The incredible utility of causal models

UMC Methods Meeting

Oisín Ryan

Department of Methodology and Statistics
Utrecht University

November 7, 2022

Drug	No drug

	Drug	No drug
Male		
Female		

	Drug	No drug
Male	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Female	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)

	Drug	No drug
Male	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Female	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

	Drug	No drug
Male	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Female	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Should we prescribe the drug? `app.wooclap.com/CBRFDD`

	Drug	No drug
Low Blood Pressure	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High Blood Pressure	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

	Drug	No drug
Low Blood Pressure	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High Blood Pressure	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Should we prescribe the drug? `app.wooclap.com/CBRFDD`

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population

Lord's Paradox

The relationship between a categorical exposure and a continuous outcome is reversed when we condition on a third variable

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population

Lord's Paradox

The relationship between a categorical exposure and a continuous outcome is reversed when we condition on a third variable

Confusing, but **not a paradox**

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Berksons Paradox

Two phenomena which are statistically *independent* in the general population are statistically *dependent* in a sub-population

Lord's Paradox

The relationship between a categorical exposure and a continuous outcome is reversed when we condition on a third variable

Confusing, but **not a paradox**

You're asking a question that statistical inference alone is not equipped to answer

Estimand **Estimator** **Estimate**

Estimand **Estimator** **Estimate**

Statistical inference in a nutshell

Estimand

Estimator

1 **Prepare Chocolate Cake Batter**

Preheat oven to 350 degrees, and prepare Yo's Ultimate Chocolate Cake batter. Prepare your pans with parchment. Pour 2 ½ lbs into each 7" round pan, 1 ½ lbs into your 6" round pan, and divide the remaining batter evenly between your 5" round pans.

2 **Bake Cakes**

Bake your 7" round cakes for 50 minutes, your 6" round cake for 40 minutes, and your 5" round cakes for 30 minutes, or until a toothpick comes out clean. Set aside to cool completely in their pans on a wire rack.

3 **Prepare Fillings & Simple Syrup**

Prepare your dark chocolate ganache, Italian meringue buttercream, and simple syrup. Set aside until you're ready to decorate.

4 **Level Cakes**

Remove your cooled cakes from their pans and level them with a ruler and serrated knife.

Estimate

Statistical inference in a nutshell

Estimand

Estimator

1 Prepare Chocolate Cake Batter

Preheat oven to 350 degrees, and prepare Yo's Ultimate Chocolate Cake batter. Prepare your pans with parchment. Pour 2 ½ lbs into each 7" round pan, 1 ½ lbs into your 6" round pan, and divide the remaining batter evenly between your 5" round pans.

2 Bake Cakes

Bake your 7" round cakes for 50 minutes, your 6" round cake for 40 minutes, and your 5" round cakes for 30 minutes, or until a toothpick comes out clean. Set aside to cool completely in their pans on a wire rack.

3 Prepare Fillings & Simple Syrup

Prepare your dark chocolate ganache, Italian meringue buttercream, and simple syrup. Set aside until you're ready to decorate.

4 Level Cakes

Remove your cooled cakes from their pans and level them with a ruler and serrated knife.

Estimate

Credit to Peter Tennant @PWGTennant

	Drug	No drug
Male	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Female	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

	Drug	No drug
Male	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Female	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Conditional Probabilities:

$$P(R = r | D = d, S = s)$$

	Drug	No drug
Male	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Female	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Marginal Probabilities:

$$P(R = r | D = d)$$

Statistical inference in a nutshell

Estimand	Estimator	Estimate
$P(R = 1 \mid D = 1, S = 0)$	# Recovered takers Male / # Drug takers Male	$.93$
$P(R = 1 \mid D = 1)$	# Recovered drug takers / # Drug takers	$.78$

What's the paradox?

Two different sets of **estimands** yield two different sets of **estimates**

- No paradox there!

What's the paradox?

Two different sets of **estimands** yield two different sets of **estimates**

- No paradox there!

We are not interested in either of these estimands *for their own sake*

What's the paradox?

Two different sets of **estimands** yield two different sets of **estimates**

- No paradox there!

We are not interested in either of these estimands *for their own sake*

We are interested in a **causal effect**

- Does taking the drug cause recovery?
- **Causal Estimand**
- But we have no way of expressing this in the language of statistical inference

Causal Inference

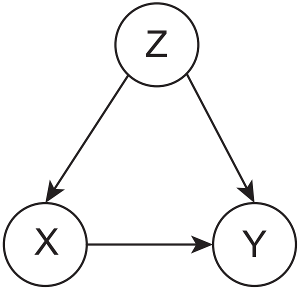
Causal Graphs

A causal graph is a diagram representing (our beliefs about) which variables share causal relations with each other

Causal Graphs

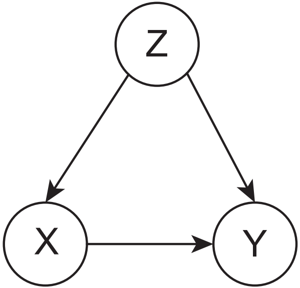
A causal graph is a diagram representing (our beliefs about) which variables share causal relations with each other

- The arrow X $\rightarrow$ Y represents our belief that X is a direct cause of Y
- We omit an arrow if expert knowledge tells us that one variable does not directly cause another. The *absence* of an arrow is a strong statement



A causal graph is a diagram representing (our beliefs about) which variables share causal relations with each other

- The arrow X $\rightarrow$ Y represents our belief that X is a direct cause of Y
- We omit an arrow if expert knowledge tells us that one variable does not directly cause another. The *absence* of an arrow is a strong statement



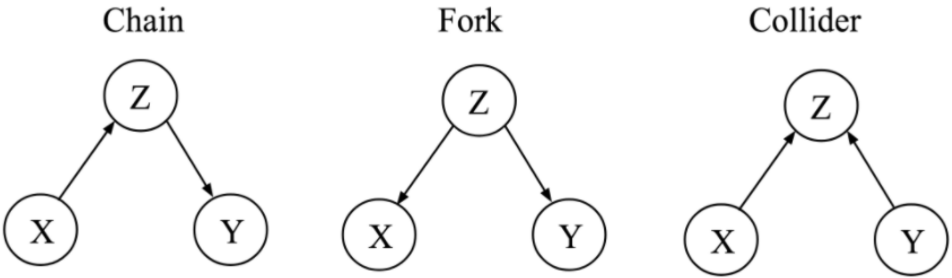
Directed Acyclic Graph (DAG) or Bayesian Network

Why Causal Models?

This machinery is useful for three important and closely related reasons:

1. Causal models map causal dependencies onto statistical dependencies
   - *Regardless* of distributions and functional forms

Graphs and statistical dependencies



Chain

X: Smoking
Z: Tar
Y: Cancer

$$X \not\perp\!\!\!\perp Y$$
$$X \perp\!\!\!\perp Y \mid Z$$

Fork

X: Storks
Z: Environment
Y: Babies

$$X \not\perp\!\!\!\perp Y$$
$$X \perp\!\!\!\perp Y \mid Z$$

Collider

X: Attractiveness
Z: Being Single
Y: Intelligence

$$X \perp\!\!\!\perp Y$$
$$X \not\perp\!\!\!\perp Y \mid Z$$

Why Causal Models?

This machinery is useful for three important and closely related reasons:

1. Causal models map causal dependencies onto statistical dependencies
    - *Regardless* of distributions and functional forms
2. Causal models allow us to define **causal effects** in the language of interventions and probabilities

Causal Effects

The **do-operator** $do(X = x)$ represents a "surgical intervention", forcing X to x.

We can use this to define our **causal estimand**

Causal Effect of Drug-Taking on Recovery:

$$CE = P[R \mid do(D = 1)] - P[R \mid do(D = 0)]$$

Causal Effects

The **do-operator** $do(X = x)$ represents a "surgical intervention", forcing X to x.

We can use this to define our **causal estimand**

Causal Effect of Drug-Taking on Recovery:

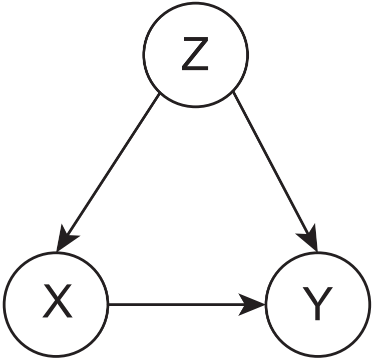
$$CE = P[R \mid do(D = 1)] - P[R \mid do(D = 0)]$$

Inference problem: "Seeing" is not always the same as "doing"

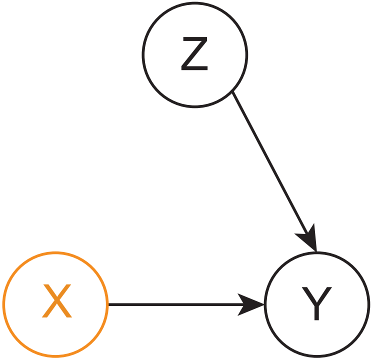
Observing $\neq$ Intervening:

$P[Y \mid X = x]$ is not **generally** the same as $P[Y \mid do(X = x)]$

Observing



Intervening



Why Causal Models?

This machinery is useful for three important and closely related reasons:

1. Causal models map causal dependencies onto statistical dependencies
    - *Regardless* of distributions and functional forms
2. Causal models allow us to define **causal effects** in the language of interventions and probabilities
3. Causal models tell us which when and how statistical estimands can act as causal estimands

	Drug	No drug
Male	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Female	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Statistical inference in a nutshell

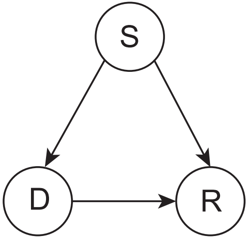
Statistical Estimand	Estimator	Estimate
$P(R=1 \mid D=1, S=1)$	# Recovered takers Male / # Drug takers Male	$.93$
$P(R=1 \mid D=1)$	# Recovered drug takers / # Drug takers	$.78$

Causal Inference in a nutshell

Causal Estimand

$P[R \mid do(D = 1)] -$
$P[R \mid do(D = 0)]$

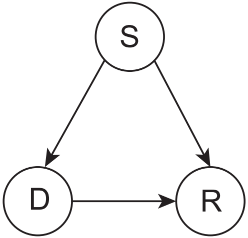
Causal Model



Causal Estimand

$$P[R \mid do(D=1)] - P[R \mid do(D=0)]$$

Causal Model



Statistical Estimand

$$P(R|D,S)$$

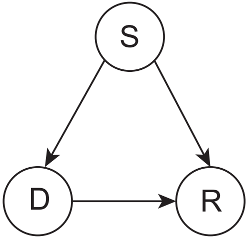
$$P(R|D)$$

Causal Estimand

Causal Model

Statistical Estimand

$$P[R \mid do(D=1)] - P[R \mid do(D=0)]$$

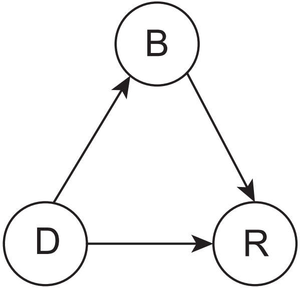


$$P(R|D,S)$$

$$\cancel{P(R|D)}$$

Post-Treatment Blood Pressure:

- Statistical information is exactly the same
- The drug works in part by decreasing blood pressure
- We should **not** condition on blood pressure



	Drug	No drug
Low Blood Pressure	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High Blood Pressure	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Example from Pearl, Glymour & Jewell, 2016

Causal Estimand

Causal Model

Statistical Estimand

Estimator

1. **Prepare Chocolate Cake Batter**
   Preheat oven to 350 degrees, and prepare Yo's Ultimate Chocolate Cake batter. Prepare your pans with parchment. Pour 2 ½ lbs into each 7" round pan, 1 ½ lbs into your 6" round pan, and divide the remaining batter evenly between your 5" round pans.
2. **Bake Cakes**
   Bake your 7" round cakes for 50 minutes, your 6" round cake for 40 minutes, and your 5" round cakes for 30 minutes, or until a toothpick comes out clean. Set aside to cool completely in their pans on a wire rack.
3. **Prepare Fillings & Simple Syrup**
   Prepare your dark chocolate ganache, Italian meringue buttercream, and simple syrup. Set aside until you're ready to decorate.
4. **Level Cakes**
   Remove your cooled cakes from their pans and level them with a ruler and serrated knife.

Estimate

Simpsons Paradox

Statistical phenomena where a relationship which is present when aggregating over the population may be reversed or absent when looking at sub-populations

Absolutely not a paradox.

- Confusion comes from a lack of clarity regarding our **causal estimand** and **causal model**

Statistical information *alone* cannot provide the answer

- Different DAGs can produce the exact same statistical dependencies in observational data

Causal models provide immediate conceptual clarity

- Miguel Hernan: Draw your assumptions before your conclusions!

Beyond Toy Examples

Causal Inference in Practice

Two broad frameworks for causal modeling

- Graphical Causal Models (Structural Causal Models)
- Potential Outcomes framework

Many tools available in these frameworks

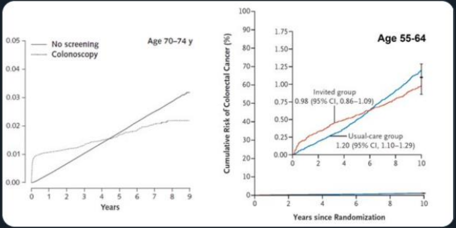
- Causal Mediation analysis
- Cyclic Causal Models
- Instrumental Variables / Mendellian Randomization
- Synthetic Control & Interrupted Time Series
- “Target Trial” emulation

Miguel Hernán @_MiguelHernan

1/
Six years ago we emulated a #TargetTrial of screening colonoscopy for colorectal #cancer using observational #Medicare data:
pubmed.ncbi.nlm.nih.gov/27669524/

Today we publish preliminary findings from a truly randomized trial:
nejm.org/doi/full/10.10...

Their main results side by side 👇



Annals of Int Med and 3 others

Statistical vs Causal Learning

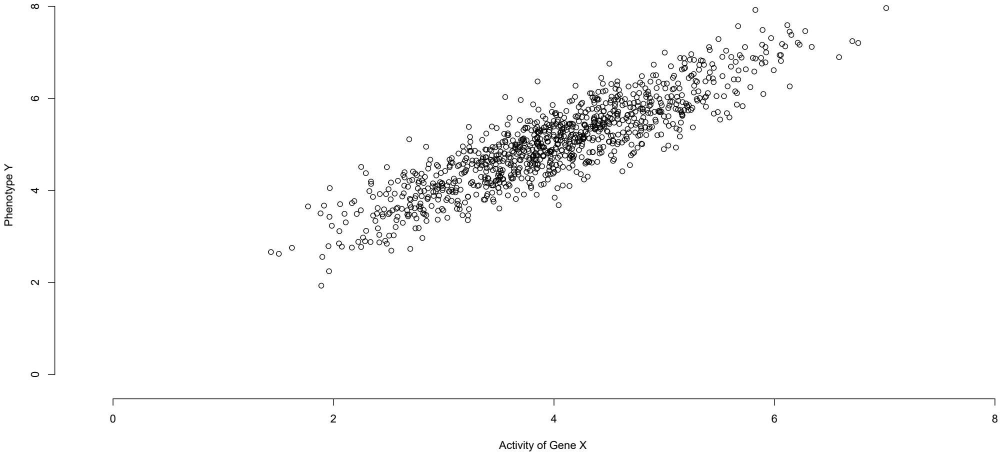
Statistical Learning / Machine Learning / "Data Science"

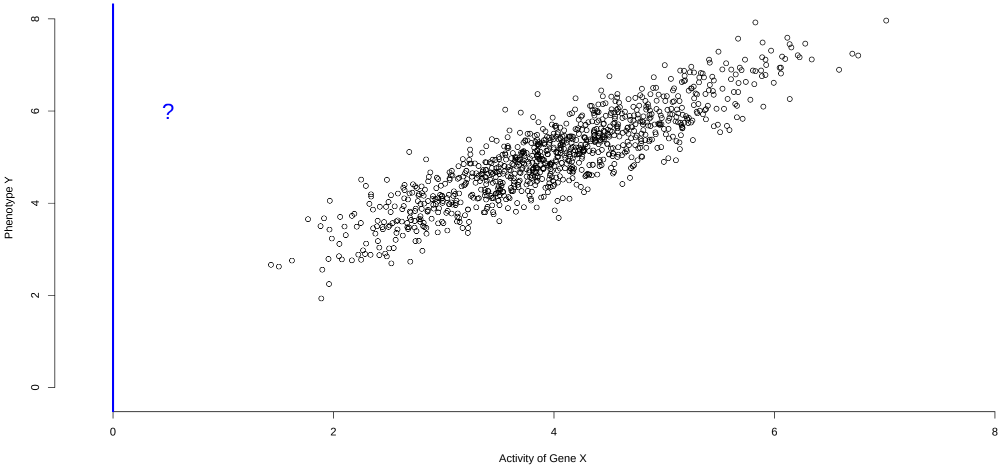
- Incredibly useful tools for learning *certain kinds* of representations of probability densities
- Usually: Those that minimize out-of-sample prediction error. But this is actually quite specific
- If I observe a new data point in *exactly the same circumstances* as I observed my training set, what types of values are likely to (co-)occur?

But is this what we actually want in most cases?

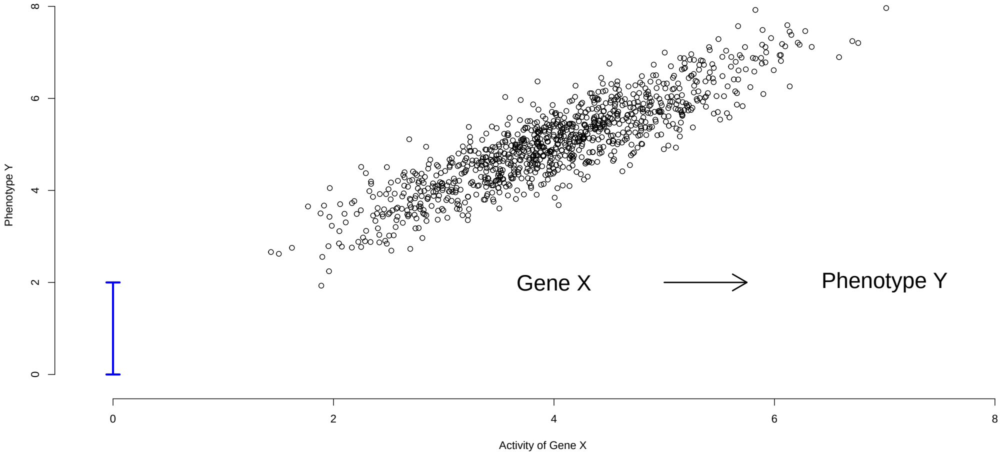
- Which treatment assignment would be best for this individual?
- What would happen if I change some policy, e.g., make more funding available for after-school programs?
- How do I make decisions which avoid gender bias?
- How do I make sure my ML model doesn't learn to classify data points based on some "undesirable" factor?

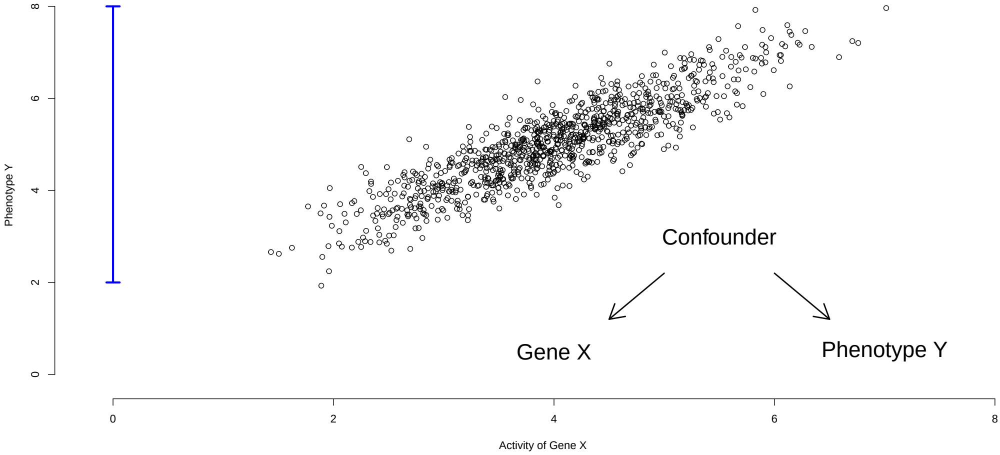
I would argue that answering these questions requires us to learn a *causal model*, not a purely statistical or predictive one


Phenotype Y
0
2
4
6
8
Activity of Gene X
0
2
4
6
8

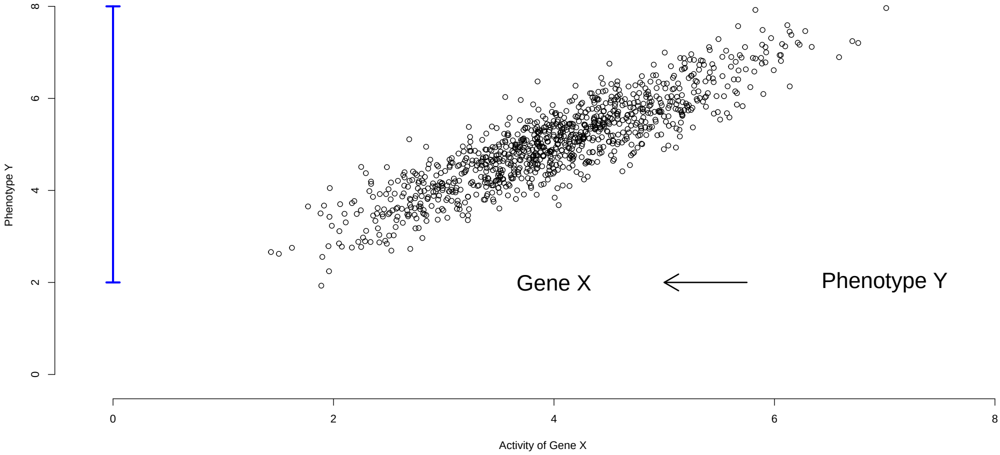


?
Phenotype Y
Activity of Gene X
0
2
4
6
8

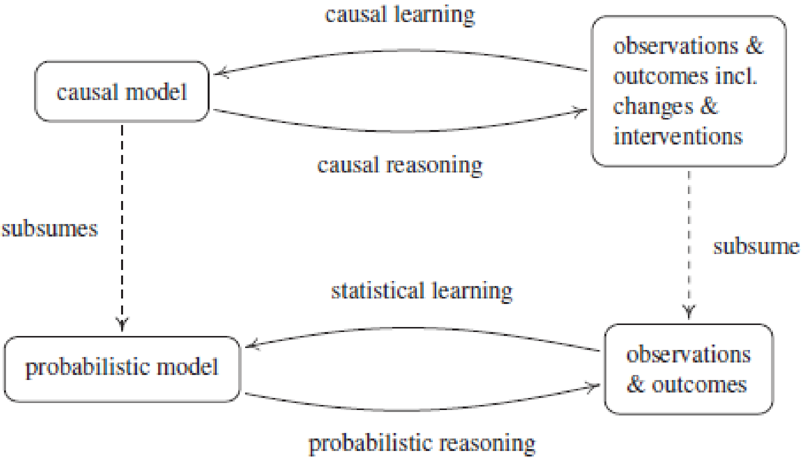

Gene X
Phenotype Y
Phenotype Y
Activity of Gene X
0
2
4
6
8



Phenotype Y
0
2
4
6
8
Confounder
Gene X
Phenotype Y
0
2
4
6
8
Activity of Gene X



Phenotype Y
0
2
4
6
8
Gene X
Phenotype Y
0
2
4
6
8
Activity of Gene X



Peters, Janzig, Scholkopf (2017) Elements of Causal Inference

INVITED PAPER

Toward Causal Representation Learning

This article reviews fundamental concepts of causal inference and relates them to crucial open problems of machine learning, including transfer learning and generalization, thereby assaying how causality can contribute to modern machine learning research.

By Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio

Conclusions

Inappropriate reliance on (advanced) statistical modeling or statistIcal learning techniques, with no clear link to causal estimands or models

- Paradoxes, confusion, poor decisions result

Causal modeling can be powerful in reshaping how we approach statistical modeling

- Judea Pearl, Don Rubin, Jamie Robins, Miguel Hernan, Angrist & Imbens
- Controlling for as many variables as possible is **an obviously terrible idea** when estimating causal effects
- Minimizing OOS PE / classification error does not in any obvious way licence causal inferences (counterfactual predictions)

Researchers make causal inferences based on observational data **all the time**

- Better to be explicit and open about this so we can move forward

Thanks!

(o.ryan@uu.nl | oisinryan.org)

Shameless plug

- Statistical network analysis as ad-hoc causal discovery (e.g. Ryan, Bringmann, Schuurman, 2022)
- Causal models of dynamical systems from cross-sectional data (Ryan* & Dablander*, pre-print 2022)
- Estimating the effect of a hypothetical language training program on study success using administrative data. A pre-registered causal inference approach (Spit, Andringa, Ryan. In-principle acceptance)
- Causal and statistical estimands (e.g. Haslbeck*, Ryan*, Dablander* 2021)
- Constructing theories (Haslbeck*, Ryan*, Robinaugh*, Waldorp, Borsboom, 2021)
- Synthetic Control and Interrupted Time Series with large-scale administrative data (in progress)
- Causal Discovery of cyclic models (in progress)
- Teaching materials and workshops