CAUSAL INFERENCE: LOGICAL FOUNDATION AND NEW RESULTS

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

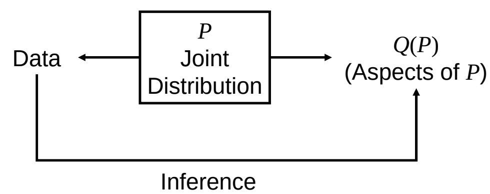
OUTLINE

- Statistical vs. Causal vs. Counterfactual inference: syntax and semantics
- The logical equivalence of Structural Equation Models (SEM) and potential outcomes (Neyman-Rubin)
- Where do they differ?
  1. Define, 2. Assume, 3. identify, 4. Estimate, 5. Test
- Which covariates should be measured?
- Causation without manipulation
- The Mediation Formula and its ramifications
- Measurement bias and effect restoration.

3-LEVEL HIERARCHY OF CAUSAL MODELS

1. Probabilistic Knowledge $P(y \mid x)$
   Bayesian networks, graphical models
2. Interventional Knowledge $P(y \mid do(x))$
   Causal Bayesian Networks (CBN)
   (Agnostic graphs, manipulation graphs)
3. Counterfactual Knowledge $P(Y_x = y, Y_{x'} = y')$
   Structural equation models, physics, functional graphs, “Treatment assignment mechanism”

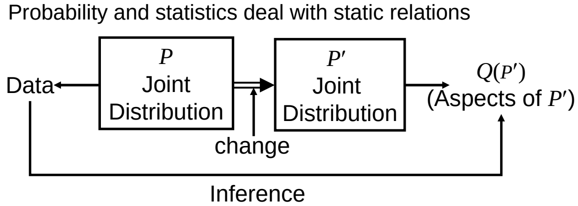
TRADITIONAL STATISTICAL INFERENCE PARADIGM



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

FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

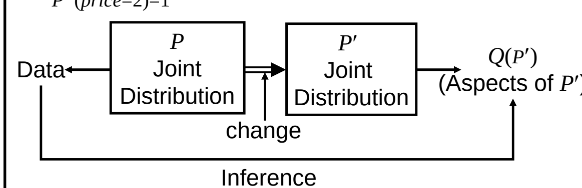
Probability and statistics deal with static relations



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.

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

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.

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)

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

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

4.

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)

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

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

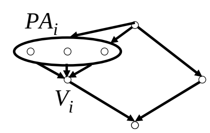
4. Non-standard mathematics:
   a) Structural equation models (Wright, 1920; Simon, 1960)
   b) Counterfactuals (Neyman-Rubin (Y_x), Lewis $(x \;\square\!\!\rightarrow Y)$)

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)

CAUSAL BAYESIAN NETWORK (CBN)
(Agnostic graphs, 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:

i. $P_x(v)$ is Markov relative to G;

ii. $P_x(v_i)$=1 for all $V_i \in X$ if v_i agrees with $X{=}x$.

iii. $P_x(v_i \mid pa_i) = P(v_i \mid pa_i)$ for all $V_i \notin X$ whenever pa_i agrees with $X{=}x$, i.e., each $P(v_i \mid pa_i)$ remains INVARIANT to interventions not involving V_i.

Algebraic representation

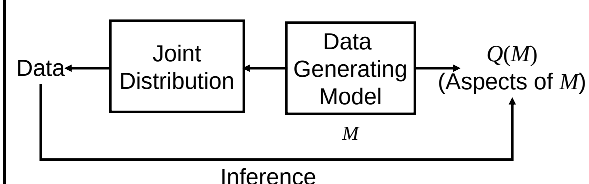
$$P_x(v) = \prod_{\{i \mid V_i \notin X\}} P(v_i \mid pa_i) \text{ for all } v \text{ consistent with } x.$$

(Truncated product, G-computation, manipulation theorem)

Graphical representation

Surgery: remove incoming arrows to X.

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!"

STRUCTURAL CAUSAL MODELS

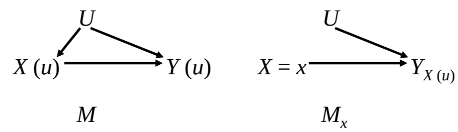
Definition: A structural causal model is a 4-tuple $\langle V, U, F, P(u)\rangle$, where

- $V = \{V_1,...,V_n\}$ are endogeneas 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

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.



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:

$$\boxed{Y_X(u) = Y_{M_X}(u)}$$

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:

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

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

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 $Y_x, X_z, Z_y, \ldots$?

N-R model	Structural model
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 $Y_X, Z_y,...$	$P(Y_X = y) = P_{M_X}(Y = y)$

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.

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).

19

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.

$$\text{ATE} \stackrel{\Delta}{=} E(Y \mid do(x_1)) - E(Y \mid do(x_0))$$

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).

20

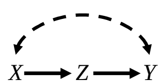
FORMULATING ASSUMPTIONS THREE LANGUAGES

1. English: Smoking (X), Cancer (Y), Tar (Z), Genotypes (U)
2. Counterfactuals: $Z_x(u) = Z_{yx}(u)$,

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

Not too friendly:
consistent? complete? redundant? arguable?

3. Structural:



21

IDENTIFYING CAUSAL EFFECTS IN POTENTIAL-OUTCOME FRAMEWORK

Define: Express the target quantity Q as a counterfactual formula, e.g., $E(Y(1) - Y(0))$

Assume: Formulate causal assumptions using the distribution:

$$P^* = P(X \mid Y, Z, Y(1), Y(0))$$

Identify: Determine if Q is identifiable using P^* and $Y = x\, Y(1) + (1 - x)\, Y(0)$.

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

22

GRAPHICAL – COUNTERFACTUALS SYMBIOSIS

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

1. Missing arrows $Y \leftarrow Z$ $\qquad Y_{x,z}(u) = Y_x(u)$
2. Missing arcs $Y \quad Z$ $\qquad Y_x \perp\!\!\!\perp Z_y$

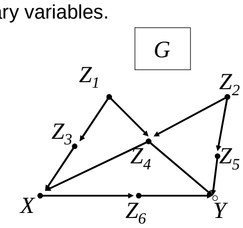
consistent, and readable from the graph.

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

23

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.

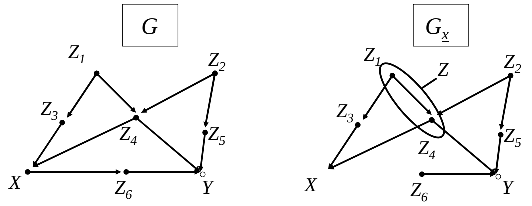


Can $P(y|do(x))$ be estimated if only a subset, Z, can be measured?

24

ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

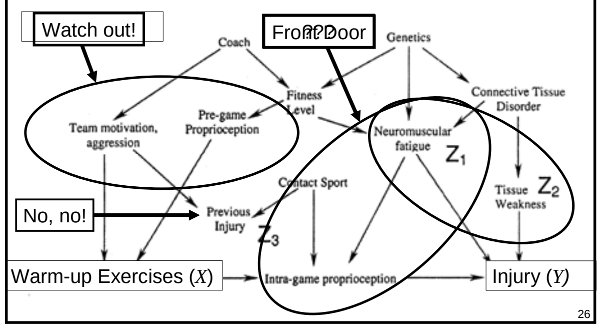
$P(y \mid do(x))$ is estimable if there is a set Z of variables such that Z d-separates X from Y in $G_{\underline{x}}$.



- 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

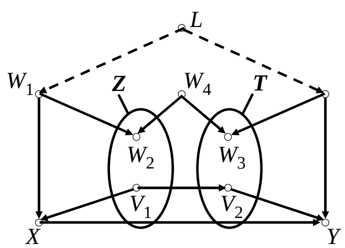
25

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



26

CONFOUNDING EQUIVALENCE WHEN TWO MEASUREMENTS ARE EQUALLY VALUABLE



$Z \overset{?}{\sim} T$

27

CONFOUNDING EQUIVALENCE WHEN TWO MEASUREMENTS ARE EQUALLY VALUABLE

Definition:
T and Z are c-equivalent if

$$\sum_t P(y \mid x, t)P(t) = \sum_z P(y \mid x, z)P(z) \qquad \forall x, y$$

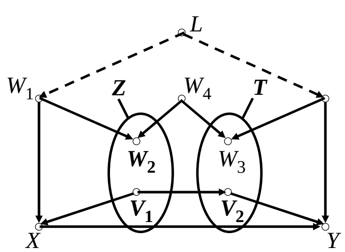
Definition (Markov boundary):
Markov boundary S_m of S (relative to X) is the minimal subset of S that d-separates X from all other members of S.

Theorem (Pearl and Paz, 2009)
Z and T are c-equivalent iff

1. $Z_m = T_m$, or
2. Z and T are admissible (i.e., satisfy the back-door condition)

28

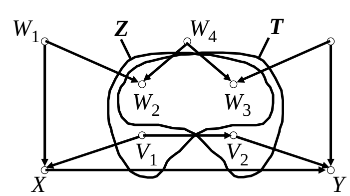
CONFOUNDING EQUIVALENCE WHEN TWO MEASUREMENTS ARE EQUALLY VALUABLE



$Z \not\approx T$

29

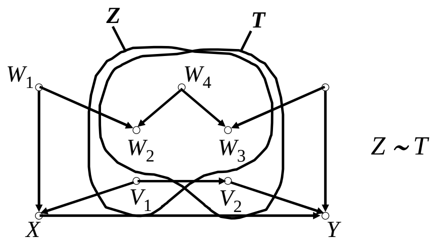
CONFOUNDING EQUIVALENCE WHEN TWO MEASUREMENTS ARE EQUALLY VALUABLE



$Z \not\approx T$

30

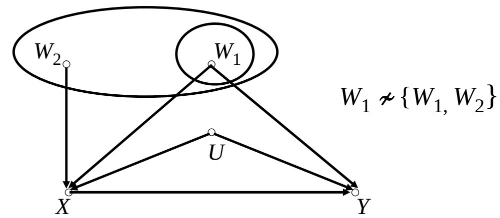
CONFOUNDING EQUIVALENCE WHEN TWO MEASUREMENTS ARE EQUALLY VALUABLE



$Z \sim T$

31

BIAS AMPLIFICATION BY INSTRUMENTAL VARIABLES



$W_1 \not\sim \{W_1, W_2\}$

- 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

32

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?

33

WHY DECOMPOSE EFFECTS?

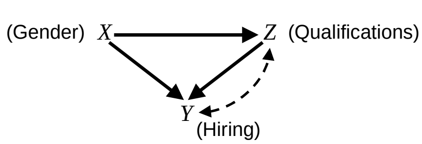
1. To understand how Nature works
2. To comply with legal requirements
3. To predict the effects of new type of interventions:

   Signal routing, rather than variable fixing

34

LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



What is the direct effect of X on Y ?

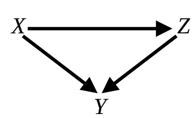
$$E(Y \mid do(x_1), do(z)) - E(Y \mid do(x_0), do(z))$$

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

35

NATURAL INTERPRETATION OF AVERAGE DIRECT EFFECTS

Robins and Greenland (1992) – "Pure"



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

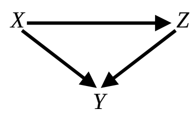
Natural Direct Effect of X on Y: $DE(x_0, x_1; Y)$
The expected change in Y, when we change X from x_0 to x_1 and, for each u, we keep Z constant at whatever value it attained before the change.

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

In linear models, DE = Controlled Direct Effect

36

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.

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

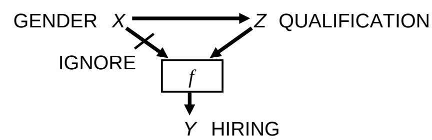
In linear models, *IE = TE - DE*

37

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

38

MEDIATION FORMULAS

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

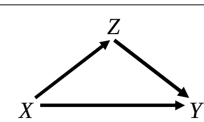
$$DE = \sum_z [E(Y \mid do(x_1, z)) - E(Y \mid do(x_0, z))] P(z \mid do(x_0)).$$

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

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

39

MEDIATION FORMULAS IN UNCONFOUNDED MODELS



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

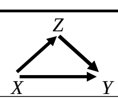
$$IE = \sum_z [E(Y \mid x_0, z)[P(z \mid x_1) - P(z \mid x_0)]$$

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

$IE =$ Fraction of responses explained by mediation
$TE - DE =$ Fraction of responses owed to mediation

40

COMPUTING THE MEDIATION FORMULA



	X	Z	Y	$E(Y\|x,z)=g_{xz}$	$E(Z\|x)=h_x$
n_1	0	0	0	$\frac{n_2}{n_1+n_2} = g_{00}$	$\frac{n_3+n_4}{n_1+n_2+n_3+n_4} = h_0$
n_2	0	0	1		
n_3	0	1	0	$\frac{n_4}{n_3+n_4} = g_{01}$	
n_4	0	1	1		
n_5	1	0	0	$\frac{n_6}{n_5+n_6} = g_{10}$	$\frac{n_7+n_8}{n_5+n_6+n_7+n_8} = h_1$
n_6	1	0	1		
n_7	1	1	0	$\frac{n_8}{n_7+n_8} = g_{11}$	
n_8	1	1	1		

$$DE = (g_{10} - g_{00})(1 - h_0) + (g_{11} - g_{01}) h_0$$

$$IE = (h_1 - h_0)(g_{01} - g_{00})$$

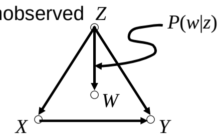
41

RAMIFICATION OF THE MEDIATION FORMULA

- *DE* should be averaged over mediator levels,
  IE should NOT be averaged over exposure levels.
- *TE-DE* need not equal *IE*
  TE-DE = proportion for whom mediation is necessary
  IE = proportion for whom mediation is sufficient
- *TE-DE* informs interventions on indirect pathways
  IE informs intervention on direct pathways.

42

MEASUREMENT BIAS AND EFFECT RESTORATION



$P(y \mid do(x))$ is identifiable from measurement of W, if $P(w \mid z)$ is given (Selen, 1986; Greenland & Lash, 2008)

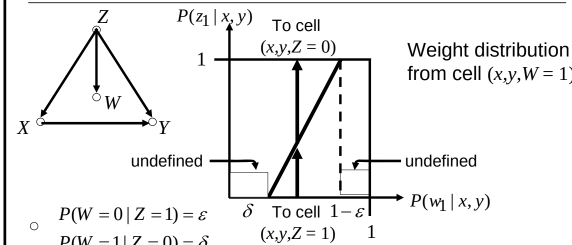
$$P(w \mid x,y,z) = P(w \mid z) \quad \text{(local independence)}$$

$$\begin{aligned} P(x,y,w) &= \sum_z P(x,y,z,w) \\ &= \sum_z P(w \mid x,y,z)P(x,y,z) \\ &= \sum_z P(w \mid z)P(x,y,z) \end{aligned}$$

$$P(x,y,z) = \sum_w I(z,w)P(x,y,w)$$

43

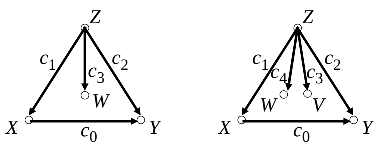
EFFECT RESTORATION IN BINARY MODELS



- $P(W=0 \mid Z=1) = \varepsilon$
  $P(W=1 \mid Z=0) = \delta$
- $P(y \mid do(x)) \cong \frac{P(x,y,w_1)}{P(x|w_1)}\left[1-\delta\left(\frac{1}{P(w_1 \mid x,y)} - \frac{1-P(x)}{P(w_1)}\right)\right] + \frac{P(x,y,w_0)}{P(x|w_0)}\left[1-\varepsilon\left(\frac{1}{P(w_0 \mid x,y)} - \frac{1-P(x)}{P(w_0)}\right)\right]$

44

EFFECT RESTORATION IN LINEAR MODELS



(a) (b)

- $c_0 = \dfrac{\sigma_{xy} - \sigma_{xw}\sigma_{yw}/k}{\sigma_{xx}\sigma_{xw}^2/k} \qquad k = c_3^2\sigma_{zz}$
- $c_0 = \dfrac{cov(XY)cov(XV) - cov(YW)cov(WV)}{cov(XV)var(X) - cov(XW)cov(WV)}$
- Correlated proxies (Cai & Kuroki, 2008)

45

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.

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

(Aristotle 384-322 B.C.)

From Charlie Poole

46

QUESTIONS???

They will be answered

47