When are conditional randomization tests also permutation tests?

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

Econometrics and Statistics
University of Chicago, Booth School of Business

IMSI Workshop on "Permutation and Causal Inference"
Chicago, IL, 2023

Introduction

Standard causal inference assumes no interference;
i.e., a unit's treatment cannot affect other units.

This describes a simple, static world.

In many interesting problems, units interact in a complex way.
—spillovers, peer effects, contagion, equilibrium effects, etc.

Pervasive in most social studies.

New methods and tools are needed. Many applications:
e.g., policy making, marketplace algorithms, climate science, healthcare.

Overview

Many current approaches tend to be heavily model-based.

In complex domains, this causes problems with inference and even with identification (e.g., "Perils of peer effects" by J. Angrist)

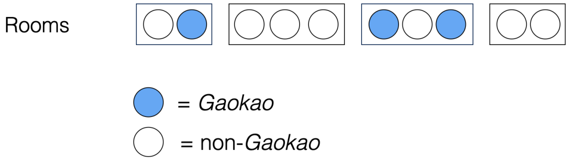
Randomization tests are nonparametric procedures that are **model-agnostic** and **finite-sample exact**.

However, they are limited in scope.

A lot of recent research work in extending the scope of randomization tests to complex domains. I will present such a line of work today.

Motivation: Peer effects (Li et al, 2019)

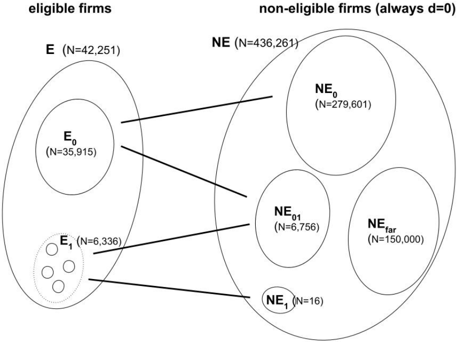
- Consider an experiment where students in a Chinese university are randomly assigned into dorm rooms.
- Each student has a binary attribute depending on whether they passed an entrance exam ($A_i = 1$) known as *Gaokao*.



- Is there an effect on academic outcome of being roommates with a Gaokao student?
- Can we test this *via permutations*?

Spillovers in a large interfirm network

- In an ongoing field experiment — with M Best, F Grosset (Columbia) — we need to study spillover effects between firms from tax audit notices.
- This is a highly complex setting with > 400,000 units, 8 million edges (firm connections).
- Fast procedures is a requirement, not a luxury!



Setup

- Units $i = 1, 2 \ldots, N$.
- $\mathbf{L} = (L_1, \ldots, L_N)$ — room (label) assignment
- $\mathbf{A} = (A_1, \ldots, A_N)$ — binary attributes.
- $\mathbf{Y} = (Y_1, \ldots, Y_N) \in \mathbb{R}^N$ — outcomes (e.g., grade improvement)
- Will use $(\mathbf{Y}^*, \mathbf{L}^*)$ for counterfactual outcomes-treatments.

Moreover,

- $Y_i(\boldsymbol{\ell})$ — potential outcome of i under room assignment $\boldsymbol{\ell}$.
- $P(\mathbf{L})$ is known and under our control (experimental study).

No Interference

- In classical causal inference, every unit i has only two potential outcomes, namely "$Y_i(0), Y_i(1)$" for control and treatment, respectively.
- However, in many problems there is interference.
- In our setting, a unit is exposed to "something more" than just a room assignment, perhaps a sum effect from the attributes of its roommates, and/or neighors, etc.
- Think also of a vaccine trial. A control unit (unvaccinated) is still "protected" by treated units (vaccinated) in its proximity.

Effective treatments

- Model interference as receiving a combined treatment exposure :

$$W_i = w_i(\mathbf{L}) \in \mathbb{W}.$$

- Examples:
  - $W_i = \sum_{j \neq i} A_j 1(L_i = L_j)$ — no. Gaokao roommates.
  - $W_i =$% of Gaokao roommates; etc.

Effective treatments

- Model interference as receiving a combined treatment exposure :

$$W_i = w_i(\mathbf{L}) \in \mathbb{W}.$$

- Examples:
  - $W_i = \sum_{j \neq i} A_j 1(L_i = L_j)$ — no. Gaokao roommates.
  - $W_i =$% of Gaokao roommates; etc.

- Although not necessary, it is useful to think that the exposure is the "effective treatment" (Manski, 2013); i.e.,

Assumption

$$Y_i(\boldsymbol{\ell}) = Y_i(\boldsymbol{\ell}') \quad \textit{for all } \boldsymbol{\ell}, \boldsymbol{\ell}', i \textit{ if } W_i = W_i'.$$

- Notation: Under the assumption, we can use $Y_i^{\omega}(\mathrm{w})$ to denote potential outcomes under $\mathbf{L}$ that generates exposure $W_i = \mathrm{w}$.

Hypotheses under interference

A large class of hypotheses under interference may be expressed as (main focus of this talk):

$$H_0 : Y_i^{\omega}(\mathsf{w}) = Y_i^{\omega}(\mathsf{w}') \text{ for all } i, \text{ and } \mathsf{w}, \mathsf{w}' \in \mathbb{W}_0 \subseteq \mathbb{W}.$$

(Manski, 2009), (Aronow, 2012), (T. and Kao, 2013), (Bowers et al., 2013), (Athey et al., 2019), Basse et al, 2019), (Puelz et al, 2021).

- e.g., $\mathbb{W}_0 = \{ 0, 1 \}$ whereas $\mathbb{W} = \{0, 1, \ldots\}$. That is, "no difference in outcomes from having 0 or 1 Gaokao roommate".
- Testing such hypotheses via randomization tests is **challenging**, however. Naive randomization / permutation can fail.

Review: Fisher's Randomization Test

Let's start with a simple problem.

When $\mathbb{W}_0 = \mathbb{W}$, then all exposures give identical outcomes under the null. This is **equivalent to** the global null of no effect:

$$H_0 : Y_i^{\omega}(\mathsf{w}) = Y_i^{\omega}(\mathsf{w}') \text{ for all } \mathsf{w}, \mathsf{w}', i.$$

This can be tested through Fisher's randomization test (Fisher, 1935),

1. Calculate test statistic, $T = t(\mathbf{W}, \mathbf{Y})$; e.g., regression coefficient
2. pval $= \mathsf{E}[t(\mathbf{W}^*, \mathbf{Y}) > T], \quad \mathbf{W}^* = w(\mathbf{L}^*), \mathbf{L}^* \sim P$.

Review: Fisher's Randomization Test

Let's start with a simple problem.

When $\mathbb{W}_0 = \mathbb{W}$, then all exposures give identical outcomes under the null. This is **equivalent to** the global null of no effect:

$$H_0 : Y_i^{\omega}(\mathsf{w}) = Y_i^{\omega}(\mathsf{w}') \text{ for all } \mathsf{w}, \mathsf{w}', i.$$

This can be tested through Fisher's randomization test (Fisher, 1935),

1. Calculate test statistic, $T = t(\mathbf{W}, \mathbf{Y})$; e.g., regression coefficient
2. pval $= \mathsf{E}[t(\mathbf{W}^*, \mathbf{Y}) > T]$, $\mathbf{W}^* = w(\mathbf{L}^*), \mathbf{L}^* \sim P$.

♠ Works because $t(\mathbf{W}^*, \mathbf{Y}) \overset{H_0}{=} t(\mathbf{W}^*, \mathbf{Y}^*) \overset{d}{=} T$.

An assessment of FRT

Main advantages:

- The test is exact in finite samples. No asymptotics.
- Not necessary to have correct Y-model specification.
- The test is robust. Same answer under transformations of Y.

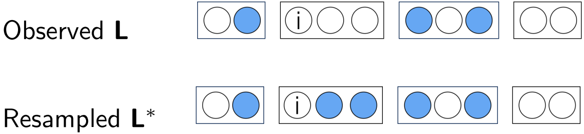
Common criticism:

- Can only test "strong" hypotheses. This is changing though. (This talk. Also, a lot of related research activity recently).
- Cannot generalize to population.

* How can we extend this to testing for spillovers?

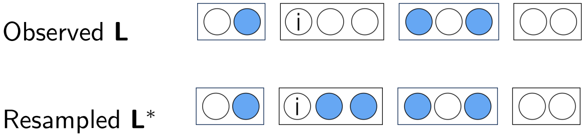
FRT problems under interference

For testing $H_0 : Y_i^{\omega}(0) = Y_i^{\omega}(1)$, suppose we naively resample $\mathbf{L}^*$ in the FRT as shown below:



FRT problems under interference

For testing $H_0 : Y_i^\omega(0) = Y_i^\omega(1)$, suppose we naively resample **L*** in the FRT as shown below:



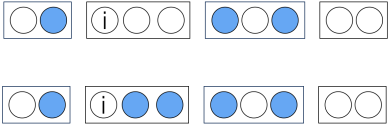
- Under **L** we observed $Y_i^\omega(0)$ for unit i.
- Under **L***, the unit has outcome $Y_i^\omega(2)$. But this outcome is **unknown** even under the null hypothesis (the "null is weak").

FRT problems under interference

For testing $H_0 : Y_i^{\omega}(0) = Y_i^{\omega}(1)$, suppose we naively resample $\mathbf{L}^*$ in the FRT as shown below:

Observed $\mathbf{L}$

Resampled $\mathbf{L}^*$



- Under $\mathbf{L}$ we observed $Y_i^{\omega}(0)$ for unit i.
- Under $\mathbf{L}^*$, the unit has outcome $Y_i^{\omega}(2)$. But this outcome is **unknown** even under the null hypothesis (the "null is weak").

$\star$ The test should employ a subset of units/assignments for which imputation is possible. How?

A recent development

Athey et al (2019), Basse et al (2019) recently proposed a general approach to apply FRTs under interference.

- Let $\mathbf{U} = (U_1, \ldots, U_N) \in \{0, 1\}^N$ denote a subset of units.
- Then, the idea is to run FRT on the subset of these *focal units*:

$$(Y_i, W_i, .. : U_i = 1)$$

under the following requirements:

1. The potential outcomes of *all* focal units should be *imputable* under the null, H_0.
2. (optional?) The resulting conditional randomization test should be easy to implement.

A general procedure

Specifically:

1. $P(\mathbf{U}) \sim$ Unif; i.e., pick focal units uniformly at random.
2. Enumerate:

$$\mathcal{W}_U = \left\{ \mathbf{W}' : Y_i^{\omega}(W_i') \text{ imputable under } H_0 \text{ for all } i \text{ with } U_i = 1. \right\}$$

3. Test statistic should depend only on data from focals: e.g.,

$$Y_i \sim W_i + X_i + ... \ (i : U_i = 1).$$

4. Run FRT by resampling from:

$$P(\mathbf{W}^* \mid \mathbf{U}) \propto 1(\mathbf{W}^* \in \mathcal{W}_U) P(\mathbf{W}^*). \tag{1}$$

A general procedure

Specifically:

1. $P(\mathbf{U}) \sim$ Unif; i.e., pick focal units uniformly at random.
2. Enumerate:

$$\mathcal{W}_U = \left\{\mathbf{W}' : Y_i^{\omega}(W_i') \text{ imputable under } H_0 \text{ for all } i \text{ with } U_i = 1.\right\}$$

3. Test statistic should depend only on data from focals: e.g.,

$$Y_i \ \sim W_i + X_i + ... \ (i : U_i = 1).$$

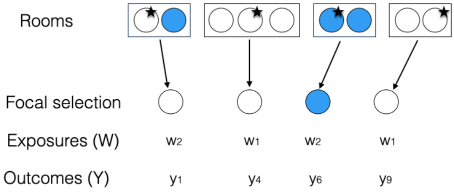
4. Run FRT by resampling from:

$$P(\mathbf{W}^* \mid \mathbf{U}) \propto 1(\mathbf{W}^* \in \mathcal{W}_U) P(\mathbf{W}^*). \tag{1}$$

- Uniform $P(\mathbf{U})$ is limiting — Basse et al (2019) extended to $P(\mathbf{U} \mid \mathbf{Z})$.
- Constructing $\mathcal{W}_U$ is challenging. Puelz et al (2021) developed a method to automate it, but still requires computation.
- **(this talk)**: Distribution (1) does not generally imply a permutation test. Such tests are ideal for computation.

Permutation test for spillovers?

- $H_0 : Y_i^\omega(\mathrm{w}_1) = Y_i^\omega(\mathrm{w}_2)$, with $\mathrm{w}_1 = 0$ and $\mathrm{w}_2 = 1$.



- This approach requires enumerating all assignments for which the focal units are exposed to $\{\mathrm{w}_1, \mathrm{w}_2\}$. This grows exponentially in N.
- Couldn't we just run a permutation test between Y and W on the focal units shown above?

Naive permutation fails

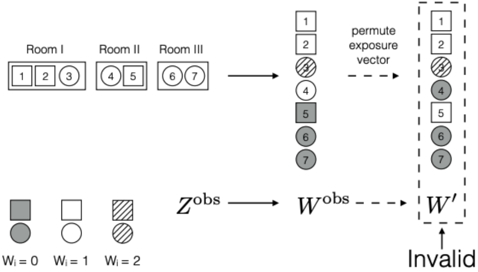


Figure: square=Gaokao student; circle = non-Gaokao

- In this example, we permute the exposures of units 4 and 5.
- However, the resulting $\mathbf{W}'$ is invalid. It cannot be generated from the design since it would require that 1,2, and 5 (Gaokao students) all have exactly one Gaokao roommate.

The problem, in summary

To summarize:

- Room assignment $\mathbf{L}$ according to known design, $P(\mathbf{L})$.
- Eff. treatment due to interference: $\mathbf{W} = w(\mathbf{L}) = (W_1, \ldots, W_N)$.
  But $w()$ can be **arbitrary!** (user-defined)
- Generalized SUTVA: $Y_i(\mathbf{L}) = Y_i^{\omega}(W_i)$.
- Goal is to test $H_0 : Y_i^{\omega}(\mathrm{w}) = Y_i^{\omega}(\mathrm{w}')$ for all $\mathrm{w}, \mathrm{w}' \in \mathbb{W}_0$.

$\star$ Can we test H_0 via permutations on (a subvector of) $\mathbf{W}$?

Main theorem

Theorem (informal)

Let $U = u(\mathbf{L})$ be the focal selection function. Let $\mathsf{S}_{A,U}$ be the permutation subgroup that leaves **A** *(attributes) and* **U** *(focals) unchanged. Suppose:*

- (a) $P(\mathbf{L}) = P(\pi\mathbf{L})$ *for all* $\pi \in \mathsf{S}_{A,U}$.
- (b) $w(\mathbf{L})$ *is* **equivariant** *with respect to* $\mathsf{S}_{A,U}$*; i.e.,* $w(\pi\mathbf{L}) = \pi w(\mathbf{L})$.
- (c) $u(\mathbf{L})$ *is equivariant with respect to* $\mathsf{S}_{A,U}$.

Then, **W** *is uniformly distributed within an orbit generated by* $\mathsf{S}_{A,U}$.

- The theorem shows that the procedure that permutes the exposures of focal units *stratified by attribute* (i.e., permutations in $\mathsf{S}_{A,U}$) is finite-sample valid.
- Note the "interaction" between (a) the design, (b) the exposure mapping, and (c) the conditioning (focal selection).

Proof sketch

Let $\mathcal{O} = \{\pi \mathbf{W}^{obs} : \pi \in \mathsf{S}_{A,U}\}$ = orbit generated by observed exposure.
Goal is to show that $\mathbf{W} \mid \mathbf{U}$ is uniform in $\mathcal{O}$ (implies permutation test!)

Proof sketch

Let $\mathcal{O} = \{\pi \mathbf{W}^{obs} : \pi \in \mathsf{S}_{A,U}\}$ = orbit generated by observed exposure.
Goal is to show that $\mathbf{W} \mid \mathbf{U}$ is uniform in $\mathcal{O}$ (implies permutation test!)

- Then,

$$\begin{aligned} P(\mathbf{W} \in \mathcal{O}, \mathbf{U} \mid \pi\mathbf{L}) &= 1\{w(\pi\mathbf{L}) \in \mathcal{O}\}\ 1\{\mathbf{U} = u(\pi\mathbf{L})\} \\ &= 1\{\pi w(\mathbf{L}) \in \mathcal{O}\}\ 1\{\mathbf{U} = \pi u(\mathbf{L})\} \qquad \text{conditions (b),(c)} \\ &= 1\{w(\mathbf{L}) \in \mathcal{O}\}\ 1\{\pi^{-1}\mathbf{U} = u(\mathbf{L})\} \qquad \text{orbit property} \\ &= 1\{w(\mathbf{L}) \in \mathcal{O}\}\ 1\{\mathbf{U} = u(\mathbf{L})\} \qquad \pi^{-1}\mathbf{U} = \mathbf{U} \text{ since } \pi \in \mathsf{S}_{A,U} \\ &= P(\mathbf{W} \in \mathcal{O}, \mathbf{U} \mid \mathbf{L}) \end{aligned}$$

Proof sketch

Let $\mathcal{O} = \{\pi \mathbf{W}^{obs} : \pi \in \mathsf{S}_{A,U}\}$ = orbit generated by observed exposure.
Goal is to show that $\mathbf{W} \mid \mathbf{U}$ is uniform in $\mathcal{O}$ (implies permutation test!)

- Then,

$$\begin{aligned} P(\mathbf{W} \in \mathcal{O}, \mathbf{U} \mid \pi\mathbf{L}) &= 1\{w(\pi\mathbf{L}) \in \mathcal{O}\}\ 1\{\mathbf{U} = u(\pi\mathbf{L})\} \\ &= 1\{\pi w(\mathbf{L}) \in \mathcal{O}\}\ 1\{\mathbf{U} = \pi u(\mathbf{L})\} \qquad \text{conditions (b),(c)} \\ &= 1\{w(\mathbf{L}) \in \mathcal{O}\}\ 1\{\pi^{-1}\mathbf{U} = u(\mathbf{L})\} \qquad \text{orbit property} \\ &= 1\{w(\mathbf{L}) \in \mathcal{O}\}\ 1\{\mathbf{U} = u(\mathbf{L})\} \qquad \pi^{-1}\mathbf{U} = \mathbf{U} \text{ since } \pi \in \mathsf{S}_{A,U} \\ &= P(\mathbf{W} \in \mathcal{O}, \mathbf{U} \mid \mathbf{L}) \end{aligned}$$

- From Bayes and condition (a), this implies that

$$P(\pi\mathbf{L} \mid \mathbf{W} \in \mathcal{O}, \mathbf{U}) = P(\mathbf{L} \mid \mathbf{W} \in \mathcal{O}, \mathbf{U}).$$

  That is, the design "maintains" its invariance even *conditional on the focals* within the subspace $\mathsf{S}_{A,U}$.

- Equivariance of $w(\mathbf{L})$ then implies the theorem.

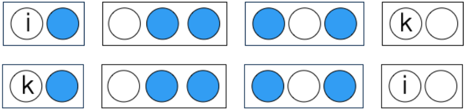
Application of theory — condition (b)

As exposure we defined

$$w_i(\mathbf{L}) = f(\{A_j : L_j = L_i, j \neq i\}) \quad \text{(for some known } f\text{)}.$$

Then, $w(\mathbf{L})$ satisfies equivariance (b).

To see this, if we swap the rooms of i, k (with $A_i = A_k$), then the exposures of i, k are transposed but all other exposures are unchanged.



Application of theory — condition (c)

Suppose we define

$$\mathbf{U} = u(\mathbf{L}) = 1\{w(\mathbf{L}) \in \mathbb{W}_0\};$$

i.e., "Focus on units that are exposed to the null exposure levels."

Then, $u(\mathbf{L})$ satisfies equivariance (c).

To see this:

$$u(\pi\mathbf{L}) = 1\{w(\pi\mathbf{L}) \in \mathbb{W}_0\} = 1\{\pi w(\mathbf{L}) \in \mathbb{W}_0\} = \pi 1\{w(\mathbf{L}) \in \mathbb{W}_0\} = \pi u(\mathbf{L}).$$

Condition (a): Design symmetry

The last condition is $P(\pi\mathbf{L}) = P(\mathbf{L})$. This is very mild under reasonable randomized designs; e.g.,

- Completely randomized design with a fixed number of units assigned to each room.
- Stratified randomized design with a fixed number of units assigned to each pair (room, attribute).
- ... etc

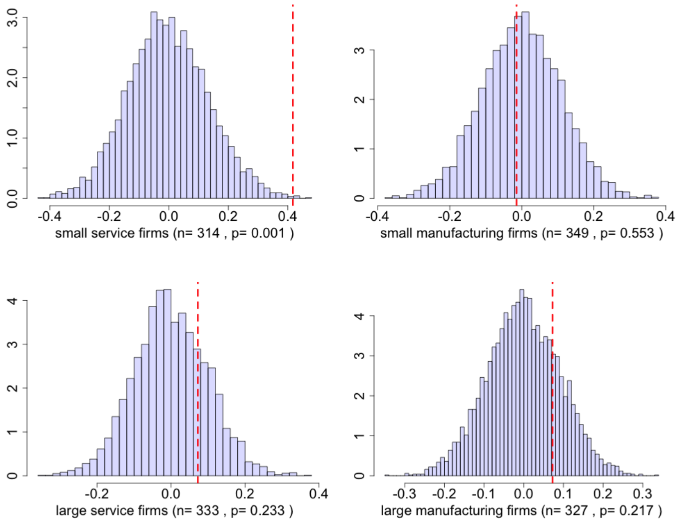
Application: (Cai and Szeidl,2017)

- Cai and Szeidl (2017) wanted to study the effect of business networks on firm performance.
- They randomized CEOs of various firms into working groups that met monthly for a year, and then tracked various firm performance metrics.
- The design was fairly complex (1,323 firms), but units were exchangeable given their size, sector, and location.
- Here, the salient attribute is 3-dim, $A_i = (\text{size}_i, \text{sector}_i, \text{region}_i)$. Our methodology can still be applied.

Analysis: Heterogeneity in peer effects

- An interesting feature of our method is that it allows analysis of heterogeneous peer effects. This can be done via conditional FRTs stratified by firm types.
- Cai and Szeidl (2017) also studied heterogeneity but in "direct effects". They showed that larger firms benefited more from the meetings.
- Our analysis complements this picture by showing that the impact of larger peers was concentrated mainly among small service firms.
- Regression specifications, such as (Cai and Szeidl, 2017), cannot easily capture peer effect heterogeneity due to model saturation.

Randomization test results



Concluding remarks

- Randomization tests are robust and finite-sample exact. They can (and should) be extended to problems with interference.
- This extension is made possible by clever conditioning procedures.
- However, conditional FRTs are computationally demanding.
- Permutation tests are computationally simple, but conditional FRTs are not always permutation tests.
- We given sufficient theoretical conditions to make the connection. But, are these conditions necessary?
- What about other designs? (e.g., two-stage, cluster)

Thank you!

- (*) Basse, Ding, Feller, Toulis "Randomization tests for group formation experiments" , (R&R, 2023)
- Puelz, Basse, Feller, Toulis "A graph-theoretic approach to randomization tests of causal effects under interference" , (JRSS-B, 2021)
- Basse, Feller, Toulis, "Randomization tests of causal effects under interference" (Biometrika, 2019)