JSM-2016
presented: 8.1.16

CAUSAL INFERENCE IN STATISTICS:

A Gentle Introduction

Judea Pearl
Departments of Computer Science and Statistics
UCLA

OUTLINE

1. The causal revolution – from statistics to policy intervention to counterfactuals
2. The fundamental laws of causal inference
3. From counterfactuals to problem solving (gems)

Old gems:
- a) policy evaluation ("treatment effects"…)
- b) attribution – "but for"
- c) mediation – direct and indirect effects

New gems:
- d) generalizability – external validity
- e) selection bias – non-representative sample
- f) missing data

FIVE LESSONS FROM THE THEATRE OF CAUSAL INFERENCE

1. Every causal inference task must rely on judgmental, extra-data assumptions (or experiments).
2. We have ways of encoding those assumptions mathematically and test their implications.
3. We have a mathematical machinery to take those assumptions, combine them with data and derive answers to questions of interest.
4. We have a way of doing (2) and (3) in a language that permits us to judge the scientific plausibility of our assumptions and to derive their ramifications swiftly and transparently.
5. Items (2)-(4) make causal inference manageable, fun, and profitable.

WHAT EVERY STUDENT SHOULD KNOW

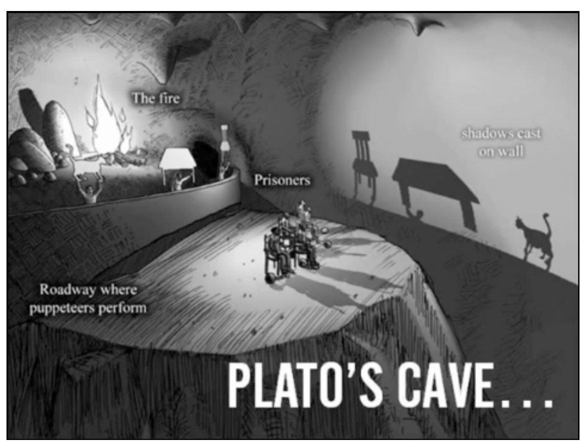
The five lessons from the causal theatre, especially:

3. We have a mathematical machinery to take meaningful assumptions, combine them with data, and derive answers to questions of interest.

5. This makes causal inference FUN !

WHY NOT STAT-101? THE STATISTICS PARADIGM 1834–2016

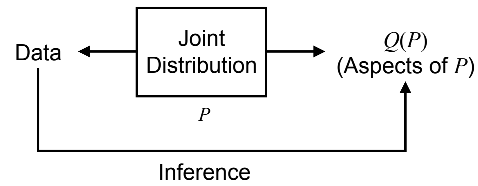
- "The object of statistical methods is the reduction of data" (Fisher 1922).
- Statistical concepts are those expressible in terms of joint distribution of observed variables.
- All others are: "substantive matter," "domain dependent," "metaphysical," "ad hockery," i.e., outside the province of statistics, ruling out all interesting questions.
- Slow awakening since Neyman (1923) and Rubin (1974).
- Traditional Statistics Education = Causalophobia



THE CAUSAL REVOLUTION

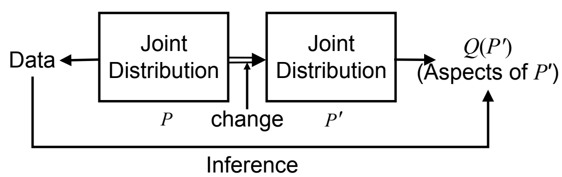
1. "More has been learned about causal inference in the last few decades than the sum total of everything that had been learned about it in all prior recorded history."
   (Gary King, Harvard, 2014)
2. From liability to respectability
   - JSM 2003 – 13 papers
   - JSM 2013 – 130 papers
3. The gems – for Fun and Profit
   - Its fun to solve problems that Pearson, Fisher, Neyman, and my professors . . . were not able to articulate.
   - Problems that users pay for.

TRADITIONAL STATISTICAL INFERENCE PARADIGM



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

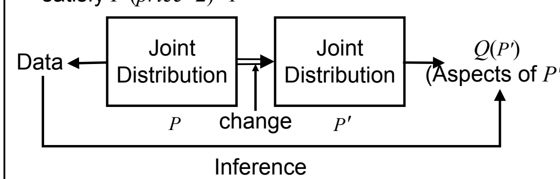
FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES



e.g., Estimate $P'(sales)$ if we double the price.
How does P change to P'? New oracle
e.g., Estimate $P'(cancer)$ if we ban smoking.

FROM STATISTICAL TO CAUSAL ANALYSIS: 1. THE DIFFERENCES

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

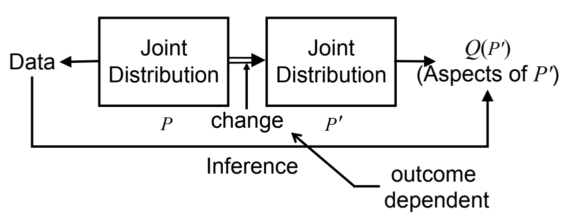


Note: $P'(sales) \neq P\,(sales \mid price = 2)$

e.g., Doubling price ≠ seeing the price doubled.

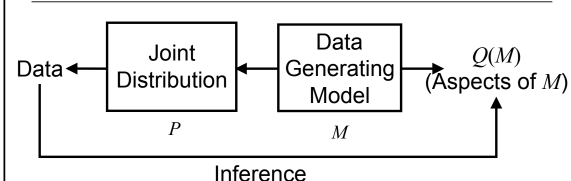
P does not tell us how it ought to change.

FROM STATISTICS TO COUNTERFACTUALS: RETROSPECTION



What happens when P changes?
e.g., Estimate the probability that a customer who bought A would buy A if we were to double the price.

STRUCTURAL CAUSAL MODEL THE NEW ORACLE



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

P – model of data, M – model of reality

WHAT KIND OF QUESTIONS SHOULD THE NEW ORACLE ANSWER THE CAUSAL HIERARCHY

- Observational Questions:
  "What if we see A" (What is?) $P(y \mid A)$
- Action Questions:
  "What if we do A?" (What if?) $P(y \mid do(A))$
- Counterfactuals Questions:
  "What if we did things differently?" (Why?) $P(y_{A'} \mid A)$
- Options:
  "With what probability?"

SYNTACTIC DISTINCTION

WHAT KIND OF QUESTIONS SHOULD THE NEW ORACLE ANSWER THE CAUSAL HIERARCHY

- Observational Questions:
  "What if we see A" Bayes Networks
- Action Questions:
  "What if we do A?" Causal Bayes Networks
- Counterfactuals Questions: Functional Causal
  "What if we did things differently?" Diagrams
- Options:
  "With what probability?"

GRAPHICAL REPRESENTATIONS

FROM STATISTICAL TO CAUSAL ANALYSIS: 2. THE SHARP BOUNDARY

1. Causal and associational concepts do not mix.

CAUSAL	ASSOCIATIONAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
"Holding constant" / "Fixing"	"Controlling for" / Conditioning
Confounding / Effect	Odds and risk ratios
Instrumental variable	Collapsibility / Granger causality
Ignorability / Exogeneity	Propensity score

2.

3.

4.

FROM STATISTICAL TO CAUSAL ANALYSIS: 3. THE MENTAL BARRIERS

1. Causal and associational concepts do not mix.

CAUSAL	ASSOCIATIONAL
Spurious correlation	Regression
Randomization / Intervention	Association / Independence
"Holding constant" / "Fixing"	"Controlling for" / Conditioning
Confounding / Effect	Odds and risk ratios
Instrumental variable	Collapsibility / Granger causality
Ignorability / Exogeneity	Propensity score

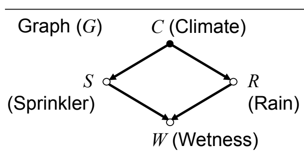
2. No causes in – no causes out (Cartwright, 1989)

   data
   causal assumptions (or experiments) } $\Rightarrow$ 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 ($x \;\Box\!\!\rightarrow Y$))

A MODEL AND ITS GRAPH

Graph (G)



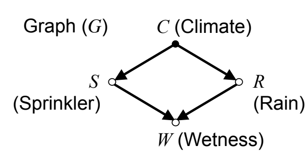
Model (M)

$C = f_C(U_C)$
$S = f_S(C, U_S)$
$R = f_R(C, U_R)$
$W = f_W(S, R, U_W)$

17

DERIVING COUNTERFACTUALS FROM A MODEL

Graph (G)

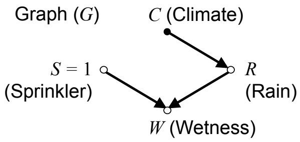


Model (M)

$C = f_C(U_C)$
$S = f_S(C, U_S)$
$R = f_R(C, U_R)$
$W = f_W(S, R, U_W)$

Would the pavement be wet HAD the sprinkler been ON?

DERIVING COUNTERFACTUALS FROM A MODEL

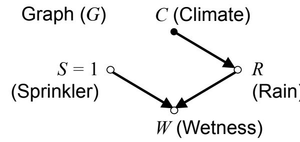


Mutilated Model ($M_{S=1}$)

$C = f_C(U_C)$
$S = 1$
$R = f_R(C, U_R)$
$W = f_W(S, R, U_W)$

Would the pavement be wet had the sprinkler been ON?
Find if $W = 1$ in $M_{S=1}$
Find if $f_W(S = 1, R, U_W) = 1$ or $W_{S=1} = 1$
What is the probability that we find the pavement is wet if we turn the sprinkler ON?
Find if $P(W_{S=1} = 1) = P(W = 1 \mid do(S = 1))$

DERIVING COUNTERFACTUALS FROM A MODEL

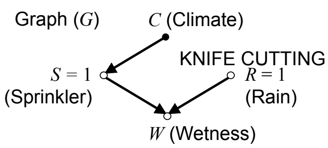


Mutilated Model ($M_{S=1}$)

$C = f_C(U_C)$
$S = 1$
$R = f_R(C, U_R)$
$W = f_W(S, R, U_W)$

Would it rain if we turn the sprinkler ON?
Not necessarily, because $R_{S=1} = R$

DERIVING COUNTERFACTUALS FROM A MODEL



Mutilated Model ($M_{R=1}$)

$C = f_C(U_C)$
$S = f_S(C, U_S)$
$R = 1$
$W = f_W(S, R, U_W)$

Would the pavement be wet had the rain been ON?
Find if $W = 1$ in $M_{R=1}$
Find if $f_W(S, R = 1, U_W) = 1$

EVERY COUNTERFACTAUL HAS A VALUE IN M

THE TWO FUNDAMENTAL LAWS OF CAUSAL INFERENCE

1. The Law of Counterfactuals (and Interventions)

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

(M generates and evaluates all counterfactuals.)

and all interventions

$$ATE = E_u[Y_x(u)] = E[Y \mid do(x)]$$

THE TWO FUNDAMENTAL LAWS OF CAUSAL INFERENCE

1. The Law of Counterfactuals (and Interventions)

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

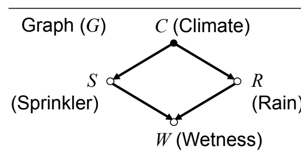
(M generates and evaluates all counterfactuals.)

2. The Law of Conditional Independence (d-separation)

$$(X \text{ sep } Y \mid Z)_{G(M)} \Rightarrow (X \perp\!\!\!\perp Y \mid Z)_{P(v)}$$

(Separation in the model ⇒ independence in the distribution.)

THE LAW OF CONDITIONAL INDEPENDENCE



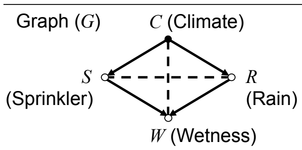
Model (M)

$C = f_C(U_C)$
$S = f_S(C, U_S)$
$R = f_R(C, U_R)$
$W = f_W(S, R, U_W)$

Gift of the Gods

If the U's are independent, the observed distribution $P(C,R,S,W)$ satisfies constraints that are:
(1) independent of the f's and of $P(U)$,
(2) readable from the graph.

D-SEPARATION: NATURE'S LANGUAGE FOR COMMUNICATING ITS STRUCTURE



Model (M)

$C = f_C(U_C)$
$S = f_S(C, U_S)$
$R = f_R(C, U_R)$
$W = f_W(S, R, U_W)$

Every missing arrow advertises an independency, conditional on a separating set.

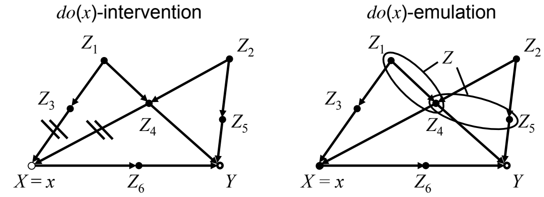
e.g., $C \perp\!\!\!\perp W \mid (S,R)$ $\qquad S \perp\!\!\!\perp R \mid C$

Applications:

1. Model testing
2. Structure learning
3. Reducing "what if I do" questions to symbolic calculus
4. Reducing scientific questions to symbolic calculus

ELIMINATING CONFOUNDING BIAS THE BACK-DOOR CRITERION

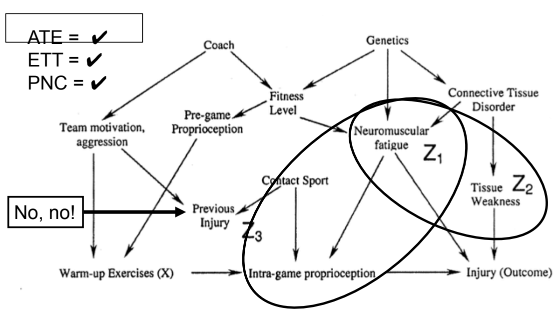
$P(y \mid do(x))$ is estimable if there is a set Z of variables that if conditioned on, would block all X-Y paths that are severed by the intervention and none other.



Moreover, $P(y \mid do((x)) = \sum_z P(y \mid x, z) P(z)$ (Adjustment)

$Back\text{-}door \Longrightarrow Y_x \perp\!\!\!\perp X \mid Z \Longrightarrow (Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}}$

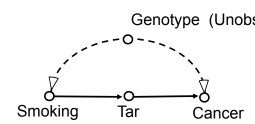
WHAT IF VARIABLES ARE UNOBSERVED? EFFECT OF WARM-UP ON INJURY (Shrier & Platt, 2008)



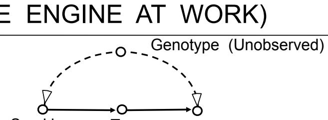
GOING BEYOND ADJUSTMENT

Goal: Find the effect of *Smoking* on *Cancer*, $P(c \mid do(s))$, given samples from $P(S, T, C)$, when latent variables confound the relationship S-C.

Query $\qquad$ Data



IDENTIFICATION REDUCED TO CALCULUS (THE ENGINE AT WORK)



Query

$P(c \mid do(s)) = \Sigma_t P(c \mid do(s), t) P(t \mid do(s))$	Probability Axioms
$= \Sigma_t P(c \mid do(s), do(t)) P(t \mid do(s))$	Rule 2
$= \Sigma_t P(c \mid do(s), do(t)) P(t \mid s)$	Rule 2
$= \Sigma_t P(c \mid do(t)) P(t \mid s)$	Rule 3
$= \Sigma_{s'} \Sigma_t P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s)$	Probability Axioms
$= \Sigma_{s'} \Sigma_t P(c \mid t, s') P(s' \mid do(t)) P(t \mid s)$	Rule 2
$= \Sigma_{s'} \Sigma_t P(c \mid t, s') P(s') P(t \mid s)$	Rule 3

Estimand

DO-CALCULUS (THE WHEELS OF THE ENGINE)

The following transformations are valid for every interventional distribution generated by a structural causal model M:

Rule 1: Ignoring observations

$P(y \mid do(x), z, w) = P(y \mid do(x), w),$

if $\quad (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}}$

Rule 2: Action/observation exchange

$P(y \mid do(x), do(z), w) = P(y \mid do(x), z, w),$

if $\quad (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}}$

Rule 3: Ignoring actions

$P(y \mid do(x), do(z), w) = P(y \mid do(x), w),$

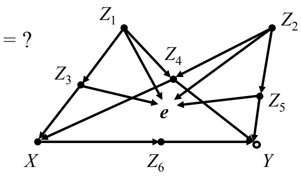
if $\quad (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\overline{Z(W)}}}$

GEM 1: THE IDENTIFICATION PROBLEM IS SOLVED (NONPARAMETRICALLY)

- The estimability of any expression of the form
  $Q = P(y_1, y_2, ..., y_n \mid do(x_1, x_2, ..., x_m), z_1, z_2, ..., z_k)$
  can be decided in polynomial time.
- If Q is estimable, then its estimand can be derived in polynomial time.
- The algorithm is complete.
- Same for ETT (Shpitser 2008).

PROPENSITY SCORE ESTIMATOR (Rosenbaum & Rubin, 1983)

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



Can e replace $\{Z_1, Z_2, Z_3, Z_4, Z_5\}$?

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

Theorem: $$\sum_z P(y \mid z, x) P(z) = \sum_e P(y \mid e, x) P(e)$$

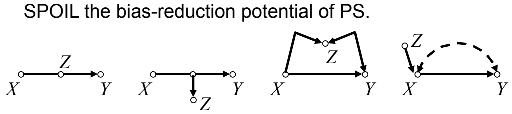
Adjustment for $e\,(z)$ replaces Adjustment for Z

WHAT PROPENSITY SCORE (PS) PRACTITIONERS NEED TO KNOW

$$e(z) = P(X=1 \mid Z=z)$$

$$\sum_z P(y \mid z, x) P(z) = \sum_e P(y \mid e, x) P(e)$$

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 PS.



3. In particular, instrumental variables tend to amplify bias.
4. Choosing sufficient set for PS, requires causal knowledge, which PS alone cannot provide.

33

DAGS VS. POTENTIAL COUTCOMES AN UNBIASED PERSPECTIVE

1. Semantic Equivalence
2. Both are abstractions of Structural Causal Models (SCM).

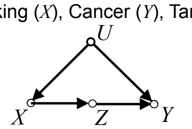
$$Y_x(u) = Y_{M_x}(u)$$

$$X \to Y \qquad y = f(x, z, u)$$

$Y_x(u)$ = All factors that affect Y when X is held constant at $X{=}x$.

CHOOSING A LANGUAGE TO ENCODE ASSUMPTIONS

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

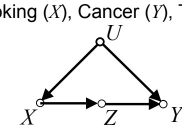


2. Potential Outcome:
   $Z_x(u) = Z_{yx}(u),$
   $X_y(u) = X_{zy}(u) = X_z(u) = X(u),$
   $Y_z(u) = Y_{zx}(u), \quad Z_x \perp\!\!\!\perp \{Y_z, X\}$

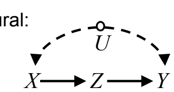
Not too friendly:
Consistent?, complete?, redundant?, plausible?, testable?

CHOOSING A LANGUAGE TO ENCODE ASSUMPTIONS

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



2. Counterfactuals: $Z_x(u) = Z_{yx}(u),$
   $X_y(u) = X_{zy}(u) = X_z(u) = X(u),$
   $Y_z(u) = Y_{zx}(u), \quad Z_x \perp\!\!\!\perp \{Y_z, X\}$
3. Structural:



$x = f_1(u, \varepsilon_1) \qquad y = f_3(z, u, \varepsilon_3)$
$z = f_2(x, \varepsilon_2) \qquad \varepsilon_1 \perp\!\!\!\perp \varepsilon_2 \perp\!\!\!\perp \varepsilon_3$

GEM 2: ATTRIBUTION

- Your Honor! My client (Mr. A) died BECAUSE he used that drug.

-

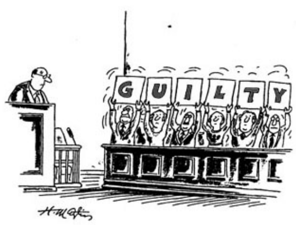
GEM 2: ATTRIBUTION

- Your Honor! My client (Mr. A) died BECAUSE he used that drug.

- Court to decide if it is MORE PROBABLE THAN NOT that A would be alive BUT FOR the drug!
- $PN = P(alive_{\{no\ drugs\}} \mid dead, drug) \geq 0.50$

CAN FREQUENCY DATA DETERMINE LIABILITY?

Sometimes:



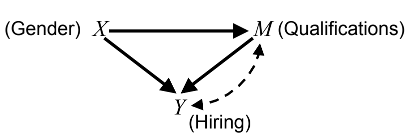
- WITH PROBABILITY ONE $1 \leq PN \leq 1$
- Combined data tell more that each study alone

GEM 3: MEDIATION 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 re-routing and mechanism deactivating, rather than variable fixing

LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?



What is the direct effect of X on Y ?

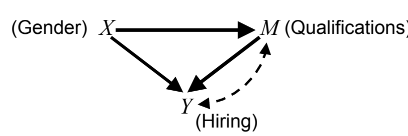
$$CDE = E(Y|do(x_1), do(m)) - E(Y|do(x_0), do(m))$$

(m-dependent) Adjust for *M*? No! No!

CDE identification is completely solved

LEGAL DEFINITION OF DISCRIMINATION

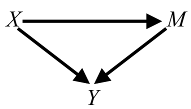
Can data prove an employer guilty of hiring discrimination?



The Legal Definition:
Find the probability that “the employer would have acted differently had the employee been of different sex and qualification had been the same.”

NATURAL INTERPRETATION OF AVERAGE DIRECT EFFECTS

Robins and Greenland (1992) – Pearl (2001)



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

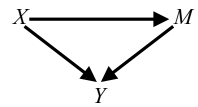
Natural Direct Effect of X on Y: $DE(x_0, x_1; Y)$
The expected change in Y, when we change X from x_0 to x_1 and, for each u, we keep M constant at whatever value it attained before the change.

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

Note the 3-way symbiosis

43

DEFINITION OF INDIRECT EFFECTS



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

No controlled indirect effect

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 M change to whatever value it would have attained had X changed to x_1.

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

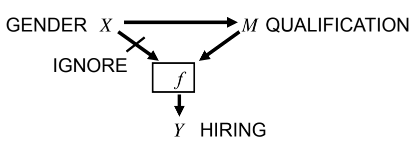
In linear models, *IE = TE - DE*

44

POLICY IMPLICATIONS OF INDIRECT EFFECTS

What is the indirect effect of X on Y?

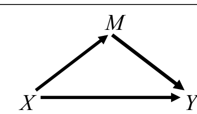
The effect of Gender on Hiring if sex discrimination is eliminated.



Deactivating a link – a new type of intervention

45

THE MEDIATION FORMULAS IN UNCONFOUNDED MODELS



$m = f(x, u_1)$
$y = g(x, m, u_2)$
u_1 independent of u_2

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

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

$$TE = E(Y \mid x_1) - E(Y \mid x_0) \qquad TE \neq DE + IE$$

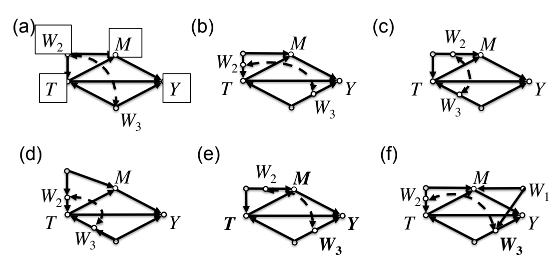
$IE =$ Fraction of responses explained by mediation (sufficient)

$TE - DE =$ Fraction of responses owed to mediation (necessary)

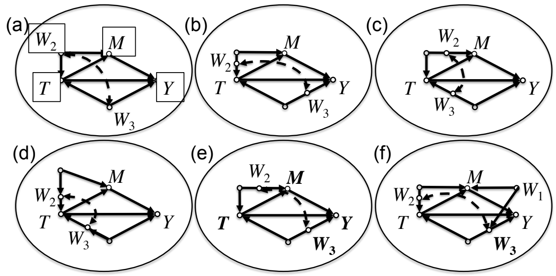
SUMMARY OF MEDIATION (GEM 3) Identification is a solved problem

- The nonparametric estimability of natural (and controlled) direct and indirect effects can be determined in polynomial time given any causal graph G with both measured and unmeasured variables.
- If NDE (or NIE) is estimable, then its estimand can be derived in polynomial time.
- The algorithm is complete and was extended to any path-specific effects (Shpitser, 2013).

WHEN CAN WE IDENTIFY MEDIATED EFFECTS?



WHEN CAN WE IDENTIFY MEDIATED EFFECTS?



GEM 4: GENERALIZABILITY AND DATA FUSION

The problem

- How to combine results of several experimental and observational studies, each conducted on a different population and under a different set of conditions,
- so as to construct a valid estimate of effect size in yet a new population, unmatched by any of those studied.

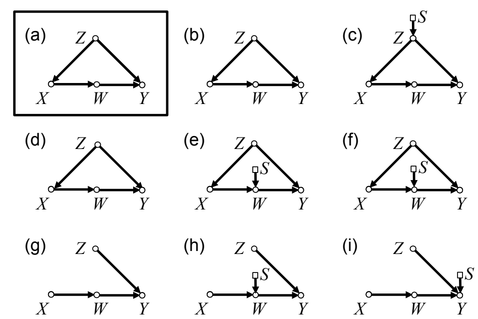
THE PROBLEM IN REAL LIFE

Target population Π^* Query of interest: $Q = P^*(y \mid do(x))$

(a) Arkansas Survey data available	(b) New York Survey data Resembling target	(c) Los Angeles Survey data Younger population
(d) Boston Age not recorded Mostly successful lawyers	(e) San Francisco High post-treatment blood pressure	(f) Texas Mostly Spanish subjects High attrition
(g) Toronto Randomized trial College students	(h) Utah RCT, paid volunteers, unemployed	(i) Wyoming RCT, young athletes

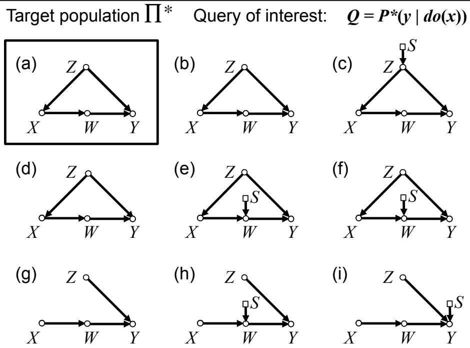
THE PROBLEM IN MATHEMATICS

Target population Π^* Query of interest: $Q = P^*(y \mid do(x))$



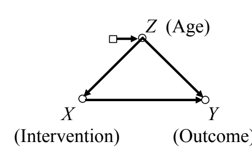
THE SOLUTION IS IN ALGORITHMS

Target population Π^* Query of interest: $Q = P^*(y \mid do(x))$



THE TWO–POPULATION PROBLEM

WHAT CAN EXPERIMENTS IN LA TELL US ABOUT NYC?



Π (LA) $\longrightarrow \Pi^*$ (NY)

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

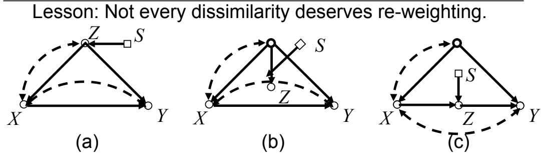
Observational study in NYC
Measured: $P^*(x,y,z)$
$P^*(z) \neq P(z)$

Needed: $Q = P^*(y \mid do(x)) = \; ? \; = \sum_z P(y \mid do(x), z) P^*(z)$

Transport Formula: $Q = F(P, P_{do}, P^*)$

TRANSPORT FORMULAS DEPEND ON THE CAUSAL STORY

Lesson: Not every dissimilarity deserves re-weighting.



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

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

TRANSPORTABILITY REDUCED TO CALCULUS

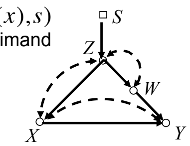
Theorem

A causal relation $\boldsymbol{R}$ is transportable from Π to Π^* if and only if it is reducible, using the rules of ***do***-calculus, to an expression in which $\boldsymbol{S}$ is separated from ***do***( ).

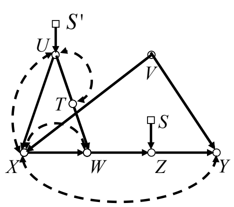
Query

$$R(\Pi^*) = P^*(y \mid do(x)) = P(y \mid do(x), s)$$
$$= \sum_w P(y \mid do(x), s, w) P(w \mid do(x), s)$$
$$= \sum_w P(y \mid do(x), w) P(w \mid s)$$
$$= \sum_w P(y \mid do(x), w) P^*(w)$$

Estimand



RESULT: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE



INPUT: Annotated Causal Graph

S ⊡→ Factors creating differences

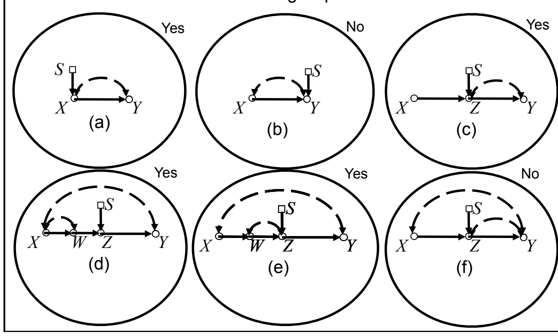
OUTPUT:

1. Transportable or not?
2. Measurements to be taken in the experimental study
3. Measurements to be taken in the target population
4. A transport formula
5. Completeness (Bareinboim, 2012)

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

WHICH MODEL LICENSES THE TRANSPORT OF THE CAUSAL EFFECT $X \rightarrow Y$

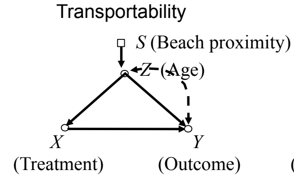
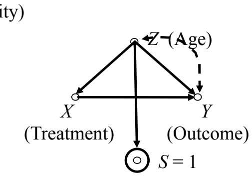
S ⊡→ External factors creating disparities



SUMMARY OF TRANSPORTABILITY RESULTS

- Nonparametric transportability of experimental results from multiple environments can be determined provided that commonalities and differences are encoded in selection diagrams.
- When transportability is feasible, the transport formula can be derived in polynomial time.
- The algorithm is complete.

GEM 5: RECOVERING FROM SAMPLING SELECTION BIAS

Transportability	Selection Bias
	
S = disparity-producing factors	S = sampling mechanism
Nature-made	Man-made
Non-estimable	Non-estimable

RECOVERING FROM SELECTION BIAS

Query: Find $P(y \mid do(x))$

Data: $P(y \mid do(x), z, S=1)$ from study

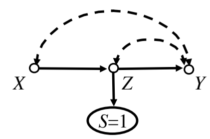
$P(y, x, z)$ from survey

Theorem:
A query Q can be recovered from selection biased data iff Q can be transformed, using *do*-calculus to a form provided by the data, i.e.,

(i) All *do*-expressions are conditioned on S = 1
(ii) No *do*-free expression is conditioned on S = 1

RECOVERING FROM SELECTION BIAS

Example:



$$P(y \mid do(x)) = \sum_z P(y \mid do(x), z) P(z \mid do(x))$$
$$= \sum_z P(y \mid do(x), z) P(z \mid x) \quad \text{(Rule 2)}$$
$$= \sum_z P(y \mid do(x), z, S=1) P(z \mid x) \quad \text{(Rule 1)}$$

GEM 6: MISSING DATA: A STATISTICAL PROBLEM TURNED CAUSAL

Sam-ple #	X	Y	Z
1	1	0	0
2	1	0	1
3	1	m	m
4	0	1	m
5	m	1	m
6	m	0	1
7	m	m	0
8	0	1	m
9	0	0	m
10	1	0	m
11	1	0	1
-			

Question:
Is there a consistent estimator of $P(X,Y,Z)$? That is, is $P(X,Y,Z)$ estimable (asymptotically) as if no data were missing.

Conventional Answer:
Run imputation algorithm and, if missingness occurs at random (MAR), (a condition that is untestable and uninterpretable), then it will coverage to a consistent estimate.

GEM 6: MISSING DATA: A STATISTICAL PROBLEM TURNED CAUSAL

Sam-ple #	X	Y	Z
1	1	0	0
2	1	0	1
3	1	m	m
4	0	1	m
5	m	1	m
6	m	0	1
7	m	m	0
8	0	1	m
9	0	0	m
10	1	0	m
11	1	0	1
-			

Question:
Is there a consistent estimator of $P(X,Y,Z)$? That is, is $P(X,Y,Z)$ estimable (asymptotically) as if no data were missing.

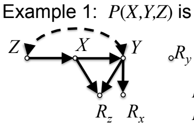
Model-based Answers:

1. There is no Model-free estimator, but,
2. Given a missingness model, we can tell you yes/no, and how.
3. Given a missingness model, we can tell you whether or not it has testable implications.

SMART ESTIMATION OF $P(X,Y,Z)$

Sam-ple #	X	Y	Z
1	1	0	0
2	1	0	1
3	1	m	m
4	0	1	m
5	m	1	m
6	m	0	1
7	m	m	0
8	0	1	m
9	0	0	m
10	1	0	m
11	1	0	1
-			

Example 1: $P(X,Y,Z)$ is estimable



$R_x = 0 \Rightarrow X$ observed
$R_x = 1 \Rightarrow X$ missing

$$P(X,Y,Z) = P(Z \mid X,Y,R_x=0,R_y=0,R_z=0)\, P(X \mid Y,R_x=0,R_y=0)\, P(Y \mid R_y=0)$$

Testable implications:

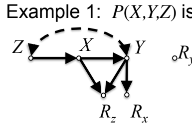
$Z \perp\!\!\!\perp R_y \mid R_z = 0$

$R_z \perp\!\!\!\perp R_x \mid Y, R_y = 0$

SMART ESTIMATION OF $P(X,Y,Z)$

Sam-ple #	X	Y	Z
1	1	0	0
2	1	0	1
3	1	m	m
4	0	1	m
5	m	1	m
6	m	0	1
7	m	m	0
8	0	1	m
9	0	0	m
10	1	0	m
11	1	0	1
-			

Example 1: $P(X,Y,Z)$ is estimable



$R_x = 0 \Rightarrow X$ observed
$R_x = 1 \Rightarrow X$ missing

$$P(X,Y,Z) = P(Z \mid X,Y,R_x=0,R_y=0,R_z=0)\, P(X \mid Y,R_x=0,R_y=0)\, P(Y \mid R_y=0)$$

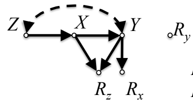
Testable implications:

$X \perp\!\!\!\perp R_x \mid Y$ is not testable because X is not fully observed.

SMART ESTIMATION OF $P(X,Y,Z)$

Sample #	X	Y	Z
1	1	0	0
2	1	0	1
3	1	m	m
4	0	1	m
5	m	1	m
6	m	0	1
7	m	m	0
8	0	1	m
9	0	0	m
10	1	0	m
11	1	0	1
-			

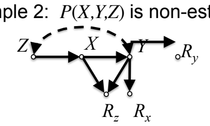
Example 1: $P(X,Y,Z)$ is estimable



$R_x = 0 \Rightarrow X$ observed
$R_x = 1 \Rightarrow X$ missing

$$P(X,Y,Z) = P(Z \mid X,Y,R_x = 0, R_y = 0, R_z = 0)$$
$$P(X \mid Y, R_x = 0, R_y = 0)$$
$$P(Y \mid R_y = 0)$$

Example 2: $P(X,Y,Z)$ is non-estimable



WHAT MAKES MISSING DATA A CAUSAL PROBLEM?

The knowledge required to guarantee consistency is causal i.e., it comes from our understanding of the mechanism that causes missingness (not from hopes for fortunate conditions to hold).

Graphical models of this mechanism provide:
1. Tests for MCAR and MAR,
2. consistent estimates for large classes of MNAR,
3. testable implications of missingness models,
4. closed-form estimands, bounds, and more.
5. Query-smart estimation procedures.

CONCLUSIONS

- A revolution is judged by the gems it spawns.
- Each of the six gems of the causal revolution is shining in fun and profit.
- More will be learned about causal inference in the next decade than most of us imagine today.
- Because statistical education is about to catch up with Statistics.

Refs: http://bayes.cs.ucla.edu/jp_home.html

Thank you

Joint work with:
Elias Bareinboim
Karthika Mohan
Ilya Shpitser
Jin Tian
Many more . . .

Time for a short commercial

Gems 1-2-3 can be enjoyed here: