CAUSES AND COUNTERFACTUALS IN THE EMPIRICAL SCIENCES

Judea Pearl
University of California
Los Angeles
(www.cs.ucla.edu/~judea/jsm09)

OUTLINE

- Inference: Statistical vs. Causal, distinctions, and mental barriers
- Unified conceptualization of counterfactuals, structural-equations, and graphs
- Inference to three types of claims:
  1. Effect of potential interventions
  2. Attribution (Causes of Effects)
  3. Direct and indirect effects
- Frills

TRADITIONAL STATISTICAL INFERENCE PARADIGM

Data → $P$ 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

Data → $P$ Joint Distribution → $P'$ Joint Distribution → $Q(P')$ (Aspects of $P'$)

Inference

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 (CONT)

1. Causal and statistical concepts do not mix.
   - CAUSAL: Spurious correlation, Randomization / Intervention, Confounding / Effect, Instrumental variable, Strong Exogeneity, Explanatory variables
   - STATISTICAL: Regression, Association / Independence, "Controlling for" / Conditioning, Odd and risk ratios, Collapsibility / Granger causality, Propensity score

2. $P'(v) \neq P(v \mid price = 2)$
   - $P$ 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)

Data → $P$ Joint Distribution → $P'$ Joint Distribution → $Q(P')$ (Aspects of $P'$)

Inference

Note: $P'(v) = P(v \mid price = 2)$

$P$ 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:
2. MENTAL BARRIERS

1. Causal and statistical concepts do not mix.
   CAUSAL
   Spurious correlation
   Randomization / Intervention
   Confounding / Effect
   Instrumental variable
   Strong Exogeneity
   Explanatory variables
   STATISTICAL
   Regression
   Association / Independence
   "Controlling for" / Conditioning
   Odd and risk ratios
   Collapsibility / Granger causality
   Propensity score

2. No causes in – no causes out (Cartwright, 1989)
   statistical assumptions + data
   causal assumptions
   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 \( Y \leftarrow X \), Lewis \( x \rightarrow Y \))

WHY CAUSALITY NEEDS SPECIAL MATHEMATICS

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

\[ Y = 2X \]

Correct notation:

\[ X = 1 \]
\[ Y = 2 \]

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

\[ Q(M) \] (Aspects of \( M \))

\[ \underline{M} \]

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

“Think Nature, not experiment!”

FAMILIAR CAUSAL MODEL
ORACLE FOR MANIPULATION

INPUT

\[ X \]

\[ Y \]

\[ Z \]

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

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

**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 $f_j$ determining $X$ with the constant functions $X=x$, to create a mutilated model $M_x$.

Example: Price – Quantity equations in economics

\[
\begin{align*}
q &= b_1 p + d_1 i + u_1 \\
p &= b_2 q + d_2 w + u_2
\end{align*}
\]

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

COUNTERFACTUALS AT WORK

ETY – EFFECT OF TREATMENT ON THE TREATED

1. Regret:
   I took a pill to fall asleep.
   Perhaps I should not have?

2. Program evaluation:
   What would terminating a program do to those enrolled?

\[ P(Y_x = y | x') \]

THE LOGIC OF CAUSAL ANALYSIS

\( A \) - CAUSAL ASSUMPTIONS

CAUSAL MODEL \( (M) \)

\( A^* \) - Logical implications of \( A \)

Queries of interest \( Q \)

\( Q(P) \) - Identified estimands

Data \( (D) \)

\( \hat{Q} \) - Estimates of \( Q(P) \)

\( g(T) \) - Goodness of fit

Causal inference

Statistical inference

MODEL TESTING

CONDITIONAL CLAIMS

GOODNESS OF FIT

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 \( M \).

\[ \text{ATE} \triangleq E(Y | do(x_1)) - E(Y | do(x_0)) \]

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).

THE FIVE NECESSARY STEPS FOR EFFECT OF TREATMENT ON THE TREATED

Define: Express the target quantity \( Q \) as a function \( Q(M) \) that can be computed from any model \( M \).

\[ \text{ETT} \triangleq P(Y_x = y | X = x') \]

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).

IDENTIFICATION IN SCM

Find the effect of \( X \) on \( Y \), \( P(y | do(x)) \), given the causal assumptions shown in \( G \), where \( Z_1, ..., Z_k \) are auxiliary variables.

Can \( P(y | do(x)) \) be estimated if only a subset, \( Z \), can be measured?
ELIMINATING CONFOUNDING BIAS
THE BACK-DOOR CRITERION

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

Moreover, \( P(y \mid do(x)) = \sum P(y \mid x, z) P(z) = \sum \frac{P(x, y, z)}{P(x \mid z)} \)

("adjusting" for \( Z \) \rightarrow Ignorability)

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

FROM IDENTIFICATION TO ESTIMATION

Define: Express the target quantity \( Q \) as a function \( Q(M) \) that can be computed from any model \( M \).

Assume: Formulate causal assumptions using ordinary scientific language and represent their structural part in graphical form.

Identify: Determine if \( Q \) is identifiable.

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

WHAT PROPENSITY SCORE (PS) PRACTITIONERS NEED TO KNOW

\[ L(z) = P(X = 1 \mid Z = z) \]

\[ \sum P(y \mid z, x) P(z) = \sum P(y \mid l, x) P(l) \]

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 others.
3. In particular, instrumental variables tend to amplify bias.
4. Choosing sufficient set for PS, requires knowledge of the model.

REGRESSION VS. STRUCTURAL EQUATIONS
(THE CONFUSION OF THE CENTURY)

Regression (claimless, nonfalsifiable):
\[ Y = ax + \varepsilon \]

Structural (empirical, falsifiable):
\[ Y = bx + u_Y \]

Claim: (regardless of distributions):
\[ E(Y \mid do(x)) = E(Y \mid do(x), do(z)) = bx \]

The mothers of all questions:
Q. When is \( b \) estimable by regression methods?
A. When all back-door paths are blocked, \( (u_Y \perp X) \)
Q. When would \( b \) equal \( a \)?
A. Graphical criteria available
TWO PARADIGMS FOR CAUSAL INFERENCE

Observed: $P(X, Y, Z, ...)$
Conclusions needed: $P(Y_{x} = y), P(X_{y} = x | Z = z), ...$

How do we connect observables, $X, Y, Z, ...$, to counterfactuals $Y_{x}, X_{y}, Z_{y}, ...$?

N-R model
Counterfactuals are primitives, new variables
Super-distribution
$P^{*}(X, Y, ..., Y_{x}, X_{z}, Z_{y}, ...)$

Structural model
Counterfactuals are derived quantities
Subscripts modify the model and distribution

$P(Y_{x} = y) = P_{M_{x}}(Y = y)$

“SUPER” DISTRIBUTION IN N-R MODEL

<table>
<thead>
<tr>
<th>$X$</th>
<th>$Y$</th>
<th>$Z$</th>
<th>$Y_{x=0}$</th>
<th>$X_{z=1}$</th>
<th>$X_{z=1}$</th>
<th>$X_{y=0}$</th>
<th>$U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$u_{1}$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>$u_{2}$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$u_{3}$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>$u_{4}$</td>
</tr>
</tbody>
</table>

Inconsistency: $x = 0 \Rightarrow Y_{x=0} = Y = x Y + (1-x) Y_{0}$

Defines:

$P^{*}(X, Y, Z, ..., Y_{x}, Z_{y}, Y_{x}, Z_{x}, Z_{y}, ...)$

$P^{*}(Y_{x} = y | Z, X_{z})$

$Y_{x} \perp X | Z_{y}$

AXIOMS OF STRUCTURAL COUNTERFACTUALS

$Y_{x}(u)=y$: $Y$ would be $y$, had $X$ been $x$ (in state $U = u$) (Galles, Pearl, Halpern, 1998):

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
   \[ X_{y}(u) = x \]
4. Composition (generalized consistency)
   \[ X_{y}(u) = x \Rightarrow Y_{w}(u) = Y_{w}(u) \]
5. Reversibility
   \[ (Y_{w}(u) = y) \& (W_{xy}(u) = w) \Rightarrow Y_{w}(u) = y \]

FORMULATING ASSUMPTIONS THREE LANGUAGES

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

2. Counterfactuals:
   \[ Z_{x}(u) = Z_{x}(u), \]
   \[ X_{y}(u) = X_{y}(u) = X_{y}(u) = X(u), \]
   \[ Y_{z}(u) = Y_{z}(u), \]
   \[ Z_{x} \perp Y_{y}, X \]

3. Structural:
   \[ x = f_{1}(u, e_{1}) \]
   \[ z = f_{2}(x, e_{2}) \]
   \[ y = f_{3}(z, u, e_{3}) \]

GRAPHICAL – COUNTERFACTUALS SYMBIOSIS

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

1. Missing arrows $Y \not\rightarrow Z$
2. Missing arcs consistent, and readable from the graph.
   \[ \text{• Express assumption in graphs} \]
   \[ \text{• Derive estimands by graphical or algebraic methods} \]

ARE THE TWO PARADIGMS EQUIVALENT?

• Yes (Galles and Pearl, 1998; Halpern 1998)
• In the N-R paradigm, $Y_{x}$ is defined by consistency:
  \[ Y = x Y_{1} + (1-x) Y_{0} \]
• In SCM, consistency is a theorem.
• Moreover, a theorem in one approach is a theorem in the other.
• Difference: Clarity of assumptions and their implications
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!

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

THE ATtribution PROBLEM

Definition:

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:

\[ PN(x, y) = P(Y' = y' | x, y) \]

Computable from \( M \)

THE ATTRIBUTION PROBLEM

Definition:

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.”

Identification:

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

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

• Identifiability under monotonicity (Combined data)

\[
PN = \frac{P'(y | x) - P'(y' | x)}{P'(y | x)} \leq \frac{P'(y' | x) - P'(y | x)}{P'(y | x)}
\]

corrected Excess-Risk-Ratio

CAN FREQUENCY DATA DECIDE LEGAL RESPONSIBILITY?

<table>
<thead>
<tr>
<th>Experimental ( \hat{d}(x) )</th>
<th>Nonexperimental ( \hat{d}(x') )</th>
<th>( \frac{\hat{d}}{\hat{d}'} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deaths (( y ))</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Survivals (( y' ))</td>
<td>984</td>
<td>986</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{Deaths (\( y \))} & = 16, 14, 28 \\
\text{Survivals (\( y' \))} & = 984, 986, 998 \\
\text{Combined data tell more that each study alone}
\end{align*}
\]

SOLUTION TO THE ATTRIBUTION PROBLEM

• WITH PROBABILITY ONE \( 1 \leq P(y' | x, y) \leq 1 \)

• Combined data tell more that each study alone
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?

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?

\[
\begin{array}{c}
X \\
\Rightarrow \\
Z \\
\Rightarrow \\
Y
\end{array}
\]

(Gender) X (Qualifications) 
(hiring)

What is the direct effect of \(X\) on \(Y\)?
\[
E(Y | do(x1), do(z)) - E(Y | do(x0), do(z))
\]
(averaged over \(z\)) Adjust for \(Z\)? No! No!

NATURAL INTERPRETATION
OF AVERAGE DIRECT EFFECTS

Robins and Greenland (1992) – “Pure”

\[
\begin{array}{c}
X \\
\Rightarrow \\
Z \\
\Rightarrow \\
Y
\end{array}
\]

\[
\begin{align*}
z &= f(x, u) \\
y &= g(x, z, u)
\end{align*}
\]

Natural Direct Effect of \(X\) on \(Y\):
\[
DE(x_0, x_1; Y) = E[Y_{x_1}Z_{x_0} - Y_{x_0}]
\]
In linear models, \(DE = \text{Controlled Direct Effect}\)

DEFINITION AND IDENTIFICATION
OF NESTED COUNTERFACTUALS

Consider the quantity
\[
Q = E_u[Y_{XZ^*}(u)]
\]

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

Given \(u, Z^*_X(u)\) is the solution for \(Z\) in \(M^u\), call it \(z\)
\[
Y_{Z^*_X(u)}(u)
\]

Given \(u, Z^*_X(u)\) is the solution for \(Y\) in \(M_{xz}\)
Can \(Q\) be estimated from \(\{\text{experimental} \cup \text{nonexperimental}\}\) data?

Experimental: nest-free expression
Nonexperimental: subscript-free expression

DEFINITION OF
INDIRECT EFFECTS

\[
\begin{array}{c}
X \\
\Rightarrow \\
Z \\
\Rightarrow \\
Y
\end{array}
\]

\[
\begin{align*}
z &= f(x, u) \\
y &= g(x, z, u)
\end{align*}
\]

Indirect Effect of \(X\) on \(Y\):
\[
IE(x_0, x_1; Y) = E[Y_{x_1}Z_{x_0} - Y_{x_0}]
\]
In linear models, \(IE = TE - DE\)
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.

GENDER X → Z QUALIFICATION

Blocking a link – a new type of intervention

MEDIATION FORMULAS

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

\[ DE = \sum_z [E(Y | do(x_1, z)) - E(Y | do(x_0, z))] P(z | do(x_0)) \]
\[ IE = \sum_z E(Y | do(x_0, z))[P(z | do(x_1)) - P(z | do(x_0))] \]
\[ TE = DE - EI \]

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

WHY TE ≠ DE + IE

In linear systems

\[ IE(\text{rev}) = -IE \]
\[ TE = \beta + m_1 m_2 \]
\[ DE = \beta \]
\[ IE = m_1 m_2 = TE - DE \]

\[ IE = \text{Effect sustained by mediation alone} \]
\[ TE - DE = \text{Effect prevented by disabling mediation} \]

TRANSPORTABILITY -- WHEN CAN WE EXTRAPOLATE EXPERIMENTAL FINDINGS TO DIFFERENT POPULATIONS?

Experimental study in LA
Measure: \( P(x, y, z) \)
Problem: We find \( P(z) \neq P^*(z) \)
(LA population is younger)
What can we say about \( P^*(y | do(x)) \)?
Intuition: \( P^*(y | do(x)) = \sum_z P(y | do(x), z) P^*(z) \)

Observational study in NYC
Measure: \( P(y | do(x), z) \)

MEDIATION FORMULAS IN UNCONFOUNDED MODELS

\[ DE = \sum_z [E(Y | x_1, z) - E(Y | x_0, z)] P(z | x_0) \]
\[ IE = \sum_z E(Y | x_0, z)[P(z | x_1) - P(z | x_0)] \]
\[ TE = E(Y | x_1) - E(Y | x_0) \]

\[ IE = \text{Fraction of responses explained by mediation} \]
\[ TE - DE = \text{Fraction of responses owed to mediation} \]

TRANSPORT FORMULAS DEPEND ON THE STORY

a) \( Z \) represents age
\[ P^*(y | do(x)) = \sum_z P(y | do(x), z) P^*(z) \]
b) \( Z \) represents language skill
\[ P^*(y | do(x)) = ?? \]
c) \( Z \) represents a bio-marker
\[ P^*(y | do(x)) = ?? \]
TRANSPORTABILITY
(Pearl and Bareinboim, 2010)

Definition 1 (Transportability)
Given two populations, denoted \(\Pi\) and \(\Pi^*\), characterized by probability distributions \(P\) and \(P^*\), and causal diagrams \(G\) and \(G^*\), respectively, a causal relation \(R\) is said to be transportable from \(\Pi\) to \(\Pi^*\) if
1. \(R(\Pi)\) is estimable from the set \(I\) of interventional studies on \(\Pi\), and
2. \(R(\Pi^*)\) is identified from \(I\), \(P\), \(P^*\), \(G\), and \(G^*\).

TRANSPORT FORMULAS DEPEND ON THE STORY

\(a)\) \(Z\) represents age
\(P^* (y \mid do(x)) = \sum_z P(y \mid do(x), z)P^*(z)\)

\(b)\) \(Z\) represents language skill
\(P^* (y \mid do(x)) = P(y \mid do(x))\)

\(c)\) \(Z\) represents a bio-marker
\(P^* (y \mid do(x)) = \sum_z P(y \mid do(x), z)P^*(z \mid x)\)

WHICH MODEL LICENSES THE TRANSPORT OF THE CAUSAL EFFECT

\(X\ Y\ S\)
\(X\ Y\ S\)
\(X\ Y\ S\)
\(X\ W\ Y\ S\)
\(X\ W\ Z\ Y\ S\)

DETERMINE IF THE CAUSAL EFFECT IS TRANSPORTABLE

What measurements need to be taken in the study and in the target population?

The transport formula
\(P^* (y \mid do(x)) = \sum_z P(y \mid do(x), z)\sum w P(w \mid do(x), z) P^*(z \mid x)\)

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.
- No other method can do better (theorem)

CONCLUSIONS

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

(Aristotle 384-322 B.C.)
QUESTIONS???

They will be answered