Machine Learning for econometrics

Causal perspective

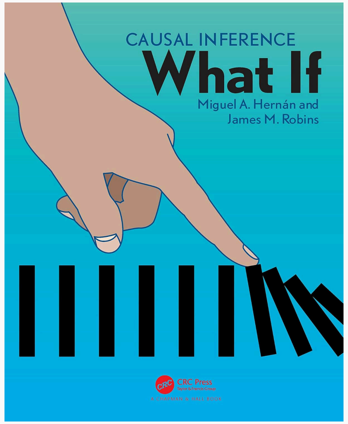
Matthieu Doutreligne

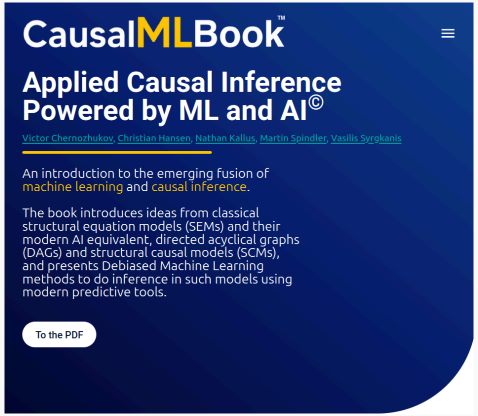
January 10, 2025

Table of contents

Introduction

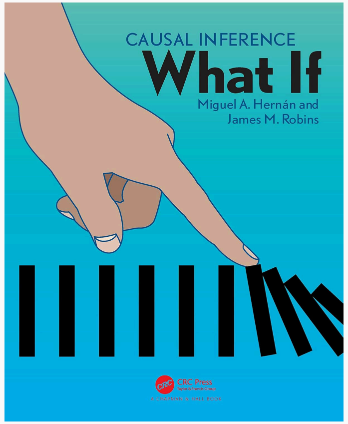
Causal inference: subfield of statistics dealing with "why questions"

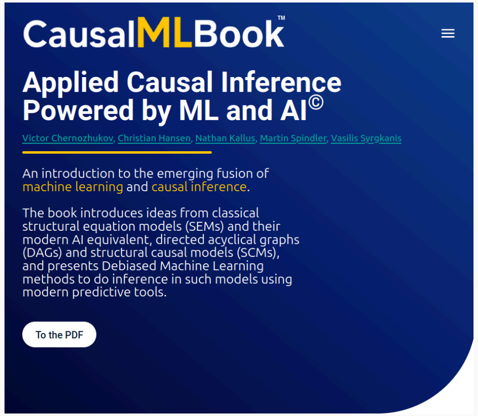




At the center of epidemiology (Hernán & Robins, 2016), econometrics (Chernozhukov et al., 2024), social sciences,

Causal inference: subfield of statistics dealing with "why questions"

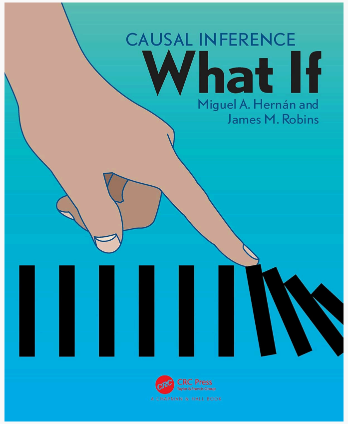


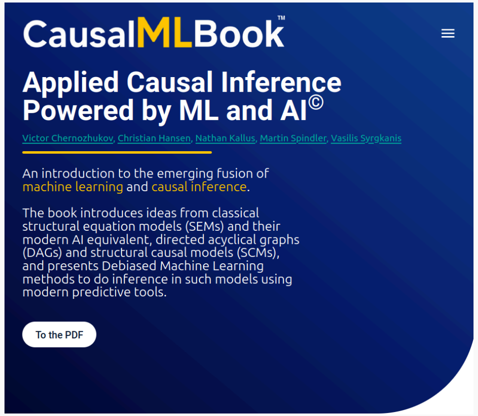


At the center of epidemiology (Hernán & Robins, 2016), econometrics (Chernozhukov, Hansen, Kallus, Spindler, & Syrgkanis, 2024), social sciences, machine learning…

Now, bridging with machine learning (Kaddour et al., 2022) : Fairness, reinforcement learning, causal discovery, causal inference for LLM, causal representations…

Causal inference: subfield of statistics dealing with "why questions"





At the center of epidemiology (Hernán & Robins, 2016), econometrics (Chernozhukov, Hansen, Kallus, Spindler, & Syrgkanis, 2024), social sciences,

This course:

- Basis of causal inference using ML appraoches (semi-parametric),
- Inspiration from epidemiology,
- Application for applied econometrics.

What is a "why question"?

- Economics: How does supply and demand (causally) depend on price?
- Policy: Are job training programmes actually effective?
- Epidemiology: How does this threatment affect the patient’s health?
- Public health : Is this prevention campaign effective?
- Psychology: What is the effect of family structure on children’s outcome?
- Sociology: What is the effect of social media on political opinions?

This is different from a predictive question

- What will be the weather tomorrow?
- What will be the outcome of the next election?
- How many people will get infected by flue next season?
- What is the cardio-vacular risk of this patient?
- How much will the price of a stock be tomorrow?

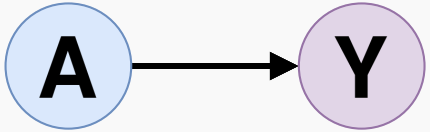
Why is prediction different from causation? (1/2)

Prediction (most part of ML): What usually happens in a given situation?

Why is prediction different from causation? (1/2)

Prediction (most part of ML): What usually happens in a given situation?

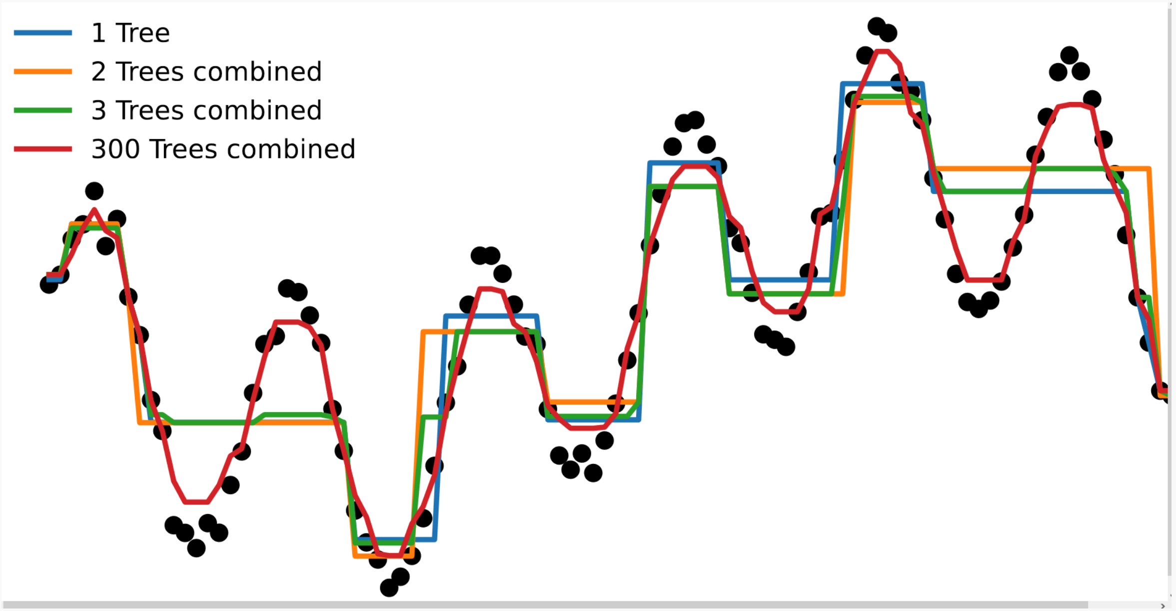
Assumption Train and test data are drawn from the same distribution.



Prediction models (X, Y)

Machine learning is (basically) pattern matching

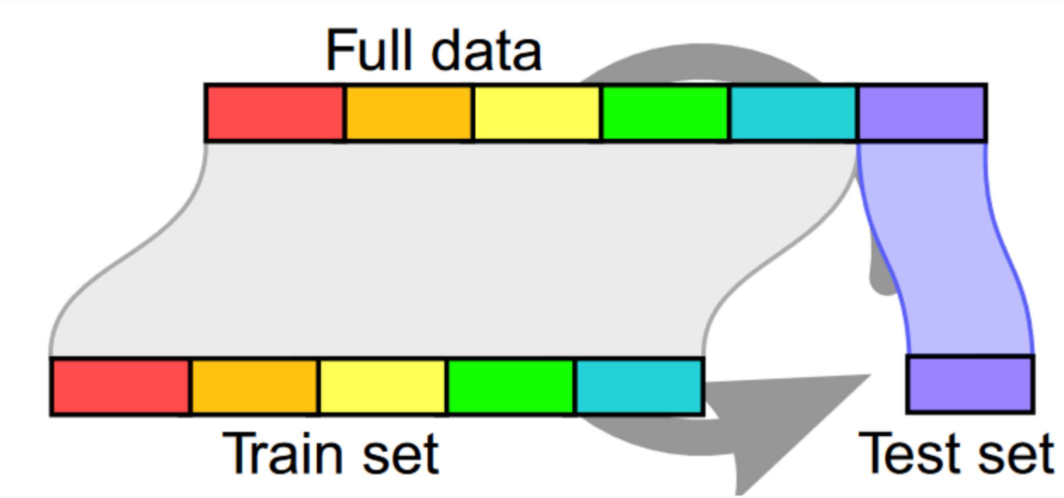
Find an estimator $f : x \to y$ that approximates the true value of y so that $f(x) \approx y$



Boosted trees : iterative ensemble of decision trees

Machine learning is pattern matching that generalizes to new data

Select models based on their ability to generalize to new data : (train, test) splits and cross validation (Stone, 1974).



“Cross validation” (Varoquaux et al., 2017)

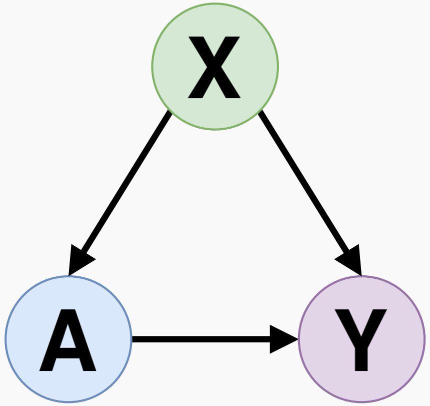
Why is prediction different from causation? (2/2)

Causal inference (most part of economists) : What would happen if we changed the system ie. under an intervention?

Why is prediction different from causation? (2/2)

Causal inference (most part of economists) : What would happen if we changed the system ie. under an intervention?

Assumption: No unmeasured variables influencing both treatment and outcome $\rightarrow$ confounders.



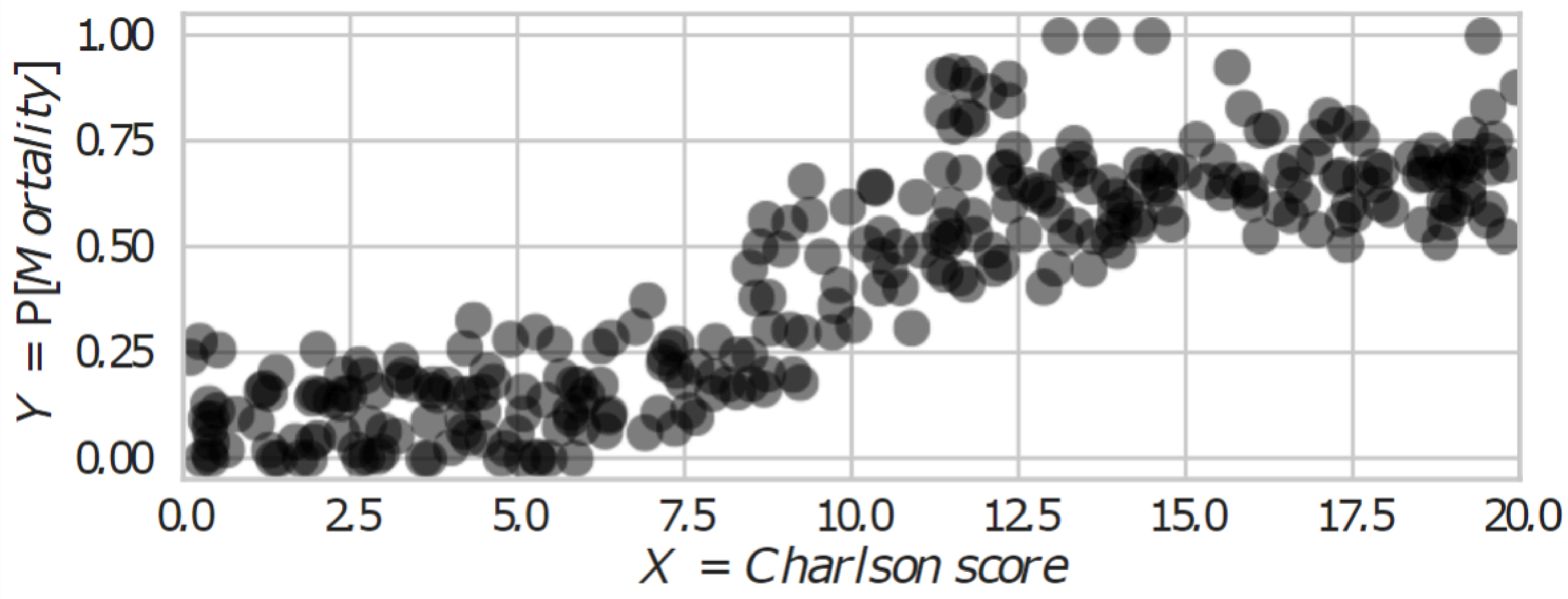
Causal inference models
$(X, A, Y(A=1), Y(A=0))$
the covariate shift between treated and control units.

Illustration of the fundamental problem of causal inference

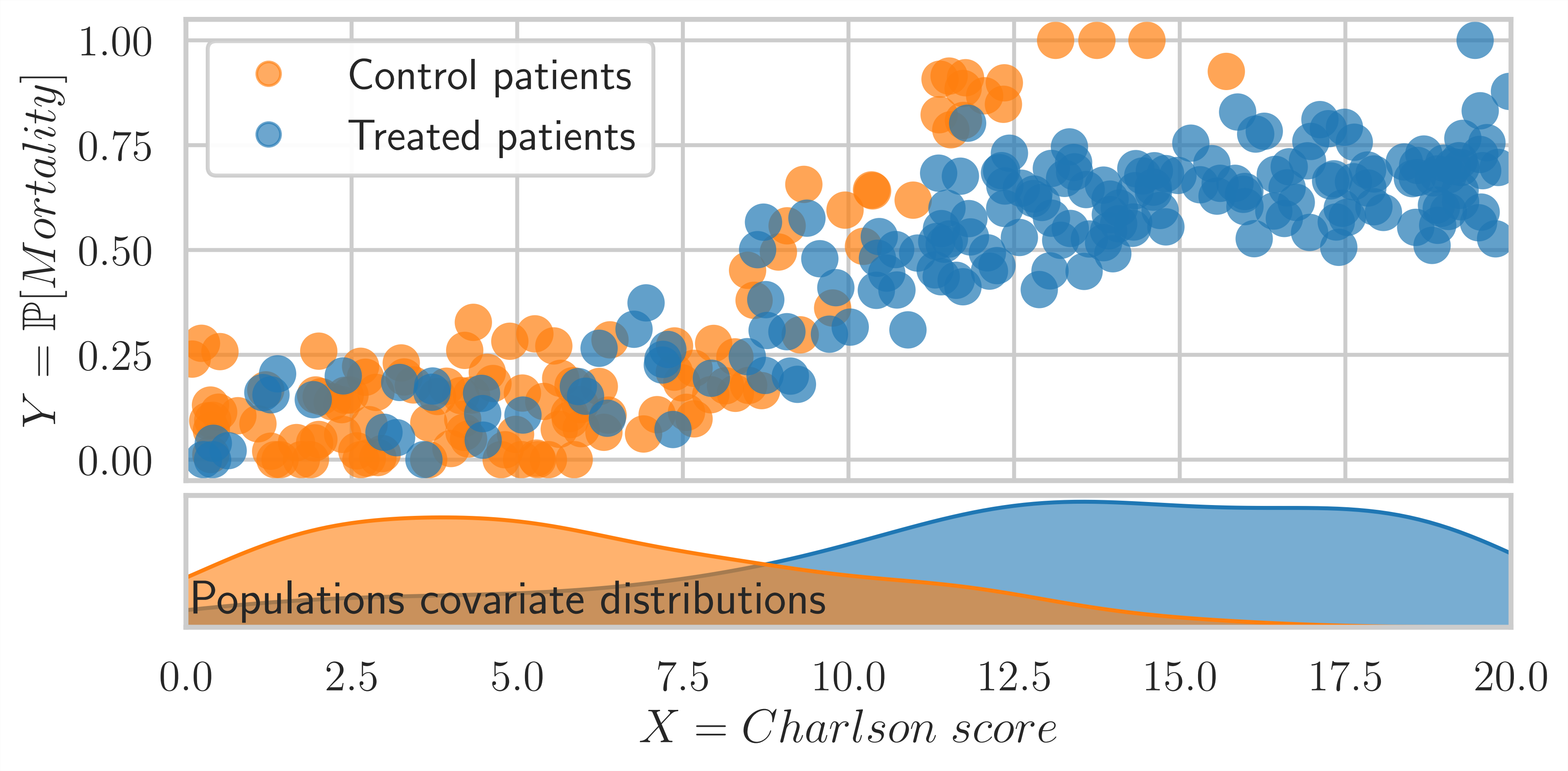
Consider an example from epidemiology:

- Population: patients experiencing a stroke
- Intervention $A = 1$: patients had access to a MRI scan **in less than 3 hours** after the first symptoms
- Comparator $A = 0$: patients had access to a MRI scan **in more than 3 hours** after the first symptoms
- $Y = \mathbb{P}[\text{Mortality}]$: the mortality at 7 days
- $X = \mathbb{P}[\text{Charlson score}]$: a comorbidity index summarizing the overall health state of the patient. Higher is bad for the patient.

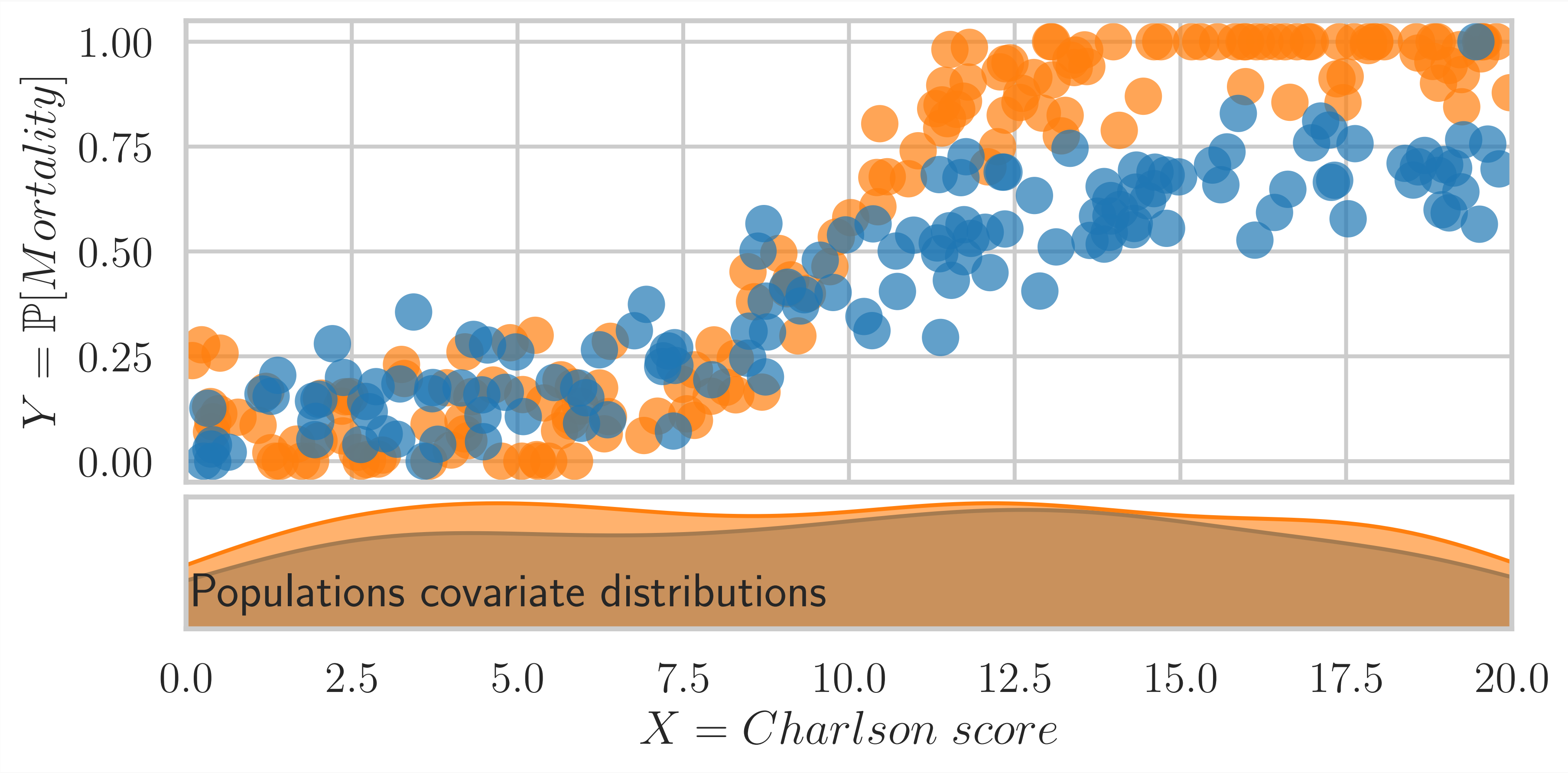
Without treatment status



With treatment status

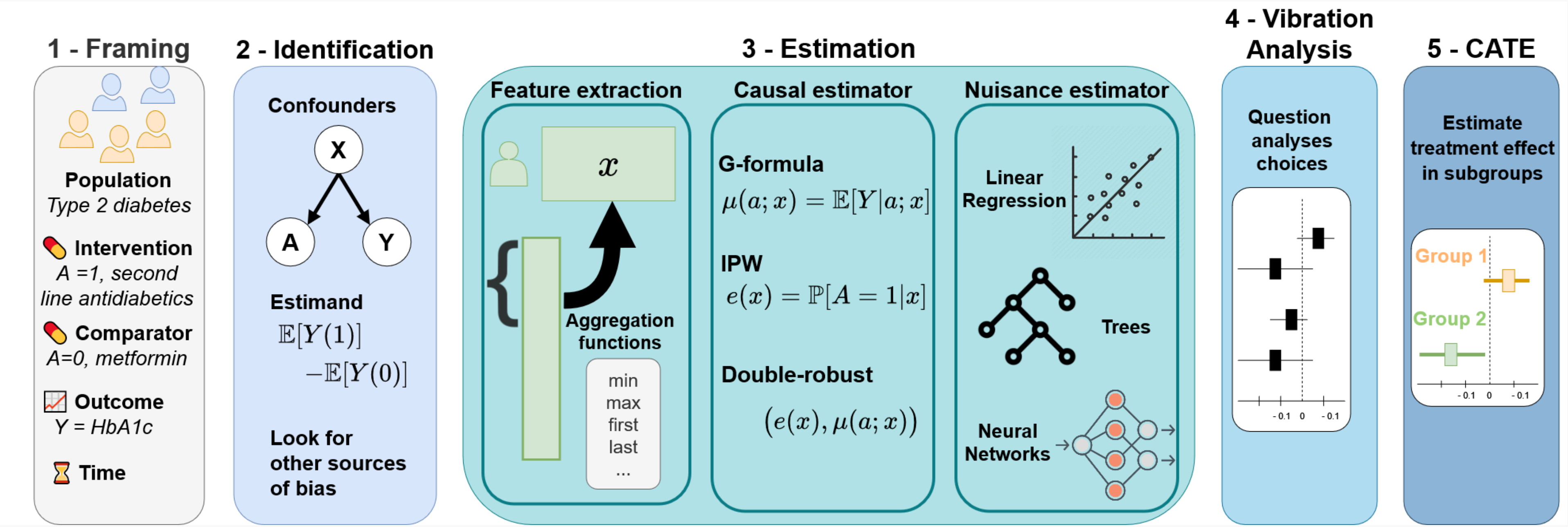


RCT case: Example in one dimension (1/2)



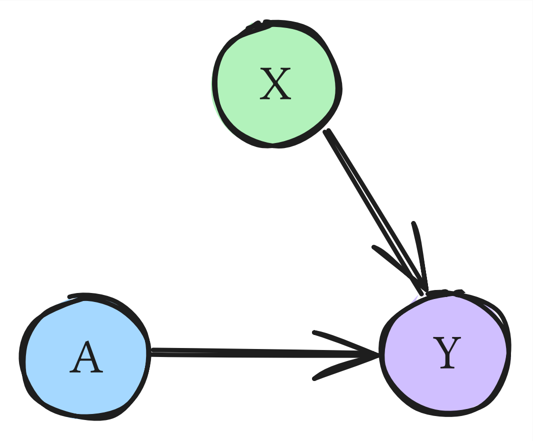
Four steps of causal inference : Framing, identification, statistical inference, vibration analysis

Complete inference flow



Randomized controlled trial (RCT) principle

- Random assignment of treatment
- Force $Y(1), Y(0) \perp A$



DAG for a RCT: the treatment is independent of the confounders

Framing: How to ask a sound causal question

Identify the target trial

What would be the ideal randomized experiment to answer the question? (Hernán & Robins, 2016)

PICO framework (Richardson et al., 1995)

- Population : Who are we interested in?
- Intervention : What treatment/intervention do we study?
- Comparison : What are we comparing it to?
- Outcome : What are we interested in?

PICO framework (Richardson, Wilson, Nishikawa, & Hayward, 1995)

- Population : Who are we interested in?
- Intervention : What treatment/intervention do we study?
- Comparison : What are we comparing it to?
- Outcome : What are we interested in?

Example with the job dataset (LaLonde, 1986)

Built to evaluate the impact of the National Supported Work (NSW) program. The NSW is a transitional, subsidized work experience program targeted towards people with long-standing employment problems.

The PICO framework

Component	Description	Example
Population	What is the target population of interest?	People with longstanding employment problems
Intervention	What is the intervention?	On-the-job training lasting between nine months and a year
Control	What is the relevant comparator?	No training
Outcome	What are the outcomes?	Earnings in 1978
Time	Is the start of follow-up aligned with intervention assignment?	The period of follow-up for the earning is the year after the intervention

Identification

Potential outcomes, (Neyman, 1923; Rubin, 1974)

The Neyman-Rubin model, let:

- Y be the outcome,
- A the (binary) treatment

For each individual, we have two potential outcomes: $Y(1)$ and $Y(0)$. But only one is observed, depending on the treatment assignment: $Y(A)$.

Directed acyclic graphs (DAG)

A tool to reason about causality

What are the causal status of each variable?

PICO framework, link to the potential outcomes

Component	Description	Notation	Example
Population	What is the target population of interest?	$X \sim P(X)$	People with longstanding employment problems
Intervention	What is the intervention?	$A \sim P(A = 1) = p_A$	On-the-job training lasting between nine months and a year
Control	What is the relevant comparator?	$1 - A \sim 1 - p_A$	No training
Outcome	What are the outcomes?	$Y(1), Y(0) \sim P(Y(1), Y(0))$	Earnings in 1978
Time	Is the start of follow-up aligned with intervention assignment?	N/A	The period of follow-up for the earning is the year after the intervention

Causal estimand: What is the targeted quantity (with potential outcomes)?

Causal estimand: What is the targeted quantity (with potential outcomes)?

- Average treatment effect (ATE)
  $\mathbb{E}[Y(1) - Y(0)]$
- Conditional average treatment effect (CATE)
  $\mathbb{E}[Y(1) - Y(0) \mid X]$

Causal estimand: What is the targeted quantity (with potential outcomes)?

- Average treatment effect on the treated (ATT): $\mathbb{E}[Y(1) - Y(0) \mid A = 1]$
- Conditional average treatment effect on the treated (CATT): $\mathbb{E}[Y(1) - Y(0) \mid A = 1, X]$

Causal estimand: What is the targeted quantity (with potential outcomes)?

Other estimands (more used in epidemiology) cover:

- Risk ratio (RR): $\frac{\mathbb{E}[Y(1)]}{\mathbb{E}[Y(0)]}$
- Odd ratio (OR) for binary outcome: $\left(\frac{\mathbb{P}[Y(1)=1]}{\mathbb{P}[Y(1)=0]}\right) / \left(\frac{\mathbb{P}[Y(0)=1]}{\mathbb{P}[Y(0)=0]}\right)$

See (Colnet et al., 2023) for a review of the different estimands and the impact on generalization.

Identification: assumptions

- What can we learn from the data?
- Knowledge based
- Cannot be validated with data

Identification: proofs

Causal Estimator

Statistical inference

Session summary

Going further

Resources

- https://web.stanford.edu/~swager/stats361.pdf
- https://www.mixtapesessions.io/
- https://alejandroschuler.github.io/mci/
- https://theeffectbook.net/index.html

Bibliography

Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., & Syrgkanis, V. (2024). Applied causal inference powered by ML and AI. Arxiv Preprint Arxiv:2403.02467. https://causalml-book.org/

Colnet, B., Josse, J., Varoquaux, G., & Scornet, E. (2023). Risk ratio, odds ratio, risk difference… Which causal measure is easier to generalize?. Arxiv Preprint Arxiv:2303.16008.

Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. American Journal of Epidemiology, 183(8), 758–764.

Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J., & Silva, R. (2022). Causal machine learning: A survey and open problems. Arxiv Preprint Arxiv:2206.15475.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. The American Economic Review, 604–620.

Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. Roczniki Nauk Rolniczych, 10(1), 1–51.

Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. (1995). The well-built clinical question: a key to evidence-based decisions. ACP Journal Club, 123(3), A12–3.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5), 688–689.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111–133.

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. Neuroimage, 145, 166–179.