HANDOUT – March 2, 2023

SCHEINES: CAUSATION

PHILOSOPHICAL FOUNDATIONS OF ML

Conny Knieling

INTRODUCTION AND OVERVIEW

I. Introduction

Event causation/actual causation/token causation:

- Particular or single events
- Transitive, anti-symmetric, irreflexive
- Used in: Legal cases, accident investigation

Type causation/causal generalization/causation among variables:

- Kinds of events
- Usually transitive, a-symmetric, not irreflexive
- Used for: policy makers, statisticians, social scientists

II. Overview

- Socratic philosophical theories (e.g. Causal Analysis)
- Euclidean philosophical theories (e.g. DAGs)

"More metaphysical mysterious than causation itself, or were circular, or were simply unable to account for the asymmetry of causation or separate spurious from real causation" (p. 2)

Axiomatic and epistemological turn in work on causality since 1985

THE AGE OF CAUSAL ANALYSIS (1970s-early 1980s)

I. The Counterfactual Theory

David Lewis: Possible Worlds, Similarity Metric

Negative: Misses cases of overdetermination or pre-emption, can't handle asymmetry nor spurious causation

II. Regularity Account

John Mackie: necessary and sufficient conditions, INUS condition

Negative: can't handle asymmetry nor spurious causation

III. Probabilistic Causality

Patrick Suppes: reduce causality to probability, associated and independent events, and: time and conditional (in)dependence to handle asymmetry and spurious causation

Handling spurious causation by: looking for other events Z that screen off C and E

Negative: replaces it with (philos.) mystery of probability, time being used this explicitly, we need more than probabilities (Cartwright)

IV. Physical Process Theory

Wes Salmon: connection to explanation, causal interaction, causal processes and exchange of invariant quantity

Negative: pseudo-processes, time is used explicitly, can't handle spurious causation

V. Manipulability Theories

Intervention on/manipulation of the cause

Handles asymmetry very well

Negative: circular, intervention not always possible

THE AXIOMATIC AND EPISTEMOLOGICAL TURN (1985-2004)

Alternative emerged: rich axiomatic theory, uses manipulation in a Woodward way, connects it to probability

Interdisciplinary and fundamentally epistemological

I. Path Analysis

Sewall Wright: causal graphs, path analytic model (statistical model and path diagram)

Problem: can't handle asymmetry

II. Philosophy

David Papineau/Dan Hausman: causal connection; bring asymmetry into the picture, association in virtue of C, independent conditional on C, associated conditional on E

Two quantities are associated if they are causally connected, and independent if they are not

III. Statistics and Comp Sci

Graphical models, directed and undirected graphs

Judea Pearl: directed acyclic graphs (DAGs), Bayes networks, indistinguishability

IV. Philosophy, again

SGS: axiomatized relationship between causation and probabilistic independence, connect it to Pearl and statistics, empirical indistinguishability

"What can and cannot be learned about causal structure from statistics concerning independence and association" (p. 12)

Causal Markow Axiom: Every variable is probabilistically independent of all of its non-effects (direct or indirect), conditional on its immediate causes

Problem: Causal Markow Axiom is debated (common causes might not screen off their effects)

Reply: conditioning on all common causes instead

Problem: other controversial assumptions, other philosophical problems

Reply: Woodward's work, independent manipulation vs. "fat hand"

CONCLUSION

No reductive analysis of causation has emerged

Axiomatic approach has emerged and seems more successful, especially for statistical and scientific practice

Recent interdisciplinary theory of causation:

- Takes direct causation as a primitive
- Defines intervention from direct causation
- Connects causal systems to probabilities and statistical evidence through axioms

Causal Diagrams for Empirical Research

By Judea Pearl

Bing:

- Here is a summary of Judea Pearl's paper "Causal diagrams for empirical research" based on the search results:
- The paper shows how graphical models can be used as a mathematical language for integrating statistical and subject-matter information. The paper introduces the notion of a causal diagram, which is a directed acyclic graph (DAG) that encodes causal assumptions about a set of variables. The paper also defines a do-operator, which represents an intervention or manipulation on a variable. The paper demonstrates how causal diagrams can be used to answer questions about causation, confounding, identification, and estimation. The paper also discusses some applications of causal diagrams to empirical research in various field

My Summary:

- Empirical research usually works in the setting where they have observed variables X, unobserved variables Z and target variables Y
- The goal is often to figure out what "caused" Y to take on its value or what is the effect of X on Y
- These variables can interact in different ways this paper formalises and gives a language to express these relations between variables as a causal diagram graph. Using this graph we can judge whether we can use our observed values to gain a good estimate of the causatory relationship between variables.
- Philosophy takeaway: We can express causal relationships by referring to the outcome of interventions in a scenario (All else held equal, change one thing)

Causal Graphs are Directed Acyclic Graphs where:

- Every variable of concern is a node
- An edge from node A to node B exists if and only if we consider A to have a direct causative effect on B
    - Causation: autonomous physical mechanisms among the corresponding quantities

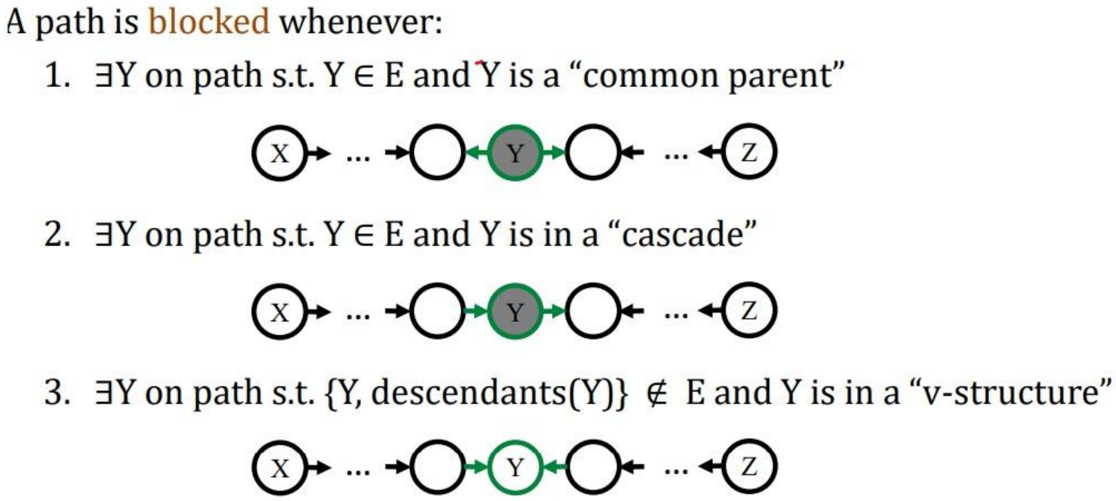
This means in a causal diagram node X is fully determined by the values of its parents (with some independent noise).

$$X = f(pa(X), \epsilon)$$

This means that the graph represents conditional independence assumptions between variables: X is independent of all its predecessors given its parents.

D-Separation is the criteria for more general conditional independence:

- Two sets of nodes are conditionally independent if the observations block all paths between them
- A blocked path is given by the following three scenarios



A path is blocked whenever:

1. $\exists Y$ on path s.t. $Y \in E$ and Y is a “common parent”

2. $\exists Y$ on path s.t. $Y \in E$ and Y is in a “cascade”

3. $\exists Y$ on path s.t. $\{Y, \text{descendants}(Y)\} \notin E$ and Y is in a “v-structure”

Img Credits: 10-708

With this we can perform interventions:

- Atomic set: Force an X to take a specific value while holding all other causal mechanisms acting on X equal (disconnects X from its causal parents) and propagate the effects of the new value of X through all its descendants.
- We use this to define causal effect: the causal effect of X on Y is a function:

$$P(y|x):\ X \times Y \rightarrow [0,\ 1]$$

  Which is the probability that Y takes the values y if X is set to x

Computing this function is called identifying the causal effect of X on Y.

Given the structure of a causal graph in which all values are observed, we can infer post intervention distributions from pre-intervention distributions - exactly what we want. This means the causal effect of X on Y is identifiable

What if not everything is observed?

Back Door: variable set S satisfies backdoor criteria with respect to sets X and Y if for any x, y in X and Y

1. No node in S is a descendant of x **and**
2. S blocks every undirected path between x and y that contains an arrow into x

Intuition: Information cannot enter through the “back door”

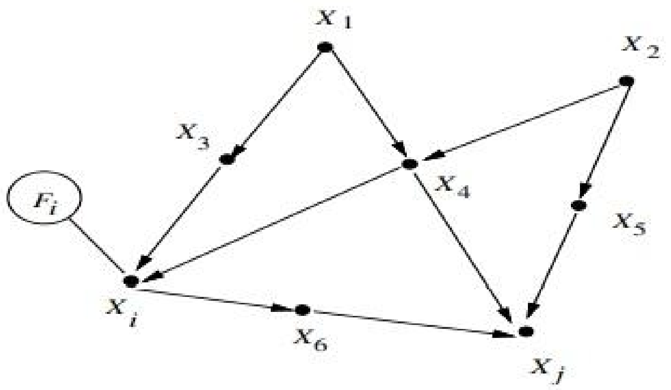


Figure 3:
A DAG representing the back-door criterion; adjusting for variables $\{X_3, X_4\}$ *(or* $\{X_4, X_5\}$*) yields an unbiased estimate of* $P(x_j|\hat{x}_i)$.

If variables S satisfy the back door conditions with respect to X and Y, then the causal effect of X on Y is identifiable given S is observed.

Front Door: set Z satisfies front door criteria with respect to X and Y if

- Z intercepts all directed paths from X to Y
- There is no back door path from X to Z
- All back door paths from Z to Y are blocked by Z

If the front door criteria is met then the causal effect of X on Y is identifiable given Z

The rest of the paper explains how you can establish a calculus of intervention: conditions under which certain interventions do not change the distribution of P(Y|X, Z) (changing Z which are in some way irrelevant etc),

This allows practitioners to judge whether or not their variable effects will be identifiable

Discussion:

- Relies heavily on causal assumptions in graph
- There are issues when extending this to cyclical graphs

Actual Causality and the Art of Modeling - Joseph Y. Halpern and Christopher Hitchcock

February 2023

- They focus on Judea Pearl's work on **actual causation**, in which **structural equations** are used to represent causal relationship between variables
  - Philosophical literature on causality focuses on actual causality, but Pearl placed this in a much wider context by formalizing notions around causality
  - They also add to actual causation the idea of normality
- Structural equations express the idea of *interventions*, e.g. what happens to a patient upon taking X drug, what happens to a bottle when it gets struck...
- Actual causation can be subjective, because it depends on the variables that we choose to examine in the structural equations
  - E.g. what's the cause of a traffic incident?
    - Traffic engineer: bad road design
    - Educator: poor drivers' education
    - Sociologist: the pub where the driver got drunk
    - Psychologist: the driver's recent breakup with his girlfriend
    - ...
- Informal defn: A is a cause of B if B counterfactually depends on A under some contingency
  - For instance, if Suzy and Billy throw rocks at a bottle, and Suzy's rock hits the bottle first, breaking the bottle, we can say that Suzy's throw is the cause of the bottle breaking because this counterfactually depends on Suzy's throw under the contingency that Billy doesn't throw
  - If not modelled properly, we can also say that Billy's throw is the cause of the rock breaking...

Causal Models

- A **causal model** $\mathcal{M}$ is composed of a signature $\mathcal{S}$ and a set of modifiable structural equations $\mathcal{F}$
  - A **signature** explicitly lists endogenous variables $\mathcal{V}$ (variables internal to the model), exogenous variables $\mathcal{U}$ (inputs external to the model), and a function $\mathcal{R}$ that associates possible values for each variable
  - A function $F_X \in \mathcal{F}$ specifies the value of $X \in \mathcal{V}$ depending on the other variables, e.g. $F_X(u, y, z) = u + y$
- E.g. For a forest fire, we could have FF standing for forest fire, where $FF = 1$ is a forest fire and $FF = 0$ is no forest fire, similarly L stands for lightning and ML stands for match lit. Then $F_{FF} = \max(L, ML)$ means that the forest fire happens if either lightning strikes or a match is lit
- Visualizing the equations, they only consider DAGs, i.e. there can't be any cycles in the equation dependencies
- (There's a lot of notation introduced that I'm not going to type but I'm just going to explain the actual causation definition directly since that's the important thing)
- (Defn) **Actual Causation**: the setting of variables $\mathbf{X} = \mathbf{x}$ is an actual cause of event ϕ in the causal model M with context $\mathbf{u}$, $(M, \mathbf{u})$, if the three conditions hold:
  1. $(M, \mathbf{u}) \models (\mathbf{X} = \mathbf{x})$ and $(M, \mathbf{u}) \models \phi$

     This just says that ϕ has to actually happen in this context, and $\mathbf{X} = \mathbf{x}$ is actually true in this situation
  2. Let's partition the set of endogenous variables into two subsets $\mathbf{Z}$ and $\mathbf{W}$ with $\mathbf{X} \subseteq \mathbf{Z}$. We can think of the variables in $\mathbf{Z}$ as the ones on the causal path from X to ϕ, in other words, if we change the values of some variables in $\mathbf{X}$, this will result in some other variables in $\mathbf{Z}$ changing in a chain until finally ϕ is true. Let $\mathbf{x}^{'}$ be an alternate setting of variables in $\mathbf{X}$ and $\mathbf{w}$ be some setting of variables in $\mathbf{W}$.

     The variables $\mathbf{W}$ are the ones that are not in the causal chain, but may still have an indirect effect on what happens. Then the two subconditions should hold:

     (a) $(M, \mathbf{u}) \models [\mathbf{X} \leftarrow \mathbf{x}^{'}, \mathbf{W} \leftarrow \mathbf{w}]\neg\phi$

     This means that if $\mathbf{X}$ had a different value, then ϕ would be false. However, for this effect to play out, we can also use a different setting of the variables in W. for instance, in the context where both lightning strikes and someone drops a match, we can say that lightning is still a cause of the fire, since in the counterfactual world where there was no lightning ($x^{'}$), and no one dropped a match (w), then the fire would have been prevented.

(b) $(M, \mathbf{u}) \models [\mathbf{X} \leftarrow \mathbf{x}, \mathbf{W}^{'} \leftarrow \mathbf{w}, \mathbf{Z}^{'} \leftarrow \mathbf{z}^{*}]\phi$ for all subsets $\mathbf{W}^{'}$ of $\mathbf{W}$ and all subsets $\mathbf{Z}^{'}$ of $\mathbf{Z}$

This basically means that when condition $\mathbf{x}$ is met, no matter what other factors we change that are not on the causal path, ϕ will still be true. For instance, if we drop the match, and lightning doesn't strike, the forest will still be on fire.

3. $\mathbf{X}$ is minimal

   This means that no smaller subset of events in X is enough to trigger ϕ, e.g. in the forest fire example, if we had an irrelevant variable "sneezing", sneezing + lightning is not a cause of forest fire since the lightning is already enough by itself

- Philosophical issues: it seems like it's straightforward to reason about this after the equations have been set up and the variables are chosen. However, this has to be done so carefully that it seems practically not worth it to go through this analysis. You can also change the conclusion by modelling things in different ways by choosing different variables to include, or changing the structure of the equations

Normality

- The definition of actual causation still doesn't seem to capture important intuitions about causation. For instance, if an assassin plans to murder someone by putting poison in his coffee, but changes his mind at the last minute, and a bodyguard suspects the plot and puts antidote in the coffee, by the definition above this would mean the bodyguard was the cause of the man surviving, but this seems wrong since there isn't actually poison in the coffee.

- They try to patch this counterexample by introducing a normality ranking function that ranks possible worlds, and adding to part 2a the fact that the contingencies to be considered have to be ranked as "more or equally normal" to the context world

  - For instance, people don't usually put poison in coffee, so we shouldn't consider the situation in which someone poisons the coffee when reasoning about counterfactuals

Discussion/general comments

I didn't actually read Judea Pearl's and Halpern's work before so I liked reading the formalization of an intuitive notion and reading the examples. However, the actual design considerations for the model seem to be very open, and different people may have different ideas about what variables to include and the relation of these variables. I think there is no clear way to come to a consensus on this since we don't know everything about how various factors interact in the real world and it may be too complex to work with even if we did know (e.g. a butterfly flapping its wings eventually causing a hurricane).

So once we have the DAG, we can reason more clearly about a situation, but most of the work is done as background. It may also be unclear what would happen in counterfactual cases outside the clear examples presented, making it impossible to apply the definition. I think the normality condition seems tacked-on to address the counterexample and suffers from similar problems in that what is considered normal differs for everyone, and I don't think a normality ranking function exists.