Causality for Distributional Robustness

DSTS 50 year anniversary

Nikolaj Thams
PhD Student at MATH, Univ. Copenhagen

Using Causality for Distributional Robustness

- The road from full to partial exogeneity
- Some detours about our work

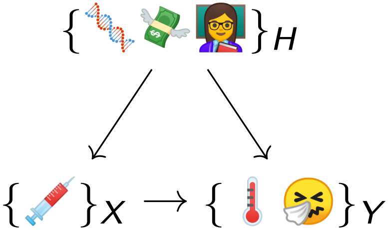
Perfect exogeneity: Randomized experiments

Goal: Learn effect X on Y.

$$\{💉\}_X \longrightarrow \{🌡️\ 🤧\}_Y$$

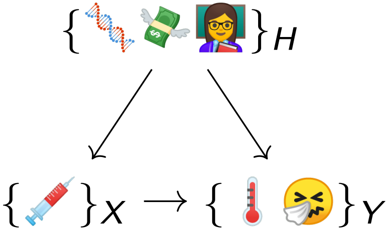
Perfect exogeneity: Randomized experiments

Goal: Learn effect X on Y.
Problem: Latent confounding factors H.



Perfect exogeneity: Randomized experiments

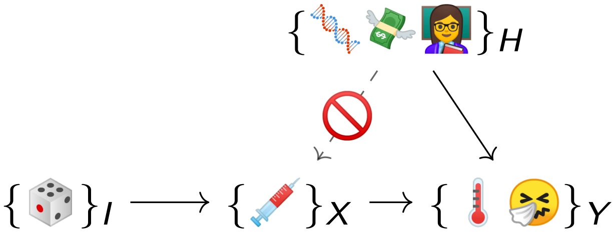
Goal: Learn effect X on Y.
Problem: Latent confounding factors H.
Solution: Use randomization to break confounding.



Why care about causal instead of confounded effect? Generalization!

Perfect exogeneity: Randomized experiments

Goal: Learn effect X on Y.
Problem: Latent confounding factors H.
Solution: Use randomization to break confounding.



Why care about causal instead of confounded effect? Generalization!

Enough exogeneity: Instrumental variables

We can't always randomize:

- Allocation unethical (e.g. smoking 🚬 or seatbelts 🔥🏎️)
- Interventions hard to define (e.g. data is pixels of image 🖼️)

Enough exogeneity: Instrumental variables

We can't always randomize:

- Allocation unethical (e.g. smoking 🚬 or seatbelts 🔥🏎️)
- Interventions hard to define (e.g. data is pixels of image 🖼️)

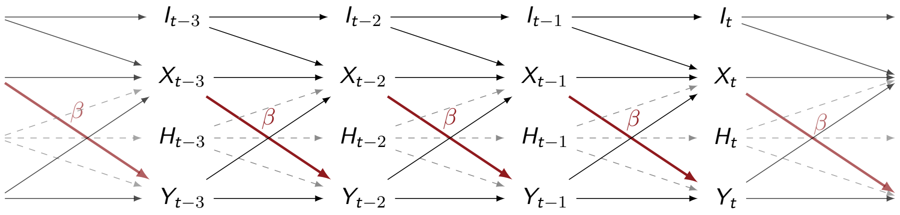
Instead: Instrumental variable (IV): I 1) independent of H and 2) only indirect effect on Y

$$\begin{tikzcd} & & H \arrow[dl] \arrow[dr] & \\ I \arrow[r] & X \arrow[rr] & & Y \end{tikzcd}$$

OLS regression: $Y - X\hat{\beta} \perp\!\!\!\perp X \longrightarrow$ IV regression: $Y - X\hat{\beta} \perp\!\!\!\perp I$.

$\rightarrow$ Causal effect β can still be estimated (Imbens and Rubin 2015) under assumptions, e.g. high rank I.

Detour 1: Instruments in time series



Currently, we are working on methods in systems with memory

- Proper adjustment for memory
- Using also lagged instruments as instruments, to increase rank of instrument,

(Thams et al. 2021).

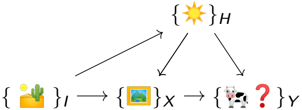
Some exogeneity: Invariance

{🖼️}$_X$ → {🐄❓}$_Y$

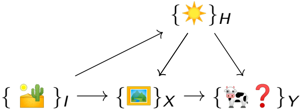
Some exogeneity: Invariance

$$\begin{tikzcd} & \{\text{☀️}\}_H \arrow[dl] \arrow[dr] & \\ \{\text{🖼️}\}_X \arrow[rr] & & \{\text{🐄❓}\}_Y \end{tikzcd}$$

Some exogeneity: Invariance

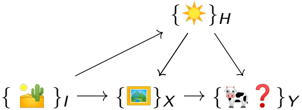


Some exogeneity: Invariance



- Causal solution: $Y - X\beta \perp\!\!\!\perp I$. IV require $\dim(I) \geq \dim(X)$.
  Image data: $\dim(X) = 1024^2$.

Some exogeneity: Invariance

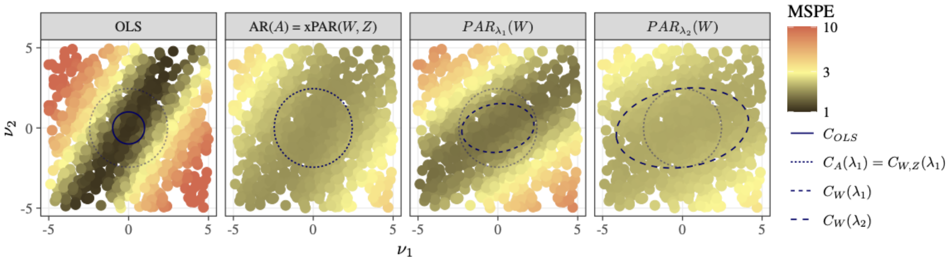


- Causal solution: $Y - X\beta \perp\!\!\!\perp I$. IV require $\dim(I) \geq \dim(X)$.
  Image data: $\dim(X) = 1024^2$.
- Instead, we can penalize our regression

$$\hat{\beta} \in \arg\min_{\beta} \|Y - X\beta\|^2 + \lambda \|\operatorname{cov}(Y - X\beta, I)\|^2,$$

Rothenhäusler et al. 2021 show that $\hat{\beta}$ generalizes well (but less than with full causal knowledge).

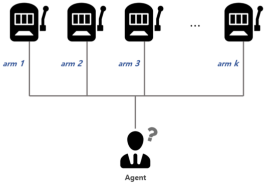
Detour 2: Generalization with proxy variables

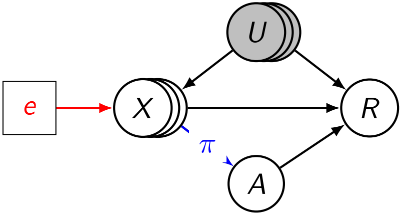


In Oberst et al. 2021, we show that

- If I is unobserved, but we have a noisy measurement (proxy) W of I, we still obtain generalization, although less.
- If we observe two proxies W, Z of I, we generalize as well as with I itself.

Detour 3: Using environments to generalize in decision making





Contextual (= covariates) + Bandits

In Saengkyongam et al. 2021,

- We develop invariant policies, that is policies that do not use confounded features.
- To do so, we use exogeneous environments, e.g. hospitals.
- Under assumptions, we generalize to new environments.

Conclusions

- If strong exogeneity is present, we can estimate causal effects 🎲
- If weak exogeneity is present, we can not estimate causal effects, but still get methods that generalize better 🎰
- We considered linear effects, but much can be generalized to non-linear functions 📈(Christiansen et al. 2021)

References I

- Christiansen, Rune, Niklas Pfister, Martin Emil Jakobsen, Nicola Gnecco, and Jonas Peters (2021). "A causal framework for distribution generalization". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1.
- Imbens, Guido W and Donald B Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Oberst, Michael, Nikolaj Thams, Jonas Peters, and David Sontag (2021). "Regularizing towards Causal Invariance: Linear Models with Proxies". In: *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. Proceedings of Machine Learning Research. PMLR, pp. 8260–8270.
- Rothenhäusler, Dominik, Nicolai Meinshausen, Peter Bühlmann, and Jonas Peters (2021). "Anchor regression: Heterogeneous data meet causality". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 83.2, pp. 215–246.
- Saengkyongam, Sorawit, Nikolaj Thams, Jonas Peters, and Niklas Pfister (2021). "Invariant Policy Learning: A Causal Perspective". In: *arXiv preprint arXiv:2106.00808*.

References II

Thams, Nikolaj, Rikke Nielsen, Sebastian Weichwald, and Jonas Peters (2021). "Instrumental Time Series: Correcting for the Past, Identifiability, and Learning". In: *Work in progress.*

Questions?