Randomized Experiments and Randomization Inference

Dr. Florian Foos

Department of Political Economy
King's College London

florian.foos@kcl.ac.uk
@florianfoos

February 19, 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.

Random Assignment

Random Assignment

- Types of random assignment:

Random Assignment

- Types of random assignment:
  1. Simple random assignment: Each subject is allocated to the treatment group with probability m/N, where m is the number of subjects assigned to treatment.
  2. Complete random assignment (exactly m units are assigned to treatment)

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)?

Full schedule of potential outcomes

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

Definition of a subject-level treatment effect

- The individual level treatment effect τ_i for a given subject i is defined as:

Definition of a subject-level treatment effect

- The individual level treatment effect τ_i for a given subject i is defined as: $\tau_i = Y_i(1) - Y_i(0)$.

Full schedule of potential outcomes

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

Full schedule of potential outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	3	4.5	1.5
2	5	5	0
3	5	4.5	-0.5
4	4.5	5	0.5
5	4	5.5	1.5
6	6	6	0

Definition of Average Treatment Effect

$\mu(Y(1)) - \mu(Y(0))$,

where

$\mu(Y(1))$ is the average value of $Y_i(1)$ for all subjects, and

$\mu(Y(0))$ is the average value of $Y_i(0)$ for all subjects.

Full schedule of potential outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ tutored	τ_i Treatment effect
1	3	4.5	1.5
2	5	5	0
3	5	4.5	-0.5
4	4.5	5	0.5
5	4	5.5	1.5
6	6	6	0
Average			

Full schedule of potential outcomes

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	3	4.5	1.5
2	5	5	0
3	5	4.5	-0.5
4	4.5	5	0.5
5	4	5.5	1.5
6	6	6	0
Average	4.58	5.08	0.5

Potential and observed outcomes

Potential and observed outcomes

- The $Y_i(1)s$ are observed for subjects who are treated, and the $Y_i(0)s$ are observed for subjects who are not treated. For any given subject, we observe either $Y_i(1)$ or $Y_i(0)$, never both at the same time.

Potential and observed outcomes

- The $Y_i(1)s$ are observed for subjects who are treated, and the $Y_i(0)s$ are observed for subjects who are not treated. For any given subject, we observe either $Y_i(1)$ or $Y_i(0)$, never both at the same time.
- The fact that we observe only one potential outcome is called the "fundamental problem of causal inference" (Holland 1986).

Potential and observed outcomes

- The $Y_i(1)s$ are observed for subjects who are treated, and the $Y_i(0)s$ are observed for subjects who are not treated. For any given subject, we observe either $Y_i(1)$ or $Y_i(0)$, never both at the same time.
- The fact that we observe only one potential outcome is called the "fundamental problem of causal inference" (Holland 1986).
- A subject's treatment effect is unobserved.

Potential and observed outcomes

- The connection between the observed outcome Y_i and the underlying potential outcomes is given by the "switching equation":

$$Y_i = d_i Y_i(1) + (1 - d_i) Y_i(0)$$

Independence assumption

Treatment status is statistically independent of potential outcomes and background attributes (X):

$$Y_i(0), Y_i(1), X \perp\!\!\!\perp D_i$$

Independence assumption

Treatment status is statistically independent of potential outcomes and background attributes (X):

$$Y_i(0), Y_i(1), X \perp\!\!\!\perp D_i$$

If a subject is randomly assigned to treatment, knowing whether a subject is treated provides no information about the subject's potential outcomes, or background attributes.

Excludability assumption

Excludability assumption

- When we only define two potential outcomes, $Y_i(1)$ and $Y_i(0)$, based on whether the treatment is administered, we assume that the only relevant causal agent is receipt of the treatment.

Excludability assumption

- When we only define two potential outcomes, $Y_i(1)$ and $Y_i(0)$, based on whether the treatment is administered, we assume that the only relevant causal agent is receipt of the treatment.
- The exclusion restriction breaks down if treatment assignment z_i sets in motion causes of Y_i other than the treatment d_i.

Non-interference assumption

Non-interference assumption

- The value of the potential outcomes for subject i depend only on whether the subject itself is treated (whether d equals 1 or 0).

Non-interference assumption

- The value of the potential outcomes for subject i depend only on whether the subject itself is treated (whether d equals 1 or 0).
- Each subject is unaffected by the treatments and assignments of other units.

Expectations

The expectation of a discrete random variable is defined as

$$E[X] = \sum xPr[X = x],$$

Expectations

The expectation of a discrete random variable is defined as

$$E[X] = \sum xPr[X = x],$$

where $Pr[X = x]$ denotes the probability that X takes on the value x, and where the summation is taken over all possible values of x.

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

$$E[Y_i(0) \mid D_i = 1] = E[Y_i(0) \mid D_i = 0]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

$$E[Y_i(0) \mid D_i = 1] = E[Y_i(0) \mid D_i = 0] = E[Y_i(0)]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

$$E[Y_i(0) \mid D_i = 1] = E[Y_i(0) \mid D_i = 0] = E[Y_i(0)]$$

$$E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]$$

Definition of Average Treatment Effect

- Under random assignment:

$$E[Y_i(1) \mid D_i = 1] = E[Y_i(1) \mid D_i = 0] = E[Y_i(1)]$$

$$E[Y_i(0) \mid D_i = 1] = E[Y_i(0) \mid D_i = 0] = E[Y_i(0)]$$

$$E[Y_i(1) \mid D_i = 1] - E[Y_i(0) \mid D_i = 0]$$

$$= E[Y_i(1)] - E[Y_i(0)] = E[\tau_i] = ATE.$$

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

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

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!
- 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

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ tutored	τ_i Treatment 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 our of 6 subjects to the treatment group?

Sampling distribution of the ATE

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

Sampling distribution of the ATE

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

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

Sampling distribution of the ATE

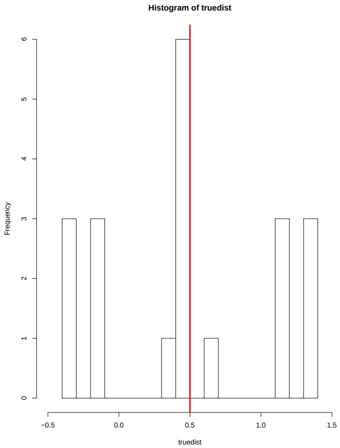
- In our example, how many different ways are there of assigning 3 our of 6 subjects to the treatment group?

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

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 zero for each subject (Fisher's exact test).

Randomization inference

- Sharp null hypothesis: The treatment effect is 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 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: Count the number of ATEs under the sharp null that are as large or larger (as small or smaller) than the actual ATE we obtain from our experiment

Randomization inference

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

Randomization inference

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

Randomization inference

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

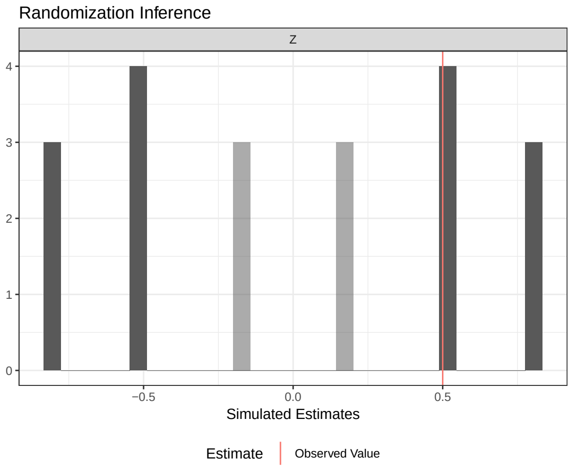
Imputing the sharp null

subject i	$Y_i(0)$ Test score if not tutored	$Y_i(1)$ Test score if tutored	τ_i Treatment effect
1	4.5	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

$\widehat{ATE}$ 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



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