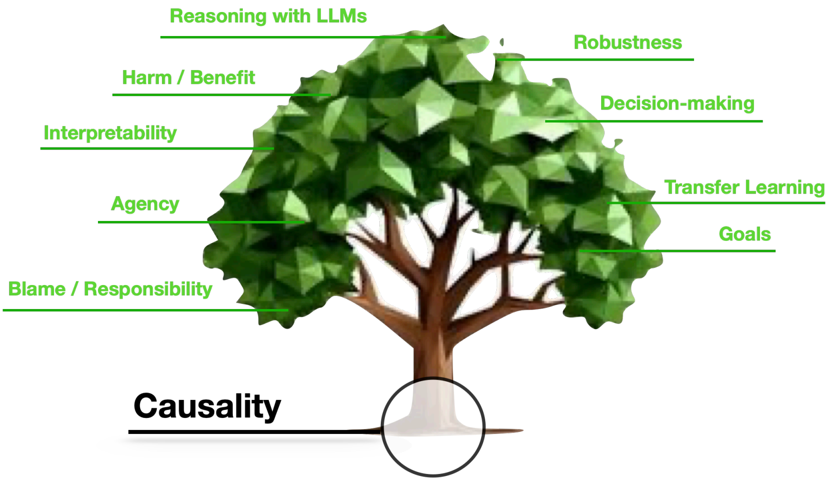
An Introduction to Causal Inference

Alexis Bellot
Slides at alexisbellot.github.io

Google DeepMind



Reasoning with LLMs
Robustness
Harm / Benefit
Decision-making
Interpretability
Transfer Learning
Agency
Goals
Blame / Responsibility
Causality

Two Approaches to Making Causal Conclusions

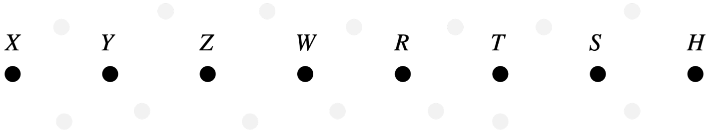
The ***logical*** approach:

The task is to automate the derivation of (sound) causal conclusions from typically very **well-articulated assumptions**

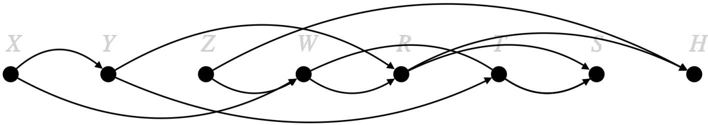
The ***human-centered*** approach:

The task is to generate causal conclusions in a **human-like** way, typically from (possibly inconsistent) **subjective causal judgments** obtained from the rich human experience

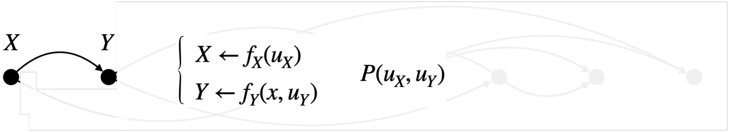
You get to measure a set of attributes of some system



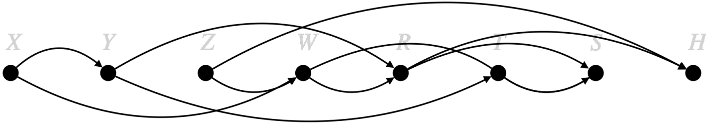
You get to measure a set of attributes of some system that are inter-connected in a complex way



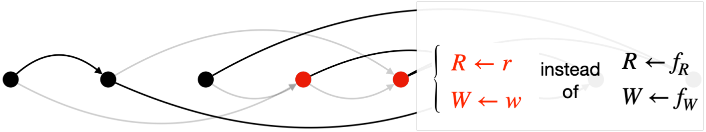
Graph is a **representation** of an underlying (structural) model M:
a collection of functions $\{f_V\}_{V \in \boldsymbol{V}}$ and distributions $P(\boldsymbol{U})$



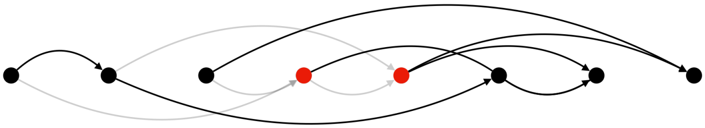
If you sample from M you get evidence of the **observational** probability of events e.g. $P(X = x, Y = y)$ or simply $P(x, y)$



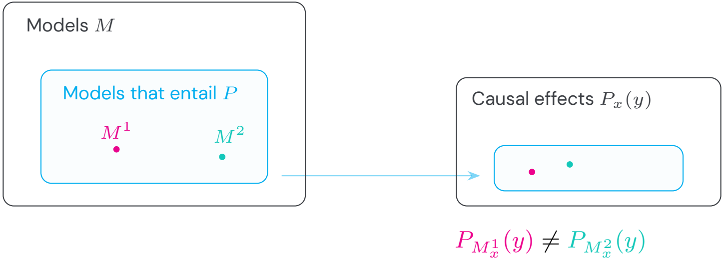
Might want to reason about the effect of **interventions** of your system M



Interventional probabilities of events, e.g. $P_x(y)$,
also written $P(y_x), P_{M_x}(y), P(y \mid do(x))$



Observational data does not uniquely determine the **effect of interventions** *ever*, see e.g. (Bareinboim et al., 2022, Causal Hierarchy Theorem).

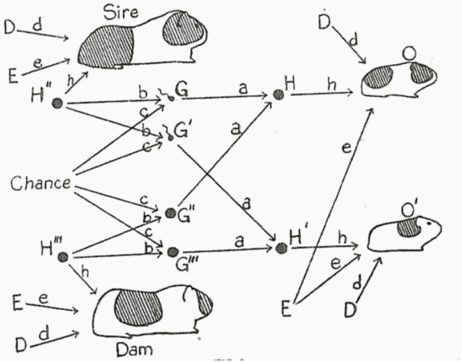


How then do we arrive at Causal Conclusions?

Wright (1920) was able to *predict the effect of interventions* by relating the parameters of an **assumed model of heredity** to the **correlations in data**.

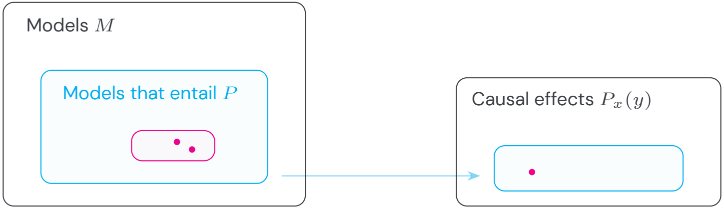
How then do we arrive at Causal Conclusions?

Wright (1920) was able to *predict the effect of interventions* by relating the parameters of an **assumed model of heredity** to the **correlations in data**.



How then do we arrive at Causal Conclusions?

Wright (1920) was able to *predict the effect of interventions* by relating the parameters of an **assumed model of heredity** to the **correlations in data**.



Models that **entail** P and **induce Wright's heredity model**

Research Program: Causal Inference

Logical Approach (Pearl, 1995; Rosenbaum and Rubin, 1983; Rubin, 1974).

(1) Query

e.g. $P_x(y)$

(2) Data distribution

e.g. $P(x, y, z)$

(3) Assumption

e.g. Causal diagram, independent noise terms, ignorability assumptions, etc.

→ **Causal inference engine** →

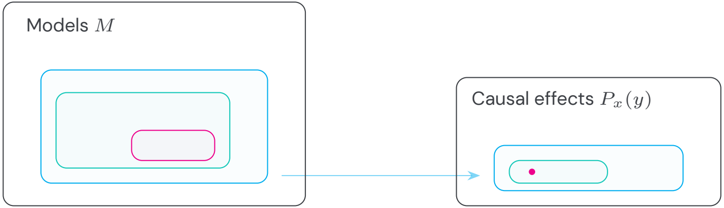
Uniquely computable from P?

If so, how?

If not, what is the best bound?

Research Program: Causal Inference

Assumptions and data constrain the space of possible models and, as a consequence, the set of possible causal effects.



Models that entail P

Models that entail P and induce causal diagram $\mathcal{G}$

Models that entail P and induce causal diagram $\mathcal{G}$ and are linear with non-Gaussian error terms

An *Identifiable* Example

(1) Query

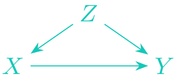
e.g. $P_x(y)$

(2) Data distribution

e.g. $P(x, y, z)$

(3) Assumption

$\mathcal{G}$



In the *observational* regime $\mathcal{G}$ induces,

$$P(y, z, x) = P(y \mid x, z)P(x \mid z)P(z)$$

In the **interventional** regime,

$$P_x(y, z) = P(y \mid x, z)1\{X = x\}P(z)$$

The ***query*** may be evaluated using,

$$\begin{aligned} P_x(y) &= \sum_z P_x(y, z) \\ &= \sum_z P(y \mid x, z)P(z) \end{aligned}$$

Foundational Result: *Truncated Product*

In the interventional regime,

$$P_x(y, z) = P(y \mid x, z)1\{X = x\}P(z)$$

Theorem *(Truncated Factorization)*. Given a causal diagram $\mathcal{G}$ (no unobserved confounding) we can always predict the effect of an intervention on $\boldsymbol{X} \leftarrow \boldsymbol{x}$,

$$P_{\boldsymbol{x}}(\boldsymbol{V} = \boldsymbol{v}) = \prod_{V \in \boldsymbol{V} \setminus \boldsymbol{X}} P(V = v \mid Pa_V = pa_V)$$

The *Back-door Adjustment* Formula

$$P_x(y) = \sum_z P(y \mid x, z) P(z)$$

Theorem *(Back-door Adjustment)*. Given a causal diagram $\mathcal{G}$ (no unobserved confounding), we may evaluate the effect of an intervention $\boldsymbol{X} \leftarrow \boldsymbol{x}$ by adjustment on the variables $\boldsymbol{Z}$ not affected by $\boldsymbol{X}$ (its non-descendants),

$$P_{\boldsymbol{x}}(\boldsymbol{y}) = \sum_{\boldsymbol{z}} P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z}) P(\boldsymbol{z}), \quad \mathbb{E}_{P_{\boldsymbol{x}}}[\boldsymbol{Y}] = \sum_{\boldsymbol{z}} \mathbb{E}_P[\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{z}] P(\boldsymbol{z})$$

The *Conditional Exogeneity* Restriction

$$P_x(y) = \sum_z P(y \mid x, z)P(z)$$

Theorem *(Counterfactual / Potential Outcomes Restrictions)* If $\boldsymbol{Y}_{\boldsymbol{X}} \perp\!\!\!\perp \boldsymbol{X} \mid \boldsymbol{Z}$ then,

$$P_{\boldsymbol{x}}(\boldsymbol{y}) = \sum_{\boldsymbol{z}} P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z})P(\boldsymbol{z}), \quad \mathbb{E}_{P_{\boldsymbol{x}}}[\boldsymbol{Y}] = \sum_{\boldsymbol{z}} \mathbb{E}_P[\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{z}]P(\boldsymbol{z})$$

Identification versus *Estimation*

$$\mathbb{E}_{P_{\boldsymbol{x}}}[\boldsymbol{Y}] = \sum_{\boldsymbol{z}} \mathbb{E}_P[\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{z}] P(\boldsymbol{z})$$

1. The ***regression estimator***:

$$\mathbb{E}_{P_x}[Y] = \sum_z \mathbb{E}_P[Y \mid x, z] P(z) = \mathbb{E}_{z \sim P} \left[\, \mathbb{E}_P[Y \mid x, z] \,\right]$$

2. The ***probability-weighted estimator***:

$$\mathbb{E}_{P_x}[Y] = \sum_{x,y,z} y \frac{P(z) 1\{X = x\}}{P(x, z)} P(x, y, z) = \mathbb{E}_P \left[\frac{P(z) 1\{X = x\}}{P(x, z)} Y \right] = \mathbb{E}_P \left[\frac{1\{X = x\}}{P(x \mid z)} Y \right]$$

A *non-Identifiable* Example

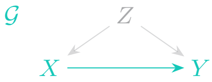
(1) Query

e.g. $P_x(y)$

(2) Data distribution

e.g. $P(x, y, z)$

(3) Assumption



In the *observational* regime $\mathcal{G}$ induces,

$$P(y, x) = \sum_z P(y, z, x)$$

$$= \sum_z P(y \mid x, z) P(x \mid z) P(z)$$

In the **interventional** regime,

$$P_x(y) = \sum_z P_x(y, z)$$

$$= \sum_z P(y \mid x, z) 1\{X = x\} P(z)$$

A *non-Identifiable* Example

(1) Query

e.g. $P_x(y)$

(2) Data distribution

e.g. $P(x, y, z)$

(3) Assumption

$\mathcal{G}$

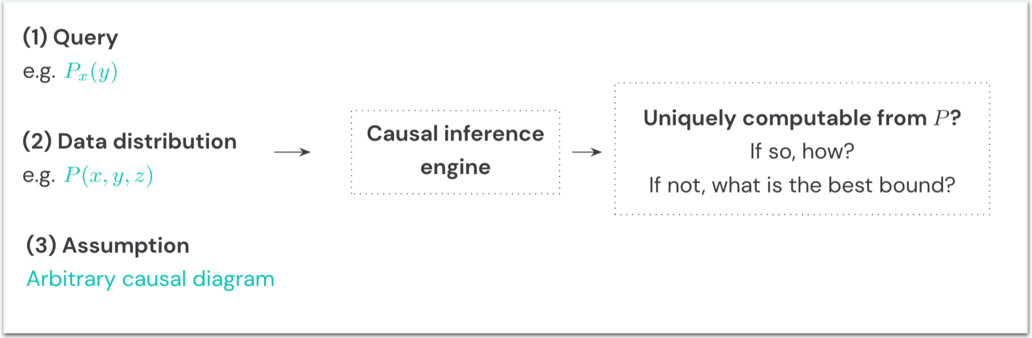
In the *observational* regime $\mathcal{G}$ induces,

$$P(y, x) = \sum_z P(y, z, x)$$
$$= \sum_z P(y \mid x, z) P(x \mid z) P(z)$$

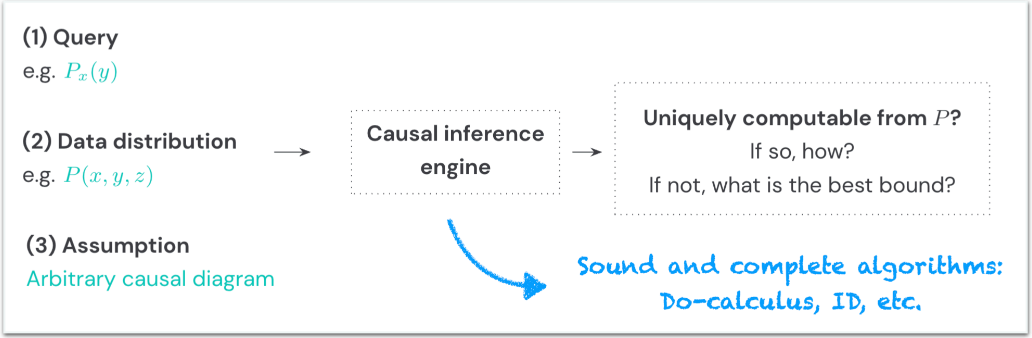
In the **interventional** regime,

$$P_x(y) = \sum_z P_x(y, z)$$
$$= \sum_z P(y \mid x, z) 1\{X = x\} P(z)$$

Research Program: Causal Inference



Research Program: Causal Inference



(Wider) Research Program: Causal Inference in Medicine and AI

Counterfactuals, domain generalization, policy evaluation and optimization, etc.

(1) Query

e.g. $P_x(y)$

(2) Data distribution →

e.g. $P(x, y, z)$

Causal inference engine →

Uniquely computable from P?

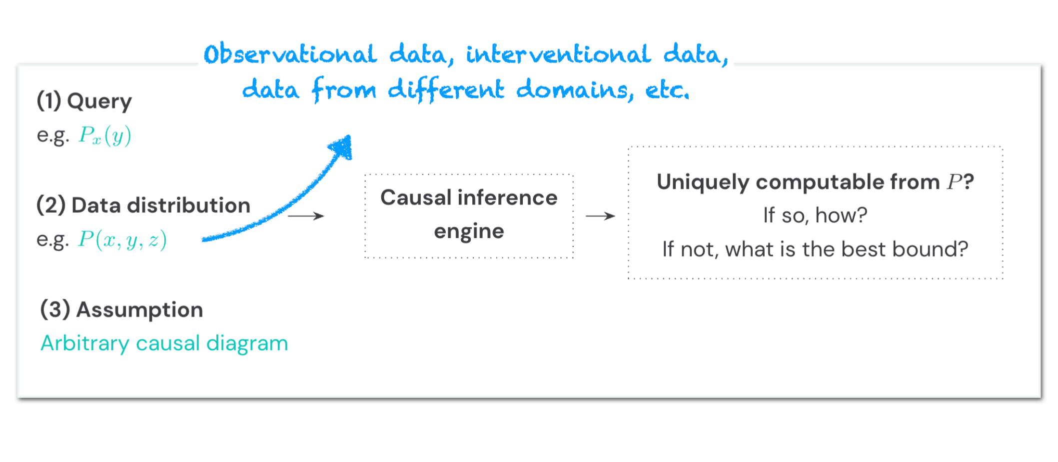
If so, how?

If not, what is the best bound?

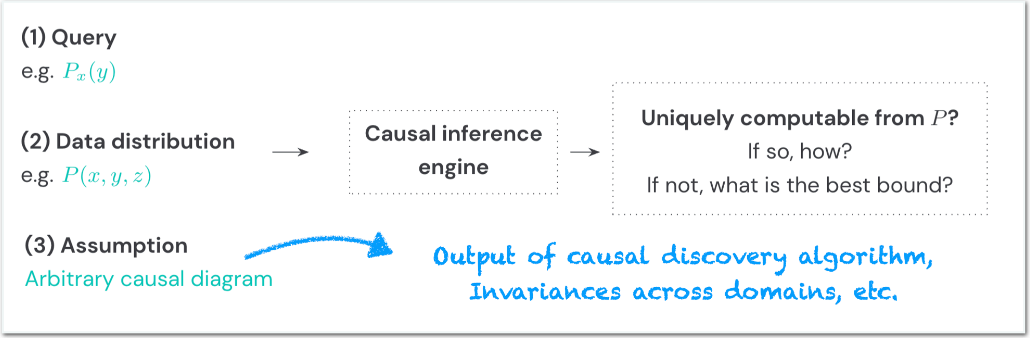
(3) Assumption

Arbitrary causal diagram

(Wider) Research Program: Causal Inference in Medicine and AI



(Wider) Research Program: Causal Inference in Medicine and AI



Typical Causal Inference Questions

What ***effect*** can we expect from a treatment given to patients with stage III cancer?

What fraction of health-care expenditure can be ***attributed*** to respiratory illnesses?

I have been suffering from obesity for two years, would my BMI be different ***had I*** adhered to a vegan diet?

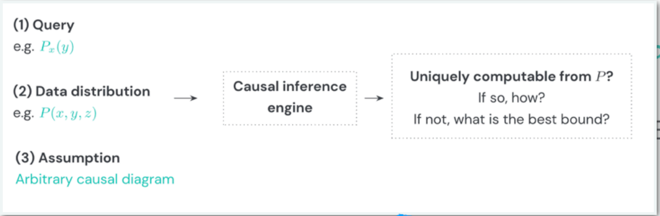
Can hospital admission statistics prove systematic ***discrimination*** against a given minority group?

Typical Causal Inference Questions

What *effect* can we expect from a treatment given to patients with stage III cancer?

What f ... *d* to respiratory illnesses?

I have ... BMI be different *had I*
adhere ...



Can hospital admission statistics prove systematic *discrimination* against a given minority group?

Concluding Remarks

Many questions are ***provably*** difficult to answer from data

In practice, most of the work is in the definition of plausible assumptions rather than modelling the data as the target for estimation depends a lot on the causal structure of the variables involved in your problem

Two paradigms for Data Science: **Data-driven** versus ***Model-based***

Appendix: External validity

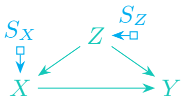
(1) Query

e.g. $P_x^*(y)$

(2) Data distribution

e.g. $P^*(z), P_x(z, y)$

(3) Assumption



The experimental study does not immediately apply in our target domain as $P_x^*(y) \neq P_x(y)$ but it can be computed by re-weighting according to $P^*(z)$:

$$
\begin{aligned}
P_x^*(y) &= \sum_z P_x^*(y, z) \\
&= \sum_z P_x^*(y \mid z) P_x^*(z) \\
&= \sum_z P_x(y \mid z) P^*(z) \\
&= \sum_z P_x(y, z) \frac{P^*(z)}{\sum_y P_x(z, y)}
\end{aligned}
$$

Bibliography I

E. Bareinboim, J. D. Correa, D. Ibeling, and T. Icard. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, page 507–556. Association for Computing Machinery, NY, USA, 1st edition, 2022.

J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Bibliography II

S. Wright. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences*, 6(6):320–332, 1920.