perform a test by *permuting* on those focal units.

**Test 2:**
1. Block households in terms of size.
2. Choose units as in test 1 within blocks.
3. Permute the treatment of selected units within blocks.

Benefits of this approach:
⇒ More testing power + valid permutation test.
How do we get a permutation test?

- In figure above, one focal unit is exposed and one is in control, with equal probabilities.

- Conditional on focal selection, treatment status is a coin flip ⇒ permutation test.

- **Important:** No need to calculate/know what these probabilities are since we get a coin flip by conditioning. Permutation tests require such symmetry.
In the paper, we formalize conditioning of the randomization test as a conditioning event $C$. Analyst can decide $P(C|Z)$, the conditioning mechanism.

For validity, the randomization distribution should be:

$$P(Z|C) = \underbrace{P(C|Z)}_{\text{conditioning mechanism}} \times \underbrace{pr(Z)}_{\text{experiment design}}$$

- $P(C|Z)$ should correct asymmetries in the design, so that a permutation test is possible.

- Full characterization is hard. Depends on the application. But symmetries in design are usually evident (e.g., randomize on household level $\Rightarrow$ symmetry wrt to households, etc.)
Distribution of p-values over choices of focals for primary effect (left) and spillovers (right).
Power simulations

We set 500 households with 10 units each. Outcome model: $Y_i(1, 1) = Y_i(0, 0) + \tau$, and $Y_i(0, 0) \sim \mathcal{N}(0, \sigma^2)$. 
Randomization tests are appealing – minimal assumptions.

Conditioning mechanisms can lead to powerful permutation tests with appropriate construction to produce necessary symmetries.

Current and future work:
- Aggregate p-values across many possible conditioning events.
- Extend to more complicated interference (e.g., peer effects in general networks).
