SISCER Module 15
Lecture 1: Randomization inference

Ting Ye & Qingyuan Zhao

University of Washington & University of Cambridge

July 2022

Plan

- General theory for randomization inference for randomized experiments.
- Example 1: Fisher's exact test for 2×2 contigency tables.
- Example 2: Stepped-wedge cluster randomized trials.
- Example 3: Matched observational studies.

Recommended references for this lecture

Rosenbaum (2002, Chap. 2); Zhang and Zhao (2022).

Setup

Suppose there are n units (e.g. clinics or patients) in an experiment.

- **Covariates** $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots \boldsymbol{X}_n)$.
- **Treatment** $\boldsymbol{Z} \in \mathcal{Z}$ is randomly determined (e.g. by tossing coins or using the RNG in R).
- **Exposure** $\boldsymbol{A} = (A_1, \ldots, A_n)$ is determined by $\boldsymbol{Z}$.
  - Semantically, "treatment" speaks from the investigator's perspective and "exposure" from the experimental unit's perspective.
  - Often (but not always), $\boldsymbol{A} = \boldsymbol{Z}$ and these terms are used interchangeably.
- **Outcome** $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$.

Conceptulization of causality

Every possible treatment assignment $\boldsymbol{z}$ corresponds to a vector of **potential outcomes** $\boldsymbol{Y}(\boldsymbol{z}) = (Y_1(\boldsymbol{z}), \ldots, Y_n(\boldsymbol{z}))$.

Assumption: Consistency

The potential outcomes are related to the observed outcomes via $\boldsymbol{Y} = \boldsymbol{Y}(\boldsymbol{Z})$.

Assumption: Validity of exposure

The potential outcome of each unit only only depends on the exposure it receives.

$$Y_i(\boldsymbol{z}) = Y_i(\tilde{\boldsymbol{z}}) \text{ for all } i \text{ and } \boldsymbol{z}, \tilde{\boldsymbol{z}} \text{ such that } A_i(\boldsymbol{z}) = A_i(\tilde{\boldsymbol{z}}).$$

Simplification: The Neyman-Rubin causal model

- We often talk about potential outcomes under an exposure and denote it as $Y_i(a)$.
- We often further assume that the individual exposure is binary, so the **individual treatment effect** of unit i is defined as as $Y_i(1) - Y_i(0)$.

Fundamental problem of causal inference

- Only one potential outcome can ever be observed.
- But we would like to infer the full *potential outcomes schedule* $(\boldsymbol{Y}(\boldsymbol{z}))_{\boldsymbol{z}\in\mathcal{Z}}$.

i	$Y_i(0)$	$Y_i(1)$	A_i	Y_i
1	?	**1**	1	1
2	**0**	?	0	0
3	?	**0**	1	0
⋮	⋮	⋮	⋮	⋮

The role of randomization

Assumption: Exogeneity of randomization

The treatment is independent of the potential outcomes schedule given the covariates:

$$\boldsymbol{Z} \perp\!\!\!\perp (\boldsymbol{Y}(\boldsymbol{z}))_{\boldsymbol{z}\in\mathcal{Z}} \mid \boldsymbol{X}. \tag{1}$$

Furthermore, the conditional distribution of $\boldsymbol{Z}$ given $\boldsymbol{X}$ is given (often called the *randomization scheme* or *treatment assignment mechanism*).

Remarks

- If the randomization scheme does not use $\boldsymbol{X}$, it is not necessary to condition on $\boldsymbol{X}$ in (1).
- This assumption allows us to infer aspects of the potential outcomes schedule.

Imputation of potential outcomes

Next we will explore, in the Neyman-Rubin causal model, how to use randomization to test the *sharp null hypothesis* $H_0 : Y_i(0) = Y_i(1)$ for all i.

Key insight

Under H_0, we may impute all the potential outcomes by $Y_i(0) = Y_i(1) = Y_i$.

Example

i	$Y_i(0)$	$Y_i(1)$	A_i	Y_i
1	**1**	1	1	1
2	0	**0**	0	0
3	**0**	0	1	0
⋮	⋮	⋮	⋮	⋮

Randomization distribution

- Consider any test statistic $T = T(\boldsymbol{A}, \boldsymbol{Y})$.
- Under a potential treatment assignment $\boldsymbol{a} = (a_1, \ldots, a_n)$, the corresponding statistic is $T(\boldsymbol{a}) = T(\boldsymbol{A}, \boldsymbol{Y}(\boldsymbol{a}))$.
- The last insight suggests that under H_0, we know the value of $T(\boldsymbol{a})$ for every $\boldsymbol{a}$.
- The *randomization distribution* is that of $T(\boldsymbol{A})$ under the randomization scheme.

Example: An simple estimator of the average treatment effect

$$T = \frac{\sum_{i=1}^n A_i Y_i}{\sum_{i=1}^n A_i} - \frac{\sum_{i=1}^n (1 - A_i) Y_i}{\sum_{i=1}^n (1 - A_i)}.$$

i	$Y_i(0)$	$Y_i(1)$	A_i	Y_i
1	**1**	1	1	1
2	0	**0**	0	0
3	**0**	0	1	0

Equal probability (1/3) on $T(1,0,1) = 1/2$; $T(1,1,0) = 1/2$; $T(0,1,1) = -1$.

Randomization tests

We may reject the null hypothesis H_0 if the observed T is too extreme when compared to its potential values.

- The p-value is the probability that $T(\boldsymbol{A})$ exceeds the observed value.
- This is a valid test (controls type I error) due to an elementary fact in probability theory: Let $F(\cdot)$ be the cumulative distribution function (CDF) of a random variable T. Then $F(T)$ is (almost) uniformly distributed on $[0, 1]$.

Some remarks

- The key difference between this and the usual formulation of hypothesis testing (e.g. t-test and F-test) is that the **reference distribution is entirely based on randomization introduced by the experiment**.
- Without exogeneity of randomization $\boldsymbol{Z} \perp\!\!\!\perp (\boldsymbol{Y}(\boldsymbol{z}))_{\boldsymbol{z}\in\mathcal{Z}} \mid \boldsymbol{X}$, we do not know the randomization distribution of $T(\boldsymbol{A})$.
- The randomization distribution can be approximated by Monte Carlo.
- Randomization tests are particularly attractive for small sample sizes and complex designs/sampling (e.g. repeated measurements, individuals from the same household).

Example: Fisher's exact test and lady tasting tea

- A basic yet important statistical problem is hypothesis testing in 2×2 contingency tables.
- This is illustrated by a famous example in Fisher's 1935 book *The Design of Experiments*.

  A lady declares that by tasing a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup... Our experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgment in a random order. The subject has been told in advance of what the test will consist... Her task is to divide the 8 cups into two sets of 4.

- Exercise: What are the units, treatment, exposure, and outcome in this experiment?

2×2 contigency tables

- Let A_i be the exposure of the i-th cup (0/1 if milk/tea was added first).
- Let Y_i be the outcome of the i-th cup (0/1 if the lady guesses milk/tea was added first).
- Let N_{ay} be the number of cups with $A_i = a$ and $Y_i = y$, $a, y = 0, 1$.
- The outcome of this experiment can be summarized by the following 2×2 table.

		Outcome Y 0	1	Total
Treatment A	0	N_{00}	N_{01}	$N_{0\cdot}$
	1	N_{10}	N_{11}	$N_{1\cdot}$
	Total	$N_{\cdot 0}$	$N_{\cdot 1}$	N

Fisher's exact test

		Outcome Y 0	1	Total
Treatment A	0	N_{00}	N_{01}	$N_{0\cdot}$
	1	N_{10}	N_{11}	$N_{1\cdot}$
	Total	$N_{\cdot 0}$	$N_{\cdot 1}$	N

- Null hypothesis $H_0 : Y_i(0) = Y_i(1)$ for all i, meaning the lady's guess is random.
- $N_{0\cdot} = N_{1\cdot} = 4$ by design and $N_{\cdot 0} = N_{\cdot 1}$ by H_0.
- So there is only one degree of freedom: Given N_{00}, the entire table is known.
- Fisher showed that the probability of observing $(N_{00}, N_{01}, N_{10}, N_{11})$ is given by

$$\frac{N_{0\cdot}!N_{1\cdot}!N_{\cdot 0}!N_{\cdot 1}!}{N_{00}!N_{01}!N_{10}!N_{11}!N!}$$

- So we may reject H_0 if N_{00} is large (compared to this hypergeometric distribution).

Quasi-randomization tests

		Outcome Y 0	1	Total
Treatment A	0	N_{00}	N_{01}	$N_{0\cdot}$
	1	N_{10}	N_{11}	$N_{1\cdot}$
	Total	$N_{\cdot 0}$	$N_{\cdot 1}$	N

- Hypothesis testing is often presented in a different setup about an underlying distribution.
- For 2×2 tables, this is the "two binomials" problem: suppose $N_{01} \sim \text{Bin}(N_{0\cdot}, \pi_0)$, $N_{11} \sim \text{Bin}(N_{1\cdot}, \pi_1)$, and we would like to test $H_0' : \pi_0 = \pi_1$.
- It turns out that Fisher's exact test is also valid for this problem.
- But this same test has different epistemic basis in these two settings.
- Fisher's exact test for the two binomials problem is a **quasi-randomization** test.

Example: Stepped-wedge design and a real clinical trial

Haines et al., *PLOS Medicine*, 2017, DOI:10.1371/journal.pmed.1002412.

- The goal was to investigate the impact of disinvestment from weekend allied health services across acute medical and surgical wards.
- 12 wards in 2 hospitals were randomized to switch from an old model of weekend allied health services to no services, before adopting a new model of services. (You can visualize the design in Figure 1 of the article.)
- During this trial, a number of patient characteristics were collected. Of interest is the average length of stay in these wards.
- Exercise: What are the units, treatment, exposure, and outcome in this experiment?
- This example will be further explored in the R practical.

Example: Matched observational studies

- By matching units in an observational studies with very similar covariates, the hope is that we reconstruct a block randomized experiment.
- Consider the Neyman-Rubin causal model. Suppose treated observation $i = 1, \ldots, n$ is matched to control observation $i + n$. Define

$$\mathcal{M} = \{\boldsymbol{a}_{[2n_1]} \in \{0,1\}^{2n_1} \mid a_i + a_{i+n_1} = 1, \forall i \in [n_1]\}$$

- Randomization analysis of matched observational studies assumes

$$\mathbb{P}\left(\boldsymbol{A} = \boldsymbol{a} \,\middle|\, \boldsymbol{X}, \boldsymbol{A} \in \mathcal{M}\right) = \begin{cases} 2^{-n_1}, & \text{if } \boldsymbol{a} \in M, \\ 0, & \text{otherwise.} \end{cases}$$

- This would be satisfied if $(\boldsymbol{X}_i, A_i)$ are drawn i.i.d. (independent and identically distributed) from a population and the matching is exact.
- By further assuming no unmeasured confounders $A_i \perp\!\!\!\perp Y_i(a) \mid \boldsymbol{X}_i$ for all a, randomization tests can be constructed in the same way as before.

Rosenbaum, P. R. (2002). *Observational Studies*. Springer Series in Statistics. Springer, New York.

Zhang, Y. and Zhao, Q. (2022). What is a randomization test?