(Casual) Causality Course 2025
Session 2

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!

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
  - Causal inference methods

Content week 1

- **Session 1** Tue 25/03

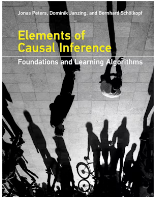
  Part I **Intro to causality and causal methods**
  motivation, causal questions, what is causality?
  experiments, interventions and counterfactuals
  structural causal models and graphs

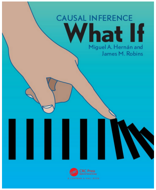
  Part II **Basics of causal inference**
  causal effect, randomized experiments
  observational studies, identifiability conditions
  graphical representation, confounding, selection bias
  random variability and measurement error

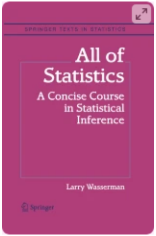
- **Session 2** Thu 27/03
  - Causal inference methods

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]







What is causality?

- Causality in Law
  https://en.wikipedia.org/wiki/Causation_(law)
  Causation—Low of Tort playlist on youtube, first 3 videos
- Rethinking Actual Causation in Tort Law
  https://harvardlawreview.org/print/vol-130/rethinking-actual-causation-in-tort-law/
- 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
- Rethinking Actual Causation in Tort Law
  https://harvardlawreview.org/print/vol-130/rethinking-actual-causation-in-tort-law/
- 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)

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

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
- 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
- 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
- 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

Computing interventional distributions in SCM

truncated factorization [Pearl, 1993], G-computation formula [Robins, 1986], manipulation theorem [Spirtes et al., 2000]

Given an SCM $\mathcal{C}$ and an intervened SCM $\tilde{\mathcal{C}}$, obtained from $\mathcal{C}$ by intervening on some X_k with $k \neq j$, we have that

$$P^{\tilde{\mathcal{C}}}(X_j | X_{pa(j)}) = P^{\mathcal{C}}(X_j | X_{pa(j)})$$

- Combining the above property and the assumption of SCM we can sometimes compute interventional distribution from observational quantities

Computing interventional distributions in SCM

truncated factorization [Pearl, 1993], G-computation formula [Robins, 1986], manipulation theorem [Spirtes et al., 2000]

Given an SCM $\mathcal{C}$ and an intervened SCM $\tilde{\mathcal{C}}$, obtained from $\mathcal{C}$ by intervening on some X_k with $k \neq j$, we have that

$$P^{\tilde{\mathcal{C}}}(X_j | X_{pa(j)}) = P^{\mathcal{C}}(X_j | X_{pa(j)})$$

- Combining the above property and the assumption of SCM we can sometimes compute interventional distribution from observational quantities
- Thus in practical terms we will be able sometimes to estimate interventional objects, such as treatment effects, from observational data alone

Computing interventional distributions in SCM

truncated factorization [Pearl, 1993], G-computation formula [Robins, 1986], manipulation theorem [Spirtes et al., 2000]

Given an SCM $\mathcal{C}$ and an intervened SCM $\tilde{\mathcal{C}}$, obtained from $\mathcal{C}$ by intervening on some X_k with $k \neq j$, we have that

$$P^{\tilde{\mathcal{C}}}(X_j | X_{pa(j)}) = P^{\mathcal{C}}(X_j | X_{pa(j)})$$

- Combining the above property and the assumption of SCM we can sometimes compute interventional distribution from observational quantities
- Thus in practical terms we will be able sometimes to estimate interventional objects, such as treatment effects, from observational data alone
- This requires the *knowledge of the causal graph*

Confounding and adjusting

- Consider an SCM $\mathcal{C}$, the causal effect from X to Y is called confounded if $P^{\mathcal{C},do(X=x)}(y) \neq P^{\mathcal{C}}(y)$

Confounding and adjusting

- Consider an SCM $\mathcal{C}$, the causal effect from X to Y is called confounded if $P^{\mathcal{C},do(X=x)}(y) \neq P^{\mathcal{C}}(y)$
- **Z** is called a valid adjustment set for X, Y if

$$P^{\mathcal{C},do(X=x)}(y) = \sum P^{\mathcal{C}}(Y|X, \mathbf{Z} = z)P^{\mathcal{C}}(\mathbf{Z} = \mathbf{z})$$

Confounding and adjusting

- Consider an SCM $\mathcal{C}$, the causal effect from X to Y is called confounded if $P^{\mathcal{C},do(X=x)}(y) \neq P^{\mathcal{C}}(y)$
- $\mathbf{Z}$ is called a valid adjustment set for X, Y if

$$P^{\mathcal{C},do(X=x)}(y) = \sum P^{\mathcal{C}}(Y|X, \mathbf{Z} = z)P^{\mathcal{C}}(\mathbf{Z} = \mathbf{z})$$

- Valid asjustment sets are:
  1. **parent adjustment** PA_X
  2. **backdoor criterion** Any $\mathbf{Z}$ such that i) contains no descendant of X and ii) blocks all backdoor paths $\rightarrow X$
  3. **towards necessity** ...

Confounding and adjusting

- Consider an SCM $\mathcal{C}$, the causal effect from X to Y is called confounded if $P^{\mathcal{C},do(X=x)}(y) \neq P^{\mathcal{C}}(y)$
- **Z** is called a valid adjustment set for X, Y if

$$P^{\mathcal{C},do(X=x)}(y) = \sum P^{\mathcal{C}}(Y|X, \mathbf{Z} = z)P^{\mathcal{C}}(\mathbf{Z} = \mathbf{z})$$

- Valid asjustment sets are:
  1. **parent adjustment** PA_X
  2. **backdoor criterion** Any **Z** such that i) contains no descendant of X and ii) blocks all backdoor paths $\rightarrow X$
  3. **towards necessity** ...
- Valid adjustment sets ensure conditional exchangeability, thus we can use standardization or stratification to compute average or conditional causal effect

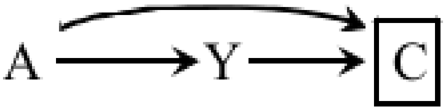
Confounding and adjusting

- Consider an SCM $\mathcal{C}$, the causal effect from X to Y is called confounded if $P^{\mathcal{C},do(X=x)}(y) \neq P^{\mathcal{C}}(y)$
- $\mathbf{Z}$ is called a valid adjustment set for X, Y if

$$P^{\mathcal{C},do(X=x)}(y) = \sum P^{\mathcal{C}}(Y|X, \mathbf{Z} = z) P^{\mathcal{C}}(\mathbf{Z} = \mathbf{z})$$

- Valid asjustment sets are:
  1. **parent adjustment** PA_X
  2. **backdoor criterion** Any $\mathbf{Z}$ such that i) contains no descendant of X and ii) blocks all backdoor paths $\rightarrow X$
  3. **towards necessity** ...
- Valid adjustment sets ensure conditional exchangeability, thus we can use standardization or stratification to compute average or conditional causal effect
- viceversa there are techniques that can handle confounding problems without relying on exchangeability: e.g. difference-in-differences, instrumental variables and the front door criterion

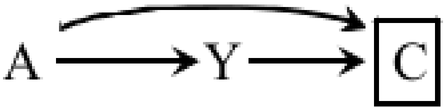
Selection bias



- If we condition on C there are two open paths between A and Y

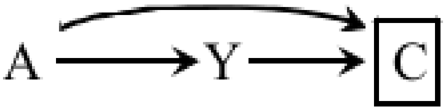
- If we condition on C there are two open paths between A and Y
- This can happen for example: differential loss to follow-up, missing data bias, nonresponse bias, healthy worker bias, self-selection bias, volunteer bias, and selection affected by treatment received before study entry

Selection bias



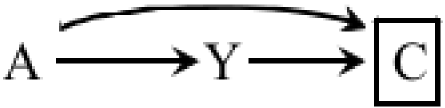
- If we condition on C there are two open paths between A and Y
- This can happen for example: differential loss to follow-up, missing data bias, nonresponse bias, healthy worker bias, self-selection bias, volunteer bias, and selection affected by treatment received before study entry
- Selection bias leads to a lack of exchangeability

Selection bias



- If we condition on C there are two open paths between A and Y
- This can happen for example: differential loss to follow-up, missing data bias, nonresponse bias, healthy worker bias, self-selection bias, volunteer bias, and selection affected by treatment received before study entry
- Selection bias leads to a lack of exchangeability
- IPW or stratification can be used to control for selection bias

Selection bias



- If we condition on C there are two open paths between A and Y
- This can happen for example: differential loss to follow-up, missing data bias, nonresponse bias, healthy worker bias, self-selection bias, volunteer bias, and selection affected by treatment received before study entry
- Selection bias leads to a lack of exchangeability
- IPW or stratification can be used to control for selection bias
- randomization does not protect from selection bias

Do-calculus

- an interventional distribtuion in an SCM is called identifiable if it can be computed from observational quantities and properties of the graph structure

Do-calculus

- an interventional distribtuion in an SCM is called identifiable if it can be computed from observational quantities and properties of the graph structure
- Pearl has developed the so-called **do-calculus** that consists in a set of three rules: Insertion/deletion of observations, Action/observation exchange, Insertion/deletion of actions

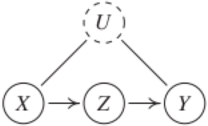
Do-calculus

- an interventional distribtuion in an SCM is called identifiable if it can be computed from observational quantities and properties of the graph structure
- Pearl has developed the so-called **do-calculus** that consists in a set of three rules: Insertion/deletion of observations, Action/observation exchange, Insertion/deletion of actions
- do-caluclus is complete, every identifiable interventional distribution can be obtained

Do-calculus

- an interventional distribtuion in an SCM is called identifiable if it can be computed from observational quantities and properties of the graph structure
- Pearl has developed the so-called **do-calculus** that consists in a set of three rules: Insertion/deletion of observations, Action/observation exchange, Insertion/deletion of actions
- do-caluclus is complete, every identifiable interventional distribution can be obtained
- one corollary of the do-calculus theorem is the **front-door adjustment**

Example 6.46 (Front-door adjustment) Let $\mathfrak{C}$ be an SCM with corresponding graph



If we do not observe U, we cannot apply the backdoor criterion. In fact, there is no valid adjustment set. But still, provided that $p^{\mathfrak{C}}(x,z) > 0$, the *do*-calculus provides us with

$$p^{\mathfrak{C};\, do(X:=x)}(y) \;=\; \sum_{z} p^{\mathfrak{C}}(z\,|\,x) \sum_{\tilde{x}} p^{\mathfrak{C}}(y\,|\,\tilde{x},z)\; p^{\mathfrak{C}}(\tilde{x}). \tag{6.23}$$

The fact that observing Z in addition to X and Y here reveals causal information nicely shows that causal relations can also be explored by observing the “channel” (here Z) that carries the “signal” from X to Y. □

Bibliography I

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.

Judea Pearl. Belief networks revisited. *Artificial Intelligence*, 59:49–56, 1993.

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

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12): 1393–1512, 1986.

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

Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.

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