

# Lecture 3: Randomized Experiments

Tianpei Xie

Sep. 16th., 2022

## Contents

<b>1</b>	<b>Randomized Experiments</b>	<b>2</b>
<b>2</b>	<b>Association is causation in Randomized Experiments</b>	<b>2</b>
2.1	Comparability and covariate balance . . . . .	2
2.2	Exchangeability . . . . .	3
2.3	No backdoor path . . . . .	3
<b>3</b>	<b>Causal Inference in Randomized Experiments</b>	<b>4</b>
3.1	Hypothesis testing . . . . .	4
3.1.1	Testing the null hypothesis of no treatment effect . . . . .	4
3.1.2	Noparametric tests . . . . .	4
3.2	Attributable effects . . . . .	4

# 1 Randomized Experiments

- **Randomized experiments** are noticeably different from **observational studies**.
    - In *randomized experiments*, the experimenter has **complete control** over the *treatment assignment mechanism* (how treatment is assigned); Ideally, the treatment assignment is not related to pre-treatment variables, or covariates.
- In randomized experiments, **association is causation**. This is because randomized experiments are special in that they guarantee that there is **no confounding**.
- In *observational studies*, the treatment is almost always a function of some covariate(s).

## 2 Association is causation in Randomized Experiments

### 2.1 Comparability and covariate balance

- Ideally, the treatment and control groups would be **the same, in all aspects**, *except for treatment*. This would mean they only differ in the treatment they receive (i.e. they are **comparable**).
- **Definition (Covariate Balance)** [Neal, 2020]  
We have **covariate balance** if the distribution of covariates is the same across treatment groups. More formally,

$$P(X|T = 1) = P(X|T = 0) \quad (1)$$

where  $X$  is the set of covariates and  $T$  is the treatment assignment.

- **Randomization implies covariate balance**, across **all** covariates, ***even unobserved ones***. Intuitively, this is because the treatment  $T$  is chosen at random, regardless of  $X$  i.e.  $(T \perp\!\!\!\perp X)$ .
- Under the covariate balance, we can show that  $p(y|do(t)) = p(y|t)$ , that is, association is causation.

**Proof:** First, let  $X$  be a sufficient adjustment set that potentially contains unobserved variables (randomization also balances unobserved covariates). By back-door adjustment, we have

$$p(y|do(t)) = \sum_x p(y|t, x) p(x)$$

Multiplying  $\frac{p(t|x)}{p(t|x)}$  we have

$$\begin{aligned} p(y|do(t)) &= \sum_x \frac{p(y|t, x) p(t|x) p(x)}{p(t|x)} \\ &= \sum_x \frac{p(y, t, x)}{p(t)} \text{ since } T \perp\!\!\!\perp X \\ &= \frac{p(y, t)}{p(t)} = p(y|t). \quad \blacksquare \end{aligned}$$

confounding association

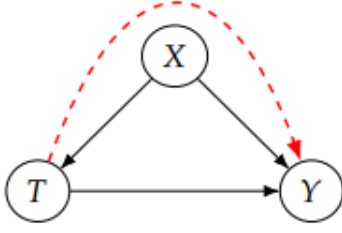


Figure 5.1: Causal structure of  $X$  confounding the effect of  $T$  on  $Y$ .

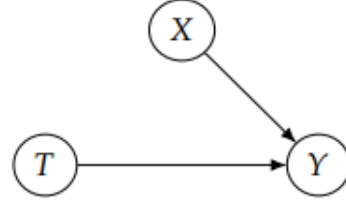


Figure 5.2: Causal structure when we randomize treatment.

Figure 1: (Left) Causal structure of  $X$  confounding the effect of  $T$  on  $Y$ . (Right) Causal structure when we randomize treatment. [Neal, 2020]

## 2.2 Exchangeability

- **Assumption 2.1** (*Ignorability / Exchangeability*) [Neal, 2020]

$$(Y(1), Y(0)) \perp\!\!\!\perp T \quad (2)$$

- Exchangeability means that the **treatment groups are exchangeable** in the sense that if they were **swapped**, the new treatment group would observe the **same outcomes** as the old treatment group, and the new control group would observe the same outcomes as the old control group. In this sense, our choice of test subjects are nothing special for treatment vs. control.

$$\begin{aligned} \mathbb{E}[Y(1)] &= \mathbb{E}[Y(1) | T = 0] = \mathbb{E}[Y(1) | T = 1] \\ \mathbb{E}[Y(0)] &= \mathbb{E}[Y(0) | T = 0] = \mathbb{E}[Y(0) | T = 1] \end{aligned}$$

- Intuitively, the randomized experiments are exchangeable since we can simply swap the definition of head and tail for the treatment assignment.

## 2.3 No backdoor path

- The final perspective that we will look at to see why association is causation in randomized experiments is that of *graphical causal models*. In regular observational data, there is almost always confounding. See Figure 1 (Left).
- However, if we **randomize**  $T$ ,  $T$  **no longer has any causal parents**. See Figure 1 (Right). This is because  $T$  is purely random. It doesn't depend on anything other than the output of a coin toss.
- Because  $T$  has no incoming edges, under randomization, **there are no backdoor paths**. So the **empty set is a sufficient adjustment set**.

This means that *all of the association that flows from  $T$  to  $Y$  is causal*. Using back-door adjustment for empty adjustment set, we have

$$P(Y | do(T = t)) = P(Y | T = t)$$

### 3 Causal Inference in Randomized Experiments

Under *completely randomized experiments*, association is causation, so *causal inference is the same as statistical inference* and the causal identification step is directly completed. We can use hypothesis testing to answer causal questions such as "if treatment has an effect or not".

#### 3.1 Hypothesis testing

##### 3.1.1 Testing the null hypothesis of no treatment effect

##### 3.1.2 Noparametric tests

#### 3.2 Attributable effects

## References

- Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.
- Brady Neal. Introduction to causal inference from a machine learning perspective. *Course Lecture Notes (draft)*, 2020.
- Judea Pearl. Causal inference. *Causality: objectives and assessment*, pages 39–58, 2010.
- Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.