Causal Inference

Sebastian Weichwald

sweichwald.de @sweichwald

Lviv Data Science Summer School Online
2020-07-20
Learn more on Causality at Lviv Summer School 2020

- Fundamentals of Causal Learning
- Wednesday, July 22
- 8:00–12:30 UTC
- Marharyta Aleksandrova
More online lectures on Causality

• 4 lectures on causality by J Peters (8 h)
  *MIT Statistics and Data Science Center, 2017* [stat.mit.edu/news/four-lectures-causality](stat.mit.edu/news/four-lectures-causality)

• causality tutorial by D Janzing and S Weichwald (4 h)
  *Conference on Cognitive Computational Neuroscience 2019* [sweichwald.de/ccn2019](sweichwald.de/ccn2019)

• course on causality by S Bauer and B Schölkopf (3 h)
  *Machine Learning Summer School 2020* [youtube.com/watch?v=btmJtThWmhA](youtube.com/watch?v=btmJtThWmhA)

• course on causality by D Janzing and B Schölkopf (3 h)
  *Machine Learning Summer School 2013* [mlss.tuebingen.mpg.de/2013/speakers.html](mlss.tuebingen.mpg.de/2013/speakers.html)
“All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word "cause" never occurs. [...] To me, it seems that [...] the reason why physics has ceased to look for causes is that, in fact, there are no such things. The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.”

— B Russell (1913), On the Notion of Cause
“Fortunately, very few physicists paid attention to Russell’s enigma. They continued to write equations in the office and talk cause–effect in the cafeteria; with astonishing success they smashed the atom, invented the transistor and the laser.

The same is true for engineering.”

— J Pearl (2009), Causality
Causal questions require causal answers. 

“Correlation does not imply causation.”

Correlation(s) may tell us something about causation.

Causal Inference: assumptions, data, explicit, algorithmic
Causal questions require causal answers.
Eating chocolate produces Nobel prize winners, says study

By Oliver Nieburg, 11-Oct-2012

Related tags: noble prize, nobel laureate, Einstein, Marie Curie, chocolate, brain, Switzerland, Sweden, candy

Messerli duly points out that correlation does not prove causation, but, he writes, "since chocolate consumption has been documented to improve cognitive function, it seems most likely that in a dose-dependent way, chocolate intake provides the abundant fertile ground needed for the sprouting of Nobel laureates. Obviously, these findings are hypothesis-generating only and will have to be tested in a prospective, randomized trial."
Figure 1. Correlation between Countries’ Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.
Kim goes on a cruise to another country.

...and reports back that year's chocolate consumption.

...and brings enormous amounts of chocolate for a year.

Can we predict a country's Nobel Laureates?

Kim goes on a cruise to another country...
Kim goes on a cruise to another country.

**SEEING:** ..and reports back that year’s chocolate consumption.
Kim goes on a cruise to another country..

**SEEING:** ..and reports back that year’s chocolate consumption.

**DOING:** ..and brings enormous amounts of chocolate for a year.
Kim goes on a cruise to another country.

**SEEING:** ..and reports back that year’s chocolate consumption.

**DOING:** ..and brings enormous amounts of chocolate for a year.

〜 Can we predict #country’s Nobel Laureates?
Causal questions require causal answers.
💡 Causal questions require causal answers.

😊 “Correlation does not imply causation.”
⚡ Causal questions require causal answers.

😆 “Correlation does not imply causation.”

SEEING VS DOING
What’s the cause and what’s the effect?

What’s the cause and what’s the effect?

Temperature $Z \leftarrow Y$ Altitude

What’s the cause and what’s the effect?

W → Q

Sebastian Weichwald — Causal Inference — Slide 11

What’s the cause and what’s the effect?

Solar Radiation \( W \) \( \rightarrow \) \( Q \) Temperature

Causal questions require causal answers.

“Correlation does not imply causation.”
Causal questions require causal answers.

“Correlation does not imply causation.”

Correlation(s) may tell us something about causation.
Causal questions require causal answers.

“Correlation does not imply causation.”

Correlation(s) may tell us something about causation.

Causal Inference: assumptions, data, explicit, algorithmic
CAUSATION AS A PROGRAMMER’S NIGHTMARE

Input:
1. “If the grass is wet, then it rained”
2. “If we break this bottle, the grass will get wet”

Output: “If we break this bottle, then it rained”
NEEDED: ALGEBRA OF DOING

Available: algebra of seeing

- e.g., What is the chance it rained if we see the grass wet?
  \[ P(\text{rain} | \text{wet}) = ? \]

\[ \{ = P(\text{wet} | \text{rain}) \frac{P(\text{rain})}{P(\text{wet})} \} \]

Needed: algebra of doing

- e.g., What is the chance it rained if we make the grass wet?
  \[ P(\text{rain} | \text{do(wet)}) = ? \]

\[ \{ = P(\text{rain}) \} \]
Formalizing the difference between seeing and doing

- observational probabilities:
  \( p(y|x) \) probability for \( Y = y \), given that we observed \( X = x \)

- interventional probabilities:
  \( p(y|\text{do}(x)) \) probability for \( Y = y \), given that we have set \( X \) to \( x \)

Confusing \( p(y|x) \) with \( p(y|\text{do}(x)) \) is the reason for most of the common misconceptions about causality!
“Normal” Probabilistic Model:

\[ M_X : \theta \mapsto P_{\theta} \]
"Normal" Probabilistic Model:

\[ M_X : \theta \mapsto \mathbb{P}_\theta \]

Causal Model:

\[ M_X : \theta \mapsto \{\mathbb{P}^{\text{do}(i)}_\theta : i \in I_X\} \]

\( I_X \) is set of interventions.
Causal Models

\[
P(X) = P^{\text{do}}(A = 0)
\]

\[
P(X) = P^{\text{do}}(A = 0, C = 0)
\]

\[
P(X) = P^{\text{do}}(C = 0)
\]

\[
P(X) = P^\emptyset
\]

has partial ordering structure

\[
M = \{P^{\text{do}}(i) : i \in I\}
\]

implies the poset of distributions

\[
\leq
\]
Causal Models

\[ \mathcal{I}_X \text{ has partial ordering structure} \]
Causal Models

\[ I_X \text{ has partial ordering structure} \]

\[ M_X \text{ implies the poset of distributions } P_X := \left( \left\{ P_{X}^{\text{do}(i)} : i \in I_X \right\} , \leq_X \right) \]
Structural Causal Models

\[ M_X = (S_X, I_X, P_{E_X}) \]
Structural Causal Models

\[ \mathcal{M}_X = (S_X, I_X, P_{E_X}) \]

- \[ S_X = \begin{cases} 
X_1 = E_1 \\
X_2 = X_1 + E_2 
\end{cases} \]

- \[ S_X = \begin{cases} 
X_1 = E_1 \\
X_2 = X_1 + E_2 
\end{cases} \]
Structural Causal Models

\[ M_X = (S_X, I_X, \mathbb{P}_{E_X}) \]

- \( S_X = \begin{cases} 
X_1 = E_1 \\
X_2 = X_1 + E_2 
\end{cases} \)
- \( I_X = \{ \emptyset, \text{do}(X_1 = 5), \text{do}(X_2 = 3) \} \)
Structural Causal Models

\[ M_X = (S_X, I_X, \mathbb{P}_{E_X}) \]

- \( S_X = \begin{cases} 
X_1 = E_1 \\
X_2 = X_1 + E_2
\end{cases} \)
- \( I_X = \{ \varnothing, \ do(X_1 = 5), \ do(X_2 = 3) \} \)
- \( E \sim \mathcal{N}(0, I) \)
Structural Causal Models

\[ M_X = (S_X, I_X, \mathbb{P}_{E_X}) \]

- \[ S_X = \begin{cases} 
X_1 = E_1 \\
X_2 = X_1 + E_2 \end{cases} \]
- \[ I_X = \{\emptyset, \ do(X_1 = 5), \ do(X_2 = 3)\} \]
- \[ E \sim \mathcal{N}(0, I) \]

Observational

\[ \mathbb{P}_{X_1}^\emptyset \sim \mathcal{N}(0, 1) \]
\[ \mathbb{P}_{X_2}^\emptyset \sim \mathcal{N}(0, 2) \]
Structural Causal Models

\[ M_X = (S_X, I_X, P_{E_X}) \]

- \( S_X = \begin{cases} 
  X_1 = E_1 \\
  X_2 = X_1 + E_2 
\end{cases} \)

- \( I_X = \{ \emptyset, \text{do}(X_1 = 5), \text{do}(X_2 = 3) \} \)

- \( E \sim \mathcal{N}(0, 1) \)

**Observational**

- \( P_{X_1}^\emptyset \sim \mathcal{N}(0, 1) \)

- \( P_{X_2}^\emptyset \sim \mathcal{N}(0, 2) \)

**Intervention on \( X_1 \)**

- \( P_{X_1}^{\text{do}(X_1=5)} \equiv 5 \)

- \( P_{X_2}^{\text{do}(X_1=5)} \sim \mathcal{N}(5, 1) \)
Structural Causal Models

\[ M_X = (S_X, I_X, P_{E_X}) \]

- \( S_X = \begin{cases} 
  X_1 = E_1 \\
  X_2 = X_1 + E_2 
\end{cases} \)
- \( I_X = \{\emptyset, \text{do}(X_1 = 5), \text{do}(X_2 = 3)\} \)
- \( E \sim \mathcal{N}(0, I) \)

observational

\[ P_{X_1}^{\emptyset} \sim \mathcal{N}(0, 1) \quad \quad P_{X_1}^{\text{do}(X_1=5)} \equiv 5 \]

intervention on \( X_1 \)

\[ P_{X_2}^{\emptyset} \sim \mathcal{N}(0, 2) \quad \quad P_{X_2}^{\text{do}(X_1=5)} \sim \mathcal{N}(5, 1) \]

intervention on \( X_2 \)

\[ P_{X_1}^{\text{do}(X_2=3)} \sim \mathcal{N}(0, 1) \quad \quad P_{X_2}^{\text{do}(X_2=3)} \equiv 3 \]
causal discovery?

observations

causal model

\[
\{ \mathbb{P}_{X}^{\text{do}(i)} : i \in \mathcal{I}_{\text{sub}} \subseteq \mathcal{I}_{X} \}
\]

\[
\{ \mathbb{P}_{X}^{\text{do}(i)} : i \in \mathcal{I}_{X} \supseteq \mathcal{I}_{\text{sub}} \}
\]
3 models inducing the same observational yet different interventional distributions

Efi/t_ting observational data well is not enough

A = \frac{1}{2}B - \frac{1}{2}C + \sqrt{3}/2N_A
B = \sqrt{3}N_B
C = \frac{1}{3}B + \sqrt{2}/3N_C

A = \sqrt{2}N_A
B = \frac{1}{2}A + C + \sqrt{3}/2N_B
C = N_C

A = h + N_A
B = h + C + N_B
C = N_C
$A = \frac{1}{2}B - \frac{1}{2}C + \sqrt{\frac{3}{2}}N_A$
$B = \frac{1}{2}A + C + \sqrt{\frac{3}{2}}N_B$
$C = \frac{1}{3}B + \sqrt{\frac{2}{3}}N_C$

$A = \sqrt{2}N_A$  
$B = \frac{1}{2}A + C + \sqrt{\frac{3}{2}}N_B$  
$C = N_C$

$A = h + N_A$  
$B = h + C + N_B$  
$C = N_C$

$\rightsquigarrow$ 3 models inducing the same observational yet different interventional distributions
\begin{align*}
A &= \frac{1}{2}B - \frac{1}{2}C + \sqrt{\frac{3}{2}}N_A \\
B &= \sqrt{3}N_B \\
C &= \frac{1}{3}B + \sqrt{\frac{2}{3}}N_C
\end{align*}

\begin{align*}
A &= \sqrt{2}N_A \\
B &= \frac{1}{2}A + C + \sqrt{\frac{3}{2}}N_B \\
C &= N_C
\end{align*}

\begin{align*}
A &= h + N_A \\
B &= h + C + N_B \\
C &= N_C
\end{align*}

\( \Rightarrow \) 3 models inducing the same observational yet different interventional distributions

\( \exists \) fitting observational data well is not enough \( \exists \)

Sebastian Weichwald — Causal Inference — Slide 20
Reichenbach’s principle of common cause (1956)

If two variables $X$ and $Y$ are statistically dependent then either

• every statistical dependence is due to a causal relation

• cases I, II, and III can also occur simultaneously

• distinction between the 3 cases is a key problem in scientific reasoning
Reichenbach’s principle of common cause (1956)

If two variables $X$ and $Y$ are statistically dependent then either

- every statistical dependence is due to a causal relation
- cases I, II, and III can also occur simultaneously
- distinction between the 3 cases is a key problem in scientific reasoning
Reichenbach’s principle of common cause (1956)

If two variables $X$ and $Y$ are statistically dependent then either

- Case I: $X \rightarrow Y$
- Case II: $X \rightarrow Z \rightarrow Y$
- Case III: $X$ and $Y$ are directly connected

• every statistical dependence is due to a causal relation
• cases I, II, and III can also occur simultaneously
• distinction between the 3 cases is a key problem in scientific reasoning
Reichenbach’s principle of common cause (1956)

If two variables $X$ and $Y$ are statistically dependent then either

I. $X \rightarrow Y$
II. $X \rightarrow Z \rightarrow Y$
III. $X \leftrightarrow Y$

- every statistical dependence is due to a causal relation
- cases I, II, and III can also occur simultaneously
- distinction between the 3 cases is a key problem in scientific reasoning
Reichenbach’s principle of common cause (1956)

If two variables $X$ and $Y$ are statistically dependent then either

- every statistical dependence is due to a causal relation
Reichenbach’s principle of common cause (1956)

If two variables $X$ and $Y$ are statistically dependent then either

- every statistical dependence is due to a causal relation
- cases I, II, and III can also occur simultaneously
Reichenbach’s principle of common cause (1956)

If two variables $X$ and $Y$ are statistically dependent then either

- every statistical dependence is due to a causal relation
- cases I, II, and III can also occur simultaneously
- distinction between the 3 cases is a key problem in scientific reasoning
Functional model of causality

• every node $X_j$ is a function of its parents and an unobserved noise term $E_j$

$$X_j = f_j(\text{PA}_j, E_j)$$

• all noise terms $E_j$ are statistically independent (causal sufficiency)

• which properties of $P(X_1, \ldots, X_n)$ follow?
Causal Markov condition (4 equivalent statements)

- existence of a functional model
- local Markov condition: every node is conditionally independent of its non-descendants, given its parents

- global Markov condition: describes all independences via d-separation
- Markov factorization: $P(X_1, \ldots, X_n) = \prod_j P(X_j|PA_j)$
Metaphor for the local Markov condition

If someone knows the genes of X’s parents, neither the genes of the grandmother nor the genes of the brother contain additional information about X.
Given observations of $W$, $E$, and $S$, what is the causal structure?
Given observations of $W$, $E$, and $S$ (ランチ), what is the causal structure?

- Which variable pairs are (in)dependent? (All dependent given the 3rd.)
Given observations of \( W, E, \) and \( S \), what is the causal structure?

- Which variable pairs are (in)dependent? (All dependent given the 3rd.)
- Assume existence of a functional model.

\[ \text{\includegraphics[width=\textwidth]{scatterplots.png}} \]
Given observations of $W$, $E$, and $S$, what is the causal structure?

- Which variable pairs are (in)dependent? (All dependent given the 3rd.)
- Assume existence of a functional model.
- Which causal structures are possible? (7)
Given observations of $W$, $E$, and $S$, what is the causal structure?

- Which variable pairs are (in)dependent? (All dependent given the 3rd.)
- Assume existence of a functional model.
- Which causal structures are possible? (7)
- Further assumption to narrow it down?
Assume existence of a functional model (causal sufficiency) & faithfulness.
Assume existence of a functional model (causal sufficiency) & faithfulness.

dependent  dependent  independent
Assume existence of a functional model (causal sufficiency) & faithfulness.
Assume existence of a functional model (causal sufficiency) & faithfulness.
Pearl’s do-operator

How to compute $p(x_1, \ldots, x_n \mid \text{do}(x_i^*))$:

- Write $p(x_1, \ldots, x_n)$ as
  $$
  \prod_{k=1}^{n} p(x_k \mid \text{parents}(x_k))
  $$

- and replace $p(x_i \mid \text{parents}(x_i))$ with $\delta_{x_i, x_i^*}$

  $$
  p(x_1, \ldots, x_n \mid \text{do}(x_i^*)) = \delta_{x_i, x_i^*} \prod_{k \neq i} p(x_k \mid \text{parents}(x_k))
  $$
Examples

1) Interventional and observational probabilities coincide (seeing = doing)
\[ p(y | \text{do}(x)) = p(y | x) \]

2) Intervening on \( x \) does not change \( y \)
\[ p(y | \text{do}(x)) = p(y) \neq p(y | x) \]

3) Intervening on \( x \) does not change \( y \)
\[ p(y | \text{do}(x)) = p(y) \neq p(y | x) \]
Confounder correction

\[
p(y|\text{do}(x)) = \sum_z p(y|x, z) p(z) \neq \sum_z p(y|x, z) p(z|x) = p(y|x)
\]
SMOKING AND CANCER: HANDLING COMPETING MODELS

   \[ P(c \mid do(s)) \approx P(c \mid s) \]
   Smoking \rightarrow Cancer

2. Tobacco Industry:
   Genotype (unobserved)
   \[ P(c \mid do(s)) = P(c) \]
   Smoking \rightarrow Cancer

3. Combined:
   \[ P(c \mid do(s)) = \text{noncomputable} \]
   Smoking \rightarrow Cancer

4. Combined and refined:
   \[ P(c \mid do(s)) = \text{computable} \]
   Smoking \rightarrow Tar \rightarrow Cancer
Typical Derivation in Causal Calculus

\[
P(c \mid do\{s\}) = \Sigma_t P(c \mid do\{s\}, t) P(t \mid do\{s\}) \quad \text{Probability Axioms}
\]

\[
= \Sigma_t P(c \mid do\{s\}, do\{t\}) P(t \mid do\{s\}) \quad \text{Rule 2}
\]

\[
= \Sigma_t P(c \mid do\{s\}, do\{t\}) P(t \mid s) \quad \text{Rule 2}
\]

\[
= \Sigma_t P(c \mid do\{t\}) P(t \mid s) \quad \text{Rule 3}
\]

\[
= \Sigma_{s'} \Sigma_t P(c \mid do\{t\}, s') P(s' \mid do\{t\}) P(t \mid s) \quad \text{Probability Axioms}
\]

\[
= \Sigma_{s'} \Sigma_t P(c \mid t, s') P(s' \mid do\{t\}) P(t \mid s) \quad \text{Rule 2}
\]

\[
= \Sigma_{s'} \Sigma_t P(c \mid t, s') P(s') P(t \mid s) \quad \text{Rule 3}
\]
<table>
<thead>
<tr>
<th>Level (Symbol)</th>
<th>Typical Activity</th>
<th>Typical Questions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Association $P(y</td>
<td>x)$</td>
<td>Seeing</td>
<td>What is? How would seeing $X$ change my belief in $Y$?</td>
</tr>
<tr>
<td>2. Intervention $P(y</td>
<td>do(x), z)$</td>
<td>Doing Intervening</td>
<td>What if? What if I do $X$?</td>
</tr>
<tr>
<td>3. Counterfactuals $P(y_x</td>
<td>x', y')$</td>
<td>Imagining, Retrospection</td>
<td>Why? Was it $X$ that caused $Y$? What if I had acted differently?</td>
</tr>
</tbody>
</table>
Challenges

Process:
1. Autocorrelation
2. Time delays
3. Nonlinear dependencies
4. Chaotic state-dependence
5. Different time scales
6. Noise distributions

Data:
7. Variable extraction
8. Unobserved variables
9. Time subsampling
10. Time aggregation
11. Measurement errors
12. Selection bias
13. Discrete data
14. Dating uncertainties

Computational/statistical:
15. Sample size
16. High dimensionality
17. Uncertainty estimation
Causal questions require causal answers.

"Correlation does not imply causation.

\[ \text{Correlation(s) may tell us something about causation.} \]
Causal questions require causal answers.
Causal questions require causal answers.

“Correlation does not imply causation.”
Causal questions require causal answers.

“Correlation does not imply causation.”
Causal questions require causal answers.

“Correlation does not imply causation.”

Correlation(s) may tell us something about causation.
Causal questions require causal answers.

“Correlation does not imply causation.”

Correlation(s) may tell us something about causation.

Causal Inference: assumptions, data, explicit, algorithmic
Causal Inference: assumptions, data, explicit, algorithmic
Causal Inference: assumptions, data, explicit, algorithmic

- 4 lectures on causality by J Peters (8 h)
  MIT Statistics and Data Science Center, 2017 [stat.mit.edu/news/four-lectures-causality](http://stat.mit.edu/news/four-lectures-causality)

- causality tutorial by D Janzing and S Weichwald (4 h)
  Conference on Cognitive Computational Neuroscience 2019 [sweichwald.de/ccn2019](http://sweichwald.de/ccn2019)

- course on causality by S Bauer and B Schölkopf (3 h)
  Machine Learning Summer School 2020 [youtube.com/watch?v=btmJtThWmhA](https://youtube.com/watch?v=btmJtThWmhA)

- course on causality by D Janzing and B Schölkopf (3 h)
  Machine Learning Summer School 2013 [mlss.tuebingen.mpg.de/2013/speakers.html](http://mlss.tuebingen.mpg.de/2013/speakers.html)