Segment 1: Fundamentals of Causal Inference

Section 03: The Potential Outcomes Paradigm

What is Causal Inference Methodology

Causal inference methodology (in this class) is a framework for:

1. Carefully defining a causal effect of interest (e.g., of a decision, action, treatment, exposure, policy, etc.)
2. Determining whether/how such an effect can be estimated from observed data

This sequential framing is intentional. Another way to put it is to separate

1. "The Science"
   - Definition of the object of inference, what we wish to learn
2. The Estimation
   - What we can learn about the science
   - How we go about learning it

The Potential Outcomes Framework

Sometimes called the Rubin Causal Model owing to foundational work in Rubin (1974, 1976, 1977, 1979, 1990)
Rooted in ideas dating back to Fisher (1918, 1925) and Neyman (1923)
Three main components of the framework:

1. Formulate *potential outcomes* corresponding to various levels of a "treatment"
2. Formulate the *assignment mechanism* governing how treatments are assigned

The Potential Outcomes Framework

Sometimes called the Rubin Causal Model owing to foundational work in Rubin (1974, 1976, 1977, 1979, 1990)
Rooted in ideas dating back to Fisher (1918, 1925) and Neyman (1923)
Three main components of the framework:

1. Formulate *potential outcomes* corresponding to various levels of a "treatment"
2. Formulate the *assignment mechanism* governing how treatments are assigned
3. Formulate a *model* for "the science" represented by all of the potential outcomes and covariates
   - This last part is actually optional...

Contrast with The Classical Paradigm

Contrast with The Classical Paradigm

In the potential outcomes paradigm . . .

- A **cause** is an *action* or *state* that occurs, relative to some other *action* or *state*
  - E.g., a person can smoke or not smoke
  - E.g., a school can increase teacher salaries or cut teacher pay
  - E.g., a product can be discounted or not
- An **effect** is a consequence of that explicitly defined cause, defined as a comparison between *what would potentially happen under competing states*

Thought Experiment: Did Pollution Cause Heart Attack?

- Martin has lived his whole life in part of Los Angeles where the air pollution concentration has always been high
- At age 70, Martin has a heart attack
- Did the pollution *cause* Martin's heart attack?

Thought Experiment: Did Pollution Cause Heart Attack?

- Martin has lived his whole life in part of Los Angeles where the air pollution concentration has always been high
- At age 70, Martin has a heart attack
- Did the pollution *cause* Martin's heart attack?
- **Question:** What would we need to know in order to conclude whether the pollution caused the heart attack?

Thought Experiment: Did Pollution Cause Heart Attack?

- Martin has lived his whole life in part of Los Angeles where the air pollution concentration has always been high
- At age 70, Martin has a heart attack
- Did the pollution *cause* Martin's heart attack?
- **Question:** What would we need to know in order to conclude whether the pollution caused the heart attack?

- What if we knew the following
  - If pollution had actually been low in Los Angeles for Martin's entire life
  - If Martin would have otherwise lived his life *exactly* the same
  - Martin would **not** have had a heart attack at age 70.
- Now would we conclude that the pollution *caused* Martin's heart attack?

Potential Outcomes and Definition of a Causal Effect

In order to know whether a particular action or state *caused* an outcome, we need to know:

1. What *would potentially happen* if the action or state occurred
2. What *would potentially happen* under some alternative action or state

A **causal effect** is a comparison between would potentially happen under competing actions or states:

$$\text{Outcome(if Action A) vs. Outcome (if Action B)}$$

Some questions....

- What counts as an *action* or *state* that can be considered a cause?
- How can we possibly know what would happen under two different actions/states?
- What about of many possible actions?
- What about studies with more than one person?
- How is this different from how we usually think about statistics?

Foundational Concept : Units

A population or universe, U, of **units**

- Objects of study on which causes or treatments may act
- Typically corresponds to standard statistical notions of the population from which a sample is drawn
- E.g., human subjects, laboratory equipment, households, plots of land
- For each $i \in U$, there are associated values:
  - *Response variables*, Y_i, of scientific interest
    - Overall goal may be to understand why values of Y vary over the units
  - *Attributes*, X_i, other variables describing the units
  - *Treatments*, Z_i, the cause to which each unit is exposed
    - Assume binary treatment $Z_i \in \{t, c\}$ for simplicity

What Statistics *Usually* Does

Associational Inference

- Write a statistical model that describes associations between random quantities
- Make inferences (estimates, tests, confidence intervals, posterior distributions, etc.) about unknown parameters in the model

- Units within a population $i \in U$
- Two types of variables defined on $i \in U$:
  1. "Response", Y_i
  2. "Attribute", X_i

     No logical distinction, both are just variables defined on U

What Statistics *Usually* Does....

Associational Inference

- Write a statistical model that describes associations between random quantities
- Make inferences (estimates, tests, confidence intervals, posterior distributions, etc.) about unknown parameters in the model

- *Associational parameters* are determined by $Pr(Y = y, X = x)$
  - E.g., $Pr(Y = y|X = x)$ would describe how Y changes with X
  - Regression $E[Y|X = x]$ is an associational parameter
- "Static" in the sense that associational parameters do not necessarily tell us about what would happen under a change in conditions

Towards Causal Inference

from Holland (1986)

When can understanding the joint distribution tell us about causal effects?

- "Causal analysis goes one step further; its aim is to infer aspects of the data generating process. With the help of such aspects, one can deduce not only the likelihood of events under static conditions, but also the dynamics of events under changing conditions.The additional information needed for making such predictions is provided by causal assumptions."
- "...one cannot substantiate causal claims from associations alone...behind every causal conclusion there must lie some causal assumption that is not testable."

Foundational Concept: Treatment

I.e., what can count as a "cause"

A *cause* is a quantity measured on each unit that could (possibly hypothetically) taken a different value

Helpful to think of "treatments" in the context of a (hypothetical) experiment

- Anything that could be reasonably applied or withheld from a unit
- **Key feature** is that the value of Z_i for each $i \in U$ *could have been different*
- Contrast with *attributes*, X_i

Question: What is the role of time? Where do X and Y fit relative to the time at which Z occurs?

What Types of "Causes"?

- *Put as bluntly and as contentiously as possible, in this article I take the position that causes are only those things that could, in principle, be treatments in experiments* - Holland (1986)

- *I do not intend to limit the discussion to activities within a controlled randomized study. I do it to emphasize...the fact that the effect of a cause is always relative to another cause.* - Holland (1986)

What Types of "Causes"?

- *Put as bluntly and as contentiously as possible, in this article I take the position that causes are only those things that could, in principle, be treatments in experiments* - Holland (1986)

- *I do not intend to limit the discussion to activities within a controlled randomized study. I do it to emphasize...the fact that the effect of a cause is always relative to another cause.* - Holland (1986)

Implications:

- Comparative questions
  - The effect of a cause is always defined relative to some other cause
- Invite the framing/language of experiments
- Forces us to be extremely explicit about what (mathematically) is meant by "causal effect"

"No causation without manipulation"

From Holland (1986)

Three statements, all using different meanings of "because" to explain the same "effect:"

(A) She did well on the exam because she was a woman.

(B) She did well on the test because she studied.

(C) She did well on the test because she was coached.

Which of (A), (B), (C) above can be a "cause" in this framework?

From Holland (1986)

Three statements, all using different meanings of "because" to explain the same "effect:"

(A) She did well on the exam because she was a woman.

- "Cause" as an attribute she possesses

(B) She did well on the test because she studied.

- "Cause" is a voluntary activity that was performed

(C) She did well on the test because she was coached.

- "Cause" is an activity that was imposed

Which of (A), (B), (C) above can be a "cause" in this framework?

Foundational Concept: Potential Outcomes

In order for Y to represent an effect of Z, it must be "post-exposure" or "post-treatment"

For the model to represent faithfully this state of affairs we need not a single variable, Y, to represent a response, but two variables, Y^t and Y^c...The interpretation of these two values, Y_i^t and Y_i^c for a given unit i, is that Y_i^t is the value of the response that would be observed if the unit were exposed to t and Y_i^c is the value that would be observed on the same unit if it were exposed to c - Holland (1986) [edited notation]

That is,

Outcome(if Action A) and Outcome (if Action B)

Foundational Concept: Potential Outcomes

Key departure from "associational inference" is the maintenance of the distinction between:

- Y_i^t the *potential outcome* that would be observed if i had exposure to $Z_i = t$
- Y_i^c the *potential outcome* that would be observed if i had exposure to $Z_i = c$
- Y_i^Z the *observed outcome* for unit i
  - $Y = Y^{obs} = Y_i^Z = Y_i^t$ when $Z_i = t$
  - $Y = Y^{obs} = Y_i^Z = Y_i^c$ when $Z_i = c$
- Z determines which of Y^t, Y^c is observed for a given $i \in U$
- Even though the model contains variables Z, Y^t, Y^c, the process of observation only involves Z, Y^Z

Potential Outcomes Framing

- $i = 1, 2, \ldots, n$ indexes *units*
- A "treatment" (or "action" or "exposure" or ...) is observed for each i
  - $Z = t, c$ corresponding to receipt (or not) of treatment
- Observed outcome of interest
  - Y, or Y^{obs}

Potential Outcomes Framing

- $i = 1, 2, \ldots, n$ indexes *units*
- A "treatment" (or "action" or "exposure" or ...) is observed for each i
  - $Z = t, c$ corresponding to receipt (or not) of treatment
- Observed outcome of interest
  - Y, or Y^{obs}
- **Potential Outcomes**
  - $Y_i^{Z=c} = Y_i^c$: What *would occur* of $Z_i = c$
  - $Y_i^{Z=t} = Y_i^t$: What *would occur* of $Z_i = t$

(need to keep track of both of these)

Notes on Potential Outcomes

$Y_i^{Z=c} = Y_i^c$: What *would occur* of $Z_i = c$
$Y_i^{Z=t} = Y_i^t$: What *would occur* of $Z_i = t$

Key Idea: Both potential outcomes exist

- Innate characteristics that exist before application of treatment
- Assignment to Z determines which Y^Z becomes observed, but does not affect the value of Y^t or Y^c
- Relationship between potential outcomes and observed data:

$$Y_i = Y_i^{obs} = Y_i^c \times \mathbb{1}(Z_i = c) + Y_i^t \times \mathbb{1}(Z_i = t)$$

Notes on Notation

- Potential outcomes notation can get quite cumbersome
- To make matters worse, there are different conventions for denoting potential outcomes
- "Easy" notational differences
  - Treatment variable is commonly denoted as S, W, Z, A
- "Headache-inducing" notational differences...

Foundational Concept: Causal Effect

A *causal effect* is defined as a contrast between outcomes for a given unit:

$$Y_i^t \quad \text{vs.} \quad Y_i^c$$

For example, t causes the effect $Y_i^t - Y_i^c$ (relative to c)

Note that the definition of the effect **does not** depend on: how many units, which outcome might actually be observed, parameters in a statistical model for the data

But what is the problem here?

Foundational Concept: Causal Effect

A *causal effect* is defined as a contrast between outcomes for a given unit:

$$Y_i^t \quad \text{vs.} \quad Y_i^c$$

For example, t causes the effect $Y_i^t - Y_i^c$ (relative to c)

Note that the definition of the effect **does not** depend on: how many units, which outcome might actually be observed, parameters in a statistical model for the data

But what is the problem here?

Impossible to observe both Y_i^t and Y_i^c!

An Outline for Practice of Causal Inference

1. Define (at least) two specific actions: 'Action A' vs. 'Action B' constituting a **treatment** or a **cause**
   - Doesn't have to be an "action" per se, could be a state or condition
   - Action A: Martin is exposed to high pollution
   - Action B: Martin is exposed to low pollution
2. Define the **effect** of 'Action A' vs. 'Action B' on outcome(s) of interest

   $1 + 2$ formalizes the question of interest
   (primarily conceptual, no statistics)

3. Design a study tailored to answer the question with data
   - "Design" can refer to prospective experimental design *or* a particular analysis of data that has already been observed (possibly non-experimental)
4. Estimate the effect, possibly with a statistical model

Separate "The Science" from "The Statistics"

The Science consists of how underlying potential outcomes encode how certain phenomena respond to intervention

- Y_i^t, Y_i^c are considered innate characteristics of units

All we can do as empirical scientists is *intervene* at a particular point in time to reveal some (but not all) of the underlying potential outcomes.

How we intervene (or what we assume about how certain treatments have been applied to units) will dictate how reliably we can "uncover" the underlying science and make causal inference

Example: Formulating Causal Questions

Question: Does the medicine prevent a heart attack?

Question: Does air pollution cause heart attack?

Question: Does pushing more advertisements make customers buy more?