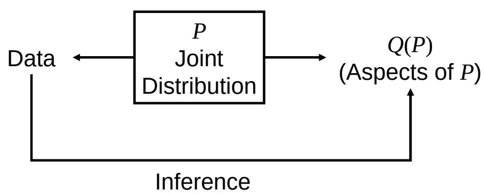
THE MATHEMATICS OF CAUSAL INFERENCE IN STATISTICS

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

OUTLINE

- Modeling: Statistical vs. Causal
- Causal Models and Identifiability
- Inference to three types of claims:
  1. Effects of potential interventions
  2. Claims about attribution (responsibility)
  3. Claims about direct and indirect effects

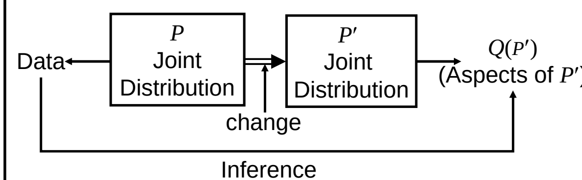
TRADITIONAL STATISTICAL INFERENCE PARADIGM



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

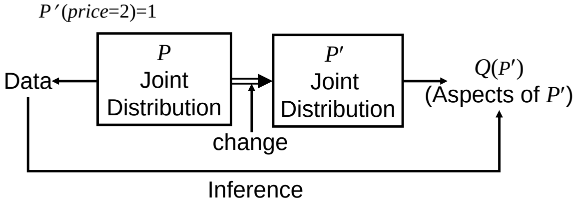
Probability and statistics deal with static relations



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	Propensity score

2.

3.

4.

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	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.

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	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 ($x \;\square\!\!\rightarrow Y$))

WHY CAUSALITY NEEDS SPECIAL MATHEMATICS

SEM Equations are Non-algebraic:

$Y = 2X$	$X = 1$
$X = 1$	$Y = 2$
<u>Process information</u>	<u>Static information</u>

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

FROM STATISTICAL TO CAUSAL ANALYSIS: 2. THE MENTAL BARRIERS

1. Every exercise of causal analysis must rest on untested, judgmental causal assumptions.
2. Every exercise of causal analysis must invoke non-standard mathematical notation.

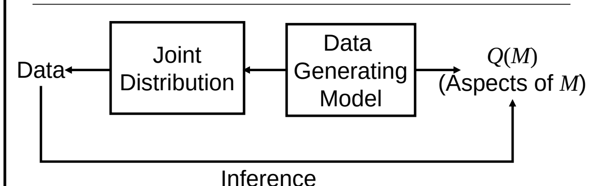
TWO PARADIGMS FOR CAUSAL INFERENCE

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, $Z_y,\ldots$?

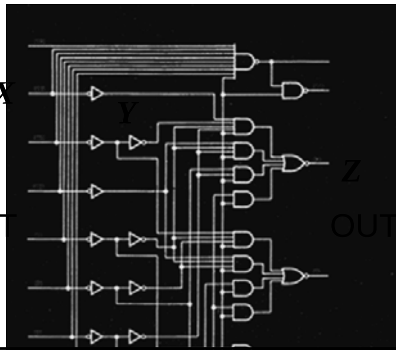
<u>*N-R* model</u>	<u>Structural model</u>
Counterfactuals are primitives, new variables	Counterfactuals are derived quantities
Super-distribution $P^*(X, Y,\ldots, Y_x, X_z,\ldots)$	Subscripts modify a data-generating model
X, Y, Z constrain $Y_x, Z_y,\ldots$	

THE STRUCTURAL MODEL PARADIGM



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

FAMILIAR CAUSAL MODEL ORACLE FOR MANIPILATION



INPUT OUTPUT

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

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. Parsimonious, consistent, and transparent

APPLICATIONS

1. Predicting effects of actions and policies
2. Learning causal relationships from assumptions and data
3. Troubleshooting physical systems and plans
4. Finding explanations for reported events
5. Generating verbal explanations
6. Understanding causal talk
7. Formulating theories of causal thinking

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

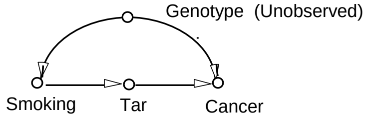
RULES OF CAUSAL CALCULUS

Rule 1: Ignoring observations
$P(y \mid do\{x\}, z, w) = P(y \mid do\{x\}, w)$
if $(Y \perp\!\!\!\perp Z | X,W)_{G_{\overline{X}}}$

Rule 2: Action/observation exchange
$P(y \mid do\{x\}, do\{z\}, w) = P(y \mid do\{x\},z,w)$
if $(Y \perp\!\!\!\perp Z | X,W)_{G_{\overline{X}\underline{Z}}}$

Rule 3: Ignoring actions
$P(y \mid do\{x\}, do\{z\}, w) = P(y \mid do\{x\}, w)$
if $(Y \perp\!\!\!\perp Z | X,W)_{G_{\overline{X}\,\overline{Z(W)}}}$

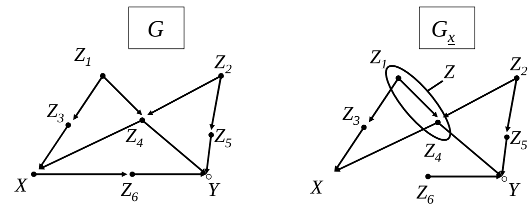
DERIVATION IN CAUSAL CALCULUS



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

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

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

- 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

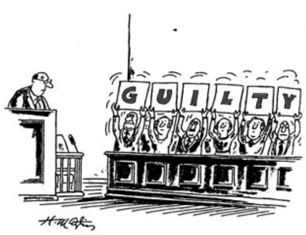
CAN FREQUENCY DATA DECIDE LEGAL RESPONSIBILITY?

	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?

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

SOLUTION TO THE ATTRIBUTION PROBLEM



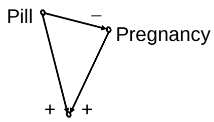
- WITH PROBABILITY ONE $1 \le P(y'_{x'} \mid x,y) \le 1$
- Combined data tell more that each study alone

EFFECT DECOMPOSITION

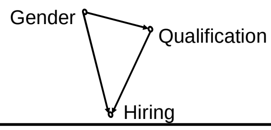
- What is the semantics of direct and indirect effects?
- What are their policy-making implications?
- Can we estimate them from data? Experimental data?

WHY DECOMPOSE EFFECTS?

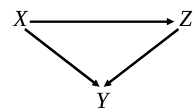
1. Direct (or indirect) effect may be more transportable.
2. Indirect effects may be prevented or controlled.



3. Direct (or indirect) effect may be forbidden



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



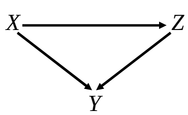
$z = f(x, \varepsilon_1)$
$y = g(x, z, \varepsilon_2)$

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

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

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

THE OPERATIONAL MEANING OF DIRECT EFFECTS



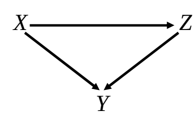
$z = f(x, \varepsilon_1)$
$y = g(x, z, \varepsilon_2)$

"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*.

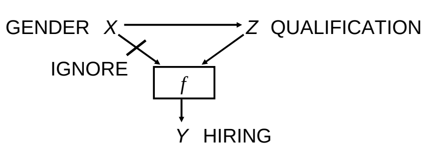
$$E[Y_{x_0 Z_{x_1}} - Y_{x_0}]$$

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

POLICY IMPLICATIONS OF INDIRECT EFFECTS

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

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



SEMANTICS AND IDENTIFICATION OF NESTED COUNTERFACTUALS

Consider the quantity

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

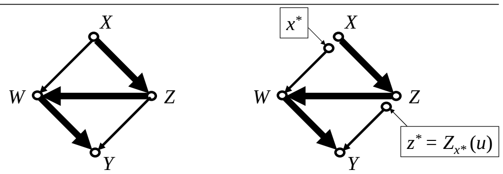
Given $\langle M, P(u) \rangle$, Q is well defined

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

$Y_{x Z_{x^*}(u)}(u)$ 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?

GENERAL PATH-SPECIFIC EFFECTS (Def.)



Form a new model, M_g^*, specific to active subgraph g

$$f_i^*(pa_i, u; g) = f_i(pa_i(g), pa_i^*(\bar{g}), u)$$

Definition: g-specific effect

$$E_g(x, x^*; Y)_M = TE(x, x^*; Y)_{M_g^*}$$

Nonidentifiable even in Markovian models

EFFECT DECOMPOSITION SUMMARY

- Graphical conditions for estimability from experimental / nonexperimental data.
- Graphical conditions hold in Markovian models
- Useful in answering new type of policy questions involving mechanism blocking instead of variable fixing.

CONCLUSIONS

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

- Complete formal basis for causal reasoning
- Powerful and friendly causal calculus
- Lays the foundations for asking more difficult questions: What is an action? What is free will? Should robots be programmed to have this illusion?