Causal frameworks and framing causal problems

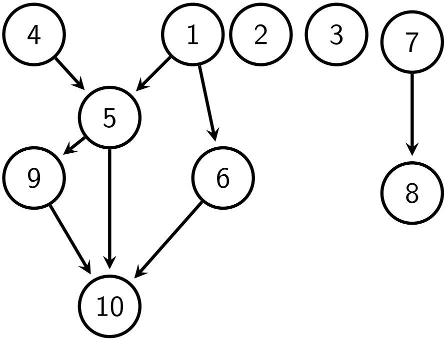
Gherardo Varando

IPL

04 May 2023

Summary of last session

- statistical models
- BN/DAG, d-separation and conditional independences



Causal models desiderata

- Represent data, similar to a statistical model
- Model what happen when changes/experiment/interventions
- Reason on and explore the causal relationships
- Represent causal relationships

Causal regression models

- given a regression model $Y = f(X, \varepsilon)$ we could interpret this as a causal model
- in the sense that we imagine the physical/true generating process to be modeled by this equation

Causal regression models

- given a regression model $Y = f(X, \varepsilon)$ we could interpret this as a causal model
- in the sense that we imagine the physical/true generating process to be modeled by this equation
- If we make *experiments* by changing the value of X we know that the associated value for Y is generated accordingly to $f(X, \varepsilon)$

Structural Causal Models

Definition [Peters et al., 2017]

A SCM over variables $X_1, \ldots, X_p$ with noise variables $\varepsilon_1, \ldots, \varepsilon_p$ is a collection of **structural assignments**:

$$X_i = f_i(X_{pa(i)}, \varepsilon_i)$$

where ε_i are assumed jointly independent and f_i are fixed deterministic functions.

- $X_{pa(i)}$ are called the parents or the **direct causes** of X_i
- we say X_i is a direct effect of its direct causes
- we assume the associated graph G to be a DAG

Bareinboim et al. [2022]

Structural Causal Models

Definition [Peters et al., 2017]

A SCM over variables $X_1, \dots, X_p$ with noise variables $\varepsilon_1, \dots, \varepsilon_p$ is a collection of **structural assignments**:

$$X_i = f_i(X_{pa(i)}, \varepsilon_i)$$

where ε_i are assumed jointly independent and f_i are fixed deterministic functions.

- $X_{pa(i)}$ are called the parents or the **direct causes** of X_i
- we say X_i is a direct effect of its direct causes
- we assume the associated graph G to be a DAG
- A SCM defines a unique distribution P over the variables $X_1, \dots, X_p$

Bareinboim et al. [2022]

Structural Causal Models

Definition [Peters et al., 2017]

A SCM over variables $X_1, \ldots, X_p$ with noise variables $\varepsilon_1, \ldots, \varepsilon_p$ is a collection of **structural assignments**:

$$X_i = f_i(X_{pa(i)}, \varepsilon_i)$$

where ε_i are assumed jointly independent and f_i are fixed deterministic functions.

- $X_{pa(i)}$ are called the parents or the **direct causes** of X_i
- we say X_i is a direct effect of its direct causes
- we assume the associated graph G to be a DAG
- A SCM defines a unique distribution P over the variables $X_1, \ldots, X_p$
- (G, P) is a Bayesian network

Bareinboim et al. [2022]

Interventions in SCM

- given a SCM we define an experiment, or intervention when we **replace one or several of the structural assignments** to obtain a new SCM

Interventions in SCM

- given a SCM we define an experiment, or intervention when we **replace one or several of the structural assignments** to obtain a new SCM
- the interventional distribution under this change is the new entailed probability distribution defined by the new SCM

Interventions in SCM

- given a SCM we define an experiment, or intervention when we **replace one or several of the structural assignments** to obtain a new SCM
- the interventional distribution under this change is the new entailed probability distribution defined by the new SCM
- e.g if changing one of the assignment (for X_k) we can write $P^{do(X_k=\tilde{f}_k(X_{\tilde{pa}(k)},\tilde{\varepsilon}))}$ the new interventional distribution
- sometimes you can find $P(\cdot|do())$

Interventions in SCM

- given a SCM we define an experiment, or intervention when we **replace one or several of the structural assignments** to obtain a new SCM
- the interventional distribution under this change is the new entailed probability distribution defined by the new SCM
- e.g if changing one of the assignment (for X_k) we can write $P^{do(X_k=\tilde{f}_k(X_{\tilde{pa}(k)},\tilde{\varepsilon}))}$ the new interventional distribution
- sometimes you can find $P(\cdot|do())$
- we say there is a **total causal effect** from X to Y in a SCM if and only if there exist a random variable $\tilde{N}$ such that

$$X \not\perp\!\!\!\perp Y \text{ in } P^{do(X=\tilde{N})}$$

- SCM defines an **observational distribution** plus **interventional distributions** for each of the possible interventions
- The set of observational and interventional distributions can also be represented with a Causal BN

- SCM defines an **observational distribution** plus **interventional distributions** for each of the possible interventions
- The set of observational and interventional distributions can also be represented with a Causal BN
- but SCMs allow an additional causal reasoning step: **counterfactuals**

Counterfactual statements

- Given some observed state/event,what would have happened if ... ?

Counterfactual statements

- Given some observed state/event,what would have happened if ... ?
- I have headache, I take aspirin and one hour later the pain disappear. Would the pain had disappeared if I had not taken aspirin?

Counterfactual statements

- Given some observed state/event,what would have happened if ... ?
- I have headache, I take aspirin and one hour later the pain disappear. Would the pain had disappeared if I had not taken aspirin?
- this is different from interventions since we are not just doing an experiment, but rather imagining interventions for alternative *worlds*, that is alternative ways the system could have evolved but conditioning on everything else being the same

Counterfactuals in SCM

- SCMs allow the computation of counterfactual probabilities
- Counterfactual statements can be seen as intervention/do-statements in a counterfactual SCM which is obtained from an initial SCM by replacing the noise distribution with $P(\varepsilon|X = x)$

Peters et al. [2017], Bareinboim et al. [2022]

Counterfactuals in SCM

- SCMs allow the computation of counterfactual probabilities
- Counterfactual statements can be seen as intervention/do-statements in a counterfactual SCM which is obtained from an initial SCM by replacing the noise distribution with $P(\varepsilon | X = x)$
- SCM can induce same observational and interventional distribution (Causal BN) but different counterfactuals

Peters et al. [2017], Bareinboim et al. [2022]

Example from Peters et al. [2017]

Example 6.19 Let $N_1, N_2 \sim \mathrm{Ber}(0.5)$, and $N_3 \sim \mathrm{U}(\{0,1,2\})$, such that the three variables are jointly independent. That is, N_1, N_2 have a Bernoulli distribution with parameter 0.5 and N_3 is uniformly distributed on $\{0,1,2\}$. We define two different SCMs. First consider $\mathfrak{C}_A$:

$$\begin{aligned}
X_1 &:= N_1 \\
X_2 &:= N_2 \\
X_3 &:= (1_{N_3>0} \cdot X_1 + 1_{N_3=0} \cdot X_2) \cdot 1_{X_1 \neq X_2} + N_3 \cdot 1_{X_1 = X_2}.
\end{aligned}$$

If X_1 and X_2 have different values, depending on N_3 we either choose $X_3 = X_1$ or $X_3 = X_2$. Otherwise $X_3 = N_3$. Now, $\mathfrak{C}_B$ differs from $\mathfrak{C}_A$ only in the latter case:

$$\begin{aligned}
X_1 &:= N_1 \\
X_2 &:= N_2 \\
X_3 &:= (1_{N_3>0} \cdot X_1 + 1_{N_3=0} \cdot X_2) \cdot 1_{X_1 \neq X_2} + (2 - N_3) \cdot 1_{X_1 = X_2}.
\end{aligned}$$

Both SCMs entail the same observational distribution; and for any possible intervention they entail the same intervention distributions, too.[9] But the two models differ in a counterfactual statement. Suppose, we have made an observation $(X_1, X_2, X_3) = (1,0,0)$ and we are interested in the counterfactual question "what would X_3 have been if X_1 had been 0?" From both SCMs, it follows that $N_3 = 0$, and thus the two SCMs $\mathfrak{C}_A$ and $\mathfrak{C}_B$ "predict" different values for X_3 under a counterfactual change of X_1 (namely 0 and 2, respectively). $\square$

Pearl Causal Hierarchy (PCH)

	Layer (Symbolic)	Typical Activity	Typical Question	Example	Machine Learning
$\mathcal{L}_1$	Associational $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?	Supervised / Unsupervised Learning
$\mathcal{L}_2$	Interventional $P(y\|do(x),c)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured?	Reinforcement Learning
$\mathcal{L}_3$	Counterfactual $P(y_x\|x',y')$	Imagining	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?	

Table 1.1: Pearl's Causal Hierarchy.

Bareinboim et al. [2022]

Potential outcomes - Counterfactual model

- consider a binary **treatment variable** A (1: treated, 0: untreated)
- and a binary **outcome** Y (1: death, 0: survival)
- A, Y are random variables that take possible different values for each individual

Hernán and Robins [2020], Wasserman [2004]

Potential outcomes - Counterfactual model

- consider a binary **treatment variable** A (1: treated, 0: untreated)
- and a binary **outcome** Y (1: death, 0: survival)
- A, Y are random variables that take possible different values for each individual
- denote with $Y^{a=1}$ (Y under treatment $a = 1$) the outcome variable that would have been observed under treatment $a = 1$, and similarly $Y^{a=0}$
- $Y^{a=1}$ and $Y^{a=0}$ are called **potential outcomes** or **counterfactual outcomes**

Hernán and Robins [2020], Wasserman [2004]

Potential outcomes - Counterfactual model

- consider a binary **treatment variable** A (1: treated, 0: untreated)
- and a binary **outcome** Y (1: death, 0: survival)
- A, Y are random variables that take possible different values for each individual
- denote with $Y^{a=1}$ (Y under treatment $a = 1$) the outcome variable that would have been observed under treatment $a = 1$, and similarly $Y^{a=0}$
- $Y^{a=1}$ and $Y^{a=0}$ are called **potential outcomes** or **counterfactual outcomes**
- for each individual, only one of the potential outcomes if actually observed/factual.

$$Y = Y^{a=A} \quad \text{(consistency equation)}$$

Hernán and Robins [2020], Wasserman [2004]

Causal effects

Definition

Individual causal effects The treatment A has a causal effect on an individual's outcome Y if $Y^{a=1} \neq Y^{a=0}$

Definition

Average causal effects An average causal effect of treatment A on outcome Y is present if

$$P(Y^{a=1} = 1) \neq P(Y^{a=0} = 1)$$

Other causal frameworks

- Granger causality

Other causal frameworks

- Granger causality
- Causal interpretation of dynamical system, ODE and SDE [Mooij et al., 2013, Peters et al., 2022, Hansen and Sokol, 2014]

Other causal frameworks

- Granger causality
- Causal interpretation of dynamical system, ODE and SDE [Mooij et al., 2013, Peters et al., 2022, Hansen and Sokol, 2014]
- Cyclic causal models [Hyttinen et al., 2012, Bongers et al., 2016]

Other causal frameworks

- Granger causality
- Causal interpretation of dynamical system, ODE and SDE [Mooij et al., 2013, Peters et al., 2022, Hansen and Sokol, 2014]
- Cyclic causal models [Hyttinen et al., 2012, Bongers et al., 2016]
- Equilibrium models [Young et al., 2019, Varando and Hansen, 2020]

Other causal frameworks

- Granger causality
- Causal interpretation of dynamical system, ODE and SDE [Mooij et al., 2013, Peters et al., 2022, Hansen and Sokol, 2014]
- Cyclic causal models [Hyttinen et al., 2012, Bongers et al., 2016]
- Equilibrium models [Young et al., 2019, Varando and Hansen, 2020]
- Dynamic Causal Models (DCM) [Friston et al., 2003]

Other causal frameworks

- Granger causality
- Causal interpretation of dynamical system, ODE and SDE [Mooij et al., 2013, Peters et al., 2022, Hansen and Sokol, 2014]
- Cyclic causal models [Hyttinen et al., 2012, Bongers et al., 2016]
- Equilibrium models [Young et al., 2019, Varando and Hansen, 2020]
- Dynamic Causal Models (DCM) [Friston et al., 2003]
- Local Independence Graphs [Didelez, 2006, Mogensen and Hansen, 2020]

Causal problems

- Discovering causal relationships (causal discovery)

Causal problems

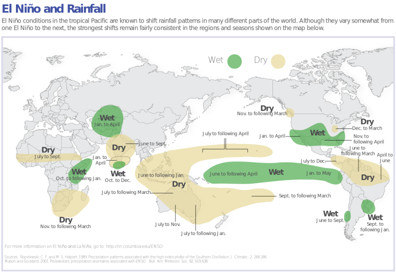
- Discovering causal relationships (causal discovery)
- Estimating causal effects (causal inference)

Causal problems

- Discovering causal relationships (causal discovery)
- Estimating causal effects (causal inference)
- Exploiting causal information in ML/prediction/forecasting/RL

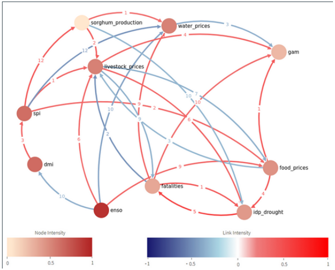
Causal Discovery

- Which are the regions most affected by ENSO?



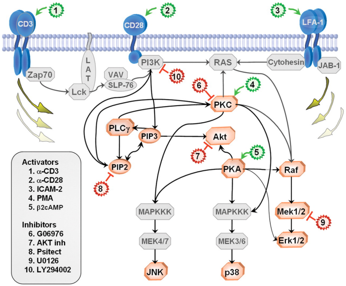
Causal Discovery

- Which are the regions most affected by ENSO?
- Which are the causal drivers and the causal relationships involving food insecurity?



Causal Discovery

- Which are the regions most affected by ENSO?
- Which are the causal drivers and the causal relationships involving food insecurity?
- Retrieve the reaction network between genes and proteins from single-cell data



Causal Discovery

- Learn the structure/graph of the causal relationships between variables

Causal Discovery

- Learn the structure/graph of the causal relationships between variables
- If only two variables are considered: bivariate causal discovery

Causal Discovery

- Learn the structure/graph of the causal relationships between variables
- If only two variables are considered: bivariate causal discovery
- it always require assumptions

Causal Discovery

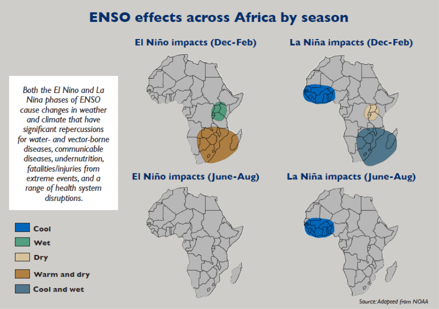
- Learn the structure/graph of the causal relationships between variables
- If only two variables are considered: bivariate causal discovery
- it always require assumptions
- can use only observational data, or even interventions

Causal Discovery

- Learn the structure/graph of the causal relationships between variables
- If only two variables are considered: bivariate causal discovery
- it always require assumptions
- can use only observational data, or even interventions
- usually the output is not a complete description of the causal relationships (depends on the assumptions, the type of data etc..)

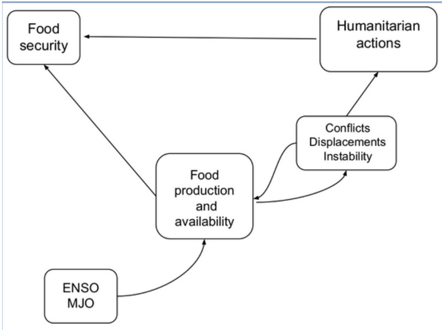
Causal Inference

- Estimate the strength of the effect of ENSO on vegetation greenness



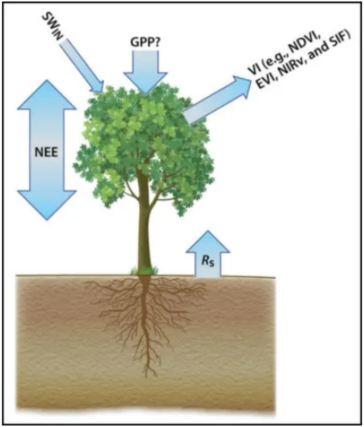
Causal Inference

- Estimate the strength of the effect of ENSO on vegetation greenness
- How effective are humanitarian actions to fight food insecurity?



Causal Inference

- Estimate the strength of the effect of ENSO on vegetation greenness
- How effective are humanitarian actions to fight food insecurity?
- What is the effect of radiation on NEE/GPP ? Or the effect of temperature on RECO?



Causal Inference

- Estimate average treatment effect (ATE) or conditional average treatment effect (CATE)

Causal Inference

- Estimate average treatment effect (ATE) or conditional average treatment effect (CATE)
- Usually provide confidence intervals and/or test hypothesis

Causal Inference

- Estimate average treatment effect (ATE) or conditional average treatment effect (CATE)
- Usually provide confidence intervals and/or test hypothesis
- Consider possible biases: confounding, selection biases, estimation biases

Causal ML

- Spurious relationships
- Causal representation learning
- Causal feature extraction/ feature selection
- robustness to intervention
- causal XAI
- Causal RL

´

Kaddour et al. [2022]

Elias Bareinboim, Juan D Correa, Duligur Ibeling, and Thomas Icard. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: the works of judea pearl*, pages 507–556. 2022.

Stephan Bongers, Jonas Peters, Bernhard Schölkopf, and Joris M Mooij. Theoretical aspects of cyclic structural causal models. *arXiv preprint arXiv:1611.06221*, 2016.

V Didelez. Asymmetric separation for local independence graphs. In *23rd Annual Conference on Uncertainty in Artifical Intelligence*, 2006.

Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302, 2003.

Niels Hansen and Alexander Sokol. Causal interpretation of stochastic differential equations. 2014.

MA Hernán and JM Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

Antti Hyttinen, Frederick Eberhardt, and Patrik O Hoyer. Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, 13(1):3387–3439, 2012.

Jean Kaddour, Aengus Lynch, Qi Liu, Matt J Kusner, and Ricardo Silva. Causal machine learning: A survey and open problems. *arXiv preprint arXiv:2206.15475*, 2022.

Søren Wengel Mogensen and Niels Richard Hansen. Markov equivalence of marginalized local independence graphs. 2020.

Joris M Mooij, Dominik Janzing, and Bernhard Schölkopf. From ordinary differential equations to structural causal models: the deterministic case. *arXiv preprint arXiv:1304.7920*, 2013.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

Jonas Peters, Stefan Bauer, and Niklas Pfister. Causal models for dynamical systems. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 671–690. 2022.

Gherardo Varando and Niels Richard Hansen. Graphical continuous lyapunov models. In *Conference on Uncertainty in Artificial Intelligence*, pages 989–998. PMLR, 2020.

Larry Wasserman. *All of statistics: a concise course in statistical inference*, volume 26. Springer, 2004.

William Chad Young, Ka Yee Yeung, and Adrian E Raftery. Identifying dynamical time series model parameters from equilibrium samples, with application to gene regulatory networks. *Statistical modelling*, 19(4):444–465, 2019.