CAUSAL INFERENCE IN THE EMPIRICAL SCIENCES

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

• Inference: Statistical vs. Causal distinctions and mental barriers
• Formal semantics for counterfactuals: definition, axioms, graphical representations
• Inference to three types of claims:
  1. Effect of potential interventions
  2. Attribution (Causes of Effects)
  3. Direct and indirect effects
TRADITIONAL STATISTICAL INFERENCE PARADIGM

Data ➔ Joint Distribution ➔ $Q(P)$ (Aspects of $P$)

Inference

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 $P'(\text{price}=2)=1$

Data

$P$
Joint Distribution

change

$P'$
Joint Distribution

$Q(P')$
(Aspects of $P'$)

Inference

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

$P$ does not tell us how it ought to change
e.g. Curing symptoms vs. curing diseases
e.g. Analogy: mechanical deformation

$P'$
Joint Distribution
FROM STATISTICAL TO CAUSAL ANALYSIS: 
1. THE DIFFERENCES (CONT)

1. Causal and statistical concepts do not mix.
   **CAUSAL**
   - Spurious correlation
   - Randomization
   - Confounding / Effect
   - Instrument
   - Holding constant
   - Explanatory variables

   **STATISTICAL**
   - Regression
   - Association / Independence
   - “Controlling for” / Conditioning
   - Odd and risk ratios
   - Collapsibility
   - Propensity score

2.

3.

4.
FROM STATISTICAL TO CAUSAL ANALYSIS:

2. MENTAL BARRIERS

1. Causal and statistical concepts do not mix.

   **CAUSAL**
   - Spurious correlation
   - Randomization
   - Confounding / Effect
   - Instrument
   - Holding constant
   - Explanatory variables

   **STATISTICAL**
   - Regression
   - Association / Independence
   - “Controlling for” / Conditioning
   - Odd and risk ratios
   - Collapsibility
   - Propensity score

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

   \[
   \text{statistical assumptions + data} \quad \Rightarrow \quad \text{causal conclusions}
   \]


4.
FROM STATISTICAL TO CAUSAL ANALYSIS:

2. MENTAL BARRIERS

1. Causal and statistical concepts do not mix.
   - CAUSAL
     - Spurious correlation
     - Randomization
     - Confounding / Effect
     - Instrument
     - Holding constant
     - Explanatory variables
   - STATISTICAL
     - Regression
     - Association / Independence
     - “Controlling for” / Conditioning
     - Odd and risk ratios
     - Collapsibility
     - Propensity score

2. No causes in – no causes out (Cartwright, 1989)
   - statistical assumptions + data
   - causal assumptions
   \[\Rightarrow\] causal conclusions


4. Non-standard mathematics:
   a) Structural equation models (Wright, 1920; Simon, 1960)
   b) Counterfactuals (Neyman-Rubin \( Y_x \), Lewis \( x \rightarrow Y \))
WHY CAUSALITY NEEDS SPECIAL MATHEMATICS

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

e.g., Pricing Policy: “Double the competitor’s price”
Correct notation:

\[ Y = 2X \]

\[ X = 1 \]

\[ Y = 2 \]

Process information

Had \( X \) been 3, \( Y \) would be 6.
If we raise \( X \) to 3, \( Y \) would be 6.
Must “wipe out” \( X = 1 \).
WHY CAUSALITY NEEDS SPECIAL MATHEMATICS

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

e.g., Pricing Policy: “Double the competitor’s price”

Correct notation:
(or)

\[
\begin{align*}
Y & \leftarrow 2X \\
X & = 1 \\
Y & = 2
\end{align*}
\]

Process information

The solution

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

Data Generating Model - Invariant strategy (mechanism, recipe, law, protocol) by which Nature assigns values to variables in the analysis.

Joint Distribution

Inference

$Q(M)$ (Aspects of $M$)

$M$ – Invariant strategy (mechanism, recipe, law, protocol) by which Nature assigns values to variables in the analysis.
FAMILIAR CAUSAL MODEL
ORACLE FOR MANIPILATION
Definition: A **structural causal model** is a 4-tuple $\langle V, U, F, P(u) \rangle$, where

- $V = \{V_1, \ldots, V_n\}$ are observable variables
- $U = \{U_1, \ldots, U_m\}$ are background variables
- $F = \{f_1, \ldots, f_n\}$ 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.
The arguments of the functions $v_i = f_i(v,u)$ define a graph $v_i = f_i(pa_i,u_i)$, $PA_i \subseteq V \setminus V_i$, $U_i \subseteq U$.

Example: Price – Quantity equations in economics:

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

The Fundamental Equation of Counterfactuals:

$$Y_x(u) = Y_{M_x}(u)$$
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:
  \[
P(Y_x = y, Z_w = z) = \sum_{u : X_x(u) = x, Z_w(u) = z} P(u)
  \]
  The Fundamental Equation of Counterfactuals. In particular:
  \[
P(y \mid do(x)) = \frac{\Delta P(Y_x(u) = y, X_x(u) = x)}{P(Y_x(u) = y)} = \sum_{u : Y_x(u) = y} P(u)
  \]
  \[
  PN(Y_{x'} = y' \mid x, y) = \sum_{u : Y_{x'}(u) = y'} P(u \mid x, y)
  \]
AXIOMS OF CAUSAL COUNTERFACTUALS

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

1. Definiteness
   $\exists x \in X \text{ s.t. } X_y(u) = x$

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
   $(Y_{xw}(u) = y \& (W_{xy}(u) = w) \Rightarrow Y_x(u) = y$
INFERRING THE EFFECT OF INTERVENTIONS

The problem:
To predict the impact of a proposed intervention using data obtained prior to the intervention.

The solution (conditional):
Causal Assumptions + Data → Policy Claims

1. Mathematical tools for communicating causal assumptions formally and transparently.

2. Deciding (mathematically) whether the assumptions communicated are sufficient for obtaining consistent estimates of the prediction required.

3. Suggesting (if (2) is negative) a set of measurements and experiments that, if performed, would render a consistent estimate feasible.
Given $P(x,y,z)$, should we ban smoking?

**Linear Analysis**

$x = u_1,$
$z = \alpha x + u_2,$
$y = \beta z + \gamma u_1 + u_3.$

Find: $\alpha \cdot \beta$

**Nonparametric Analysis**

$x = f_1(u_1),$
$z = f_2(x, u_2),$
$y = f_3(z, u_1, u_3).$

Find: $P(y|do(x))$
Given $P(x,y,z)$, should we ban smoking?

**Linear Analysis**
\[
x = u_1, \\
z = \alpha x + u_2, \\
y = \beta z + \gamma u_1 + u_3.
\]
Find: $\alpha \cdot \beta$

**Nonparametric Analysis**
\[
x = \text{const.} \\
z = f_2(x, u_2), \\
y = f_3(z, u_1, u_3).
\]
Find: $P(y|do(x)) \triangleq P(Y=y)$ in new model
Given $P(x,y,z)$, should we ban smoking?

![Diagram showing causal relationships between Smoking, Tar in Lungs, and Cancer.](image)
Given $P(x,y,z)$, should we ban smoking?

**Pre-intervention**

$$P(x,y,z) = \sum_u P(u)P(x | u)P(z | x)P(y | z,u)$$

**Post-intervention**

$$P(y,z | do(x)) = \sum_u P(u)P(z | x)P(y | z,u)$$
EFFECT OF INTERVENTION
AN EXAMPLE (cont)

Given $P(x,y,z)$, should we ban smoking?

Pre-intervention

$$P(x,y,z) = \sum_u P(u)P(x \mid u)P(z \mid x)P(y \mid z,u)$$

Post-intervention

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

To compute $P(y,z \mid do(x))$, we must eliminate $u$. (graphical problem).
ELIMINATING CONFOUNDING BIAS
A GRAPHICAL CRITERION

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

Moreover, \( P(y \mid do(x)) = \sum_z P(y \mid x, z) P(z) \) (“adjusting” for \( Z \))
RULES OF CAUSAL CALCULUS

Rule 1: Ignoring observations
\[ P(y \mid do\{x\}, z, w) = P(y \mid do\{x\}, w) \]
if \((Y \perp Z \mid X, W)_{G_X}\)

Rule 2: Action/observation exchange
\[ P(y \mid do\{x\}, do\{z\}, w) = P(y \mid do\{x\}, z, w) \]
if \((Y \perp Z \mid X, W)_{G_XZ}\)

Rule 3: Ignoring actions
\[ P(y \mid do\{x\}, do\{z\}, w) = P(y \mid do\{x\}, w) \]
if \((Y \perp Z \mid X, W)_{G_XZ(W)}\)
DERIVATION IN CAUSAL CALCULUS

Probability Axioms

\[ P(c \mid do\{s\}) = \sum_t P(c \mid do\{s\}, t) P(t \mid do\{s\}) \]

Rule 2

\[ = \sum_t P(c \mid do\{s\}, do\{t\}) P(t \mid do\{s\}) \]

Rule 2

\[ = \sum_t P(c \mid do\{s\}, do\{t\}) P(t \mid s) \]

Rule 2

\[ = \sum_t P(c \mid do\{t\}) P(t \mid s) \]

Rule 3

\[ = \sum_{s'} \sum_t P(c \mid t, s') P(s' \mid do\{t\}) P(t \mid s) \]

Rule 2

\[ = \sum_{s'} \sum_t P(c \mid t, s') P(s') P(t \mid s) \]

Rule 3

Probability Axioms
Problem:
Predict $P(y \mid do(x))$ from a study in which only $Z$ can be controlled.

Solution:
Determine if $P(y \mid do(x))$ can be reduced to a mathematical expression involving only $do(z)$. 
COMPLETENESS 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).
THE CAUSAL RENAISSANCE: VOCABULARY IN ECONOMICS


Trend
THE CAUSAL RENAISSANCE: 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. Imperfect experiments (universal bounds for IV)
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!

\[ PN = P(\text{?} | A \text{ is dead, took the drug}) > 0.50 \]
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:

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

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

\[
\max \left\{ \frac{0}{P(x,y)} - \frac{P(y|x')}{P(x,y)} \right\} \leq PN \leq \min \left\{ \frac{1}{P(x,y)} \right\}
\]

• 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?

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Nonexperimental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$do(x)$</td>
<td>$do(x')$</td>
</tr>
<tr>
<td>Deaths ($y$)</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Survivals ($y'$)</td>
<td>984</td>
<td>986</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>1,000</td>
</tr>
</tbody>
</table>

- **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?

\[ PN \triangleq P(Y_{x'} = y'| x, y) > 0.50 \]
SOLUTION TO THE ATTRIBUTION PROBLEM

- WITH PROBABILITY ONE \( 1 \leq P(y'_{x'} | x, y) \leq 1 \)
- Combined data tell more than each study alone
EFFECT DECOMPOSITION
(direct vs. indirect effects)

1. Why decompose effects?

2. What is the semantics 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?
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
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!
**NATURAL SEMANTICS OF AVERAGE DIRECT EFFECTS**

Robins and Greenland (1992) – “Pure”

\[ z = f(x, u) \]
\[ y = g(x, z, u) \]

Average 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, \( DE = \text{Controlled Direct Effect} \)
Consider the quantity \( Q \triangleq E_u [Y_{xZx*}(u)] \)

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 \( z \)

\( Y_{xZx*}(u) \) is the solution for \( Y \) in \( M_{xz} \)

Can \( Q \) be estimated from \{ experimental, nonexperimental \} data?

Experimental: nest-free expression
Nonexperimental: subscript-free expression
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 $x_1$.

$$E[Y_{x_0Z_{x_1}} - Y_{x_0}]$$

In linear models, $IE = TE - DE$
What is the indirect effect of \( X \) on \( Y \)?

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

Blocking a link – a new type of intervention
Theorem 5: The total, direct and indirect effects obey the following equality

\[ TE(x, x^*; Y) = DE(x, x^*; Y) - IE(x^*, x; Y) \]

In words, the total effect (on \( Y \)) associated with the transition from \( x^* \) to \( x \) is equal to the difference between the direct effect associated with this transition and the indirect effect associated with the reverse transition, from \( x \) to \( x^* \).
Theorem: If there exists a set $W$ such that

$$Y_{xz} \perp Z_{x^*} \mid W \quad \text{for all } z \text{ and } x$$

Then the average direct effect

$$DE(x, x^*; Y) = E(Y_x, Z_{x^*}) - E(Y_{x^*})$$

Is identifiable from experimental data and is given by

$$DE(x, x^*; Y) = \sum_{w, z} [E(Y_{xz} \mid w) - E(Y_{x^*z} \mid w)]P(Z_{x^*} = z \mid w)P(w)$$
Theorem: If there exists a set $W$ such that

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

then,

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

Example:
GENERAL PATH-SPECIFIC EFFECTS (Def.)

Form a new model, $M^*_g$, specific to active subgraph $g$

$$f^*_i(pa_i,u;g) = f_i(pa_i(g), pa^*_i(\overline{g}), u)$$

Definition: $g$-specific effect

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

Nonidentifiable even in Markovian models
SUMMARY OF RESULTS

1. Formal semantics of path-specific effects, based on signal blocking, instead of value fixing.
2. Path-analytic techniques extended to nonlinear and nonparametric models.
3. Meaningful (graphical) conditions for estimating direct and indirect effects from experimental and nonexperimental data.
CONCLUSIONS

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

• Complete formal basis for causal and counterfactual reasoning
• Unifies the graphical, potential-outcome and structural equation approaches
• Provides friendly and formal solutions to century-old problems and confusions.