9.S918: Quantitative Inference in Brain and Cognitive Sciences

Week 2 Day 12: Causal inference continued

Roger Levy
Dept. of Brain & Cognitive Sciences
Massachusetts Institute of Technology

February 12, 2025

Introductory causal inference

- You have probably had previous exposure to both probability and statistics
- You are less likely to have had exposure to **causal inference**
- Causal inference uses probability and statistics, but it is something separate from the traditional construal of those two fields
- You can think of causal inference as being a framework extending more traditional statistics by:
    - Adding new probability-based mathematical constructs; and,
    - Developing a set of practice for statistical inference based on those constructs
- Two causal inference frameworks:
    - The **potential outcomes** framework
    - The **causal graphical models** framework

The potential-outcomes framework

- In epidemiology and many other areas of statistics, causal inference was developed out of the idea of **potential outcomes** (Neyman 1923, Rubin 1974)
- Consider an outcome, Y, and a potential **treatment** A
- **Example:**

Y: an individual survives to the end of the year (0: no, 1: yes)

A: an individual with heart disease receives a heart transplant (0: no, 1: yes)

Potential-outcome random variables

- Suppose that A is discrete; for this case, $A \in \{0,1\}$
- The **potential outcomes**, or **counterfactual outcomes**, are random variables for Y for each potential value of A

$Y^{a=0}$	The value that Y would take if A were 0
$Y^{a=1}$	The value that Y would take if A were 1

- **Counterfactual risk** is the **expected value** of each counterfactual-outcome random variable:

$$E[Y^{a=0}] \qquad\qquad E[Y^{a=1}]$$

- Expected value, or expectation, is defined as follows:

$$E[X] = \sum_x xP(X = x)$$

- So we are interested in (and likewise for $Y^{a=1}$):

$$E[Y^{a=0}] = \sum_y yP(Y^{a=0} = y) = 0 \times P(Y^{a=0} = 0) + 1 \times P(Y^{a=0} = 1) = \boxed{P(Y^{a=0} = 1)}$$

Counterfactual data and causal effects

- Suppose we knew **what would happen** for each individual in the population under each value of the treatment
- Then we could compute the counterfactual risks:

$$E[Y^{a=0}] = 0.5 \qquad E[Y^{a=1}] = 0.5$$

- The **average causal effect** of treatment A is defined as the difference of counterfactual risks:

$$E[Y^{a=1}] - E[Y^{a=0}] = 0$$

- Here, treatment is **ineffective**

(Hernan & Robins, 2020, Table 1.1)

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Cyclope	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0
$P(Y^{a=*}) = 1$	0.5	0.5

Estimating causal effects

Remember, $E[Y^{a=i}] = P(Y^{a=i} = 1)$

	L	A	Y	Y^0	Y^1
Rheia	0	0	0	0	?
Kronos	0	0	1	1	?
Demeter	0	0	0	0	?
Hades	0	0	0	0	?
Hestia	0	1	0	?	0
Poseidon	0	1	0	?	0
Hera	0	1	0	?	0
Zeus	0	1	1	?	1
Artemis	1	0	1	1	?
Apollo	1	0	1	1	?
Leto	1	0	0	0	?
Ares	1	1	1	?	1
Athena	1	1	1	?	1
Hephaestus	1	1	1	?	1
Aphrodite	1	1	1	?	1
Polyphemus	1	1	1	?	1
Persephone	1	1	1	?	1
Hermes	1	1	0	?	0
Hebe	1	1	0	?	0
Dionysus	1	1	0	?	0

- Naively, we might estimate the counterfactual risks $P(Y^{a=i} = 1)$ directly from observed A and Y:

$$\hat{P}_{MLE}(Y = 1 \,|\, A = 0) = \frac{3}{7} \quad \hat{P}_{MLE}(Y = 1 \,|\, A = 1) = \frac{7}{13}$$

- But under what circumstances $\hat{P}_{MLE}(Y \,|\, A = i) = \hat{P}_{MLE}(Y^{a=i} = 1)$?
- The following is certainly true:

$$\hat{P}_{MLE}(Y = 1 \,|\, A = i) = \frac{\text{Count}(Y = 1 \wedge A = i)}{\text{Count}(A = i)}$$

$$= \frac{\text{Count}(Y^{a=1} = 1 \wedge A = i)}{\text{Count}(A = i)}$$

$$= \hat{P}_{MLE}(Y^{a=i} = 1 \,|\, A = i)$$

Consistency: when $A = i,\, Y = Y^{a=i}$

Crucial step; make sure you understand it!

- So, the following condition suffices:

$$P(Y^{a=i} = 1 \,|\, A = i) = P(Y^{a=i} = 1)$$

- This is called **Exchangeability**:

$$Y^a \perp A \,|\, \{\}$$

Exchangeability and randomization

Goal: $\hat{P}(Y^a = 1)$

- Why is a randomized experiment so powerful?
- Recap of exchangeability criterion:

$$Y^a \perp A \,|\, \{\}$$

- If we ourselves determine A in a way that is *truly blind to* Y^a, it **imposes** exchangeability!
- We can now go ahead and estimate

$$\hat{P}(Y^{a=i} = 1) = \hat{P}(Y = 1 \,|\, A = i)$$

- Hooray!!!

Rheia
Kronos
Demeter
Hades
Hestia
Poseidon
Hera
Zeus
Artemis
Apollo
Leto
Ares
Athena
Hephaestus
Aphrodite
Polyphemus
Persephone
Hermes
Hebe
Dionysus

Does loss of randomization make things hopeless?

- In the real world, many datasets are ***not*** randomized this way
- **Example:** let's imagine some other variable that might affect whether treatment A is applied; e.g., $L =$ whether the patient was in critical condition (1=yes, 0=no)
- In general, L will be related to Y^a
  - E.g., in this example, patients in critical condition are surely more likely to die overall!

$$\cancel{A \perp Y \mid \{\}}$$

	L
Rheia	0
Kronos	0
Demeter	0
Hades	0
Hestia	0
Poseidon	0
Hera	0
Zeus	0
Artemis	1
Apollo	1
Leto	1
Ares	1
Athena	1
Hephaestus	1
Aphrodite	1
Polyphemus	1
Persephone	1
Hermes	1
Hebe	1
Dionysus	1

Conditional exchangeability

- But now suppose we have observed (i.e., it's in our dataset) the factor L that affected whether the treatment A was applied
- If the following condition holds, it can help us estimate the counterfactual risks $P(Y^a)$:

$$A \perp Y^a \mid L$$

- That is, L captures all the information available in A that is relevant to all Y^a
- This is called CONDITIONAL EXCHANGEABILITY

	L	A	$Y^{a=0}$	$Y^{a=1}$	Y
Rheia	0	0	0	1	0
Kronos	0	0	1	0	1
Demeter	0	0	0	0	0
Hades	0	0	0	0	0
Hestia	0	1	0	0	0
Poseidon	0	1	1	0	0
Hera	0	1	0	0	0
Zeus	0	1	0	1	1
Artemis	1	0	1	1	1
Apollo	1	0	1	0	1
Leto	1	0	0	1	0
Ares	1	1	1	1	1
Athena	1	1	1	1	1
Hephaestus	1	1	0	1	1
Aphrodite	1	1	0	1	1
Polyphemus	1	1	0	1	1
Persephone	1	1	1	1	1
Hermes	1	1	1	0	0
Hebe	1	1	1	0	0
Dionysus	1	1	1	0	0

Using conditional exchangeability

- Can we estimate $P(Y^{a=i} = 1 \mid L)$?
- It turns out we can!

$P(Y^{a=i} = 1 \mid L) = P(Y^{a=i} = 1 \mid L, A)$ **CONDITIONAL EXCHANGEABILITY**

$$\hat{P}_{\text{MLE}}(Y^{a=i} = 1 \mid L = j, A = k) = \frac{\text{Count}(Y^{a=i} = 1, L = j, A = k)}{\text{Count}(L = j, A = k)}$$

- In estimating this condprob:
  - when $i = k$ we use CONSISTENCY
  - WHEN $i \neq k$ we have "missing data", so ignore those instances

CONSISTENCY: when $A = i, Y = Y^{a=i}$

	L	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	0	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	1
Leto	1	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Polyphemus	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

L	A	Count(L, A)	Count$(Y^{a=0} = 1, L, A)$	Count$(Y^{a=1} = 1, L, A)$	$\hat{P}_{\text{MLE}}(Y^{a=i} = 1 \mid L)$
0	0	4	1	?	1/4
0	1	4	?	1	1/4
1	0	3	2	?	2/3
1	1	9	?	6	2/3

This is just like estimating $P(Y \mid L, A)$!

Using conditional exchangeability

- We originally characterized our goal as estimating the **counterfactual risks** $E[Y^{a=i}] = P(Y^{a=i} = 1)$
- With conditional exchangeability, we estimated $P(Y^{a=i} = 1 \,|\, L)$ – these are called **stratum-specific risks** (where each **stratum** is a value of L)
- Often, this may be all you need or want
  - If the causal effect of A depends on L, then "summarizing out" L discards information!
- But there are situations where the basic counterfactual risk $E[Y^{a=i}]$ may be of interest
  - e.g., "how many lives would it save if everyone who came to the hospital with heart disease received a heart transplant?
- But we can recover the basic counterfactual risks through **standardization** (or the mathematically equivalent **inverse probability weighting**)

Standardization

L = whether the patient was in critical condition (1=yes, 0=no)

L	A	$\hat{P}_{\text{MLE}}(Y^{a=i} \mid L)$
0	0	1/4
0	1	1/4
1	0	2/3
1	1	2/3

	L	A	Y	Y^0	Y^1
Rheia	0	0	0	0	?
Kronos	0	0	1	1	?
Demeter	0	0	0	0	?
Hades	0	0	0	0	?
Hestia	0	1	0	?	0
Poseidon	0	1	0	?	0
Hera	0	1	0	?	0
Zeus	0	1	1	?	1
Artemis	1	0	1	1	?
Apollo	1	0	1	1	?
Leto	1	0	0	0	?
Ares	1	1	1	?	1
Athena	1	1	1	?	1
Hephaestus	1	1	1	?	1
Aphrodite	1	1	1	?	1
Polyphemus	1	1	1	?	1
Persephone	1	1	1	?	1
Hermes	1	1	0	?	0
Hebe	1	1	0	?	0
Dionysus	1	1	0	?	0

$$\hat{P}_{MLE}(L=1) = \frac{12}{20} = \frac{3}{5}$$

- By the law of total probability,

We have just estimated this

We can estimate this from the data, too

$$P(Y^{a=i}) = \sum_j P(Y^{a=i} \mid L)\, P(L)$$

- Expanding the sum and plugging in our estimates we get:

$$P(Y^{a=i}=1) = P(Y^{a=i}=1 \mid L=0)P(L=0) + P(Y^{a=i}=1 \mid L=1)P(L=1)$$
$$= \frac{1}{4} \times \frac{2}{5} + \frac{2}{3} \times \frac{3}{5}$$
$$= \frac{1}{10} + \frac{2}{5}$$
$$= \frac{5}{10} = \frac{1}{2}$$

(Because $\hat{P}_{\text{MLE}}(Y^{a=i} \mid L)$ are the same for $a=0$ and $a=1$, this work gives us the result for both counterfactual treatments, and the risk ratio is 1)

IDENTIFIABILITY of causal effects

- IDENTIFIABILITY means, our assumptions allow the causal effect we are interested in to be ***uniquely estimated*** from the available data (set of observed/measured variables)
  - "Uniquely estimate": if we had an arbitrary large quantity of data, we could estimate the causal effect with arbitrarily high accuracy and precision
- Simple case of unidentifiability: Hernan & Robins's heart transplant example, if L (severity of disease) affects probability of a heart transplant and we don't measure it
  - Suppose that people with transplants have lower survival rates: $\hat{P}_{MLE}(Y|A=1) < \hat{P}_{MLE}(Y|A=0)$
    - Could be because heart transplants are dangerous
    - Or: sicker people are more likely to get transplants!

Not an issue of sample size—no amount of data would help!

The three criteria for identifiability

- **Consistency**: $Y = Y^{a=i}$ whenever $A = i$
  - Consequence: different individuals' outcomes don't affect each other
  - Consequence: there can be no "multiple versions" of the same treatment A in terms of their influence on Y
- **Conditional Exchangeability**: for all i, $Y^{a=i} \perp A \,|\, Z$ for some set of observed variables Z
  - Consequence: there can be no "hidden common causes" or "hidden mediators" of A and $Y^{a=i}$
- **Positivity**: for all i and all values of Z, $P(A = i \,|\, Z) > 0$
  - e.g., in our example, it can't be the case that individuals with heart disease are *always* given transplants
- If all three criteria hold, we can estimate causal effects

Summary of intro to potential outcomes

- The potential outcomes framework formalizes causal effects (or risks) through **counterfactual outcome** (also called **potential outcome**) **variables**
- At most one counterfactual outcome is observable in each datum, so causal effects cannot in general be naively estimated from data

	L	A	Y	Y^0	Y^1
Rheia	0	0	0	0	?
Kronos	0	0	1	1	?
Demeter	0	0	0	0	?
Hades	0	0	0	0	?
Hestia	0	1	0	?	0
Poseidon	0	1	0	?	0

- However, if the three following conditions hold, the data can be viewed as a **conditionally randomized experiment** and causal effects can be estimated

Consistency	**Conditional Exchangeability**	**Positivity**
$Y = Y^{a=i}$ whenever $A = i$	$\exists Z. \forall i. Z$ is observed $\wedge\, Y^{a=i} \perp A \mid Z$	$\forall i \,.\, \forall Z \,.\, P(A = i \mid Z) > 0$

- This analysis also sheds light on the power of randomized experiments: they offer **unconditional exchangeability**