From Statistical to Causal Models

Guido Consonni

SISBAYES 2025 September 4-5, 2025
Department of Statistical Sciences, University of Padova

Outline

Awards for Causal Reasoning

Turing award 2011

Judea Pearl

"For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning"

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021

David Card

"For his empirical contributions to labour economics"

Joshua D. Angrist and Guido W. Imbens

"For their methodological contributions to the analysis of causal relationships"

Rousseeuw Prize for Statistics 2022

James Robins, Miguel Hernán, Thomas Richardson, Andrea Rotnitzky, Eric Tchetgen Tchetgen

"For their pioneering work on Causal Inference with applications in Medicine and Public Health"

Causality and AI/ML

“The kind of causal inference seen in natural human thought can be “algorithmitized” to help produce human-level machine intelligence”
Judea Pearl, 2019, *Communications of the ACM*

“Some of the hard open problems of machine learning and AI are intrinsically related to causality, and progress may require advances in our understanding of how to model and infer causality from data”
Bernhard Schölkopf, 2022, *International Congress of Mathematicians*

Causal AI

Preserving Causal Constraints in Counterfactual Explanations for Machine Learning Classifiers
NeurIPS, 2019

Interpretable ML; feasible counterfactuals

Causal Reasoning and Large Language Models: Opening a New Frontier for Causality
Transactions on Machine Learning Research, 2024

"Behavorial" study of LLMs to benchmark their capability in generating causal arguments

Improving the accuracy of medical diagnosis with causal machine learning
Nature communications, 2020

we reformulate diagnosis as a counterfactual inference task and derive counterfactual diagnostic algorithms.In medical diagnosis a doctor aims to explain a patient's symptoms by determining the diseases causing them, while existing diagnostic algorithms are purely associative

Robust Agents Learn Causal World Models

International Conference on Learning Representations, 2024

"Any agent capable of satisfying a regret bound for a large set of distributional shifts must have learned an approximate causal model of the data generating process"

Explaining the Behavior of Black-Box Prediction Algorithms with Causal Learning

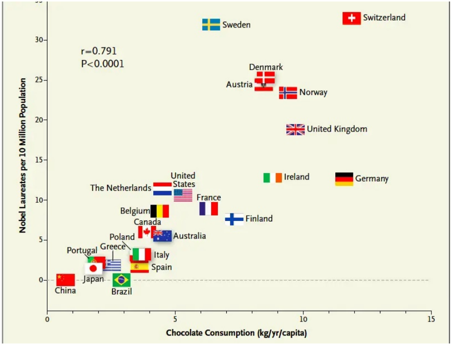
Transactions on Machine Learning Research, 2025

Causal approaches to post-hoc explainability for black-box prediction models(e.g. deep neural networks trained on image pixel data)

Table of Contents

Correlation vs causation

Correlation does not imply causation
Chocolate and Nobel prize winners



Understanding causation

- Manipulability
- Intervention

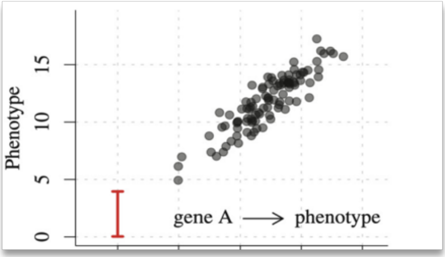
J. Woodward (2001). *Causation and manipulability*
J. Pearl (2009). *Causality: models, reasoning, and inference*. 2nd edn

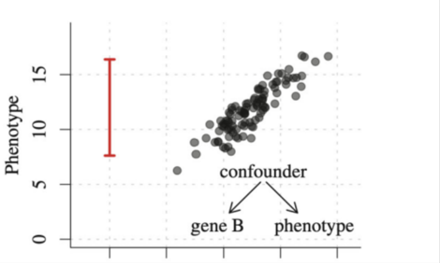
- Epidemiology
  J. M. Robins, M. A. Hernan, and B. Brumback (2000)
- Agriculture
  S. Wright (1921)
- Econometrics
  T. Haavelmo (1944); K. D. Hoover (2001)

Causal effect

Definition

A random variable X has a causal effect on a random variable Y if there exist $x \neq x'$ such that the distribution of Y after intervening on X and setting it to x differs from the distribution of Y after setting X to x'





Both gene A and B are positively correlated with the phenotype. Yet, only gene A ha a causal effect on the phenotype: knocking it out strongly reduces the phenotype. Not so for gene B.

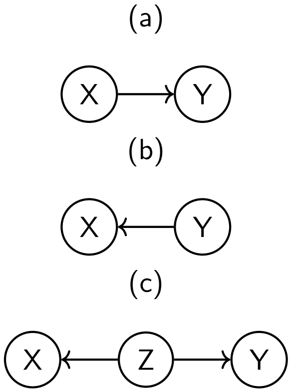
Correlation and Causation: what's the connection?

Principle

If two random variables X and Y are statistically dependent, $X \not\perp\!\!\!\perp Y$, *then there exists a random variable* Z *which causally influences both of them and which explains all their dependence that is*

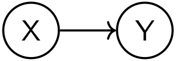
$X \perp\!\!\!\perp Y \mid Z$ *(c)*

As a special case, Z *may coincide with* X *or* Y *(a) or (b)*

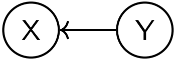


Chocolate consumption and Nobel laureates

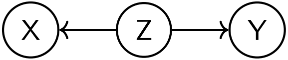
(a) ✗



(b) ✗



(c) ✓



X: Chocolate consumption (rate)
Y: Nobel laureates (rate)
Z: Confounder

- The class of observational distributions over X and Y that can be realized by these models is the same in all three cases
- Cannot distinguish among a), b) and c) through passive observation i.e., in a purely data-driven way
- Z latent confounder
  drives consumer spending and investment in education and research
  [from background knowledge]

Making Sense of Correlation

- Correlation is still useful
- Causality is not always needed
- Gene A and gene B remain useful features for making predictions
- In a passive, or observational, setting
  - we measure the activities of certain genes
    and are asked to predict the phenotype

- However, if we want to answer interventional questions
  - the outcome of a gene knockout experiment
  - the effect of a policy enforcing a job training program
- We need more than correlation
- We need a causal model

Table of Contents

Causal Graphical Model

Definition

A Causal Graphical Model (CGM) $\mathcal{M} = (G, p)$ over n random variables $X_1, \ldots, X_n$ consists of

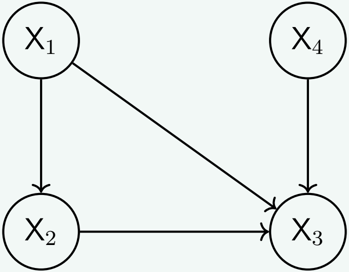
- a directed acyclic graph (DAG) G in which directed edges $(X_j \rightarrow X_i)$ represent a direct causal effect of X_j on X_i;
- a joint distribution $p(X_1, \ldots, X_n)$ which is Markovian w.r.t. G

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid PA_i); \quad PA_i = \{X_j : (X_j \rightarrow X_i) \in G\}$$

PA_i is the set of parents, or direct causes, of X_i in G

Decomposition of the joint distribution into causal conditionals

Four variables

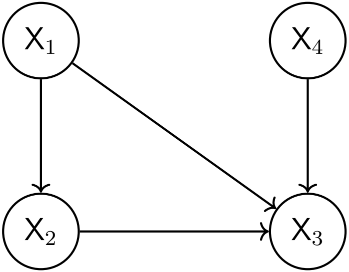


$$P(X_1, X_2, X_3, X_4) =$$
$$P(X_1)\, P(X_4)\, P(X_2 \mid X_1)\, P(X_3 \mid X_1, X_2, X_4)$$

Markov Condition

Definition

A joint distribution satisfies the Markov condition w.r.t. a DAG G if every variable is conditionally independent of its non-descendants in G given its parents in G



$X_2 \perp\!\!\!\perp X_4 \mid X_1$

$X_4 \perp\!\!\!\perp \{X_1, X_2\}$

$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i \mid PA_i)$ iff the Markov Condition holds

A CGM satisfies the Markov condition

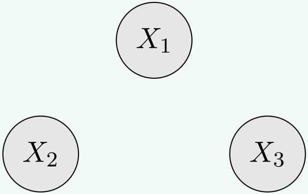
Intervention on a causal DAG

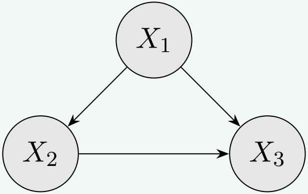
What makes a DAG "causal"

Externally forcing a variable to take on a particular value (intervention) renders the variable independent of its causes
thus breaking their causal influence

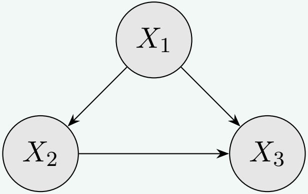
- do-operator
- graph-surgery

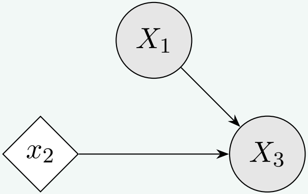
Three variables and a graph





From graph G to G'

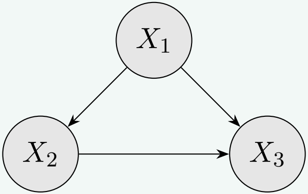




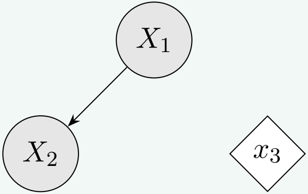
Starting graph G

Post-intervention graph G' for $do(X_2 = x_2)$.

From graph G to G''



Starting graph G



Post-intervention graph G'' for $do(X_3 = x_3)$.

More on $do(X = x)$

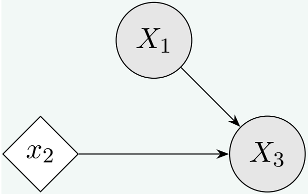
- Intervention and Conditioning are different
- Conditioning is passive
- Intervention is active
  - if a gene is knocked out, it is no longer influenced by other genes that were previously regulating it
    instead, its activity is now solely determined by the intervention

Note

This is fundamentally different from conditioning, because passively observing the activity of a gene provides information about its driving factors (i.e., its direct causes)

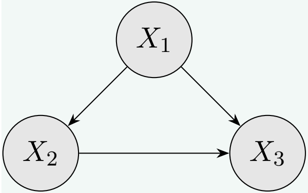
$$p(y \mid x) \neq p(y \mid do(X = x))$$

Example



DAG G' for $do(X_2 = x_2)$

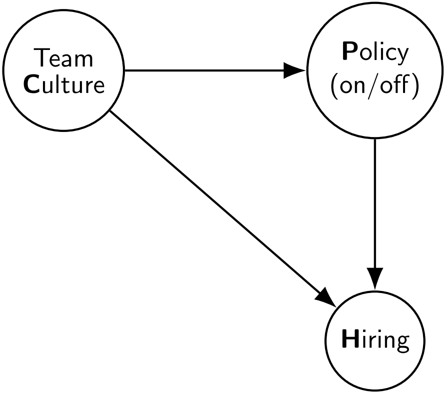
$$p(X_3|do(X_2 = x_2)) = \sum_{x_1} p(x_1)p(X_3|x_1, x_2)$$



DAG G

$$p(X_3 \mid x_2) = \sum_{x_1} p(x_1 \mid x_2)p(X_3 \mid x_1, x_2)$$

Gender-blind Hiring Policy



Goal: effect of enforcing policy
$P(H = \text{hire} \mid \mathbf{do}(P = \text{on}))$

Problem: confounding by Culture
(Culture $\rightarrow$ Policy and Culture $\rightarrow$ Hiring)

Key point: Conditioning on P (observing) keeps the path $C \rightarrow P \rightarrow H$ open while intervention $\mathbf{do}(P = \text{on})$ cuts $C \rightarrow P$.

Policy effect

What observation gives (not causal):

$$P(H{=}1 \mid P{=}\text{on}) = \sum_c P(H{=}1 \mid P{=}\text{on}, C{=}c)\ P(C{=}c \mid P{=}\text{on})$$

This can be *biased* if C (e.g., team culture) influences both policy adoption and hiring.

What intervention asks for (causal):

$$P(H{=}1 \mid \mathbf{do}(P{=}\text{on})) = \sum_c P(H{=}1 \mid P{=}\text{on}, C{=}c)\ P(C{=}c)$$

⇒ Replace $P(C{=}c \mid P{=}\text{on})$ with the *marginal* $P(C{=}c)$
[**backdoor adjustment** for the confounder C]

Interpretation. Observing $P{=}$on answers "who chose the policy?"
Intervening $\mathbf{do}(P{=}\text{on})$ answers "what if everyone were forced to use it?"

Structural Causal Model

Definition

An SCM $\mathcal{M} = (F, p_U)$ consists of

i) a set F of n assignments
(the structural equations)

$$F = \{X_i := f_i(PA_i, U_i),\, i = 1, \ldots, n\}$$

$PA_i \subseteq \{X_1, \ldots, X_n\} \setminus X_i$: causal parents
U_i's: noise variables

ii) a joint distribution $p_U(U_1, \ldots, U_n)$

Linking SCM's and CGM's

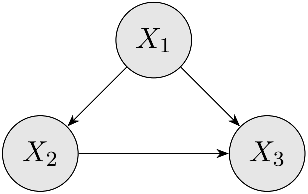
Definition

The causal graph G *induced* by an SCM is the directed graph with vertex set $\{X_1, \ldots, X_n\}$ and a directed edge from each vertex in PA_i to X_i for all i.

Example

SCM over $\{X_1, X_2, X_3\}$ with some $p_U(U_1, U_2, U_3)$

$$X_1 := f_1(U_1),\ X_2 := f_2(X_1, U_2),\ X_3 := f_3(X_1, X_2, U_3)$$



induced DAG G

Difference between CGM and SCM

SCM allows for a rich class of causal models
including models with
cyclic causal relations
not obeying the causal Markov condition (because of complex covariance structures between the noise terms)

Usually the following assumptions are added

A1) Acyclicity: the induced graph G is a DAG

A2) Causal sufficiency/no hidden confounders: the U_i's are jointly independent, i.e.

$$p_U(U_1, \ldots, U_n) = p_{U_1}(U_1) \times \ldots p_{U_n}(U_n)$$

Acyclicity and Causal sufficiency ensure that the distribution induced by an SCM *factorises* according to its induced causal graph G (and the Markov condition is satisfied w.r.t. G)

Interventions in SCM's

Definition

An intervention $do(X_i = x_i)$ in an SCM $\mathcal{M} = (F, p_U)$ is modeled by

- replacing the i-th structural equation in F by $X_i = x_i$
- remaining F_j's remain unchanged ($j \neq i$)

Result is the interventional SCM $\mathcal{M}^{do(X_i = x_i)} = (F', p_U)$.

From $\mathcal{M}^{do(X_i = x_i)} = (F', p_U)$
deduce the interventional distribution $p(X_{-i} \,|\, do(X_i = x_i))$
and the intervention graph G'

Interventions in SCM

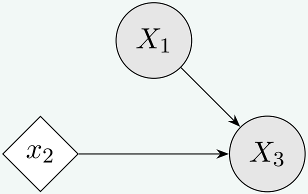
SCM $\mathcal{M} = (F, p_U)$

$$X_1 := f_1(U_1),\ X_2 := f_2(X_1, U_2),\ X_3 := f_3(X_1, X_2, U_3)$$

SCM $\mathcal{M}^{do(X_2=x_2)} = (F', p_U)$

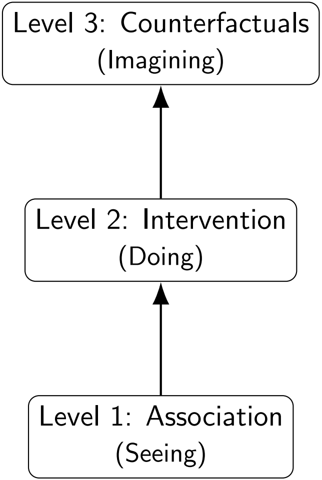
$$X_1 := f_1(U_1),\ X_2 := x_2,\ X_3 := f_3(X_1, X_2, U_3)$$

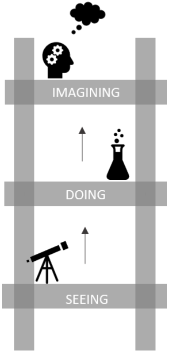
Graph G' induced by $\mathcal{M}^{do(X_2=x_2)}$



Seeing, Doing, Imagining

The ladder of causality





Observations, interventions, counterfactuals

i) observation
passively seen or measured

ii) intervention
external manipulation or experimentation

iii) counterfactual
what *would* have been, given that something else was in fact observed

Issues with counterfactuals

Cannot be observed empirically

unfalsifiable

unscientific (Popper, 1959)

problematic (Dawid, 2000)

Yet, humans seem to perform counterfactual reasoning in practice starting in early childhood (Buchsbaum et al., 2012)

Counterfactuals

"Given that patient X received treatment A and their health got worse, what would have happened if they had been given treatment B instead, *all else being equal*?"

- SCMs provide a suitable framework for counterfactual reasoning
- Observing what actually happened
  provides information about the *background state* of the system
  namely the noise terms $\{U_1, \ldots, U_n\}$ in an SCM
- This differs from an intervention where such background information is not available

Definition (Counterfactuals in SCM's)

Given evidence $X = x$ observed from an SCM $\mathcal{M} = (F, p_U)$
the counterfactual SCM $\mathcal{M}^{X=x}$ is obtained by updating p_U to $p_{U \mid X=x}$

$$\mathcal{M}^{X=x} = (F, p_{U \mid X=x})$$

Counterfactuals are then computed by performing interventions in the counterfactual SCM $\mathcal{M}^{X=x}$

Computing counterfactuals with SCM: Example

$SCM\ \mathcal{M} = (F, p_U)$

$X := U_X,\ Y := 3X + U_Y;\ U_X, U_Y \overset{iid}{\sim} N(0,1)$

We observe $X = 2$ and $Y = 6.5$
and want to answer the counterfactual question
"What would Y have been, had $X = 1$?"

We are thus interested in

$$p^{\mathcal{M}^{X=2, Y=6.5; do(X=1)}}(Y)$$

Example ctd

Recall: $X := U_X$, $Y := 3X + U_Y$, $U_X, U_Y \overset{iid}{\sim} N(0,1)$

- Update the noise distribution $p_U \to p_{U \mid X=2, Y=6.5}$

$$U_X \sim \delta(2),\; U_Y \sim \delta(0.5)$$

- Obtain the updated SCM $\mathcal{M}^{X=2,Y=6.5} = (F, p_{U \mid X=2,Y=6.5})$
- Perform the intervention $do(X = 1)$ on $\mathcal{M}^{X=2,Y=6.5}$

$$p^{\mathcal{M}^{X=2,Y=6.5;do(X=1)}}(Y) = \delta(3.5)$$

- Above differs from the interventional distribution $Y \mid do(X = 1) \sim N(3,1)$

Factorizations

Altitude and Temperature

- Disentangled factorization

$$p(A,T) = p(A)p(T \mid A)$$

- Entangled factorization

$$p(A,T) = p(T)p(A \mid T)$$

In the disentangled factorization some components generalize across domains

- Austria and Switzerland (CH)

$$p_{Austria}(A,T) = p_{Austria}(A)p(T \mid A)$$
$$p_{CH}(A,T) = p_{CH}(A)p(T \mid A)$$

$p(T \mid A)$ is likely to be the same across these two countries

$p(A)$ is country-specific

Independent causal mechanisms

For a model to correctly predict the effect of interventions, it needs to have components that are robust when moving from an observational distribution to certain interventional distributions.

Principle (Independent Causal Mechanisms (ICM))

The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other.

In the two-variable case, say (A, T), it reduces to independence between

- the cause distribution $p(A)$
- the mechanism producing the effect from the cause $p(T \mid A)$

Principle (Sparse Mechanism Shift)

Small distribution changes manifest in a sparse or local way in the causal/disentangled factorization; i.e., they should usually not affect all factors simultaneously.

Table of Contents

Causal discovery

- So far we assumed a given causal DAG, possibly based on domain knowledge
- Often domain knowledge only incomplete or unavailable
- Need to learn the causal DAG
  Typically using observational (passive) data which are abundant
- Hopeless?
- Surprisingly the problem becomes *easier* when the number of variables becomes *higher*
  because there are nontrivial *conditional independence* properties among the variables implied by the causal structure
- Several methods
  constraint-based methods
  score-based methods
- Also can use observational and interventional data

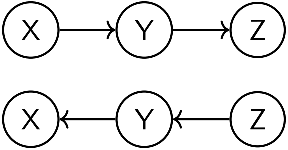
Markov equivalence

Definition (Markov equivalence)

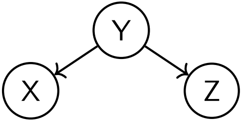
Two DAGs are said to be Markov equivalent if they encode the same conditional independence (CI) statements.
The set of all DAGs encoding the same CI's is called a Markov equivalence class

Chains, forks and colliders

(a) Chains

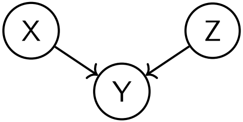


(b) Fork



(a) and (b) imply $X \perp\!\!\!\perp Z \mid Y$
(and no others)

(c) Collider
(v-structure)



(c) implies $X \perp\!\!\!\perp Z$
(but $X \not\perp\!\!\!\perp Z \mid Y$)

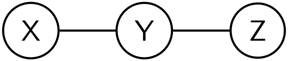
$\{(a)\} \bigcup \{(b)\}$: Markov equiv class
(c): distinct equivalence class

Markov equivalence: characterization

Result

Two DAG's are Markov equivalent iff they have the same skeleton and the same v-structures

Skeleton of Chains (a), Fork (b) and Collider (c)



v-structures

(a) and (b)	NO
(c)	YES

Table of Contents

Causal Reasoning

Causal reasoning starts from a known (or postulated) causal graph and answers causal queries of interest

Two steps

(i) *identify* the query, i.e., derive an estimand
[a well-defined expression in terms of *observable* quantities]

(ii) *make inference* on the estimand using data

Definition (Treatment effect)

Outcome Y and binary treatment T

$$\tau := \mathbb{E}[Y \mid do(T = 1)] - \mathbb{E}[Y \mid do(T = 0)]$$

Outcome Y and continuous X

$$\tau(x') := \left[\frac{\partial}{\partial x} \mathbb{E}[Y \mid do(X = x)] \right]_{x=x'}$$

Treatment effects involve interventional expressions
Causal reasoning answers queries
using observational data together with a causal model

Identification

Given a causal graph and no hidden confounders

The causal effect can be identified through the interventional distribution

$$p(X_1, \ldots, X_n \mid do(X_i = x_i)) = \delta(x_i) \prod_{j \neq i} p(X_j \mid PA_j) \qquad \text{(g)}$$

The interventional distribution of any X_h $(h \neq i)$ can be obtained by *marginalization*

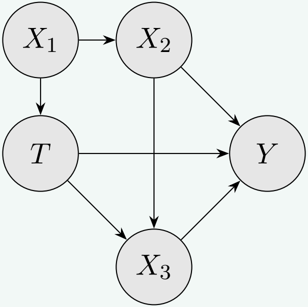
Remarks

Formula (g) has been named

- *g-formula*
  Robins (1986)
- *truncated factorization*
  Pearl (2000, 2009)

It relies on the independence of causal mechanisms
i.e. intervening on a variable leaves the remaining causal conditionals unaffected

Evaluation of treatment effect with three covariates $\{X_1, X_2, X_3\}$



Factorization of interventional distribution

$$p(y, t, x_1, x_2, x_3 \mid do(T = t)) = \delta(t)p(x_1)p(x_2 \mid x_1)p(y \mid x_2, x_3, t)p(x_3 \mid x_2, t)$$

Target distribution $\quad p(y \mid do(T = t))$

Adjustment set

$$
\begin{aligned}
p(y \mid do(T=t)) &= \sum_{x_1,x_2,x_3} p(y,t,x_1,x_2,x_3 \mid do(T=t)) \\
&= \sum_{x_2}\sum_{x_1} p(x_2 \mid x_1)p(x_1) \sum_{x_3} p(y \mid x_2,x_3,t)p(x_3 \mid x_2,t) \\
&= \sum_{x_2} p(x_2)p(y \mid x_2,t)
\end{aligned}
$$

x_2 is a valid adjustment set

More adjustment sets

It can be proved using graphical criteria or otherwise that

$$Y \perp\!\!\!\perp X_1 \mid \{T, X_2\} \tag{1.a}$$

$$X_2 \perp\!\!\!\perp T \mid X_1 \tag{1.b}$$

$$p(y \mid do(T = t)) = \sum_{x_1, x_2} p(x_1, x_2) p(y \mid x_1, x_2, t), \text{ using (1.a)} \tag{2.a}$$

$$= \sum_{x_1} p(x_1) \sum_{x_2} p(x_2 \mid x_1, t) p(y \mid x_1, x_2, t), \text{ using (1.b)}$$

$$= \sum_{x_1} p(x_1) p(y \mid x_1, t) \tag{2.b}$$

Both $\{x_1, x_2\}$ by (2.a) and $\{x_1\}$ by (2.b) are valid adjustment sets. However $\{x_1, x_3\}$ is not.

Adjustment sets

Whenever

$$p(y \mid do(T = t)) = \sum_z p(z) p(y \mid z, t) \quad (3)$$

z is called a valid adjustment set

Under causal sufficiency (no hidden variables) there exist graphical criteria to find valid adjustment sets

To estimate the involved quantities in (3)
additional assumptions are required
in particular *overlap*:
for any t and feature values $\mathbf{x}$, $\mathbf{X}$, $0 < p(T = t \mid \mathbf{X} = \mathbf{x}) < 1$

Optimal adjustment sets

Henckel *et al* 2022

Back to Counterfactual Inference

Recall the basic **three steps**

1. **Abduction:** Update beliefs about background variables U given evidence
2. **Action:** Modify the model by applying the intervention (replace structural equation(s) for intervened variables)
3. **Prediction:** Propagate the modified model forward using the updated distribution for U

Twin network for counterfactuals

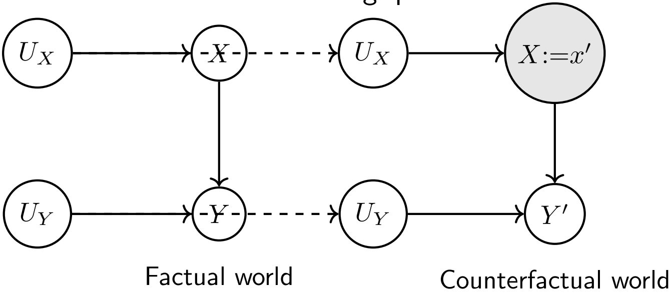
Abduction step requires large computational resources.

Even if we start with U_i's mutually independent, conditioning on evidence typically destroys this independence.

Necessary to carry over a full description of the joint distribution.

A solution is represented by a twin network

It consists of two interlinked networks, one representing the real world and the other the counterfactual world being queried.



Dashed edges: the same exogenous variables U feed both worlds.
Counterfactual world differs only by the intervention $\mathbf{do}(X{=}x')$.

Deep Twin Networks

Nature Machine Intelligence | Volume 5 | February 2023 | 159–168 **160**

Article https://doi.org/10.1038/s42256-023-00611-x

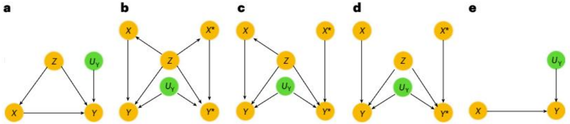


Fig. 1 | Construction and interventions on Twin Networks. Orange nodes are observed, green latent. **a**, Example SCM; **b**, twin network of **a**; **c**, intervention in the twin network on node X^*; **d**, interventions in the twin network on X and X^*; **e**, unconfounded version of **a**.

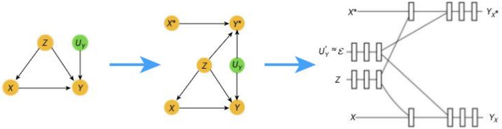


Fig. 2 | From DAG to twin network DAG to deep neural network architecture for binary *X*, *Y*. Rectangular blocks are neural network blocks, such as fully convolutional network layers or lattices; forward intersections are concatenations of features.

nature machine intelligence

Article

https://doi.

Estimating categorical counterfactuals via deep twin networks

Received: 23 May 2022

Accepted: 4 January 2023

Published online: 20 February 2023

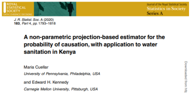
Check for updates

Athanasios Vlontzos [1,5], Bernhard Kainz [1,2] & Ciarán M. Gilligan-Lee [3,4]

Counterfactual inference is a powerful tool, capable of solving challenging problems in high-profile sectors. To perform counterfactual inference, we require knowledge of the underlying causal mechanisms. However, causal mechanisms cannot be uniquely determined from observations and

- idea of counterfactual ordering for causal models with categorical variables
  [posits desirable properties that causal mechanisms should possess]
  [avoids counterfactual prediction that can conflict with domain knowledge]
- proof that counterfactual ordering is equivalent to specific functional constraints on the Structural Causal Model
- ML computation *via* deep twin networks
  [deep neural networks that can perform twin network counterfactual inference.]

Kenyan Water dataset



J. R. Statist. Soc. A (2020)
183, Part 4, pp. 1793–1818

A non-parametric projection-based estimator for the probability of causation, with application to water sanitation in Kenya

Maria Cuellar
University of Pennsylvania, Philadelphia, USA
and Edward H. Kennedy
Carnegie Mellon University, Pittsburgh, USA

nature machine intelligence

Article

Estimating categorical counterfactuals via deep twin networks

Received: 23 May 2022
Accepted: 4 January 2023

Athanasios Vlontzos, Bernhard Kainz & Ciarán M. Gilligan-Lee

Counterfactual inference is a powerful tool, capable of solving challenging problems in high-profile sectors. To perform counterfactual inference, we require knowledge of the underlying causal mechanisms. However, causal mechanisms cannot be uniquely determined from observations and

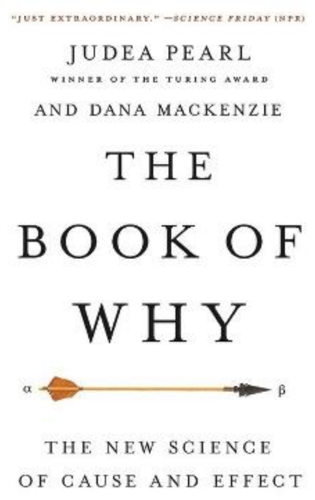
- Problem: Children in Western Kenya were exposed to high bacterial concentration in their drinking water and developed diarrheal disease. What is the probability that it was the bacterial exposure (and not something else) that caused the disease?

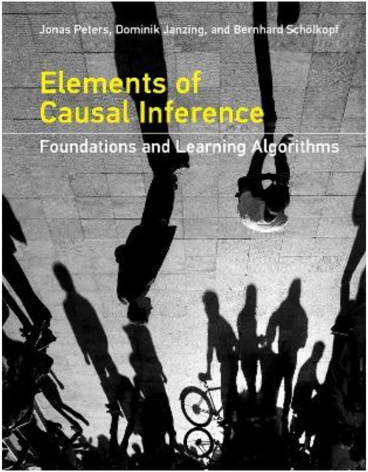
  This is the probability of causation (PC).
- We are *not* asking whether the exposure causes the outcome *on average* (the usual randomized-trial question), which is also important.
- Knowing whether the exposure *actually* caused the outcome (for those exposed who had the outcome) matters for policy impact on the target population.
- In this case policy means: protecting water springs by installing pipes and concrete containers
- It was found that the PC was very low suggesting that protecting water springs is not likely to have an effect on the development of children diarrhea in these populations and that the source of the disease is not related to water.

Table of Contents

- Over the recent years ML and AI have achieved remarkable success to make accurate predictions
- As these systems become increasingly integrated into high-stakes applications
  [medical diagnosis, autonomous driving]
  they face severe limitations
  [changes in the environment used in the training set or changes in the data generating system through external intervention]
- To face these challenges causal models are needed
- Causal graphical models promise to be especially suitable when applied to AI systems to enable
  robustness to changes in the environment
  [sparse causal mechanism]
  interpretability
  [transparent reasoning and human oversight]
  explainability
  [allowing counterfactual questions such as “Would the outcome have been different if we had acted otherwise?]

Selected References





From Statistical to Causal Learning

Bernhard Schölkopf
Max Planck Institute for Intelligent Systems, Tübingen, Germany
bs@tuebingen.mpg.de

Julius von Kügelgen
Max Planck Institute for Intelligent Systems, Tübingen, Germany
University of Cambridge, United Kingdom
jvk@tuebingen.mpg.de

April 4, 2022

Abstract

We describe basic ideas underlying research to build and understand artificially intelligent systems: from symbolic approaches via statistical learning to interventional models relying on concepts of causality. Some of the hard open problems of machine learning and AI are intrinsically related to causality, and progress may require advances in our understanding of how to model and infer causality from data.*

Mathematics Subject Classification 2020

Primary 68T05; Secondary 68Q32, 68T01, 68T10, 68T30, 68T37

Keywords

Causal inference, machine learning, causal representation learning