Causal Discovery

Introduction to Control and Machine Learning

Summer Semester 2024


Apurwa Agrawal
(23161645)


Rahul Ramraje
(23081510)


Neelima Kawatra
(22931138)


Sachin
(23276228)


Swarnendu Saha
(22929020)

Objective

- **What?** Introduction to Causal AI
- **Oh no!** Assumptions in Causal Discovery
- **Why?** Causal Discovery Methods
- **How?** Calculus of Causality
- **Where?** Applications
- **Alas!** Limitations
- **Wow!** Future work

Motivation: Answering what?

(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

- Figure: Recognition algorithms generalize poorly to new environments.
- Top five labels and confidence produced by ClarifAI.com shown.
- Interpolation is great, but we need to extrapolate.

Introduction to our hero: Causality!

Causality is central notion in science, decision-taking and daily life.

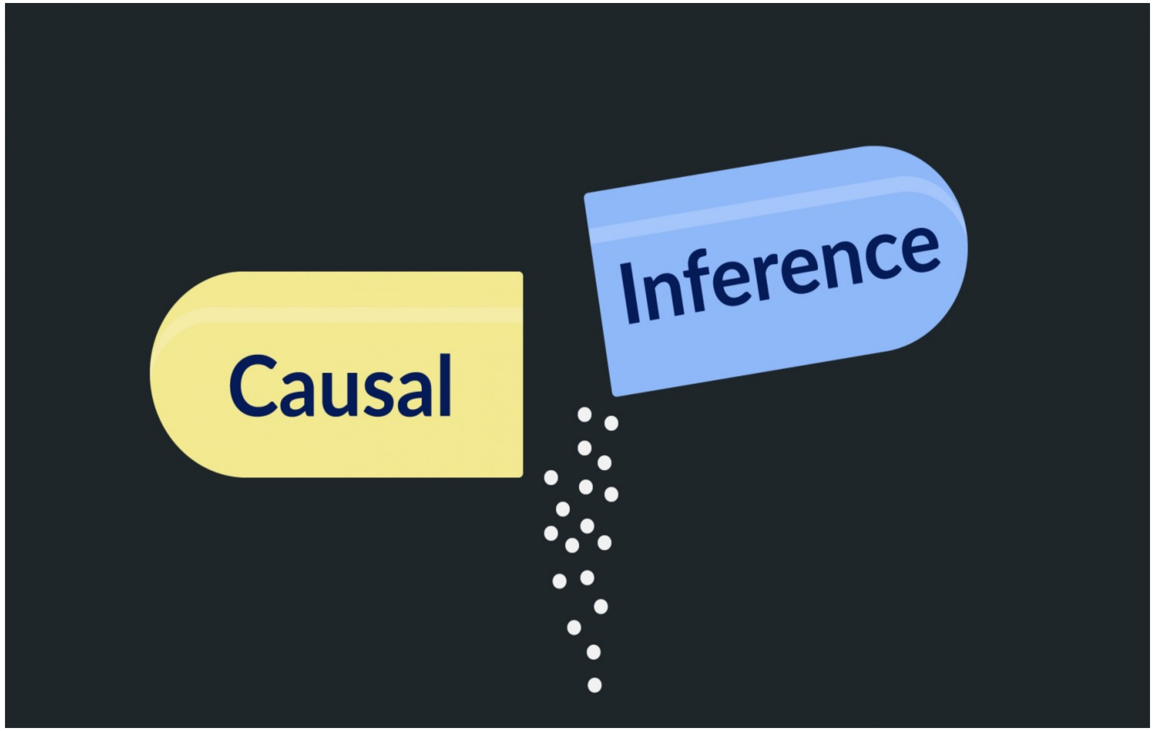
How do you define cause and effect?

- **Cause and Effect:**
  - Cause: An event or factor that brings about a change.
  - Effect: The change that results from a cause.
  - Example: A medication (cause) leads to improved health outcomes (effect).

Judea Pearl : “A is the cause of B if B ‘listens to’ A”

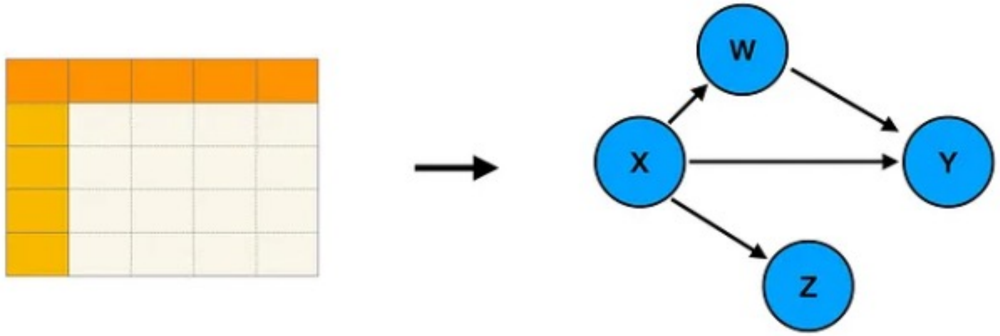
Causal Inference

- Causal Inference refers to the process of determining the effect of one variable on another. It aims to establish a cause-and-effect relationship using statistical methods, often from observational or experimental data.
- It helps in determining whether and how a change in one variable influences another.



Causal Discovery

- Causal discovery aims to uncover the underlying causal structure from the data.
- It involves identifying which variables are causes and which are effects without necessarily having prior knowledge of the relationships.
- It involves using algorithms and statistical methods to identify potential causal relationships.



Big picture goal of causal discovery: translate data into a causal model. Image by author.

Assumptions!

Faithfulness Assumption

Markov assumption: $X \perp\!\!\!\perp_{G} Y \mid Z \Longrightarrow X \perp\!\!\!\perp_{P} Y \mid Z$

Faithfulness Assumption

Markov assumption: $X \perp\!\!\!\perp_{\mathrm{G}} Y \mid Z \Longrightarrow X \perp\!\!\!\perp_{\mathrm{P}} Y \mid Z$

Causal graph $\longrightarrow$ Data

Causal graph $\longleftarrow$ Data

Faithfulness Assumption

Markov assumption: $X \perp\!\!\!\perp_G Y \mid Z \Rightarrow X \perp\!\!\!\perp_P Y \mid Z$

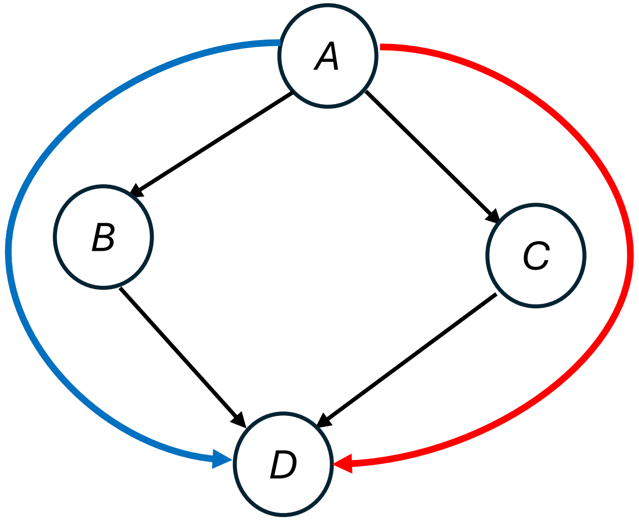
Causal graph $\longrightarrow$ Data

Causal graph $\longleftarrow$ Data

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Leftarrow X \perp\!\!\!\perp_P \mathrm{Y} \mid \mathrm{Z}$

Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Leftarrow X \perp\!\!\!\perp_P Y \mid Z$



Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Longleftarrow X \perp\!\!\!\perp_P Y \mid Z$

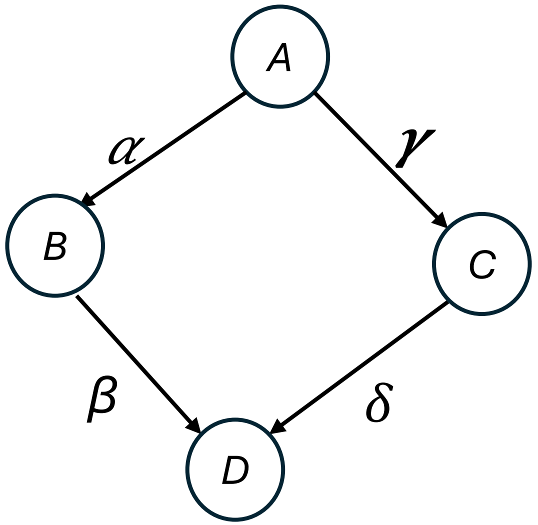
B := αA

C := γA

D := $\beta B + \delta C$

$A \perp\!\!\!\perp D$

but A and D aren't d-separated

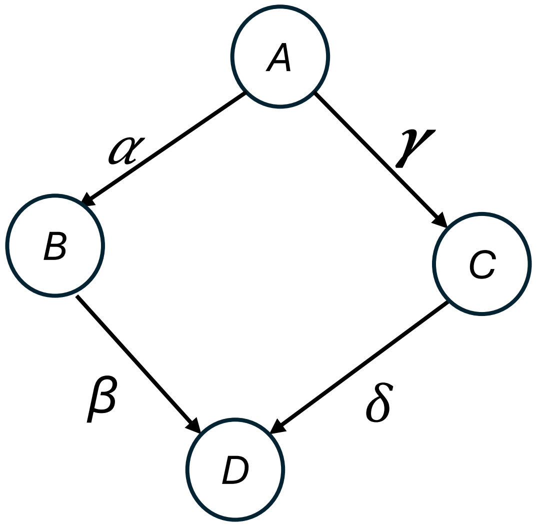


Violation of Faithfulness

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \Leftarrow X \perp\!\!\!\perp_P Y \mid Z$

$A \perp\!\!\!\perp D$

but A and D aren't d-separated



$B := \alpha A$

$C := \gamma A$

$D := \beta B + \delta C$

$D = (\delta \beta + \gamma \delta) A$

Paths cancel if $\delta \beta = - \gamma \delta$

Causal Sufficiency and Acyclicity

Causal Sufficiency:

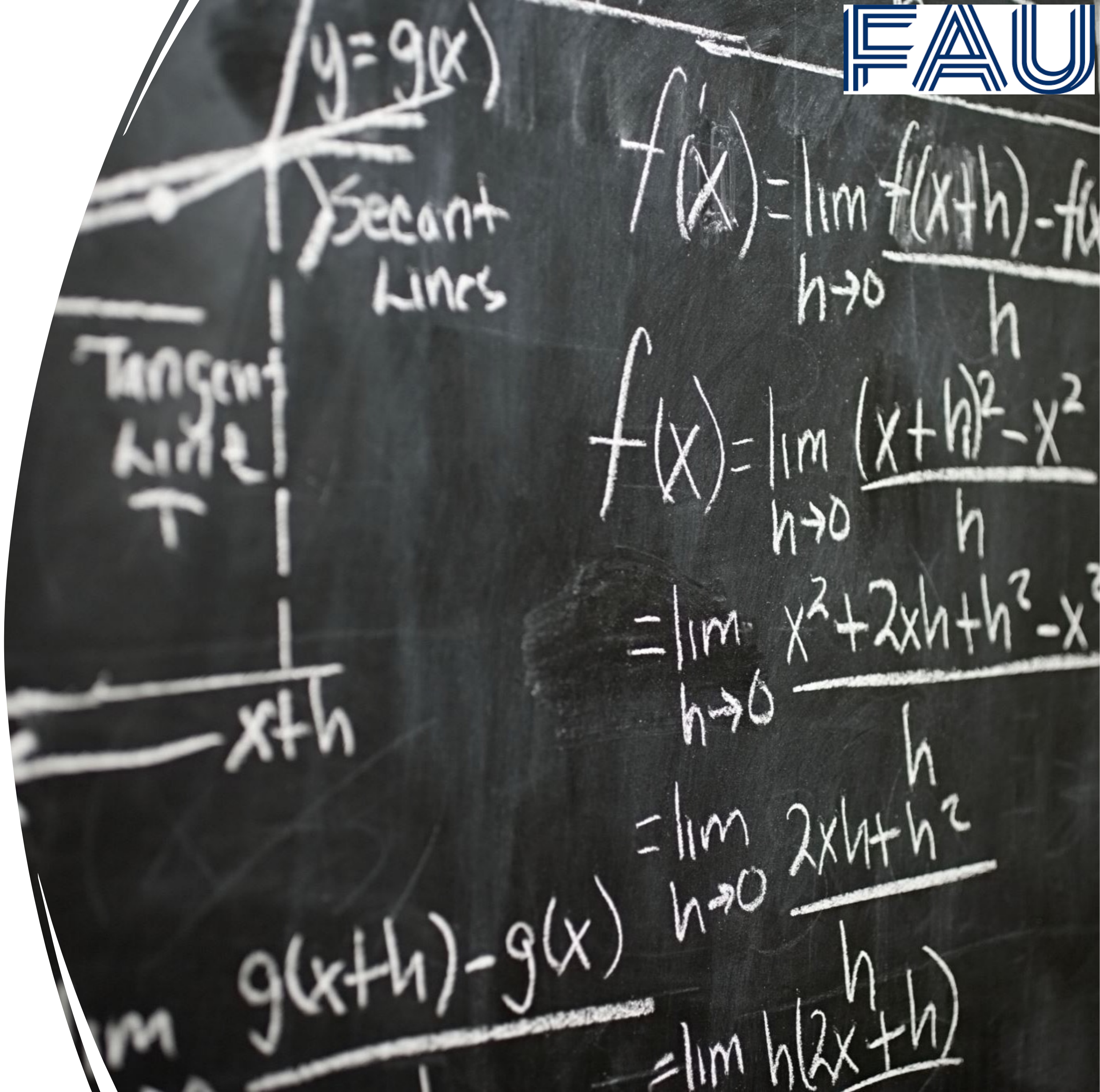
there are no unobserved confounders of any of the variables in the graph.

Acyclicity:

still assuming there are no cycles in the graph.

Causal discovery methods

- Constraint-based
- Score-based



Constraint-based method

CONSTRAINT-BASED METHODS AIM TO CONSTRUCT A CAUSAL GRAPH BY EXPLOITING THE CONDITIONAL INDEPENDENCIES OBSERVED IN THE DATA.

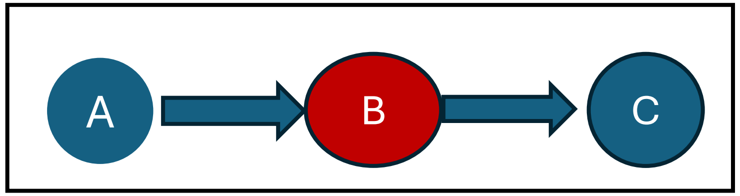
ASSUME FAITHFULNESS, WHERE DATA INDEPENDENCE RELATIONSHIPS MATCH THE GRAPH.

PERFORM STATISTICAL TESTS TO IDENTIFY CIS AND STRUCTURE THE CAUSAL GRAPH.

PC ALGORITHM

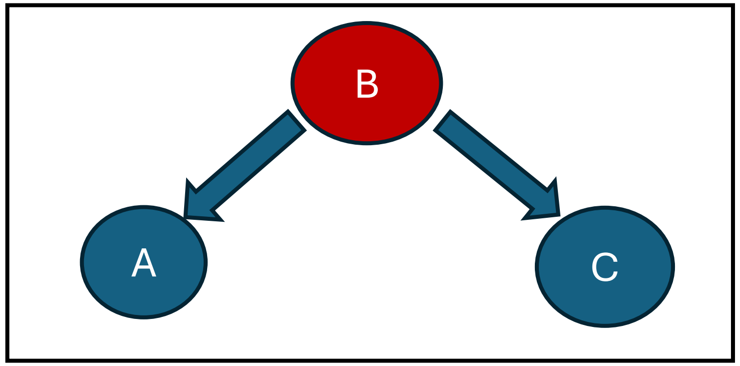
D-separation

$A \perp\!\!\!\perp C \mid B$



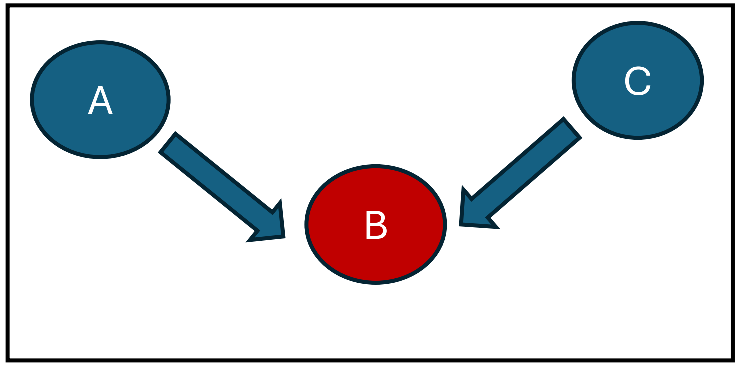
Chain

$A \perp\!\!\!\perp C \mid B$



Fork

$A \not\perp\!\!\!\perp C \mid B$



Blocked

Score based method

- Greedy Equivalent Search (GES) Algorithm

- Score-based algorithms aim to find the best causal graph by maximizing a fitness measure *S(G,D)* across possible graphs *G*.

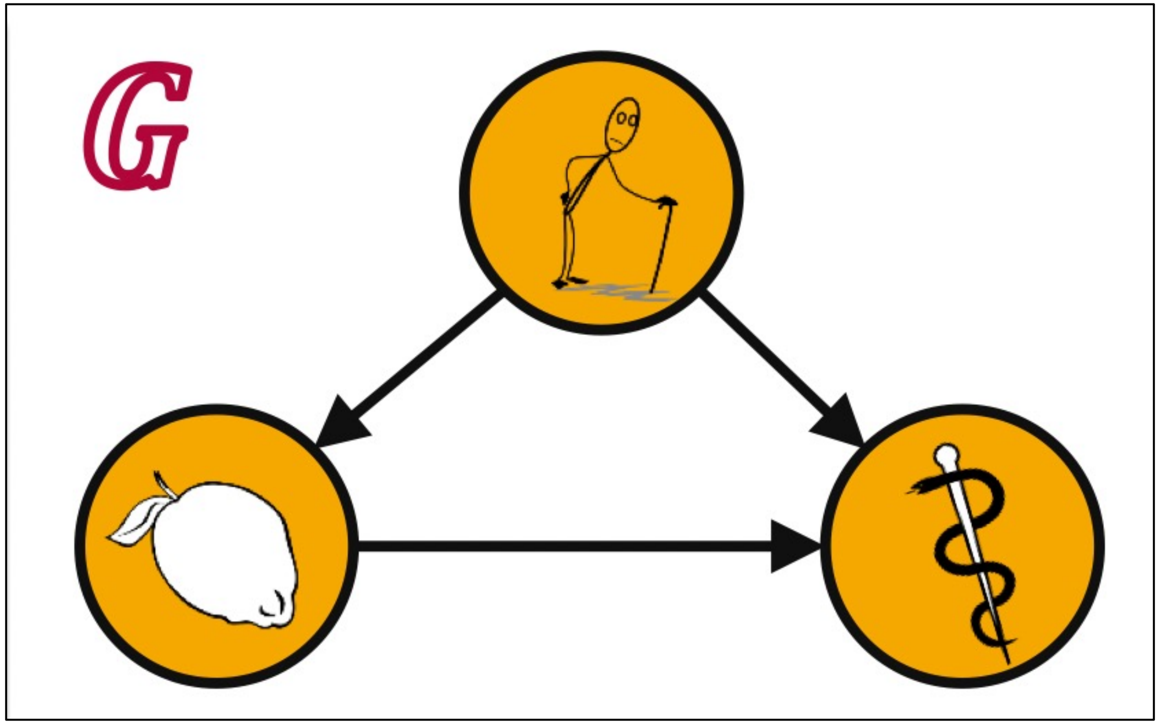
- This measure evaluates how well a graph explains the relationships observed in the data.

$$G = \underset{G \in \mathcal{G}}{\operatorname{argmax}}\, S(G, D)$$

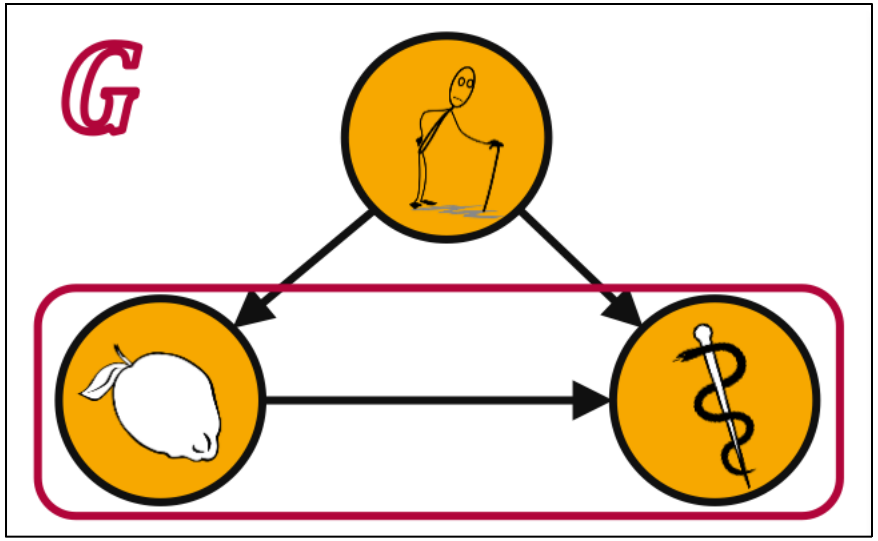
The Calculus of Causality!!

The Calculus of causality

- Causality – not well defined ?
  - Probability theory has an associational, and not a causal nature
- Assume that the data is generated by model G
  - The recovery of scurvy is casually influenced by the treatment with lemons
  - But now both the recovery of scurvy as well as the treatment with lemons are causally influenced by the age of the sailors
- The question remains:
  - Should we treat scurvy with Lemons?



The Calculus of causality



- We run an experiment w.r.t. the model G, i.e., we favor old sailors for treatment with lemons
- The observed data for all sailors:

Combined	Recovery	No Recovery	Total	Recovery Rate
No Lemons	20	20	40	50 %
Lemons	16	24	40	40 %
Total	36	44	80	

- Hence, we see that

P(*recovery* | *lemons*) < ***P***(*recovery* | *no lemons*)

The Calculus of causality

The observed data for old sailors:

Combined	Recovery	No Recovery	Total	Recovery Rate
No Lemons	2	8	10	20 %
Lemons	9	21	30	30 %
Total	11	29	40	

P*(recovery | lemons,old)* > ***P****(recovery | no lemons, old)*

The observed data for young sailors:

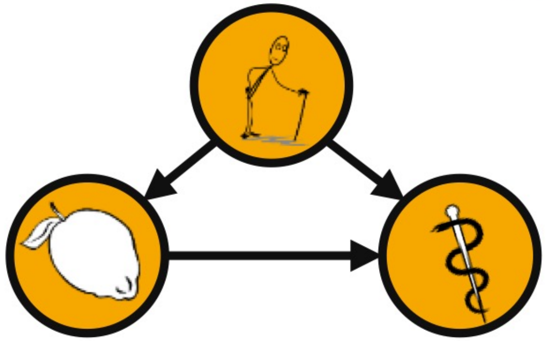
Combined	Recovery	No Recovery	Total	Recovery Rate
No Lemons	18	12	30	60 %
Lemons	7	3	10	70 %
Total	25	15	40	

P*(recovery|lemons, young)* > ***P****(recovery | no lemons, young)*

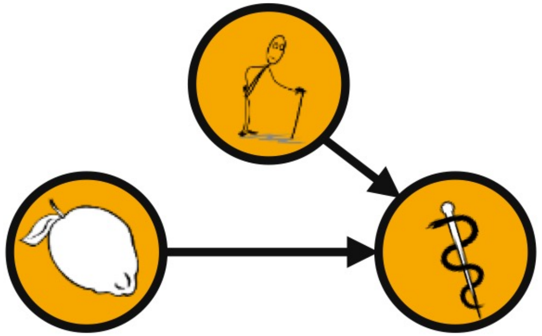
The Calculus of causality

- Simpson's Paradox : Reversal association between two variables after considering the third variable .
- In an interventional regime, all influences stemming from "natural causes" of the exposure variable are removed.
- New operator – **Pearl's do-operator**

Observational Regime



Interventional Regime



The Calculus of causality

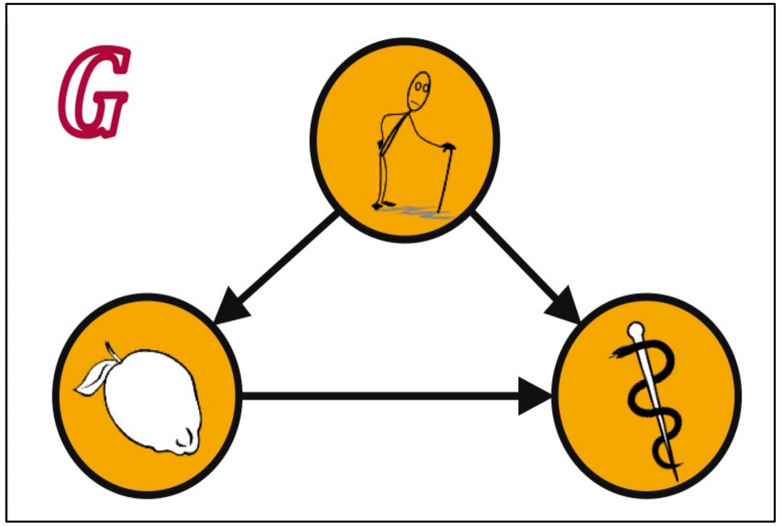
- Resolution of the Simpson's paradox
  - We should treat scurvy with lemons if

$$\boxed{\boldsymbol{P}(\textit{recovery}|\textit{do}(\textit{lemons})) > \boldsymbol{P}(\textit{recovery}|\textit{do}(\textit{no lemons}))}$$

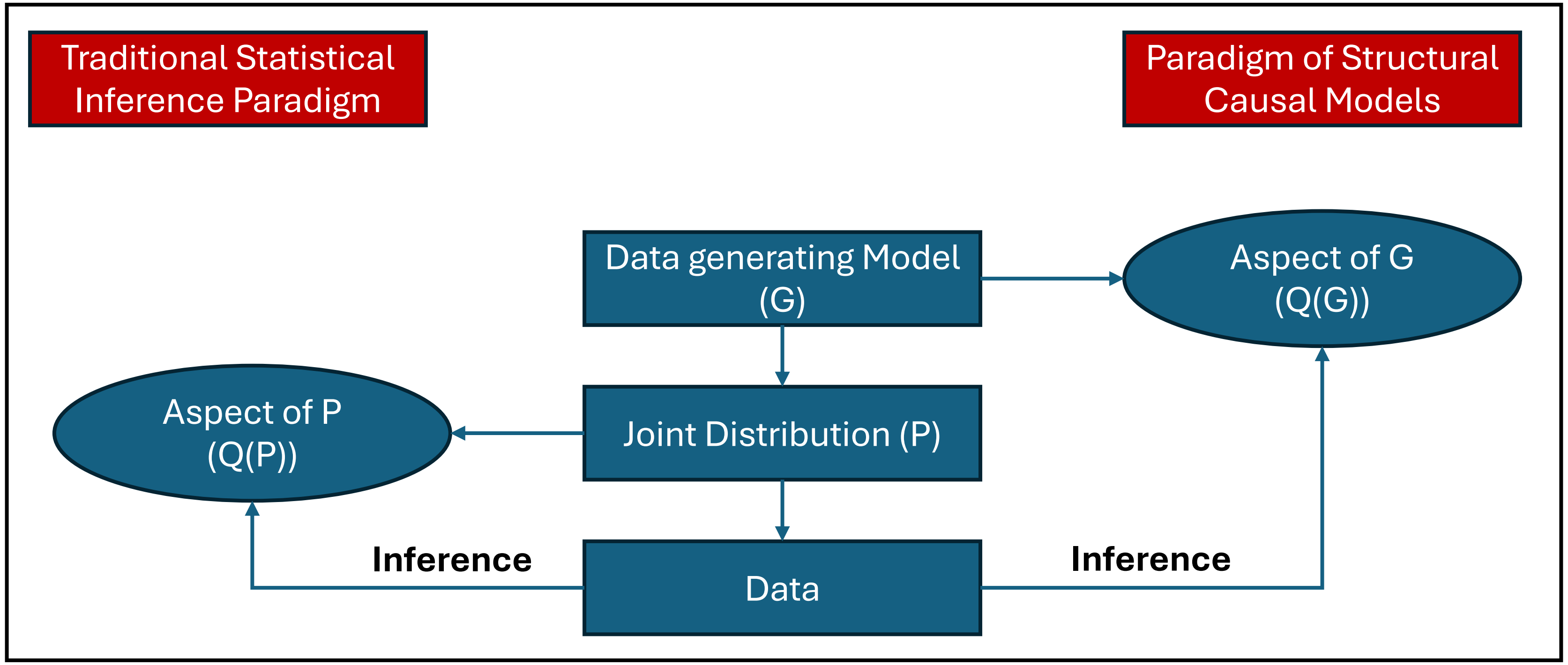
- Derivation of do-operator
  - If identifiable,

    P(.|do(.)) can be calculated from G and observational Data
  - In our example, we have

$$\boxed{\begin{aligned}&P(\text{recovery}|\text{do}(\text{lemons})) = \sum P(\text{age})\, P(\text{recovery}|\text{age},\text{lemons}) = 0.5\\ &P(\text{recovery}|\text{do}(\text{no lemons})) = \sum P(\text{age})\, P(\text{recovery}|\text{age},\text{no lemons}) = 0.4\end{aligned}}$$



The Concept



What is the sailors' probability of recovery when **we see** a treatment with lemons?

***Q(P)* = *P*(*recovery* | *lemons*)**

What is the sailors' probability of recovery if **we do** treat them with lemons?

***Q(P)* = *P*(*recovery* | do(*lemons*))**

Applications

- **Healthcare**
  - Personalized treatment plans, drug discovery.
- **Finance**
  - Enhanced investment strategies, financial decision-making.
- **Marketing**
  - Customer behavior analysis, marketing strategy optimization.

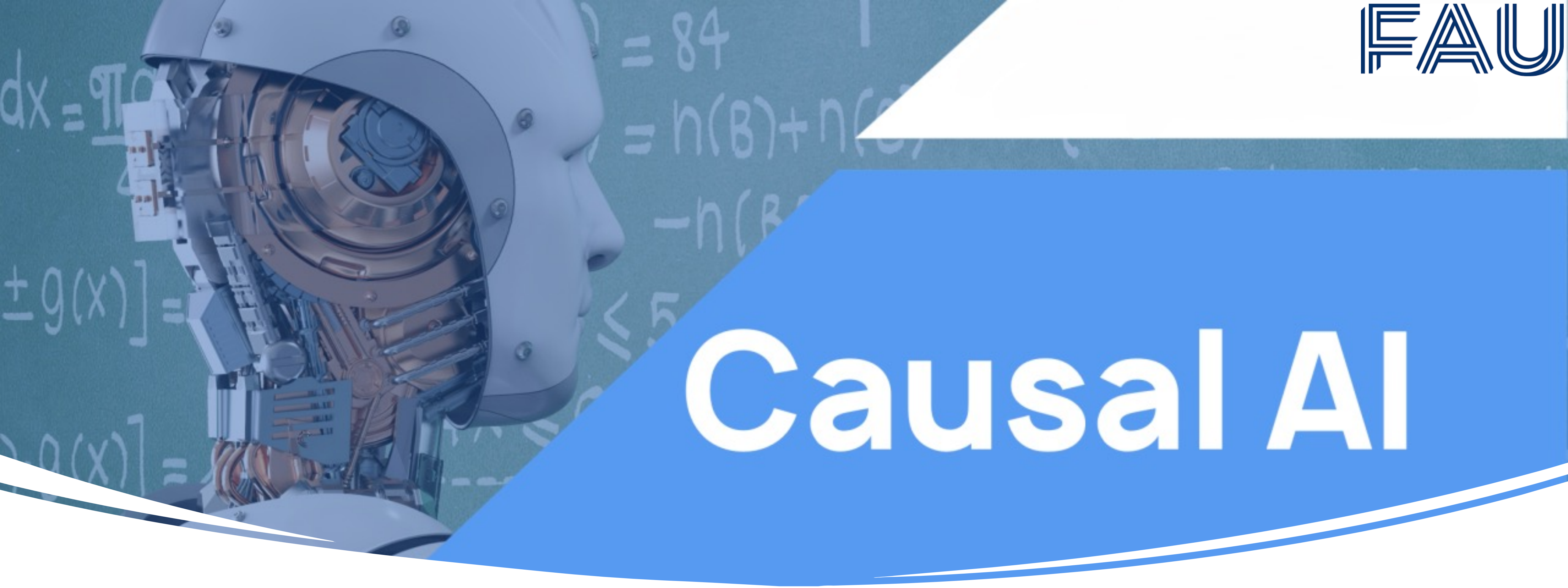
Limitations and Future Works

Limitations:

- **Data Dependency:** Requires high-quality, unbiased data.
- **Assumption Sensitivity:** Incorrect assumptions can lead to wrong conclusions.
- **Limited Generalizability:** Models may not work well in different contexts.

Future Solutions:

- **Enhanced Data Collection and Cleaning:** Advanced tools and techniques to ensure unbiased, high-quality data.
- **Dynamic Assumption Testing:** Continuous validation and updating of assumptions with real-time data.
- **Efficient Algorithms:** Development of more efficient algorithms.
- **Transfer Learning and Diverse Training Data:** Use transfer learning and diverse datasets to improve model adaptability.



Conclusion

- **Improves decision accuracy** by understanding true cause-and-effect relationships.
- **Enhances outcome across industries** like healthcare, finance and, more.

References

- Pearl, J., Glymour, M., & Jewell, N. P. (2016). **Causal Inference in Statistics: A Primer. John Wiley & Sons.** https://doi.org/10.1002/9781119186847

- Plattner, H. (2017). **Causal Inference: Theory and Applications.** https://hpi.de/fileadmin/user_upload/fachgebiete/plattner/teaching/CausalInference/04_17_Causal_Inference-Theory_and_Applications.pdf

- **CMPT 727 Spring 2023. Mathematical, Statistical, and Probabilistic Foundations of Machine Learning.** https://www.youtube.com/playlist?list=PL_5SuHtr8fsrK9NqWWSL4YL8urMAHLsvU

- Peters, J., Janzing, D., & Schölkopf, B. (2017). **Elements of Causal Inference: Foundations and Learning Algorithms**. MIT Press. https://doi.org/10.7551/mitpress/11444.001.0001

THANK YOU

We are open for questions now!