Causality in the Social Sciences II

Renjie Yang

COMPHI LAB for Data Science

June 2020

Causality in the Social Sciences

- Science seems capable of understanding the hidden causes of nature.
  - Newton: fundamental causes of real motion
  - Higgs field is responsible for mass
  - DNA molecules replicate to preserve traits
- Our social systems seem to be at once the products of free human action and causal mechanisms that run behind our backs.
  - I am free to act as I choose.
  - Companies can't deliver their goods: social reorganization would cause a different result.

Outline

Causal Claims in Science Reporting

- https://www.independent.co.uk/news/science/diet-red-meat-health-impact-consequences-a7518146.html
- https://www.theguardian.com/technology/2016/aug/08/positive-link-between-video-games-and-academic-perform
- https://www.mirror.co.uk/science/prostate-cancer-risk-soars-quarter-9260548

Causal Claims in Science Reporting

- https://behavioralscientist.org/why-triggering-emotions-wont-lead-to-lasting-behavior-
- https://behavioralscientist.org/venture-capitalists-are-using-the-wrong-tools-to-impro
- https://behavioralscientist.org/teachers-like-you-like-you-how-finding-common-ground-c

Confusions Over Causation

- Spurious correlation.http://tylervigen.com/view_correlation?id=1703
- Anecdotes. Bill Smith lived to be 105 years old. He said the secret to his longevity was eating one turnip a day.
- Correlation vs. causation
- Reverse Causality. Does green space in urban environments cause people to exercise more?

The Science of Causal Inference

The field of causal inference or causal modeling attempts to clarify the above confusions by proposing:

- formal definitions of causal effects
- assumptions necessary to identify causal effects from data
- rules about what variables need to be controlled for
- sensitivity analyses to determine the impact of violations of assumptions on conclusions

The Literature of Causal Inference

- Differs in distinct academic disciplines
- Statisticians started working on causal modeling since at least 1920s (Wright 1921; Neyman 1923)
- It became its own area of statistical research since about the 1970s
- Rubin's potential outcome framework(Rubin 1974); Causal graph framework (Greenland and Robins 1986; Pearl 2000)

Treatments and Outcomes

Suppose we are interested in the causal effect of some treatment A on some outcome Y.

Treatment Examples:

- $A = 1$ if receive influenza vaccine; $A = 0$ otherwise
- $A = 1$ if assigned to a large class; $A = 0$ if assigned to a small one

Outcome Examples:

- $Y = 1$ if develop cardiovascular disease within 2 years; $Y = 0$ otherwise
- $Y =$ student performance

Potential Outcomes

- Notation: Y^a is the outcome that would be observed if treatment was set to $A = a$

Potential Outcomes Example

- Suppose treatment is influenza vaccine and the outcome is the time until the individual gets the flu.
- Y^1: time until the individual would get the flu if they received the flue vaccine
- Y^0: time until the individual would get the flu if they did not receive the flue vaccine

Counterfactuals

Did influenza vaccine prevent me from getting the flu?
What actually happened:

- I got the vaccine and did not get sick
- My actual exposure was $A = 1$
- My observed outcome was $Y = Y^1$

What would have happened (but not happened in reality):

- Had I not gotten the vaccine, would I have gotten sick?
- My counterfactual exposure is $A = 0$
- My counterfactual outcome is Y^0

Potential Outcomes and Counterfactuals

- Before the treatment decision is made, any outcome is a potential outcome: Y^0 and Y^1
- After the study, there is an observed outcome, $Y = Y^A$, and counterfactual outcomes Y^{1-A}
- Counterfactual outcomes Y^0 and Y^1 are typically assumed to be the same as potential outcomes Y^0 and Y^1.

Qualifications of Treatments

- There might be many potential ways to intervene on a variable. These different ways might also be associated with different outcomes. Example: losing weight.
- Immutable variables such as age or gender do not fit cleanly in the potential outcome framework.
- We would focus on treatments that could be thought of as interventions, which are those that we can imagine being randomized or manipulated in a hypothetical trial.

Causal Effects

- A had a causal effect on Y if Y^1 differs from Y^0.
- Example: "I took ibuprofen and my headache is gone, therefore the medicine worked."
- The statement is equivalent to $Y^1 = 1$
- What would have happened had you not taken ibuprofen?
- Causal effect inquires if $Y^1 \neq Y^0$ and how to measure its strength.

Causal Inference and Statistical Inference

- Causal inference is substantially more difficult than statistical inference
- The fundamental problem of causal inference is that we can only observe one potential outcome for each person.
- Hopeless: what would have happened to me had I not taken ibuprofen?
- Possible: What would the rate of headache remission be if everyone took ibuprofen when they had a headache versus if no one did?
- With certain assumptions, we can estimate population level causal effects.

Average Causal Effect

- $E(Y^1 - Y^0)$
- Average value of Y if everyone was treated with $A = 1$ minus average value of Y if everyone was treated with $A = 0$
- Picture Illustration

Average Causal Effect Example

- Example: Regional $(A = 1)$ vs. general $(A = 0)$ anesthesia for hip fracture surgery on risk of major pulmonary complications
- $E(Y^1 - Y^0) = -0.1$
- Probability of major pulmonary complications is lower by 0.1 if given regional anesthesia compared with general anesthesia
- If 1000 people were going to have hip fracture surgery, we would expect 100 fewer people to have pulmonary complications under regional anesthesia compared with general anesthesia

Conditioning vs. Manipulating

- In general, $E(Y^1 - Y^0) \neq E(Y|A=1) - E(Y|A=0)$
- $E(Y|A=1)$: mean of Y among people with $A=1$
- $E(Y^1)$: mean of Y if the whole population was treated with $A=1$
- $E(Y|A=1) - E(Y|A=0)$ is generally not a causal effect, because it is comparing two different populations of people.
- $E(Y^1 - Y^0)$ is a causal effect, because it is comparing what would happen if the same people were treated with $A=1$ vs. if the same people were treated with $A=0$

The Fundamental Problem of Causal Inference

- We only observe one treatment and one outcome for each person. How do we use observed data to connect observed outcomes and potential outcomes?

Assumptions for Causal Inference

- Stable Unit Treatment Value Assumption (SUTVA)
- Consistency
- Ignorability
- Positivity

Stable Unit Treatment Value Assumption (SUTVA)

- Units do not interfere with each other
- Treatment assignment of one unit does not affect the outcome of another unit
- There is only one version of treatment

Therefore, SUTVA allows us to write potential outcome for the ith person in terms of only that person's treatment.

Consistency Assumption

- The potential outcome under treatment $A = a$, Y^a, is equal to the observed outcome if the actual treatment received is $A = a$

Ignorability Assumption

- Given pre-treatment covariates X, treatment assignment is independent from the potential outcomes.
- $Y^0, Y^1 \perp\!\!\!\perp A|X$
- Among people with the same values of X, we can think of treatment A as being randomly assigned.
- Example: X is the variable “age”

Positivity Assumption

- For every set of values for X, treatment assignment was not deterministic
- $P(A = a|X = x) > 0$ for all a and x
- If for some values of X treatment was deterministic, then we would have no observed values of Y for one of the treatment groups for those values of X.

Observed Data and Potential Outcomes

$$
\begin{aligned}
E(Y^a|X=x) &= E(Y^a|A=a, X=x), ignorability \\
&= E(Y|A=a, X=x), consistency
\end{aligned}
$$

$$
E(Y^a) = \Sigma_x E(Y|A=a, X=x) P(X=x)
$$

Example

Compute the causal effect of saxagliptin versus sitagliptin. The outcome is Major Adverse Cardiac Event (MACE). Saxagliptin users were more likely to have past use of other antidiabetic drug.

	MACE =yes	MACE= no	Total
Saxa=yes	50	950	1000
Saxa=no	200	3800	4000
Total	250	4750	5000

	MACE =yes	MACE= no	Total
Saxa=yes	300	2700	3000
Saxa=no	300	2700	3000
Total	600	5400	6000

Confounding

- Confounders are often defined as variables that affect treatment and affect the outcome.
- If I assign treatment based on a coin flip, then that affects treatment but should not affect outcome.
- If older are at higher risk of cardiovascular disease (the outcome) and are more likely to receive statins (the treatment), then age is a confounder.
- Causal inference need to identify a set of variables X that will make the ignorability assumption hold, and statistically estimate causal effects by controlling these variables.

The Graphical Model Framework

- It would be more intuitive if we could use graphs to identify a set of variables X that will make the ignorability assumption $Y^0, Y^1 \perp\!\!\!\perp A | X$ hold.
- In order to achieve this goal, we need to connect probability with graphs.
- The graphical modeling framework is well suited for this task.
- In general, the graphical framework and the potential outcome framework lead to the same formulas for causal effects. They are two different languages for the same thing.

Probability and Directed Acyclic Graphs (DAGs)

Definition 1 (Factorization)

Let G be a DAG with vertices $V = (X_1, ..., X_d)$. If P is a distribution for V with probability function $p(x)$, we say that P is Markov to G, or that G represents P, if

$$p(x) = \prod_{j=1}^{d} p(x_j | \pi_{x_j})$$

where π_{x_j} is the set of parent nodes of X_j. The set of distributions represented by G is denoted by $M(G)$.

Probability and Directed Acyclic Graphs (DAGs)

Theorem 2 (Local Markov Property)

For a graph $G = (V, E)$, a distribution $P \in M(G)$ if and only if the following Markov condition holds: for every variable W,

$$W \perp\!\!\!\perp \widetilde{W} | \pi_W$$

where $\widetilde{W}$ denotes all the other variables except the parents and descendants of W.

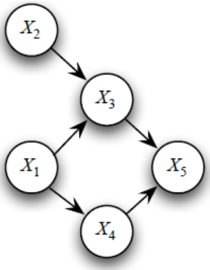
Probability and Directed Acyclic Graphs (DAGs)

Theorem 3 (Global Markov Property)

(Verma and Pearl (1988)) If sets A and B are d-separated by C in a DAG G, then A is independent of B conditional on C in every distribution compatible with G. Conversely, if A and B are not d-separated by C in a DAG G, then A and B are dependent conditional on C in at least one distribution compatible with G.

Why Global

- $X_1 \perp\!\!\!\perp X_2$,
- $X_2 \perp\!\!\!\perp \{X_1, X_4\}$, $X_3 \perp\!\!\!\perp X_4 | \{X_1, X_2\}$,
- $X_4 \perp\!\!\!\perp \{X_2, X_3\} | X_1$, $X_5 \perp\!\!\!\perp \{X_1, X_2\} | \{X_3, X_4\}$,
- $X_2 \perp\!\!\!\perp \{X_4, X_5\} | \{X_1, X_3\} \leftarrow$



Probability and Directed Acyclic Graphs (DAGs)

Definition 4 (d-separation)

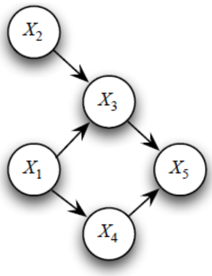
Let X and Y be distinct vertices and let W be a set of vertices not containing X or Y. Then X and Y are d-separated given W if there exists no undirected path U between X and Y such that

1. every collider on U has a descendant in W
2. no other vertex on U is in W.

If A,B, and W are distinct sets of vertices and A and B are not empty, then A and B are d-separated given W if for every $X \in A$ and $Y \in B$, X and Y are d-separated given W

Causal Interpretation

- $X_1 \sim P(X_1)$
- $X_2 \sim P(X_2)$
- $X_3|X_1, X_2 \sim P(X_3|X_1, X_2)$
- $X_4|X_1 \sim P(X_4|X_1)$
- $X_5|X_3, X_4 \sim P(X_5|X_3, X_4)$

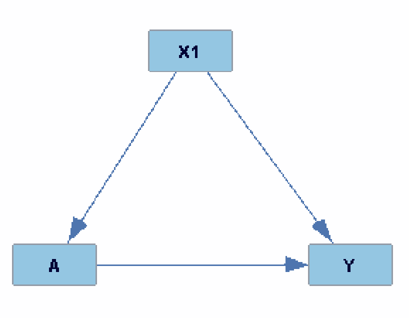


Outline

Confounding

- Confounders are often defined as variables that affect treatment and affect the outcome.
- If I assign treatment based on a coin flip, then that affects treatment but should not affect outcome.
- If older are at higher risk of cardiovascular disease (the outcome) and are more likely to receive statins (the treatment), then age is a confounder.
- Causal inference need to identify a set of variables X that will make the ignorability assumption hold, and statistically estimate causal effects by controlling these variables.

Matching



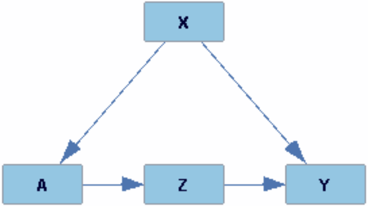
X is sufficient to control for confounding. Ignorability assumption holds: $Y^0, Y^1 \perp\!\!\!\perp A|X$

Matching

- Matching is a method that attempts to make an observational study more like a randomized trial.
- The idea is to match individuals in the treated group ($\mathsf{A} = 1$) to individuals in the control group ($\mathsf{A} = 0$) on the covariate X.

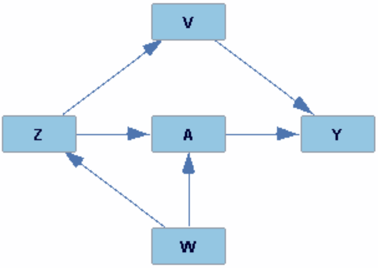
Frontdoor Paths

- A frontdoor path from A to Y is one that begins with an arrow emanating out of A.
- If we are interested in the causal effect of A on Y, we should not control for variables in the frontdoor paths, because they capture effects of treatment.



Backdoor Paths

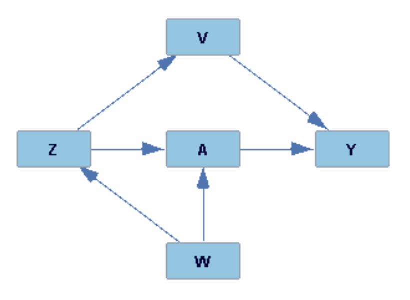
- A backdoor path from A to Y is one that travel though arrows going into A.
- Backdoor paths confound the relationship between A and Y
- To sufficiently control for confounding, must identify a set of variables that block all backdoor paths from treatment to outcome.



Backdoor Criterion

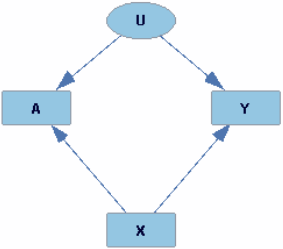
A set of variables X is sufficient to control for confounding if

- it blocks all backdoor paths from treatment to the outcome
- it does not include any descendants of treatment



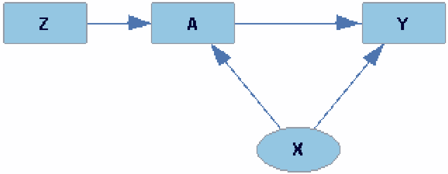
Unmeasured Confounding

- Unmeasured variables U affects both A and Y
- Ignorability assumption violated
- Estimation of causal effect would be biased



Instrumental Variables

- Instrumental variables (IV) is a causal inference method that does not rely on the ignorability assumption.
- Instrument variable Z affects the treatment but not the outcome.
- The purpose is to infer $E(Y^{A=1}) - E(Y^{A=0})$ from $E(Y^{Z=1}) - E(Y^{Z=0})$



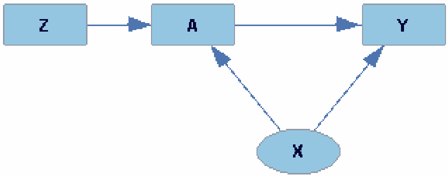
Instrumental Variables

Assumptions of instrumental variables (not enough for identification)

- it is associated with the treatment;
- exclusion restriction: it affects the outcome only through its effect on treatment;

The assumptions imply that

- $E(Y^{Z=1}) - E(Y^{Z=0}) = E(Y|Z=1) - E(Y|Z=0)$



Instrumental Variable Example

- A: smoking during pregnancy (yes/no)
- Y: birthweight
- X: mother's age, weight, parity, etc.
- Z: randomize to either receive encouragement to stop smoking ($Z = 1$) or receive usual care ($Z = 0$)

Sexton and Hebel (1984) "A clinical trial of change in maternal smoking and its effect on birth weight." *The Journal of the American Medical Association*

Complier Average Causal Effect (CACE)

$$
\begin{aligned}
E(Y|Z=1,A^0=0,A^1=1)-E(Y|Z=0A^0=0,A^1=1) \\
= E(Y^{Z=1}-Y^{Z=0}|compliers) \\
= E(Y^{A=1}-Y^{A=0}|compliers)
\end{aligned}
$$

A^0	A^1	Label
0	0	Never-takers
0	1	Compliers
1	0	Defiers
1	1	Always-takers

Observed Data

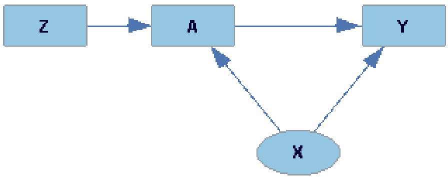
For each person we observe an A and a Z, not (A^0, A^1)

Z	A	A0	A1	Class
0	0	0	?	Never-takers or compliers
0	1	1	?	Always-takers or defiers
1	0	?	0	Never-takers or defiers
1	1	?	1	Always-takers or compliers

Instrumental Variables

Assumptions of instrumental variables (enough for identification)

- it is associated with the treatment;
- exclusion restriction: it affects the outcome only through its effect on treatment;
- monotonicity assumption: there are no defiers



Identification of Causal Effects

$$
\begin{aligned}
E(Y|Z=1) = E(Y|Z=1, always\ takers)P(always\ takers) \\
+E(Y|Z=1, never\ takers)P(never\ takers) \\
+E(Y|Z=1, compliers)P(compliers)
\end{aligned}
$$

$$
\begin{aligned}
E(Y|Z=0) = E(Y|Z=0, always\ takers)P(always\ takers) \\
+E(Y|Z=0, never\ takers)P(never\ takers) \\
+E(Y|Z=0, compliers)P(compliers)
\end{aligned}
$$

Identification of Causal Effects

$$
\begin{aligned}
E(Y|Z=1) - E(Y|Z=0) = E(Y|Z=1, compliers)P(compliers) \\
-E(Y|Z=0, compliers)P(compliers)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{E(Y|Z=1) - E(Y|Z=0)}{P(compliers)} \\
&= E(Y|Z=1, compliers) - E(Y|Z=0, compliers) \\
&= E(Y^{a=1}|compliers) - E(Y^{a=0}|compliers)
\end{aligned}
$$

Identification of Causal Effects

$$P(compliers) = E(A|A|Z = 1) - E(A|Z = 0)$$

$$CACE = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{E(A|Z = 1) - E(A|Z = 0)}$$

Two stage Least Square Estimator

2SLS estimator is a consistent estimator of CACE.

- State 1: $A_i = \alpha_0 + Z_i \alpha_1 + \epsilon_i$
- $\hat{A}_i$ is estimate of $E(A|Z)$
- Stage 2: $Y_i = \beta_0 + \hat{A}_i \beta_1 + \epsilon_i$
- The causal effect is represented by β_1

Outline

Case Study Methodology

- One of the limitations of Bayesian Networks is that they require a predetermined set of variables.
- Before the logic of inference can begin, the social scientist must already have determined what the possible causes might be.
- If we do not already know the causes, how do we know that we are measuring the right variables?
- Case study methods respond to this problem by taking an intensive look at a small number of events, social structures, or institutions.

Social Scientific Experimentation

- Experimentation is one of the iconic practices of science, but it has been largely shunned by the social sciences.
- Experimentation seems to be an exemplary way of identifying causal mechanisms.
- Is the lack of experimentation an indication of the methodological weakness of the social sciences relative to the natural sciences?
- Or is there something about the social sciences that makes experimentation difficult, or even inappropriate?

Roles of Scientific Experimentation

Standard accounts of the epistemology of experimentation identify two roles for experiments in scientific inquiry.

- Causal relations can only be identified through regularities. These are expressed by the generalizations or laws of a theory. Experiments can be used to test theory.
- The goal of an experiment is to isolate and manipulate a purported cause. A successful experiment thus identifies a causal relationship between two factors, X and Y, by showing that changes in X are correlated with changes in Y

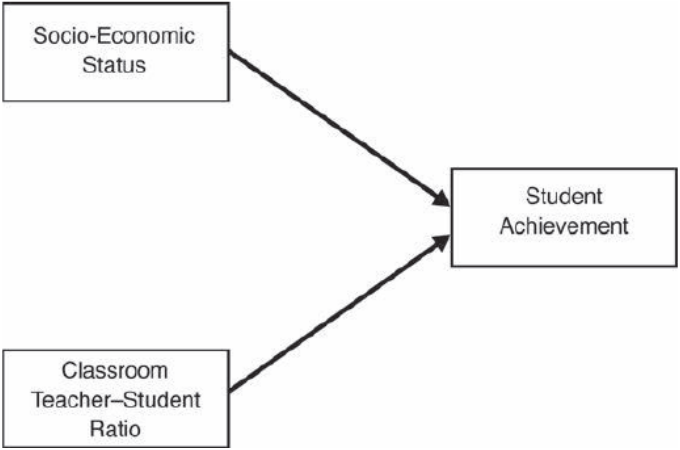
Social Scientific Experimentation

- John Stuart Mill famously argued against social scientific experimentation in his System of Logic (Mill 1987 [1872], 70–1).
- The root of the problem, he argued, was in the way that the causal factors combine and influence one another. The interactions are so complex that it is impossible to isolate a possible causal factor in the way demanded by experiment.
- Mill's concern applies to experiments conceived either as theory tests or as manipulation of causes.
- Small experiments would not do: by isolating the system one is unable to identify the real causes of social events.

Reliable Production of Phenomena

- Ian Hacking argued that the relationship between theory and experiment had been misunderstood by philosophers of science (Hacking 1983).
- Experiments rarely test hypotheses derived from theory, nor do they typically confirm the causal mechanisms predicted by theory.
- Experimentation is often dedicated to the production of "effects" or "phenomena," which are events that experimenters learn to reliably reproduce.
- Like the effects of physics, the phenomena discovered by social scientific experiments become the subject of social scientific theories.

The Coleman Report (1966)



Project STAR

- Project STAR (Student–Teacher Achievement Ratio), 1980s involved more than 6,000 school children in classes from kindergarten to third grade.
- It divided children into two groups: classes of 13 to 17, and classes of 22 to 25.
- In each participating school, the classrooms were assigned so that some children went to small classes and others went to large classes.
- The study found that smaller classes had a significant, positive effect on educational outcomes, and later phases of the study showed that it lasted beyond the early elementary school years.

Quasi-Experiment and RCT

- The interventionist conception of causality can be satisfied when there is no human intervention at all. The groups that are identical except one independent variable can grow naturally. Methods that try to discover causes in this way are known as “natural” or “quasi-” experiments.
- In an RCT (randomized controlled trial), the two groups—usually called “treatment” and “control” in this context are entirely determined by random selection.
- These methods are difficult in reality because of cases like sample bias.

Reference

- Mark Risjord (2014) *Philosophy of Social Science: A Contemporary Introduction*, Routledge.
- Larry Wasserman's lecture notes on Graphical Models.
- Jason Roy's slides on Causal Inference