Segment 2: Randomized Studies

Section 02: Randomization-Based Inference

Randomization-Based Inference

For *population-level* inference we think of a sampling distribution dictated by:

- Randomly sampling units from a population
- Randomly assigning $\mathbf{Z}$ in accordance with the assignment mechanism

For *sample-level* inference

- We are interested in inferring something about potential outcomes/causal effects in a given sample
  - (e.g, τ_{SATE})
- We are not focused on repeated samples from a population
  - So what is random? Where does variability in a given estimate come from?
- View $\mathbf{Z}$ as the random quantity
- Envision re-randomization of the sample units in accordance with the assignment mechanism

Randomization-Based Inference

When estimating a sample average causal effect (τ_{SATE}):

- X, Y^c, Y^t in the sample are regarded as *fixed*
- $\mathbf{Z}$ is random
- "Sampling" different values of $\mathbf{Z}$ for this sample generates variability
  - From the assignment mechanism
  - Every value of $\mathbf{Z}$ gives different set of *observed* outcomes
- Will dictate a *randomization distribution* that plays a role analogous to a *sampling distribution*
- Will develop some intuition of how *study design* can dictate *uncertainty of (causal) estimates*

Randomization-Based Inference

Unbiasedness Under Completely Randomized Designs

Assignment mechanism:

$$Pr(\mathbf{Z}|\mathbf{X}, \mathbf{Y}^c, \mathbf{Y}^t) = \binom{n}{n_t}^{-1} \text{ if } \sum_{i=1}^{n} = n_t$$

and 0 otherwise, where n is the number of units in the sample and n_t is the desired number to receive the treatment

An example of an *ignorable* or *unconfounded* treatment design:

$$Z \perp\!\!\!\perp Y^t, Y^c$$

Observed-data comparison $\hat{\tau} = \frac{1}{n_t} \sum Y_i - \frac{1}{n_c} \sum Y_i$, is *unbiased* for τ_{PATE} and τ_{SATE}

How Design can Dictate Uncertainty

The rationale of randomization for unbiasedness is clear, but what about uncertainty?

- Unbiasedness tells us we should get the "right" answer *on average*
- Does not guarantee the estimate for any particular random assignment will be close to the truth
  - Could be "unlucky" and get poor *balance*
  - Particularly with small samples
- More "unlucky" randomizations $\Rightarrow$ more variability in estimates across $\mathbf{Z}$ $\Rightarrow$ more uncertainty
- Reducing the chances of "unlucky" randomizations $\rightarrow$ decrease the variance of effect estimates
- The comparison between Bernoulli randomization and the Completely Randomized Design is one example

Fisher's Randomization Inference

- RA Fisher was the first to graps the importance of randomization for assessing causal effects (1920s)
- **Basic idea** is to make inference based *solely on the assignment mechanism*
  - No modeling assumptions for the potential outcomes
  - Truly nonparametric
  - The assignment mechanism completely determines the randomization distribution of any specified test statistic
- Fisher focused on settings where
  - The assignment mechanism is **known**
  - Interest lies in a **sharp null hypothesis** of no causal effect whatsoever
  - Distinct from being interested in a "typical" or "average" causal effect

Pieces of Fisher's Exact Method

- A *known* randomized assignment mechanism
- Observed data $\mathbf{Z}$, $\mathbf{X}$, $\mathbf{Y}^{obs}$
- A *test statistic*, T
  - A function of observed data $(Y^{obs}, \mathbf{X})$ and the (stochastic) assignment vector $(\mathbf{Z})$
- A *sharp null hypothesis*
  - A null hypothesis that is sufficiently specific to "fill in" a hypothetical value for each unit's missing potential outcome
- The *randomization distribution*
  - Distribution of possible values of the test statistic under all possible assignments according to the assignment mechanism, assuming the null hypothesis is true
- A *p-value under the randomization distribution given the sharp null hypothesis*
  - Probability of observing a value of the test statistic at least as extreme under the null hypothesis

Implementing a Fisher's Exact Test

of the sharp null hypothesis

- Decide on a test statistic, T
- Specify the *sharp null hypothesis* of no causal effect whatsoever
  - $H_0 : Y_i^0 = Y_i^1 \quad \forall i$
  - Stronger than a hypothesis about the *average* causal effect
  - Unobserved potential outcome is *implied* for every unit under the sharp null:$Y_i^{miss} = Y_i^{obs}$
- Calculate the test statistic in the observed data, T^{obs}
- "Re-randomize" units from the assignment mechanism
  - Calculate T^{sim} for each simulation from the assignment mechanism, $\mathbf{Z}^{sim}$
- Compare T^{obs} to the distribution of T^{sim}
  - Calculate the exact p-value: $Pr(|T| > T^{obs})$
  - If T^{obs} looks very different than most of $T^{sim}, \Rightarrow$ reject H_0

Fisher's Exact Test: Toy Example

Observed Data, Completely Randomized Experiment

Unit	Z	Y^0	Y^1
1	0	66	?
2	0	0	?
3	0	0	?
4	1	?	0
5	1	?	607
6	1	?	436

Fisher's Exact Test: Toy Example

Observed Data, Completely Randomized Experiment

Unit	Z	Y^0	Y^1
1	0	66	?
2	0	0	?
3	0	0	?
4	1	?	0
5	1	?	607
6	1	?	436

- Test statistic, $T = T(\mathbf{Z}, Y^{obs})$:

$$\sum_i Z_i Y_i^{obs} / n_t - \sum_i (1 - Z_i) Y_i^{obs} / n_c$$

- $T^{obs} = 325.6$

Fisher's Exact Test: Toy Example

One "Re-Randomization", Completely Randomized Experiment

Unit	Z^{sim}	Y^0	Y^1
1	0	66	?
2	0	0	?
3	1	0	?
4	1	?	0
5	0	?	607
6	1	?	436

- $H_0 : Y_i^0 = Y_i^1$ for $i = 1, 2, \ldots, 6$
- Simulate $\mathbf{Z}$ from the randomization distribution
- "Fill in" potential outcomes *under the null hypothesis*
- Recalculate $T = T(\mathbf{Z}, Y^{obs})$
- $T^{sim} = -79.0$

Basis of test: Where does T^{obs} fall in the distribution of would-be values of T^{sim} under every possible assignment to $\mathbf{Z}$, assuming that H_0 is true?

The Randomization Distribution

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	T
0	0	0	1	1	1	325.6
0	0	1	0	1	1	325.6
0	0	1	1	0	1	-79.0
0	0	1	1	1	0	35.0
0	1	0	0	1	1	325.6
0	1	0	1	0	1	-79.0
0	1	0	1	1	0	35.0
0	1	1	0	0	1	-79.0
0	1	1	0	1	0	35.0
0	1	1	1	0	0	-369.6
1	0	0	0	1	1	369.6
1	0	0	1	0	1	-35.0
1	0	0	1	1	0	79.0
1	0	1	0	0	1	-35.0
1	0	1	0	1	0	79.0
1	0	1	1	0	0	-325.6
1	1	0	0	0	1	-35.0
1	1	0	0	1	0	79.0
1	1	0	1	0	0	-325.6
1	1	1	0	0	0	-325.6

Test Result

- How likely (under the randomization distribution) are we to have observed a value of the test statistic at least as large as T^{obs}?
- **Exact p-value**: $Pr(|T| \geq T^{obs}) = 8/20 = 0.40$ for the sharp null
- Under the sharp null of absolutely no effect, the observed difference could be due to chance

Comments and Extensions

- With large n, impossible to enumerate all possible $\mathbf{Z}^{sim}$
  - Simulate a (large) random sample of $\mathbf{Z}^{sim}$ instead
- Extensions to other randomization schemes straightforward
  - If we know the **assignment mechanism** we can always simulate $\mathbf{Z}^{sim}$ from it
- Nonzero null: $H_0 : Y_i^1 - Y_i^0 = \tau$
  - Equivalently, $H_0 : Y^1 = Y^0 + \tau$
- Fisher interval estimates
  - Calculate Fisher p-value under a range of possible τ
  - The uncertainty interval for τ will be all values for which we fail to reject $H_0 : Y^0 = Y^1 + \tau$
- Choice of test statistics
  - Many options, context/problem dependent