Randomized Experiments and Randomization Inference

Dr. Florian Foos

Department of Government
London School of Economics & Political Science

f.foos@lse.ac.uk
@florianfoos

September 16, 2019

Random assignment

Random assignment

- "Experiment": a study in which subjects are assigned to treatment (and control) with a known probability between 0 and 1.

Random assignment

- "Experiment": a study in which subjects are assigned to treatment (and control) with a known probability between 0 and 1.
- Random assignment: the probability of assignment to treatment (and control) is equal for each subject.

Random assignment

- "Experiment": a study in which subjects are assigned to treatment (and control) with a known probability between 0 and 1.
- Random assignment: the probability of assignment to treatment (and control) is equal for each subject.
- That means no subjects has a higher probability to be treated than another subject.

Recap: Potential outcomes

Recap: Potential outcomes

- $Y_i(1)$

Recap: Potential outcomes

- $Y_i(1)$ is the potential outcome if the ith subject was treated.

Recap: Potential outcomes

- $Y_i(1)$ is the potential outcome if the ith subject was treated.
- $Y_i(0)$

Recap: Potential outcomes

- $Y_i(1)$ is the potential outcome if the ith subject was treated.
- $Y_i(0)$ is the potential outcome if the ith subject was not treated.

Recap: Conditional potential outcomes

Recap: Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$

Recap: Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.

Recap: Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.
- $Y_i(1) \mid D_i = 0$

Recap: Conditional potential outcomes

- $Y_i(0) \mid d_i = 1$: untreated potential outcome for subjects that receive the treatment.
- $Y_i(1) \mid D_i = 0$: treated potential outcome for subjects that would not receive the treatment under a hypothetical random assignment.

Recap: Treatments as random variables

Recap: Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.

Recap: Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.
- D_i is a random variable (the ith subject might be treated in one hypothetical study and not in another).

Recap: Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.
- D_i is a random variable (the ith subject might be treated in one hypothetical study and not in another).
- $Y_i(0) \mid D_i = 1$

Recap: Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.
- D_i is a random variable (the ith subject might be treated in one hypothetical study and not in another).
- $Y_i(0) \mid D_i = 1$: untreated potential outcome for subjects that would receive the treatment under a hypothetical random assignment.

Recap: Treatments as random variables

- We distinguish between d_i the treatment that a given subject receives and D_i, the treatment that could be administered hypothetically.
- D_i is a random variable (the ith subject might be treated in one hypothetical study and not in another).
- $Y_i(0) \mid D_i = 1$: untreated potential outcome for subjects that would receive the treatment under a hypothetical random assignment.
- We use D_i when talking about the statistical properties of treatments.

Example

- What's the effect of private tutoring on exam scores (ranging from 1 to 6)?

Estimator and estimand

Estimator and estimand

- Without prior knowledge, an estimate from just one experiment is only a best guess about the true value of the ATE. Any one ATE might be a little too high or a little too low.

Estimator and estimand

- Without prior knowledge, an estimate from just one experiment is only a best guess about the true value of the ATE. Any one ATE might be a little too high or a little too low.
- Our data set is just one of many possible data sets that could have been created via random assignment. If we would redo the exact same random assignment procedure, different units would be allocated to treatment and control groups!

Estimator and estimand

Estimator and estimand

- The average estimated ATE across all possible random assignments is equal to the true ATE.

Estimator and estimand

- The average estimated ATE across all possible random assignments is equal to the true ATE.
- On average we recover the true ATE. Our *estimator* is unbiased.

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?
Average	4.5	5	

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?
Average	4.5	5	0.5

Observed Outcomes

	$Y_i(0)$	$Y_i(1)$	τ_i
	Test score if		Treatment
subject i	not tutored	tutored	effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	?	5.5	?
6	6	?	?

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	?	5.5	?
6	6	?	?
Average	5.17	4.83	

Observed Outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	?	5.5	?
6	6	?	?
Average	5.17	4.83	-0.32

Sampling distribution of the ATE

- In our example, how many different ways are there of assigning 3 out of 6 subjects to the treatment group?

Sampling distribution of the ATE

- In our example, how many different ways are there of assigning 3 out of 6 subjects to the treatment group?

Sampling distribution of the ATE

- In our example, how many different ways are there of assigning 3 out of 6 subjects to the treatment group?

$$\frac{N!}{m!(N-m)!} =$$

Sampling distribution of the ATE

- In our example, how many different ways are there of assigning 3 out of 6 subjects to the treatment group?

$$\frac{N!}{m!(N-m)!} = \frac{6!}{3!3!} =$$

Sampling distribution of the ATE

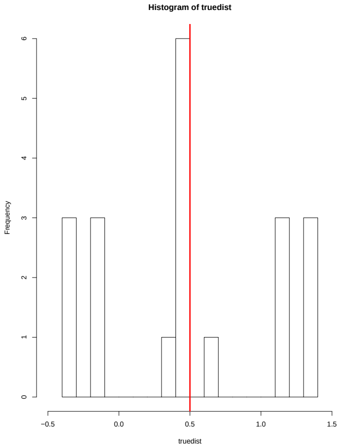
- In our example, how many different ways are there of assigning 3 out of 6 subjects to the treatment group?

$$\frac{N!}{m!(N-m)!} = \frac{6!}{3!3!} = \frac{720}{36} = 20$$

ATEs

	Estimated ATEs	Frequency
1	-0.33	3
2	-0.17	3
3	0.33	1
4	0.50	6
5	0.67	1
6	1.17	3
7	1.33	3
	0.5	20

Sampling Distribution of estimated ATEs



Randomization inference

Randomization inference

- Sharp null hypothesis: The treatment effect is exactly zero for each subject (Fisher's exact test).

Randomization inference

- Sharp null hypothesis: The treatment effect is exactly zero for each subject (Fisher's exact test).
- If the sharp null hypothesis is true, then $Y_i(0) = Y_i(1)$.

Randomization inference

- Sharp null hypothesis: The treatment effect is exactly zero for each subject (Fisher's exact test).
- If the sharp null hypothesis is true, then $Y_i(0) = Y_i(1)$.
- Under the sharp null hypothesis, we can take the observed outcomes in our data set, and impute the counterfactual potential outcomes, re-assigning subjects to treatment and control group over and over again.

Randomization inference

Randomization inference

- Simulated randomizations provide the exact sampling distribution of the estimated ATE under the sharp null.

Randomization inference

- Simulated randomizations provide the exact sampling distribution of the estimated ATE under the sharp null.
- Now we can calculate the number of times we obtain an estimated ATE at least as large as the one we obtained from our actual experiment if the treatment effect was zero for every subject.

Randomization inference

Randomization inference

- One-tailed test of sharp null:

Randomization inference

- One-tailed test of sharp null:

Randomization inference

- One-tailed test of sharp null:
    1. Count the number of ATEs under the sharp null that are as large or larger (or as small or smaller) than the actual ATE we obtain from our experiment.

Randomization inference

- One-tailed test of sharp null:
    1. Count the number of ATEs under the sharp null that are as large or larger (or as small or smaller) than the actual ATE we obtain from our experiment.
    2. Divide by the number of random assignments.

Randomization inference

- One-tailed test of sharp null:
  1. Count the number of ATEs under the sharp null that are as large or larger (or as small or smaller) than the actual ATE we obtain from our experiment.
  2. Divide by the number of random assignments.
  3. -> One-tailed p-value.

Randomization inference

- One-tailed test of sharp null:
  1. Count the number of ATEs under the sharp null that are as large or larger (or as small or smaller) than the actual ATE we obtain from our experiment.
  2. Divide by the number of random assignments.
  3. -> One-tailed p-value.
- Two-tailed test of sharp null:

Randomization inference

- One-tailed test of sharp null:
  1. Count the number of ATEs under the sharp null that are as large or larger (or as small or smaller) than the actual ATE we obtain from our experiment.
  2. Divide by the number of random assignments.
  3. -> One-tailed p-value.

- Two-tailed test of sharp null:
  1. Count the *absolute* number of ATEs under the sharp null that are as large or larger than the actual ATE we obtain from our experiment

Randomization inference

- One-tailed test of sharp null:
  1. Count the number of ATEs under the sharp null that are as large or larger (or as small or smaller) than the actual ATE we obtain from our experiment.
  2. Divide by the number of random assignments.
  3. -> One-tailed p-value.

- Two-tailed test of sharp null:
  1. Count the *absolute* number of ATEs under the sharp null that are as large or larger than the actual ATE we obtain from our experiment
  2. Divide by the number of random assignments.

Randomization inference

- One-tailed test of sharp null:
  1. Count the number of ATEs under the sharp null that are as large or larger (or as small or smaller) than the actual ATE we obtain from our experiment.
  2. Divide by the number of random assignments.
  3. -> One-tailed p-value.

- Two-tailed test of sharp null:
  1. Count the *absolute* number of ATEs under the sharp null that are as large or larger than the actual ATE we obtain from our experiment
  2. Divide by the number of random assignments.
  3. -> Two-tailed p-value.

Imputing the sharp null

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	?
2	5	?	?
3	?	4.5	?
4	4.5	?	?
5	4	?	?
6	?	6	?
Average	4.5	5	0.5

Imputing the sharp null

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	?	4.5	0
2	5	?	0
3	?	4.5	0
4	4.5	?	0
5	4	?	0
6	?	6	0
Average	4.5	5	0

Imputing the sharp null

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	4.5	4.5	0
2	5	5	0
3	4.5	4.5	0
4	4.5	4.5	0
5	4	4	0
6	6	6	0
Average	4.75	4.75	0

Re-assign subjects to treatment and control

Re-assign subjects to treatment and control

- Perform random assignment

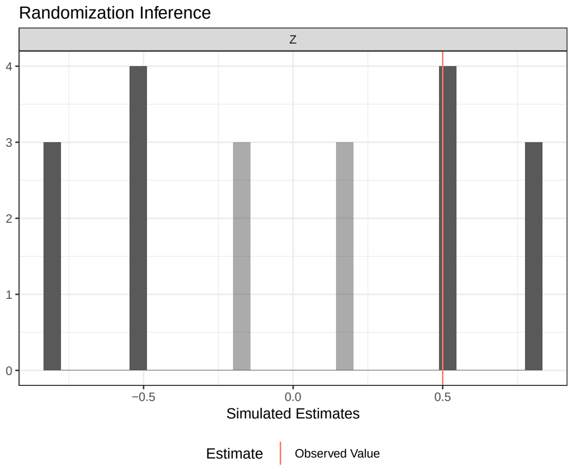
Re-assign subjects to treatment and control

- Perform random assignment
- Subjects reveal their outcomes under the sharp null.

$\widehat{ATE}$ under sharp null

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1		4.5	
2	5		
3		4.5	
4	4.5		
5		4	
6	6		
Average	5.17	4.33	-0.84

Samling distribution of ATEs if sharp null is true



Flexibility of approach

Flexibility of approach

- We can test different sharp hypotheses, not only the sharp null, is the effect 2 for each subject?

Flexibility of approach

- We can test different sharp hypotheses, not only the sharp null, is the effect 2 for each subject?
- We can use test statistics other than DIM, e.g. f-statistic, log-likelihood statistic, difference-in-variance.

Flexibility of approach

- We can test different sharp hypotheses, not only the sharp null, is the effect 2 for each subject?
- We can use test statistics other than DIM, e.g. f-statistic, log-likelihood statistic, difference-in-variance.
- When would we want to do that?
  1. Balance checks
  2. Testing interaction effects / treatment effect heterogeneity

Let's do randomization inference using the ri2 package in R.