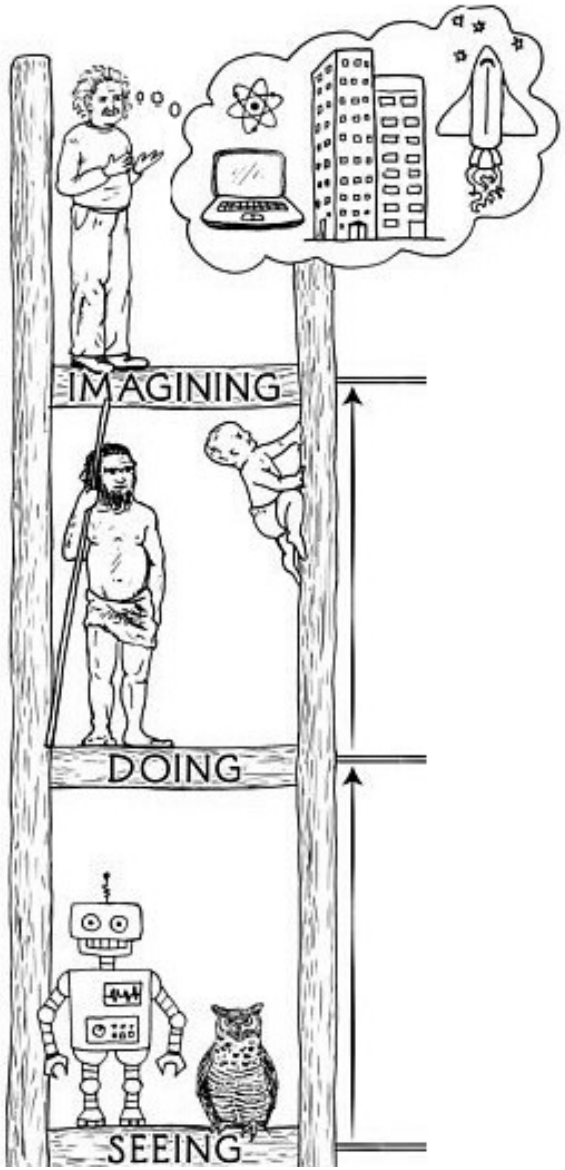
Estimating causal effects

Charles K. Assaad, Emilie Devijver, Eric Gaussier

emilie.devijver@univ-grenoble-alpes.fr

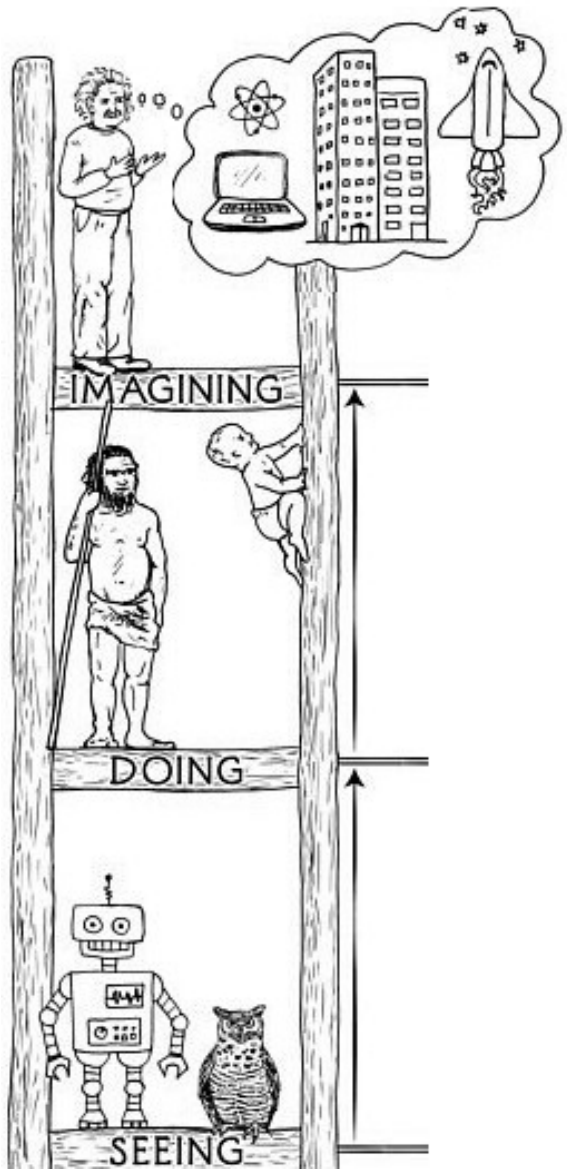
Table of content



Counterfactuals

Interventions

Associations



Counterfactuals

I took an aspirin, and my headache is gone: would I have had a headache had I not taken that aspirin?

Interventions

It I take an aspirin now, will I wake up with a headache?

P(*headache*|do(*aspirin*))

Associations

I took an aspirin after dinner, will I wake up with a headache?

Potential Outcomes (PO)

- Set of n units indexed by i (individuals)
- T_i be the value of a treatment assigned to individual i

Definition (Potential outcomes)

The *potential outcome* under treatment level t, denoted by $Y_i(t)$, is the value that the outcome would have taken were T_i set to t, possibly contrary to the fact.

- For binary T_i, $Y_i(0)$ is the potential outcome if the unit i does not receive the treatment (control), and $Y_i(1)$ is the potential outcome if the unit i does receive the treatment (treated).
  - Treatment Group: The group of subjects or units that receive the specific intervention or treatment being studied.
  - Control Group: The group of subjects or units that do not receive the treatment, serving as a comparison to assess the treatment's effectiveness.

Potential Outcomes (PO)

- Set of n units indexed by i (individuals)
- T_i be the value of a treatment assigned to individual i

Definition (Potential outcomes)

The *potential outcome* under treatment level t, denoted by $Y_i(t)$, is the value that the outcome would have taken were T_i set to t, possibly contrary to the fact.

- For binary T_i, $Y_i(0)$ is the potential outcome if the unit i does not receive the treatment (control), and $Y_i(1)$ is the potential outcome if the unit i does receive the treatment (treated).
  - Treatment Group: The group of subjects or units that receive the specific intervention or treatment being studied.
  - Control Group: The group of subjects or units that do not receive the treatment, serving as a comparison to assess the treatment's effectiveness.

Potential Outcomes (Contd.)

Key Points about Potential Outcomes:

- Fundamental for understanding causal effects and comparing the effects of interventions.
- Crucial for estimating the impact of treatments or interventions in causal inference studies.
- Potential outcomes enable us to translate causal questions into the estimation of a causal estimand.

History of the concept

- Started from Neyman (1923) and Fisher's (1935) work on understanding experiments
- Formalized by Rubin in a series of papers (from 1974)
- Potential outcomes has evolved into an entire framework for causal inquiry.

Can be seen as an alternative way to express counterfactuals by the do operator (Pearl, 2000).

Potential Outcomes (Contd.)

Key Points about Potential Outcomes:

- Fundamental for understanding causal effects and comparing the effects of interventions.
- Crucial for estimating the impact of treatments or interventions in causal inference studies.
- Potential outcomes enable us to translate causal questions into the estimation of a causal estimand.

History of the concept

- Started from Neyman (1923) and Fisher's (1935) work on understanding experiments
- Formalized by Rubin in a series of papers (from 1974)
- Potential outcomes has evolved into an entire framework for causal inquiry.

Can be seen as an alternative way to express counterfactuals by the do operator (Pearl, 2000).

Example of Treatment and Control Groups

Illustrative Example:

- Consider a clinical trial evaluating the effectiveness of a new drug for a specific medical condition.
- The patients receiving the actual drug constitute the treatment group, while those receiving a placebo or standard treatment form the control group.
- By comparing the outcomes between the two groups, researchers can assess the causal impact of the new drug on the patients' health outcomes.

Counterfactuals for Estimating Causal Effects

Using Counterfactuals to Estimate Causal Effects:

- Counterfactuals provide a hypothetical comparison of what would have happened under different treatment conditions.
- Used to estimate the causal effect of an intervention by comparing the observed outcome with the hypothetical outcome that would have occurred without the intervention.

Application of Counterfactuals in Causal Inference:

- Essential for determining the causal impact of treatments, policies, or interventions in observational and experimental studies.
- Enable researchers to evaluate the effectiveness of interventions by comparing the actual outcomes with the hypothetical outcomes in the absence of the intervention.

Counterfactuals for Estimating Causal Effects

Using Counterfactuals to Estimate Causal Effects:

- Counterfactuals provide a hypothetical comparison of what would have happened under different treatment conditions.
- Used to estimate the causal effect of an intervention by comparing the observed outcome with the hypothetical outcome that would have occurred without the intervention.

Application of Counterfactuals in Causal Inference:

- Essential for determining the causal impact of treatments, policies, or interventions in observational and experimental studies.
- Enable researchers to evaluate the effectiveness of interventions by comparing the actual outcomes with the hypothetical outcomes in the absence of the intervention.

Illustrative Example: Counterfactuals in a Study

Example Scenario:

- Consider a study evaluating the impact of a new teaching method on student performance in a particular subject.
- The counterfactual comparison involves assessing the performance of students who received the new teaching method with the hypothetical performance they would have had if they had not received the new method.
- By comparing the actual performance with the hypothetical performance, researchers can estimate the causal effect of the new teaching method on student achievement.

Individual Treatment Effect (ITE)

Represents the causal effect of a treatment or intervention on an individual unit within a study.

Definition (ITE)

For each individual i,

$$ITE_i = Y_{1i} - Y_{0i}$$

where:

- ITE_i is the Individual Treatment Effect for the ith unit,
- Y_{1i} is the PO for the ith unit under the treatment,
- Y_{0i} is the PO for the ith unit under the control.

We can consider other quantities (ratio, percentage increase...) but always some contrast measure between two POs.

Individual Treatment Effect (ITE)

Represents the causal effect of a treatment or intervention on an individual unit within a study.

Definition (ITE)

For each individual i,

$$ITE_i = Y_{1i} - Y_{0i}$$

where:

- ITE_i is the Individual Treatment Effect for the ith unit,
- Y_{1i} is the PO for the ith unit under the treatment,
- Y_{0i} is the PO for the ith unit under the control.

We can consider other quantities (ratio, percentage increase...) but always some contrast measure between two POs.

The fundamental Problem of Causal Inference

- Can we do something to estimate the ITE?
- **The fundamental Problem of Causal Inference** (Holland, 1986)
  It is impossible to observe the value of $Y_i(1)$ and $Y_i(0)$ for the same unit, therefore it is impossible to observe the ITE.

Average Treatment Effect (ATE)

- Represents the average causal effect of a treatment or intervention on the outcome variable within a population.
- Provides an overall assessment of the treatment's impact on the entire population under study.

Definition (ATE)

$$ATE = E[Y_1 - Y_0]$$

where:

- Y_1 is the potential outcome under the treatment,
- Y_0 is the potential outcome under the control,
- $E[\cdot]$ denotes the expectation or average over the entire population.

Differences between ATE and ITE

Distinguishing ATE and ITE:

- ATE provides the average treatment effect for the entire study population, while ITE focuses on the specific effects for individual units.
- ATE assesses the overall impact of a treatment at a population level, while ITE emphasizes individual-level variations in treatment effects.
- ATE is used for evaluating the general effectiveness of interventions, whereas ITE is crucial for understanding personalized treatment effects.

Example of Average Treatment Effect (ATE)

Real-World Scenario:

- A study assessing the impact of a new educational program on student performance.
- ATE is calculated by comparing the average test scores of students who participated in the program with those who did not. We need assumptions here!
- The difference in average scores provides an estimate of the average effect of the educational program on the overall student population.

Example of Individual Treatment Effect (ITE)

Real-World Scenario:

- A clinical trial investigating the efficacy of a new drug for a specific medical condition in a diverse patient population.
- ITE is computed by analyzing the individual response to the drug compared to the response they would have had without the treatment.
- The variation in treatment effects among different patient subgroups helps in identifying specific patient characteristics that influence the drug's effectiveness.

Real-World Examples of Treatment Effects (Contd.)

Key Takeaways from Real-World Examples:

- ATE provides insights into the overall impact of interventions on a study population, guiding policy and program decisions.
- ITE helps in understanding the heterogeneous responses to treatments among individuals, enabling personalized treatment strategies and interventions.

How to estimate ATE?

- **Hypothetical world**: we observe every potential outcome for every individual.
- **In reality**: we observe one (at most) for each individual.

Can we average the observations from control and treatment?
Yes but under very stringent assumptions!

How to estimate ATE?

- **Hypothetical world**: we observe every potential outcome for every individual.
- **In reality**: we observe one (at most) for each individual.

Can we average the observations from control and treatment?
Yes but under very stringent assumptions!

Assumptions

SUTVA

Definition (SUTVA: Stable Unit Treatment Value Assumption)

Observed outcome = potential outcome of the observed treatment: $Y_i(t) = Y_i$ if $T_i = t$.

For a binary treatment, this writes

$$Y_i = Y_i(1) \times T_i + Y_i(0) \times (1 - T_i)$$

- No interference: manipulating another unit's tratment does not affect a unit's PO
- Consistency: for each unit, no different form or version of each treatment level, which lead to different PO

Definition (SUTVA: Stable Unit Treatment Value Assumption)

Observed outcome = potential outcome of the observed treatment: $Y_i(t) = Y_i$ if $T_i = t$.

For a binary treatment, this writes

$$Y_i = Y_i(1) \times T_i + Y_i(0) \times (1 - T_i)$$

- No interference: manipulating another unit's tratment does not affect a unit's PO
- Consistency: for each unit, no different form or version of each treatment level, which lead to different PO

Assumptions

Positivity and ignorability

Definition

Positivity: we assume that, for all units i and treatment levels t,

$$P(T_i = t) > 0$$

Ignorability: we assume that, for all treatment levels, t,

$$Y_i(t) \perp\!\!\!\perp T_i$$

This means that the average outcome in the treated group is representative of what we would see on average if everyone got treated (same for the controls).

$$E(Y_i(1)) = E(Y_i(1)|T_i = 1) = E(Y_i(1)|T_i = 0)$$

First estimator of ATE

Under SUTVA, positivity, ignorability,

$$\begin{aligned} E(Y_i|T_i=1) - E(Y_i|T_i=0) &= E(Y_i(1)|\ T_i=1) - E(Y_i(0)|T_i=0) \\ &= E(Y_i(1)) - E(Y_i(0)) \\ &= ATE \end{aligned}$$

Sample means estimator

$$\widehat{\text{ATE}} = \frac{1}{n_t}\sum_{i=1}^{n} Y_i T_i - \frac{1}{n_c}\sum_{i=1}^{n} Y_i(1-T_i)$$

Estimator unbiased, consistent and asymptotically gaussian.

First estimator of ATE

Under SUTVA, positivity, ignorability,

$$\begin{aligned} E(Y_i|T_i=1) - E(Y_i|T_i=0) &= E(Y_i(1)|\ T_i=1) - E(Y_i(0)|T_i=0) \\ &= E(Y_i(1)) - E(Y_i(0)) \\ &= ATE \end{aligned}$$

Sample means estimator

$$\widehat{\text{ATE}} = \frac{1}{n_t}\sum_{i=1}^{n} Y_i T_i - \frac{1}{n_c}\sum_{i=1}^{n} Y_i(1-T_i)$$

Estimator unbiased, consistent and asymptotically gaussian.

Randomized Controlled Trials (RCTs)

The easiest way to collect data that satisfies those assumptions is to perform a randomized experiment.

Definition (Randomized Experiment)

An experiment is a study in which the probability of treatment assignment $P(T_i = t)$ is directly under the control of a researcher.

Definition and Application:

- RCTs are experimental studies where participants are randomly assigned to either the treatment or control group.
- They are considered the gold standard for estimating causal effects as randomization helps control for both observed and unobserved confounding variables.

Challenges: non-compliance, arm switches, ...

Randomized Controlled Trials (RCTs)

Example:

- Clinical trials for testing the efficacy of a new drug in a controlled setting.
- Analyzing the impact of a policy change on employment using survey data and statistical controls.

Drawbacks: not always possible (unethical, need for a large sample size, bias in population selection, lack of follow-up...)

Challenges and Solutions:

- Observational studies use data from naturally occurring settings and are prone to confounding and bias.
- Techniques such as multivariate regression, stratification, and sensitivity analysis help control for confounding factors and improve causal inference from observational data.

Randomized Controlled Trials (RCTs)

Example:

- Clinical trials for testing the efficacy of a new drug in a controlled setting.
- Analyzing the impact of a policy change on employment using survey data and statistical controls.

Drawbacks: not always possible (unethical, need for a large sample size, bias in population selection, lack of follow-up...)

Challenges and Solutions:

- Observational studies use data from naturally occurring settings and are prone to confounding and bias.
- Techniques such as multivariate regression, stratification, and sensitivity analysis help control for confounding factors and improve causal inference from observational data.

Randomized Controlled Trials (RCTs)

Example:

- Clinical trials for testing the efficacy of a new drug in a controlled setting.
- Analyzing the impact of a policy change on employment using survey data and statistical controls.

Drawbacks: not always possible (unethical, need for a large sample size, bias in population selection, lack of follow-up...)

Challenges and Solutions:

- Observational studies use data from naturally occurring settings and are prone to confounding and bias.
- Techniques such as multivariate regression, stratification, and sensitivity analysis help control for confounding factors and improve causal inference from observational data.

Another set of assumptions

Definition

Conditional positivity: we assume that, for all units i and treatment levels t,

$$P(T_i = t | X_i = x) > 0$$

for all x in the domain.
Conditional ignorability: we assume that
$Y_i(1), Y_i(0) \perp\!\!\!\perp T_i | X_i = x$ for all x and t.

- Knowing a unit's covariate values will never determine what treatment that unit gets with certainty
- The covariate tell the whole story of the treatment assignment process, and within levels of X_i, treatment is assigned as-if-random.

How to pick the good set of covariates?

Another set of assumptions

Definition

Conditional positivity: we assume that, for all units i and treatment levels t,

$$P(T_i = t | X_i = x) > 0$$

for all x in the domain.
Conditional ignorability: we assume that
$Y_i(1), Y_i(0) \perp\!\!\!\perp T_i | X_i = x$ for all x and t.

- Knowing a unit's covariate values will never determine what treatment that unit gets with certainty
- The covariate tell the whole story of the treatment assignment process, and within levels of X_i, treatment is assigned as-if-random.

How to pick the good set of covariates?

Another set of assumptions

Definition

Conditional positivity: we assume that, for all units i and treatment levels t,

$$P(T_i = t | X_i = x) > 0$$

for all x in the domain.
Conditional ignorability: we assume that
$Y_i(1), Y_i(0) \perp\!\!\!\perp T_i | X_i = x$ for all x and t.

- Knowing a unit's covariate values will never determine what treatment that unit gets with certainty
- The covariate tell the whole story of the treatment assignment process, and within levels of X_i, treatment is assigned as-if-random.

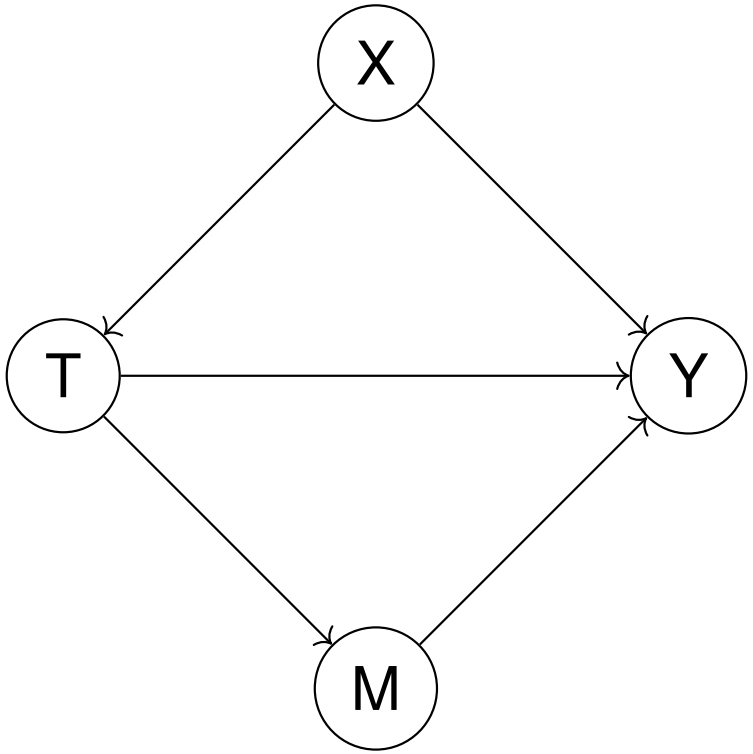
How to pick the good set of covariates?

How to pick the good set of covariates?

It comes back to the causal graph!

We look for variables that

- block all non-causal paths from T to Y
- let all causal paths from T to Y open.



Setting on observational dataset

- Dataset $(T_i, Y_i, X_i)_{1 \leq i \leq n}$
- Assumptions: SUTVA, conditional ignorability wrt X, conditional positivity wrt X

Post-stratification

- Within strate, we can identify the ATE by the difference-in-means
- $\text{CATE}(x) = E(Y(1) - Y(0)|X = x)$
- Come back to ATE:

$$\begin{aligned} & E(Y(1) - Y(0)) \\ =& E(E(Y(1) - Y(0)|X)) \\ =& \sum_x (E(Y|T = 1, X = x) - E(Y|T = 0, X = x))P(X = x) \end{aligned}$$

estimated by

$$\widehat{\text{ATE}} = \sum_x \widehat{\text{CATE}}(x)\frac{n_x}{n}$$

Drawbacks:

- If too many strates, too few units in each strate
- continuous covariates...

Post-stratification

- Within strate, we can identify the ATE by the difference-in-means
- $\text{CATE}(x) = E(Y(1) - Y(0)|X = x)$
- Come back to ATE:

$$\begin{aligned} & E(Y(1) - Y(0)) \\ =& E(E(Y(1) - Y(0)|X)) \\ =& \sum_x (E(Y|T = 1, X = x) - E(Y|T = 0, X = x))P(X = x) \end{aligned}$$

estimated by

$$\widehat{\text{ATE}} = \sum_x \widehat{\text{CATE}}(x)\frac{n_x}{n}$$

Drawbacks:

- If too many strates, too few units in each strate
- continuous covariates...

Inverse Probability of Treatment Weighting (IPTW)

Fact: in a randomized experiment, covariate distribution are balanced across treatment groups, but not in observational studies.

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n} w_i(Y_i T_i - Y_i(1-T_i))$$

Which weight?

Definition (Propensity score)

The propensity score is the probability of receiving the treatment:

$$e(x) = P(T = 1|X = x)$$

IPTW (or Horowitz-Thompson estimator): weighted estimator with $w_i = (e(X_i))^{-1}$ for treated units and $w_i = (1 - e(X_i))^{-1}$ for control units

Inverse Probability of Treatment Weighting (IPTW)

Fact: in a randomized experiment, covariate distribution are balanced across treatment groups, but not in observational studies.

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n} w_i(Y_i T_i - Y_i(1-T_i))$$

Which weight?

Definition (Propensity score)

The propensity score is the probability of receiving the treatment:

$$e(x) = P(T = 1 | X = x)$$

IPTW (or Horowitz-Thompson estimator): weighted estimator with $w_i = (e(X_i))^{-1}$ for treated units and $w_i = (1 - e(X_i))^{-1}$ for control units

Estimating the Propensity Score

Logistic Regression Formula:

$$\text{logit}(e) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

Where:

- e represents the estimated propensity score,
- $X_1, X_2, \ldots, X_p$ represent the observed covariates or confounding variables,
- $\beta_0, \beta_1, \ldots, \beta_p$ are the coefficients of the logistic regression model.

The estimated propensity scores are used in the matching process to create balanced groups for comparison and analysis.

Estimating the propensity score

More recently, many more ML methods

- Boosting
- NN
- RF

Matching

Definition and Application: how to fill in the missing outcome for each unit?

- Propensity score matching is used to estimate the causal effect of a treatment or intervention by balancing the distribution of observed covariates between the treatment and control groups.
- It involves matching treated and untreated units

Definition (Matching)

For each unit i, find the unit j with opposite treatment and most similar covariate values and use their outcome as the missing one for i.

Which similarity? Euclidian, propensity score-based, ...

Example in Social Sciences

Research Question: Does participation in a mentoring program improve academic performance in at-risk students?
Propensity Score Matching Process:

1. Collect demographic data, socioeconomic background, and previous academic performance of at-risk students.
2. Estimate the propensity scores using logistic regression, considering relevant covariates.
3. Match treated students who participated in the mentoring program with similar untreated students who did not participate, based on their propensity scores.
4. Compare the academic performance of the matched groups to evaluate the impact of the mentoring program on the students' academic outcomes.

Regression-based methods

We blocked the non-causal open path from T to Y by adjusting on X.
How to model the relationship between $Y_i(t)$ and X_i?
Regression model:

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma$$

Do $\hat{\beta}$ is an estimate of the treatment effect?
Only if (the model is correctly specified and) the treatment effect is constant across units $\rightarrow$ **homogeneous treatment effect**

Heterogeneous treatment effect

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma + T_i X_i\lambda$$

Then ATE $= \beta + E(X_i)\lambda$.

Regression-based methods

We blocked the non-causal open path from T to Y by adjusting on X.
How to model the relationship between $Y_i(t)$ and X_i?
Regression model:

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma$$

Do $\hat{\beta}$ is an estimate of the treatment effect?
Only if (the model is correctly specified and) the treatment effect is constant across units $\rightarrow$ **homogeneous treatment effect**

Heterogeneous treatment effect

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma + T_i X_i\lambda$$

Then ATE $= \beta + E(X_i)\lambda$.

Regression-based methods

We blocked the non-causal open path from T to Y by adjusting on X.
How to model the relationship between $Y_i(t)$ and X_i?
Regression model:

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma$$

Do $\hat{\beta}$ is an estimate of the treatment effect?
Only if (the model is correctly specified and) the treatment effect is constant across units → **homogeneous treatment effect**

Heterogeneous treatment effect

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma + T_i X_i\lambda$$

Then ATE $= \beta + E(X_i)\lambda$.

Regression-based methods

We blocked the non-causal open path from T to Y by adjusting on X.
How to model the relationship between $Y_i(t)$ and X_i?
Regression model:

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma$$

Do $\hat{\beta}$ is an estimate of the treatment effect?
Only if (the model is correctly specified and) the treatment effect is constant across units $\rightarrow$ **homogeneous treatment effect**

Heterogeneous treatment effect

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma + T_i X_i\lambda$$

Then ATE $= \beta + E(X_i)\lambda$.

Regression-based methods

We blocked the non-causal open path from T to Y by adjusting on X.
How to model the relationship between $Y_i(t)$ and X_i?
Regression model:

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma$$

Do $\hat{\beta}$ is an estimate of the treatment effect?
Only if (the model is correctly specified and) the treatment effect is constant across units $\rightarrow$ **homogeneous treatment effect**

Heterogeneous treatment effect

$$E(Y_i|T_i, X_i) = \alpha + \beta T_i + X_i\gamma + T_i X_i\lambda$$

Then ATE $= \beta + E(X_i)\lambda$.

Regression-based methods

$$\text{ATE} = \beta + E(X_i)\lambda$$

One would like $E(X_i) = 0$: de-meaning covariates
Then in very specific cases with very strong assumptions, you can express the ATE with the regression coefficients!

Other solution:

S-learner $\quad \mu(t, x) = E(Y|T = t, X = x)$

T-learner $\quad \mu(1, x) = E(Y|T = 1, X = x)$

$\mu(0, x) = E(Y|T = 0, X = x)$

Regression-based methods

$$\text{ATE} = \beta + E(X_i)\lambda$$

One would like $E(X_i) = 0$: de-meaning covariates
Then in very specific cases with very strong assumptions, you can express the ATE with the regression coefficients!

Other solution:

S-learner	$\mu(t, x) = E(Y \mid T = t, X = x)$
T-learner	$\mu(1, x) = E(Y \mid T = 1, X = x)$
	$\mu(0, x) = E(Y \mid T = 0, X = x)$

Double robustness

- Most estimator (IPTW, S-learner, T-learner) are sensitive to model misspecification
- Doubly robust estimator: combine them! For example, augmented IPW:

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i) + T_i \frac{Y_i - \hat{\mu}_{(1)}(X_i)}{\hat{e}(X_i)} - (1 - T_i)\frac{Y_i - \hat{\mu}_{(0)}(X_i)}{1 - \hat{e}(X_i)}$$

- Even with double robustness, needs of correct specification

Double robustness

- Most estimator (IPTW, S-learner, T-learner) are sensitive to model misspecification
- Doubly robust estimator: combine them! For example, augmented IPW:

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i) + T_i \frac{Y_i - \hat{\mu}_{(1)}(X_i)}{\hat{e}(X_i)} - (1 - T_i)\frac{Y_i - \hat{\mu}_{(0)}(X_i)}{1 - \hat{e}(X_i)}$$

- Even with double robustness, needs of correct specification

Double Machine Learning

$$Y = g_0(T, X) + U$$
$$ATE = E(g_0(1, X) - g_0(0, X))$$

g_0 is learnt by ML

- flexibility,
- heterogenous treatment effects,
- high dimensional X

1st method

1. $\hat{g}_0$ using ML
2. plug in predictions to estimate ATE

Caution! ML methods address the variance-bias trade-off, which leads to a bias of the causal estimate

Double Machine Learning

$$Y = g_0(T, X) + U$$
$$ATE = E(g_0(1, X) - g_0(0, X))$$

g_0 is learnt by ML

- flexibility,
- heterogenous treatment effects,
- high dimensional X

1st method

1. $\hat{g}_0$ using ML
2. plug in predictions to estimate ATE

Caution! ML methods address the variance-bias trade-off, which leads to a bias of the causal estimate

Double Machine Learning

Definition (Neyman Orthogonality)

The error terms that arise due to regularization do not affect the causal estimate. For ψ a score function, $\mathcal{D}$ a dataset, η the nuisance part,

$$E(\psi(\mathcal{D}; \mathsf{ATE}, \eta)) = 0.$$

Sample splitting: split the sample into two parts: one for the ML estimation, one for the causal estimation ATE.

Theorem

Central limit theorem for the double ML estimator under regularity conditions with known covariance matrix.

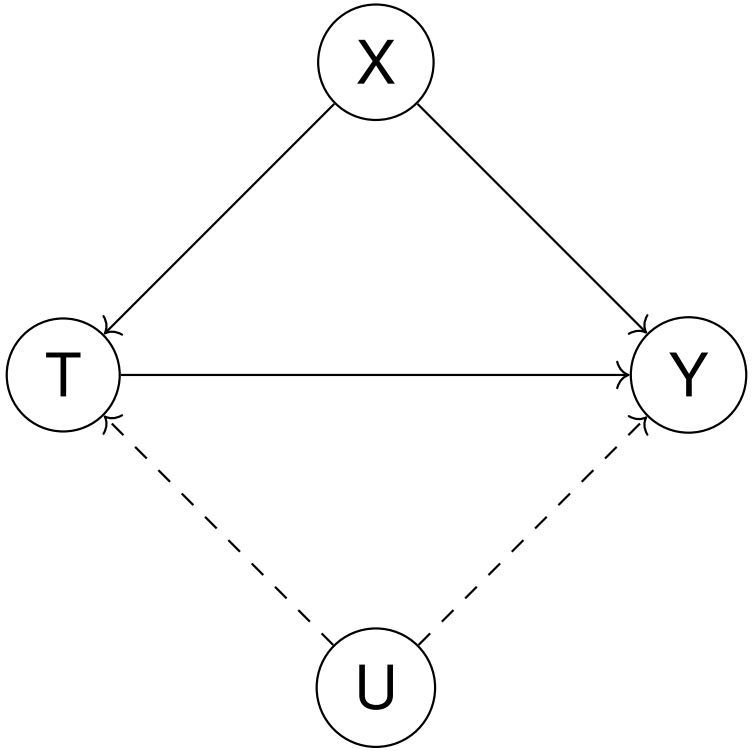
Causal forest

- A causal tree is constructing leaves such that the individuals *i* come from a randomized experiment. Then, sum over the leaves.
- Causal forest: ensemble of causal trees
- This is a consistent estimator of CATE
- Variable importance deduced from causal RFs

Unmeasured confounding

If we suspect some unmeasured confounding,

- Conditional ignorability does not hold with respect to X
- We assume that there is an unmeasured variable U such that conditional ignorability hold with respect to X and U



Instrumental Variable Analysis

Definition and Application:

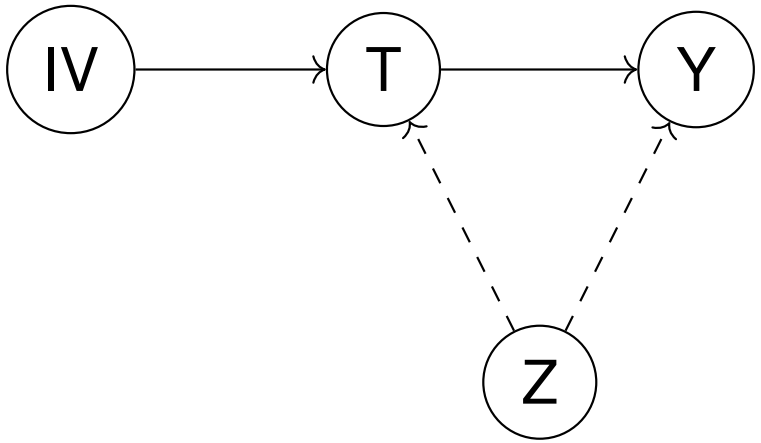
- Instrumental variables IV: correlated with the treatment but not directly associated with the outcome, allowing for the estimation of causal effects in the presence of unobserved biases.
- Key aspect of an IV: should be strongly correlated with the endogenous variable (X) but uncorrelated with the error term and any unobserved confounders (Z). This helps address potential issues of endogeneity and omitted variable bias, improving the validity of your causal inference.
- In other words: we want to split the variation in X_i that is uncorrelated with the noise, to estimate the causal effect.

Instrumental Variable Analysis

Definition (Instrumental Variable)

An instrumental variable IV must satisfy three conditions:

1. **Relevance** IV has a causal effect on T
2. **Exclusion restriction** the causal effect of IV on Y is fully mediated by Y
3. **Instrumental unconfoundedness** The relationship between IV and Y is unconfounded of confounded only by variables we measure and can adjust on.



Instrumental Variable Analysis

Estimation through IV

Under the linear model,

$$
\begin{aligned}
Z &= \varepsilon_Z \\
IV &= \varepsilon_{IV} \\
T &= \beta_{z,t} Z + \beta_{iv,t} IV + \varepsilon_T \\
Y &= \beta_{z,y} Z + \beta_{t,y} T + \varepsilon_Y .
\end{aligned}
$$

Rewriting the equation, we can identify $\beta_{t,y}$ through the estimation of two linear regressions.

Two-stage least squares estimator (2SLS estimator)

Caution! this can have large variance if the value $\beta_{IV,T}$ is near zero. In such cases, IV is called a weak instrument.

Instrumental Variable Analysis

Estimation through IV
Under the linear model,

$$\begin{aligned} Z &= \varepsilon_Z \\ IV &= \varepsilon_{IV} \\ T &= \beta_{z,t} Z + \beta_{iv,t} IV + \varepsilon_T \\ Y &= \beta_{z,y} Z + \beta_{t,y} T + \varepsilon_Y . \end{aligned}$$

Rewriting the equation, we can identify $\beta_{t,y}$ through the estimation of two linear regressions.
Two-stage least squares estimator (2SLS estimator)
Caution! this can have large variance if the value $\beta_{IV,T}$ is near zero. In such cases, IV is called a weak instrument.

Example in Social Sciences

Research Question: Does increased spending on education lead to improved long-term economic outcomes for individuals?
Instrumental Variable Analysis Process:

1. Identify an instrumental variable, such as a policy change affecting education spending at the regional level.
2. Verify that the instrumental variable is correlated with education spending but not directly associated with individual economic outcomes.
3. Use the instrumental variable to estimate the causal effect of education spending on long-term economic outcomes, addressing the endogeneity issue.

Differences-in-Differences (Diff-in-Diff)

- Diff-in-Diff is a quasi-experimental technique that compares the changes in outcomes between a treatment group and a control group before and after an intervention.
- It helps estimate the causal effect of the intervention by accounting for the common time-related trends in both groups.
- The method is widely used in economics, public policy, and social sciences to evaluate the impact of policy changes and interventions.

Assumptions

- SUTVA at both time points
- Parallel trend assumption $(Y_1(0) - Y_0(0)) \perp\!\!\!\perp T$
- No pre-treatment effect assumption: $E(Y_0(1)|T = 1) = E(Y_0(0)|T = 1)$
- We can identify the average treatment effect in the treated

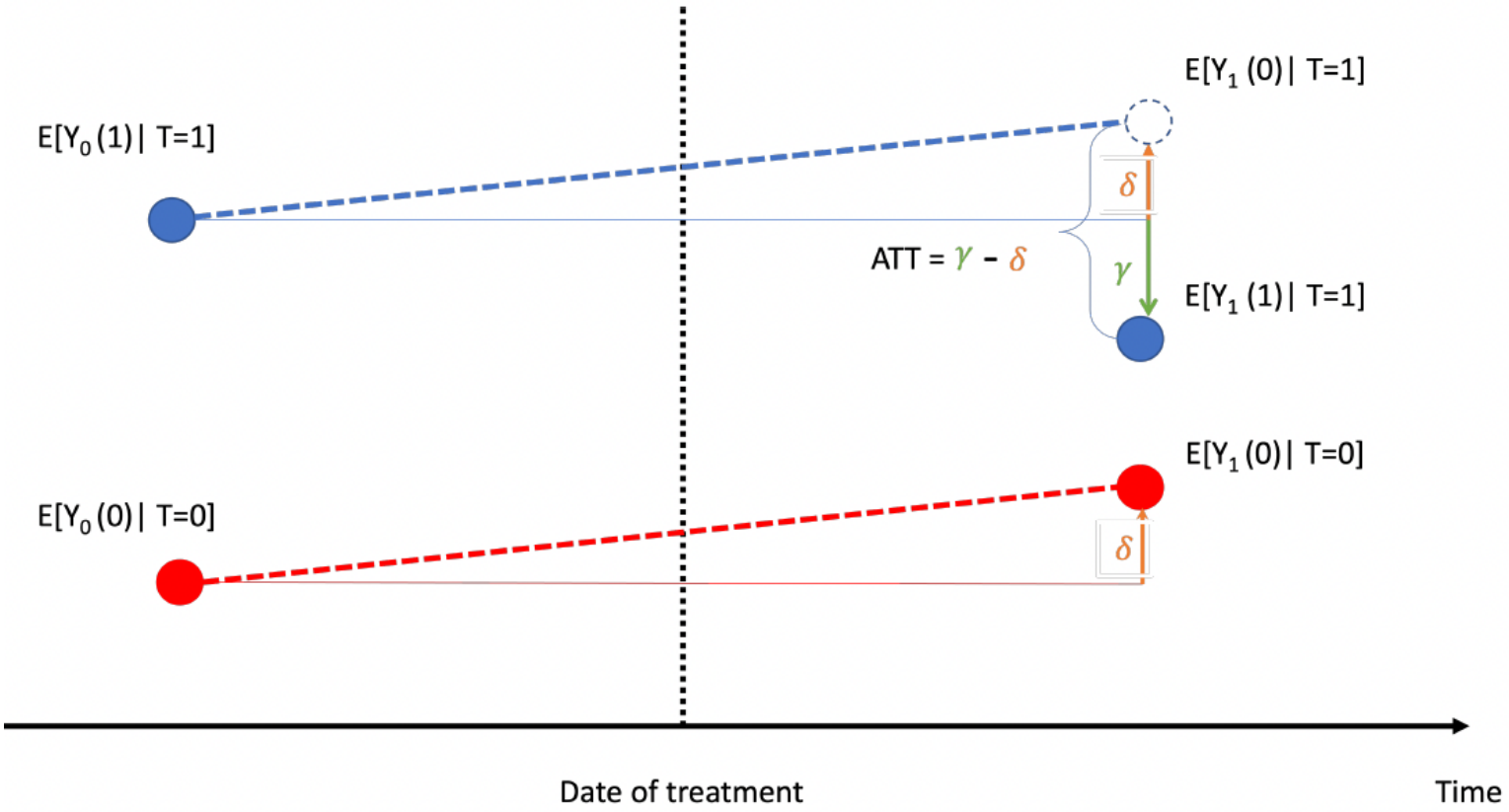
Differences-in-Differences (Diff-in-Diff)

- Diff-in-Diff is a quasi-experimental technique that compares the changes in outcomes between a treatment group and a control group before and after an intervention.
- It helps estimate the causal effect of the intervention by accounting for the common time-related trends in both groups.
- The method is widely used in economics, public policy, and social sciences to evaluate the impact of policy changes and interventions.

Assumptions

- SUTVA at both time points
- Parallel trend assumption $(Y_1(0) - Y_0(0)) \perp\!\!\!\perp T$
- No pre-treatment effect assumption: $E(Y_0(1)|T = 1) = E(Y_0(0)|T = 1)$
- We can identify the average treatment effect in the treated

Differences-in-Differences (Diff-in-Diff) Method



Differences-in-Differences (Diff-in-Diff) Method

Consider the following model:

$$Y_{it} = \beta_0 + \beta_1 \cdot T_i + \beta_2 \cdot \text{Post}_t + \beta_3 \cdot (T_i \times \text{Post}_t) + \epsilon_{it}$$

where:

- Y_{it} represents the outcome variable for unit i at time t,
- T_i is a dummy variable for the treatment group,
- Post_t is a dummy variable for the post-treatment period,
- β_0, β_1, β_2, and β_3 are the regression coefficients,
- ϵ_{it} is the error term.

The Diff-in-Diff method allows the estimation of the causal effect by comparing the differential changes in the outcome variable over time between the treatment and control groups before and after the treatment implementation.

Example in Social Sciences

Research Question: What is the impact of a minimum wage increase on employment in a specific industry?

Diff-in-Diff Analysis Process:

1. Select treatment and control groups from the same industry.
2. Analyze the employment trends before and after the minimum wage increase in both groups.
3. Compare the differences in employment changes between the treatment and control groups to estimate the causal effect of the policy change.

Sum up

- Causal sufficiency
  - Under SUTVA, positivity, ignorability: sample means estimator

$$\widehat{\text{ATE}} = \frac{1}{n_t} \sum_{i=1}^{n} Y_i T_i - \frac{1}{n_c} \sum_{i=1}^{n} Y_i (1 - T_i)$$

  - When those assumptions hold? ... only RCTs ...
  - Lightning assumptions: SUTVA, conditional positivity, conditional ignorability, wrt backdoor/frontdoor set
    - Post-stratification:
    - IPTW: propensity score as weights,
    - Matching
    - Regression-based methods - double ML
- Hidden confounders
  - Instrumental variable analysis
  - Differences in differences

Sum up

- Causal sufficiency
  - Under SUTVA, positivity, ignorability: sample means estimator

$$\widehat{\text{ATE}} = \frac{1}{n_t} \sum_{i=1}^{n} Y_i T_i - \frac{1}{n_c} \sum_{i=1}^{n} Y_i (1 - T_i)$$

  - When those assumptions hold? ... only RCTs ...
  - Lightning assumptions: SUTVA, conditional positivity, conditional ignorability, wrt backdoor/frontdoor set
    - Post-stratification:

$$\widehat{\text{ATE}} = \sum_{x} \widehat{\text{CATE}}(x) \frac{n_x}{n}$$

    - IPTW: propensity score as weights,

$$\widehat{\text{ATE}} = \frac{1}{n} \sum_{i=1}^{n} w_i (Y_i T_i - Y_i (1 - T_i))$$

    - Matching
    - Regression-based methods - double ML
- Hidden confounders
  - Instrumental variable analysis

Sum up

- Causal sufficiency
  - Under SUTVA, positivity, ignorability: sample means estimator

$$\widehat{\text{ATE}} = \frac{1}{n_t} \sum_{i=1}^{n} Y_i T_i - \frac{1}{n_c} \sum_{i=1}^{n} Y_i (1 - T_i)$$

  - When those assumptions hold? ... only RCTs ...
  - Lightning assumptions: SUTVA, conditional positivity, conditional ignorability, wrt backdoor/frontdoor set
    - Post-stratification:
    - IPTW: propensity score as weights,
    - Matching
    - Regression-based methods - double ML
- Hidden confounders
  - Instrumental variable analysis
  - Differences in differences

References

- *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*, Donald Rubin, Journal of Educational Psychology, 1974
- *Causal Inference: The Mixtape*, Scott Cunningham, 2021
- *Causal inference in statistics, social, and biomedical sciences*, Imbens, G. W., & Rubin, D. B., Cambridge University Press, 2015
- *Double/debiased machine learning for treatment and structural parameters*, Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, James Robins, The Econometrics Journal, 2018
- *Estimation and Inference of Heterogeneous Treatment Effects using Random Forests*, Stefan Wager & Susan Athey, Journal of the American Statistical Association, 2018