**CAUSES AND COUNTERFACTUALS:**
CONCEPTS, PRINCIPLES AND TOOLS

Judea Pearl
Elias Bareinboim
University of California, Los Angeles
{judea, eb}@cs.ucla.edu

NIPS 2013 Tutorial

---

**OUTLINE**

Concepts:
* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:
* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:
* Policy evaluation
* Transportability
* Mediation
* Missing Data

---

**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

Data → Joint Distribution → \( Q(P') \) (Aspects of \( P' \)) → Inference

* 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.

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

---

**FROM STATISTICAL TO CAUSAL ANALYSIS:**

1. THE DIFFERENCES

Data → Joint Distribution → Joint Distribution → \( Q(P') \) (Aspects of \( P' \)) → Inference

* Note: \( P'(sales) \neq P(sales \mid price = 2) \)
  * e.g., Doubling price \neq \) seeing the price doubled.
  * \( P \) does not tell us how it ought to change.

---

**FROM STATISTICAL TO COUNTERFACTUALS:**

1. THE DIFFERENCES

Data → Joint Distribution → Joint Distribution → \( Q(P') \) (Aspects of \( P' \)) → Inference

* 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.
### THE STRUCTURAL MODEL PARADIGM

- **Joint Distribution**
- **Data Generating Model**
- **Q(M)** (Aspects of M)

\[ P \rightarrow Q(M) \rightarrow M \]

Inference

\( M \) – 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?) \( P(y | A) \)
- **Action Questions:**
  - “What if we do A?” (What if?) \( P(y | do(A)) \)
- **Counterfactuals Questions:**
  - “What if we did things differently?” (Why?) \( P(y_{A'} | A) \)
- **Options:**
  - “With what probability?”

SYNTACTIC DISTINCTION

---

### FROM STATISTICAL TO CAUSAL ANALYSIS: 2. THE SHARP BOUNDARY

1. Causal and associational concepts do not mix.
   - **CAUSAL**
     - Spurious correlation
     - Randomization / Intervention
     - “Holding constant” / “Fixing”
     - Confounding / Effect
     - Instrumental variable
     - Ignorability / Exogeneity
   - **ASSOCIATIONAL**
     - Regression
     - Association / Independence
     - “Controlling for” / Conditioning
     - Odds and risk ratios
     - Collapsibility / Granger causality
     - Propensity score

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

\[ data \Rightarrow \text{causal conclusions} \]


4. Non-standard mathematics:
   - Structural equation models (Wright, 1920; Simon, 1960)
   - Counterfactuals (Neyman-Rubin \( \{Y_i\} \), Lewis \( \{\text{D} \Rightarrow Y\} \))

---

### THE NEW ORACLE: STRUCTURAL CAUSAL MODELS

**THE WORLD AS A COLLECTION OF SPRINGS**

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 \( V \),
  - \( y = f(v, u) \) e.g., \( y = \alpha + \beta x + u \)
  - Not regression!!!
- \( P(u) \) is a distribution over \( U \)

\( P(u) \) and \( F \) induce a distribution \( P(v) \) over observable variables
The Fundamental Equation of Interventions:

\[ M \xymatrix{ X(u) \ar[r]^U & Y(u) \ar[r] & \Delta Y(u) } \]

The Fundamental Equation of Counterfactuals:

\[ Y_u(u) \Delta = \Delta Y_{M_u}(u) \]

COUNTERFACTUALS ARE EMBARRASSINGLY SIMPLE

Definition:
Given a SCM model \( M \), the potential outcome \( Y_u(u) \) for unit \( u \) is equal to the solution for \( Y \) in a mutilated model \( M_u \), in which the equation for \( X \) is replaced by \( X = x \).

\[ \begin{align*}
M & \xymatrix{ X(u) \ar[r]^U & Y(u) \ar[r] & \Delta Y(u) } \\
M_u & \xymatrix{ X(u) \ar[r]^U & Y(u) \ar[r] & \Delta Y_u(u) }
\end{align*} \]

The Fundamental Equation of Interventions:

\[ P(Y = y \mid do(X = x)) = P_{M_u}(Y = y) \]

COMPUTING THE EFFECTS OF INTERVENTIONS

\[ \begin{align*}
P(Y = y \mid do(X = x)) & = P_{M_u}(Y = y) \\
P(x, y, u) & = P(u)P(x \mid y, u) \\
P(y, u \mid do(x)) & = P(u)P(y \mid x, u) \quad \text{Truncated product} \\
P(y \mid do(x)) & = \sum_u P(y \mid x, u)P(u) \quad \text{Adjustment formula}
\end{align*} \]

THE TWO FUNDAMENTAL LAWS OF CAUSAL INference

1. The Law of Counterfactuals (and Interventions)

\[ Y_u(u) = Y_{M_u}(u) \]

(M generates and evaluates all counterfactuals.)

2. The Law of Conditional Independence (\( \perp \mid \))

\[ (X \perp Y \mid Z)_{M(M)} \Rightarrow (X \perp Y \mid Z)_{P(v)} \]

(Separation in the model \( \Rightarrow \) independence in the distribution.)

THE LAW OF CONDITIONAL INDEPENDENCE

<table>
<thead>
<tr>
<th>( S ) (Sprinkler)</th>
<th>( C ) (Climate)</th>
<th>( R ) (Rain)</th>
<th>( W ) (Wetness)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph (G)</td>
<td>Model (M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( C = \Phi(U_C) )</td>
<td>( S = \Phi(C \cup \mathcal{Z}) )</td>
<td>( R = \Phi(C \cup \mathcal{R}) )</td>
<td>( W = \Phi(S \cup R \cup U_W) )</td>
</tr>
</tbody>
</table>

Gift of the Gods

If the \( U \)'s are independent, the observed distribution \( P(C \cup R \cup S \cup W) \) satisfies constraints that are:

1. independent of the \( Y \)'s and of \( P(U) \),
2. readable from the graph.

\( \perp \):

- Every missing arrow advertises an independency, conditional on a separating set.
- For example, \( C \perp W \mid S, R \)

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

\( \Rightarrow \) : nature’s language for communicating its structure

\( \Phi(C \cup \mathcal{U}_C) \) : \( C \) is a function of \( \mathcal{U}_C \), a set of variables

\( \Phi(C \cup \mathcal{R}) \) : \( C \) is a function of \( \mathcal{R} \), a set of variables

\( \Phi(C \cup \mathcal{Z}) \) : \( C \) is a function of \( \mathcal{Z} \), a set of variables

\( \Phi(C \cup \mathcal{W}) \) : \( C \) is a function of \( \mathcal{W} \), a set of variables

\( \Phi(S \cup R \cup U_W) \) : \( S \cup R \cup U_W \) is a function of \( \Phi \), a set of variables

\( \Phi(C \cup \mathcal{R}) \) : \( C \) is a function of \( \mathcal{R} \), a set of variables

\( \Phi(C \cup \mathcal{Z}) \) : \( C \) is a function of \( \mathcal{Z} \), a set of variables
OUTLINE

Concepts:
* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:
* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:
* Policy evaluation
* Transportability
* Mediation
* Missing Data

FIRST LAYER OF THE CAUSAL HIERARCHY

PROBABILITIES
(What if I see $X=x$?)

THE EMERGENCE OF THE FIRST LAYER

Data $\rightarrow$ Joint Distribution $\rightarrow$ Data Generating Model $\rightarrow$ $P(v)$ $\rightarrow$ $M$

Theorem (PV, 1991). Every Markovian structural causal model $M$ (recursive, with independent disturbances) induces a passive distribution $P(v_1, \ldots, v_n)$ that can be factorized as

$$P(v_1, v_2, \ldots, v_n) = \prod_j P(v_j \mid p_{a_j})$$

where $p_{a_j}$ are the (values of) the parents of $V_j$ in the causal diagram associated with $M$.

THE SECOND LAYER ON CAUSAL HIERARCHY:

CAUSAL EFFECTS
(What if I do $X=x$?)

TOOL 1. GRAPH SEPARATION
(D-SEPARATION)

(normal valve  abnormal valve)

$X \rightarrow z \rightarrow y$  $X \rightarrow z \leftarrow y$  $X \rightarrow z \rightarrow y$

$X \rightarrow z \leftarrow y$  $X \rightarrow z \rightarrow y$

$X \rightarrow z \rightarrow y$  $\uparrow w$

$X$ Cls: (Wet $\rightarrow$ Sprinkler)
$X$ Cls: (Wet $\rightarrow$ Season $\mid$ Sprinkler)
$\checkmark$ Cls: (Rain $\rightarrow$ Slippery $\mid$ Wet)
$\checkmark$ Cls: (Season $\rightarrow$ Wet $\mid$ Sprinkler, Rain)
$X$ Cls: (Sprinkler $\rightarrow$ Rain $\mid$ Season, Wet)

season  sprinkler  rain

wet  slippery
Observation 1:
The distribution alone tells us nothing about change; it just describes static conditions of a population (under a specific regime).

Observation 2:
We need to be able to represent “change,” or how the population reacts when it undergoes change in regimes.

CAUSAL INFEERENCE: MOVING BETWEEN REGIMES

- What happens when $P$ changes?
e.g., Infer whether less people would get cancer if we ban smoking.

$Q(P) = P(\text{Cancer} = \text{true} \mid \text{do(Smoking} = \text{no}))$

Not an aspect of $P$.

THE BIG PICTURE: THE CHALLENGE OF CAUSAL INFERENCE

- Goal: how much $Y$ changes with $X$ if we vary $X$ between two different constants free from the influence of $Z$.
- This is the definition of causal effect.

METHOD FOR COMPUTING CAUSAL EFFECTS: RANDOMIZED EXPERIMENTS

PROBLEM 1. COMPUTING EFFECTS FROM OBSERVATIONAL DATA

Questions:
* What is the relationship between $P(z, x, w, y)$ and $P(y \mid \text{do}(x))$?
* Is $P(y \mid \text{do}(x)) = P(y) x$?
COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:
Q₁ = Pr(wet | Sprinkler = on)
Q₂ = Pr(wet | do(Sprinkler = on))

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:
Q₁ = Pr(wet | Sprinkler = on)
Q₂ = Pr(wet | do(Sprinkler = on))

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:
Q₁ = Pr(wet | Sprinkler = on)
Q₂ = Pr(wet | do(Sprinkler = on))

COMPUTING CAUSAL EFFECTS FROM OBSERVATIONAL DATA

Queries:
Q₁ = Pr(wet | Sprinkler = on)
Q₂ = Pr(wet | do(Sprinkler = on))

= P(p₁) + P(p₂)

= P(p₁)

= P(p₁) + P(p₂)

∑ₚₑ,ₚₛ,ₚ₁,ₚ₂ P(Se) P(Sp | Se) P(Ra | Se) P(We | Sp, Ra) P(Sl | We)
OUTLINE

Concepts:
* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:
* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:
* Policy evaluation
* Transportability
* Mediation
* Missing Data

TOOL 2. TRUNCATED FACTORIZATION PRODUCT (OPERATIONALIZING INTERVENTIONS)

Corollary (Truncated Factorization, Manipulation Thm., G-comp.):
The distribution generated by an intervention \( \text{do}(X = x) \)
in a Markovian model \( M \) is given by the truncated factorization:

\[
P(v_1, v_2, \ldots, v_n | \text{do}(x)) = \prod_{i | V_i \notin X} P(v_i | pa_i) \bigg| x = x
\]

OUTLINE

Concepts:
* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:
* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:
* Policy evaluation
* Transportability
* Mediation
* Missing Data

NO FREE LUNCH: ASSUMPTIONS ENCODED IN CBNs

Definition (Causal Bayesian Network):

\( P(v) \): observational distribution
\( P(v | \text{do}(x)) \): experimental distribution
\( P^* \): set of all observational and experimental distributions

A DAG \( G \) is called a Causal Bayesian Network compatible with \( P^* \) if and only if the following three conditions hold for every \( P(v | \text{do}(x)) \in P^* \):

i. \( P(v | \text{do}(x)) \) is Markov relative to \( G \);
ii. \( P(v_i | \text{do}(x)) = 1 \), for all \( V_i \notin X \);
iii. \( P(v_i | pa_i, \text{do}(x)) = P(v_i | pa_i) \), for all \( V_i \notin X \).

IF SEASON IS LATENT, IS THE EFFECT STILL COMPUTABLE?

Queries:

\( Q_1 = P(\text{wet} | \text{Sprinkler = on}) = P(p_1) + P(p_2) \)

\( Q_2 = P(\text{wet} | \text{do}(<\text{Sprinkler = on}>) = P(p_1) \)

\[
\sum_{\text{Se,Ra,Sl}} P(\text{Se}) P(p_1 | \text{Se}) P(\text{Ra} | \text{Se}) P(\text{We} | \text{Sp}, \text{Ra}) P(\text{Sl} | \text{We})
\]

\( = \sum_{\text{Se}, \text{Ra}} P(\text{We} | \text{Sp}, \text{Se}) P(\text{Se}) = \sum_{\text{Se}} P(\text{We} | \text{Sp}, \text{Ra}) P(\text{Ra}) \)

Adjustment formula

TOOL 3. BACK-DOOR CRITERION (THE PROBLEM OF CONFOUNDING)

Goal: Find the effect of \( X \) on \( Y \), \( P(y | \text{do}(x)) \), given measurements on auxiliary variables \( Z_1, \ldots, Z_k \)

Diagram:

\( X \)
\( Y \)
\( Z_1 \)
\( Z_2 \)
\( Z_3 \)
\( G \)
ELIMINATING CONFOUNDING BIAS
THE BACK-DOOR CRITERION

\( P(y \mid do(x)) \) is estimable if there is a set \( Z \) of variables that \( d \)-separates \( X \) from \( Y \) in \( G \).

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

GOING BEYOND ADJUSTMENT

Goal: Find the effect of \( S \) on \( C, P(c \mid do(s)) \), given measurements on auxiliary variable \( T \), and when latent variables confound the relationship \( S-C \).

OUTLINE

- Causal inference – a paradigm shift
- The two fundamental laws

Basic tools:
- Graph separation
- The truncated product formula
- The back-door adjustment formula
- The do-calculus

Capabilities:
- Policy evaluation
- Transportability
- Mediation
- Missing Data

TOOL 3. CAUSAL CALCULUS
(IDENTIFIABILITY REDUCED TO CALCULUS)

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

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

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

- Rule 3: Ignoring actions
  \[ P(y \mid do(x), do(z), w) = P(y \mid do(x), w) \]
  \( \text{if } (Y \perp Z \mid X, W) \perp \!
\)

TECHNICAL NOTE.
THE IDENTIFIABILITY PROBLEM

ID PROBLEM (decision): Given two models \( M_1 \) and \( M_2 \), compatible with \( G \) that agree on the observable distribution over \( V \), \( P_i(v) = P_i(v) \), decide whether they also agree in the target quantity \( Q = P(Y \mid do(x)) \), i.e., whether the effect \( P(Y \mid do(x)) \) is identifiable from \( G \) and \( P(v) \).
WHAT CAN EXPERIMENTS ON DIET REVEAL ABOUT THE EFFECT OF CHOLESTEROL ON HEART ATTACK?

G: $Z \rightarrow X \rightarrow Y$ $\leftarrow Z$

$Z$: Diet
$X$: Cholesterol level
$Y$: Heart Attack

Measured:
- Observational study: $P(x, y, z)$
- Experimental study: $P(x, y \mid do(z))$

Needed: $Q = P(y \mid do(x)) = ?$ $= P(x, y \mid do(z)) / P(x \mid do(z))$

(i.e., $\exists f, f: P(v), P(v \mid do(z)) \rightarrow P(y \mid do(x))$)

OUTLINE

Concepts:
* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:
* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:
* Policy evaluation
* Transportability
* Mediation
* Missing Data

SUMMARY OF POLICY EVALUATION RESULTS

- The estimability of any expression of the form
  $Q = P(y_1, y_2, \ldots, y_n \mid do(x_1, x_2, \ldots, x_m), z_1, z_2, \ldots, z_k)$
  can be determined given any causal graph $G$ containing measured and unmeasured variables.
- If $Q$ is estimable, then its estimand can be derived in polynomial time (by estimable we mean either from observational or from experimental studies.)
- The algorithm is complete.
- The causal calculus is complete for this task.

OUTLINE

Concepts:
* Causal inference – a paradigm shift
* The two fundamental laws

Basic tools:
* Graph separation
* The truncated product formula
* The back-door adjustment formula
* The do-calculus

Capabilities:
* Policy evaluation
* Transportability
* Mediation
* Missing Data

PROBLEM 2. GENERALIZABILITY AMONG POPULATIONS BREAK (TRANSPORTABILITY)

Question:

Is it possible to predict the effect of $X$ on $Y$ in a certain population $\Pi^*$, where no experiments can be conducted, using experimental data learned from a different population $\Pi$?

Answer: Sometimes yes.

HOW THIS PROBLEM IS SEEN IN OTHER SCIENCES? (e.g., external validity, meta-analysis, ...)

- Extrapolation across studies requires “some understanding of the reasons for the differences.” (Cox, 1958)
- “External validity’ asks the question of generalizability: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?” (Shadish, Cook and Campbell, 2002)
- “An experiment is said to have “external validity” if the distribution of outcomes realized by a treatment group is the same as the distribution of outcome that would be realized in an actual program.” (Manski, 2007)
**MOVING FROM THE “LAB” TO THE “REAL WORLD” ...**

![Diagram](image)

Everything is assumed to be the same, trivially transportable!

Everything is assumed to be different, not transportable...

**TRANSPORT FORMULAS DEPEND ON THE CAUSAL 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
\[ P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z \mid x) \]

**TRANSPORTABILITY REDUCED TO CALCULUS**

Theorem
A causal relation \( R \) is transportable from \( \Pi \) to \( \Pi' \) if and only if it is reducible, using the rules of \( do \)-calculus, to an expression in which \( S \) is separated from \( do() \).

\[
R(\Pi') = P^*(y \mid do(x)) = P(y \mid do(x), s) = \sum_w P(y \mid do(x), s, w) P(w \mid do(x), s) = \sum_w P(y \mid do(x), w) P(w | s) = \sum_w P(y \mid do(x), w) P^*(w)
\]

**RESULT: ALGORITHM TO DETERMINE IF AN EFFECT IS TRANSPORTABLE**

Input: Annotated Causal Graph
- 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 $X \rightarrow Y$

FROM META-ANALYSIS TO META-SYNTHESIS

The problem
How to combine results of several experimental and observational studies, each conducted on a different population and under a different set of experimental conditions, so as to construct an aggregate measure of effect size that is "better" than any one study in isolation.

META-SYNTHESIS AT WORK

SUMMARY OF TRANSPORTABILITY RESULTS

• Nonparametric transportability of experimental results from multiple environments and limited experiments 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.
• The causal calculus is complete for this task.

OUTLINE

CONCEPTS:
• Causal inference – a paradigm shift
• The two fundamental laws

BASIC TOOLS:
• Graph separation
• The truncated product formula
• The back-door adjustment formula
• The do-calculus

CAPABILITIES:
• Policy evaluation
• Transportability
• Mediation
• Missing Data

MEDIATION: A GRAPHICAL-COUNTERFACTUAL SYMBIOSIS

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: deactivate a mechanism, rather than fix a variable

LEGAL IMPLICATIONS OF DIRECT EFFECT

Can data prove an employer guilty of hiring discrimination?

\[ \text{(Gender) } X \rightarrow Z \rightarrow Y \text{ (Hiring)} \]

What is the direct effect of \( X \) on \( Y \)? (CDE)

\[ E(Y|do(x_1), do(z)) - E(Y|do(x_0), do(z)) \]

(z-dependent) Adjust for \( Z \)? No! No!

Identification is completely solved (Tian & Shpitser 2006)

NATURAL INTERPRETATION OF AVERAGE DIRECT EFFECTS


\[ X \rightarrow Z \rightarrow Y \]

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 \( Z \) constant at whatever value it attained before the change.

\[ E(Y_{x_1|Z=x_0} - Y_{x_0}) \]

In linear models, \( DE = \text{Controlled Direct Effect} = \beta(x_1 - x_0) \)

DEFINITION OF INDIRECT EFFECTS

\[ X \rightarrow Z \rightarrow Y \]

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 \( x_1 \).

\[ E(Y_{x_1|Z=x_1} - Y_{x_0}) \]

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.

\[ \text{GENDER } X \rightarrow Z \rightarrow Y \text{ QUALIFICATION} \]

Deactivating a link – a new type of intervention

THE MEDIATION FORMULAS IN UNCONFONDED MODELS

\[ X \rightarrow Z \rightarrow Y \]

\[ z = f(x, u_1) \]

\[ y = g(x, z, u_2) \]

\[ u_1 \text{ independent of } u_2 \]

\[ DE = \sum_z [E(Y|X_1, z) - E(Y|X_0, z)]P(z|X_0) \]

\[ IE = \sum_z [E(Y|X_0, z)]P(z|X_1) - P(z|X_0)] \]

\[ TE = E(Y|X_1) - E(Y|X_0) \]

\[ IE = \text{Fraction of responses explained by mediation (sufficient)} \]

\[ TE - DE = \text{Fraction of responses owed to mediation (necessary)} \]
THE MEDIATION FORMULAS IN UNCONFOUNDED MODELS

\[ z = f(x, u_1) \]
\[ y = g(x, z, u_2) \]
\[ u_1 \text{ independent of } u_2 \]

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

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

TE = \( E(Y \mid x_1) - E(Y \mid x_0) \)

Complete identification conditions for confounded models with multiple mediators.

WEAKER AND TRANSPARENT CONDITIONS FOR NDE IDENTIFICATION

No confounding

\[ X \rightarrow M \rightarrow Y \rightarrow W \]

(a)

(b)

There exists a set \( W \) such that:
A-1 No member of \( W \) is a descendant of \( X \).
A-2 \( W \) blocks all back-door paths from \( M \) to \( Y \), disregarding the one through \( X \).
A-3 The \( W \)-specific effect of \( X \) on \( M \) is identifiable.
\( P(m \mid \text{do}(x), w) \)
A-4 The \( W \)-specific effect of \( \{X, M\} \) on \( Y \) is identifiable.
\( P(y \mid \text{do}(x, m), w) \)

WHEN CAN WE IDENTIFY MEDIATED EFFECTS?

SUMMARY OF RESULTS ON MEDIATION

- Ignorability is not required for identifying natural effects
- 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 effect by Shpitser (2013).

OUTLINE

- Concepts:
  - Causal inference – a paradigm shift
  - The two fundamental laws
- Basic tools:
  - Graph separation
  - The truncated product formula
  - The back-door adjustment formula
  - The do-calculus
- Capabilities:
  - Policy evaluation
  - Transportability
  - Mediation
  - Missing Data
MISSING DATA: A CAUSAL INFERENCE PERSPECTIVE
(Mohan, Pearl & Tian 2013)

- Pervasive in every experimental science.
- Huge literature, powerful software industry, deeply entrenched culture.
- Current practices are based on statistical characterization (Rubin, 1976) of a problem that is inherently causal.
- Needed: (1) theoretical guidance, (2) performance guarantees, and (3) tests of assumptions.

WHAT CAN CAUSAL THEORY DO FOR MISSING DATA?

Q-1. What should the world be like, for a given statistical procedure to produce the expected result?
Q-2. Can we tell from the postulated world whether any method can produce a bias-free result? How?
Q-3. Can we tell from data if the world does not work as postulated?
- To answer these questions, we need models of the world, i.e., process models.
- Statistical characterization of the problem is too crude, e.g., MCAR, MAR, MNAR, testable, recoverable, non-recoverable

Graphical Models for Inference With Missing Data
(From Mohan et al., NIPS-2013)

Recoverability of Query (Q)

Is \( Q = P(X,Y,Z) \) recoverable?
\[ Q = P(X,Y)P(X,Y) \]
\[ = P(Z|D = 0, X, Y)P(X,Y) \]
\[ = P(Z^*|R_2 = 0, X, Y)P(X,Y) \]

WHY GRAPHS?

\[ x \rightarrow y \rightarrow z \rightarrow w \]
\[ z \perp x | y \wedge w | y \Rightarrow x \perp w z | y \]

1. Match the organization of human knowledge
   1a. Guard veracity of assumptions
   1b. Assure transparency of assumptions
   1c. Assure transparency of their logical ramifications
2. Blueprints for simulation
3. Unveil testable implications

RECOVERABILITY AND TESTABILITY

<table>
<thead>
<tr>
<th>Assumption Type</th>
<th>Recoverable</th>
<th>Testable</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCAR</td>
<td>recoverable</td>
<td>almost testable</td>
<td></td>
</tr>
<tr>
<td>MAR</td>
<td>recoverable</td>
<td>uncharted</td>
<td></td>
</tr>
<tr>
<td>MNAR</td>
<td>uncharted</td>
<td>uncharted</td>
<td></td>
</tr>
</tbody>
</table>

Distribution with missing values

Graph depicting the missingness process

WHY GRAPHS?

\[ x \rightarrow y \rightarrow z \rightarrow w \]
\[ z \perp x | y \wedge w | y \Rightarrow x \perp w z | y \]

1. Match the organization of human knowledge
   1a. Guard veracity of assumptions
   1b. Assure transparency of assumptions
   1c. Assure transparency of their logical ramifications
2. Blueprints for simulation
3. Unveil testable implications

RECOVERABILITY AND TESTABILITY

<table>
<thead>
<tr>
<th>Assumption Type</th>
<th>Recoverable</th>
<th>Testable</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCAR</td>
<td>recoverable</td>
<td>almost testable</td>
<td></td>
</tr>
<tr>
<td>MAR</td>
<td>recoverable</td>
<td>uncharted</td>
<td></td>
</tr>
<tr>
<td>MNAR</td>
<td>uncharted</td>
<td>uncharted</td>
<td></td>
</tr>
</tbody>
</table>
1. Constructing the graph requires less knowledge than deciding whether a problem lies in MCAR, MAR or MNAR.
2. Understanding what the world should be like for a given procedure to work is a precondition for deciding when model’s details are not necessary.
3. Knowing whether non-convergence is due to theoretical impediment or local optima, is extremely useful.
4. Graphs unveil when a model is testable.

CONCLUSIONS

1. Think nature, not data, not even experiment.
2. Think hard, but only once – the rest is mechanizable.
3. Speak a language in which the veracity of each assumption can be judged by users, and which tells you whether any of those assumptions can be refuted by data.

Thank you

TUTORIAL BIBLIOGRAPHY