Introduction to Potential Outcomes

Thomas Richardson
University of Washington

Simons Causal Bootcamp
20 January 2022

Outline

- A brief History of Causation
- Potential Outcomes and Counterfactuals
- Randomized experiments
- Average Causal Effect (ACE)
- Observation Studies

Causation

Democritus (460-390 BC)
(aka the laughing philosopher because he emphasized the value of cheerfulness)

"I would rather discover a single causal relationship than be king of Persia"

The potential outcomes framework: philosophy

Hume (1748) *An Enquiry Concerning Human Understanding*:

We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second, ...

... where, if the first object had not been the second never had existed.

Note: this is not one of the 3(!) causal theories Hume is famous for.

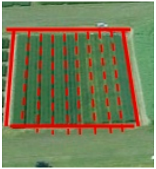
Causation

Agricultural field trials: wish to know which seed varieties produce (cause) the greatest yield... but different plots (of land) have different fertility, drainage etc.,

The potential outcomes framework: crop trials

Jerzy Neyman (1923):

To compare v varieties [on m plots] we will consider numbers:



$$\overbrace{\begin{matrix} U_{11} & \ldots & U_{1m} \\ \vdots & & \vdots \\ U_{v1} & \ldots & U_{vm} \end{matrix}}^{\text{plots}} \left.\vphantom{\begin{matrix} U_{11} \\ \vdots \\ U_{v1} \end{matrix}}\right\} \text{varieties}$$

U_{ij} is crop yield that would be observed if variety i were planted in plot j.

Physical constraints only allow one variety to be planted in a given plot in any given growing season $\Rightarrow$ Observe only one number per col.

Application to clinical trials

- Each patient in study is assigned to either:
  - Treatment (aka Drug) ($X = 1$)
  - Control (aka Placebo) ($X = 0$)
- For each patient we observe one outcome (Y), either:
  - Good e.g. Recover ($Y = 1$)
  - Bad e.g. Die ($Y = 0$)

Plots in a field $\Rightarrow$ *Patients*; *Kg of wheat* $\Rightarrow$ *Live or Die*

Potential outcomes with binary treatment and outcome

For binary treatment X, we define two potential outcome variables:

- $Y(x = 0)$: the value of Y that *would* be observed for a given unit *if* assigned $X = 0$;
- $Y(x = 1)$: the value of Y that *would* be observed for a given unit *if* assigned $X = 1$;

$Y(x = 0)$ and $Y(x = 1)$ are two different random variables (not different realizations of the same variable).

Notation: We will use $Y(x_i)$ as an abbreviation for $Y(x = i)$

Popularized by Rubin (1974); sometimes called the 'Neyman-Rubin causal model'.

Alternative notations for $Y(x = i)$ used by other authors: $Y^{x=i}$ or $Y_{x=i}$.

Potential Outcomes

Unit	Potential Outcomes	
	$Y(x=0)$	$Y(x=1)$
1	0	1
2	0	1
3	0	0
4	1	1
5	1	0

Potential Outcomes

Unit	Potential Outcomes		Observed	
	$Y(x = 0)$	$Y(x = 1)$	X	Y
1	0	1	1	
2	0	1	0	
3	0	0	1	
4	1	1	1	
5	1	0	0	

Potential Outcomes

Unit	Potential Outcomes		Observed	
	$Y(x=0)$	$Y(x=1)$	X	Y
1	0	1	1	1
2	0	1	0	0
3	0	0	1	0
4	1	1	1	1
5	1	0	0	1

Consistency Axiom

$$Y = (1 - X) \cdot Y(x = 0) + X \cdot Y(x = 1)$$

equivalently:

$$X = x \qquad \Rightarrow \qquad Y = Y(x).$$

In words, we have the following tautology:

For an individual who has $X = x$*, their observed response* Y *is equal to the response* $Y(x)$ *that would be observed had* X *been* x*.*

Drug Response Types:

In the simplest case where Y is a binary outcome we can think of patients as belonging to one of 4 'types':

$Y(x_0)$	$Y(x_1)$	Name
0	0	*Never Recover*
0	1	*Helped*
1	0	*Hurt*
1	1	*Always Recover*

Actual vs. Potential outcomes

Key Distinction

- X is the treatment that a given patient gets;
  thus far, this need not be randomly assigned, and could result from doctor and patient choices;

- Y is the observed response for a given patient;

- $Y(x)$ is the response that would be observed for a given paitent if (possibly counter to fact) they received $X = x$.

Potential Outcomes and Missing Data

Fundamental Problem of Causal Inference:
We never observe both $Y(x=0)$ and $Y(x=1)$.

Unit	Potential Outcomes		Observed	
	$Y(x=0)$	$Y(x=1)$	X	Y
1	?	1	1	1
2	0	?	0	0
3	?	0	1	0
4	?	1	1	1
5	1	?	0	1

Stable Unit Treatment Value Assumption (SUTVA)

- $Y(x = 0)$: the value of Y that *would* be observed for a given unit *if* assigned $X = 0$;
- $Y(x = 1)$: the value of Y that *would* be observed for a given unit *if* assigned $X = 1$;

Implicit Assumption: these outcomes, $Y(x = 0)$, $Y(x = 1)$ are 'well-defined'. Specifically:

- Only one version of $X = 1$ and $X = 0$;
  (only one version of 'drug' and 'placebo')
- Subject's outcome only depends on what they receive:
  no 'interference' between units (SUTVA).
  (Might not hold in a vaccine trial for an infectious disease if subjects are in contact.)

Average Causal Effect (ACE) of X on Y

$$\begin{aligned} \text{ACE}(X \to Y) &\equiv E[Y(x_1) - Y(x_0)] \\ &= p(\textit{Helped}) - p(\textit{Hurt}) \qquad \in [-1, 1] \end{aligned}$$

Thus $\text{ACE}(X \to Y)$ is the difference in % recovery if everybody treated ($X = 1$) vs. if nobody treated ($X = 0$).

Identification of the ACE under randomization

If X is assigned randomly then

$$X \perp\!\!\!\perp Y(x_0) \quad \text{and} \quad X \perp\!\!\!\perp Y(x_1) \tag{1}$$

$$\begin{aligned} P(Y(x_i) = 1) &= P(Y(x_i) = 1 \mid X = i) \quad \text{(Why?)} \\ &= P(Y = 1 \mid X = i) \quad \text{(Why?)} \end{aligned}$$

Thus:

$$\begin{aligned} \text{ACE}(X \to Y) &= E[Y(x_1) - Y(x_0)] \\ &= E[Y(x_1)] - E[Y(x_0)] \\ &= E[Y(x_1) \mid X = 1] - E[Y(x_0) \mid X = 0] \\ &= E[Y \mid X = 1] - E[Y \mid X = 0]. \end{aligned}$$

Thus if (1) holds then $\text{ACE}(X \to Y)$ is identified from $P(Y \mid X)$.

Two-way Table

Under randomization, the relationship between the counterfactual distribution $P(Y(x_0), Y(x_1))$ and the observed distributions $\{P(Y \mid x_0), P(Y \mid x_1)\}$ is:

		col sums	
		$P(Y{=}0 \mid X{=}0)$	$P(Y{=}1 \mid X{=}0)$
row sums	$P(Y{=}0 \mid X{=}1)$	$P(Y(x_0){=}0, Y(x_1){=}0)$	$P(Y(x_0){=}1, Y(x_1){=}0)$
	$P(Y{=}1 \mid X{=}1)$	$P(Y(x_0){=}0, Y(x_1){=}1)$	$P(Y(x_0){=}1, Y(x_1){=}1)$

Here $P(Y{=}i \mid X{=}j) = P(Y(x_j){=}i)$ due to randomization.

Equivalently we may write this in terms of types

	$P(Y{=}0 \mid X{=}0)$	$P(Y{=}1 \mid X{=}0)$
$P(Y{=}0 \mid X{=}1)$	$P(\text{NR})$	$P(\text{HU})$
$P(Y{=}1 \mid X{=}1)$	$P(\text{HE})$	$P(\text{AR})$

Identification Problem

Want: $P(Y(x_0), Y(x_1))$; Given: $P(Y \mid X=0), P(Y \mid X=1)$

Under randomization, as before: $X \perp\!\!\!\perp Y(x_i)$ implies:

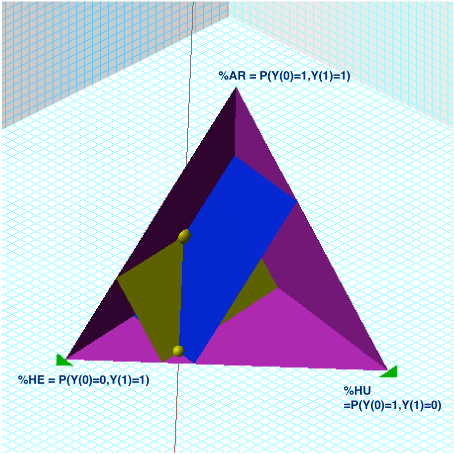
$$P(Y(x_i) = 1) = P(Y(x_i) = 1 \mid X = i) = P(Y = 1 \mid X = i).$$

Thus the observed joint $P(Y|X)$ puts two restrictions on $P(Y(x_0), Y(x_1))$:

$$\begin{aligned} P(Y=1 \mid X=0) &= P(Y(x_0)=1, Y(x_1)=0) + P(Y(x_0)=1, Y(x_1)=1) \\ P(Y=1 \mid X=1) &= P(Y(x_0)=0, Y(x_1)=1) + P(Y(x_0)=1, Y(x_1)=1). \end{aligned}$$

Each restriction implies a 2-d subset in Δ_3.
Intersection forms a 1-d subset on which ACE is constant.

Graphing Calculator Plot

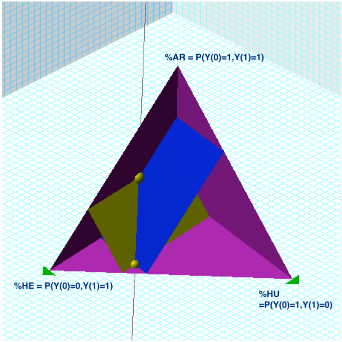


In this plot:

$P(Y{=}1 \mid X{=}0) = P(Y(x_0) = 1) = \%HU + \%AR =$ 0.3, (yellow)

$P(Y{=}1 \mid X{=}1) = P(Y(x_1) = 1) = \%HE + \%AR =$ 0.6, (blue)

Fréchet inequalities



Equation for line segment in simplex:

$$\left\{\begin{array}{lll} P(1,1) & = & t \\ P(1,0) & = & c_0 - t \\ P(0,1) & = & c_1 - t \\ P(0,0) & = & 1 - c_0 - c_1 + t \end{array} \quad \begin{array}{r} t \in \left[\max\{0, (c_0 + c_1) - 1\}, \min\{c_0, c_1\}\right] \\ c_0 \equiv P(Y=1 \mid X=0) \\ c_1 \equiv P(Y=1 \mid X=1) \end{array}\right\}$$

Extreme points are given by ‘Fréchet inequalities’.

Big Picture: Connecting Distributions in Experiment

Counterfactual *Observed*

$$\Delta_3 \rightarrow \Delta_1 \times \Delta_1$$

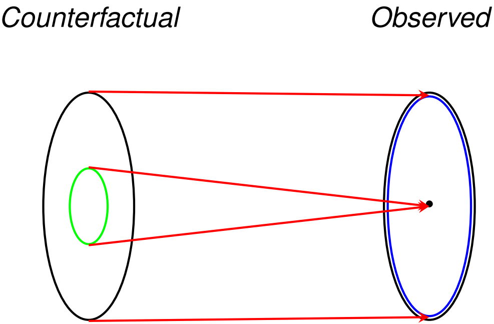
$$P(Y(x_0), Y(x_1)) \mapsto \{P(Y \mid x_0), P(Y \mid x_1)\}$$

$$= \{P(Y(x_0)), P(Y(x_1))\}$$

(by Randomization)

$$\{\%HE, \%HU, \%NR, \%AR\} \mapsto \{\%HU + \%AR,\ \%HE + \%AR\}$$

Identification Problem under Experiment



$$\Delta_3 \quad \to \quad \Delta_1 \times \Delta_1$$

$$P(Y(x_0), Y(x_1)) \quad \mapsto \quad \{P(Y \mid x_0), P(Y \mid x_1)\}$$
$$= \{P(Y(x_0)), P(Y(x_1))\}$$

(by Randomization)

$$\{\%HE, \%HU, \%NR, \%AR\} \quad \mapsto \quad \{\%HU + \%AR,\ \%HE + \%AR\}$$

Observational study; no randomization

Suppose that we do not know that $X \perp\!\!\!\perp Y(x_0)$ and $X \perp\!\!\!\perp Y(x_1)$.
What can be inferred about the ACE?

P(X,Y)	*Placebo* $X = 0$	*Drug* $X = 1$
Die: $Y = 0$	7/20	4/20
Live: $Y = 1$	3/20	6/20

What is:

- The largest proportion of people of type *Helped*,
  $P(Y(x_0)=0, Y(x_1)=1)$? $(6 + 7)/20 = 0.65$
- The smallest proportion of people of type *Hurt*,
  $P(Y(x_0)=1, Y(x_1)=0)$? 0

  $\Rightarrow$ Max value of ACE: $(6 + 7)/20 - 0 = 0.65$

Similar logic:

$\Rightarrow$ Min value of ACE: $0 - (4 + 3)/20 = -0.35$

(Note, as before, $P(Y = 1 \mid X = 0) = 0.3$, $P(Y = 1 \mid X = 1) = 0.6$.)

Inference for the ACE without randomization

Suppose that we do not know that $X \perp\!\!\!\perp Y(x_0)$ and $X \perp\!\!\!\perp Y(x_1)$.

What can be inferred from the observed distribution $P(X, Y)$?

General case:

$$
\begin{aligned}
-(P(X{=}0, Y{=}1) + P(X{=}1, Y{=}0)) \\
\leqslant \quad ACE(X \rightarrow Y) \\
\leqslant P(X{=}0, Y{=}0) + P(X{=}1, Y{=}1)
\end{aligned}
$$

$\Rightarrow$ Bounds will always include zero.

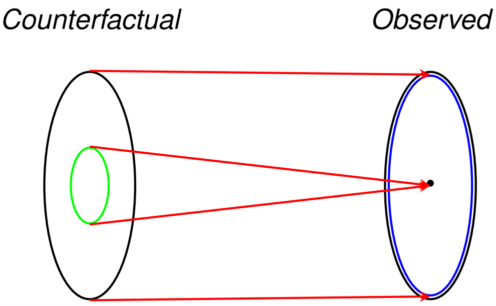
What further information can we obtain?

Observational study: one-way table!

Observed	Counterfactual	
$p(X=0, Y=0)$	$p(X=0, Y(x_0)=0, Y(x_1)=0)$	$p(X=0, Y(x_0)=0, Y(x_1)=1)$
$p(X=0, Y=1)$	$p(X=0, Y(x_0)=1, Y(x_1)=0)$	$p(X=0, Y(x_0)=1, Y(x_1)=1)$
$p(X=1, Y=0)$	$p(X=1, Y(x_0)=0, Y(x_1)=0)$	$p(X=1, Y(x_0)=1, Y(x_1)=0)$
$p(X=1, Y=1)$	$p(X=1, Y(x_0)=0, Y(x_1)=1)$	$p(X=1, Y(x_0)=1, Y(x_1)=1)$

Observed	Counterfactual	
$p(X=0, Y=0)$	$p(X=0, \text{NR})$	$p(X=0, \text{HE})$
$p(X=0, Y=1)$	$p(X=0, \text{HU})$	$p(X=0, \text{AR})$
$p(X=1, Y=0)$	$p(X=1, \text{NR})$	$p(X=1, \text{HU})$
$p(X=1, Y=1)$	$p(X=1, \text{HE})$	$p(X=1, \text{AR})$

Identification Problem



$$\Delta_7 \quad \rightarrow \quad \Delta_3$$

$$P(X, Y(x_0), Y(x_1)) \quad \mapsto \quad P(X, Y)$$

Wish to know set of $P(Y(x_0), Y(x_1))$ margins of distns $P(X, Y(x_0), Y(x_1))$ mapping to a given observed distribution $P(X, Y)$.

Want: $P(Y(x_0), Y(x_1))$; Given: $P(X, Y)$

Bounds on joints $P(Y(x_0), Y(x_1))$

Observed	Counterfactual	
$p(X=0, Y=0)$	$p(X=0, \text{NR})$	$p(X=0, \text{HE})$
$p(X=0, Y=1)$	$p(X=0, \text{HU})$	$p(X=0, \text{AR})$
$p(X=1, Y=0)$	$p(X=1, \text{NR})$	$p(X=1, \text{HU})$
$p(X=1, Y=1)$	$p(X=1, \text{HE})$	$p(X=1, \text{AR})$

$$
\begin{aligned}
0 \leqslant \quad \%HE \quad &\leqslant P(X=0, Y=0) + P(X=1, Y=1) \\
0 \leqslant \quad \%HU \quad &\leqslant P(X=0, Y=1) + P(X=1, Y=0) \\
0 \leqslant \quad \%NR \quad &\leqslant P(X=0, Y=0) + P(X=1, Y=0) \ = \ P(Y=0) \\
0 \leqslant \quad \%AR \quad &\leqslant P(X=0, Y=1) + P(X=1, Y=1) \ = \ P(Y=1)
\end{aligned}
$$

Bounds on margins $P(Y(x_i))$

Observed	Counterfactual	
$p(X=0, Y=0)$	$p(X=0, \text{NR})$	$p(X=0, \text{HE})$
$p(X=0, Y=1)$	$p(X=0, \text{HU})$	$p(X=0, \text{AR})$
$p(X=1, Y=0)$	$p(X=1, \text{NR})$	$p(X=1, \text{HU})$
$p(X=1, Y=1)$	$p(X=1, \text{HE})$	$p(X=1, \text{AR})$

We also have the following inequalities on the marginals:

$$P(Y(x_0) = 1) = P(\text{HU}) + P(\text{AR})$$

$$P(Y(x_1) = 1) = P(\text{HE}) + P(\text{AR})$$

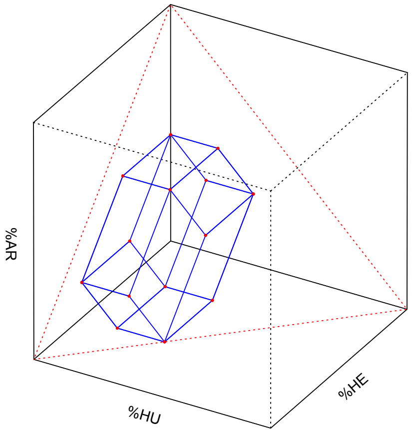
$$P(X = 0, Y = 1) \leqslant P(Y(x_0) = 1) \leqslant 1 - P(X = 0, Y = 0)$$

$$P(X = 1, Y = 1) \leqslant P(Y(x_1) = 1) \leqslant 1 - P(X = 1, Y = 0)$$

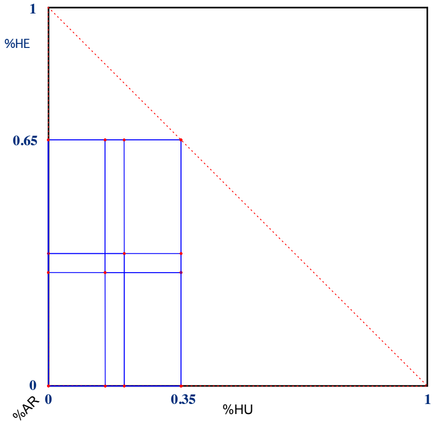
Thus we have 6 pairs of parallel planes.

Polytope for observational study

Set of margins $P(Y(x_0), Y(x_1))$ compatible with the Obs. Study.



Checking ACE bounds



This confirms the ACE bounds we derived earlier.

(But why is this helpful!?)

Summary so far

- Causal contrasts compare the *potential* outcomes of the same units under different treatments:
- In our observed data, for each unit one outcome will be 'actual'; the others will be 'counterfactual'.
  (Exceptions in fields where cross-over designs are possible.)
- The potential outcome framework allows
  Causation to be 'reduced' to *Missing Data*
  $\Rightarrow$ Conceptual progress!
- The ACE is identified if $X \perp\!\!\!\perp Y(x_i)$ for all values x_i.
- Randomization of treatment assignment implies $X \perp\!\!\!\perp Y(x_i)$.
- Without independence the ACE is not identified, and cannot be bounded away from zero.