Introduction to Causal Inference

Adèle H. Ribeiro

https://adele.github.io/  |  adele.ribeiro@uni-marburg.de

Faculty of Mathematics and Computer Science
Philipps-Universität Marburg

Institute of Medical Informatics
University of Münster

Lisbon Machine Learning School (LxMLS)
July 17, 2024
AI system is better than human doctors at predicting breast cancer

A fairer way forward for AI in health care

Without careful implementation, artificial intelligence could widen health-care inequality.

Linda Nordling
Chat GPT - Impressive Abilities:

- Debug code
- Play Games
- Translate text
- Explain complex topics
- Create AI art
- Grade essays
- Write a resume
- Brainstorm ideas
- Play games with ChatGPT
- Create chatbots
- Language tutor
- Create jokes
- Solve coding errors
- Write poetry

This classic answer engine still outsmarts AI chatbots

For questions meaning hard data and math calculations, 15-year-old WolframAlpha is a fast, accurate alternative to inaccurate AI chatbots.

GPT-4 still can’t reason. But it’s a phenomenally useful tool anyway.

Konstantine Arkoudas · Follow
12 min read · Apr 11, 2024
Causality: A Missing Link to Reasoning in AI

The ability to understand cause-and-effect relationships is crucial for deeper understanding and decision-making processes.

Why artificial intelligence needs to understand consequences

A machine with a grasp of cause and effect could learn more like a human, through imagination and regret.
The Mathematical Framework of Causal Data Science
Judea Pearl — Causality

Director of the Cognitive Systems Laboratory at the University of California, Los Angeles.

In 2011, he won the A. M. Turing Award (the highest distinction in computer science and a $250,000 prize)

“for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.”
— Association for Computing Machinery (ACM)

“Deep learning has instead given us machines with truly impressive abilities but no intelligence. The difference is profound and lies in the absence of a model of reality.”
— The Book of Why: The New Science of Cause and Effect
In 2021, Angrist & Imbens won the Nobel Prize in Economics “for their methodological contributions to the analysis of causal relationships”
Yoshua Bengio — Deep Learning

Professor at the University of Montreal, and the Founder and Scientific Director of Mila – Quebec AI Institute

In 2018, he won the A. M. Turing Award, with Geoffrey Hinton, and Yann LeCun

“for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.”

— Association for Computing Machinery (ACM)

“Causality is very important for the next steps of progress of machine learning,” — interview with IEEE Spectrum.
Why causality is so important?

Causality allows important capabilities such as

**Causal Effect:** can determine the effect of *unrealized* interventions rather than just predicting an outcome (i.e., can distinguish between association and causation)

- Causal Effect Identification and Estimation

**Explainability:** provides a better understanding of the underlying mechanisms

- Causal Discovery

**Fairness:** captures and disentangles any mechanisms of discrimination that may be present, including direct, indirect-mediated, and indirect-confounded.

**Generalizability:** allows the transportability of causal effects across different domains.

**Data Fusion:** provides language and theory to cohesively combine prior knowledge and data from multiple and heterogeneous studies.
Causality Theory by Judea Pearl

https://causality101.net/
Causality Theory by Judea Pearl

The Book of Why

Causal Inference in Statistics

Causality

The New Science of Cause and Effect

A Primer

Models, Reasoning, and Inference

Judea Pearl

Madelyn Glymour

Nicholas P. Jewell

Wiley

SECOND EDITION

11
Prediction vs Effect of Interventions

Statistical Association vs Causation
**Task:** Can I guess the size of a fire by observing the number of firefighters?

**Yes!**

- **$X$: Number of firefighters in action**
- **$Y$: Size of the (initial) fire**

\[
\rho_{XY} \neq 0 \implies X \text{ is a good predictor of } Y
\]

\[
P(Y = y | X = x) \neq P(Y = y)
\]

**Observational Probability Distribution**

**Correlation between severity of fire and number of firefighters in action**

\[
y = 2.3x - 1 \\
R^2 = 0.92
\]

**Positive Correlation:**

The more firefighters, the stronger the fire!
Prediction ⇒ Decision-Making?

Should we reduce the number of firefighters to decrease the size of the fire?

**Misleading correlation:** It is the size of the fire that determines the number of firefighters needed, not the other way around.
Causal Effect \equiv \text{Effect of an Intervention}

The causal direction is determined by understanding the underlying reality.

\( X \): Number of firefighters in action
\( Y \): (Initial) Severity of the fire

\[
\begin{align*}
X &= f_X(Y, U_X) \\
Y &= f_Y(U_Y)
\end{align*}
\]

Underlying Structural Causal Model (SCM)

\( Y \) is not a function of \( X \)

In other words, \( X \) is not a cause of \( Y \)

Changing the number of firefighters through an action/intervention on \( X \), \( do(X = x) \), does not affect the initial size of the fire \( (Y) \).
Structural Causal Model (SCM)

EXPLAINABILITY AND THE DATA GENERATING MODEL
**Definition:** A structural causal model $\mathcal{M}$ (or, data generating model) is a tuple $\langle V, U, \mathcal{F}, P(u) \rangle$, where

- $V = \{V_1, \ldots, V_n\}$: are endogenous variables
- $U = \{U_1, \ldots, U_m\}$: are exogenous variables
- $\mathcal{F} = \{f_1, \ldots, f_n\}$: are functions determining $V$, i.e., $v_i \leftarrow f_i(pa_i, u_i)$, where
  - $Pa_i \subseteq V$ are endogenous causes (parents) of $V_i$
  - $U_i \subseteq U$ are exogenous causes of $V_i$.
- $P(U)$ is the probability distribution over $U$.

**Assumption:** $\mathcal{M}$ is recursive, i.e., there are no feedback (cyclic) mechanisms.
Structural Equation Model (SEM)

- Pre-specified causal order
- Linear functions
- Normal distribution
- Markovianity / Causal Sufficiency: Error terms in $U$ are independent of each other (diagonal covariance matrix).

Full specification of an SCM requires parametric and distributional assumptions. Estimation of such models usually requires strong assumptions (e.g., Markovianity).
Statistical Association vs Causation

Pre-Interventional/Observational SCM
\[
\mathcal{M} = \begin{cases} 
V = \{X, Y\} \\
U = \{U_{XY}, U_X, U_Y\} \\
\mathcal{F} = \begin{cases} 
X = f_X(U_X, U_{XY}) \\
Y = f_Y(X, U_Y, U_{XY}) 
\end{cases} \\
P(U) 
\end{cases}
\]
Observational Distribution
\[P(V) \doteq P_{\mathcal{M}}(V)\]

Post-Interventional/Interventional SCM
\[
\mathcal{M}_x = \begin{cases} 
V = \{X, Y\} \\
U = \{U_{XY}, U_X, U_Y\} \\
\mathcal{F} = \begin{cases} 
X = x \\
Y = f_Y(x, U_Y, U_{XY}) \\
P(U) 
\end{cases} 
\end{cases}
\]
Interventional Distribution
\[P(V | do(X = x)) \doteq P_{\mathcal{M}_x}(V)\]

Can we predict better the value of \(Y\) after observing that \(X = x\)?
\[P(Y = y | X = x) \neq P(Y = y) \implies X \text{ is correlated to } Y\]

Can we predict better the value of \(Y\) after making an intervention \(do(X = x)\)?
\[\exists x \text{ s.t. } P_{\mathcal{M}_x}(Y = y) \neq P(Y = y) \implies X \text{ is a cause of } Y\]
Statistical Association vs Causation

Pre-Interventional/Observational SCM
\[ \mathcal{M} = \begin{cases} \mathcal{V} = \{X, Y\} \\ \mathcal{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} X = f_X(U_X, U_{XY}) \\ Y = f_Y(X, U_Y, U_{XY}) \end{cases} \\ P(U) \end{cases} \]

Observational Data
Observational Distribution
Observational Causal Diagram

Post-Interventional/Interventional SCM
\[ \mathcal{M}_x = \begin{cases} \mathcal{V} = \{X, Y\} \\ \mathcal{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} X = x \\ Y = f_Y(x, U_Y, U_{XY}) \end{cases} \\ P(U) \end{cases} \]

Interventional Data
Interventional Distribution
Interventional Causal Diagram

Loss of Information
Randomized Experiments

A well accepted way to access $P(Y \mid do(X = x))$ is through a perfectly realized Randomized Experiments / Control Trials (e.g. RCT):

$$\mathbb{E}[Y \mid do(X = x_0)] - \mathbb{E}[Y \mid do(X = x_1)]$$

Average Causal Effect: $\mathbb{E}[Y \mid do(X = x_0)] - \mathbb{E}[Y \mid do(X = x_1)]$
What if we cannot conduct randomized experiments?

(for example due to ethical concerns, practical limitations, or logistical challenges)
Markovian Parametrization

Data

Potential Causal Diagrams

Potential SCMs

\[ \mathcal{M}_{11} = \langle V, U_1, \mathcal{F}_{11}, P_{11}(u_1) \rangle \]

\[ \mathcal{M}_{1k_1} = \langle V, U_1, \mathcal{F}_{1k_1}, P_{1k_1}(u_1) \rangle \]

\[ \mathcal{M}_{21} = \langle V, U_2, \mathcal{F}_{21}, P_{21}(u_2) \rangle \]

\[ \mathcal{M}_{2k_2} = \langle V, U_2, \mathcal{F}_{2k_2}, P_{2k_2}(u_2) \rangle \]

\[ \mathcal{M}_{31} = \langle V, U_3, \mathcal{F}_{31}, P_{31}(u_3) \rangle \]

\[ \mathcal{M}_{3k_3} = \langle V, U_3, \mathcal{F}_{3k_3}, P_{3k_3}(u_3) \rangle \]

\[ \mathcal{M}_{41} = \langle V, U_4, \mathcal{F}_{41}, P_{41}(u_4) \rangle \]

\[ \mathcal{M}_{4k_4} = \langle V, U_4, \mathcal{F}_{4k_4}, P_{4k_4}(u_4) \rangle \]

\[ \mathcal{M}_{51} = \langle V, U_5, \mathcal{F}_{51}, P_{51}(u_5) \rangle \]

\[ \mathcal{M}_{5k_5} = \langle V, U_5, \mathcal{F}_{5k_5}, P_{5k_5}(u_5) \rangle \]

Enclosed Knowledge / Assumptions

Markovian Parametrization

True Model
Multiple models / neural nets fit the data equally well, leading to different causal explanations!
Pearl’s Causal Hierarchy (PCH)

The Three Inferential Layers
Ladder of Causation

<table>
<thead>
<tr>
<th>Layer</th>
<th>Task / Language</th>
<th>Typical Question</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counterfactual</td>
<td>Structural Causal Model</td>
<td>What if I had acted differently?</td>
<td>Was it the aspirin that stopped my headache?</td>
</tr>
<tr>
<td>Interventional</td>
<td>ML- Reinforcement (Causal Bayes Net)</td>
<td>What if I do X?</td>
<td>Will my headache be cured if I take aspirin?</td>
</tr>
<tr>
<td>Associational</td>
<td>ML- (Un)Supervised (Bayesian Networks, Decision Trees, Deep Neural Networks)</td>
<td>What if I see?</td>
<td>What does a symptom tell us about the disease?</td>
</tr>
</tbody>
</table>

Cross-layer inferences:

- **Doing**: most of the inferences are about causal effects (policies, treatments, decisions)
- **Seeing**: most of the available data is observational, passively collected

---

Ladder of Causation

Cross-layer inferences:

Doing

most of the inferences are about causal effects
(policies, treatments, decisions)

Causal Hierarchy Theorem: The ladder almost never collapses. That is, for almost any SCM, the rungs of the ladder remain distinct.

Association vs Causation

Will we be able to decide the true relationship just by seeing more data?

Which type of data would maybe provide us more definite conclusion?
Bayesian Network

A DAG, possibly with latent confounders (ADMG), representing the \textit{conditional independences} implied by an SCM.
Encoding Conditional independencies

Fork
- \( Z \) as a common cause

Chain
- \( Z \) as a mediator

V-Structure
- \( Z \) as a collider or common effect

In both cases, \( Z \) is a non-collider!
Active and Inactive Triplets

**Definition (inactive):** A triplet \( \langle V_i, V_m, V_j \rangle \) is said to be *inactive* relative to a set \( Z \) if the middle node \( V_m \):

1. Is a non-collider and is in \( Z \); or
2. Is a collider and neither it nor any of its descendants in \( Z \).

\[
\begin{align*}
\text{W is non-collider} \quad &\quad \text{W is (descendant of) a} \\
\text{and } W \in Z \quad &\quad \text{collider and } W, A \notin Z \\
X \to W \to Y \quad &\quad X \to W \leftarrow Y \\
X \leftarrow W \leftarrow Y \quad &\quad X \to W \leftarrow Y \\
X \leftarrow \cdots \leftarrow W \to Y \quad &\quad X \to W \leftarrow \cdots \leftarrow Y \\
X \leftarrow \cdots \leftarrow W \leftarrow Y \\
\end{align*}
\]
**Definition (d-separation):** A path $p$ in an ADMG $G$ is said to be **d-separated** (or blocked) by a set of variables $Z$ if and only if $p$ contains an inactive triplet in it.

A set $Z$ d-separates $X$ and $Y$ if and only if $Z$ blocks every path between a node in $X$ and a node in $Y$. We denote that by $(X \perp \perp Y \mid Z)_G$.

Does $Z$ d-separate $X$ and $Y$?

\[
\begin{align*}
X & \leftarrow B \rightarrow W \rightarrow Y & \text{Z:} & \times \{\} \checkmark \{B\} \checkmark \{W\} \checkmark \{B, W\} \\
X & \rightarrow B \rightarrow W \leftarrow Y & \text{Z:} & \checkmark \{\} \checkmark \{B\} \checkmark \{W\} \times \{B, W\} \\
X & \rightarrow B \rightarrow W \rightarrow Y & \text{Z:} & \times \{\} \times \{B\} \times \{W\} \times \{B, W\}
\end{align*}
\]

**Global Markov property:** $(X \perp \perp Y \mid Z)_G \Rightarrow (X \perp \perp Y \mid Z)_P$
Bayesian Networks (BN) are **Minimal Independence Maps**: 

\[(X \perp \!
\!
\perp Y \mid Z)_G \Rightarrow (X \perp \!
\!
\perp Y \mid Z)_P\]

**Observational Distribution**

\[P(V) \doteq P_\mathcal{M}(V) = \sum_{u} \prod_{V_i \in V} P(v_i \mid p_{a_i}, u_i)P(u)\]

Edges have no causal semantics!

**Factorization obtained by Chain Rule and conditional independencies implied by the SCM \(\mathcal{M}\).**

\[P(v) = P(w \mid z, x, y, a) P(z \mid x, y, a) P(x \mid y, a) P(y \mid a) P(a)\]

\[= P(w \mid z) P(z \mid x, y) P(x \mid a) P(y \mid a) P(a)\]

\[W \perp X, Y, A \mid Z \quad A \perp Z \mid X, Y \quad Y \perp X \mid A\]
Bayesian Networks (BN) are **Minimal Independence Maps**:

\[
(X \perp Y \mid Z)_G \Rightarrow (X \perp Y \mid Z)_P
\]

No edges of \( G \) can be removed without ceasing such a property.

**Observational Distribution**

\[
P(V) \doteq P_M(V) = \sum \prod_{u \in V} P(v_i \mid pa_i, u_i)P(u)
\]

Factorization obtained by Chain Rule and conditional independencies implied by the SCM \( M \).

Edges have no causal semantics!

\[
P(v) = P(w \mid z, x, y, a) P(z \mid x, y, a) P(x \mid y, a) P(y \mid a) P(a)
\]

\[
= P(w \mid z) P(z \mid x, y) P(x \mid a) P(y \mid a) P(a)
\]

\[
W \perp X, Y, A \mid Z \quad A \perp Z \mid X, Y \quad Y \perp X \mid A
\]
Markov Equivalence Class

\[
\mathcal{M}_1 = \begin{cases} 
V = \{X, Y\} \\
U = \{U_x, U_y\} \\
\mathcal{F} = \begin{cases} f_X(U_x) \\
f_Y(X, U_y) \\
P(U) \end{cases} \\
P(U) \end{cases}
\]

\[
\mathcal{M}_{N-1} = \begin{cases} 
V = \{X, Y\} \\
U = \{U_x, U_y, U_{XY}\} \\
\mathcal{F} = \begin{cases} f_X(U_x, U_{XY}) \\
f_Y(U_y, U_{XY}) \\
P(U) \end{cases} \\
P(U) \end{cases}
\]

\[
\mathcal{M}_N = \begin{cases} 
V = \{X, Y\} \\
U = \{U_x, U_y\} \\
\mathcal{F} = \begin{cases} f_X(U_x) \\
f_Y(U_y) \\
P(U) \end{cases} \\
P(U) \end{cases}
\]

Data

\[
P(x, y) = \sum_{u_x, u_y} P(x | y) P(y) P(u_x, u_y)
\]

\[
P(x, y) = \sum_{u_x, u_y} P(y | x) P(x) P(u_x, u_y)
\]

\[
\vdots
\]

Causation does not imply correlation!
Equivalent Bayesian Networks

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Factorization</th>
<th>Bayesian Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X, Y, Z)$</td>
<td>$P(x, y, z) = P(y</td>
<td>x, z)P(z</td>
</tr>
<tr>
<td>with $P(Y</td>
<td>X, Z) = P(Y</td>
<td>X)$</td>
</tr>
<tr>
<td>i.e., $X \perp Y</td>
<td>Z$</td>
<td>$P(x, y, z) = P(x</td>
</tr>
<tr>
<td></td>
<td>$= P(x</td>
<td>z)P(z</td>
</tr>
<tr>
<td></td>
<td>$P(x, y, z) = P(y</td>
<td>x, z)P(x</td>
</tr>
<tr>
<td></td>
<td>$= P(y</td>
<td>z)P(x</td>
</tr>
</tbody>
</table>

Two models are considered *Markov equivalent* if they imply the same conditional independencies.
### Equivalent Bayesian Networks

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Factorization</th>
<th>Bayesian Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X, Y, Z)$</td>
<td>$P(x, y, z) = P(y \mid x, z)P(z \mid x)P(x)$</td>
<td>![Bayesian Network Diagram]</td>
</tr>
<tr>
<td>with $P(Y \mid X, Z) = P(Y \mid X)$</td>
<td>$= P(y \mid z)P(z \mid x)P(x)$</td>
<td>![Bayesian Network Diagram]</td>
</tr>
<tr>
<td>i.e., $X \perp Y \mid Z$</td>
<td>$P(x, y, z) = P(x \mid y, z)P(y \mid z)P(z)$</td>
<td>![Bayesian Network Diagram]</td>
</tr>
<tr>
<td></td>
<td>$= P(x \mid z)P(z \mid y)P(y)$</td>
<td>![Bayesian Network Diagram]</td>
</tr>
</tbody>
</table>

**Invariance:**
Z is never a collider (either ancestor of X and Y).
Equivalent Bayesian Networks

Distribution

\[ P(X, Y, Z) \]

with \( P(Y \mid X) = P(Y) \)

i.e., \( X \perp\!\!\!\!\!\!\!\perp Y \)

Factorization

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

Bayesian Networks

Identifiable if no latent confounder

Markov Equivalent
Equivalent Bayesian Networks

**Distribution**

\[ P(X, Y, Z) \]

with \( P(Y | X) = P(Y) \)

i.e., \( X \perp\!
\perp Y \)

**Factorization**

\[ P(x, y, z) = P(z | x, y)P(x | y)P(y) \]

\[ = P(z | x, y)P(x)P(y) \]

**Bayesian Networks**

Identifiable if no latent confounder

Markov Equivalent

**Invariance:**

\( Z \) is always a collider (non-ancestor of \( X \) and \( Y \)).
Causal Bayesian Network

A DAG, possibly with latent confounders (ADMG), representing the **causal and confounding relationships** implied by an SCM
CBN: Encoder of Structural Causal Knowledge

Structural Causal Model (SCM)
\[ \mathcal{M} = \langle V, U, \mathcal{F}, P(u) \rangle \]

\[ \mathcal{M} = \begin{cases} 
  V = \{A, B, C, D\} \\
  U = \{U_A, U_B, U_C, U_D, U_{CD}\} \\
  \mathcal{F} = \begin{cases} 
    A \leftarrow f_A(U_A) \\
    B \leftarrow f_B(A, D, U_B) \\
    D \leftarrow f_Z(U_D, U_{CD}) \\
    C \leftarrow f_X(B, U_C, U_{CD}) \\
  \end{cases} \\
  P(U) 
\end{cases} \]

An SCM \( \mathcal{M} = \langle V, U, \mathcal{F}, P(u) \rangle \) induces a causal diagram such that, for every \( V_i, V_j \in V \):

\[ V_i \rightarrow V_j, \text{ if } V_i \text{ appears as argument of } f_j \in \mathcal{F}. \]
CBN: Encoder of Structural Causal Knowledge

Structural Causal Model (SCM)
$$\mathcal{M} = \langle \mathcal{V}, \mathcal{U}, \mathcal{F}, P(u) \rangle$$

$$\mathcal{M} = \begin{cases} \mathcal{V} = \{A, B, C, D\} \\ \mathcal{U} = \{U_A, U_B, U_C, U_D, U_{CD}\} \\ \mathcal{F} = \begin{cases} A \leftarrow f_A(U_A) \\ B \leftarrow f_B(A, D, U_B) \\ D \leftarrow f_Z(U_D, U_{CD}) \\ C \leftarrow f_X(B, U_C, U_{CD}) \end{cases} \\ P(U) \end{cases}$$

An SCM $$\mathcal{M} = \langle \mathcal{V}, \mathcal{U}, \mathcal{F}, P(u) \rangle$$ induces a causal diagram such that, for every $$V_i, V_j \in \mathcal{V}$$:

$$V_i \rightarrow V_j$$, if $$V_i$$ appears as argument of $$f_j \in \mathcal{F}.$$
An SCM $\mathcal{M} = \langle V, U, \mathcal{F}, P(u) \rangle$ induces a causal diagram such that, for every $V_i, V_j \in V$:

$V_i \rightarrow V_j$, if $V_i$ appears as argument of $f_j \in \mathcal{F}$.

$V_i \leftrightarrow V_j$ if the corresponding $U_i, U_j \in U$ are correlated or $f_i, f_j$ share some argument $U \in U$. 
An SCM $\mathcal{M} = \langle V, U, \mathcal{F}, P(u) \rangle$ induces a causal diagram such that, for every $V_i, V_j \in V$:

$V_i \rightarrow V_j$, if $V_i$ appears as argument of $f_j \in \mathcal{F}$.

$V_i \leftrightarrow V_j$ if the corresponding $U_i, U_j \in U$ are correlated or $f_i, f_j$ share some argument $U \in U$. 

Structural Causal Model (SCM) 

$\mathcal{M} = \langle V, U, \mathcal{F}, P(u) \rangle$

$\mathcal{M} = \{V = \{A, B, C, D\}, U = \{U_A, U_B, U_C, U_D, U_{CD}\}, \mathcal{F} = \{A \leftarrow f_A(U_A), B \leftarrow f_B(A, D, U_B), D \leftarrow f_D(U_D, U_{CD}), C \leftarrow f_C(B, U_C, U_{CD})\}, P(U)\}$

CBN: Encoder of Structural Causal Knowledge

Induced Causal Bayesian Network (CBN)
CBN: Encoder of **Structural Causal Knowledge**

Structural Causal Model (SCM)
\[ \mathcal{M} = \langle V, U, \mathcal{F}, P(u) \rangle \]

\[ \mathcal{M} = \begin{cases} 
V = \{A, B, C, D\} \\
U = \{U_A, U_B, U_C, U_D, U_{CD}\} \\
\mathcal{F} = \begin{cases} 
A \leftarrow f_A(U_A) \\
B \leftarrow f_B(A, D, U_B) \\
D \leftarrow f_D(U_D, U_{CD}) \\
C \leftarrow f_C(B, U_C, U_{CD}) \\
P(U) 
\end{cases} 
\end{cases} \]

An SCM \( \mathcal{M} = \langle V, U, \mathcal{F}, P(u) \rangle \) induces a causal diagram such that, **for every** \( V_i, V_j \in V \):

- \( V_i \to V_j \), if \( V_i \) appears as argument of \( f_j \in \mathcal{F} \).
- \( V_i \leftrightarrow V_j \) if the corresponding \( U_i, U_j \in U \) are correlated or \( f_i, f_j \) share some argument \( U \in U \).
CBN: Encoder of **Structural Causal Knowledge**

Let \( P_* \) be the collection of all interventional distributions \( P(V \mid do(x)) \), \( X \subseteq V \), including the null (observational) distribution \( P(V) \).

An Acyclic Directed Mixed Graph (ADMG) \( G \) is a CBN for \( P_* \) if for every intervention \( do(X = x) \), \( X \subseteq V \), if it hold:

\[
P(V \mid do(X = x)) \doteq P_{\mathcal{M}_x}(V)
\]

\[
= \sum_u \prod_{V_i \in V \setminus X} P(v_i \mid pa_i, u_i)P(u) \bigg|_{X=x}
\]

Interventional Distribution

Truncated factorization implied by the SCM \( \mathcal{M}_{x^*} \).

Semi-Markov relative to \( G_X \)
Causal Effect Identification from Causal Diagrams / CBNs
Causal Pipeline from a Causal Diagram

1. Query
   \[ P(y \mid do(x)) \]

2. Causal Constraints
   \[
   \begin{align*}
   P(y \mid do(x)) &= \sum_z P(y \mid x, z)P(z) \\
   \mathbb{E}(y \mid do(x)) &= \sum_z \mathbb{E}(y \mid x, z)P(z)
   \end{align*}
   \]

3. Observational Distribution
   \[
   \hat{\mathbb{E}}(y \mid do(x)) = \sum_z \hat{\mathbb{E}}((y \mid x, z)\hat{P}(z))
   \]

Causal Modeling / Causal Discovery → Causal Effect Identification → Causal Effect Estimation
Causal Effect

The **causal effect** of a (set of) treatment variable(s) $X$ on a (set of) outcome variable(s) $Y$ is a quantity derived from $P(Y \mid do(X))$ that tells us how much $Y$ changes due to an intervention $do(X = x)$.

**Examples:**

- **Average Treatment Effect (ATE)** for discrete treatments:

$$
\mathbb{E}[Y \mid do(X = x')] - \mathbb{E}[Y \mid do(X = x)],
$$

defined for two treatment levels $x'$ and $x$ of $X$.

- **Average Treatment Effect (ATE)** for continuous treatments,

$$
\frac{\partial \mathbb{E}[Y_i \mid do(X_j = x_j)]}{\partial x_j}, \text{ for all } Y_i \in Y, \text{ and } X_j \in X.
$$

The derivative shows the rate of change of $Y$ w.r.t. $do(X = x)$

where

$$
\mathbb{E}[Y \mid do(X = x)] = \sum_{y \in \Omega_Y} y P(y \mid do(x))
$$

Jacobian of $\mathbb{E}[Y \mid do(X = x)]$, where

$$
\mathbb{E}[Y \mid do(X = x)] = \int_{\Omega_Y} y P(y \mid do(x))dy,
$$

and $\Omega_Y$ is the space of all possible values that $Y$ might take on.
Classical Causal Effect Identification

1. Query
   \[ P(y \mid do(x)) \]

2. Causal Constraints
   \[ X \rightarrow M \rightarrow Y \]

3. Probability Distributions
   \[ P(x, m, y) \quad \text{Observational Distribution} \]

\[ P(y \mid do(x)) = \sum_m P(m \mid x) \sum_{x'} P(y \mid m, x') P(x') \]

Inference Engine

Solution: yes/no

Available Distributions

Interventional Distribution

Structural knowledge available

The Effect Identification Problem

**Causal Effect Identifiability:** The causal effect of a (set of) treatment variable(s) $X$ on a (set of) outcome variable(s) $Y$ is said to be identifiable from a causal diagram $G$ and the probability of the observed variables $P(V)$ if the interventional distribution $P(Y \mid do(X))$ is uniquely computable, i.e., if for every pair of SCMs $\mathcal{M}_1$ and $\mathcal{M}_2$ that induce $G$ and $P^{\mathcal{M}_1}(V) = P^{\mathcal{M}_2}(V) = P(V) > 0$, $P^{\mathcal{M}_1}(Y \mid do(X)) = P^{\mathcal{M}_2}(Y \mid do(X)) = P(Y \mid do(X))$.

In words, causal effect identifiability means that, no matter the form of true SCM, for all models $\mathcal{M}$ agreeing with $\langle G, P(V) \rangle$, they also agree in $P(y \mid do(x))$. 

![Diagram of causal effect identifiability](image-url)
**Causal Effect Identifiability:** The causal effect of a (set of) treatment variable(s) $X$ on a (set of) outcome variable(s) $Y$ is said to be identifiable from a causal diagram $G$ and the probability of the observed variables $P(V)$ if the interventional distribution $P(Y | do(X))$ is uniquely computable, i.e., if for every pair of SCMs $\mathcal{M}_1$ and $\mathcal{M}_2$ that induce $G$ and $P_{\mathcal{M}_1}(V) = P_{\mathcal{M}_2}(V) = P(V) > 0$, $P_{\mathcal{M}_1}(Y | do(X)) = P_{\mathcal{M}_2}(Y | do(X)) = P(Y | do(X))$.

In words, causal effect identifiability means that, no matter the form of true SCM, for all models $\mathcal{M}$ agreeing with $\langle G, P(V) \rangle$, they also agree in $P(y | do(x))$. 

Identifiable | Non-Identifiable
--- | ---
Models inducing $P(y)$ | Models inducing $P(y)$
Same $P(y | do(x))$ | Different $P(y | do(x))$
Models compatible with $G$ | All models

In words, causal effect identifiability means that, no matter the form of true SCM, for all models $\mathcal{M}$ agreeing with $\langle G, P(V) \rangle$, they also agree in $P(y | do(x))$. 

In words, causal effect identifiability means that, no matter the form of true SCM, for all models $\mathcal{M}$ agreeing with $\langle G, P(V) \rangle$, they also agree in $P(y | do(x))$.
Identification Via Adjustment over Parents

Let $G$ be a causal graph with all parents observed.

Then, the effect of $X$ on $Y$ is given by:

$$P(y \mid do(x)) = \sum_{pa_x} P(y \mid x, pa_x) P(pa_x)$$

Proof follows from the truncated factorization for Markovian models. Try at home!

$X = \{X\}$  \hspace{1cm}  $Y = \{Y\}$  \hspace{1cm}  $Pa_X = \{Z_1, Z_2\}$

$$P(y \mid do(x)) = \sum_{z_1, z_2} P(y \mid x, z_1, z_2) P(z_1, z_2)$$
Identification via Backdoor Criterion

Let $X$ be a set of treatment variables and $Y$ a set of outcome variables in the causal graph $G$.

If there exists a set $Z$ such that:

1. $Z$ d-separates $X$ and $Y$ in the graph $G_X$, i.e., the graph resulting from cutting the arrows out of $X$

2. no node in $Z$ is a descendant of a variable $X \in X$ in $G$ (all variables in $Z$ are pre-treatment)

Then, $Z$ satisfies the backdoor criterion for $(X, Y)$ and, then the effect of $X$ on $Y$ is given by:

$$P(y \mid do(x)) = \sum_{Z} P(y \mid x, z) P(z)$$

$Z$, a set of covariates, admissible for backdoor adjustment

In $G_X$, all non-backdoor paths are severed

Identification via Backdoor Criterion

Let $\mathbf{X}$ be a set of treatment variables and $\mathbf{Y}$ a set of outcome variables in the causal graph $G$.

If there exists a set $\mathbf{Z}$ such that:

1. $\mathbf{Z}$ d-separates $\mathbf{X}$ and $\mathbf{Y}$ in the graph $G_X$, i.e., the graph resulting from cutting the arrows out of $\mathbf{X}$

2. no node in $\mathbf{Z}$ is a descendant of a variable $X \in \mathbf{X}$ in $G$ (all variables in $\mathbf{Z}$ are pre-treatment)

Then, $\mathbf{Z}$ satisfies the backdoor criterion for $(\mathbf{X}, \mathbf{Y})$ and, then the effect of $\mathbf{X}$ on $\mathbf{Y}$ is given by:

$$P(y \mid do(x)) = \sum_z P(y \mid x, z) \cdot P(z)$$

In $G_X$, all non-backdoor paths are severed.

Admissible Sets for BD Adjustment

\(Z\) satisfies the **backdoor criterion** for or \((X, Y)\) in the causal graph \(G\) if:

1. \(Z\) d-separates \(X\) and \(Y\) in the graph \(G_X\), i.e., the graph resulting from cutting the arrows out of \(X\)
2. no node in \(Z\) is a descendant of a variable \(X \in X\) in \(G\) (all variables in \(Z\) are pre-treatment)

Minimal BD Adjustment Sets

\(\{Z_1\}\),  
\(\{Z_2\}\),  
\(\{Z_1, Z_2\}\)

\(\{Z_2\}\),  
\(\{Z_1, Z_2\}\)

\(\{\}\),  
\(\{\}\)

\[
P(y \mid do(x)) = \sum_{z_1} P(y \mid x, z_1) P(z_1)
\]

\[
P(y \mid do(x)) = P(y \mid x)
\]

\[
P(y \mid do(x)) = P(y \mid x)
\]

There is no BD Adjustment Set!

\(P(y \mid do(x))\) is non-identifiable
\[ P(Y|do(X)) = \sum Z P(Y|X, Z) P(Z) \]
\[ P(Y|do(X)) \text{ is not identifiable from } P(X,Y,Z). \]
**Theorem:** If a set $Z$ satisfies the *backdoor criterion* w.r.t. the ordered pair $(X, Y)$, then, for all $x$, it holds that $Y_x \perp X \mid Z$.

Although the satisfiability of $Z$ to the *backdoor criterion* can be tested given a causal diagram or a PAG, the condition $Y_x \perp X \mid Z$ is sometimes framed as an assumption, referred to as *(conditional) ignorability, exchangeability* or *unconfoundedness*. 

*Counterfactual Interpretation of Backdoor*
Many Scenarios Beyond Adjustment!

\[
P(y \mid do(x)) = \sum_{m} P(m \mid x) \sum_{x'} P(y \mid m, x')P(x')
\]

Front-Door

\[
E(y \mid do(x)) = \sum_{m,z} P(m \mid x, z)P(z \mid x) \sum_{x'} E(y \mid m, x', z)P(x' \mid z)
\]

Conditional Front-Door

\[
P(y \mid do(x)) = \frac{\sum_{z_2} P(x, y \mid z_1, z_2)P(z_2)}{\sum_{z_2} P(x \mid z_1, z_2)P(z_2)}
\]

Napkin

\[
P(y \mid do(x)) = \sum_{z_2, z_3} P(y \mid x, z_1, z_2, z_3)P(z_2)
\]

Unnamed

\[
\sum_{z_1} P(z_3 \mid x, z_1)P(z_1)
\]

And many others…. 
Tools for Causal Identification

1. Markovian Models (No Unobserved Confounders)
   i. Truncated Factorization / G-computation or G-formula

2. Adjustment over Parents (No Unobserved Parents)

3. Non-Markovian Models (Under the Presence of Unobserved Confounders)
   i. Graphical criteria (Backdoor Adjustment, Generalized Adjustment, Front-door Adjustment)
   ii. Do-Calculus (a.k.a Causal Calculus)
   iii. Identify Algorithm (a.k.a. ID algorithm)


Advances on Effect Identification given a Causal Diagram

Identification from observational and experimental data:


Identification of stochastic/soft (and possibly imperfect) interventions:


Identification and Estimation via Deep Neural Networks:


What if domain knowledge does not allow you construct a causal diagram?
Super-Exponential Growth

The space of DAGs grows super-exponentially with the number $n$ of variables, as shown by the following recurrence relation (Robinson, 1973):

$$|DAG(n)| = \sum_{i=1}^{n} \binom{n}{1} 2^{i(n-i)} |DAG(n - 1)|$$

Inference through enumeration is not a good idea!

| $n$  | $|DAG(n)|$         |
|------|-------------------|
| 2    | 3                 |
| 3    | 27                |
| 4    | 729               |
| 5    | 59,049            |
| 6    | $1.4349 \times 10^7$ |
| 7    | $1.0460 \times 10^{10}$ |
| 8    | $2.2877 \times 10^{13}$ |
Super-Exponential Growth

The space of ADMGs also grows super-exponentially with the number $n$ of variables, and it is much bigger than the space of DAGs:

$$|ADMG(n)| = |DAG(n)| \times 2^{n(n-1)/2}$$

| $n$ | $|DAG(n)|$ | $|ADMG(n)|$ |
|-----|----------|------------|
| 2   | 3        | 6          |
| 3   | 27       | 216        |
| 4   | 729      | 46,656     |
| 5   | 59,049   | $6.0457 \times 10^7$ |
| 6   | $1.4349 \times 10^7$ | $4.7019 \times 10^{11}$ |
| 7   | $1.0460 \times 10^{10}$ | $2.1936 \times 10^{16}$ |
| 8   | $2.2877 \times 10^{13}$ | $6.1410 \times 10^{21}$ |
Many models are statistically indistinguishable without additional parametric / distributional assumptions.

In non-parametric settings, causal discovery algorithms can only learn a graphical representation of its *Markov equivalence class* (MEC)!

**Fast Causal Inference (FCI):** Sound and complete causal discovery algorithm, even in the presence of unobserved confounders and selection bias.

Causal Discovery: Learning Structural Invariances

\[ \mathcal{M}_1 = \begin{cases} V = \{X, Y, Z\} \\ U = \{U_x, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_X) \\ Z \leftarrow f_Z(X, Y, U_2) \\ Y \leftarrow f_Y(U_Y) \end{cases} \\ P(U) \end{cases} \]

\[ \vdots \]

\[ \mathcal{M}_{N-1} = \begin{cases} V = \{X, Y, Z\} \\ U = \{U_{XZ}, U_{YZ}, U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{XZ}, U_X) \\ Z \leftarrow f_Z(Y, U_{XZ}, U_2) \\ Y \leftarrow f_Y(U_Y) \end{cases} \\ P(U) \end{cases} \]

\[ \mathcal{M}_N = \begin{cases} V = \{X, Y, Z\} \\ U = \{U_{XZ}, U_{YZ}, U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{XZ}, U_X) \\ Z \leftarrow f_Z(U_{XZ}, U_{YZ}, U_2) \\ Y \leftarrow f_Y(U_{YZ}, U_Y) \end{cases} \\ P(U) \end{cases} \]

Markov Equivalence Class (MEC)

Data

Conditional (in)dependencies

\[ P(v) \]

\[ X \perp Y \]

\[ X \rightarrow Z \]

\[ Z \rightarrow Y \]

\[ X \rightarrow Y | Z \]

\[ \vdots \]
FCI Algorithm - Pipeline

Unknown Reality

True causal diagram

Complete Graph

X \rightarrow Z \rightarrow W \rightarrow Y

X \perp W
X \perp Y | Z, W

Skeleton

Conditional Independence Tests

X \perp W
X \perp Y | Z, W

Partial Ancestral Graph (PAG)

FCI Rules
(R1) – (R10)

Implied by the PAG using m-separation

X \perp W
X \perp Y | Z, W

Implied by the ADMG using d-separation

X \perp W
X \perp Y | Z, W

X \not\perp W

By faithfulness, are correctly observed in the data

A \rightarrow B \quad \Rightarrow \quad B \text{ non-ancestor of } A
A \rightarrow B \quad \Rightarrow \quad A \text{ ancestor of } B
A \leftrightarrow B \quad \Rightarrow \quad \text{spurious association}
A \leftarrow B \quad \Rightarrow \quad \text{selection bias}

Z is not an ancestor of X or W.
Z \text{ and } W \text{ are ancestors of } Y.
Z \text{ is not confounded with } Y.
Conditional Independence Tests

Gaussian errors and independent observations: partial correlation test

R package: https://cran.r-project.org/web/packages/pcalg/

Kernel-based non-parametric test:

R package: https://cran.r-project.org/web/packages/CondIndTests

Continuous (conditional Gaussian) or Discrete (Binary, Ordinal, Multinomial) - Linear Regression

- R package: https://cran.r-project.org/web/packages/MXM/

Gaussian errors and correlated observations (family data):

Statistics in Medicine.
R package: https://github.com/adele/FamilyBasedPGMs
**PAG: Representation of the Markov Equivalence Class**

Partial Ancestral Graph (PAG)

- $Z$ is not an ancestor of $X$ or $W$.
- $Z$ and $W$ are ancestors of $Y$.
- $Z$ is not confounded with $Y$.

True (unknown) causal diagram

- $X \perp W$
- $X \perp Y \mid Z, W$
Fast Causal Inference (FCI) Algorithm

Underlying Causal Diagram

Partial Ancestral Graph

Data → FCI → E.C.
Developments in Causal Discovery with Unobserved Confounding

**Going Beyond the Markov Equivalence Class:**

1. Causal Discovery with Interventional Data


Developments in Causal Discovery with Unobserved Confounding

Going *Beyond* the Markov Equivalence Class:

2. Causal Discovery with Prior Knowledge


3. Human-in-the-Loop Probabilistic Causal Discovery

Developments in Causal Discovery with Unobserved Confounding

Going *Beyond* the Markov Equivalence Class:

4. Causal Discovery in Linear Models

5. Causal Discovery for Additive Noise Models
Developments in Causal Discovery with Unobserved Confounding

Learning Dynamic Systems:

1. Causal Discovery with Cycles


2. Causal Discovery from Time-Series Data

Can we identify causal effects from the equivalence class?

**Effect Identification:**

For Covariate Adjustment, we can use the Generalized Adjustment Criterion.

Recently, we proposed complete calculus and algorithms for the identification of marginal and conditional causal effect in PAGs!


Effect Identification in Markov Equivalence Classes

1. Query
   \[ P(y \mid do(x)) \]

2. PAG
   - Available (Observational) Distribution
   - Inferred (Interventional) Distribution

3. Data
   \[ P(x, y, z, w) \]

\[
P(y \mid do(x)) = \sum_{z} P(y \mid x, z) P(z)
\]

Solution

Yes / No

Can be constructed in a fully data-driven way!
Identification via Adjustment in Markov Equivalence Classes

Identification is possible only when the Generalized Adjustment Criterion applies.

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

General Identification in Markov Equivalence Classes

1. Query
   \[ P(y \mid do(x)) \]

2. PAG
   \[
   \begin{align*}
   &X_2 \leftrightarrow Y_2 \leftrightarrow Y_3 \leftrightarrow Y_4 \\
   &X_1 \leftrightarrow Y_1 \leftrightarrow Y_5
   \end{align*}
   \]

3. Data
   \[ P(x, y, z, w) \]  
   \text{Observational Distribution}

\[ P(y \mid do(x)) = \frac{P(y_1, y_4, y_5 \mid x_1) \cdot P(y_2 y_3, y_4, y_5 \mid x_2)}{P(y_4, y_5)} \]

Available (Observational) Distribution

Inferred (Interventional) Distribution

Solution

yes / no

Complete algorithms, available at the PAGId R package:
https://github.com/adele/PAGId

Effect Identifiability given a PAG

An effect identifiable in a PAG $\mathcal{P}$ is identifiable in all causal diagrams $G$ in the Markov Equivalence Class using the same identification formula!
Effect Non-Identifiability given a PAG

An effect not identifiable in a PAG $\mathcal{P}$ is not identifiable in at least one causal diagrams $G$ in the Markov Equivalence Class.
Causal Inference Workflow

Continuous Process of Scientific Discovery and Causal Hypothesis Refinement

SCM $M^*$ (Unobserved Nature)

$Z \leftarrow f_Z(U_z)$
$X \leftarrow f_X(Z, U_x)$
$Y \leftarrow f_Y(X, Z, U_y)$

$P(U_z, U_x, U_y)$

knowledge(t)

data(t) → distributions(t)

causal hypothesis (t)

new discoveries (t+1)

queries(t)

Experimental validation(t)

new insights

not answerable

answerable

new challenges

perform new observations and/or experiments(t+1)

A Statistical Learning  B Causal Learning  C Causal Inference  D Causal Exp. Design
Many other Topics in Causal Inference

1. Causal Representation Learning & Causal Abstraction
2. Causal Reinforcement Learning
3. Fairness & Mediation Analysis
4. Individual Treatment Effect (ITE) Estimation
5. Data-Driven Covariate Selection for Adjustment
6. Partial Effect Identification
7. Many more…
Causal Representation Learning & Causal Abstraction

Toward Causal Representation Learning

This article reviews fundamental concepts of causal inference and relates them to crucial open problems of machine learning, including transfer learning and generalisation, thereby assaying how causality can contribute to modern machine learning research.

By Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio

Coarse-grained causal models:

Causal Consistency of Structural Equation Models

| Paul K. Rohsenow†, Sebastian Wiederhold‡, Stephan Zobbe§, Jodie M. Mood‡ | Dominik Janzing, Merete Grosse-Wentrup, Bernhard Schölkopf
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>†Equal contribution</td>
<td></td>
</tr>
</tbody>
</table>

The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)

Causal Effect Identification in Cluster DAGs

Tara V. Anand†, Adele H. Ribeiro‡, Jin Tian§, Elias Bareinboim†

†Department of Biomedical Informatics, Columbia University
‡Department of Computer Science, Columbia University
§Department of Computer Science, Iowa State University

tara.v.anand@columbia.edu, adele@cs.columbia.edu, jitian@iastate.edu, eb@cs.columbia.edu

Neural Causal Abstractions

Kevin Xia and Elias Bareinboim

Causal Artificial Intelligence Lab
Columbia University
{kevinxia, eb}@cs.columbia.edu

Causal Reinforcement Learning

http://crl.causalai.net

**TASK 1**
Generalized Policy Learning
combining online + offline learning
Learn policy $\Pi$ by systematically combining offline ($L_1$) and online ($L_2$) modes of interaction.

**TASK 2**
When and Where to Intervene?
refining the policy space
Identify subset of $L_2$ to refine the policy space do($\Pi$($X$)) based on topological constraints implied by $M$ on $G$.

**TASK 3**
Counterfactual Decision-Making
changing optimization function based on intentionality, free will, and autonomy
Optimization criterion based on counterfactuals and $L_3$-based randomization (instead of $L_2$/do()-counterpart).

**TASK 4**
Generalizability & Robustness of Causal Claims
transportability & structural invariances
Generalize policy based on structural invariances shared across training (SCM $M$) and deployment environments ($M^*$).

**TASK 5**
Learning Causal Models
discovering the causal structure with observation and experiments
Learn the causal graph $G$ (of $M$) by systematically combining observations ($L_1$) and experimentation ($L_2$).

**TASK 6**
Causal Imitation Learning
policy learning with unobserved rewards
Construct $L_2$-policy based on partially observable $L_1$-data coming from an expert with unknown reward function.

By Elias Bareinboim’s Research Group
A Causal Framework for Decomposing Spurious Variations

Fairness and Mediation Analysis

D. Plecko and E. Bareinboim
Department of Computer Science
Columbia University
dp3144@columbia.edu, eb@cs.columbia.edu


Causal Fairness Analysis

A Causal Toolkit for Fair Machine Learning


Drago Plečko
Seminar für Statistik, ETH Zürich
drago.plecko@stat.math.ethz.ch

Elias Bareinboim
Department of Computer Science, Columbia University
eb@cs.columbia.edu
Generalization Bounds and Representation Learning for Estimation of Potential Outcomes and Causal Effects

Fredrik D. Johansson  
Chalmers University of Technology  
Göteborg, 412 96, Sweden  
FREDRIK.JOHANSSON@CHALMERS.SE

Uri Shalit  
Technion - Israel Institute of Technology  
Haifa, 3200003, Israel  
URISHALIT@TECHNION.AC.IL

Nathan Kallus  
Cornell University  
New York, NY 10044, USA  
KALLUS@CORNELL.EDU

David Sontag  
Massachusetts Institute of Technology  
Cambridge, MA 02139, USA  
DSONTAG@CSAIL.MIT.EDU


Other related works cited within, such as:

Estimating individual treatment effect: generalization bounds and algorithms

Uri Shalit, Fredrik D. Johansson, David Sontag  

Learning Representations for Counterfactual Inference

Fredrik D. Johansson*  
CSE, Chalmers University of Technology, Göteborg, SE-412 96, Sweden  
FREDRIK.JOHANSSON@CHALMERS.SE

Uri Shalit*  
David Sontag  
CS, New York University, 251 Mercer Street, New York, NY 10012 USA  
SHALIT@CS.NYU.EDU  
DSONTAG@CS.NYU.EDU

* Equal contribution
Finding Valid Adjustments under Non-ignorability with Minimal DAG Knowledge


Differentiable Causal Backdoor Discovery

Limor Gultchin
University of Oxford
The Alan Turing Institute

Matt J. Kusner
University College London
The Alan Turing Institute

Varun Kanade
University of Oxford
The Alan Turing Institute

Ricardo Silva
University College London
The Alan Turing Institute
Partial Effect Identification

Stochastic Causal Programming for Bounding Treatment Effects

Kirtan Padh
Helmholtz AI, Helmholtz Munich & Technical University Munich

Jakob Zeitler
University College London

David Watson
King’s College London

Matt Kusner
University College London

Ricardo Silva
University College London

Niki Kilbertus
Helmholtz AI, Helmholtz Munich & Technical University Munich

Kirtan Padh, Jakob Zeitler, David Watson, Matt Kusner, Ricardo Silva, Niki Kilbertus; Proceedings of the Second Conference on Causal Learning and Reasoning, PMLR 213:142-176

Thank you! :)

Feel free to reach out to me if you have any questions:

adele.ribeiro@uni-marburg.de