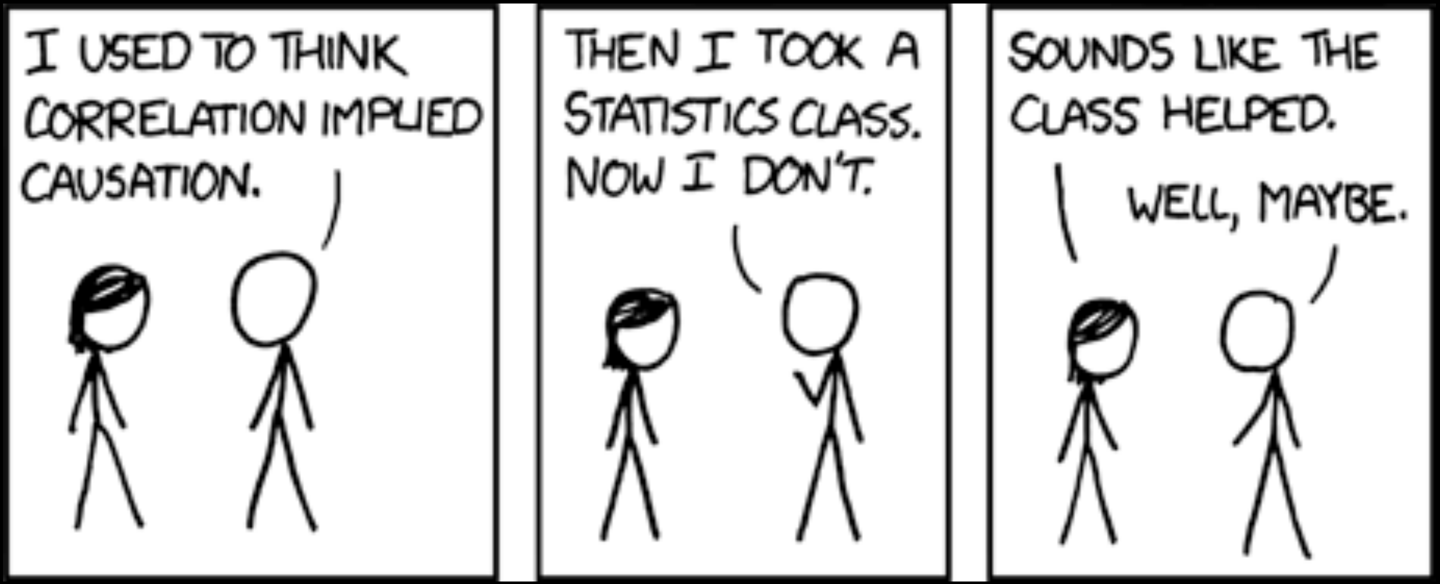
Beyond Correlation

Causality & Counterfactual Reasoning

—

Tanmayee Narendra
IBM Research

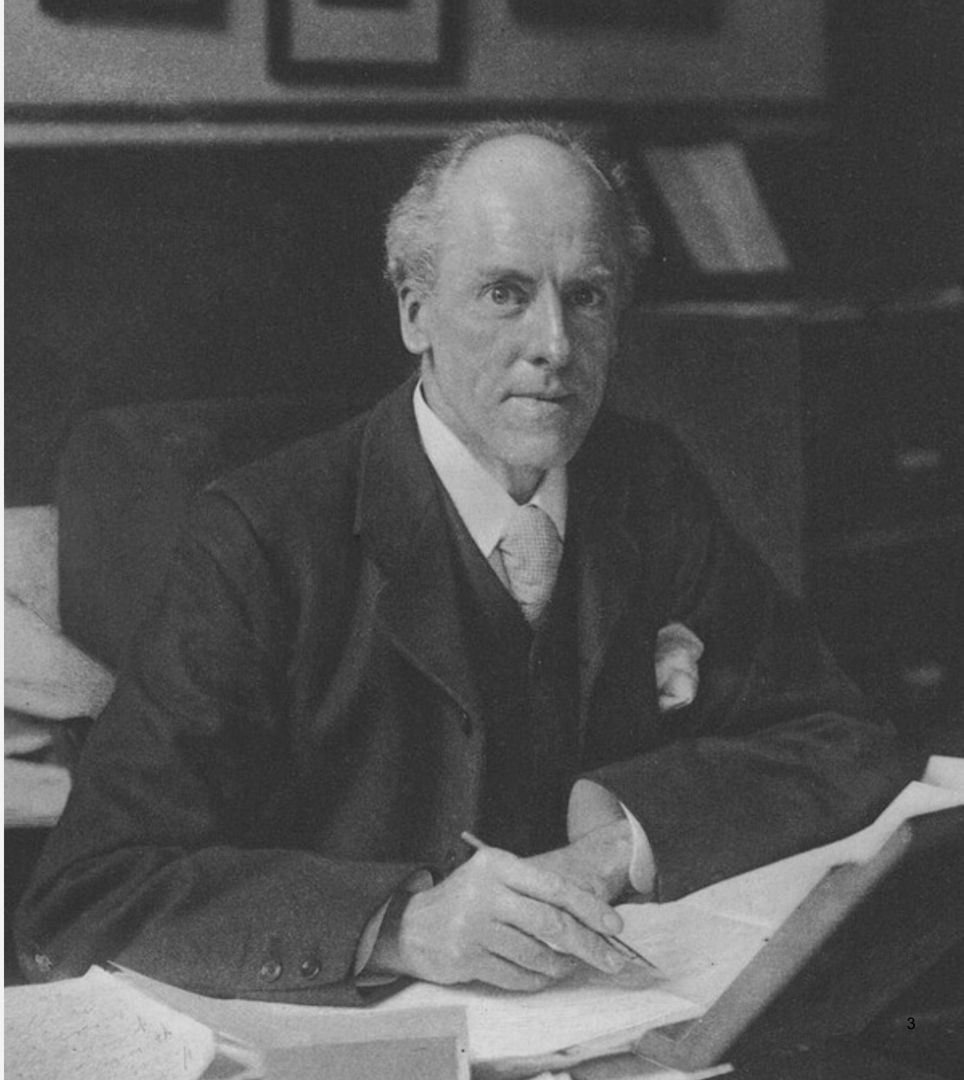
Obligatory Comic



https://xkcd.com/552/

Karl Pearson

- “Beyond such discarded fundamentals as ‘matter’ and ‘force’ lies still another fetish amidst the inscrutable arcana of modern science, namely, the category of cause and effect.”
- He categorically denied the need for an independent concept of causal relation beyond correlation.

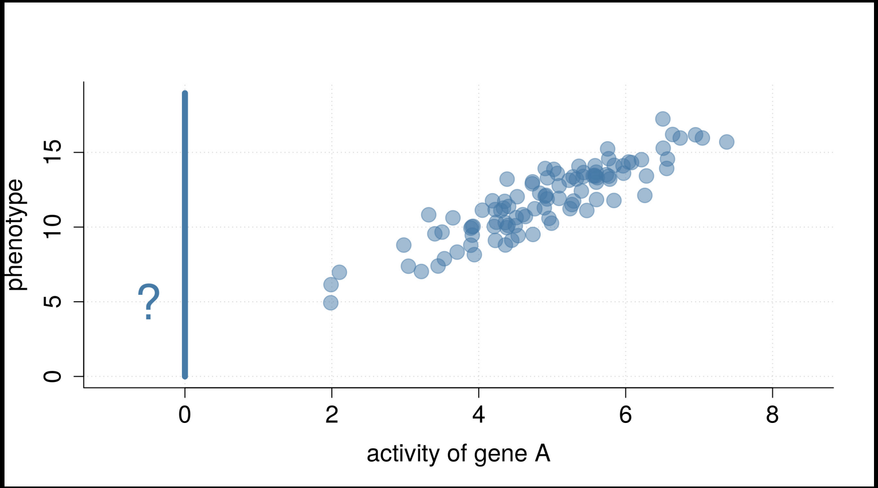


Pearl, Judea. *Causality*. Cambridge university press, 2009.

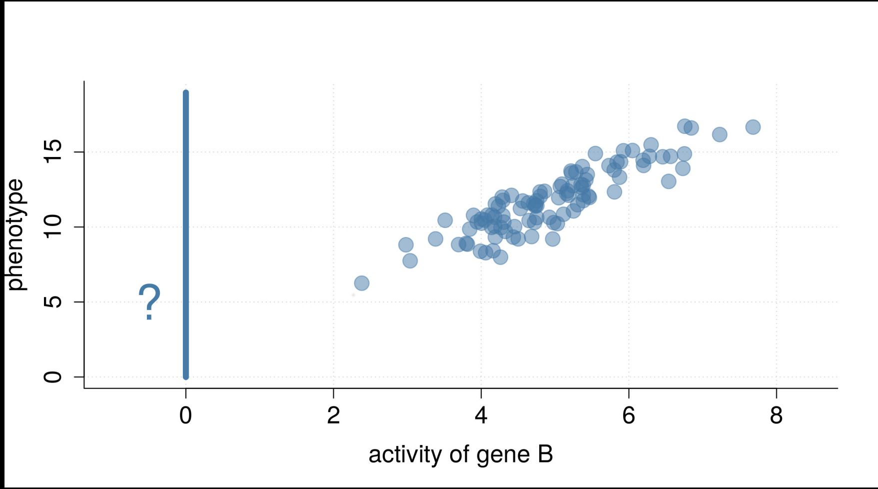
Why Causality?

An Example

Gene A



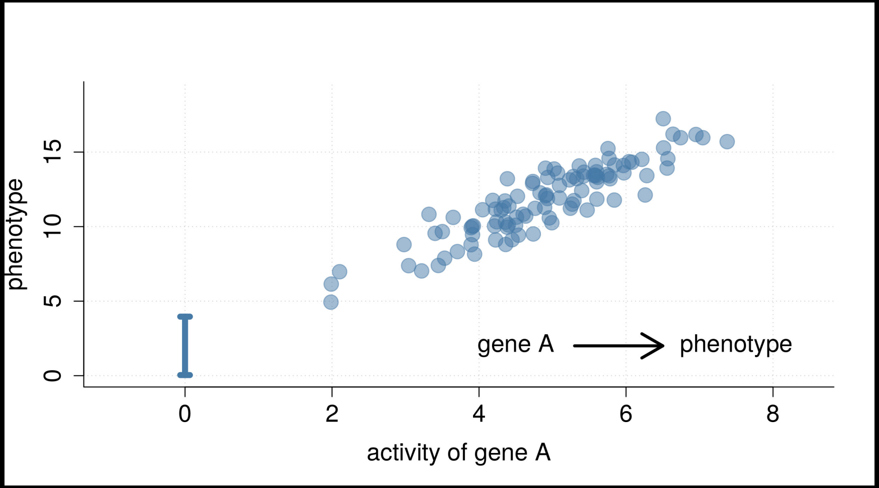
Gene B



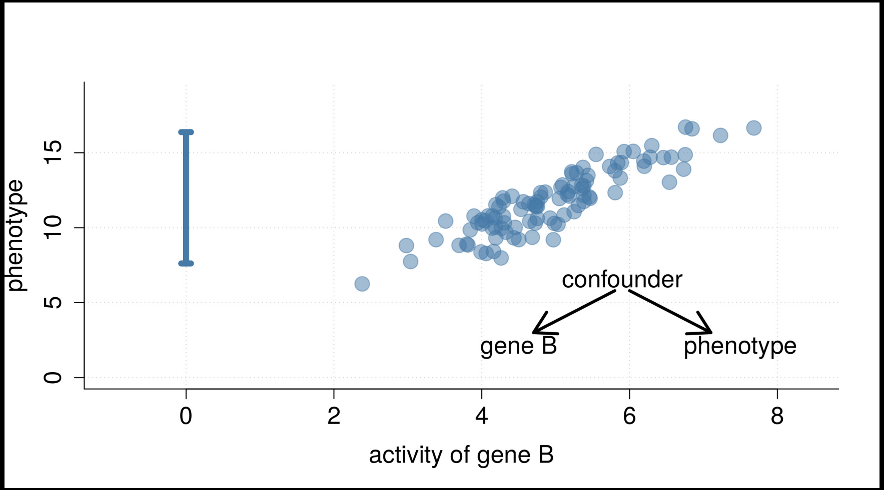
Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

An Example

Gene A

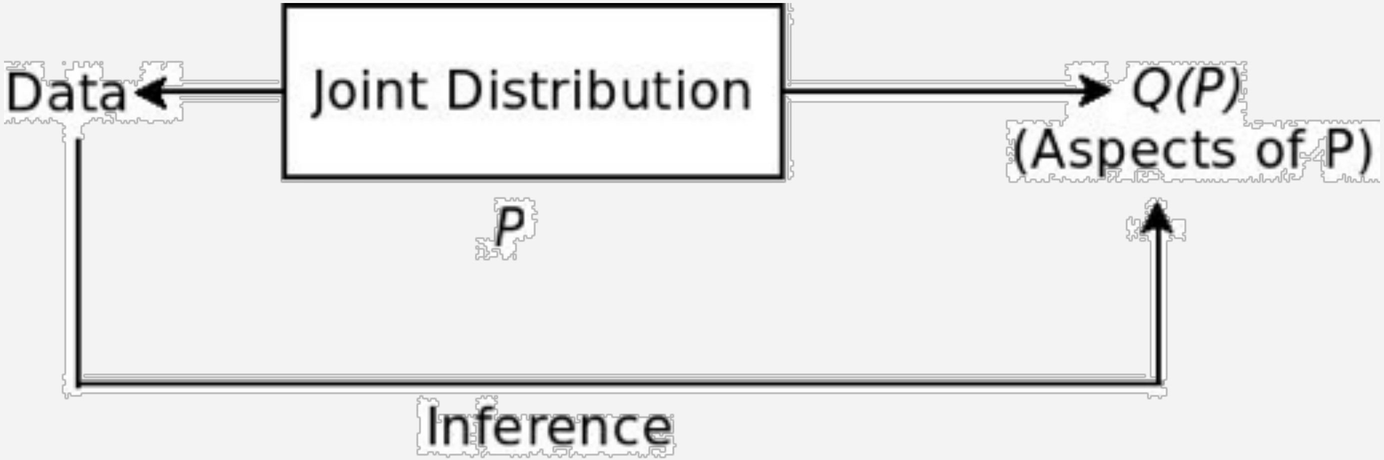


Gene B



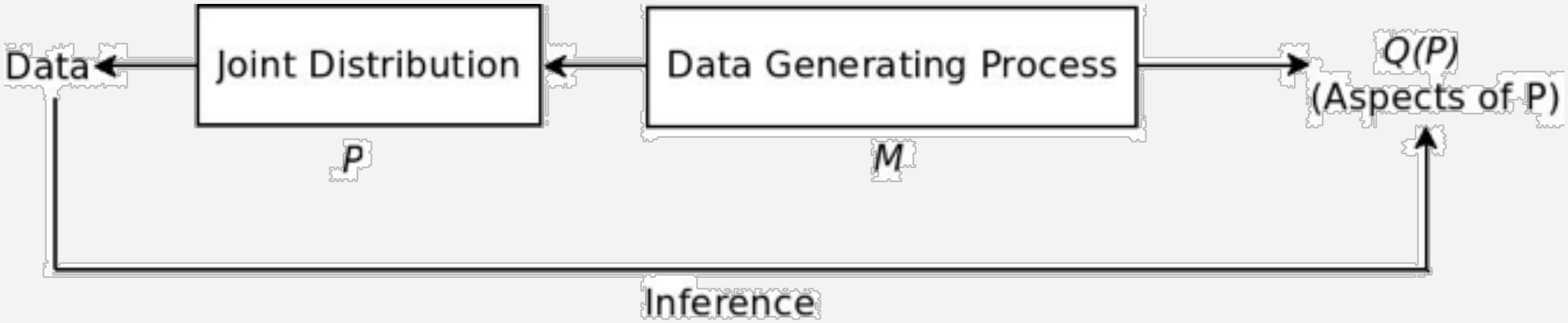
Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

Statistical Inference Paradigm



Pearl, Judea. *Causality*. Cambridge university press, 2009.

Causal Inference Paradigm



Pearl, Judea. *Causality*. Cambridge university press, 2009.

Questions a Causal Model must Answer

What if we *see* A?

(Observation)

What if we *do* A?

(Intervention)

What if we *did things differently?*

(Counterfactual)

Causal Model

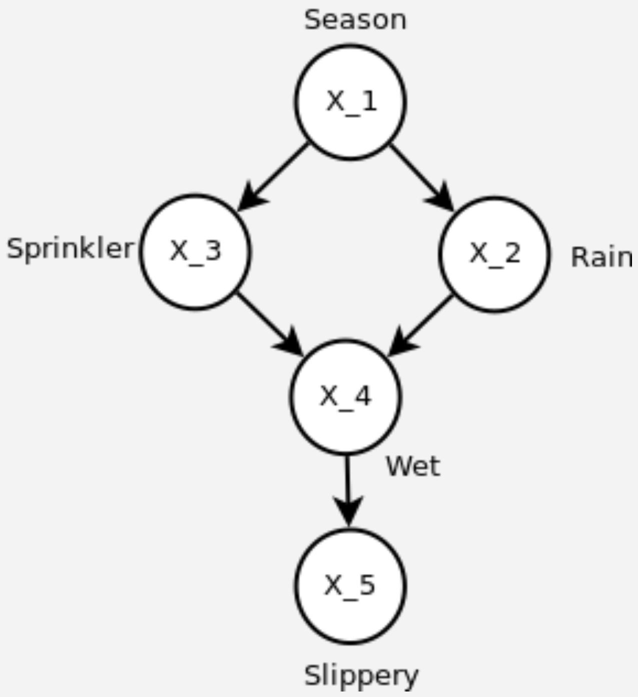
Causal Graph

Distribution

Intervention Distributions

Counterfactual

Causal Bayesian Network

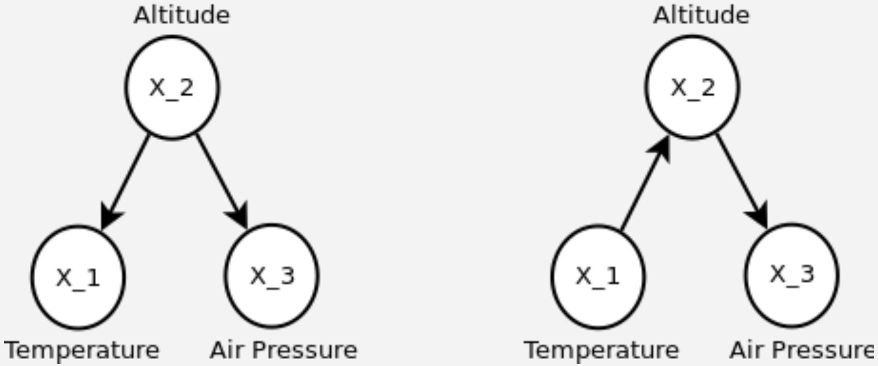


$$P(X) = \prod_{i=1}^{n} P\big(X_i | Pa(X_i)\big)$$

where $Pa(X_i)$ stands for parents of X_i

Pearl, Judea. *Causality*. Cambridge university press, 2009.

Difference between BN and CBN

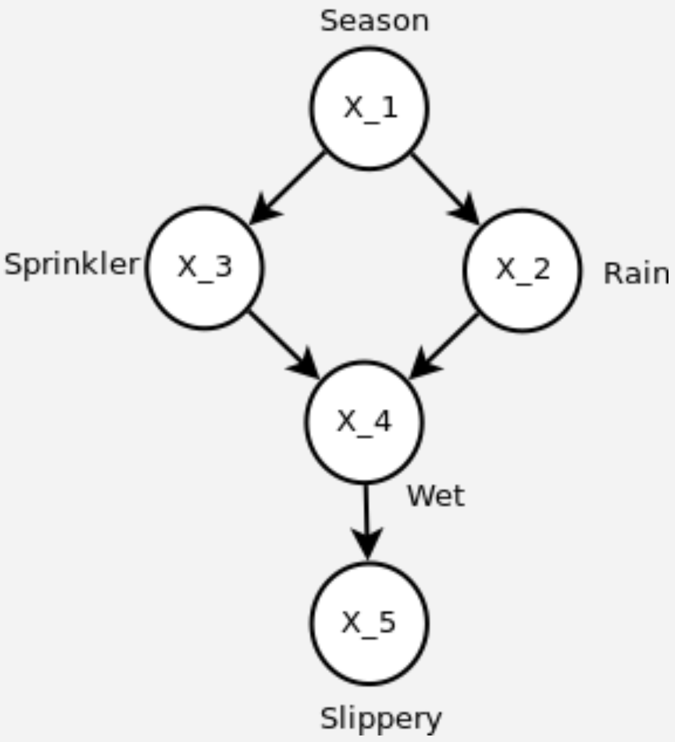


These two graphs entail the same probability distributions, but their causal interpretations are different.

Pearl, Judea. *Causality*. Cambridge university press, 2009.

Prediction

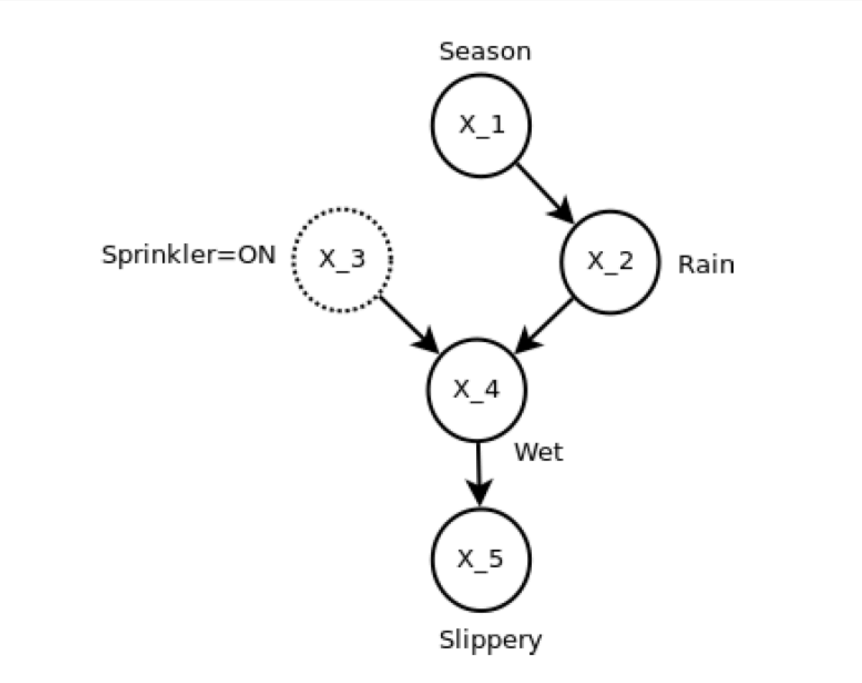
Would the pavement be slippery if we find the sprinkler off?



Pearl, Judea. *Causality*. Cambridge university press, 2009.

Intervention

Would the pavement be slippery if we make sure the sprinkler is on?



Pearl, Judea. *Causality*. Cambridge university press, 2009.

Structural Causal Models

- Each observed variable is a function of a subset of variables
- This subset corresponds to parents in the causal graph
- U_i are *unknown* variables which are jointly independent

- $X_1 := U_1$
- $X_2 := f_1(X_1, U_2)$
- $X_3 := f_1(X_1, U_3)$
- $X_4 := f_1(X_2, X_3, U_4)$
- $X_5 := f_1(X_4\ U_5)$

Causal Bayesian Networks and Structural Causal Model (SCM) are canonically equivalent

Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

Learning Causal Models from Data

Probabilistic Approaches

- Score based methods – Greedy Equivalence Search
- Conditional independence methods – PC Algorithm, SGS Algorithm
- Information Geometry approaches

Parametric Approaches

- Linear Gaussian Acyclic Models (LinGAM)
- Additive Noise Models (ANM)
- Causal Additive Models (CAM)

Some Examples

Eye Disease

Consider a treatment for eye disease –

- For 99% of patients, the treatment works and the patient is cured $(T = 1, B = 0)$
  - If untreated, they turn blind in a day $(B = 1)$
- For the remaining 1%, if given the treatment, they turn blind in a day $(T = 1, B = 1)$
  - If untreated, they regain normal vision
- The category of a patient is controlled by a rare condition $(N_B = 1)$, that is unknown to the doctor

Assume the underlying SCM $\mathfrak{C}$–

- $T := N_T$
- $B := T \cdot N_B + (1 - T) \cdot (1 - N_B)$

- Now, imagine a patient with poor eyesight who goes blind after the treatment
- What would have happened if the doctor administered treatment $T = 0$?
- We need to compute $do(T := 0)$.
- This is a counterfactual question

Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

Eye Disease

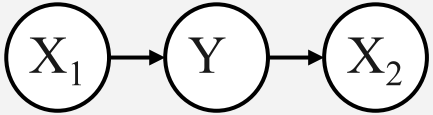
- Underlying SCM $\mathfrak{C}$ –
  - $T := N_T$
  - $B := T \cdot NB + (1 - T) \cdot (1 - NB)$
- For the given patient, $N_B = 1$
- To answer the counterfactual, first update the distribution based on the observation
  - $T := 1$
  - $B := T \cdot 1 + (1 - T) \cdot (1 - 1) = T$

- Now, calculate the effect of $do(T := 0)$
  - $T := 0$
  - $B := T$
- Clearly, $P^{\mathfrak{C}|B=1,T=1;do(T:=0)}(B = 0) = 1$
- The patient would have been cured if the doctor had not given him the treatment

Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

Correlation and Causation

- Consider the SCM $\mathfrak{C}$ –
  - $X_1 := N_1$
  - $Y := X_1 + N_Y$
  - $X_2 := Y + N_2$
- With $N_1, N_Y \sim \mathcal{N}(0,1)$ and $N_2 \sim \mathcal{N}(0,0.1)$

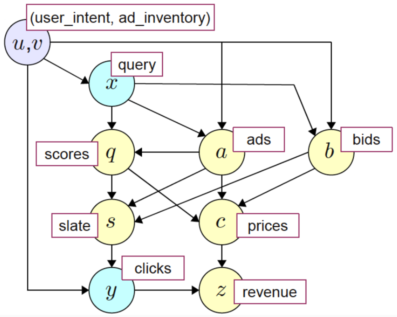


- If we are interested in predicting Y from X_1 and X_2 then, X_2 is a better predictor
- If we want to change Y however, interventions on X_2 are useless

Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. MIT press, 2017.

Computational Advertisement

$$
P\begin{pmatrix} u,v,x,a,b \\ q,s,c,y,z \end{pmatrix} = \left\{ \begin{array}{ll}
P(u,v) & \text{Exogenous vars.} \\
\times\, P(x\,|\,u) & \text{Query.} \\
\times\, P(a\,|\,x,v) & \text{Eligible ads.} \\
\times\, P(b\,|\,x,v) & \text{Bids.} \\
\times\, P(q\,|\,x,a) & \text{Scores.} \\
\times\, P(s\,|\,a,q,b) & \text{Ad slate.} \\
\times\, P(c\,|\,a,q,b) & \text{Prices.} \\
\times\, P(y\,|\,s,u) & \text{Clicks.} \\
\times\, P(z\,|\,y,c) & \text{Revenue.}
\end{array} \right.
$$



- Complex system with several ML components, and actors with varied interests
- Causal Modelling helps in design of the system, by making it principled and cheaper
- Will I get a higher click-through rate if I change my machine learning model for ad-placement?
- Changing the ad-placement algorithm is an intervention in the system
- Use the causal graph to track the consequences of these interventions

Bottou, Léon, et al. "Counterfactual reasoning and learning systems: The example of computational advertising." *The Journal of Machine Learning Research* 14.1 (2013): 3207-3260.

Exoplanet Search

- Removing systemic noise from observations of the Kepler space observatory
- New technique called Half-Sibling Regression

Schölkopf, Bernhard, et al. "Removing systematic errors for exoplanet search via latent causes." *International Conference on Machine Learning*. 2015.

Advances in Causal Inference

- UCLA - Judea Pearl
- CMU - Peter Spirtes, Clark Glymour, Richard Scheines
- Harvard - Donald Rubin
- ETH-Zurich - Jonas Peters, Peter Buhlmann, Nicolai Meinshausen, Stefen Bauer
- MPI-Tubingen - Dominik Janzing, Bernhard Scholkopf
- Others - Joris Mooij, Patrik Hoyer and many others

Thank You

Tanmayee Narendra
Software Engineer, IBM Research
—
n.tanmayee@gmail.com
triptoes1.github.io