An Introduction to Causal Inference

UW Markovia Reading Group
presented by
Michele Banko & Kevin Duh
Nov. 9, 2004

Outline

1. D-separation
2. Causal Graphs
3. Causal Markov Condition
4. Inference
5. Conclusion

Motivation

- How can we connect causal relations and probabilistic independence?
- For any directed graph, how can we compute all and only those independence and conditional independence relations that hold for all values of the parameters?

D-separation (Pearl, 1988; Spirtes, 1994)

D-Separation

- If variables *X* and *Y* are **d-separated** relative to a set of variables *Z* in a DAG *G*, then *X* and *Y* are independent conditional on *Z* in all probability distributions G can represent.
- Knowledge about *X* gives you no extra information about *Y* once you have knowledge of *Z*.
- "D" stands for "dependence" (or "directed")

Formal Definition

- If G is a DAG in which X, Y and Z are disjoint sets of vertices, then X and Y are **d-connected** by Z in G if and only if there exists an undirected path U between some vertex in X and some vertex in Y such that for every collider C on U, either C or a descendent of C is in Z, and no non-collider on U is in Z.
- A vertex *v* is a **collider** if two arrowheads meet at *v*, e.g. → *V* ←, ↔ *V* ←
- X and Y are **d-separated** by Z in G if and only if they are *not d-connected* by Z in G.

Examples

Are X and Y d-separated given Z?

X → Z → Y
There are no colliders on the path from X to Y
Z is a non-collider on the path from X to Y, but it is in the conditioning set
X and Y are NOT d-connected by Z
So they are d-separated by Z
Independence? $X \perp Y \mid Z$

X → Z ← Y
Z is a collider on the path from X to Y
There are no non-colliders
X and Y are d-connected by Z
So they are NOT d-separated given Z
Independence? $X \perp Y$

Causal Graphs

- Causal Graphs: DAGs that are interpreted causally

  (Y)ellowed Fingers ← (S)moking → Lung (C)ancer

- Tell us for any manipulation, what other variables we would expect to change

Causal Graphs

- Are incomplete
  - Do not necessarily included all of the causes of each variable present
  - May leave out variables that might come between a cause and its effect
- Are complete
  - All common causes of the variables are included.
  - All causal relations among the variables are included

"When DAGs are interpreted causally, the Markov condition + d-separation are in fact the *correct* connection between causal structure and probabilistic independence"

This assumption is called the Causal Markov Assumption

The Causal Markov Condition

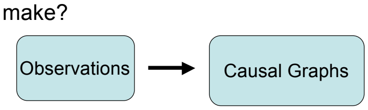
- Intuition: Ignoring its effects, all relevant probabilistic information about a variable is contained in its direct causes.
- Definition: For a variable *X* and any set of variables *Y* that does not include the effects of *X*, *X* is independent of *Y* conditional on its direct causes.

  (Y)ellowed Fingers → (S)moking → Lung (C)ancer

- Independence relations obtained:
  - Y: Y is independent of C conditional on S
  - S: Y and C are effects of S..
  - C: C is independent of Y conditional on S
- Same as when we apply d-separation

Inference

- Given statistical data, what causal statements can we make?



- This is often done in scientific research (esp. social science and epidemiology)
- Can we formalize the inference process and identify the assumptions that make causal inference valid?

Example: Big Shoe-size causes High IQ?

- To identify whether big shoe size causes high IQ, researchers conducted a controlled experiment measuring the shoe-size and IQ of 60 children from age 6 to 15.

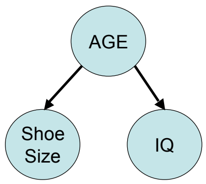
	Low IQ	High IQ
Big shoe	2	22
Small shoe	32	4

- Conclusion: Data suggests that big shoe => high IQ and small shoe => low IQ.

Source: http://www.cmu.edu/CSR/

WRONG!

- Correlation does not imply causation.
- Example assumes there are no other common causes in the experiment, e.g:



Assumptions are Important!

- Correlation and statistical independence do not always imply causation. However, under some assumptions, it can!
- More causal assumptions => more constrained set of causal graphs => stronger inferential purchase
- Depending on the case, some assumptions may be justified, some may not.

Some Assumptions

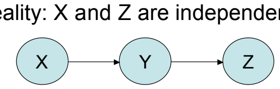
- Causally Markov
  - Any population produced by a causal graph will have the independence relations obtained by applying d-separation to it.
  - i.e. causal graph => a set of independence statements
- Reliability of Statistical Inference about Independence
  - Data is statistically ideal, so any probabilistic independence claim we make from data is correct.
- Faithfulness
- Causal Sufficiency

Faithfulness Assumption: Various Definitions

- If X and Y are not correlated, they really are independent.
- Faithfulness vs. Causally Markov
  - Causal Markov only produces a set of independence relations from a causal graph, but says nothing about whether there are additional independence relations.
  - Faithfulness says: the set of independence relations derived from Causal Markov is the exact set of independence relations.
- Graph structure vs. Distribution
  - Faithfulness says: Independence arises only from graph structure, not some “lucky” distribution

“Lucky” Distribution

- In reality: X and Z are independent given Y

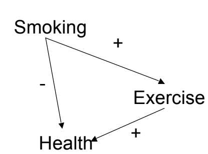


- Unluckily, we have a “lucky” distribution:
  - P(X=1) = 0.5
  - P(Y=1|X=1) = 1   P(Y=0|X=0) = 1
  - P(Z=1|Y=1) =1   P(Z=0|Y=0) = 1
  - Knowing Z automatically tells us the value of X, so conditioning on Y makes no difference. X and Z appear not independent.

Source: Glymour & Cooper, ***Computation, Causation, and Discovery***

Another “Lucky” Distribution

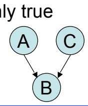
- Causal Markov condition gives no independence statement for this graph:



- But some distributions might make “Smoking” seem independent of “Health” if the positive effect from Smoking via Exercise cancels out the negative effect
  - Population is “unfaithful” to the causal graph that generated it

What do we gain from assuming Faithfulness?

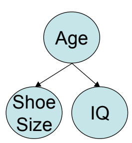
- Causal Markov condition only gives us a set of independence relations, but more may exist. Faithfulness says this set is the exact set.
- Example: A,B,C. From data, there is some connection between A and B, and B and C.
  - Markov condition only gives us A _||_ C
  - Says nothing about A _||_ B , C _||_ B , A _||_ B | C, etc.
  - Faithfulness assumes A _||_ C is the only true independence, and the rest are false.



Causal Sufficiency

- Causal sufficiency assumes that common causes of all variables are **measured**.
  - Vs. Completeness, which assumes all common causes of all variables are modeled.
- In practical inference situations, not all variables can be measured.
  - If the unmeasured variable is a common cause, relationships among its effects are harder to find.
  - If causal sufficiency is assumed, the set of possible graphs can be restricted.

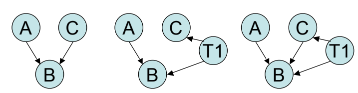
Example: No Causal Sufficiency



- Complete: Age is the only common cause for ShoeSize and IQ.
- Shoe_||_IQ|Age
- But, suppose Age is unmeasured.
  - Data will only include independence statements not conditioned on Age, which is nothing!

What to do when not assuming Causal Sufficiency

- The set of causal graph we infer from independence relations must include all graphs that have common causes we did not measure.
- E.g. Suppose we measure A,B,C and observed A_||_C. Assuming Causal Markov and Faithfulness but not Sufficiency, we have have several causal graphs, e.g.:



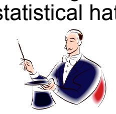
Etc.

Review of Assumptions

- By the following assumptions, we can restrict the set of causal graphs when inferring causality from independence in statistical data:
  - Causal Markov: the independence relations implied by causal graph holds in the population
  - Faithfulness: the independence relations implied by causal graph are the only ones that hold in the population.
  - Causal sufficiency: the set of measured variables include all common causes in the causal graph

Conclusion

- Is causal inference “magically pulling causal rabbits out of a statistical hat?”



- No -- it investigates what rabbits can and cannot be pulled out, depending on particular assumptions.
- Take-home lesson:
  - When inferring causality from statistical data, we must think clearly about whether our assumptions are valid before jumping to conclusions.

References

- Glymour, C., Scheines, R., Spirtes, P., and Kelly, K. (1987). *Discovering causal structure*.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan and Kaufman, San Mateo
- Spirtes, P. (1994). *Conditional Independence in Directed Cyclic Graphical Models for Feedback*. Technical Report CMU-PHIL-54, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA.
- Scheines, R. (1997), *An Introduction to Causal Inference*, in Causality in Crisis, ed. by Steven Turner and Vaughan McKim, University of Notre Dame Press.
- Glymour, C., Cooper, G. (1999), *Computation, Causation, and Discovery*, AAAI Press.
- CMU Curriculum on Causal and Statistical Reasoning. http://www.cmu.edu/CSR