Title: Causal inference yesterday, today and tomorrow (PI-IVADO-IC Special Webinar)

Speakers: Ilya Shpitser

Subject: Other

Date: September 13, 2024 - 9:30 AM

URL: https://pirsa.org/24090157

Abstract:

As part of a monthly webinar series jointly hosted by Perimeter, IVADO, and Institut Courtois, Ilya Shpitser will present an introduction to causal inference and its applications to problems in physics and computer science. This seminar will be fully on zoom and members of all three institutes are welcome.

Abstract: In this talk I will give some history of ideas of causal inference, describe the causal inference workflow, including formalizing the cause-effect question in terms of a parameter, defining (or learning) the causal model, checking if the data has information about the desired parameter via identification theory, and efficiently estimating the parameter if it is identified. I will briefly touch on connections of causal inference to other areas, discuss what machine learning and causal inference can teach each other, and describe some open problems.

Zoom TBC

Causal Inference Yesterday, Today, and Tomorrow

Ilya Shpitser

Perimeter Institute-IVADO-Institut Courtois Partnership
Webinar Series

September 13, 2024

Two Quotes on Causality from the 1740s

We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second, . . . where, if the first object had not been the second never had existed.

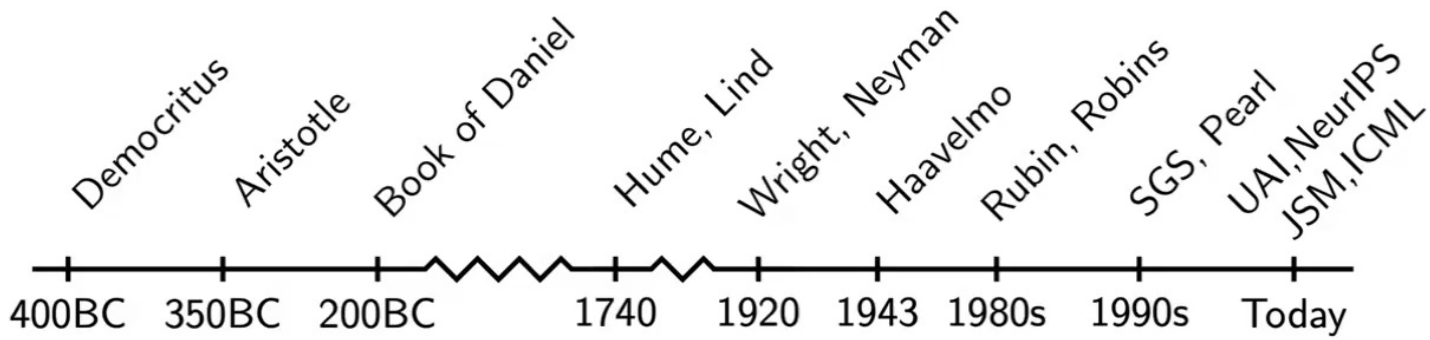
David Hume (1748)

Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of the knees. They lay together in one place, being a proper apartment for the sick in the forehold...

James Lind (1747)

Timeline

- "I would rather learn one causal law than be King of Persia."
- Aristotle's four causes (material, formal, efficient, final).
- Book of Daniel: earliest recorded mention of a comparison trial.
- Hume's definition, Lind's scurvy trial.
- Wright's path analysis, Neyman's potential outcomes.
- Haavelmo's structural equations.
- Modern methods: Rubin,Robins,Pearl,Spirtes,Glymour,Schenes.
- Today: tutorials at UAI, NeurIPS, ICML. Conference/journal papers. Causal inference a part of FDA best practices guidelines.



2/18

Fundamental Ideas

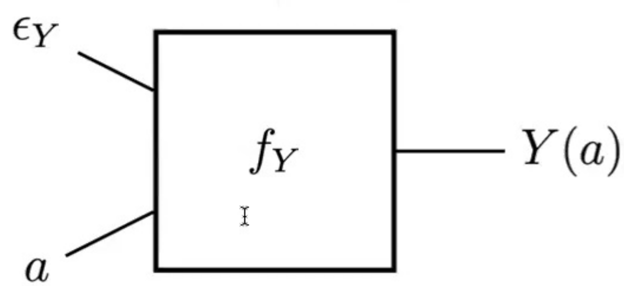
- Conceptualize cause-effect relationships by means of (random) responses to hypothetical interventions (potential outcomes):

$$Y(a) \equiv \text{“}Y \text{ if } A\text{, possibly contrary to fact, had value } a\text{.”}$$

- Notation first appears in Jerzy Neyman's Masters thesis.
- Note: very different from:

$$Y \mid a \equiv \text{“}Y \text{ if we observed } A \text{ to have value } a\text{.”}$$

- Can *equivalently* define $Y(a) \leftarrow f_Y(a, \epsilon_Y)$ in terms of *an invariant causal mechanism (structural equation)*:



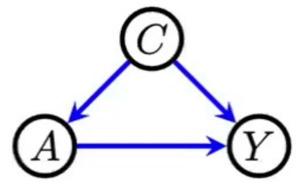
- First appeared in Sewall Wright's pedigree analysis work (for the linear case), extended by Haavelmo et al to econometrics models, generalized to the non-parametric case by Pearl et al.

Schools of Thought on Causal Modeling

- Potential outcomes (Jerzy Neyman by way of Donald Rubin):
  - Specify causal assumptions algebraically.
  - Emphasis on (potentially counterfactual) random variables.
  - Common in statistics and public health.
- Causal graphs (Sewall Wright by way of Pearl, Spirtes, Glymour, Scheines):
  - Specify causal assumptions graphically.
  - Emphasis on operators (e.g. the do() operator) and structural equations.
  - Common in computer science (and CMU philosophy).
- Despite some historical antagonism, the frameworks are complementary, and a unification exists based on *Single World Intervention Graphs* of Richardson and Robins.
- For example, it is possible (and very useful!) to express Pearl's do-calculus in terms of potential outcomes (S et al, 2019).

Example: the Conditionally Ignorable / "Backdoor" Model

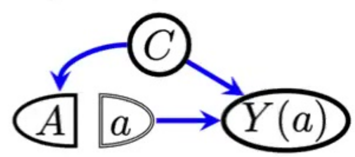
- A (drug), Y (death), $\vec{C}$ (health status).
- Model (graphical):



- Model (algebraic): $\left(Y(a) \perp\!\!\!\perp A \mid \vec{C}\right)$, $p(A|\vec{C}) > 0$.
- Informally: "A causes Y, $\vec{C}$ causes A and Y, and there are no unobserved common causes of A and Y."
- Implies the g-formula:

$$\mathbb{E}[Y(a)] - \mathbb{E}[Y(a')] = \int \left\{\mathbb{E}[Y|A=a,\vec{C}] - \mathbb{E}[Y|A=a',\vec{C}]\right\} p(\vec{C})d\vec{C}$$

- Single World Intervention Graph Unification. Graph directly implies the algebraic assumption by d-separation:



5/18

The (Modern) Causal Inference Pipeline

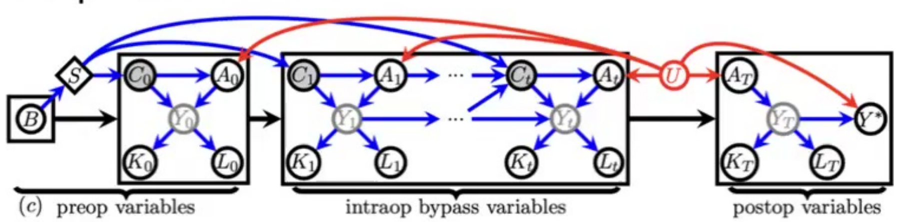
- Define a cause-effect question mathematically in terms of a causal parameter (usually as a hypothetical experiment).
- Elicit a causal model (encoding assumptions) from an expert, or *learn this model from data*.
- Identification: is the causal parameter uniquely expressible in terms of available data, given the causal model?
- Estimation: construct a procedure to estimate an identified parameter from data (efficiency? robustness?)
- Causal parameters generally *counterfactual*: emphasis on robustness of procedures to assumption violations, sensitivity analysis, etc.
- Quantifying uncertainty is very important.

Causal Inference: The (Current) State of the Art

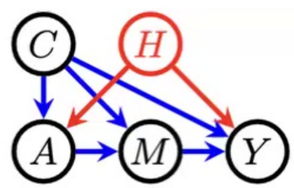
- Model elicitation: greatly simplified by Pearl's graphical modeling formalism: "Probabilistic reasoning in intelligent systems," (1988), "Causality," (2000).
- Learning models from data: "Causation, prediction and search," (2000), Spirtes, Glymour Scheines. The PC algorithm, the FCI algorithm and extensions.
- Identification theory: g-computation (Robins, 1986) for the fully observed case, the ID algorithm for the hidden variable case (Tian and Pearl, 2002, S and Pearl 2006, Huang and Valtorta 2006). Extensions to many settings: mediation analysis, counterfactual policies, network data, transportability, data fusion, selection bias, etc.
- Estimation theory: semi-parametric theory based on influence functions (Bickel et al, 1993), (van der Vaart, 1991), (Newey, 1990). Recently popular approach of "double machine learning" (using flexible nuisance models) based on earlier foundational work of Robins (2004), in the context of sequential decision-making, (Tchetgen Tchetgen et al (a) and (b), 2005) in the context of higher order influence functions, and (Zheng and van der Laan, 2011) in the context of targeted learning.

The Causal Inference Pipeline: an Example

- Intraoperative events during open heart surgery may damage kidneys (leading to acute kidney injury).
- Can we find out which events are causal?
- **Formulating the question:** $\mathbb{E}[Y^*(a_1, \ldots, a_K)]$, where Y^* is post-operative AKI, $A_1, \ldots, A_k$ are intraoperative events. This is an example of a *marginal structural model* (causal regression).
- **Eliciting the model:**
  - Complicated version:



  - Simple version:

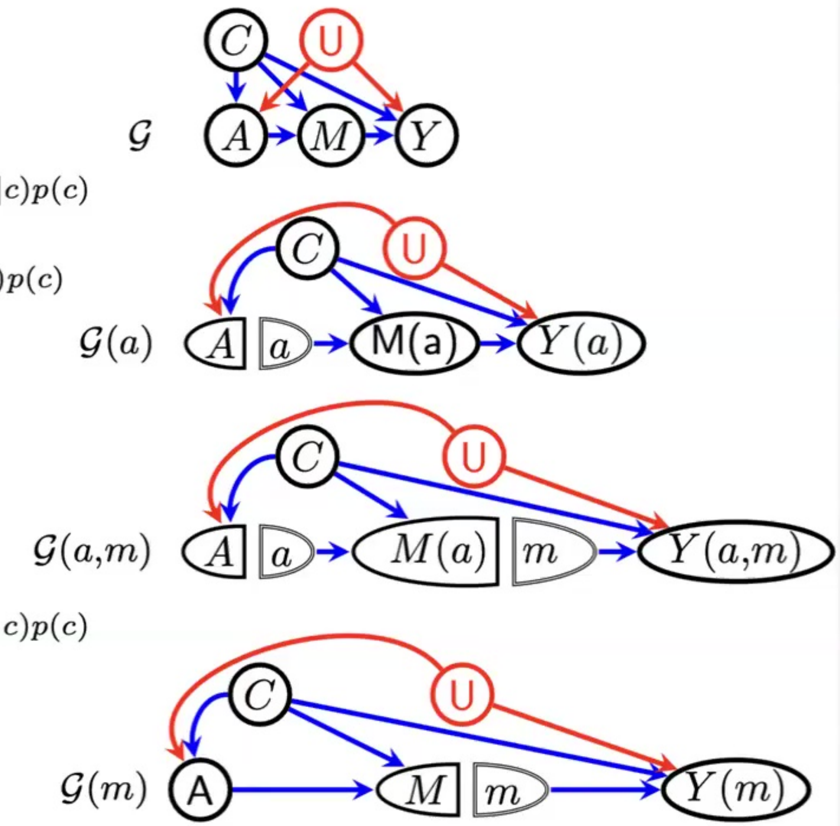


8/18

The Causal Inference Pipeline: an Example (Continued)

- **Identification of target parameters:**
- An example derivation using SWIGs and the 3 potential outcomes calculus rules:

$$
\begin{aligned}
& p(Y(a)) \\
&=^{p} \sum_{m,c} p(Y(a))|M(a)=m,c)p(M(a)=m|c)p(c) \\
&=^{2,\mathcal{G}(a)} \sum_{m} p(Y(a))|M(a)=m,c)p(m|a,c)p(c) \\
&=^{2,\mathcal{G}(a,m)} \sum_{m} p(Y(a,m)|c)p(m|a,c)p(c) \\
&=^{3,\mathcal{G}(a,m)} \sum_{m} p(Y(m)|c)p(m|a,c)p(c) \\
&=^{p} \sum_{m}\sum_{a'} p(Y(m)|a',c)p(a'|c)p(m|a) \\
&=^{2,\mathcal{G}(m)} \sum_{m}\sum_{a'} p(Y|m,a',c)p(a'|c)p(m|a,c)p(c)
\end{aligned}
$$



9/18

The Causal Inference Pipeline: an Example (Continued)

- **Functional estimation:**
- What's the best estimator for $\sum_{m,a',c} \mathbb{E}[Y|m,a',c]p(a'|c)p(m|a,c)p(c)$?
- If $p(y,m,a|c)$ is parametric, the MLE plug-in estimator.
- If not, the plug-in estimators exhibits *first order bias* (due to the mismatch between the loss we want, and the loss we used).
- If want to use flexible machine learning function learners for the above functional, want an estimator that "doesn't care" (is orthogonal to) the way nuisance functions are learned.
- Such estimators are derived in semi-parametric theory of influence functions.
- In particular, the IF based estimator for the above functional is (Fulcher, Tchetgen Tchetgen, and S, 2020):

$$\begin{aligned} U(c,a,m,y;\psi) \quad &= \quad \frac{p(m|a,c)}{p(m|a',c)}\{y - \mathbb{E}[y|m,a',c]\} \\ &+ \frac{\mathbb{I}(a'=a)}{p(a'|c)}\left\{\sum_{a'} \mathbb{E}[y|m,a',c]p(a'|c) - \sum_{m,a'} \mathbb{E}[y|m,a',c]p(m|a',c)p(a'|c)\right\} \\ &+ \sum_{m} \mathbb{E}[y|m,a',c]p(m|a,c) - \psi, \end{aligned}$$

with the estimating equation: $\frac{1}{n}\sum_{i=1}^{n} U(c_i,a_i,m_i,y_i;\hat{\psi}) = 0.$

Connections to Other Fields and Methods

- Strong connections between causal inference and missing data.
  - Connecting treatment A and outcome Y:

$$Y(a) \equiv \text{“}Y \text{ if } A\text{, possibly contrary to fact, had value } a\text{.”}$$

  - Connecting an observability indicator R and outcome Y that is potentially missing:

$$Y(r = 1) \equiv \text{“}Y \text{ if it, possibly contrary to fact, had been observed.”}$$

- Causal inference methods of great interest in many empirical disciplines: social science, psychology, public health, medicine, etc.
- Connections to quantum theory: Bell inequality is a fact about (classical) hidden variable DAGs! Lots of folks in both physics and causal inference communities are thinking about implications of structured systems with hidden variables for learning graphs from data, and falsifying causal explanations of observed (classical and quantum) phenomena.
- Lots of mutual interest between machine learning and causal inference: I think these disciplines have much to teach each other!

What Can Machine Learning Teach Causal Inference?

- Constructive optimism.
- Finite sample results.
- Emphasis on tasks and validation.
- Excellent predictive performance very helpful as a subroutine in semi-parametric estimation of causal effects (use predictors as nuisance models).
- Powerful optimization methods are likely helpful for solving estimating equations.
- Surprisingly many problems are regression problems.

What Can Causal Inference Teach Machine Learning?

- Constructive pessimism.
- Asymptotic identification and estimation results.
- Emphasis on transparency and assumptions.
- Semi-parametric theory. (Often the correct approach for many ML problems e.g. reinforcement learning).
- Parameter identifiability is important (at least some "model fragility" issues go back to lack of identification).
- Principled ways of thinking about problems that aren't supervised.
- Causal language is helpful when thinking about model stability, invariance, fairness, interpretability.
- Principled methods for missing data problems.
- Not every problem is a regression problem.

(Opinion): Important Open Problems in Causal Inference

- General and "well-behaved" hidden variable DAG likelihoods (avoiding model misspecification and singularities).
- Principled post-selection inference (how to estimate causal effects after we learn a causal model from data).
- Finite sample causal inference results.
- Getting around the "infinite regress" in problems where the loss has to be estimated.
- What observable implications does a given hidden variable causal model have?
- Causal inference with multimodal data.

Conclusions

- Causal inference ideas are quite old, modern causal inference grew out of early classical statistical analyses of experimental data.
- Some historical dichotomies in approach to causal inference are false dichotomies!
- Causal inference has many connections to other areas of statistics and machine learning, empirical science, including physics (in particular quantum theory).
- Machine learning and causal inference are complementary in area of study and "attitude" – and have much to teach each other.
- Lots of interesting open problems!

Thank you for listening!

Contact info:

Ilya Shpitser ilyas@cs.jhu.edu