Title: Counterfactual and Graphical Frameworks for Causal Modeling

Speakers: Thomas Richardson

Series: Quantum Foundations, Quantum Information

Date: September 16, 2024 - 10:50 AM

URL: https://pirsa.org/24090084

Abstract: In the Statistics literature there are three main frameworks for causal modeling: counterfactuals (aka potential outcomes), non-parametric structural equation models (NPSEMs) and graphs (aka path diagrams or causal Bayes nets). These approaches are similar and, in certain specific respects, equivalent. However, there are important conceptual differences and each formulation has its own strengths and weaknesses. These divergences are of relevance both in theory and when the approaches are applied in practice. This talk will introduce the different frameworks, and describe, through examples, both the commonalities and dissimilarities. In particular, we will see that the â€œdefaultâ€• assumptions within these frameworks lead to different identification results when quantifying mediation and, more generally, path-specific effects.

Counterfactual and Graphical Frameworks for Causal Modeling

Thomas Richardson

Causalworlds 16 September 2024

Collaborators: James Robins (Harvard); Robin Evans (Oxford); Ilya Shpitser (Johns Hopkins)

Outline

I Causal Models: Three approaches
- Potential Outcomes
- Non-Parametric Structural Equations
- Graphs

II Relations between these approaches
- Graphical unification

Causal Models link Worlds

All aim to relate two types of situation:

- An observational world:
  A 'natural' process assigns 'treatments'.
  Example: each patient chooses their own treatment.
- An experimental world
  'Treatments' assigned via an 'external' process.
  Example: each patient is given the same treatment.

Basic Inferential Tasks

- Given observational data make predictions about what would be observed in an experimental setting.
- Given experimental data predict what happens in an observational context.
  For example, where not everyone may wish to avail themselves of treatment.
- Combine experimental and observational data to predict the result of some experiment that was not performed.

High Level View of Frameworks: Ontology

All relate observational and experimental, but with different objects:

- Potential Outcomes: Neyman (1923)

  Experimental and observational distributions are margins of a single joint:

  $$P(X, Y, Y(x=0), Y(x=1)) \Rightarrow P(X, Y)\ P(Y(x_0))\ P(Y(x_1))$$

  All events defined on a single sample-space.

- Structural Equations: Haavelmo (1943)

  Eq. Model for Observed vars $\Rightarrow$ Eq. Model for intervened system

- Graphical Causal Models: Wright (1923)

  Separate experimental and observational distributions

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

  No single sample-space;

  Alternative notation: $P(X, Y)\ P(Y \mid \mathrm{do}(x = 0))\ P(Y \mid \mathrm{do}(x = 1))$.

Potential Outcomes aka Counterfactual Models

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

$Y(x = 0)$ 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 $Y(x = i)$

Rubin (1974) applied to observational data; sometimes called the 'Neyman-Rubin causal model'.

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;
  Stable Unit Treatment Value Assumption (SUTVA).
  (Might not hold in a vaccine trial for an infectious disease if subjects are in contact.)

Drug Response Types:

Simplest case: outcome taking values 0, 1;
1 indicate a good outcome
patients are 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*

Potential Outcomes

The potential outcomes describe two different experimental worlds, in which everyone receives $x = 0$ or $x = 1$:

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

Observed Outcomes

Conceptually the data for the observational world is obtained from the 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	
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

Thus:

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

equivalently:

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

or even more simply: $Y = Y(X)$.

Conceptually: $X, Y(x_0), Y(x_1)$ are primitive;
Y is derived as a deterministic fn. of $X, Y(x_0), Y(x_1)$.

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

Consequence: The distribution $P(X, Y(x_0), Y(x_1))$ is not identified.

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 the process that assigned X (in the 'observational' world) assigned X randomly then

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

So:

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

Thus:

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

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

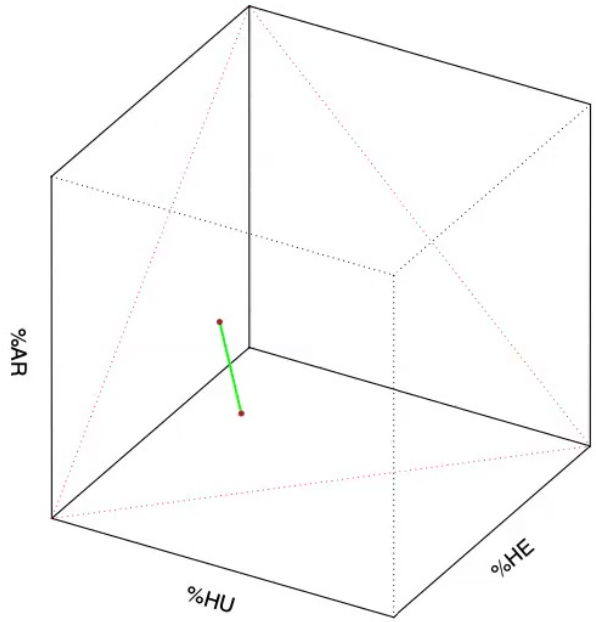
Inference: 'Observational' world ⇒ Difference of two Exp. Worlds

Random Assignment; Poss. Distn. for $P(Y(x_0), Y(x_1))$

If we know $X \perp\!\!\!\perp Y(x_0)$ and $X \perp\!\!\!\perp Y(x_1)$

$P(X, Y)$	$X = 0$	$X = 1$
$Y = 0$	0.35	0.20
$Y = 1$	0.15	0.30

$P(\text{HE}) = P(Y(x_0) = 0, Y(x_1) = 1)$,
likewise for $P(\text{HU})$, $P(\text{AR})$.



$$P(Y{=}1 \mid X{=}0) = P(Y(x_0) = 1) = \%\text{HU} + \%\text{AR} = 0.3,$$
$$P(Y{=}1 \mid X{=}1) = P(Y(x_1) = 1) = \%\text{HE} + \%\text{AR} = 0.6,$$
$$\text{ACE}(X \to Y) = 0.6 - 0.3 = 0.3.$$

Inference for the ACE without randomization

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

The ACE is not identified. We obtain these bounds:

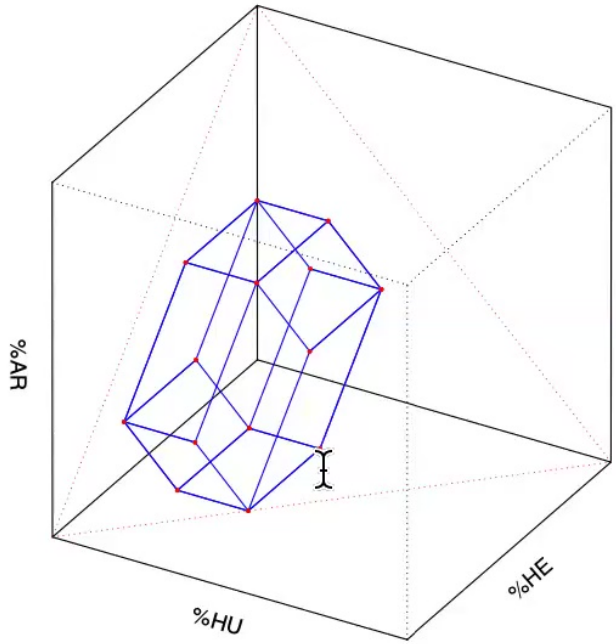
$$
\begin{aligned}
-[P(X{=}0, Y{=}1) + P(X{=}1, Y{=}0)] \\
\leqslant \quad ACE(X \to Y) \quad \leqslant \\
P(X{=}0, Y{=}0) + P(X{=}1, Y{=}1)
\end{aligned}
$$

$\Rightarrow$ Bounds will always include zero.

No Random Assignment; Poss. Distn. for $P(Y(x_0), Y(x_1))$

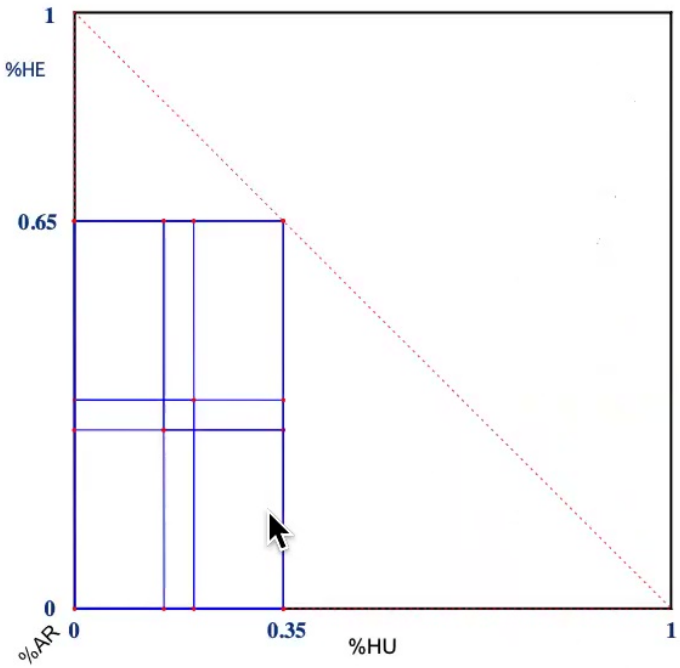
Without assuming treatment assigned randomly:

$P(X, Y)$	$X = 0$	$X = 1$
$Y = 0$	0.35	0.20
$Y = 1$	0.15	0.30



Here: $-0.35 \leqslant \mathrm{ACE}(X \to Y) \leqslant 0.65$.

Checking ACE bounds



This confirms the ACE bounds given earlier.

(But why is this helpful!?)

Combining two Obs Studies: Cholestyramine data

Z: assignment to treatment or control arm (randomized);
X: whether patient takes (more than certain amount of) drug;
Y: patient's health outcome.

Z	X	Y	count
0	0	0	158
0	0	1	14
0	1	0	0
0	1	1	0
1	0	0	52
1	0	1	12
1	1	0	23
1	1	1	78

(Data originally considered by Efron and Feldman (1991); dichotomized by Pearl.)

Combining two Obs Studies: Cholestyramine data

Z: assignment to treatment or control arm (randomized);
X: whether patient takes (more than certain amount of) drug;
Y: patient's health outcome.

Z	X	Y	count
0	0	0	158
0	0	1	14
0	1	0	0
0	1	1	0
1	0	0	52
1	0	1	12
1	1	0	23
1	1	1	78

(Data originally considered by Efron and Feldman (1991); dichotomized by Pearl.)

We wish to find $\mathrm{ACE}(X \to Y)$. Note $Z = 0 \;\Rightarrow\; X = 0$.

Combining two Obs Studies: Cholestyramine data

Z: assignment to treatment or control arm (randomized);
X: whether patient takes (more than certain amount of) drug;
Y: patient's health outcome.

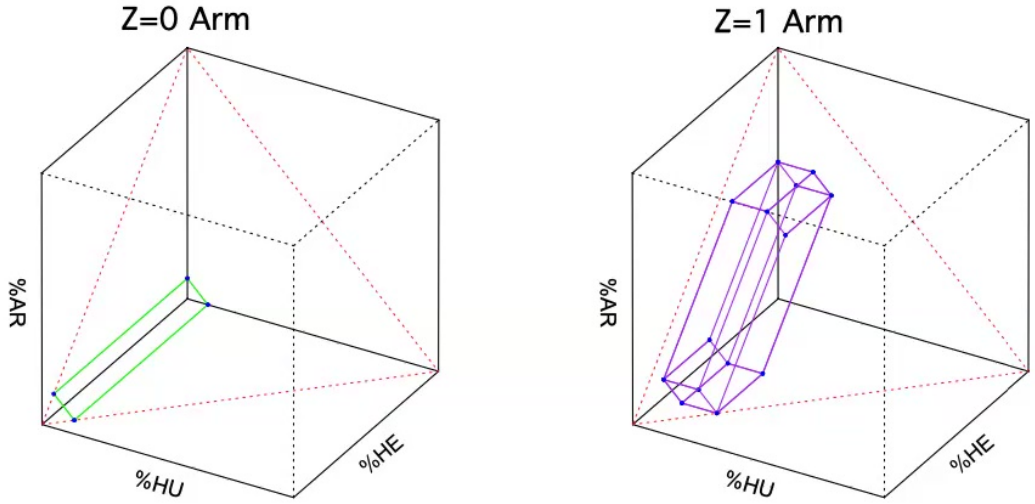
Z	X	Y	count
0	0	0	158
0	0	1	14
0	1	0	0
0	1	1	0
1	0	0	52
1	0	1	12
1	1	0	23
1	1	1	78

(Data originally considered by Efron and Feldman (1991); dichotomized by Pearl.)

We wish to find ACE$(X \to Y)$. Note $Z = 0 \Rightarrow X = 0$.

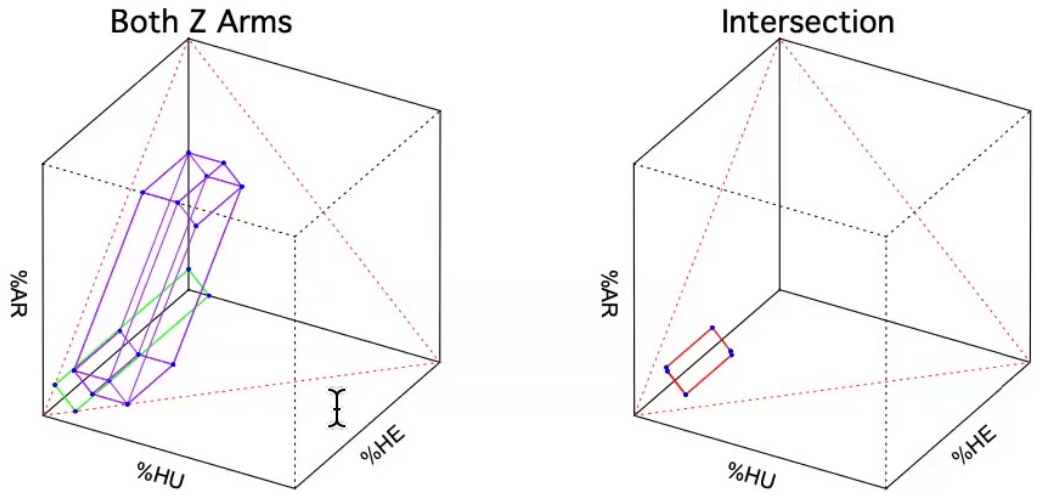
Idea: Analyze each Z arm as an observational study.

Each Z Arm

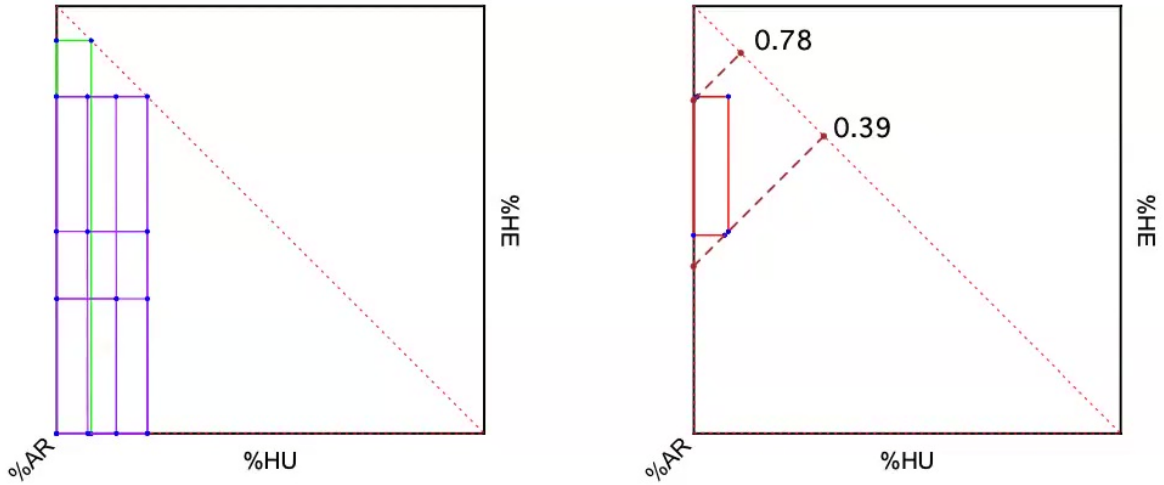


$Z = 0$ arm polytope is 2-d since $Z = 0 \Rightarrow X = 0$

Combining the Arms



Obtaining ACE bounds



Upper bound is: 0.78; lower bound is 0.39

Note: ACE bounds from each arm contain 0, but not when combined.

Why?

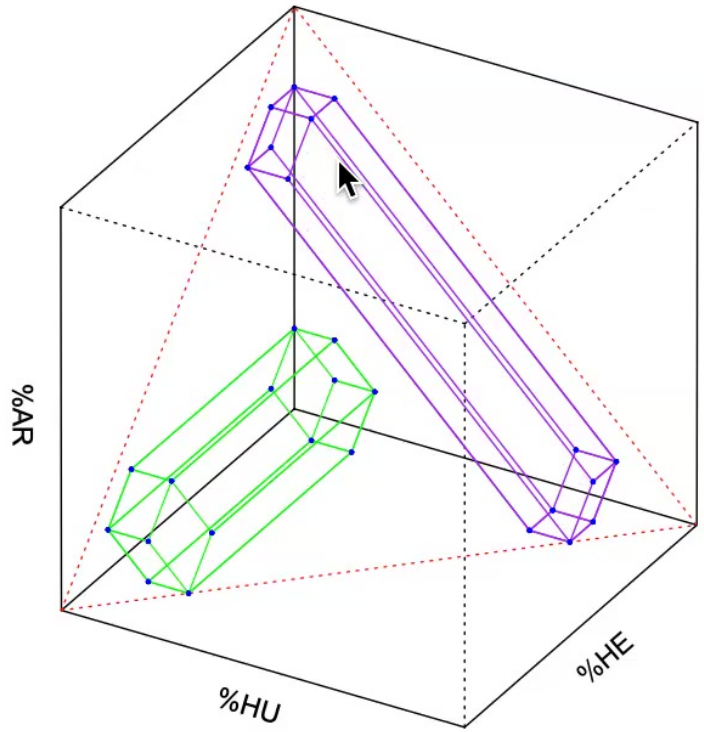
Assumptions:

- We assumed that which Z arm you are in does not affect your outcome Y except via X;
  In other words, Z has no direct effect on Y except through X:

$$Y(x, z_0) = Y(x, z_1) \equiv Y(x)$$

- Also assumed that Z is randomized: $Z \perp\!\!\!\perp Y(x_0), Y(x_1)$.
  (Aside) There are other ways to formulate this assumption in terms of a hidden variable. The bounds here can be shown to be algebraically equivalent to the CHSH inequalities.

Polytopes may not intersect



$\Rightarrow$ Model places testable constraints on $P(X, Y \mid Z)$.

Model for observables

For Z binary requiring that the polytopes intersect leads to the following:
If $p(X, Y \mid Z)$ is compatible with the binary IV model iff

$$\begin{aligned}
p(Y{=}0, X{=}0 \mid Z{=}0) + p(Y{=}1, X{=}0 \mid Z{=}1) &\leqslant 1, \\
p(Y{=}0, X{=}0 \mid Z{=}1) + p(Y{=}1, X{=}0 \mid Z{=}0) &\leqslant 1, \\
p(Y{=}0, X{=}1 \mid Z{=}0) + p(Y{=}1, X{=}1 \mid Z{=}1) &\leqslant 1, \\
p(Y{=}0, X{=}1 \mid Z{=}1) + p(Y{=}1, X{=}1 \mid Z{=}0) &\leqslant 1,
\end{aligned}$$

This describes a subset of $\Delta^3 \times \Delta^3$.

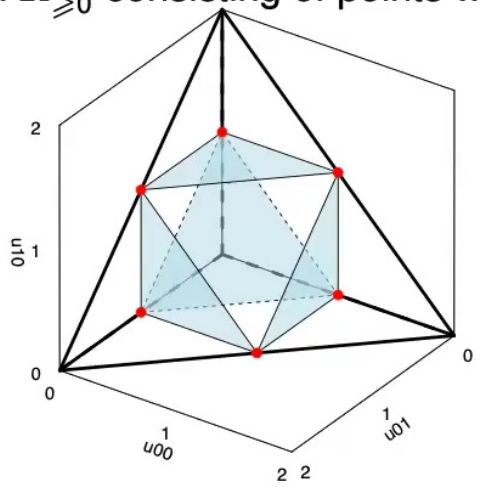
These are the IV inequalities of Pearl (1995) and Bonet (2001); they provide a *falsification* test of the binary IV model.

Visualizing the restrictions

Define the following variables:

$$
\begin{aligned}
u_{00} &\equiv p(Y=0, X=0 \mid Z=0) + p(Y=1, X=0 \mid Z=1) \leqslant 1, \\
u_{01} &\equiv p(Y=0, X=0 \mid Z=1) + p(Y=1, X=0 \mid Z=0) \leqslant 1, \\
u_{10} &\equiv p(Y=0, X=1 \mid Z=0) + p(Y=1, X=1 \mid Z=1) \leqslant 1, \\
u_{11} &\equiv p(Y=0, X=1 \mid Z=1) + p(Y=1, X=1 \mid Z=0) \leqslant 1,
\end{aligned}
$$

Since $u_{00} + u_{01} + u_{10} + u_{11} = 2$ these variables
live in a 3-d simplex of $\mathbb{R}^4_{\geqslant 0}$ consisting of points with sum $= 2$.



It follows that at most one inequality can be violated (see also Cai, Kuroki, Pearl, Tian, 2008).

Adjusting for covariates

Suppose that treatment X is assigned randomly given a covariate L

$$X \perp\!\!\!\perp Y(\tilde{x}) \mid L.$$

(sometimes called 'conditional ignorability')

Then

$$\begin{aligned} P[Y(\tilde{x}) = y \mid L = l] &= P[Y(\tilde{x}) = y \mid L = l, X = \tilde{x}] \quad \text{by indep.} \\ &= P[Y = y \mid L = l, X = \tilde{x}] \end{aligned}$$

Hence:

$$\begin{aligned} P[Y(\tilde{x}) = y] &= \sum_l P[Y(\tilde{x}) = y \mid L = l] P(L = l) \\ &= \sum_l P[Y = y \mid L = l, X = \tilde{x}] P(L = l) \end{aligned}$$

(called the 'backdoor formula' or 'standardization').

Potential Outcome Framework: Main points

- Postulates a joint distribution over outcomes in experimental and observational settings;
- (Typically) experimental outcomes are 'primary', of which observational outcomes are deterministic functions.
- \+ Rich language allowing many quantities of interest to be formulated, e.g. ETT, Natural Direct Effect,
- \+ 'Reduces' Causation to Missing Data; all outcomes 'observable' *a priori*;
- \+ Allows precise characterization of identification assumptions as conditional independence;
- \- Does not provide qualitative guidance as to when assumptions hold;
- \- Reasoning abstractly about multivariate conditional independence can be hard;
- \- Joint distribution over potential outcomes is not identified *even* from randomized experiments;
- -? Potential outcomes "do not exist" (McCullagh).

Structural Equation Models

Non-Parametric Structural Equation Models (NPSEM)

Originates in Econometrics: Haavelmo (1943), Strotz&Wold (1960)

System of equations describing the observational world:

One equation for each variable V expressing V as a function $f_V(\cdot, \cdot)$ of its direct causes and a 'disturbance' term ε_V.

Simple scenario with covariate L, treatment X and outcome Y:

$$
\begin{aligned}
L &= f_L(\varepsilon_L) \\
X &= f_X(L, \varepsilon_X) \\
Y &= f_Y(L, X, \varepsilon_Y)
\end{aligned}
$$

In general: distribution over errors induces a distribution over observed variables recursively via structural equations.
A Non-Parametric Structural Equation Model with Independent Errors (NPSEM-IE) aka Structural Causal Model (SCM) also assumes error terms are mutually independent.

Here $\varepsilon_L \perp\!\!\!\perp \varepsilon_X \perp\!\!\!\perp \varepsilon_Y$.

Experimental World derived from Observational (I)

An experimental world is then derived by removing the equation for the variable that is being fixed:

Example 1: fixing X to 0:

Obs.		Exp.
$L = f_L(\varepsilon_L)$		$L = f_L(\varepsilon_L)$
$X = f_X(L, \varepsilon_X)$	$\Rightarrow$	$X = 0$
$Y = f_Y(L, X, \varepsilon_Y)$		$Y = f_Y(L, X, \varepsilon_Y)$

Example 2: fixing L to 0:

Obs.		Exp.
$L = f_L(\varepsilon_L)$		$L = 0$
$X = f_X(L, \varepsilon_X)$	$\Rightarrow$	$X = f_X(L, \varepsilon_X)$
$Y = f_Y(L, X, \varepsilon_Y)$		$Y = f_Y(L, X, \varepsilon_Y)$

Note: this breaks the first rule of algebra!

Summary: Structural Equation Approach

- Specifies a data-generating process – with autonomous 'mechanisms' – for the observational distribution;
- Individual outcomes under intervention derived by removing equations

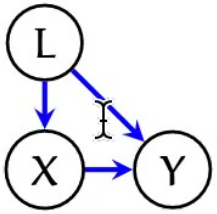
+ Intuitive specification of a generating process, encodes qualitative understanding
+ Guidance as to when assumptions will hold;

- Observational setting is primary: problematic since many examples where measurement of X is well-defined, but intervention or assignment of X is not.
- Typically assumed that interventions on *all* variables are well-defined;
- Error terms are not observable (even *a priori*);
- Assumption of independent errors is strong (more later);
- Implicitly specifies a joint distribution over actual and potential outcomes, but without notation to express distinction.

Graphical Approach

Causation via Graphs (I)

Approach due to Spirtes *et al.* (1993), Pearl (1995), relates to Sewall Wright's Path Diagrams.
Causal system represented by a directed acyclic graph (DAG).
Observational distribution factorizes according to this graph:

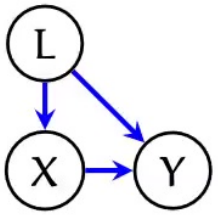


$$P(L, X, Y)$$
$$= P(L)P(X|L)P(Y|X, L)$$

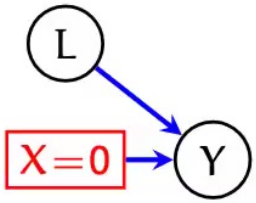
(If the DAG has missing edges) Pearl's d-separation criterion may be applied to read off conditional independence implied by the factorization.

Causation via Graphs (II)

Approach due to Spirtes *et al.* (1993), Pearl (1995), relates to Sewall Wright's Path Diagrams.
Causal system represented by a directed acyclic graph (DAG).
Observational distribution factorizes according to this graph:



$$P(L, X, Y)$$
$$= P(L)P(X \mid L)P(Y \mid X, L)$$



$$P(L, Y \mid do(X=0))$$
$$= P(L)P(Y \mid X=0, L)$$

Experimental world is obtained from Observational by removing edges into X and the term for X in the factorization.

Summary: Causal Graphical Approach

- Specifies a data-generating process – with autonomous 'mechanisms' – for the observational distribution;
- Intervention distributions derived by removing edges and factors
- \+ Intuitive specification of a generating process, encodes qualitative understanding
- \+ Guidance as to when assumptions will hold;
- \+? No joint distribution over experimental and observational settings; don't require counterfactuals to 'exist';
- \- Observational setting is primary: problematic since many examples where measurement of X is well-defined, but intervention or assignment of X is not.
- \- Typically assumed that interventions on *all* variables are well-defined;
- \- Without consistency, why should we care if $p(y|x) = p(y|do(x))$?;
- \- Without counterfactuals: no way to describe Effect of Treatment on the Treated, etc.,

How do these relate?

Are they the same?

(Acyclic) NPSEM-IE $\Rightarrow$ Causal DAG

Given an acyclic NPSEM-IE (with independent errors), can construct a DAG by adding an edge $A \rightarrow B$ if A is an arg. in f_B the function for B.

$\Rightarrow$ Observational distribution from equation system factors according to the orginal DAG

$\Rightarrow$ Distribution of remaining variables from system after removing equations factor according to the DAG with edges removed.

Notational Difference: NPSEMs and Potential Outcomes

- Counterfactual Approach: Key Distinction between:
  - Y: the outcome in the observational world;
  - $Y(x)$: the outcome in the experimental world.
- Structural Equations and Graphical Approach:

  The same variable Y is used for both;

  The context of the graph or equation system is used to make the distinction.

Potential Outcomes vs. NPSEMs

Potential outcome models:

Postulate potential outcomes; derive observed variables (via consistency).

Non-Parametric Structural Equation Models (NPSEMs):

Postulate a model for the observables; derive counterfactuals (via removing equations).

$\Rightarrow$ to naive users NPSEMs can appear to require a smaller ontological commitment.

This is an illusion: in fact, the commitments in the potential outcome model will be fewer if not all variables can be intervened on.

NPSEM $\Rightarrow$ Counterfactual Model

Simple fix:
The structural equation for V can be written as giving the potential outcome from the experimental world where all inputs (aka parents) are fixed:

$$\begin{array}{lcl} L = f_L(\varepsilon_L) & & L = f_L(\varepsilon_L) \\ X = f_X(L, \varepsilon_X) & \Rightarrow & X(l) = f_X(l, \varepsilon_X) \\ Y = f_Y(L, X, \varepsilon_Y) & & Y(l, x) = f_Y(l, x, \varepsilon_Y) \end{array}$$

observed variables are given by: $X = X(L)$, $Y = Y(L, X(L))$.

NPSEM $\Rightarrow$ Counterfactual Model

Simple fix:
The structural equation for V can be written as giving the potential outcome from the experimental world where all inputs (aka parents) are fixed:

$$\begin{array}{lcl} L = f_L(\varepsilon_L) & & L = f_L(\varepsilon_L) \\ X = f_X(L, \varepsilon_X) & \Rightarrow & X(l) = f_X(l, \varepsilon_X) \\ Y = f_Y(L, X, \varepsilon_Y) & & Y(l, x) = f_Y(l, x, \varepsilon_Y) \end{array}$$

observed variables are given by: $X = X(L)$, $Y = Y(L, X(L))$.

Writing as counterfactuals make clear equations represent relationships that are invariant under interventions on other variables:

intervening to set L *and* X *to* 0*, the value for* Y *will be:* $f_Y(0, 0, \varepsilon_Y)$.

Counterfactual Model $\Rightarrow$ NPSEM

If we are given "one step ahead potential outcomes" giving the outcome under an intervention on the inputs (aka parents) of a variable structural equations and error terms are easy to construct:

One-step ahead potential outcomes: $X; \quad M(x); \quad Y(x, m)$.

$$\begin{aligned} \varepsilon_X &= X \\ \varepsilon_M &= (M(x_0), M(x_1)) \\ \varepsilon_Y &= (Y(x_0, m_0), Y(x_0, m_1) \\ &\qquad Y(x_1, m_0), Y(x_1, m_1)) \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} X &= f_X(\varepsilon_X) \\ M(x) &= f_M(x, \varepsilon_M) \\ Y(x, m) &= f_Y(x, m, \varepsilon_Y), \end{aligned}$$

Error term ε_V corresponds to *set* of one-step ahead potential outcomes for a variable: $\{V(\mathbf{pa_V}) \mid \mathbf{pa_V} \in \mathfrak{X}_{\mathbf{pa_V}}\}$

Function f_V is a simple co-ordinate projection which selects the appropriate element from ε_V, according to the values taken by $\mathbf{pa_V}$.

Example: $f_M(x = 1, \varepsilon_M) = f_M(x = 1, (M(x_0), M(x_1))) = M(x_1)$.

NPSEM-IE $\Rightarrow$ (Untestable) Counterfactual Model

The assumption that error terms are independent becomes:

$$\mathbf{L} \perp\!\!\!\perp \{\mathbf{X}(\mathbf{l}); \mathbf{l}\} \perp\!\!\!\perp \{\mathbf{Y}(\mathbf{l}, \mathbf{x}); \mathbf{l}, \mathbf{x}\}.$$

Note that here we are assuming that *sets* of counterfactual random variables are independent.

Parts of this assumption are not testable via any randomized experiment on the variables in the system.
Further, these assumptions lead to additional identification results.

Pearl: "DAGs and Potential Outcomes are equivalent theories".

Important caveats:

- NPSEMs typically assume all variables are seen as being subject to well-defined interventions
- Users of structural equations tend to worry less about whether an intervention is well-defined.

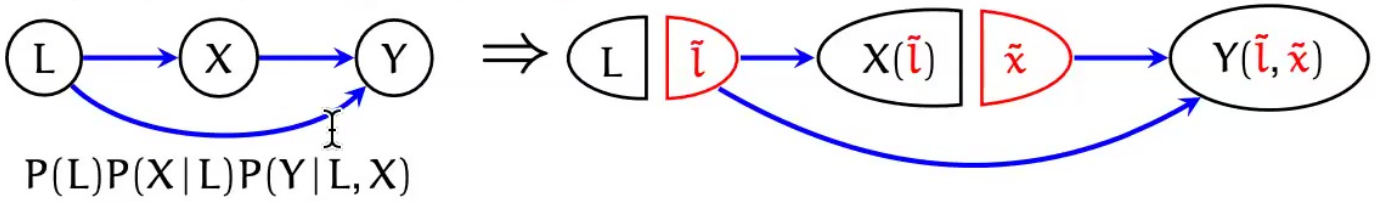
  Ex. If the variable $\mathcal{M}$ is your response to a question, how to intervene on $\mathcal{M}$?!
- Pearl's approach to unifying graphs and counterfactuals typically associates with a DAG the counterfactual model corresponding to an NPSEMs with Independent Errors (NPSEM-IEs) with DAGs.

  This assumption is not empirically testable via randomized experiments.

  More on this below.

Single World Intervention Graphs (SWIGs) R+Robins (2013)

Graphical Representation, fixing both l and x:



$$P(L)P(X \mid L)P(Y \mid L, X)$$

factorization:

$$P(L, X(\tilde{l}), Y(\tilde{l}, \tilde{x})) = P(L)P(X(\tilde{l}))P(Y(\tilde{l}, \tilde{x}))$$

'modularity':

$$P(X(\tilde{l}) = x) = P(X = x \mid L = \tilde{l}),$$
$$P(Y(\tilde{l}, \tilde{x}) = y) = P(Y = y \mid L = \tilde{l}, X = \tilde{x}).$$

we may apply d-separation (red nodes are always blocked):

$$L \perp\!\!\!\perp X(\tilde{l}) \perp\!\!\!\perp Y(\tilde{l}, \tilde{x})$$

How many additional independences in NPSEM-IE?

Assumption of independent errors implies super-exponentially many 'cross-world' counterfactual independence assumptions:

No. Obs. Vars.	2	3	4	K
Dim. $P(\mathbf{V})$	3	7	15	$2^K - 1$
No. Cnterfactual Vars.	3	7	15	$2^K - 1$
Dim. Cnterfactual Dist.	7	127	32767	$2^{(2^K-1)} - 1$
Dim. SWIG	5	113	32697	$(2^{(2^K-1)} - 1) - \sum_{j=1}^{K-1}(4^j - 2^j)$
Dim. NPSEM-IE	4	19	274	$\sum_{j=0}^{K-1}(2^{2^j} - 1)$
No. untestable indep. constrnts in NPSEM-IE	1	94	32423	$O(2^{2^K-2})$

Table: Dimensions of counterfactual models associated with complete graphs with binary variables.

Cross-world independences unnecessary for most purposes

For many purposes these extra 'cross-world' independences are irrelevant.
Specifically, the Independences arising from a SWIG imply all of the identification results that hold in the *do*-calculus of Pearl (1995); see also Spirtes *et al.* (1993):

But these extra independences do lead to additional identification results in the context of mediation and path-specific effects.

These additional identification results are not subject to experimental test even if randomized interventions on all variables are possible.

Eliminating a false trichotomy

Previously the main approach to unifying counterfactuals and graphs was via Non-Parametric Structural Equation Models with Independent Errors:

This gave causal modelers three options:

- Use graphs, and not counterfactuals (Dawid).
- Use counterfactuals, and not graphs (many Statisticians).
- Use both graphs and counterfactuals, but be forced to make 'a lot' of additional assumptions that are:
  - not experimentally testable (even in principle);
  - not necessary for most identification results.

  Require (as the default) all variables to be intervened upon.

Summary

- Potential outcomes represent the most general framework for reasoning about causality.
- An NPSEM is the special case of a counterfactual model in which we can intervene on every variable.
- An NPSEM-IE further assumes cross-world independence relations that are experimentally untestable and lead to novel identification results.

Summary

- Potential outcomes represent the most general framework for reasoning about causality.
- An NPSEM is the special case of a counterfactual model in which we can intervene on every variable.
- An NPSEM-IE further assumes cross-world independence relations that are experimentally untestable and lead to novel identification results.
- Graphs are a powerful, essential tool for reasoning about joint distributions.
- SWIGs provide a simple way to connect potential outcome models and graphs without the restrictions associated with NPSEM-IEs.

Thank You!