(Casual) Causality Course 2025
Session 1

Gherardo Varando

25 March 2025

(Casual) Causality Course 2025

Intructors

- Gherardo gherardo.varando@uv.es
- Emiliano emiliano.diaz@uv.es
- Vassilis vasileios.sitokonstantinou@uv.es

Schedule

- **week 1, Tuesday** Intro and causal inference (GV)
- **week 1, Thursday** Causal inference and robustness (GV)
- **week 2, Tuesday** Causal Discovery (ED)
- **week 2, Thursday** Causal Discovery (ED)
- **week 3** Intensive weeek with group projects!

Learning outcomes

- Understand the fundamental goals and problems of causal methods

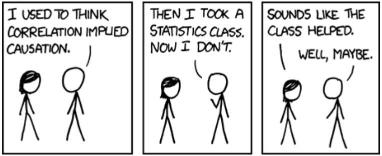


Figure: xkcd (CC BY-NC 2.5) https://xkcd.com/552

Learning outcomes

- Understand the fundamental goals and problems of causal methods
- Familiarize with the vocabulary, definitions and basic concepts of causality

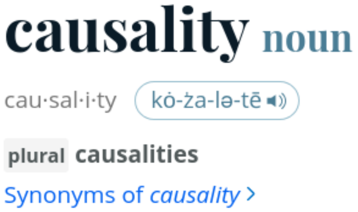


Figure: https://www.merriam-webster.com/dictionary/causality

Learning outcomes

- Understand the fundamental goals and problems of causal methods
- Familiarize with the vocabulary, definitions and basic concepts of causality
- Understand the fundamentals behind basic methodologies in causal inference and causal discovery

$$
\begin{aligned}
Y &= \theta(X) \cdot T + g(X, W) + \epsilon \qquad & \mathbb{E}[\epsilon | X, W] = 0 \\
T &= f(X, W) + \eta & \mathbb{E}[\eta \mid X, W] = 0 \\
& \mathbb{E}[\eta \cdot \epsilon | X, W] = 0
\end{aligned}
$$

Figure: https://econml.azurewebsites.net

Learning outcomes

- Understand the fundamental goals and problems of causal methods
- Familiarize with the vocabulary, definitions and basic concepts of causality
- Understand the fundamentals behind basic methodologies in causal inference and causal discovery
- How causal methods and tools are relevant in ML?

Figure: From Kaddour et al. [2022]

Content week 1

- **Session 1** Tue 25/03
  - Part I **Intro to causality and causal methods**
  - Part II **Basics of causal inference**

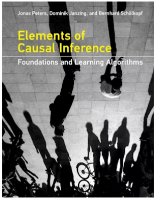
- **Session 2** Thu 27/03
  - Part I Causal inference methods
  - Part II Robustness to interventions

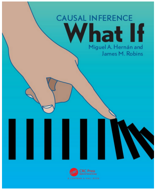
Content week 1

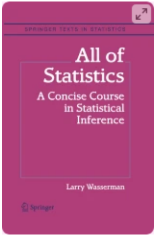
- **Session 1** Tue 25/03
  - Part I **Intro to causality and causal methods**
    motivation, causal questions, what is causality?
    experiments, interventions and counterfactuals
    structural causal models and graphs
  - Part II **Basics of causal inference**
    causal effect, randomized experiments
    observational studies, identifiability conditions
    graphical representation, confounding, selection bias
    random variability and measurement error
- **Session 2** Thu 27/03
  - Part I Causal inference methods
  - Part II Robustness to interventions

Basic references

- Elements of causal inference [Peters et al., 2017] [EC]
- Causal Inference: What If [Hernan and Robins, 2025] [Wif]
- All of Statistics [Wasserman, 2013] [AoS]







Felix qui potuit rerum cognoscere causas . . .

Publius Vergilius Maro

Georgica Book Two

Figure: By Michael van der Gucht - Public Domain

Felix qui potuit rerum cognoscere causas . . .

Publius Vergilius Maro

Georgica Book Two

Contents

Figure: By Michael van der Gucht - Public Domain

What is causality?

- Causality in Law
  https://en.wikipedia.org/wiki/Causation_(law)
  Causation—Low of Tort playlist on youtube, first 3 videos
- Causality in Physics
  https://en.wikipedia.org/wiki/Causality_(physics)
  https://www.youtube.com/watch?v=eG_eHDDMgCs
  Rovelli [2022]

Figure: Karditsa Thinker at the National Archaeological Museum, Athens

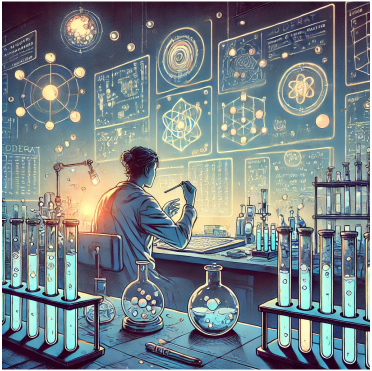
What is causality?

- Causality in Law
  https://en.wikipedia.org/wiki/Causation_(law)
  Causation—Low of Tort playlist on youtube, first 3 videos
- Causality in Physics
  https://en.wikipedia.org/wiki/Causality_(physics)
  https://www.youtube.com/watch?v=eG_eHDDMgCs
  Rovelli [2022]

Figure: Karditsa Thinker at the National Archaeological Museum, Athens

- Work in groups and briefly discuss causality in physics or law **[15 min]**, choose a speaker
- The two speakers present the main ideas **[5 min]** (whiteboard, slides)

Non-causal questions … still important!!



Non-causal questions ... still important!!

- Estimate the tree diameters or age in a forest [West, 2021]

Figure: From https://theforestguild.com/estimating-the-age-of-trees/

Non-causal questions ... still important!!

- Estimate the tree diameters or age in a forest [West, 2021]
- Counts the number of trees in the desert [unexpectedly high Brandt et al., 2020]

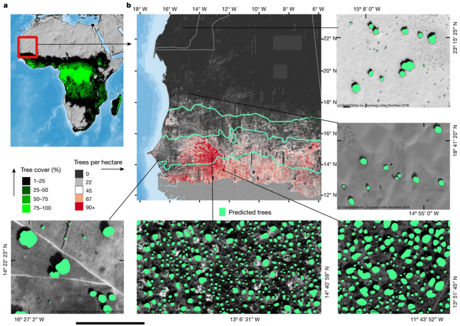


Figure: From Brandt et al. [2020] paper

Non-causal questions ... still important!!

- Estimate the tree diameters or age in a forest [West, 2021]
- Counts the number of trees in the desert [unexpectedly high Brandt et al., 2020]
- Cloud detection [Aybar et al., 2024]

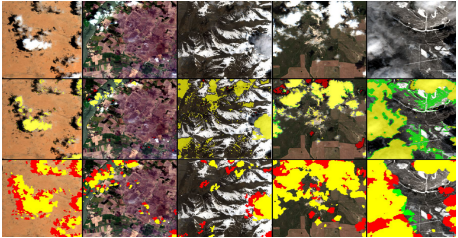


Figure: From Aybar et al. [2024] poster

Non-causal questions ... still important!!

- Estimate the tree diameters or age in a forest [West, 2021]
- Counts the number of trees in the desert [unexpectedly high Brandt et al., 2020]
- Cloud detection [Aybar et al., 2024]
- Describing patterns of wood density [Yang et al., 2024]

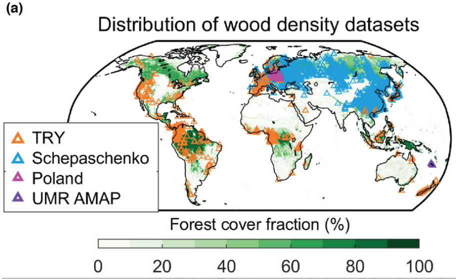
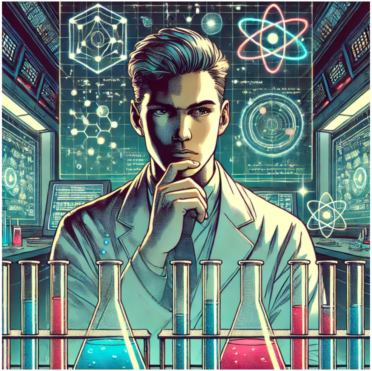


Figure: From Yang et al. [2024] paper

Causal questions



Causal questions

- How droughts affect yields

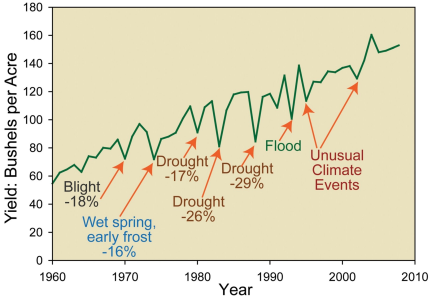


Figure: credit USDA FAS, from ESA Website (ESA Standard Licence)

Causal questions

- How droughts affect yields
- Evaluate agricultural practice [Tsoumas et al., 2023, Giannarakis et al., 2022]

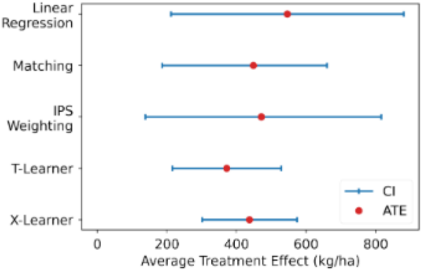


Figure: From Tsoumas et al. [2023]

Causal questions

- How droughts affect yields
- Evaluate agricultural practice [Tsoumas et al., 2023, Giannarakis et al., 2022]
- How does armed conflict influence tropical forest loss? [Christiansen et al., 2022]

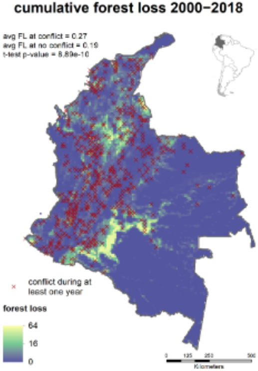


Figure: From Christiansen et al. [2022]

Causal questions

- How droughts affect yields
- Evaluate agricultural practice [Tsoumas et al., 2023, Giannarakis et al., 2022]
- How does armed conflict influence tropical forest loss? [Christiansen et al., 2022]
- Estimate the effect of ENSO on vegetation [Le, 2023]

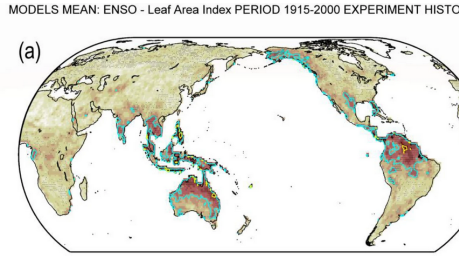


Figure: From Le [2023]

Probabilistic causation

[Wasserman, 2013, AoS, Chapter 16]

Roughly speaking, the statement "X causes Y" means that changing the value of X will change the distribution of Y.

Probabilistic causation

[Wasserman, 2013, AoS, Chapter 16]

Roughly speaking, the statement "X causes Y" means that changing the value of X will change the distribution of Y.

- How to make the previous *definition* precise and operative?

Probabilistic causation

[Wasserman, 2013, AoS, Chapter 16]

Roughly speaking, the statement "X causes Y" means that changing the value of X will change the distribution of Y.

- How to make the previous *definition* precise and operative?
- What does it mean changing the value of X?

Probabilistic causation

[Wasserman, 2013, AoS, Chapter 16]

Roughly speaking, the statement "X causes Y" means that changing the value of X will change the distribution of Y.

- How to make the previous *definition* precise and operative?
- What does it mean changing the value of X?
- various solutions are possible, we will see two possibilities: **Structural Causal Models (SCM)** and the **Potential Outcome framework**

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**:

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

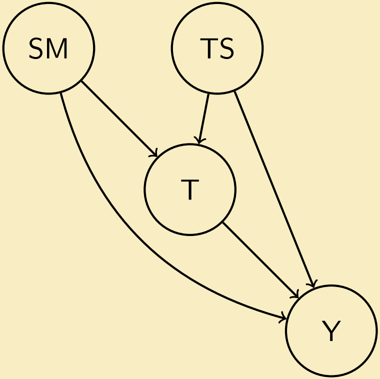
where ε_i are assumed jointly independent and f_i are fixed deterministic functions.

- $X_{pa(i)}$ 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

Example

A simple SCM for variables

- Y cotton yield
- SM soil moisture on sowing
- TS temperature on sowing
- T sowing is performed on optimal day (1) or not (0)



$$
\begin{aligned}
SM &= 0.3 + \epsilon_{SM} \\
TS &= 25 + \epsilon_{TS} \\
T &= \mathbf{1}_{>0}\left(1.2 + 0.8SM - 0.05TS + \epsilon_T > 0\right) \\
Y &= \max\{0,\ 2.5 + 1.5SM - 0.07TS + 2.0T + \epsilon_Y\}
\end{aligned}
$$

SCMs as statistical models

Every SCM define a unique distribution over the considered variables.

Figure: From https://www.quilting-in-america.com/cotton-cultivation-process.html

SCMs as statistical models

Every SCM define a unique distribution over the considered variables.

- How? ...

Figure: From https://www.quilting-in-america.com/cotton-cultivation-process.html

SCMs as statistical models

Every SCM define a unique distribution over the considered variables.

- How? ...
- We can sample from the SCM following **a topological order of the DAG**

Figure: From https://www.quilting-in-america.com/cotton-cultivation-process.html

SCMs as statistical models

Every SCM define a unique distribution over the considered variables.

- How? ...
- We can sample from the SCM following **a topological order of the DAG**
- If we can generate observations, we implicitly defined a distribution

Figure: From https://www.quilting-in-america.com/cotton-cultivation-process.html

SCMs can model interventions or experiments

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

Figure: From https://www.cottonfarming.com/feature-story/field-day-highlights/

SCMs can model interventions or experiments

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

Figure: From https://www.cottonfarming.com/feature-story/field-day-highlights/

SCMs can model interventions or experiments

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
- We indicate probabilities under interventions with e.g. $P(Y|\,\mathrm{do}(T=1))$

Figure: From https://www.cottonfarming.com/feature-story/field-day-highlights/

Definition (Total causal effect Peters et al. [2017])

We say there is a **total causal effect** from T to Y if (and only if) there exist $\tilde{N}$ such that

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

- How we intuitive interpret the definition ?

Definition (Total causal effect Peters et al. [2017])

We say there is a **total causal effect** from T to Y if (and only if) there exist $\tilde{N}$ such that

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

- How we intuitive interpret the definition ?
- If there is no direct path between T and Y then there is no total causal effect

Definition (Total causal effect Peters et al. [2017])

We say there is a **total causal effect** from T to Y if (and only if) there exist $\tilde{N}$ such that

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

- How we intuitive interpret the definition ?
- If there is no direct path between T and Y then there is no total causal effect
- Sometimes there is a direct path but no total causal effect

With SCM we can reason about counterfactuals

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

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

Figure: From https://www.nytimes.com/2011/07/12/us/12drought.html

With SCM we can reason about counterfactuals

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

- Given some observed state/event,what would have happened if ... ?
- 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 $P(\varepsilon|X = x)$

Figure: From https://www.nytimes.com/2011/07/12/us/12drought.html

Counterfactuals in SCM

Counterfactuals are 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

- SCMs allow the computation of counterfactual probabilities
- SCM can induce same observational and interventional distribution (Causal BN) but different counterfactuals (see example in Peters et al. [2017])

Pearl Causal Hierarchy (PCH)

From Bareinboim et al. [2022]

	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.

On the concept of *Harm*

- Sarvet and Stensrud [2023] interventionalist vs contrefactual
- Mueller and Pearl [2024] response

Personalized decision making

- Mueller and Pearl [2023]
- Tian and Pearl [2000]

Potential outcomes - Counterfactual model

Hernan and Robins [2025]
Wasserman [2013]

- consider a binary **treatment variable** A (1: forest management practice (thinning, controlled burns,...) , 0: wild/uncontrolled forest)
- and a binary **outcome** Y (1: burned area, 0: not burned)
- A, Y are random variables that take possible different values for each individual

Figure: From https://www.kunc.org/2024-03-15/long-term-study-finds-combination-of-prescribed-burns

Potential outcomes - Counterfactual model

Hernan and Robins [2025]
Wasserman [2013]

- denote with the outcome variable that would have been observed under treatment $a = 1$, and similarly $Y^{a=0}$

Figure: From https://www.kunc.org/2024-03-15/long-term-study-finds-combination-of-prescribed-burns

Potential outcomes - Counterfactual model

Hernan and Robins [2025]
Wasserman [2013]

- denote with 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**

Figure: From https://www.kunc.org/2024-03-15/long-term-study-finds-combination-of-prescribed-burns

Potential outcomes - Counterfactual model

Hernan and Robins [2025]
Wasserman [2013]

- denote with 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 is actually observed/factual.

Figure: From https://www.kunc.org/2024-03-15/long-term-study-finds-combination-of-prescribed-burns

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

Causal effects

Definition

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

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

or equivalently (for binary outcomes)

$$\mathbb{E}[Y^{a=1}] \neq \mathbb{E}[Y^{a=0}]$$

Causal effects

Definition

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

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

or equivalently (for binary outcomes)

$$\mathbb{E}[Y^{a=1}] \neq \mathbb{E}[Y^{a=0}]$$

- in practice we need to **measure** causal effects

Causal effects

Definition

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

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

or equivalently (for binary outcomes)

$$\mathbb{E}[Y^{a=1}] \neq \mathbb{E}[Y^{a=0}]$$

- in practice we need to **measure** causal effects
- causal risk difference $P(Y^{a=1} = 1) - P(Y^{a=0} = 1)$

Causal effects

Definition

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

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

or equivalently (for binary outcomes)

$$\mathbb{E}[Y^{a=1}] \neq \mathbb{E}[Y^{a=0}]$$

- in practice we need to **measure** causal effects
- causal risk difference $P(Y^{a=1} = 1) - P(Y^{a=0} = 1)$
- causal risk ratio $\frac{P(Y^{a=1}=1)}{P(Y^{a=0}=1)}$

Causal effects

Definition

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

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

or equivalently (for binary outcomes)

$$\mathbb{E}[Y^{a=1}] \neq \mathbb{E}[Y^{a=0}]$$

- in practice we need to **measure** causal effects
- causal risk difference $P(Y^{a=1} = 1) - P(Y^{a=0} = 1)$
- causal risk ratio $\frac{P(Y^{a=1}=1)}{P(Y^{a=0}=1)}$
- causal odds ratio $\frac{P(Y^{a=1}=1)/P(Y^{a=1}=0)}{P(Y^{a=0}=1)/P(Y^{a=0}=0)}$

Randomized experiments

- We collect data following a randomized control study: for each individual (forest unit/patch) we flip a coin and we assign the treatment variable to be $a = 1$ if heads and $a = 0$ if tails.

Randomized experiments

- We collect data following a randomized control study: for each individual (forest unit/patch) we flip a coin and we assign the treatment variable to be $a = 1$ if heads and $a = 0$ if tails.
- We then collect the outcome variable Y (e.g. burned or not after 1 year) for all individuals in the study

Randomized experiments

- assume no problem with the study, everybody is following instruction and there are no measurements problems (*ideal randomized experiment*)

Randomized experiments

- assume no problem with the study, everybody is following instruction and there are no measurements problems (*ideal randomized experiment*)
- can we say something about the causal effect of A on Y ?

Randomized experiments

- assume no problem with the study, everybody is following instruction and there are no measurements problems (*ideal randomized experiment*)
- can we say something about the causal effect of A on Y ?
- yes! we can compute the average causal effect ... formally because there is **exchangeability** between the treated ($A = 1$) and untreated ($A = 0$) groups

New problem: the firefighters do not like your randomized study they say that it is too dangerous not to manage some patches at all, and that some areas have a too high fire risk to be left completely untreated

Conditional randomized experiments

- Assume we have now a covariate L, measured before treatment was assigned (e.g. risk of fire: low, medium, high)

Conditional randomized experiments

- Assume we have now a covariate L, measured before treatment was assigned (e.g. risk of fire: low, medium, high)
- We divide the population into strata based on the levels of L and we perform randomization with different probabilities in each stratum (e.g. 0.8 for low fire risk, 0.5 for medium fire risk, and 0.2 for high fire risk)

Conditional randomized experiments

- Assume we have now a covariate L, measured before treatment was assigned (e.g. risk of fire: low, medium, high)
- We divide the population into strata based on the levels of L and we perform randomization with different probabilities in each stratum (e.g. 0.8 for low fire risk, 0.5 for medium fire risk, and 0.2 for high fire risk)
- In each stratum, we have exchangeability and we can compute average treatment effects

Conditional randomized experiments

- Assume we have now a covariate L, measured before treatment was assigned (e.g. risk of fire: low, medium, high)
- We divide the population into strata based on the levels of L and we perform randomization with different probabilities in each stratum (e.g. 0.8 for low fire risk, 0.5 for medium fire risk, and 0.2 for high fire risk)
- In each stratum, we have exchangeability and we can compute average treatment effects
- this is called *stratification*

Conditional randomized experiments

- Assume we have now a covariate L, measured before treatment was assigned (e.g. risk of fire: low, medium, high)
- We divide the population into strata based on the levels of L and we perform randomization with different probabilities in each stratum (e.g. 0.8 for low fire risk, 0.5 for medium fire risk, and 0.2 for high fire risk)
- In each stratum, we have exchangeability and we can compute average treatment effects
- this is called *stratification*
- moreover we say that this procedure ensure **conditional exchangeability** $Y^a \perp\!\!\!\perp A|L$

Computing ATE from conditionally randomized data

- From the data collected with a conditionally randomized experiment we can compute the ATE in all population.

Computing ATE from conditionally randomized data

- From the data collected with a conditionally randomized experiment we can compute the ATE in all population.
- **Standardization** consists in computing the marginal counterfactual risk as the weighted average of the stratum-specific risk.

$$P(Y^a = 1) = \sum P(Y^a = 1 | L = l) P(L = l)$$

Computing ATE from conditionally randomized data

- From the data collected with a conditionally randomized experiment we can compute the ATE in all population.
- **Standardization** consists in computing the marginal counterfactual risk as the weighted average of the stratum-specific risk.

$$P(Y^a = 1) = \sum P(Y^a = 1 | L = l) P(L = l)$$

- **Inverse Probability Weighting** is an alternative, but equivalent, procedure to compute $P(Y^a = 1)$ by weighting each individual sample by $w_l = 1/P(A = a | L = l)$ and then we compute $P(Y^a = 1) = \sum w_l P(Y | A = a, L = l)$

Observational studies

Sometimes is unethical, too expensive or simply impossible to conduct experiments, so we are left only with observational data, what can we do? (e.g. we just collect historical data on forest management practice and burned areas)

Observational studies

Sometimes is unethical, too expensive or simply impossible to conduct experiments, so we are left only with observational data, what can we do? (e.g. we just collect historical data on forest management practice and burned areas)

- Problems: confounding variables (e.g. forest are controlled more when we expect to have high risk of fire, patients are treated with more invasive/effective treatments when doctors thinks the case is more serious, ...)

Observational studies

Sometimes is unethical, too expensive or simply impossible to conduct experiments, so we are left only with observational data, what can we do? (e.g. we just collect historical data on forest management practice and burned areas)

- Problems: confounding variables (e.g. forest are controlled more when we expect to have high risk of fire, patients are treated with more invasive/effective treatments when doctors thinks the case is more serious, . . .)
- still under certain conditions and assumptions we can identify the causal effect

Observational studies

Sometimes is unethical, too expensive or simply impossible to conduct experiments, so we are left only with observational data, what can we do? (e.g. we just collect historical data on forest management practice and burned areas)

- Problems: confounding variables (e.g. forest are controlled more when we expect to have high risk of fire, patients are treated with more invasive/effective treatments when doctors thinks the case is more serious, ...)
- still under certain conditions and assumptions we can identify the causal effect
- this conditions *assure that the observational study can be used somehow as a randomized trial*

Identifiability conditions

1. **Exchangeability** the conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on measured covariates L

Identifiability conditions

1. **Exchangeability** the conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on measured covariates L
2. **Positivity** the probability of receiving every value of treatment conditional on L is greater than zero, i.e., positive

Identifiability conditions

1. **Exchangeability** the conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on measured covariates L
2. **Positivity** the probability of receiving every value of treatment conditional on L is greater than zero, i.e., positive
3. **Consistency** the values of treatment under comparison correspond to well-defined interventions that, in turn, correspond to the versions of treatment in the data

Identifiability conditions

1. **Exchangeability** the conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on measured covariates L
2. **Positivity** the probability of receiving every value of treatment conditional on L is greater than zero, i.e., positive
3. **Consistency** the values of treatment under comparison correspond to well-defined interventions that, in turn, correspond to the versions of treatment in the data

If we can assume this three conditions we can use the techniques such as IPW or standardization to compute ATE from observational data

Effect modification

- We say that V is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of V
- **Stratification** or **matching** can be used to identify effect modification

Effect modification

- We say that V is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of V
- **Stratification** or **matching** can be used to identify effect modification
- To construct our matched population we replaced the treated in the population by a subset of the treated in which the matching factor L had the same distribution as that in the untreated.

Effect modification

- We say that V is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of V
- **Stratification** or **matching** can be used to identify effect modification
- To construct our matched population we replaced the treated in the population by a subset of the treated in which the matching factor L had the same distribution as that in the untreated.
- they require exchangeability and positivity

Effect modification

- We say that V is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of V
- **Stratification** or **matching** can be used to identify effect modification
- To construct our matched population we replaced the treated in the population by a subset of the treated in which the matching factor L had the same distribution as that in the untreated.
- they require exchangeability and positivity
- Standardization (or IPW), stratification and matching measure different causal effects: Average effects in the entire population, conditional causal effects (startification) and usually causal effects in the treated and untreated for matching

Bibliography I

Cesar Aybar, Gonzalo Mateo-García, Giacomo Acciarini, Vít Růžička, Gabriele Meoni, Nicolas Longépé, and Luis Gómez-Chova. Onboard cloud detection and atmospheric correction with efficient deep learning models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:19518–19529, 2024. doi: 10.1109/JSTARS.2024.3480520.

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.

Martin Brandt, Compton J Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, et al. An unexpectedly large count of trees in the west african sahara and sahel. *Nature*, 587(7832):78–82, 2020.

Rune Christiansen, Matthias Baumann, Tobias Kuemmerle, Miguel D. Mahecha, and Jonas Peters. Toward causal inference for spatio-temporal data: Conflict and forest loss in colombia. *Journal of the American Statistical Association*, 117(538):591–601, 2022. doi: 10.1080/01621459.2021.2013241. URL https://doi.org/10.1080/01621459.2021.2013241.

Georgios Giannarakis, Vasileios Sitokonstantinou, Roxanne Suzette Lorilla, and Charalampos Kontoes. Towards assessing agricultural land suitability with causal machine learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1442–1452, June 2022.

M.A. Hernan and J.M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC Press, 2025. ISBN 9781420076165. URL https://books.google.es/books?id=_KnHIAAACAAJ.

Jean Kaddour, Aengus Lynch, Qi Liu, Matt J. Kusner, and Ricardo Silva. Causal machine learning: A survey and open problems, 2022. URL https://arxiv.org/abs/2206.15475.

Thanh Le. Increased impact of the el niño–southern oscillation on global vegetation under future warming environment. *Scientific Reports*, 13(1):14459, 2023.

Scott Mueller and Judea Pearl. Personalized decision making–a conceptual introduction. *Journal of Causal Inference*, 11(1):20220050, 2023.

Scott Mueller and Judea Pearl. Perspective on 'harm'in personalized medicine–an alternative perspective. Technical report, 2024. URL https://ftp.cs.ucla.edu/pub/stat_ser/r530.pdf.

Bibliography II

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

Carlo Rovelli. How causation is rooted into thermodynamics, 2022. URL https://arxiv.org/abs/2211.00888.

Aaron L Sarvet and Mats J Stensrud. Perspective on 'harm'in personalized medicine. *American Journal of Epidemiology*, page kwad162, 2023.

Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1):287–313, 2000.

Ilias Tsoumas, Georgios Giannarakis, Vasileios Sitokonstantinou, Alkiviadis Koukos, Dimitra Loka, Nikolaos Bartsotas, Charalampos Kontoes, and Ioannis Athanasiadis. Evaluating digital agriculture recommendations with causal inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14514–14522, 2023.

Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.

PW West. Tamm review: Sampling to estimate the frequency distribution of tree diameters or ages across large forest areas. *Forest Ecology and Management*, 488:118939, 2021.

Hui Yang, Siyuan Wang, Rackhun Son, Hoontaek Lee, Vitus Benson, Weijie Zhang, Yahai Zhang, Yuzhen Zhang, Jens Kattge, Gerhard Boenisch, Dmitry Schepaschenko, Zbigniew Karaszewski, Krzysztof Stereńczak, Álvaro Moreno-Martínez, Cristina Nabais, Philippe Birnbaum, Ghislain Vieilledent, Ulrich Weber, and Nuno Carvalhais. Global patterns of tree wood density. *Global Change Biology*, 30(3):e17224, 2024. doi: https://doi.org/10.1111/gcb.17224. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.17224. e17224 GCB-23-2908.R1.