OAK RIDGE National Laboratory

Introduction to Causality

Ayana Ghosh, Zachary Fox

Computational Sciences & Engineering Division

Oak Ridge National Laboratory, USA

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

Diagnostic Evaluation:

Let's put down our thoughts on causality

- Have you heard about this before and in which context ?
- Any related topic that might have sparked some interest ? Maybe a popular phrase starting with "correlation does not imply…"
- Is this even important ?
- Is it just statistics ?

Few *causal* statements from our daily lives

- His headache got better because he took an aspirin
- Her productivity increased because she spent less hours in social media
- She got the prestigious Fellowship because she went to Harvard

Are these statements truly *causal ?*

Let's try to imagine a counterfactual world…

- His headache got better because he took an aspirin

  He did not take an aspirin

- Her productivity increased because she spent less hours in social media

  Instead of spending less hours in social media, she could have done something different (Not quiet clear from the statement)

- She got the prestigious Fellowship because she went to Harvard

  Do we change the university she went to ? Or there are other factors ?

Solving causal questions, thinking about causality

- We are typically interested in understanding the effects of manipulating a cause and predicting what outcomes might have occurred with or without that manipulation

- Our focus is on determining the impact of changing a cause and considering the potential outcomes of having made or not made that change

Are causal questions our only focus ?

Of course, yes!! I mean, for this part of the course! ☺

- Not always we care about answering causal questions
- This is where typical statistical modeling, ML/DL approaches fit in
- Focus is on identifying patterns in the data, statistical dependencies
- E.g., shows to binge on, products to buy, ground-state energy of a material etc.

Between statistical models & causal models

- Prediction, uncertainty estimates
  - Co-occurrences, expected values, joint/marginal/conditional probabilities, statistical dependencies

- Causal models build on concepts and techniques from statistical modeling but operate at a more advanced level than statistical information

Let's take a quick example…

We take a blood sample from a random sample of a population record on the activity of gene X and blood insulin levels Y

From statistical modeling we can answer:

- Average gene activity
- Probability of observing gene activity
- If gene activity score is observed as 4, what would be the blood insulin level ?

From statistical modeling we *cannot* answer:

- What happens if I change the gene expression
- Predict insulin level from gene activity in a different scenario

Causal Inference

- Given some understanding about the system/problem of interest, can we estimate a causal effect, i.e., identify and quantify causal relationships between variables ?

- Understand whether a change in one variable directly causes a change in another, distinguishing these causal effects from mere associations or correlations

- Randomized controlled trials are considered the gold standard
- E.g., Randomly chosen participants to receive either the vaccine or a placebo to determine the vaccine's effectiveness in preventing COVID-19 infection

- How do we do it ?

Ways to do Causal Inference

- Potential outcomes

Donald B. Rubin

- Structural Causal Models

Judea Pearl

Potential outcomes

Action: Aspirin (X = 1) or No Aspirin (X = 0)

Outcome: Headache gone (Y = 1) or Headache remains (Y = 0)

- Should I take an aspirin if I have a headache ?

- Two potential outcomes are possible!

Outcome if treated : $Y^{X=1}$, outcome not treated: $Y^{X=0}$

We define the causal effect as the difference in potential outcomes!

Causal effects

$$\text{Individual causal effect} = Y_i^{X=1} - Y_i^{X=0}$$

This is where the fundamental problem in causal inference comes from.
We can only observe one potential outcome per unit

$$\text{Average causal effect} = E[Y_i^{X=1} - Y_i^{X=0}] = E[Y^1] - E[Y^0]$$

Causal effects

	Potential outcomes		Individual causal effect
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$
Ayana	1	1	0
Zach	0	0	0
Sarah	1	0	1
Dennis	1	1	0
Kate	0	1	0
Thomas	1	0	1
Ray	1	0	1
Kevin	0	0	0

Average causal effect = $E[Y^1] - E[Y^0]$ = 5/8 - 3/8 = 0.25

OAK RIDGE National Laboratory

Causal effects

But we only observe one outcome per person!

	Potential outcomes		Individual causal effect (unobserved)	Observed	
	Y_i^1	Y_i^0	$Y_i^1 - Y_i^0$	X_i	Y_i
Ayana	1	1	0	0	1
Zach	0	0	0	1	0
Sarah	1	0	1	1	1
Dennis	1	1	0	0	1
Kate	0	1	0	0	1
Thomas	1	0	1	1	1
Ray	1	0	1	0	0
Kevin	0	0	0	1	0

$E[X=1] = (0+1+1+0)/4 = 0.5$; $E[X=0] = (1+1+1+0)/4 = 0.75$

$E(Y|X=1) - E(Y|X=0) = -0.25$

Naïve conclusion: Aspirin decreases chances of headache

OAK RIDGE National Laboratory

Did you notice any problem in our exercise ?

$$E(Y|X=1) - E(Y|X=0) = E(Y^1) - E(Y^0)$$

This may not be the full story!

- Observing that $E(Y|X=1) \neq E(Y|X=0)$ does not imply a causal effect of X on Y, observing ≠ intervening
- The average value of headache levels who did versus did not take aspirin are unequal
- We can use this in RCTs: $E(Y|X=1) - E(Y|X=0)$ to estimate average causal effect

How to proceed further ?

- Under what conditions can we draw the same kinds of conclusions from observational data as we would from a randomized controlled trial (RCT) ?

- There are two objectives to accomplish:

- **Identify** : I want to infer causal effect assuming I have access to population-level statistical information (information on variables, with infinite sample size)

- **Estimate** : Once I have identified the causal effect, how to estimate this effect from sample data

Key Assumptions

- Stable unit treatment value assumption (SUTVA)

  There are NO different versions of each treatment leading to different potential outcomes

  Important to go from individual causal effect to average causal effect

What happens otherwise if there are multiple ways to go from X=0 to X=1 ?

Key Assumptions

- Exchangeability

  Relates to the missing data problem in which you only have information on observed potential outcomes

  Treatment received (X) and the potential outcome given treatment Y^X are independent

  This is harder to achieve because in a non-randomized study, treatment may depend on other factors (also known as covariates)!

  Conditional exchangeability: Treatment received (X) and the potential outcome given treatment Y^X are independent with certain levels of Z

Key Assumptions

- Positivity

  Every possible combination of values of Z in the population should include both exposed and unexposed participants

  In RCTs, this is inherently present

  For other studies, we need to check between each covariate/treatment pair etc.

OAK RIDGE National Laboratory

All together

- For identifying causal relations:
  - SUTVA
  - Conditional Exchangeability
  - Positivity

- Let's write it down with the expectation notations

OAK RIDGE National Laboratory

All together

$$E\,[(Y(1) - Y(0)]$$

Potential outcomes are from own treatment & no other individual's treatment

$$= E\,[(Y(1)] - E[Y(0)]$$

$$= E_X\,[E(Y(1)|T=1,\,X] - E[Y(0)|T=0,X]]$$

Un-Confoundedness & conditional exchangeability

$$= E_X\,[E(Y|T=1,\,X] - E[Y|T=0,\,X]]$$

Consistency, potential outcome == observed

How to proceed further ?

- Under what conditions can we draw the same kinds of conclusions from observational data as we would from a randomized controlled trial (RCT) ?

- There are two objectives to accomplish:

- Identify : I want to infer causal effect assuming I have access to population-level statistical information (information on variables, with infinite sample size)

- **Estimate** : Once I have identified the causal effect, how to estimate this effect from sample data

How do we estimate the causal effect ?

Let's define a few terminologies

- Estimand – Quantity we want to estimate
  - Causal estimand ($E[Y(1) - Y(0)]$) ; contains potential outcomes
  - Statistical estimand ($E_X\,[E(Y|T=1, X] - E[Y|T=0, X]]$); observed data
- Estimate – Approximation of some estimand using data
- Estimation – process for getting data & estimand to estimate

OAK RIDGE National Laboratory

Causal → Statistics

We use the key assumptions to achieve identification-estimation flow

Strategies for estimation

- Different Approaches: Propensity scores matching, weighting
- Rubin world
  - Individual observations → individual causal effects → average causal effect

- Identify the target, measure confounders (relate to both treatment & outcome)
- Evaluate the assumptions
- Use covariate-based techniques for creating balanced groups
- Estimate average causal effect

Ways to do Causal Inference

- Potential outcomes

Donald B. Rubin

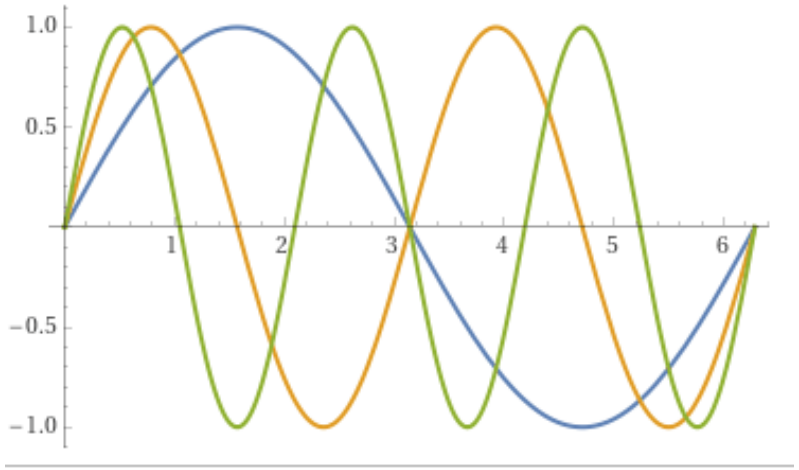
- Structural Causal Models

Judea Pearl

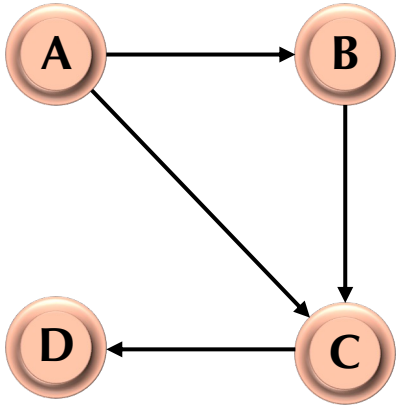
Structural Equation Models

- Graphical representation of the causal structure
- Causal relations and how to manipulate variables
- Somewhat easier to work with when we have lot of variables & many causal relations might be possible

Graph terminologies

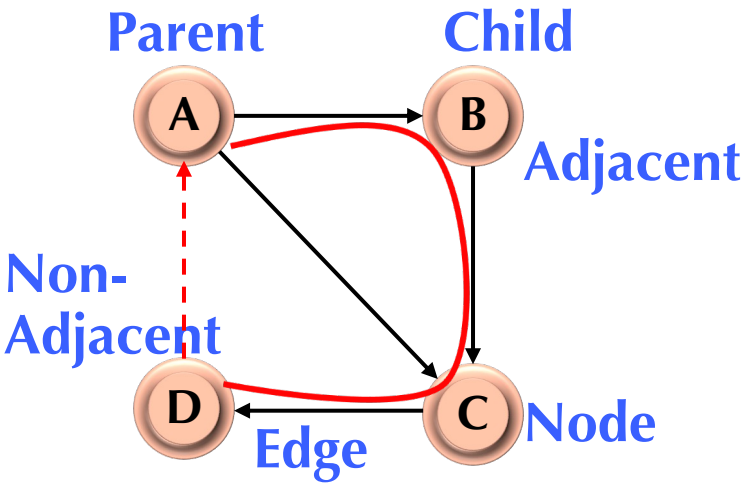


Not these ones



Something like this

Graph terminologies



Nodes

Edges

Ancestor

Descendant

Directed graph

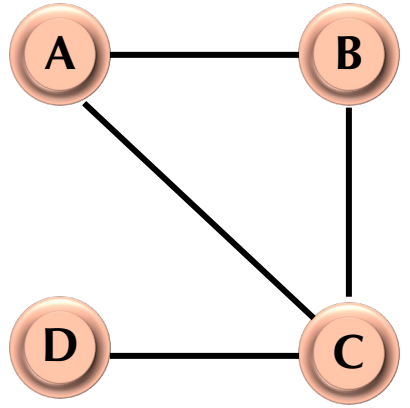
Path

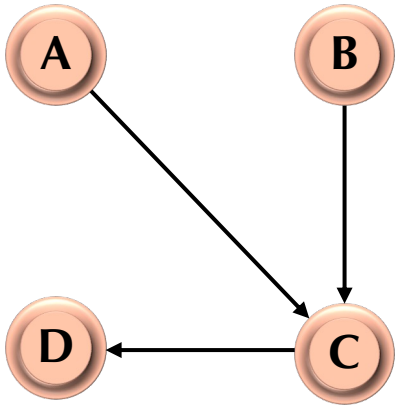
Immorality

Cycle

Acyclic

Undirected graph





Graph terminologies

Nodes

Edges

Ancestor

Descendant

Directed graph

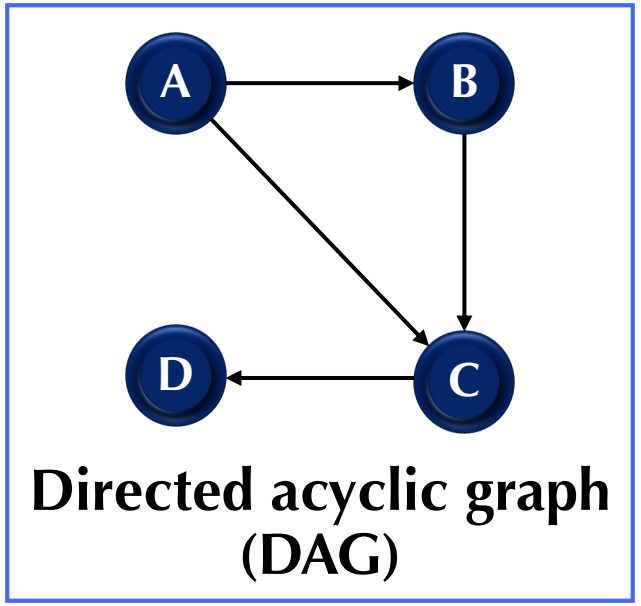
Path

Immorality

Cycle

Acyclic

Undirected graph



Directed acyclic graph (DAG)

Bayesian Networks

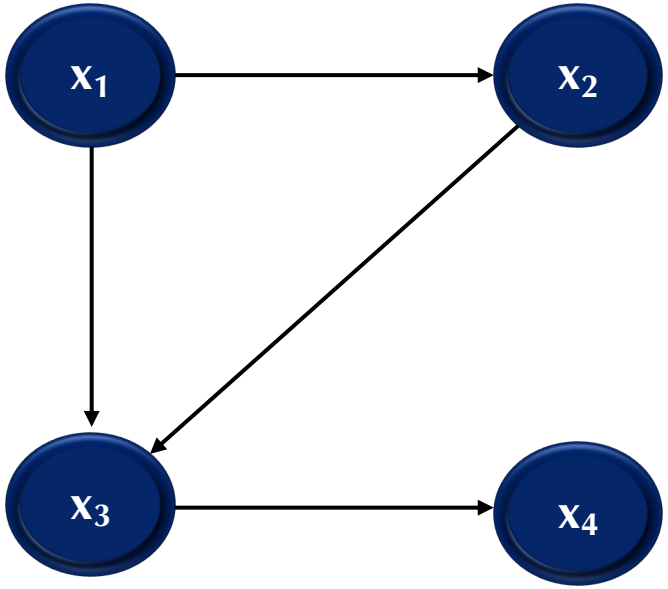
Statistical side of the story (no causality)

Modeling joint distribution $P(x_1, \ldots, x_n) = P(x_1) \Pi_i P(x_i | x_{i-1}, \ldots, x_n)$

$P(x_1, x_2, x_3, x_4) = P(x_1) P(x_2|x_1) P(x_3|x_2,x_1) P(x_4|x_3,x_2,x_1)$

If we model for dependence for x_4 : 2^{n-1} parameters

Bayesian Networks

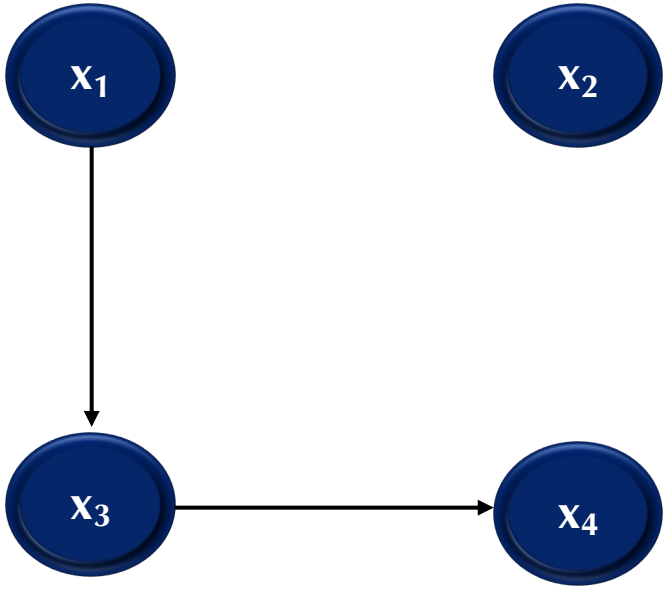


Local Markov Assumption:

Given its parents in the DAG, a node X is independent of all its non-descendants

$$P(x_1, x_2, x_3, x_4) = P(x_1)\, P(x_2|x_1)\, P(x_3|x_2,x_1)\, P(x_4|x_3)$$

Bayesian Networks



Local Markov Assumption:

Given its parents in the DAG, a node X is independent of all its non-descendants

$$P(x_1, x_2, x_3, x_4) = P(x_1)\, P(x_2)\, P(x_3|x_1)\, P(x_4|x_3)$$

OAK RIDGE National Laboratory

Bayesian Networks

Modeling joint distribution $P(x_1, \ldots, x_n) = P(x_1) \prod_i P(x_i | pa_i)$

Local Markov Assumption ➔ Bayesian network factorization

Bayesian network factorization ➔ Local Markov Assumption

Bayesian Networks

Minimality assumption

Local Markov Assumption:

1. Given its parents in the DAG, a node X is independent of all its non-descendants

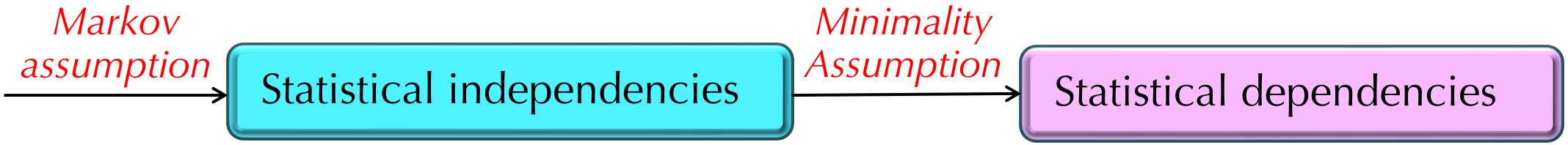
2. Adjacent nodes in the DAG are dependent

$X \rightarrow Y$ Permits distributions where $P(x,y) = P(x)\ P(y|x)$

OAK RIDGE National Laboratory

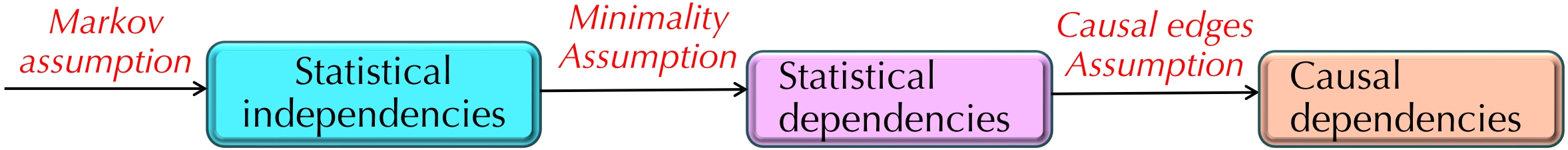
Where is the causal graph ?

So far what we have seen with BNs, they are purely statistical!



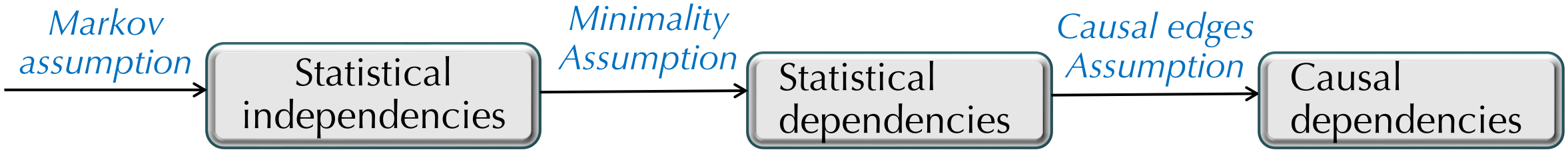
A variable X is a cause of Y if Y can change its response to changes in X

In DAG, every parent is a direct cause of all its children

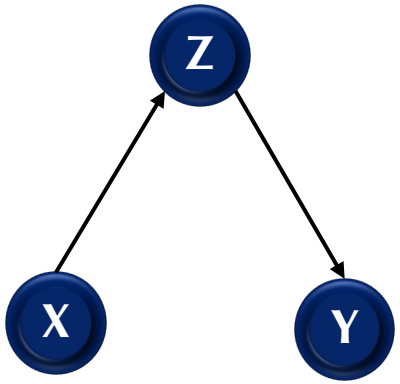


Causal graph

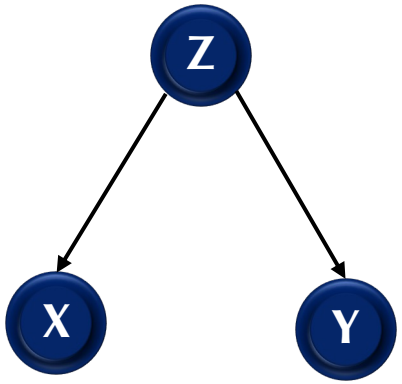
DAG + Local Markov & Causal edges assumptions



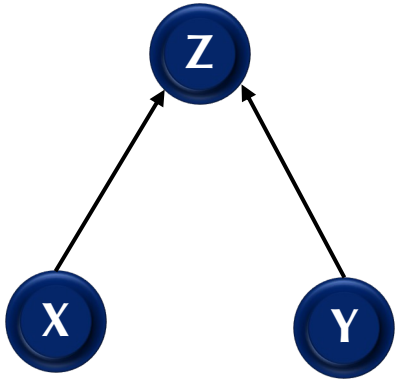
Primary graphical structures



Chain

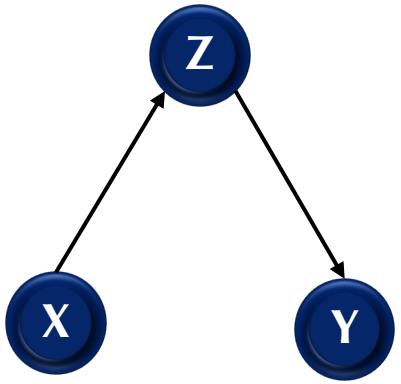


Fork



Collider

Primary graphical structures

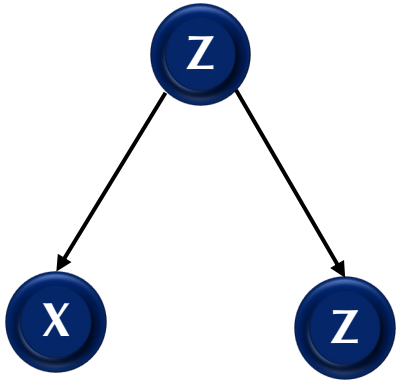


Chain

$X \perp\!\!\!\perp Y \mid Z$

- X changes Z which in turn changes Y
- This is also known as mediation
- If we condition on Z, we block transmission of causal information

Primary graphical structures



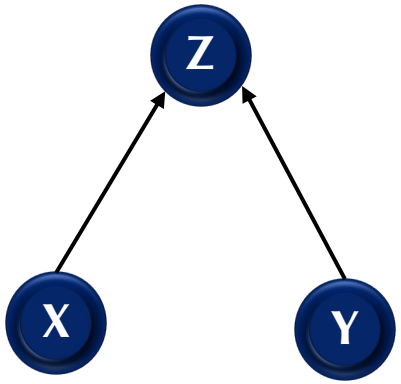
Fork

$X \perp\!\!\!\perp Y \mid Z$

- Z causes X and Y, which makes X and Y statistically dependent
- This transmits non-causal information
- If we intervene on X, it doesn't change Y
- Confounding or common cause variables, backdoor path
- Condition on Z to block non-causal information

OAK RIDGE National Laboratory

Primary graphical structures



Collider

$X \not\perp\!\!\!\perp Y \mid Z$

- X and Y are not related but cause Z
- Common effect and do not transmit causal information
- Conditioning on Z introduces spurious association between X and Y aka collider bias
- Extra: There shouldn't be any colliders if we want an unblocked path → d-separation criteria

OAK RIDGE National Laboratory

Causal Graphs and Structural Causal Models

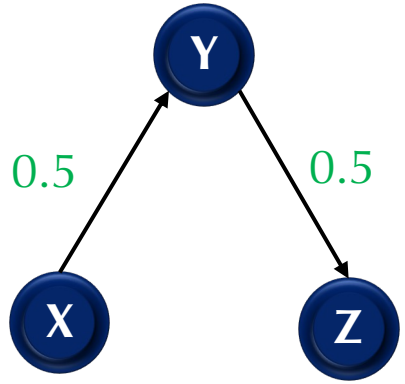
- With DAGs we visualize the beliefs on causal relationships

- SCMs are set of equations that delineate causal connections between variables, often affected by independent noise terms N (usually omitted from the graphical representation)

$$X := \epsilon_X$$

$$Y := 0.5Y + \epsilon_Y$$

$$Z := 0.5X + 0.5Y + \epsilon_Z$$

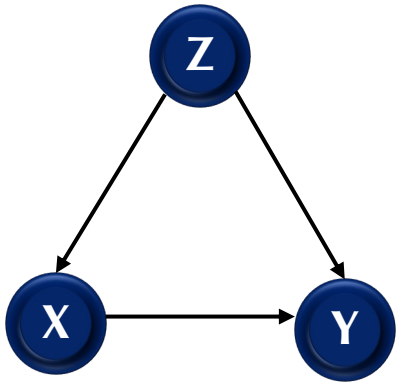


Causal Graphs and Structural Causal Models

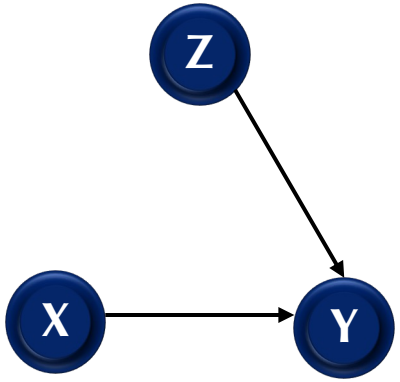
- For estimating causal effects, we introduce the do operator; do (X=x)

Average causal effect = E[Y | *do*(X=1)] - E[Y | *do*(X=0)]

- When we intervene on X, we cut-off all incoming variables edges

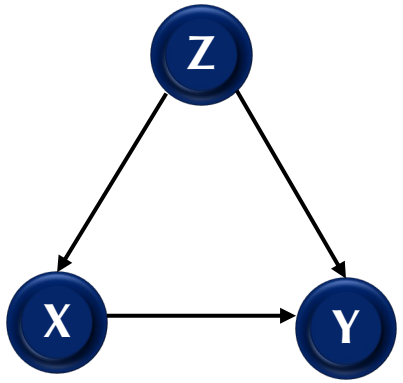


Observed



Intervened

Going back to Aspirin/Headache example



Average causal effect = E[Y | *do*(X=1)] - E[Y | *do*(X=0)]

Naïve estimate = E[Y | X=1] - E[Y | X=0]

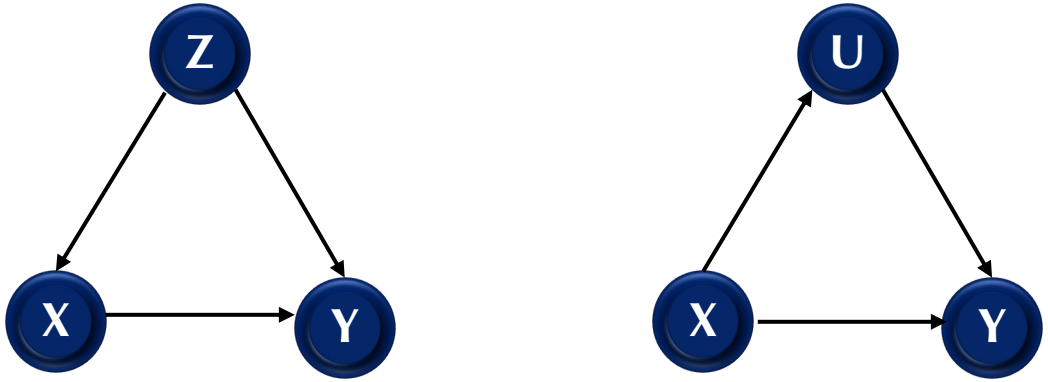
We know that observation is not equal to intervention

Correct estimate = E[Y | X=1, Z] - E[Y | X=0, Z)]

OAK RIDGE National Laboratory

Simpsons Paradox

DAGs are nice to identify the causal structure, we still need to think about which variables to condition on to provide correct estimates



Same statistical dependencies for both DAGs, hence they're observationally equivalent but these imply different interventional effects

OAK RIDGE National Laboratory

All together: Strategies for estimation

- Specify causal target
- Find DAG including observed and unobserved variables
- Find which variables to condition on (valid adjustment sets)
- Can we block all backdoor paths
- Estimate the true causal effect by conditioning on chosen variables

OAK RIDGE National Laboratory

Wrapping up I: Potential Outcomes versus SCMs for Causal Inference

Potential Outcomes

- Define target
- Emulate RCT with focus on effect of X on Y
- Understand covariates
- Utilize pre-treatment variables
- Focus is on estimation methods such as propensity score matching etc.

Structural Causal Models

- Define target
- Understand causal relations between multiple variables
- Address mediation, direct and indirect effects
- Emphasis is on identification
- Estimate is via conditioning on selected variables

Causal learning is much more difficult than doing statistical learning!

Since there would be some assumptions always, refutability checks are recommended to further determine the validity and robustness of the causal relations.

Let's do a hands-on exercise

https://drive.google.com/file/d/1tToNdY-6lqxGRctKfx12a5fZKQiQ0re8/view?usp=sharing

Causal Discovery

- Given some understanding about the system/problem of interest, can we estimate a causal effect, i.e., identify and quantify causal relationships between variables ?

- Yes, causal inference with potential outcomes, SCMs

- But what if we don't know the causal graph ?

- Causal discovery: Data → causal graph

Different Strategies

- Independence-based causal discovery
  - Find all conditional independent relations from the data
  - Construct DAG with only these independencies following d-separation rules
  - Assumptions:
    - Sufficiency: No observed common causes
    - No selection bias (conditioning on unobserved colliders are not allowed)
    - Faithfulness: Statistical dependencies are primarily due to causal relations rather than spurious relations

Different Strategies

- Semi-parametric causal discovery
  - Remember that SCMs are in principle non-parametric
  - Apply a few more assumptions
    - Linear model with additive Non-Gaussian Noise
    - Non-linear model with additive Gaussian Noise

Wrapping up II

Independence-based causal discovery

- Constraint-based, score-based methods
- From estimated conditional and marginal independencies
- Might lead to set of DAGs

Semi-parametric causal discovery

- Linear/non-linear Non-Gaussian/Gaussian
- Some prior knowledge on the variables are required
- Unique identification of DAG

Causal learning is much more difficult than doing statistical learning!

Since there would be some assumptions always, refutability checks are recommended to further determine the validity and robustness of the causal relations.

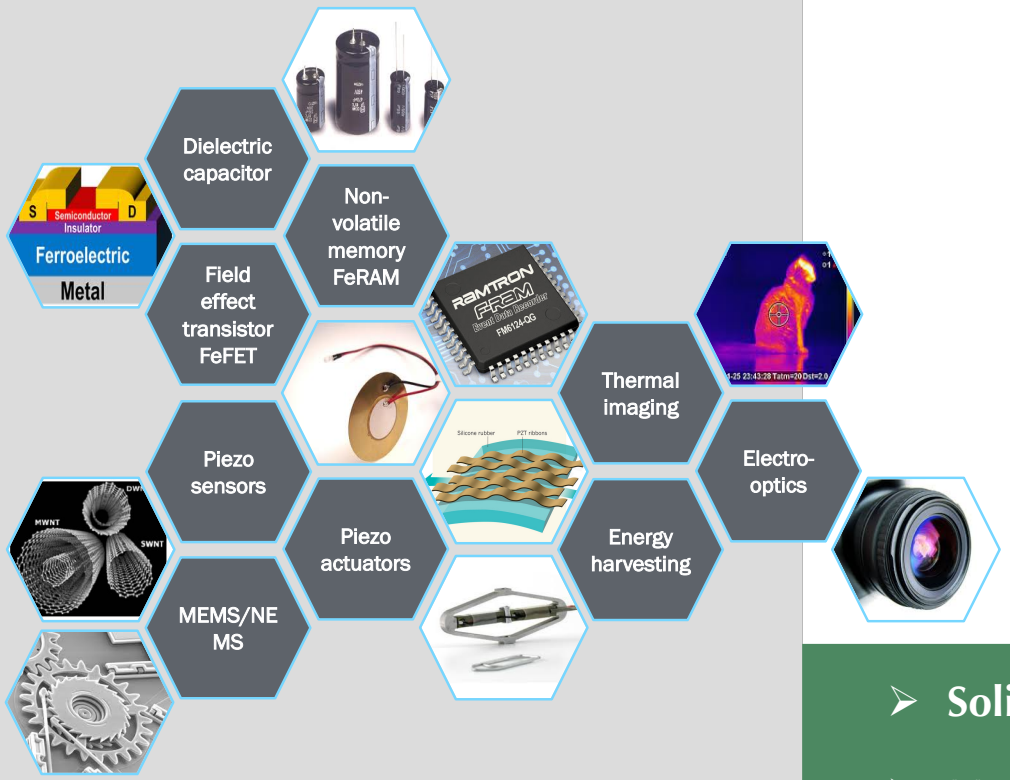
Diagnostic Evaluation:

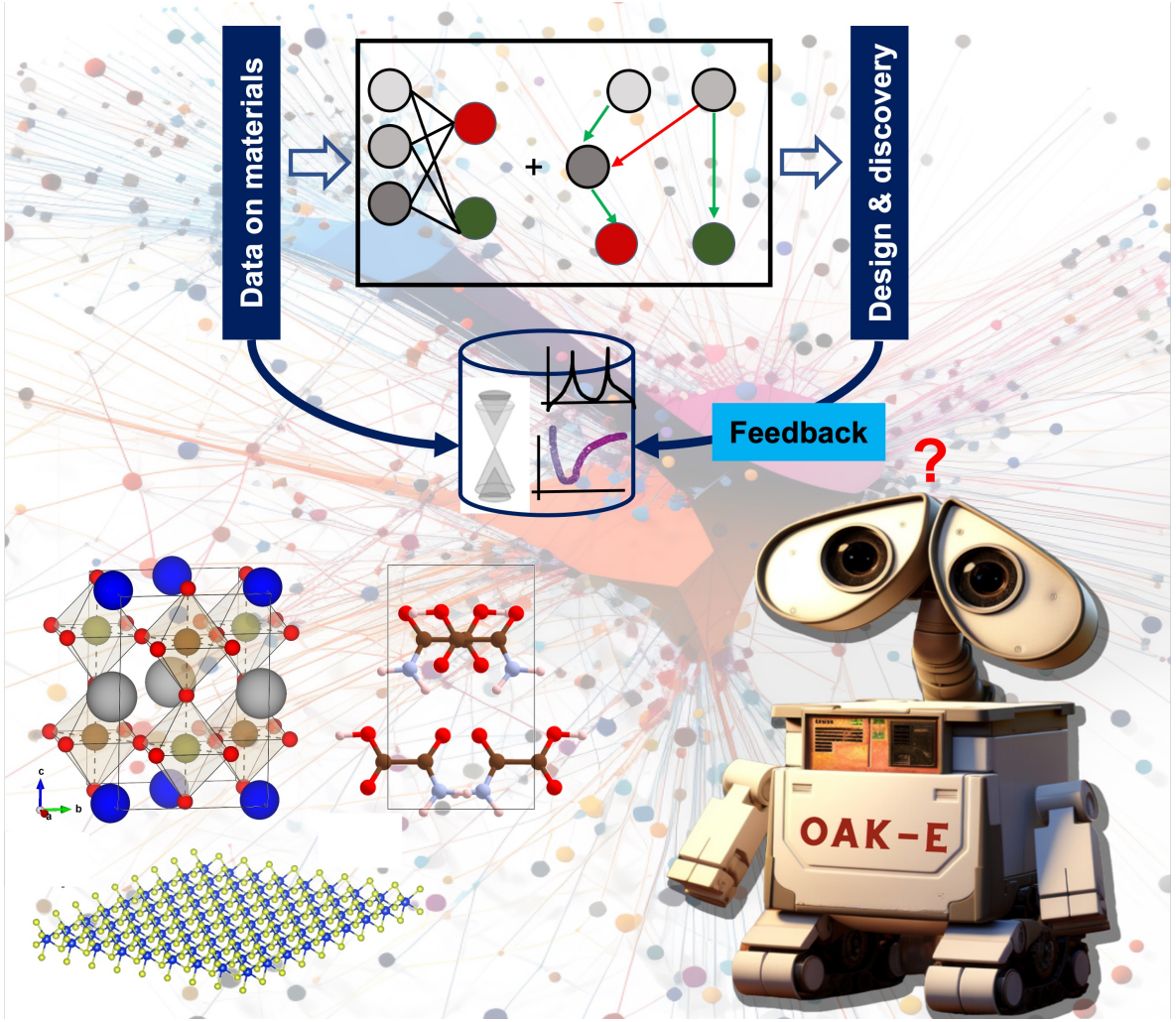
Let's put down our thoughts on causality

- Have you heard about this before and in which context ?
- Any related topic that might have sparked some interest ? Maybe a popular phrase starting with "correlation does not imply…"
- Is this even important ?
- Is it just statistics ?

OAK RIDGE National Laboratory

Causal models for Materials Science Applications





- Solid state materials for microelectronics, catalysis, photovoltaics etc.
- Molecules for drug discovery, electronics etc.
- Data from simulations & experiments over different length and time scales

Let's do a hands-on exercise

https://colab.research.google.com/drive/1sK6V5eTa12IVH46IL6BcwZgUIonpbT1f?usp=sharing

References

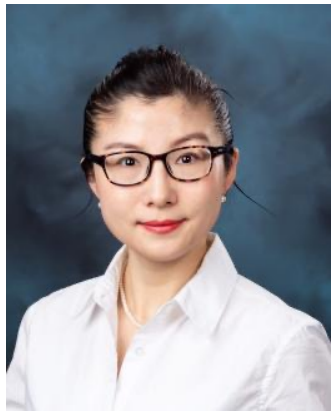
- Hern´an, M. A. (2018). The C-word: scientific euphemisms do not improve causal inference from observational data. American journal of public health, 108(5), 616-619.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. Advances in Methods and Practices in Psychological Science, 1(1), 27-42.
- Schafer, J. L., Kang, J. (2008). Average causal effects from nonrandomized studies: a practical guide and simulated example. Psychological methods, 13(4), 279.
- Hern´an MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman Hall/CRC. Free copy: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf
- Pearl, J., Glymour, M. & Jewell, N.P. (2016) Causal Inference in Statistics: A Primer
- Spirtes, Peter, Clark N. Glymour, Richard Scheines, and David Heckerman. Causation, prediction, and search. MIT press, 2000.
- Shimizu, S., Hoyer, P. O., Hyv¨arinen, A., Kerminen, A., & Jordan, M. (2006). A linear non-Gaussian acyclic model for causal discovery. Journal of Machine Learning Research, 7(10).
- Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Sch¨olkopf, B. (2016). Distinguishing cause from effect using observational data: methods and benchmarks. The Journal of Machine Learning Research, 17(1), 1103-1204.
- Peters, J., B¨uhlmann, P., & Meinshausen, N. (2016). Causal inference by using invariant prediction: identification and confidence intervals. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 947-1012.
- Peters, J., Janzing, D., & Sch¨olkopf, B. (2017). Elements of causal inference: foundations and learning algorithms. The MIT Press.

Acknowledgments

LDRD AI Initiative

Dr. Ayana Ghosh
(Advanced Computing Methods for Physical Sciences Section, CSED, CCSD)

Dr. Zachary Fox
(Advanced Computing in Health Sciences Section, CSED, CCSD)



Dr. Dan Lu
(Advanced Computing Methods for Physical Sciences Section, CSED, CCSD)

Dr. Prasanna Balaprakash
(CCSD)