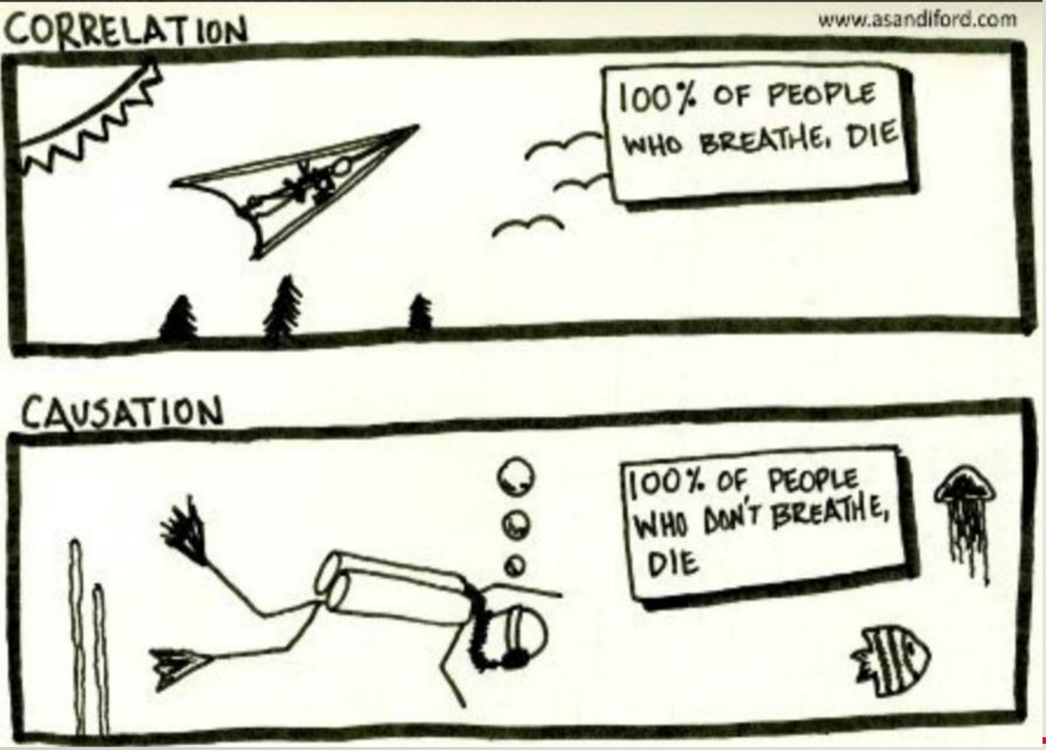


CAUSAL INFERENCE IN NATURAL LANGUAGE PROCESSING FOR ECONOMICS

WEEK 2: CAUSAL INFERENCE – TEXT AS TREATMENT

SUMMER SEMESTER 22

DR.HUYEN NGUYEN

HTTP://HUYENTTNGUYEN.COM

LOGISTICS

- Any questions with…
  - the assignment requirements & instructions?
  - Access to Slack? Open Olat?
  - Access on Github?
  - Group sign-ups? (groups of 2 to 3 students, in total 8 groups)

  https://docs.google.com/spreadsheets/d/1EEUJp1Srz05BwEUH2u3CbYwMykmqEUhZdn2SqrmeaRs/edit#gid=0

  Group sign-up Deadline: Tuesday 19.04 @ 23.59 → If not yet in group by then, I'll randomize you into groups.

AGENDA

- Causal Inference (CI) Framework
  - Refresher
  - Assumptions in detail
- CI with Text as Treatment
  - Framework
  - Challenges
  - Applications
- Coding exercise: CI A/B Test Trip Advisor Recommender System

CENTRAL POLICY QUESTION

What happens if someone receives a treatment/policy X is implemented vs. what would happen if they had not received that treatment/X is not implemented?

AVERAGE TREATMENT EFFECT (ATE)

- Based on the defined counterfactuals, an analyst must specify a quantity of interest that involves the distribution of counterfactuals.

$$ATE = E[Y(1)] - E[Y(0)]$$

- This is the average difference in counterfactual individuals who do not obtain university degrees, if we could intervene to change the obtain-university- degree "treatment", with which the individual is assigned to.

IDEAL EXPERIMENT

- If you had any resources you need at your disposal, what would you do to answer your question of interest?

- What would you need to do to get random assignment of treatment?

IDEAL EXPERIMENT

- 1)Natural experiment: Someone else/An institution/Nature does the experiment for you.
  - E.g: The impact of WWII and Nazi "purge" of scientists on knowledge production (Iaria, Schwarz and Waldinger QJE 2018)
- 2)Clever identification strategies (instrumental variables, regression discontinuity design, difference-in-difference, matching/high dimension controls)
  - E.g: The effect of liberalization on economic performance (Giavani & Tabellini JME 2005)
- For detailed, easy-to-understand readings on these techniques, I recommend the book of Angrist, Joshua and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics.

IDENTIFICATION VS. ESTIMATION VS. INFERENCE

Given the underlying data generating process (DGP), if I could take as large a sample as I want, would I eventually learn the causal effect of interest? = IDENTIFICATION question

→ The answer to question of causal interest is identified IF there exists an ideal experiment that can answer it.

(!) Causal questions for which there is no ideal experiment that can answer it are Fundamentally Unidentified Questions (in short, FUQ'ed questions…)

CAUSAL INFERENCE?

- Causal inference is the process of exploring how changes on variable T affect variable Y .
- where T = treatment and Y = outcome,
- ΔT (changes in T) = intervention.
- Causal Inference = the process of drawing a conclusion about <u>whether</u> and <u>how</u> Y changes when intervening on T.

ASSUMPTIONS OF CAUSAL INFERENCE

- Ignorability: $\big(Y(1), Y(0)\big) \perp T$

- Positivity: $0 < \Pr(T = 1|X = x) < 1$

- Consistency: $\forall i \; s.t \; T_i = t, Y_i(t) = Y_i$

ASSUMPTIONS FOR CAUSAL INFERENCE (1)

In short: How to make sure we can identify the ATE?

1) Ignorability: requires that treatment T assignment is independent of the realized counterfactual outcomes.

$$\big(Y(1), Y(0)\big) \perp T$$

(!) the most important and difficult to justify assumption.

ASSUMPTIONS FOR CAUSAL INFERENCE (1)

- Randomizing treatment assignment is one way to satisfy ignorability. Why?
- In expectation there are no systematic pretreatment differences between treated and untreated samples.

ASSUMPTIONS FOR CAUSAL INFERENCE (1)

- Randomized assignment may NOT always be feasible.
- In this case, we may need to rely on **conditional ignorability:**

$$\left(\left(Y(1), Y(0)\right) \perp\!\!\!\perp T\right) | X$$

- where X is a set of variables such that treatment T assignment and the potential outcomes is unconfounded within levels of X.
- Caveat: Conditional ignorability requires that there are no unobserved confounders. → strong assumption that analyst must carefully assess.

ASSUMPTIONS FOR CAUSAL INFERENCE (2)

2) Positivity: requires that the probability of receiving the treatment is bounded between 0 and 1.

$$0 < \Pr(T = 1|X = x) < 1$$

Under conditional ignorability, this must hold for all x such that

$$\Pr(X = x) \neq 0$$

ASSUMPTIONS FOR CAUSAL INFERENCE (2)

- Intuitively, this requires that in order to learn about causal effects, it needs to be possible to observe units under <u>both</u> types of treatment statuses.
- It also implies that the treatment status cannot be perfectly predicted given the variables X.
- Randomized treatment assignment also ensures positivity is satisfied.

ASSUMPTIONS FOR CAUSAL INFERENCE (3)

3) Consistency: requires that the observed outcome at a given treatment status for a given unit, is the same as we would observe if that unit was assigned to the treatment.

$$\forall i \; s.t \; T_i = t, Y_i(t) = Y_i$$

Consistency relates counterfactual outcomes to observed ones and captures two main assertions:

First, no interference: the outcome for a sample i is affected only by sample i's treatment status, and NOT the treatment status of other samples.

Second, there is ONLY ONE version of treatment.

ASSUMPTIONS FOR CAUSAL INFERENCE (3)

3) Consistency: requires that the observed outcome at a given treatment status for a given unit, is the same as we would observe if that unit was assigned to the treatment.

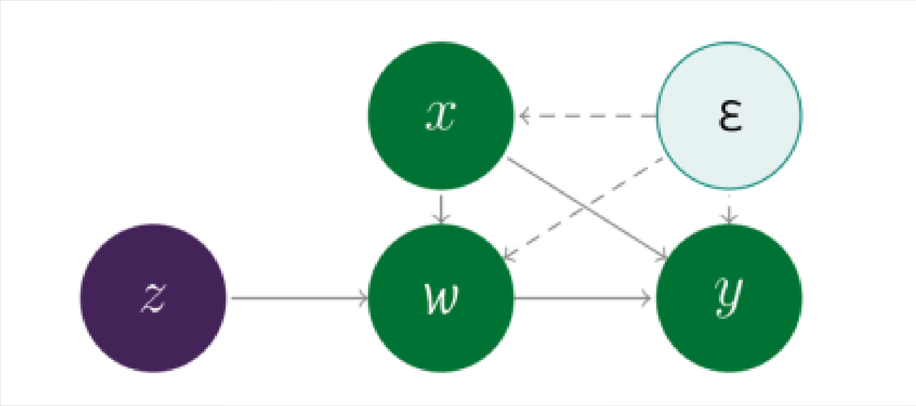
$$\forall i \ s.t \ T_i = t, Y_i(t) = Y_i$$

Consistency relates counterfactual outcomes to observed ones and captures two main assertions:

→ First, no interference: the outcome for a sample i is affected only by sample i's treatment status, and NOT the treatment status of other samples.

→ Second, there is ONLY ONE version of treatment.

THE EMPIRICAL PROBLEM



y = outcome
w = (text) treatments
x = observable covariates
ε = confounders
z = instruments

(!) spurious correlation by confounders i.e. variables that are correlated with both the treatment and outcome.

What are the likely (text) confounders ε?

TEXT AS DATA IN THE CI FRAMEWORK

Discovery is central to text/audio/image-based causal inferences because such complex, high-dimensional data types ALWAYS need simplification before it can be used meaningfully.

E.g: Take a collection of e-mails, classify into 'spam' and 'not spam.'

Let $\boldsymbol{g}$ be the function which maps the documents into our measure of interest g.

→ $\boldsymbol{g}$ = a crucial codebook that tells us how to compress our documents into categories, topics, or dimensions.

TEXT AS DATA IN THE CI FRAMEWORK

- The need to discover and iteratively define measures and concepts from data is a fundamental component of social science research.
- (!) BUT the iterative discovery process poses problems for causal inference. WHY?
- We may not know $\boldsymbol{g}$ in advance of conducting our experiment.
- AND consequently, we may not know our outcome or treatment.

→ Identification problem!

ANALYST INDUCED SUTVA VIOLATION (AISV)

- Identification problem occurs because the particular $\boldsymbol{g}$ we obtain will often depend upon the treatments and responses.
- And using this information can create a dependence across units (!)
- Most causal inference approaches assume that each unit's response depends on ONLY its treatment status, NOT on any other unit's treatment.
- This is one component of the Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1980).

ANALYST INDUCED SUTVA VIOLATION (AISV)

- But, when using the same documents for discovering $\mathbf{g}$ and estimating effects , the researcher can induce a SUTVA violation where none had previously existed.
- WHY?
- The $\mathbf{g}$ that we discover depends on the particular set of treatment assignments and responses in our sample → changing other units' treatment status = change the $\mathbf{g}$ discovered and, as a result, the measured response or intervention for a particular unit
- WHY BOTHER? AISV makes it impossible to evaluate properties of our estimator such as variance, bias or consistency without further assumptions.

OVERFITTING

- By using the same documents to discover and estimate effects, noise can be mistaken for a robust causal effect.
- Searching over $\boldsymbol{g}$ = recoding variables in an experiment to search for significance.
- This idea of overfitting also formalizes the intuition that some analysts have that
- latent-variable models are ‘baking in’ an effect

TEXT AS TREATMENT - LAB/FIELD EXPERIMENTS WITH TEXTS

Gold standard to obtain causal estimates. For instance:

- Subjects fill out open-ended survey responses before and after the experiment
- Subjects can talk to each other in a chatroom
- Subjects view randomly assigned text treatments

Applications? (1) how do different information provided to students by alumni affect career choices (Gallen & Wasserman 2021)

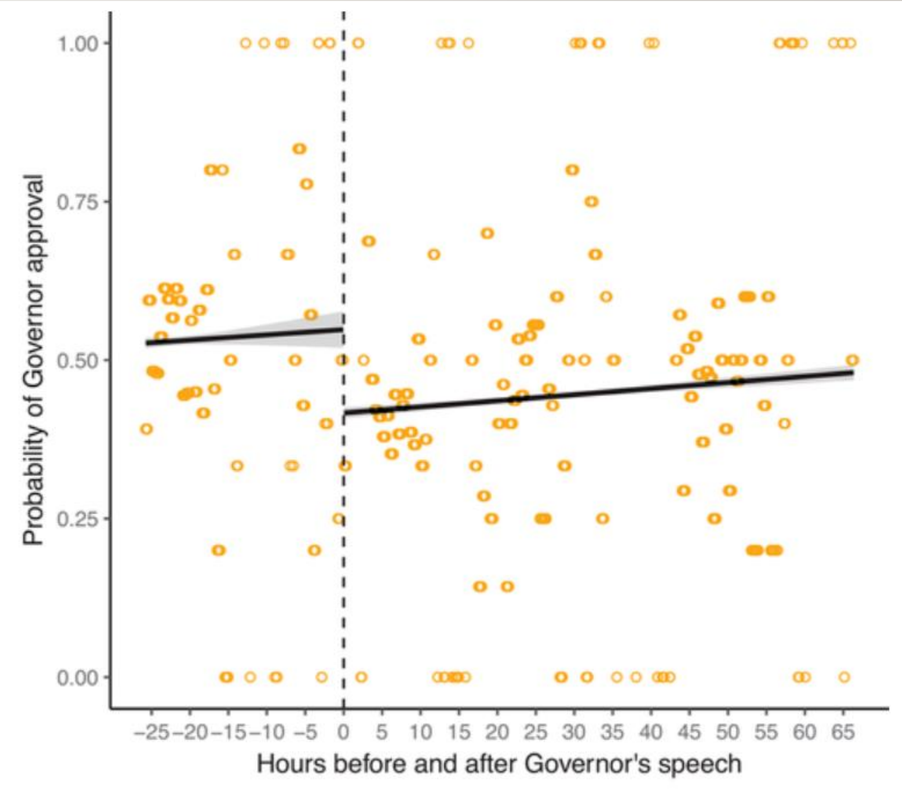
(2) how appealing to civic duty via mails impacts on voter turnout in Michigan 2006 primary election (Gerber et al 2008)

DIFFERENCE-IN-DIFFERENCE WITH TEXT

- Estimate changes in an outcome due to changes in text.
- (!) a lot of strong assumptions for this, probably infeasible in reality.
- -> More realistic approach? measure changes in a text-based metric due to a treatment
- Include fixed effects and trends for the individuals.
- Try to illustrate effect with **event-study graph**
  - Ash, Morelli, and Van Weelden (2017): higher news transparency causes senate members up for election to spend more time on divisive issues

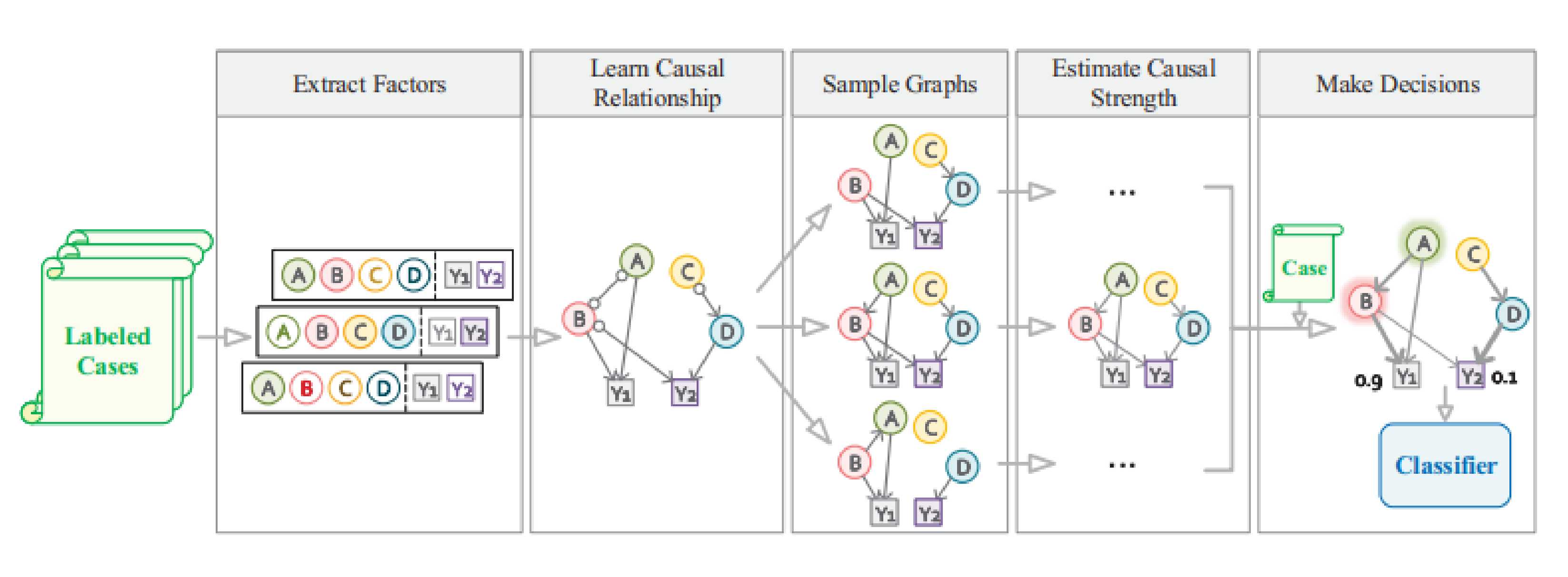
REGRESSION DISCONTINUITY WITH TEXT

- Local impact of a threshold treatment on a text-based metric.
  - E.g: Electoral RD effect on type of speech used by Congressmen
- Local impact of a threshold text treatment.
  - E.g: Electoral RD effect of governor's Trump-delegated speech on Governor's approval probability during Covid-19 pandemic (Shino & Binder 2020)



CI MODEL WITH TEXT: AN EXAMPLE

- A legal AI system assists judges to deal with complicated cases that involve multiple parties and complex events, CI could help to figure out the exact distinguishable elements that are crucial for fair and impartial judgements (Liu et al., 2021, ACL)



CI WITH TEXT AS TREATMENTS CHALLENGES

- Goal? The causal relationship that (specific aspects of the) text has on downstream decisions, behaviors, and outcomes.

- Challenges?
- 1) Ignorability is typically violated.
- Why? People often choose the text they read for reasons related to outcome of interests. (e.g.: political affiliation, mood, education level, gender)

CI WITH TEXT AS TREATMENTS CHALLENGES

Challenges?

- 2) Positivity is often violated.
- Why? Suppose we want to know if politeness causally decreases e-mail response times. In other words, we want to causally know if unpoliteness increases e-mail response times. Even if the text contains all confounders (e.g.: topic, tone, writing styles), a polite email is unlikely to contain a certain writing style (e.g., profane).

EXERCISE TIME ☺

- Open Stata (RDD refresher).
- Open your Google Colab.
- Go to Github's course channel →Exercise folder to access these files.