Lecture 2
The potential outcome framework

Outline

- Definition of concepts: intervention, potential outcomes, causal effects
- The SUTVA assumption
- Causal estimand and assignment mechanism
- Perfect doctor example
- An alternative language: Causal directed acyclic graph (DAG)
- Suggested reading: Imbens and Rubin Chapter 1.1-1.8, 1.10, 3.3 Peng's book Chapter 1.3, 2

What is causality?

- Causality is tied to an action applied to a unit (intervention, treatment, manipulation)
- **“No causation without manipulation”** (Rubin, 1975)
- Manipulation need not be performed, but should be theoretically possible

Causal questions that we have encountered

- Will a new drug that increases HDL-C concentration reduce the risk of heart disease?
- If a smoking mother stops smoking , will the baby have a higher chance to survive?
- Will two doses of the Moderna vaccine stop the infection of COVID-19?
- If someone has a longer length of education, will his earnings in the future increase?

The potential outcome framework: basic concepts

- Treatment: An active intervention applied at a particular moment in time, whose effects we wish to assess relative to no intervention (the control)
  - Example: take the aspirin to treat headache
    - What should we compare?

- Unit: A physical object at a particular moment in time
  - Example: my headache before and after deciding to take or not to take the aspirin
    - Can we compare the before and after outcome for the same me?
      - No! Me at one time and me at another time are not the same "unit"

- Potential outcomes: Outcomes that would be observed if active treatment is applied and that would be observed if control treatment is applied

Unit	Potential Outcomes		Causal Effect
	Y(Aspirin)	Y(No Aspirin)	
You	No Headache	Headache	Improvement due to Aspirin

The potential outcome framework: basic concepts

- How to define a causal effect?

1. Headache gone only with aspirin:
   $Y(\text{Aspirin}) = \text{No Headache}$, $Y(\text{No Aspirin}) = \text{Headache}$
2. No effect of aspirin, with a headache in both cases:
   $Y(\text{Aspirin}) = \text{Headache}$, $Y(\text{No Aspirin}) = \text{Headache}$
3. No effect of aspirin, with the headache gone in both cases:
   $Y(\text{Aspirin}) = \text{No Headache}$, $Y(\text{No Aspirin}) = \text{No Headache}$
4. Headache gone only without aspirin:
   $Y(\text{Aspirin}) = \text{Headache}$, $Y(\text{No Aspirin}) = \text{No Headache}$

Positive causal effect of aspirin

No causal effect of aspirin

No causal effect of aspirin

Negative causal effect of aspirin

- Treatment: $W_i = 1$ if treated or $W_i = 0$ if control
- Causal effect for unit i: $\tau_i = Y_i(1) - Y_i(0)$

The potential outcome framework: basic concepts

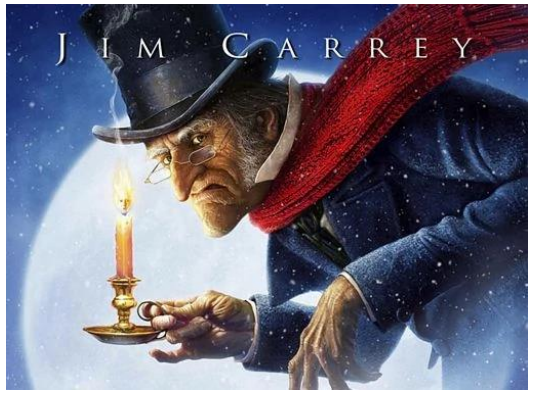
- Pre-treatment covariates: A background characteristic (measured or unmeasured) of a unit that could not have been affected by treatment assignment
  - Headache example: intensity of headache before making the decision to take aspirin or not
  - Evaluation of a job training program: age, previous educational achievement, family, socio-economic status, pre-training earnings, etc.
  - Pre-training earnings is a pre-treatment covariate, while post-training earnings are potential outcomes

Voters	Contact	Turnout		Age	Gender
i	W_i	$Y_i(1)$	$Y_i(0)$	X_{i1}	X_{i2}
1	1	1	?	20	M
2	0	?	0	55	F
3	0	?	1	40	F
⋮	⋮	⋮	⋮	⋮	⋮
n	1	0	?	62	M

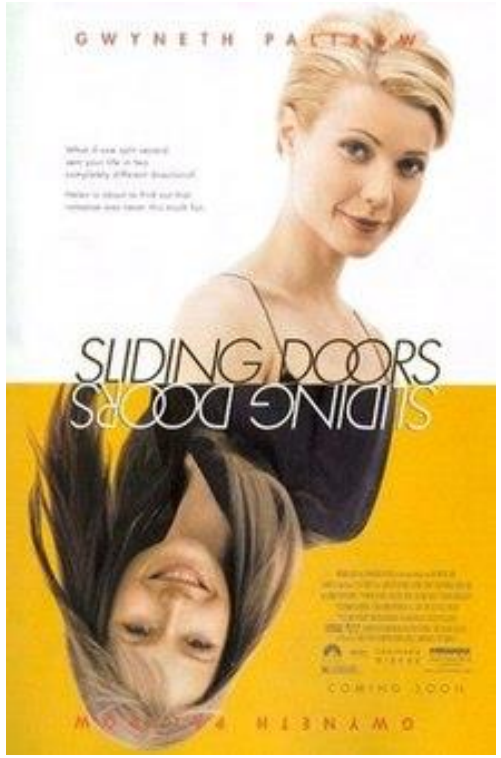
Pre-treatment covariates

- Observed outcome: $Y_i = Y_i(W_i)$. Only one of the potential outcomes is observed

The potential outcomes are natural concepts

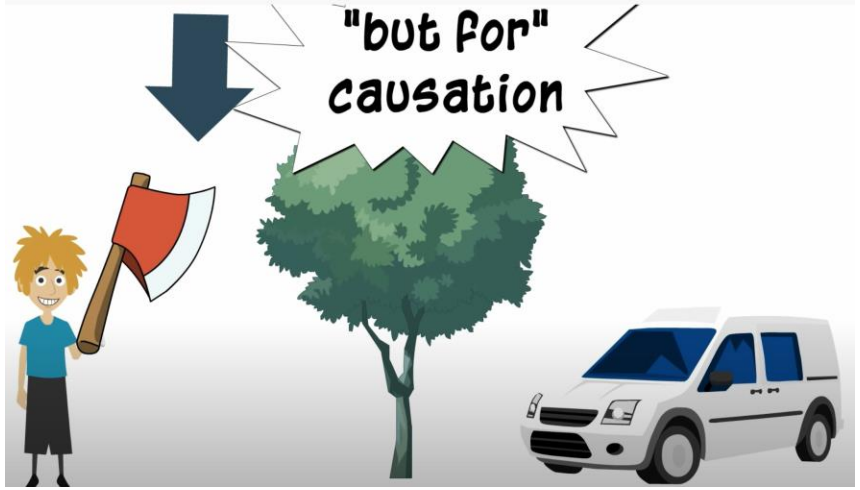


- *A Christmas Carol* (1843): a fiction by Charles Dickens
  - $Y(0)$ potential future if Scrooge continues his miserly ways
  - $Y(1)$ potential future with changed Scrooge



- *Sliding Doors* (1998): The film alternates between two storylines, showing two paths the central character's life could take depending on whether she catches a train.

The potential outcomes are natural concepts



Watch the YouTube video:
https://www.youtube.com/watch?v=P9TShT3xn4Q

But-for test in legal contexts

- The but-for test is a test commonly used in both tort law and criminal law to determine actual causation. The test asks, "but for the existence of X, would Y have occurred?"
- The Federal Judicial Center's "Reference Manual on Scientific Evidence" (1994, Chapter 3, p. 481) states:

In most cases, the analysis considers the difference ***between the plaintiff's economic position if the harmful event had not occurred and the plaintiff's actual economic position****. The damages study restates the plaintiff's position "but for" the harmful event; this part is often called but-for analysis.* ***Damages are the difference between the but-for value and the actual value.***

Definition of causal effect

- Causal effect for unit i: $\tau_i = Y_i(1) - Y_i(0)$
- We can also use other quantities to define the causal effect:
  ex. Log fold change $\log(Y_i(1)/Y_i(0))$, percentage change $\frac{Y_i(1)-Y_i(0)}{Y_i(0)} \times 100\%$.
- Any causal quantity is a function of potential outcomes
- Extension to Non-binary treatment:
  - Categorical: $Y_i(0), Y_i(1), \cdots, Y_i(K-1)$
  - Continuous: $Y_i(t)$ for any $t \in \mathbb{R}$

Fundamental problem of causal inference

Voters	Contact	Turnout		Age	Gender
i	W_i	$Y_i(1)$	$Y_i(0)$	X_{i1}	X_{i2}
1	1	1	?	20	M
2	0	?	0	55	F
3	0	?	1	40	F
⋮	⋮	⋮	⋮	⋮	⋮
n	1	0	?	62	M

- Causal inference is a missing data problem: only one potential outcome is observed for each unit
  - potential outcomes are thought to be fixed for each unit
  - potential outcomes do have a distribution across units
  - treatment variable determines which potential outcome is observed
  - observed outcomes are random for each unit because the treatment is random

- Summarize: where does the randomness in the data come from?

Underlying assumptions of the potential outcome framework: SUTVA

- The above notations implies three assumptions
  1. Causal ordering: Y_i cannot causally affect W_i
     - Potential violations: feedback effects (ex. Dosage adjustment)
     - Health status affect exercising, symptoms driving treatment assignment
  - No interference between units : each unit's potential outcomes remain the same no matter what treatments the other units receive.

    $$Y_i(W_1, \cdots, W_n) = Y_i(W_i)$$

    Example: if we have no contact with each other, whether you take an aspirin has no effect on the status of my headache
    - Counter-examples:
      - vaccine effect, effect of job training programs (peer effects)
      - **spillover:** an economic event in one context that occurs because of something else in a seemingly unrelated context. For example, if consumer spending in the United States declines, it has spillover effects on the economies that depend on the U.S. as their largest export market.

Underlying assumptions of the potential outcome framework: SUTVA

3. No different forms or versions of each treatment level (consistency)
   - By setting $W_i = 1$ for taking the aspirin, we assume that there is only one version of the tablets
   - If there are multiple versions (say different dosages), we need to redefine as sperate levels of treatments
   - Violated also if the effect defer depending on the method of administering the treatment. Assigned to receive a treatment v.s. choose to take the treatment

- **SUTVA**: Stable Unit Treatment Value Assumption 2 + 3
  - SUTVA can hold even if each unit has a different version of treatment (e.g. each person's surgeon can differ)

Causal effects of immutable Characteristics

- Immutable characteristics: sex, race, age, etc.
- Can immutable characteristics have meaningful causal effects?

- Examples:
  - **Race**: Study race in the labor market by sending fictitious resumes to help-wanted ads in Boston and Chicago newspapers. To manipulate perceived race, resumes are randomly assigned African-American- or White-sounding names. Researchers find that white names receive 50 percent more callbacks for interviews.
    [Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review*, 2004]

  - **Gender:** Since the mid-1990's, one third of Village Council head positions in India have been randomly reserved for a woman: In these councils only women could be elected to the position of head. Researchers find that leaders invest more in infrastructure that is directly relevant to the needs of their own genders.
    [Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica,* 2004]

Causal effects of immutable Characteristics

Extremely difficult with observational data

- Asian American Discrimination in Harvard Admissions
  - The organization Students for Fair Admissions and other plaintiffs filed a lawsuit against Harvard College in 2014, claiming that the college discriminates against Asian American applicants in its undergraduate admissions process.

  - Based on the admission data for Classes of 2014-2019, Economist Peter Arcidiacono from Duke developed a model to estimate the causal influence of Asian American status and concluded that "typical Asian American applicants would see their average admit rate rise by 19%, or approximately 1 percentage point, if they were treated as white applicants."

  - Using the same dataset, Economist David Card from UC Berkeley argued that Arcidiacono's models place too much emphasis on academic factors as predictors of admissions outcomes. By considering contextual factors including "high school, community and family background." in his model, Card argued that the effect of considering racial and ethnic factors doesn't result in a bias towards Asian-American students as Arcidiacono found.

Causal estimand

- Unit causal effects are difficult (impossible in most cases) to estimate
- Learning about causal effect typically requires multiple unit

- We can average the unit causal effects over n units
  - Average treatment effect : $\mathrm{SATE} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$
  - Average treatment effect for the treated: $\mathrm{SATT} = \frac{1}{n_1}\sum_{i=1}^{n} W_i\{Y_i(1) - Y_i(0)\}$ where $n_1 = \sum_{i=1}^{n} W_i$

- If we treat the units as sampled from a population
  - Population average treatment effect: $\mathrm{PATE} = \mathbb{E}(Y_i(1) - Y_i(0))$
  - Average treatment effect for the treated: $\mathrm{PATT} = \mathbb{E}(Y_i(1) - Y_i(0) \mid W_i = 1)$

- Estimand: any value that can be calculated from Science
  we want to define an estimand that is free from any model assumptions

Causal estimand

Other causal quantities of interest

- Heterogenous effects
  - Conditional average treatment effect (CATE)

$$\tau(\boldsymbol{x}) = \mathbb{E}(Y_i(1) - Y_i(0) \mid \boldsymbol{X}_i = \boldsymbol{x})$$

    - Applications in precision medicine and microtargeting
  - Quantile treatment effects
    - $Q_{Y(1)-Y(0)}(\alpha)$ where $Q_Z(\alpha)$ is the αth quantile of the random variable Z, e.x. $\alpha = 0.5$ corresponds to the median
    - (Easier to estimate) $Q_{Y(1)}(\alpha) - Q_{Y(0)}(\alpha)$

- Non-additive effect
  - Odds ratio for binary outcome

$$\frac{P(Y_i(1) = 1)/P(Y_i(1) = 0)}{P(Y_i(0) = 1)/P(Y_i(0) = 0)}$$

How to estimate the causal estimand?

How do we estimate the causal estimand from observed data?

Voters	Contact	Turnout		Age	Gender
i	T_i	$Y_i(1)$	$Y_i(0)$	X_{i1}	X_{i2}
1	1	1	?	20	M
2	0	?	0	55	F
3	0	?	1	40	F
⋮	⋮	⋮	⋮	⋮	⋮
n	1	0	?	62	M

- Causal inference is a missing data problem: only one potential outcome is observed for each unit
- Understanding the assignment mechanism is important
- Assignment mechanism: the process governing which units receive active treatment and which receive control

How to estimate the causal estimand?

- We can not simply compare between observed outcomes across units

Unit	Unknown			Known	
	Potential Outcomes		Causal Effect	Actual	Observed
	Y(Aspirin)	Y(No Aspirin)		Treatment W_i	Outcome Y_i^{obs}
You	No Headache	Headache	Improvement due to Aspirin	Aspirin	No Headache
I	No Headache	No Headache	None	No Aspirin	No Headache

- We should also not simply compare a before-after effect of the same individual (not the same unit)

Perfect doctor example

- Perfect doctor chooses the better treatment for each patient, i.e. the treatment under which the patient will live longer

The underlying fact: 0 (drug) v.s. 1(surgery)

Unit	Potential Outcomes		Causal Effect
	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
Patient #1	1	7	6
Patient #2	6	5	−1
Patient #3	1	5	4
Patient #4	8	7	−1
Average	4	6	2

Average causal effect: 2

The observed data with perfect doctors

Unit i	Treatment W_i	Observed Outcome Y_i^{obs}
Patient #1	1	7
Patient #2	0	6
Patient #3	1	5
Patient #4	0	8

Observed mean difference between two groups: -1

- What is wrong?

Perfect doctor example

Unit	W_i	$Y_i(0)$	$Y_i(1)$
1	1	?	7
2	0	6	?
3	1	?	5
4	0	8	?

Unit	W_i	$Y_i(0)$	$Y_i(1)$
1	1	7	7
2	0	6	6
3	1	7	5
4	0	8	6

- In the previous analysis, we implicitly impute the missing values by the group mean
- This is NOT reasonable for perfect doctors. We at least has the constraint that the missing potential outcomes (counterfactuals) should be no larger than the observed outcome

- More explicit answers relies on making assumptions

Experimental and observational studies

Two common types of studies

- **Randomized experiment:** the assignment probability $P(W|\boldsymbol{X})$ is known
  - Assignment can be completely random or depend on pre-treatment covariates of the individuals
  - Laboratory experiments, survey experiments, field experiments

- **Observational studies:** exact assignment probabilities may be unknown to the researcher
  - The researcher still has substantial information about the assignment mechanism
    For instance, the researchers knows what $\boldsymbol{X}$ includes. In medical decisions in some situations are solely based on patients' medical records.
  - No intervention is actually performed

- Tradeoff between interval and external validity

A brief history of the potential outcome approach

- Neyman (1923)
  - Define estimands in randomized experiment as functions of potential outcomes
  - While Neyman and others for half century restricted use of potential outcomes to randomized experiments

- Fisher (1925)
  - Proposing the necessity of physical randomization for credibly assessing causal effect

- Rubin (1974, 1975, 1977, 1978)
  - Define causal estimates with potential outcomes in all situations, not just randomized experiments
  - Discuss the assignment mechanism more extensively

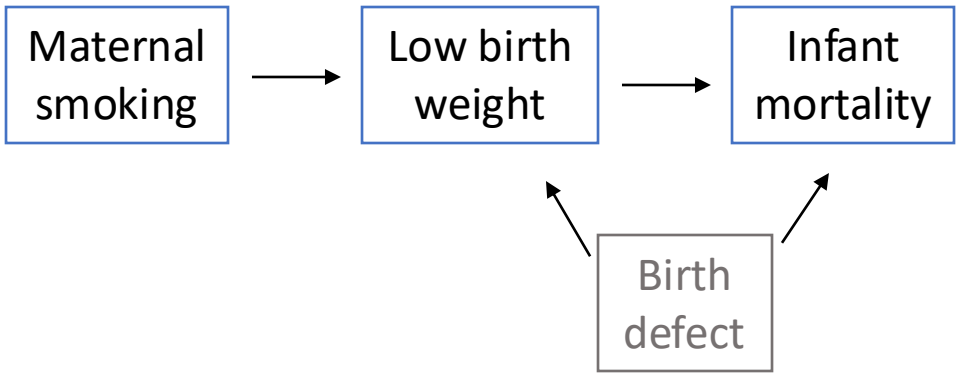
Causal directed acyclic graphs

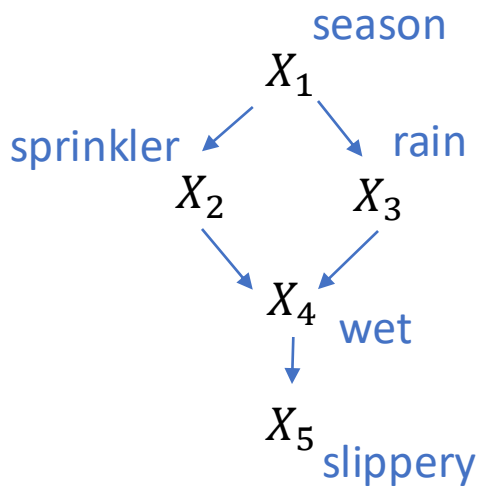
- This course focuses on using the potential outcome language, but there is another language to quantitatively describe causal inference
- Causal DAG: a directed graph with no cycles and the arrows has a causal meaning
- Nodes are random variables (may not be observed)
- PA_j: parents of X_j (nodes that have direct arrows to X_j)

$PA \rightarrow Y$

parent descendant

Examples:

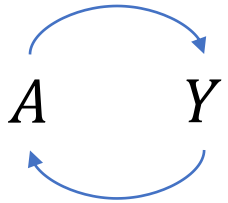




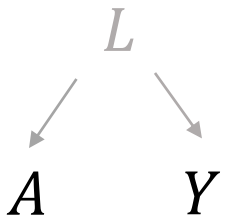
What are implicitly assumed in a DAG?

- No directed cycle:

  A can not cause itself

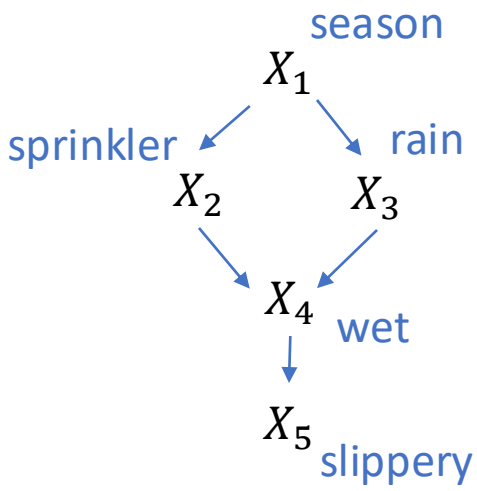


- The common causes of any pair of variables in the graph must be in the graph (either observed or unobserved)



- SUTVA: one version of treatment and no interference
- Intervention can be done on any node that has an arrow out

Connection with the potential outcome framework



For each X_j, assume existence of

- random error $E_{X_j} \perp PA_j$: include other causal factors that are not confounding factors
- A deterministic unknown function f_j for each node j
- Structural equation: function to represent causal relationship from ALL its direct parents to a node

$$X_4 = f_4(X_2, X_3, E_{X_4})$$

- Potential outcome: fix a particular value of all direct parents

$$X_j(pa_j) = f_j(pa_j, E_{X_j})$$

Potential outcomes for joint intervention on parents PA_j

- Example:

$$X_4(x_2, x_3) = f_4(x_2, x_3, E_{X_4})$$

$$X_4(x_2) = f_4(x_2, X_3(x_2), E_{X_4}) = f_4(x_2, X_3, E_{X_4})$$

Comparison between the two languages

- One reference:
  Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature*, *58*(4), 1129-79.

- As a graphical approach, DAG is superior in illustrating the causal relationships in a complex model and in clarifying some key assumptions

- With DAG, we can perform causal network discovery under additional assumptions

- DAG has more difficulties to capture and represent individual level heterogeneity

- Formal identification assumption for the causal estimand can be clearer using the potential outcome language

- In DAG, all variables are doable and the literature is silent about experiments

Reference papers to read

Causal effects of immutable Characteristics:

- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review, 94*(4), 991-1013. https://www.aeaweb.org/articles?id=10.1257/0002828042002561
- Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica, 72*(5), 1409-1443. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2004.00539.x?casa_token=1SVxHRxC5hQAAAAA:fz62pJ2QMeg41laiRMlERS5K1psbJVovA3-fgD5OTPjb2C-UkLuiQvv0CV2HSFq9z7r3WxxmH_p7NqE
- Arcidiacono, P., Kinsler, J., & Ransom, T. (2022). Asian American discrimination in Harvard admissions. *European Economic Review*, 104079. https://www.iza.org/publications/dp/13172/asian-american-discrimination-in-harvard-admissions

Lord's paradox:

- Holland, P. W., & Rubin, D. B. (1983). On Lord's paradox. *Principals of modern psychological measurement*, 3-25. https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2333-8504.1982.tb01321.x

Comparison between DAG and potential outcome framework:

- Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature*, *58*(4), 1129-79. https://www.aeaweb.org/articles?id=10.1257/jel.20191597