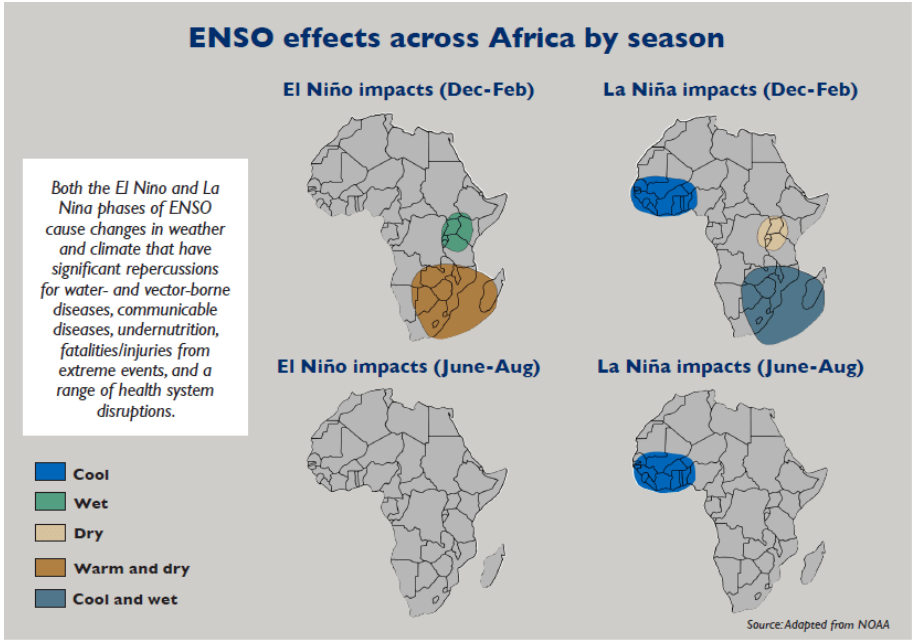
Causal Inference

Gherardo Varando

IPL
16 May 2023

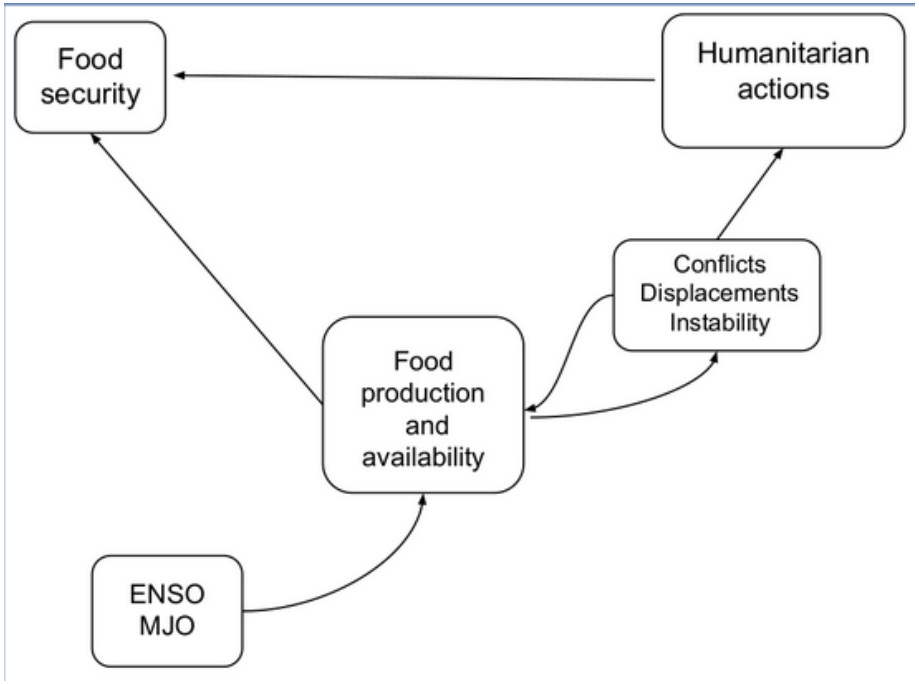
Causal Inference

- 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?

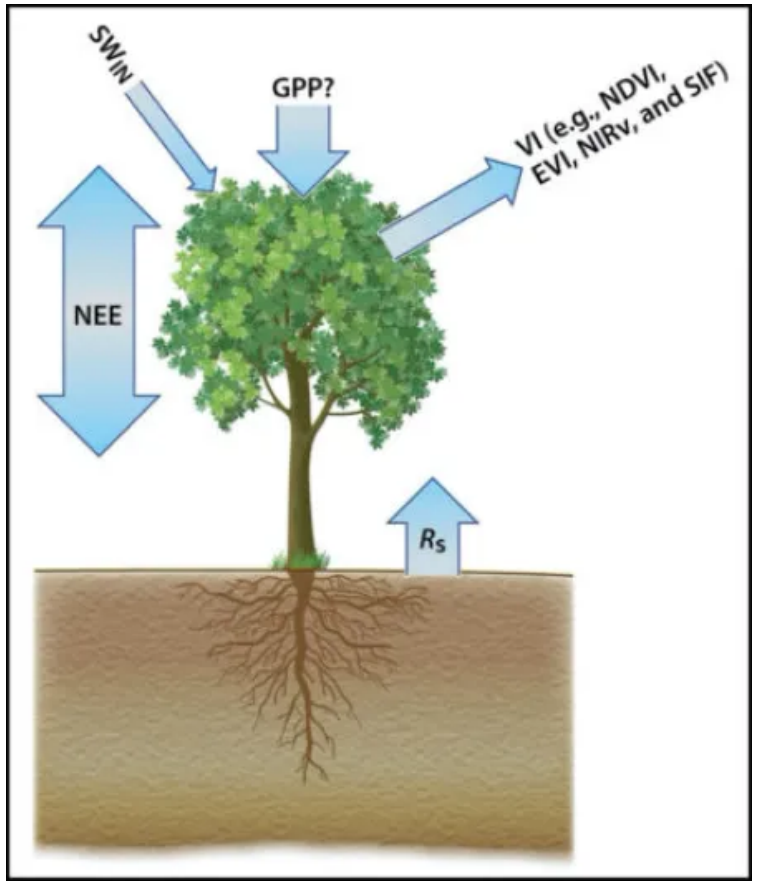


Causal Inference

- Estimate the strength of the effect of ENSO on vegetation greenness
- How effective are humanitarian actions to fight food insecurity?
- there is an effect of precipitation in Denmark into precipitation in the Mediterranean region?

Causal Inference

- Estimate the strength of the effect of ENSO on vegetation greenness
- How effective are humanitarian actions to fight food insecurity?
- there is an effect of precipitation in Denmark into precipitation in the Mediterranean region?
- What is the effect of radiation on NEE/GPP ? Or the effect of temperature on RECO?



- Causal link between forest management practices and burned area size, fire severity, post-fire regeneration
- Biomass yield drivers/what if scenarios (management alternatives to optimize productivity)
- Applying causal inference methods to study problems that feature social and climate variables
- Estimate the effect of droughts on crop production across Europe
- Estimate the conditional average treatment effect (CATE) of major crop rotations (in Europe) on yield and/or soil organic carbon using Double Machine Learning (DML)
- ...

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 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
- Usually assume total or partial knowledge of the causal graph

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
- Usually assume total or partial knowledge of the causal graph
- SCM or potential outcome frameworks

What we will discuss

1. Brief and fast summary of Part I from *Causal Inference: What If* book [Hernán and Robins, 2020]: randomized trials, observational studies and conditions for identification of causal effects, biases
2. Connections with SCM, graphical causal models
3. adjustment, backdoor criterion, do-calculus, front-door adjustment [Peters et al., 2017]
4. some examples

Potential outcomes - Counterfactual model

- 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

Hernán and Robins [2020], Wasserman [2004]

Potential outcomes - Counterfactual model

- 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
- 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: 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
- 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 is actually observed/factual.

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

Hernán and Robins [2020], Wasserman [2004]

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 (imagine infinite data or all the population is studied)

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 (imagine infinite data or all the population is studied)
- assume no problem with the study, everybody is following instruction and there are no measurements problems (*ideal randomized experiment*)

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 (imagine infinite data or all the population is studied)
- 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

- 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 (imagine infinite data or all the population is studied)
- 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

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 (imagine infinite data or all the population is studied)
- 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
- formally exchangeability is $Y^a \perp\!\!\!\perp A$ (**careful, not** $Y \perp\!\!\!\perp A$!!)

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 va = 0 if tails.
- We then coll burned or not after 1 year) nagine infinite data or all th
- assume no p is following instruction a oblems (*ideal randomized*
- can we say s of A on Y ?
- yes! we can ... formally because ther treated $(A = 1)$ and
- formally exch **not** $Y \perp\!\!\!\perp A$!!)

What if we do not have enough money for treating (managing the forest patch) in half of the population? What if we have a limit on the funding available ? can we still perform an (ideal) randomized experiment ?

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 (if) we are in an ideal randomized study setting thus we can compute average treatment effects (e.g. suppose we obtain risk ratios $P(Y^{a=1} = 1|L)/P(Y^{a=0} = 1|L)$ $0.002/0.001 = 2$, $0.001/0.005 = 0.2$ and $0.01/0.2 = 0.05$)

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 (if) we are in an ideal randomized study setting thus we can compute average treatment effects (e.g. suppose we obtain risk ratios $P(Y^{a=1} = 1|L)/P(Y^{a=0} = 1|L)$ $0.002/0.001 = 2$, $0.001/0.005 = 0.2$ and $0.01/0.2 = 0.05$)
- this is called *stratification* and since the causal effect measured are different in each stratum we say that there is *effect modification* by L

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 (if) we are in an ideal randomized study setting thus we can compute average treatment effects (e.g. suppose we obtain risk ratios $P(Y^{a=1} = 1|L)/P(Y^{a=0} = 1|L)$ $0.002/0.001 = 2$, $0.001/0.005 = 0.2$ and $0.01/0.2 = 0.05$)
- this is called *stratification* and since the causal effect measured are different in each stratum we say that there is *effect modification* by L
- 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 $P(Y^a = 1) = \sum w_l P(Y|A = a, L = l)$

Computing ATE from conditionally randomized data

- From the da… randomized experiment w… population.
- **Standardiza…** arginal counterfactu… the stratum-spec…

$$P(\ldots (L = l)$$

- **Inverse Pro…** tive, but equivalent, p… by weighting each individual sample by $w_l = 1/P(A = a|L = l)$ and then we compute $P(Y^a = 1) = \sum w_l P(Y|A = a, L = l)$

remember we obtained risk ratios $P(Y^{a=1} = 1|L)/P(Y^{a=0} = 1|L)$ $0.002/0.001 = 2$, $0.001/0.005 = 0.2$ and $0.01/0.2 = 0.05$ in strata with treatment probabilities 0.8, 0.5 and 0.2 for low, medium and high risk areas. What is the risk ratio in the entire population ?

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 pop- ulation 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 pop- ulation 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 pop- ulation 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

Counterfactual models associated with a causal DAG. In this book, a causal DAG G represents an underlying counterfactual model. To provide a formal definition of the counterfactual model represented by a DAG G, we use the following notation. For any random variable W, let $\mathcal{W}$ denote the support (i.e., the set of possible values w) of W. For any set of ordered variables $W_1, \ldots, W_m$, define $\overline{w}_m = (w_1, \ldots, w_m)$. Let R denote any subset of variables in V and let r be a value of R. Then V_m^r denotes the counterfactual value of V_m when R is set to r.

A nonparametric structural equation model (NPSEM) represented by a DAG G with vertex set V assumes the existence of unobserved random variables (errors) ϵ_m and deterministic unknown functions $f_m(pa_m, \epsilon_m)$ such that $V_1 = f_1(\epsilon_1)$ and the one-step ahead counterfactual $V_m^{\overline{v}_{m-1}} \equiv V_m^{pa_m}$ is given by $f_m(pa_m, \epsilon_m)$. That is, only the parents of V_m have a direct effect on V_m relative to the other variables on G. An NPSEM implies that any variable V_j on the graph can be intervened on, as counterfactuals in which V_j has been set to a specific value v_j are assumed to exist. Both the factual variable V_m and the counterfactuals V_m^r for any $R \subset V$ are obtained recursively from V_1 and $V_j^{\overline{v}_{j-1}}$, $m \geq j > 1$. For example, $V_3^{v_1} = V_3^{v_1, V_2^{v_1}}$, i.e., the counterfactual value $V_3^{v_1}$ of V_3 when V_1 is set to v_1 is the one-step ahead counterfactual $V_3^{v_1, v_2}$ with v_2 equal to the counterfactual value $V_2^{v_1}$ of V_2. Similarly, $V_3 = V_3^{V_1, V_2^{V_1}}$ and $V_3^{v_1, v_4} = V_3^{v_1}$ because V_4 is not a direct cause of V_3.

Robins (1986) called this NPSEM a finest causally interpreted structural tree graph (FCISTGs) "as fine as the data". Pearl (2000) showed how to represent this model with a DAG. Robins (1986) also proposed more realistic causally interpreted structural tree graphs in which only a subset of the variables are subject to intervention. For expositional purposes, we will assume that every variable can be intervened on, even though the statistical methods considered here do not actually require this assumption.

A FCISTG model does not imply that the causal Markov assumption of Technical Point 6.1 holds; additional statistical independence assumptions are needed. For example, Pearl (2000) assumed an NPSEM in which all error terms ϵ_m are mutually independent. We refer to Pearl's model with independent errors as an NPSEM-IE. In contrast, Robins (1986) only assumed that the one-step ahead counterfactuals $V_m^{\overline{v}_{m-1}} = f_m(pa_m, \epsilon_m)$ and $V_j^{\overline{v}_{j-1}} = f_j(pa_j, \epsilon_j)$, $j < m$, are jointly independent when $\overline{v}_{j-1}$ is a subvector of the $\overline{v}_{m-1}$, and referred to this as the finest fully randomized causally interpreted structured tree graph (FFRCISTG) model. Robins (1986) showed this assumption implies that the causal Markov assumption holds. An NPSEM-IE is an FFRCISTG but not vice-versa because an NPSEM-IE makes many more independence assumptions than an FFRCISTG (Robins and Richardson 2011).

A DAG represents an NPSEM but we need to specify which type. For example, the DAG in Figure 6.2 may correspond to either an NPSEM-IE that implies full exchangeability $\left(Y^{a=0}, Y^{a=1}\right) \perp\!\!\!\perp A$, or to an FFRCISTG that only implies marginal exchangeability $Y^a \perp\!\!\!\perp A$ for both $a = 0$ and $a = 1$. In this book we assume that DAGs represent FFRCISTGs whenever we do not mention the underlying counterfactual model.

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

---

Peters et al. [2017]

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

---

Peters et al. [2017]

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*

Peters et al. [2017]

Confounding and adjusting

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

Peters et al. [2017]

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

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

Peters et al. [2017]

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** ...

Peters et al. [2017]

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

---

Peters et al. [2017]

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

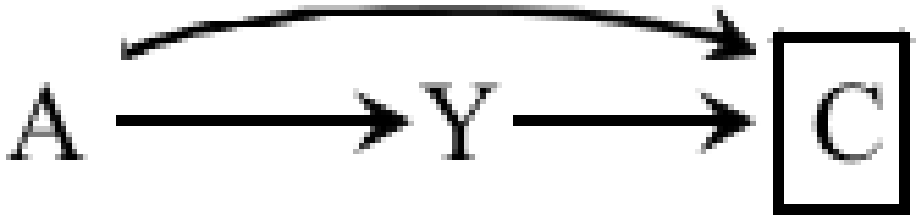
$$P^{\mathcal{C},do(X=x)}(y) = \sum P^{\mathcal{C}}(Y|X, \mathbf{Z} = z)P^{\mathcal{C}}(\mathbf{Z} = \mathbf{z})$$

- 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

---

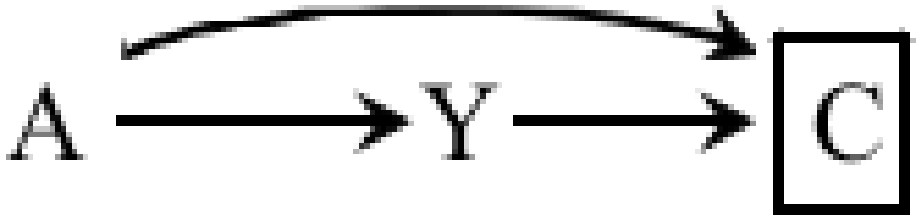
Peters et al. [2017]

Selection bias



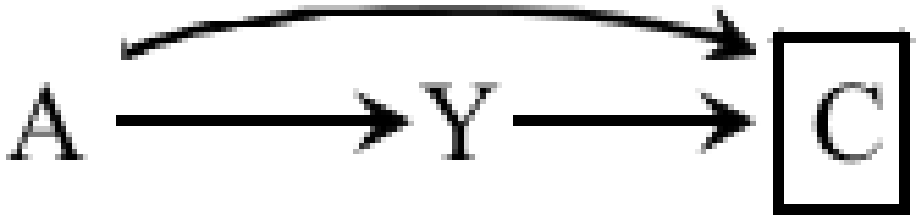
- If we condition on C there are two open paths between A and Y

Selection bias



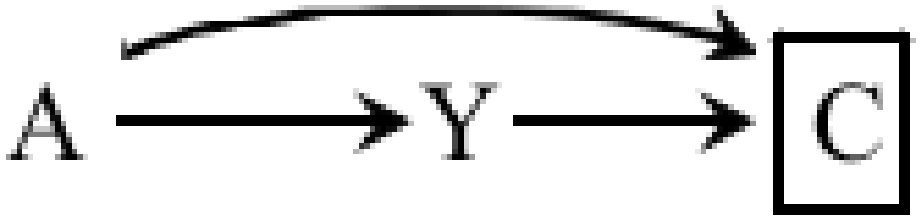
- If we condition on C there are two open paths between A and Y
- This can happen for example: differential loss to follow-up, missing data bias, nonresponse bias, healthy worker bias, self-selection bias, volunteer bias, and selection affected by treatment received before study entry

Selection bias



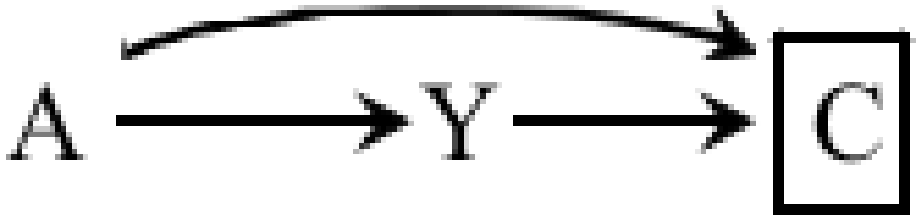
- If we condition on C there are two open paths between A and Y
- This can happen for example: differential loss to follow-up, missing data bias, nonresponse bias, healthy worker bias, self-selection bias, volunteer bias, and selection affected by treatment received before study entry
- Selection bias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

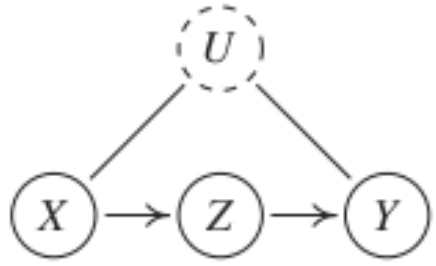
Do-calculus

- 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. □

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

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.

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*, volume 26. Springer, 2004.