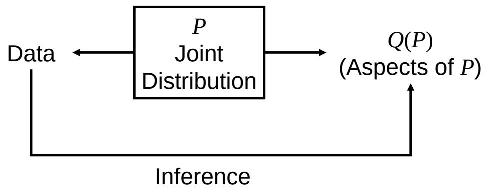
THE MATHEMATICS OF CAUSAL RELATIONS

Judea Pearl
Department of Computer Science
UCLA

OUTLINE

- Statistical vs. Causal Modeling: distinction and mental barriers
- Formal semantics for counterfactuals: definition, axioms, graphical representations
- Graphs and Algebra: Symbiosis translation and accomplishments

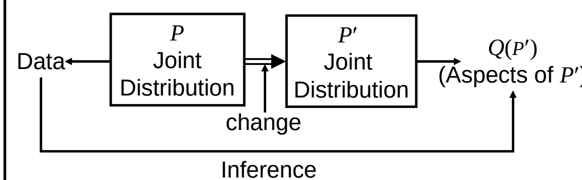
TRADITIONAL STATISTICAL INFERENCE PARADIGM



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

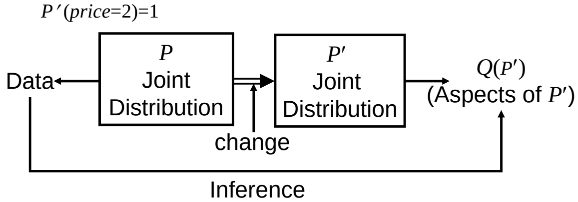
Probability and statistics deal with static relations



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
e.g. Curing symptoms vs. curing diseases
e.g. Analogy: mechanical deformation

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

1. Causal and statistical concepts do not mix.

CAUSAL	STATISTICAL
Spurious correlation	Regression
Randomization	Association / Independence
Confounding / Effect	"Controlling for" / Conditioning
Instrument	Odd and risk ratios
Holding constant	Collapsibility
Explanatory variables	

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	Association / Independence
Confounding / Effect	“Controlling for” / Conditioning
Instrument	Odd and risk ratios
Holding constant	Collapsibility
Explanatory variables	

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

   statistical assumptions + data
   causal assumptions } $\Rightarrow$ 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.

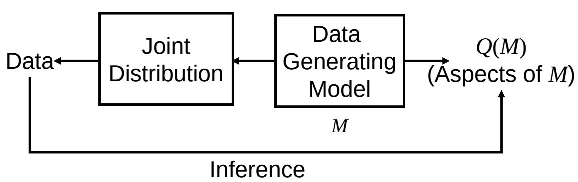
CAUSAL	STATISTICAL
Spurious correlation	Regression
Randomization	Association / Independence
Confounding / Effect	“Controlling for” / Conditioning
Instrument	Odd and risk ratios
Holding constant	Collapsibility
Explanatory variables	

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

   statistical assumptions + data
   causal assumptions } $\Rightarrow$ causal conclusions

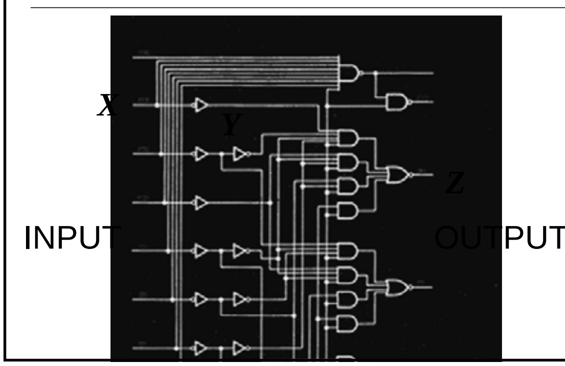
3. Causal assumptions cannot be expressed in the mathematical language of standard statistics.
4. Non-standard mathematics:
   a) Structural equation models (Wright, 1920; Simon, 1960)
   b) Counterfactuals (Neyman-Rubin (Y_x), Lewis (x □→ Y))

THE STRUCTURAL MODEL PARADIGM



M – Oracle for computing answers to Q's.
e.g.,
Infer whether customer u who bought product A would still buy A if we were to double the price.

FAMILIAR CAUSAL MODEL ORACLE FOR MANIPILATION



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 observable variables
- $U = \{U_1,...,U_m\}$ are background variables
- $F = \{f_1,..., f_n\}$ are functions determining V, $v_i = f_i(v, u)$
- $P(u)$ is a distribution over U

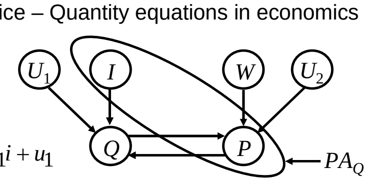
$P(u)$ and F induce a distribution $P(v)$ over observable variables

STRUCTURAL MODELS AND CAUSAL DIAGRAMS

The arguments of the functions $v_i = f_i(v,u)$ define a graph

$$v_i = f_i(pa_i,u_i) \quad PA_i \subseteq V \setminus V_i \qquad U_i \subseteq U$$

Example: Price – Quantity equations in economics

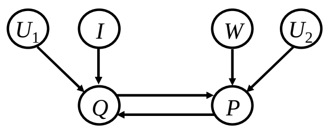


$$q = b_1 p + d_1 i + u_1$$
$$p = b_2 q + d_2 w + u_2$$

STRUCTURAL MODELS AND INTERVENTION

Let X be a set of variables in V.
The action $do(x)$ sets X to constants x regardless of the factors which previously determined X.
$do(x)$ replaces all functions f_i determining X with the constant functions X=x, to create a mutilated model M_x

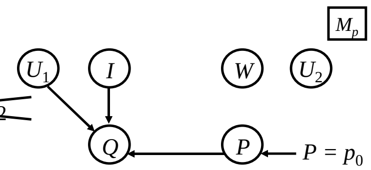
$$q = b_1 p + d_1 i + u_1$$
$$p = b_2 q + d_2 w + u_2$$



STRUCTURAL MODELS AND INTERVENTION

Let X be a set of variables in V.
The action $do(x)$ sets X to constants x regardless of the factors which previously determined X.
$do(x)$ replaces all functions f_i determining X with the constant functions X=x, to create a mutilated model M_x

$$q = b_1 p + d_1 i + u_1$$
~~$p = b_2 q + d_2 w + u_2$~~
$$p = p_0$$



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:

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

The super-distribution P^* is derived from M.
Parsimonous, consistent, and transparent

AXIOMS OF CAUSAL COUNTERFACTUALS

Y would be y, had X been x (in state $U = u$)

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
   $$W_x(u) = w \Rightarrow Y_{xw}(u) = Y_x(u)$$
5. Reversibility
   $$(Y_{xw}(u) = y \,\&\, (W_{xy}(u) = w) \Rightarrow Y_x(u) = y$$

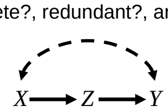
DIFFICULTIES WITH ALGEBRAIC LANGUAGE:

Consider a set of assumptions:

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

Unfriendly:
Consistent?, complete?, redundant?, arguable?

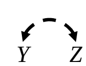
Friendly language:



GRAPHICAL – COUNTERFACTUALS SYMBIOSIS

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

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

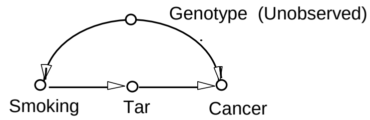


$Y_x \perp\!\!\!\perp Z_y$

consistent, and readable from the graph.

Every theorem in SEM is a theorem in Potential Response Model, and conversely.

DERIVATION IN CAUSAL CALCULUS



$P(c \mid do\{s\}) = \Sigma_t P(c \mid do\{s\}, t)\, P(t \mid do\{s\})$ 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

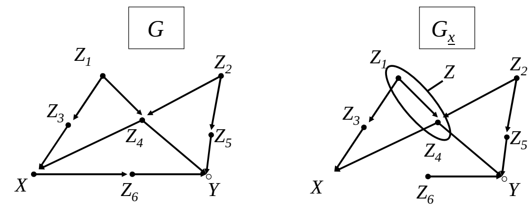
$= \Sigma_{s'} \Sigma_t P(c \mid do\{t\}, s')\, P(s' \mid do\{t\})\, P(t \mid s)$ Probability Axioms

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

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

THE BACK-DOOR CRITERION

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



Moreover, $P(y \mid do(x)) = \sum_z P(y \mid x,z)\, P(z)$
("adjusting" for Z)

RECENT RESULTS ON IDENTIFICATION

- *do*-calculus is complete
- Complete graphical criterion for identifying causal effects (Shpitser and Pearl, 2006).
- Complete graphical criterion for empirical testability of counterfactuals (Shpitser and Pearl, 2007).

STRUCTURAL ANALYSIS: SOME USEFUL RESULTS

1. Complete formal semantics of counterfactuals
2. Transparent language for expressing assumptions
3. Complete solution to causal-effect identification
4. Legal responsibility (bounds)
5. Non-compliance (universal bounds)
6. Integration of data from diverse sources
7. Direct and Indirect effects,
8. Complete criterion for counterfactual testability

DETERMINING THE CAUSES OF EFFECTS (The Attribution Problem)

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

-

DETERMINING THE CAUSES OF EFFECTS (The Attribution Problem)

- 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!
  $PN = P(? \mid A \text{ is dead, took the drug}) \geq 0.50$

THE PROBLEM

Semantical Problem:

1. What is the meaning of $PN(x,y)$:
   "Probability that event y would not have occurred if it were not for event x, given that x and y did in fact occur."

•

THE PROBLEM

Semantical Problem:

1. What is the meaning of $PN(x,y)$:
   "Probability that event y would not have occurred if it were not for event x, given that x and y did in fact occur."

Answer:

$$PN(x,y) = P(Y_{x'} = y' \mid x, y)$$

Computable from M

THE PROBLEM

Semantical Problem:

1. What is the meaning of $PN(x,y)$:
   "Probability that event y would not have occurred if it were not for event x, given that x and y did in fact occur."

Analytical Problem:

2. Under what condition can $PN(x,y)$ be learned from statistical data, i.e., observational, experimental and combined.

TYPICAL THEOREMS
(Tian and Pearl, 2000)

- Bounds given combined nonexperimental and experimental data

$$\max\left\{\begin{array}{c} 0 \\ \dfrac{P(y)-P(y_{x'})}{P(x,y)} \end{array}\right\} \le PN \le \min\left\{\begin{array}{c} 1 \\ \dfrac{P(y'_{x'})}{P(x,y)} \end{array}\right\}$$

- Identifiability under monotonicity (Combined data)

$$PN = \frac{P(y|x)-P(y|x')}{P(y|x)} + \frac{P(y|x')-P(y_{x'})}{P(x,y)}$$

corrected Excess-Risk-Ratio

CAN FREQUENCY DATA DECIDE LEGAL RESPONSIBILITY?

	Experimental		Nonexperimental	
	$do(x)$	$do(x')$	x	x'
Deaths (y)	16	14	2	28
Survivals (y')	984	986	998	972
	1,000	1,000	1,000	1,000

- Nonexperimental data: drug usage predicts longer life
- Experimental data: drug has negligible effect on survival
- Plaintiff: Mr. A is special.
  1. He actually died
  2. He used the drug by choice
- Court to decide (given both data):
  Is it more probable than not that A would be alive but for the drug?

$$PN \triangleq P(Y_{x'} = y' \mid x, y) > 0.50$$

TYPICAL THEOREMS
(Tian and Pearl, 2000)

- Bounds given combined nonexperimental and experimental data

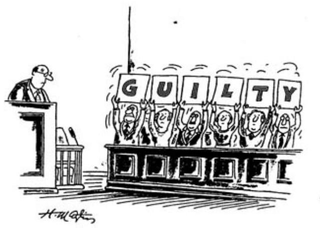
$$\max\left\{\begin{array}{c} 0 \\ \dfrac{P(y)-P(y_{x'})}{P(x,y)} \end{array}\right\} \le PN \le \min\left\{\begin{array}{c} 1 \\ \dfrac{P(y'_{x'})}{P(x,y)} \end{array}\right\}$$

- Identifiability under monotonicity (Combined data)

$$PN = \frac{P(y|x)-P(y|x')}{P(y|x)} + \frac{P(y|x')-P(y_{x'})}{P(x,y)}$$

corrected Excess-Risk-Ratio

SOLUTION TO THE ATTRIBUTION PROBLEM



- WITH PROBABILITY ONE $1 \leq P(y'_{x'} \mid x,y) \leq 1$
- Combined data tell more that each study alone

CONCLUSIONS

Structural-model semantics, enriched with logic and graphs, provides:

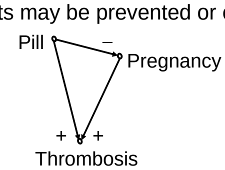
- Complete formal basis for causal reasoning
- Unifies the graphical, potential-outcome and structural equation approaches
- Powerful and friendly causal calculus

DIRECT AND INDIRECT EFFECTS

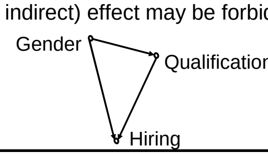
- What is the semantics of direct and indirect effects?
- Can we estimate them from data? Experimental data?

WHY DECOMPOSE EFFECTS?

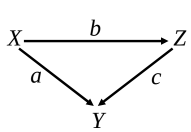
1. Direct (or indirect) effect may be more transportable.
2. Indirect effects may be prevented or controlled.



3. Direct (or indirect) effect may be forbidden



TOTAL, DIRECT, AND INDIRECT EFFECTS HAVE SIMPLE SEMANTICS IN LINEAR MODELS



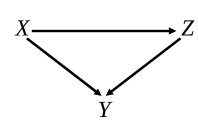
$z = bx + \varepsilon_1$
$y = ax + cz + \varepsilon_2$

$TE \triangleq \frac{\partial}{\partial x} E(Y \mid do(x)) = a + bc$

$DE \triangleq \frac{\partial}{\partial x} E(Y \mid do(x), do(z)) = a$ $\quad Z$ - independent

$IE \triangleq TE - DE = bc$

SEMANTICS BECOMES NONTRIVIAL IN NONLINEAR MODELS
(even when the model is completely specified)



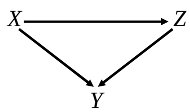
$z = f(x, \varepsilon_1)$
$y = g(x, z, \varepsilon_2)$

$TE \triangleq \frac{\partial}{\partial x} E(Y \mid do(x))$

$DE \triangleq \frac{\partial}{\partial x} E(Y \mid do(x), do(z))$ Dependent on z?

$IE \triangleq$???? Void of operational meaning?

THE OPERATIONAL MEANING OF DIRECT EFFECTS



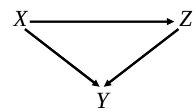
$z = f(x, \varepsilon_1)$
$y = g(x, z, \varepsilon_2)$

"Natural" Direct Effect of *X* on *Y*:
The expected change in *Y* per unit change of *X*, when we keep *Z* constant at whatever value it attains before the change.

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

In linear models, *NDE* = Controlled Direct Effect

THE OPERATIONAL MEANING OF INDIRECT EFFECTS



$z = f(x, \varepsilon_1)$
$y = g(x, z, \varepsilon_2)$

"Natural" Indirect Effect of *X* on *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 under a unit change in *X*.

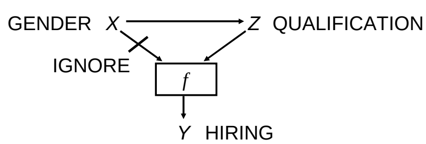
$$E[Y_{x_0 Z_{x_1}} - Y_{x_0}]$$

In linear models, *NIE* = *TE* - *DE*

POLICY IMPLICATIONS (Who cares?)

What is the ~~direct~~ indirect effect of *X* on *Y*?

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



LEGAL DEFINITIONS TAKE THE NATURAL CONCEPTION (FORMALIZING DISCRIMINATION)

``The central question in any employment-discrimination case is whether the employer would have taken the same action had the employee been of different race (age, sex, religion, national origin etc.) and everything else had been the same''

[Carson versus Bethlehem Steel Corp. (70 FEP Cases 921, 7th Cir. (1996))]

x = male, *x′* = female
y = hire, *y′* = not hire
z = applicant's qualifications

NO DIRECT EFFECT

$$Y_{x'Z_x} = Y_x, \quad Y_{xZ_{x'}} = Y_{x'}$$

SEMANTICS AND IDENTIFICATION OF NESTED COUNTERFACTUALS

Consider the quantity

$$Q \triangleq E_u[Y_{xZ_{x^*(u)}}(u)]$$

Given $\langle M, P(u) \rangle$, Q is well defined

Given u, $Z_{x^*}(u)$ is the solution for Z in M_{x^*}, call it z

$Y_{xZ_{x^*(u)}}(u)$ is the solution for Y in M_{xz}

Can Q be estimated from $\left\{ \begin{array}{c} \text{experimental} \\ \text{nonexperimental} \end{array} \right\}$ data?

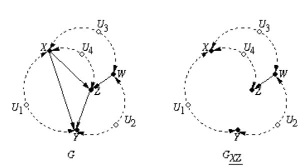
GRAPHICAL CONDITION FOR EXPERIMENTAL IDENTIFICATION OF AVERAGE NATURAL DIRECT EFFECTS

Theorem: If there exists a set *W* such that

$$(Y \perp\!\!\perp Z \mid W)_{G_{\underline{XZ}}} \text{ and } W \subseteq ND(X \cup Z)$$

$$NDE(x, x^*; Y) = \sum_{w,z} [E(Y_{xz} \mid w) - E(Y_{x^*z} \mid w)] P(Z_{x^*} = z \mid w) P(w)$$

Example:



CONCLUSIONS

Structural-model semantics, enriched with logic and graphs, provides:

- Complete formal basis for causal reasoning
- Unifies the graphical, potential-outcome and structural equation approaches
- Powerful and friendly causal calculus