REVISED
1:19 pm, Oct 12, 2010

arbor - oct 2010

CAUSAL INFERENCE: MATHEMATICAL FOUNDATIONS AND METHODOLOGICAL PRINCIPLES

Judea Pearl
University of California, Los Angeles
(www.cs.ucla.edu/~judea/)

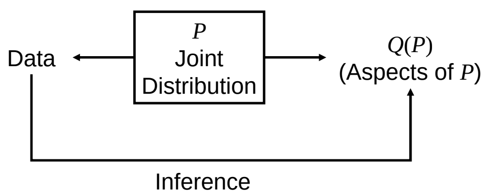
1

TOPICS TO BE COVERED

1. The three-level hierarchy of causal sentences
2. The logical equivalence of structural and counterfactual models
3. The five steps of causal inference methodology
   1. Define, 2. Assume, 3. identify, 4. Estimate, 5. Test
4. How to represent causal assumptions
5. Which covariates should be measured?
6. What predictors should enter propensity scores.
7. Why Yes to “Causation without manipulation"
8. The Mediation Formula and effect decomposition
9. Measurement bias and effect restoration.
10. Transportability across studies: "external validity."

2

TRADITIONAL STATISTICAL INFERENCE PARADIGM

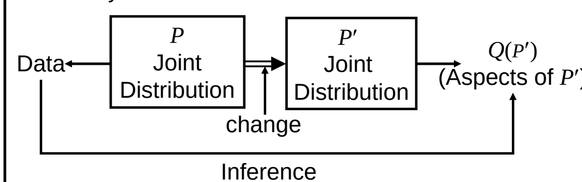


e.g.,
Infer whether customers who bought product A would also buy product B.
$Q = P(B \mid A)$

3

FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

Probability and statistics deal with static relations

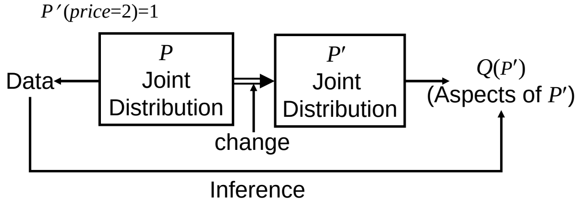


What happens when P changes?
e.g.,
Infer whether customers who bought product A would still buy A if we were to double the price.

4

FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

What remains invariant when P changes say, to satisfy $P'(price{=}2)=1$



Note: $P'(v) \neq P(v \mid price = 2)$

P does not tell us how it ought to change
Causal knowledge: what remains invariant

5

FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES (CONT)

1. Causal and statistical concepts do not mix.

CAUSAL	STATISTICAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
Confounding / Effect	“Controlling for” / Conditioning
Instrumental variable	Odd and risk ratios
Exogeneity / Ignorability	Collapsibility / Granger causality
Mediation	Propensity score

2.

3.

4.

6

FROM STATISTICAL TO CAUSAL ANALYSIS: 2. MENTAL BARRIERS

1. Causal and statistical concepts do not mix.

CAUSAL	STATISTICAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
Confounding / Effect	"Controlling for" / Conditioning
Instrumental variable	Odd and risk ratios
Exogeneity / Ignorability	Collapsibility / Granger causality
Mediation	Propensity score

2. No causes in – no causes out (Cartwright, 1989)

$$\left.\begin{array}{l}\text{statistical assumptions + data}\\ \text{causal assumptions}\end{array}\right\} \Rightarrow \text{causal conclusions}$$

3. Causal assumptions cannot be expressed in the mathematical language of standard statistics.
4.

7

FROM STATISTICAL TO CAUSAL ANALYSIS: 2. MENTAL BARRIERS

1. Causal and statistical concepts do not mix.

CAUSAL	STATISTICAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
Confounding / Effect	"Controlling for" / Conditioning
Instrumental variable	Odd and risk ratios
Exogeneity / Ignorability	Collapsibility / Granger causality
Mediation	Propensity score

2. No causes in – no causes out (Cartwright, 1989)

$$\left.\begin{array}{l}\text{statistical assumptions + data}\\ \text{causal assumptions}\end{array}\right\} \Rightarrow \text{causal conclusions}$$

3. Causal assumptions cannot be expressed in the mathematical language of standard statistics.
4. Non-standard mathematics:
   a) Structural equation models (Wright, 1920; Simon, 1960)
   b) Counterfactuals (Neyman-Rubin (Y_x), Lewis $(x \;\square\!\!\rightarrow Y)$)

8

3-LEVEL HIERARCHY OF CAUSAL MODELS

1. Probabilistic Knowledge $\boldsymbol{P(y \mid x)}$
   Bayesian networks, graphical models
2. Interventional Knowledge $\boldsymbol{P(y \mid do(x))}$
   Causal Bayesian Networks (CBN)
   (Manipulation graphs)
3. Counterfactual Knowledge $\boldsymbol{P(Y_x = y, Y_{x'} = y')}$
   Structural equation models, physics, functional graphs, regret and introspection

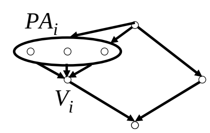
9

BAYESIAN NETWORKS

Definition

P is Markov relative to G iff:

$$V_i \perp\!\!\!\perp ND_i \mid PA_i$$



$\boxed{G}$

Algebraic Representation

$$P(v_i,...,v_n) = \prod_i P(v_i \mid pa_i)$$

Graphical Representation

$$(X \perp\!\!\!\perp Y \mid Z)_G \Rightarrow (X \perp\!\!\!\perp Y \mid Z)_P$$

(d-separation $\Rightarrow$ conditional independence)

10

CAUSAL BAYESIAN NETWORK (CBN)
(Manipulation graphs)

Definition

Let $P_x(v)$ stand for $P(v \mid do(X{=}x))$, X and x arbitrary
G is a CBN relative to P_x* iff:
each $P(v_i \mid pa_i)$ remains INVARIANT to interventions not involving V_i.

Algebraic representation

$$P_x(v) = \prod_{\{i \mid V_i \notin X\}} P(v_i \mid pa_i) \quad \text{for all } v \text{ consistent with } x.$$

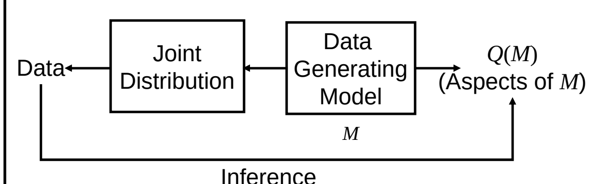
(Truncated product, G-computation, manipulation theorem)

Graphical representation

Surgery: remove incoming arrows to X.

11

THE STRUCTURAL MODEL PARADIGM



M – Invariant strategy (mechanism, recipe, law, protocol) by which Nature assigns values to variables in the analysis.

"Think Nature, not experiment!"

12

STRUCTURAL CAUSAL MODELS

Definition: A structural causal model is a 4-tuple $\langle V,U, F, P(u)\rangle$, where

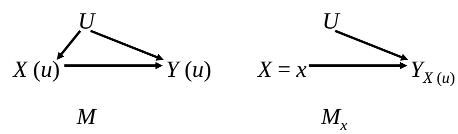
- $V = \{V_1,...,V_n\}$ are endogenous variables
- $U = \{U_1,...,U_m\}$ are background variables
- $F = \{f_1,..., f_n\}$ are functions determining V,
  $v_i = f_i(v, u)$ e.g., $y = \alpha + \beta x + u_Y$
- $P(u)$ is a distribution over U

$P(u)$ and F induce a distribution $P(v)$ over observable variables

13

CAUSAL MODELS AND COUNTERFACTUALS

Definition:
The sentence: "Y would be y (in unit u), had X been x," denoted $Y_x(u) = y$, means:
The solution for Y in a mutilated model M_x, (i.e., the equations for X replaced by $X = x$) with input $U=u$, is equal to y.



14

CAUSAL MODELS AND COUNTERFACTUALS

Definition:
The sentence: "Y would be y (in unit u), had X been x," denoted $Y_x(u) = y$, means:
The solution for Y in a mutilated model M_x, (i.e., the equations for X replaced by $X = x$) with input $U=u$, is equal to y.

The Fundamental Equation of Counterfactuals:

$$\boxed{Y_x(u) = Y_{M_x}(u)}$$

15

CAUSAL MODELS AND COUNTERFACTUALS

Definition:
The sentence: "Y would be y (in situation u), had X been x," denoted $Y_x(u) = y$, means:
The solution for Y in a mutilated model M_x, (i.e., the equations for X replaced by $X = x$) with input $U=u$, is equal to y.

- Joint probabilities of counterfactuals:

$$P(Y_x = y, Z_w = z) = \sum_{u: Y_x(u)=y, Z_w(u)=z} P(u)$$

In particular:

$$P(y \mid do(x)) \triangleq P(Y_x = y) = \sum_{u: Y_x(u)=y} P(u)$$

$$P(Y_{x'} = y' \mid x, y) = \sum_{u: Y_{x'}(u)=y'} P(u \mid x, y)$$

16

TWO PARADIGMS FOR CAUSAL INFERENCE

Observed: $P(X, Y, Z,...)$
Conclusions needed: $P(Y_x=y)$, $P(X_y=x \mid Z=z)$...

How do we connect observables, $X,Y,Z,\ldots$
to counterfactuals Y_x, X_z, Z_y,… ?

<u>*N-R* model</u>	<u>Structural model</u>
Counterfactuals are primitives, new variables	Counterfactuals are derived quantities
Super-distribution $P^*(X,Y,...,Y_x,X_z,...)$	Subscripts modify the model and distribution
X,Y,Z constrain Y_x, Z_y,...	$P(Y_x = y) = P_{M_x}(Y = y)$

17

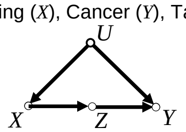
ARE THE TWO PARADIGMS EQUIVALENT?

- Yes (Galles and Pearl, 1998; Halpern 1998)
- In the *N-R* paradigm, Y_x is defined by consistency:
  $$Y = xY_1 + (1-x)Y_0$$
- In SCM, consistency is a theorem.
- Moreover, a theorem in one approach is a theorem in the other.
- Difference: Clarity of assumptions and their implications

18

FORMULATING ASSUMPTIONS THREE LANGUAGES

1. English: Smoking (X), Cancer (Y), Tar (Z), Genotypes (U)

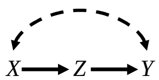


2. Counterfactuals: $Z_x(u) = Z_{yx}(u)$,

$X_y(u) = X_{zy}(u) = X_z(u) = X(u)$,

$Y_z(u) = Y_{zx}(u), Z_x \perp\!\!\!\perp \{Y_z, X\}$

3. Structural:



19

GRAPHICAL – COUNTERFACTUALS SYMBIOSIS

Every causal graph expresses counterfactuals assumptions, e.g., $X \to Y \to Z$

1. Missing arrows $Y \not\leftarrow Z$ $\quad Y_{x,z}(u) = Y_x(u)$
2. Missing arcs $Y \quad Z$ $\quad Y_x \perp\!\!\!\perp Z_y$

consistent, and readable from the graph.

- Express assumption in graphs
- Derive estimands by graphical or algebraic methods

20

THE FIVE NECESSARY STEPS OF CAUSAL ANALYSIS

Define: Express the target quantity Q as a function $Q(M)$ that can be computed from any model M.

Assume: Formulate causal assumptions A using some formal language.

Identify: Determine if Q is identifiable given A.

Estimate: Estimate Q if it is identifiable; approximate it, if it is not.

Test: Test the testable implications of A (if any).

21

THE FIVE NECESSARY STEPS FOR EFFECT ESTIMATION

Define: Express the target quantity Q as a function $Q(M)$ that can be computed from any model M.

$$\text{ATE} \triangleq E(Y \mid do(x_1)) - E(Y \mid do(x_0))$$

Assume: Formulate causal assumptions A using some formal language.

Identify: Determine if Q is identifiable given A.

Estimate: Estimate Q if it is identifiable; approximate it, if it is not.

Test: Test the testable implications of A (if any).

22

THE FIVE NECESSARY STEPS FOR EFFECT OF TREATMENT ON THE TREATED

Define: Express the target quantity Q as a function $Q(M)$ that can be computed from any model M.

$$\text{ETT} \triangleq P(Y_x = y \mid X = x')$$

Assume: Formulate causal assumptions A using some formal language.

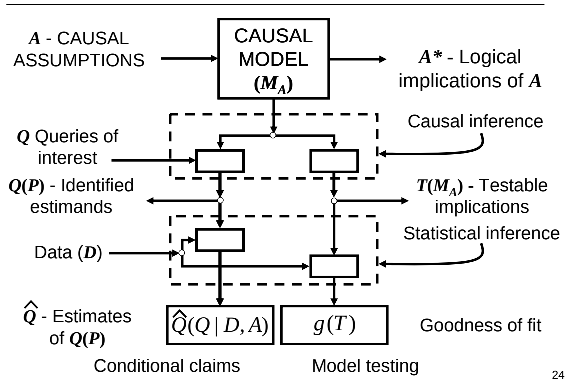
Identify: Determine if Q is identifiable given A.

Estimate: Estimate Q if it is identifiable; approximate it, if it is not.

Test: Test the testable implications of A (if any).

23

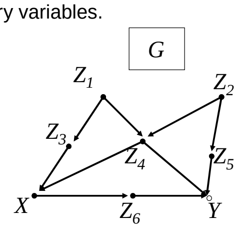
THE LOGIC OF SEM



24

IDENTIFICATION IN SCM

Find the effect of X on Y, $P(y|do(x))$, given the causal assumptions shown in G, where $Z_1,\ldots, Z_k$ are auxiliary variables.

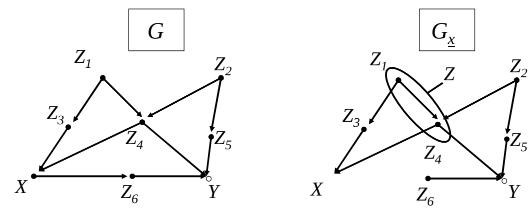


Can $P(y|do(x))$ be estimated if only a subset, Z, can be measured?

25

ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

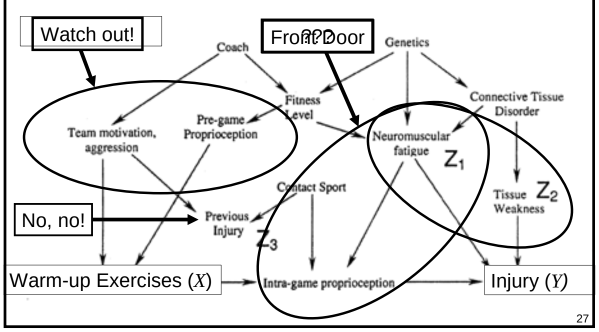
$P(y \mid do(x))$ is estimable if there is a set Z of variables such that Z d-separates X from Y in $G_{\underline{x}}$.



- Moreover, $P(y \mid do(x)) = \sum_z P(y \mid x, z) P(z) = \sum_z \frac{P(x,y,z)}{P(x \mid z)}$ ("adjusting" for Z) $\rightarrow$ Ignorability

26

EFFECT OF WARM-UP ON INJURY (After Shrier & Platt, 2008)



27

FROM IDENTIFICATION TO ESTIMATION

Define: Express the target quantity Q as a function $Q(M)$ that can be computed from any model M.

$$Q = P(y \mid do(x))$$

Assume: Formulate causal assumptions using ordinary scientific language and represent their structural part in graphical form.

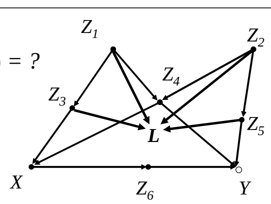
Identify: Determine if Q is identifiable.

Estimate: Estimate Q if it is identifiable; approximate it, if it is not.

28

PROPENSITY SCORE ESTIMATOR (Rosenbaum & Rubin, 1983)

$P(y \mid do(x)) = ?$



$$L(z_1, z_2, z_3, z_4, z_5) \triangleq P(X = 1 \mid z_1, z_2, z_3, z_4, z_5)$$

Theorem: $\sum_z P(y \mid z, x) P(z) = \sum_l P(y \mid L = l, x) P(L = l)$

Adjustment for L replaces Adjustment for Z

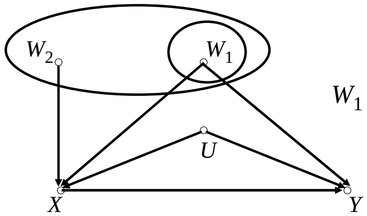
29

WHAT PROPENSITY SCORE (PS) PRACTITIONERS NEED TO KNOW

1. The asymptotic bias of PS is EQUAL to that of ordinary adjustment (for same Z).
2. Including an additional covariate in the analysis CAN SPOIL the bias-reduction potential of others.
3. In particular, instrumental variables tend to amplify bias.
4. Choosing sufficient set for PS, requires knowledge of the model.

30

BIAS AMPLIFICATION
BY INSTRUMENTAL VARIABLES



$$W_1 \nsim \{W_{1,} W_2\}$$

- Adding W_2 to Propensity Score increases bias (if such exists) (Wooldridge, 2009)
- In linear systems – always
- In non-linear systems – almost always (Pearl, 2010)
- Outcome predictors are safer than treatment predictors

31

EFFECT DECOMPOSITION
(direct vs. indirect effects)

1. Why decompose effects?
2. What is the definition of direct and indirect effects?
3. What are the policy implications of direct and indirect effects?
4. When can direct and indirect effect be estimated consistently from experimental and nonexperimental data?

32

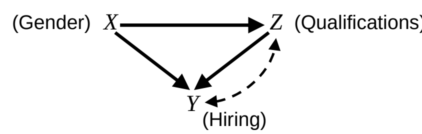
WHY DECOMPOSE EFFECTS?

1. To understand how Nature works
2. To comply with legal requirements
3. To predict the effects of new type of interventions:
   Signal routing, rather than variable fixing

33

LEGAL IMPLICATIONS
OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



What is the direct effect of X on Y ?

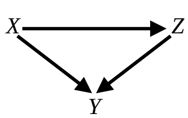
$$E(Y \mid do(x_1), do(z)) - E(Y \mid do(x_0), do(z))$$

(averaged over z) Adjust for Z? No! No!

34

NATURAL INTERPRETATION OF
AVERAGE DIRECT EFFECTS

Robins and Greenland (1992) – "Pure"



$z = f(x, u)$
$y = g(x, z, u)$

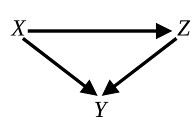
Natural Direct Effect of X on Y: $DE(x_0, x_1; Y)$
The expected change in Y, when we change X from x_0 to x_1 and, for each u, we keep Z constant at whatever value it attained before the change.

$$E[Y_{x_1 Z_{x_0}} - Y_{x_0}]$$

In linear models, *DE* = Controlled Direct Effect

35

DEFINITION OF
INDIRECT EFFECTS



$z = f(x, u)$
$y = g(x, z, u)$

Indirect Effect of X on Y: $IE(x_0, x_1; Y)$
The expected change in Y when we keep X constant, say at x_0, and let Z change to whatever value it would have attained had X changed to x_1.

$$E[Y_{x_0 Z_{x_1}} - Y_{x_0}]$$

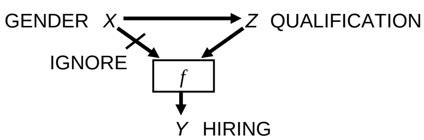
In linear models, *IE = TE - DE*

36

POLICY IMPLICATIONS OF INDIRECT EFFECTS

What is the indirect effect of X on Y?

The effect of Gender on Hiring if sex discrimination is eliminated.



Deactivating a link – a new type of intervention

37

MEDIATION FORMULAS

1. The natural direct and indirect effects are identifiable in Markovian models (no confounding),
2. And are given by:

$$DE = \sum_z [E(Y \mid do(x_1, z)) - E(Y \mid do(x_0, z))] P(z \mid do(x_0)).$$

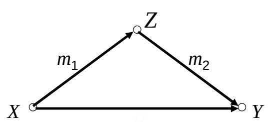
$$IE = \sum_z E(Y \mid do(x_0, z))[P(z \mid do(x_1)) - P(z \mid do(x_0))]$$

$$TE = DE - EI(rev)$$

3. Applicable to linear and non-linear models, continuous and discrete variables, regardless of distributional form.

38

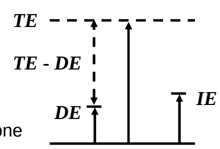
WHY $TE \neq DE + IE$



$TE = DE - IE(rev)$

In linear systems

$$TE = \beta + m_1 m_2$$
$$DE = \beta$$
$$IE = m_1 m_2 = TE - DE$$



$IE =$ Effect sustained by mediation alone

Is NOT equal to:

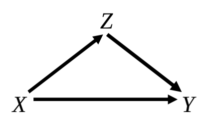
$TE - DE =$ Effect prevented by disabling mediation

Disabling mediation

Disabling direct path

39

MEDIATION FORMULAS IN UNCONFOUNDED MODELS



$$DE = \sum_z [E(Y \mid x_1, z) - E(Y \mid x_0, z)] P(z \mid x_0)$$

$$IE = \sum_z [E(Y \mid x_0, z)[P(z \mid x_1) - P(z \mid x_0)]$$

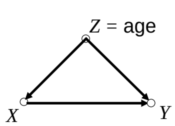
$$TE = E(Y \mid x_1) - E(Y \mid x_0)$$

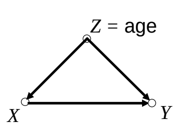
$IE =$ Fraction of responses explained by mediation

$TE - DE =$ Fraction of responses owed to mediation

40

TRANSPORTABILITY -- WHEN CAN WE EXTRPOLATE EXPERIMENTAL FINDINGS TO DIFFERENT POPULATIONS?





Experimental study in LA
Measured: $P(x, y, z)$
$P(y \mid do(x), z)$

Observational study in NYC
Measured:

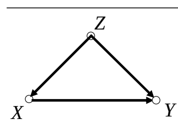
Problem: We find $P(z) \neq P^*(z)$
(LA population is younger)

What can we say about $P^*(y \mid do(x))$

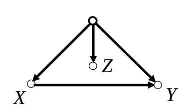
Intuition: $P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$

41

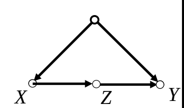
TRANSPORT FORMULAS DEPEND ON THE STORY



(a)



(b)



(c)

a) ***Z*** represents age
$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$

b) ***Z*** represents language skill
$P^*(y \mid do(x)) =$???

c) ***Z*** represents a bio-marker
$P^*(y \mid do(x)) =$???

42

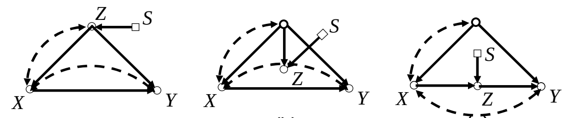
TRANSPORTABILITY
(Pearl and Bareinboim, 2010)

Definition 1 (Transportability)
Given two populations, denoted Π and Π^*, characterized by probability distributions $\boldsymbol{P}$ and $\boldsymbol{P}^*$, and causal diagrams $\boldsymbol{G}$ and $\boldsymbol{G}^*$, respectively, a causal relation $\boldsymbol{R}$ is said to be transportable from Π to Π^* if

1. $\boldsymbol{R}(\Pi)$ is estimable from the set $\boldsymbol{I}$ of interventional studies on Π, and
2. $\boldsymbol{R}(\Pi^*)$ is identified from $\boldsymbol{I}$, $\boldsymbol{P}$, $\boldsymbol{P}^*$, $\boldsymbol{G}$, and $\boldsymbol{G}^*$.

43

TRANSPORT FORMULAS DEPEND ON THE STORY

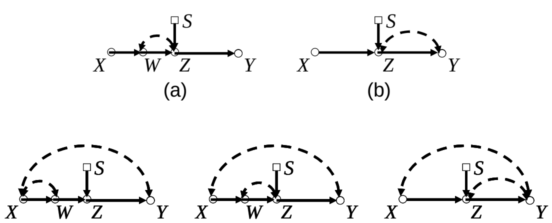


(a) (b) (c)

a) $\boldsymbol{Z}$ represents age
$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$
b) $\boldsymbol{Z}$ represents language skill
$$P^*(y \mid do(x)) = P(y \mid do(x))$$
c) $\boldsymbol{Z}$ represents a bio-marker
$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z \mid x)$$

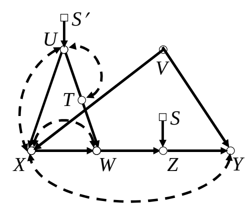
44

WHICH MODEL LICENSES THE TRANSPORT OF THE CAUSAL EFFECT



45

DETERMINE IF THE CAUSAL EFFECT IS TRANSPORTABLE



What measurements need to be taken in the study and in the target population?

The transport formula
$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) \sum_w P^*(z \mid w) \sum_t P(w \mid do(x), t) P^*(t)$$

46

CONCLUSIONS

I TOLD YOU CAUSALITY IS SIMPLE

- Formal basis for causal and counterfactual inference (complete)
- Unification of the graphical, potential-outcome and structural equation approaches
- Friendly and formal solutions to century-old problems and confusions.
- No other method can do better (theorem)

47

CONCLUSIONS

He is wise who bases causal inference on an explicit causal structure that is defensible on scientific grounds.

(Aristotle 384-322 B.C.)

48

QUESTIONS???

They will be answered

49

PHYSICS DESERVES A NEW ALGEBRA

Scientific Equations (e.g., Hooke's Law) are non-algebraic

e.g., Length (Y) equals a constant (2) times the weight (X)

Correct notation:

Process information	The solution
$Y := 2X$	$X = 1$
$X = 1$	$Y = 2$

Had X been 3, Y would be 6.
If we raise X to 3, Y would be 6.
Must "wipe out" $X = 1$.

50

REGRESSION VS. STRUCTURAL EQUATIONS (THE CONFUSION OF THE CENTURY)

Regression (claimless, nonfalsifiable):

$$Y = ax + \varepsilon_Y$$

Structural (empirical, falsifiable):

$$Y = bx + u_Y \qquad (\ := \text{assignment})$$

Claim: (regardless of distributions):

$$E(Y \mid do(x)) = E(Y \mid do(x), do(z)) = bx$$

The mothers of all questions:

Q. When would $\boldsymbol{b}$ equal $\boldsymbol{a}$?
A. When ($\boldsymbol{u_Y} \perp\!\!\!\perp \boldsymbol{X}$), read from the diagram

Q. When is $\boldsymbol{b}$ a partial regression? $\boldsymbol{b} = \beta_{YX \cdot Z}$
A. Shown in the diagram, Slide 36.

51

AXIOMS OF STRUCTURAL COUNTERFACTUALS

$Y_x(u)=y$: Y would be y, had X been x (in state $U = u$) (Galles, Pearl, Halpern, 1998):

1. Definiteness
   $$\exists x \in X \ s.t. \ X_y(u) = x$$
2. Uniqueness
   $$(X_y(u) = x) \,\&\, (X_y(u) = x') \Rightarrow x = x'$$
3. Effectiveness
   $$X_{xw}(u) = x$$
4. Composition (generalized consistency)
   $$X_w(u) = x \Rightarrow Y_{wx}(u) = Y_w(u)$$
5. Reversibility
   $$(Y_{xw}(u) = y) \,\&\, (W_{xy}(u) = w) \Rightarrow Y_x(u) = y$$

52