# CAUSAL NETWORKS

## POTENTIAL OUTCOMES

Fabio Stella
Department of Informatics, Systems and Communication,
University of Milan-Bicocca
Viale Sarca 336, 2016 Milan, ITALY
e-mail: fabio.stella@unimib.it
Twitter: FaSt@FabioAStella

In this lecture you will learn about new concepts and notations needed to clearly describe causal concepts.

In particular, the lecture presents and discusses the following:

- Potential Outcomes and Individual Treatment Effects
- The Fundamental Problem of Causal Inference
  - Average Treatment Effects and Missing Data Interpretation
  - Ignorability – Exchangeability
  - Conditional Exchangeability – Unconfoundedness
  - Positivity – Overlap – Common Support and Extrapolation
  - No interference, Consistency, and SUTVA

# PART I

# POTENTIAL OUTCOMES AND INDIVIDUAL TREATMENT EFFECTS

We will use the following notation:

- $X$ denotes the random variable for **TREATMENT**,
- $Y$ denotes the random variable for the **OUTCOME** of interest,
- **Z** denotes a set of random variables (**COVARIATES**),

In general, we will use uppercase letters $Z$ to denote random variables and lowercase letters $z$ to denote values that random variables take on.

Much of what we consider will be settings where $X$ and $Y$ are binary.

In general, we can extend things to work in settings where $X$ and $Y$ can take on more than two values or where $X$ and $Y$ are continuous.

We have the following narration, you are unhappy (😟), and consider whether or not to get a dog (🐶) to help make you happy (😀).

To answer the question to the right we need to know more.

You get the dog and become happy!!!

The dog (🐶) **CAUSED** you to be happy (😀)?

- What if I told you "*I'm certain you would have become happy also without getting the dog*"?

The dog was not necessary to make you happy, so its claim to a causal effect on your happiness is weak.

- What if instead I told you "*I'm certain you would have remained unhappy without getting the dog*"?

The dog has a pretty strong claim to a causal effect on your happiness.

We have just used the causal concept known as **POTENTIAL OUTCOMES**.

unhappy

$Y = 0$

**OUTCOME**

$(Y)$

happy

$Y = 1$

Happiness = **OUTCOME** $= Y$

Dog = **TREATMENT** $= X$

do not get

$X = 0$

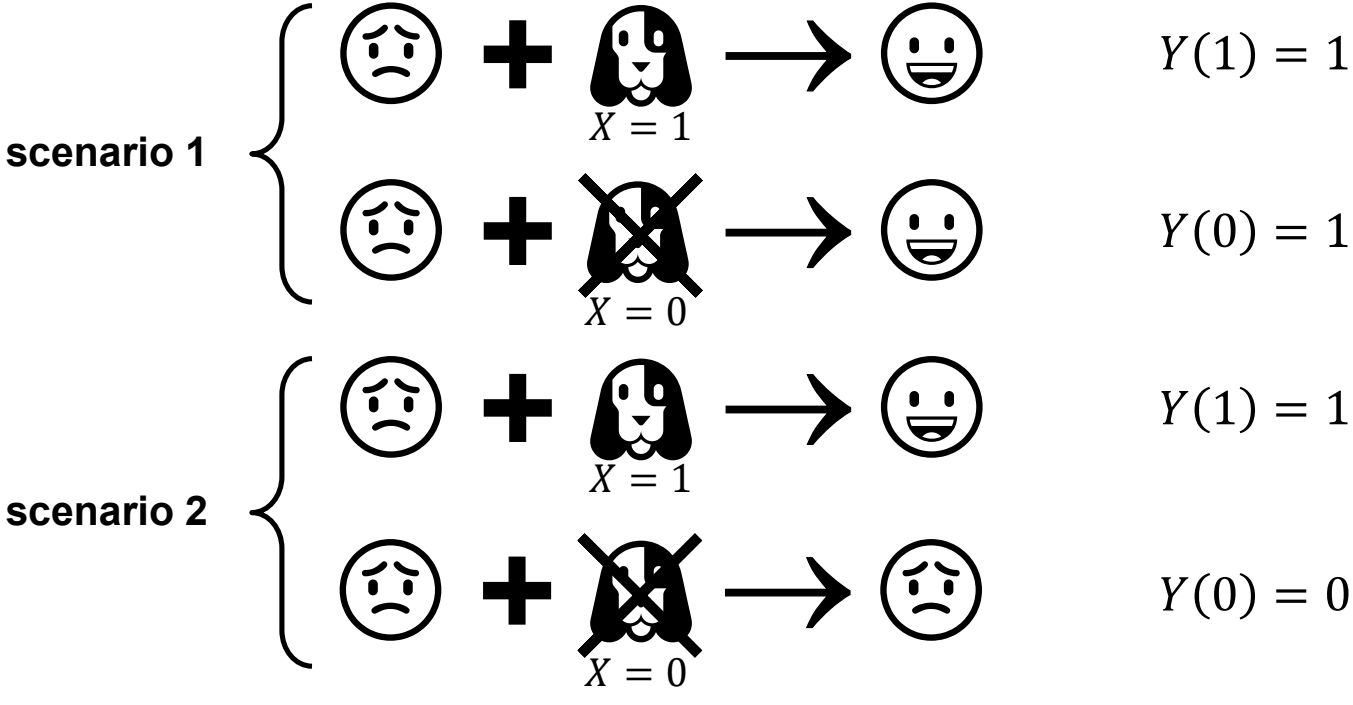
**TREATMENT**

$(X)$

get

$X = 1$

We let $Y(1)$ be the **POTENTIAL OUTCOME OF HAPPINESS** you would observe if you were to get a dog $(X = 1)$.
We let $Y(0)$ be the **POTENTIAL OUTCOME OF HAPPINESS** you would observe if you were to not get a dog $(X = 0)$.



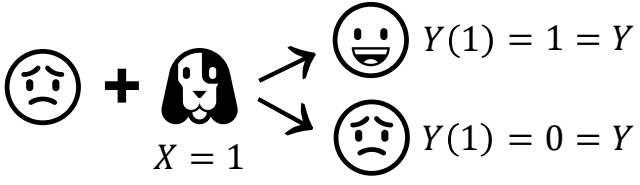

We let $Y(1)$ be the **POTENTIAL OUTCOME OF HAPPINESS** you would observe if you were to get a dog $(X = 1)$.

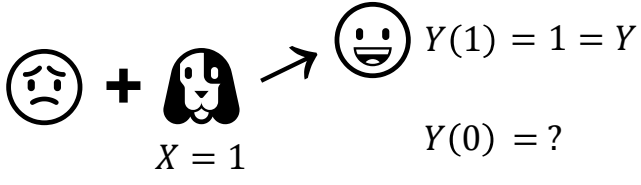
We let $Y(0)$ be the **POTENTIAL OUTCOME OF HAPPINESS** you would observe if you were to not get a dog $(X = 0)$.

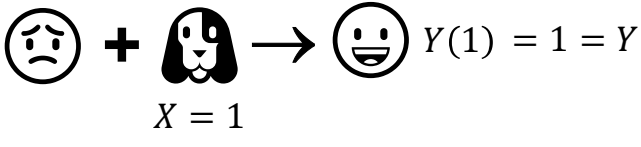
**POTENTIAL OUTCOME**

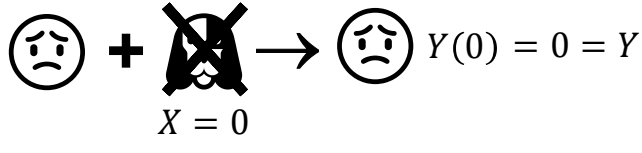
$Y(x)$ denotes what your outcome would be, if you were to take treatment $X = x$.

- A **POTENTIAL OUTCOME** $\boldsymbol{Y(x)}$ is distinct from the **OBSERVED OUTCOME** $\boldsymbol{Y}$ in that not all potential outcomes are observed.
- All potential outcomes can potentially be observed.
- The actually observed potential outcome depends on the given value $x$ of treatment $X$.



Up to now there is only a single individual in the whole population: you.

you 

However, the **POPULATION** consists of many **INDIVIDUALS** or **UNITS**.

population
individuals
units 

Each individual (unit) is tipically associated with one or more variables, referred to as **COVARIATES** $\mathbf{Z}$.

individual (unit) 
age
gender
weight

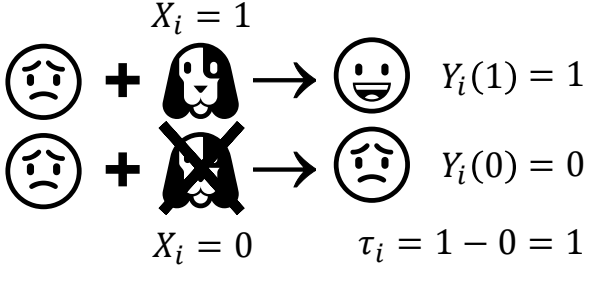
We denote **TREATMENT** $\boldsymbol{X}$, **COVARIATES** $\mathbf{Z}$ and **OUTCOME** $\boldsymbol{Y}$ of the $i^{th}$ individual (unit) as $X_i$, $\mathbf{Z}_i$ and $Y_i$.

$i^{th}$ individual (unit)  $X_i \ \mathbf{Z}_i \ Y_i$

**INDIVIDUAL TREATMENT EFFECT – (ITE)**

The individual treatment effect (ITE) for the $i^{th}$ individual (unit) is defined as follows:

$$\tau_i \triangleq Y_i(1) - Y_i(0)$$



population
individuals
units

$Y_1(x)\ Y_2(x)$ ... ... ... $Y_i(x)$ ... ...

**You**

$Y(x)$ is a random variable (different individuals have different potential outcomes).

$Y_i(x)$ is treated as a non-random variable we are conditioning on so much individualized (and context-specific) information, that we restrict our focus to a single individual (in a specific context) whose potential outcomes are deterministic.

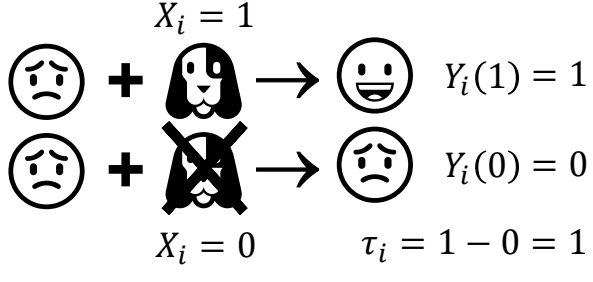
Individual treatment effects (ITEs) are some of the main quantities that we care about in causal inference.

**You** would choose to get a dog because the **CAUSAL EFFECT (ITE)** $\tau_i \triangleq Y_i(1) - Y_i(0)$ of getting a dog on your happiness is positive $\tau_i = 1$.

**INDIVIDUAL TREATMENT EFFECT – (ITE)**

The individual treatment effect (ITE) for the $i^{th}$ individual (unit) is defined as follows:

$$\tau_i \triangleq Y_i(1) - Y_i(0)$$



population
individuals
units

$Y_1(x)\ Y_2(x)$ ... ... ... ... ...

**You**

$Y_i(x)$

$Y(x)$ is a random variable (different individuals have different potential outcomes).

$Y_i(x)$ is treated as a non-random variable we are conditioning on so much individualized (and context-specific) information, that we restrict our focus to a single individual (in a specific context) whose potential outcomes are deterministic.

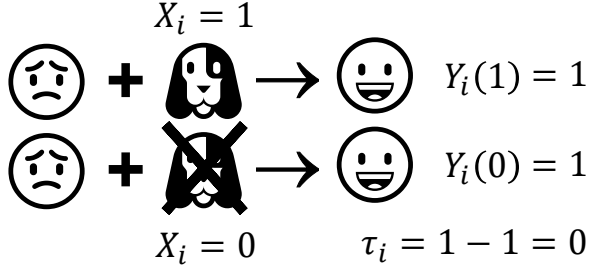
Individual treatment effects (ITEs) are some of the main quantities that we care about in causal inference.

**You** would decide to not get a dog because there is **NO CAUSAL EFFECT** of getting a dog $\tau_i \triangleq Y_i(1) - Y_i(0) = 1 - 1 = 0$.

**INDIVIDUAL TREATMENT EFFECT – (ITE)**

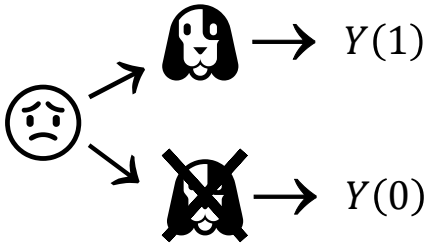
The individual treatment effect (ITE) for the $i^{th}$ individual (unit) is defined as follows:

$$\tau_i \triangleq Y_i(1) - Y_i(0)$$



# PART II

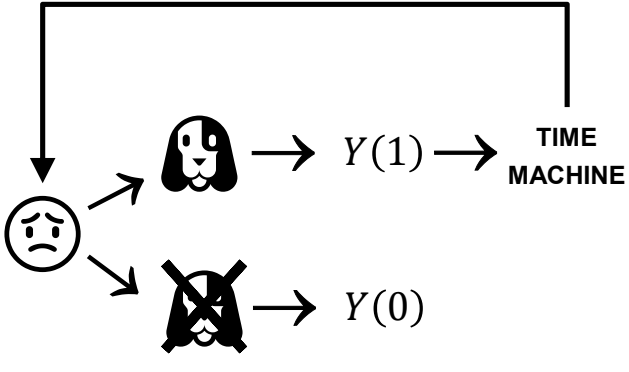
# THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

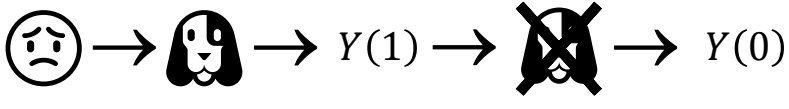



It is impossible to observe all potential outcomes for a given individual (unit).

In particular:

- You cannot observe both $Y(1)$ and $Y(0)$, unless you have a **TIME MACHINE** that would allow you to go back in time and choose the version of treatment $X$ that you didn't take the first time.




- You cannot simply get a dog, observe $Y(1)$, give the dog away, and then observe $Y(0)$ because the second observation will be influenced by all the actions you took between the two observations and anything else that changed since the first observation.

**THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE**

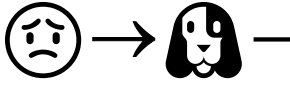
We cannot observe both $Y_i(1)$ and $Y_i(0)$, therefore we cannot observe the causal effect

$$\tau_i \triangleq Y_i(1) - Y_i(0)$$

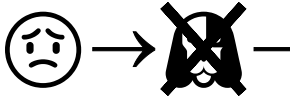
Indeed, we care about making causal claims, which are defined in terms of potential outcomes.

In machine learning, we often only care about predicting the observed outcome $Y$, so there is no need for potential outcomes, which means machine learning does not have to deal with this fundamental problem that we must deal with in causal inference.

- The potential outcomes that you do not (and cannot) observe are known as **COUNTERFACTUALS** because they are counter to fact (reality).
- A potential outcome $Y(x)$ does not become counter to fact (**COUNTERFACTUAL**) until another potential outcome $Y(x')$ is observed.
- Note that there are no counterfactuals or factuals until the outcome is observed. Before that, there are only potential outcomes.



$Y(1)$ **COUNTERFACTUAL**



$Y(0) = 1$ **FACTUAL**

$X = 0$

**THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE**

We cannot observe both $Y_i(1)$ and $Y_i(0)$, therefore we cannot observe the causal effect

$$\tau_i \triangleq Y_i(1) - Y_i(0)$$

We know that we can't access individual treatment effects (ITE), due to the fundamental problem of causal inference.

However, what about **AVERAGE TREATMENT EFFECTS**?

**AVERAGE TREATMENT EFFECT - ATE**

The average treatment effect (ATE) is obtained by taking an average over the ITEs:

$$\tau \triangleq \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y(1) - Y(0)]$$

where we recall that the average is over the individuals "$i$" if $Y_i(x)$ is deterministic.

How would we actually compute the ATE?

Let us assume that data in the right table (Table 2.1) represent the whole population of interest.

The **FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE** can be seen as a **MISSING DATA PROBLEM**, i.e., all question marks (**?**) in Table 2.1 mean that we do not observe the value for the corresponding cell.

Therefore, we can not compute directly the average treatment effect (ATE) or average causal effect (ACE):

$$\tau \triangleq \mathbb{E}[Y_i(1) - Y_i(0)] = ?$$

| $i$ | $X$ | $Y$ | $Y(1)$ | $Y(0)$ | $Y(1) - Y(0)$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | ? | 0 | ? |
| 2 | 1 | 1 | 1 | ? | ? |
| 3 | 1 | 0 | 0 | ? | ? |
| 4 | 0 | 0 | ? | 0 | ? |
| 5 | 0 | 1 | ? | 1 | ? |
| 6 | 1 | 1 | 1 | ? | ? |

Table 2.1

However, we know the following:

$$\tau \triangleq \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y(1) - Y(0)]$$

and linearity of expectation $\mathbb{E}[\cdot]$ gives:

$$\tau \triangleq \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \overset{?}{=} \mathbb{E}[Y|X = 1] - \mathbb{E}[Y|X = 0]$$

**ASSOCIATIONAL DIFFERENCE**

We could be tempted to use the associational difference

$$\mathbb{E}[Y|X = 1] - \mathbb{E}[Y|X = 0]$$

Unfortunately, this is **not true in general**. If it were, that would mean that causation is simply association.

$$\mathbb{E}[Y|X=1] - \mathbb{E}[Y|X=0]$$

is an associational quantity, while

$$\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

is a causal quantity.

| $i$ | $X$ | $Y$ | $Y(1)$ | $Y(0)$ | $Y(1) - Y(0)$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | ? | 0 | ? |
| 2 | 1 | 1 | 1 | ? | ? |
| 3 | 1 | 0 | 0 | ? | ? |
| 4 | 0 | 0 | ? | 0 | ? |
| 5 | 0 | 1 | ? | 1 | ? |
| 6 | 1 | 1 | 1 | ? | ? |

Table 2.1

In general, they are not equal due to **CONFOUNDING**.

The graphical representation of such a difference is depicted in Figure 2.1.

In particular, we say that the covariate $Z$ confounds the effect of $X$ on $Y$, because the of the following path

$$X \leftarrow Z \rightarrow Y$$

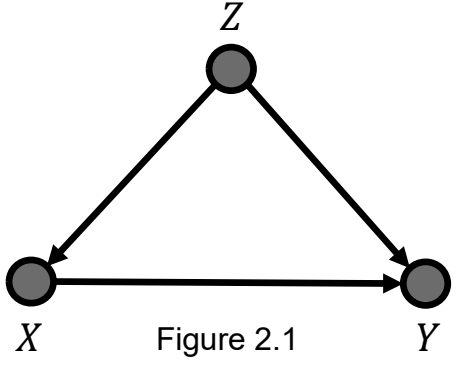



Figure 2.1

But … under which assumptions the following equality holds true?

$$\tau \triangleq \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \mathbb{E}[Y|X = 1] - \mathbb{E}[Y|X = 0]$$

In other terms, when ATE is equal to the associational difference?

What legitimates us to calculate the ATE by taking the average of the $Y(0)$ column, ignoring the question marks, and subtracting that from the average of the $Y(1)$ column, ignoring the question marks?

| $i$ | $X$ | $Y$ | $Y(1)$ | $Y(0)$ | $Y(1) - Y(0)$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | ? | 0 | ? |
| 2 | 1 | 1 | 1 | ? | ? |
| 3 | 1 | 0 | 0 | ? | ? |
| 4 | 0 | 0 | ? | 0 | ? |
| 5 | 0 | 1 | ? | 1 | ? |
| 6 | 1 | 1 | 1 | ? | ? |

Table 2.1

Ignoring question marks (missing data) is known as:

**IGNORABILITY − EXCHANGEABILITY**

$$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X$$

Assuming **IGNORABILITY − EXCHANGEABILITY** means we can ignore how a patient ended up selecting the treatment $X$ she selected and just assuming she was randomly assigned her treatment $X$.

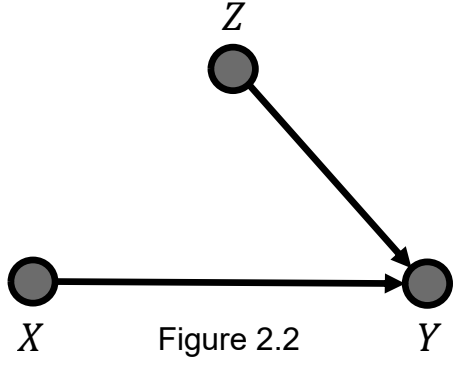



Figure 2.2

**IGNORABILITY – EXCHANGEABILITY** is fundamental because it allows us to reduce the ATE to the associational difference:

$$\tau \triangleq \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

$$= \mathbb{E}[Y(1)|X = 1] - \mathbb{E}[Y(0)|X = 0]$$

$$= \mathbb{E}[Y|X = 1] - \mathbb{E}[Y|X = 0]$$

**IGNORABILITY – EXCHANGEABILITY** means that the groups (treatment/control) are exchangeable in the sense that if they were swapped, the new treatment group would observe the same outcomes as the old treatment group, and the new control group would observe the same outcomes as the old control group.

| $i$ | $X$ | $Y$ | $Y(1)$ | $Y(0)$ | $Y(1) - Y(0)$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | ? | 0 | ? |
| 2 | 1 | 1 | 1 | ? | ? |
| 3 | 1 | 0 | 0 | ? | ? |
| 4 | 0 | 0 | ? | 0 | ? |
| 5 | 0 | 1 | ? | 1 | ? |
| 6 | 1 | 1 | 1 | ? | ? |

Table 2.1

$$\mathbb{E}[Y(1)|X = 1] = \mathbb{E}[Y(1)|X = 0]$$

$$\mathbb{E}[Y(0)|X = 0] = \mathbb{E}[Y(0)|X = 1]$$

which brings to

**IGNORABILITY – EXCHANGEABILITY**

$$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X$$

nearly equivalent to

**MEAN IGNORABILITY – EXCHANGEABILITY**

$$\mathbb{E}[Y(1)|X = x] = \mathbb{E}[Y(1)]$$

$$\mathbb{E}[Y(0)|X = x] = \mathbb{E}[Y(0)] \qquad \forall\, x$$

**IGNORABILITY − EXCHANGEABILITY** is fundamental because it allows us to reduce the ATE to the associational difference:

$$\tau \triangleq \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

$$= \mathbb{E}[Y(1)|X = 1] - \mathbb{E}[Y(0)|X = 0]$$

$$= \mathbb{E}[Y|X = 1] - \mathbb{E}[Y|X = 0]$$

| $i$ | $X$ | $Y$ | $Y(1)$ | $Y(0)$ | $Y(1) - Y(0)$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | ? | 0 | ? |
| 2 | 1 | 1 | 1 | ? | ? |
| 3 | 1 | 0 | 0 | ? | ? |
| 4 | 0 | 0 | ? | 0 | ? |
| 5 | 0 | 1 | ? | 1 | ? |
| 6 | 1 | 1 | 1 | ? | ? |

Table 2.1

**MEAN (IGNORABILITY − EXCHANGEABILITY)** is a weaker assumption than **FULL (IGNORABILITY − EXCHANGEABILITY)** (box below), because it only constrains the first moment of the distribution.

In general, **MEAN (IGNORABILITY − EXCHANGEABILITY)** is sufficient for ATE, but it is common to assume complete independence, as formally represented in the box below.

**IGNORABILITY − EXCHANGEABILITY**

$$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X$$

nearly equivalent to

$$\mathbb{E}[Y(1)|X = 1] = \mathbb{E}[Y(1)|X = 0]$$

$$\mathbb{E}[Y(0)|X = 0] = \mathbb{E}[Y(0)|X = 1]$$

which brings to

**MEAN IGNORABILITY − EXCHANGEABILITY**

$$\mathbb{E}[Y(1)|X = x] = \mathbb{E}[Y(1)]$$

$$\mathbb{E}[Y(0)|X = x] = \mathbb{E}[Y(0)] \qquad \forall\, x$$

**IGNORABILITY − EXCHANGEABILITY** is fundamental because it allows us to reduce the ATE to the associational difference:

$$\tau \triangleq \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

$$= \mathbb{E}[Y(1)|X = 1] - \mathbb{E}[Y(0)|X = 0]$$

$$= \mathbb{E}[Y|X = 1] - \mathbb{E}[Y|X = 0]$$

An important intuition to have about **IGNORABILITY − EXCHANGEABILITY** is that it guarantees that the groups are comparable.

In other words, the **TREATMENT GROUP** $(X = 1)$ and the **CONTROL GROUP** $(X = 0)$ are the same in all relevant aspects other than the treatment $X$.

| $i$ | $X$ | $Y$ | $Y(1)$ | $Y(0)$ | $Y(1) - Y(0)$ |
|---|---|---|---|---|---|
| 1 | 0 | 0 | ? | 0 | ? |
| 2 | 1 | 1 | 1 | ? | ? |
| 3 | 1 | 0 | 0 | ? | ? |
| 4 | 0 | 0 | ? | 0 | ? |
| 5 | 0 | 1 | ? | 1 | ? |
| 6 | 1 | 1 | 1 | ? | ? |

Table 2.1

$$\mathbb{E}[Y(1)|X = 1] = \mathbb{E}[Y(1)|X = 0]$$

$$\mathbb{E}[Y(0)|X = 0] = \mathbb{E}[Y(0)|X = 1]$$

which brings to

**IGNORABILITY − EXCHANGEABILITY**

$$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X$$

nearly equivalent to

**MEAN IGNORABILITY − EXCHANGEABILITY**

$$\mathbb{E}[Y(1)|X = x] = \mathbb{E}[Y(1)]$$

$$\mathbb{E}[Y(0)|X = x] = \mathbb{E}[Y(0)] \qquad \forall\, x$$

The assumption of **IGNORABILITY – EXCHANGEABILITY** allows us to **IDENTIFY CAUSAL EFFECTS**.

- To **IDENTIFY A CAUSAL EFFECT** is to reduce a causal expression to a purely statistical expression.
  - To reduce an expression from one that uses potential outcome notation to one that uses only statistical notation such as $X$, $Z$, $Y$, expectations, and conditioning.
  - We can calculate the **CAUSAL EFFECT** from just the **OBSERVATIONAL DISTRIBUTION** $P(X, Z, Y)$

**IDENTIFIABILITY**

A causal quantity (e.g. $\mathbb{E}[Y(x)]$) is identifiable if we can compute it from a purely statistical quantity (e.g. $\mathbb{E}[Y|X = x]$).

**IGNORABILITY – EXCHANGEABILITY** is extremely important, but how realistic of an assumption is it?

**IGNORABILITY – EXCHANGEABILITY**

$$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X$$

**COMPLETELY UNREALISTIC**, confounding is likely to happen in most data we observe.

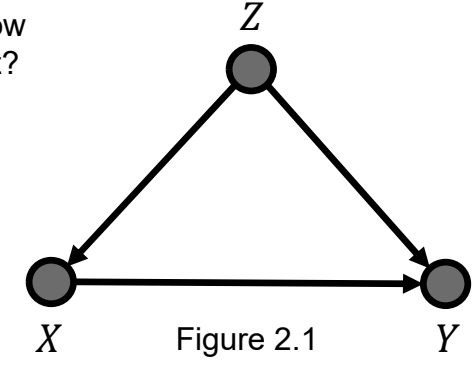


Figure 2.1

We can make the **IGNORABILITY – EXCHANGEABILITY** assumption realistic by performing **RANDOMIZED EXPERIMENTS**, which force the treatment $X$ to not be caused by anything but a coin toss, so then we have the causal structure shown in Figure 2.2.

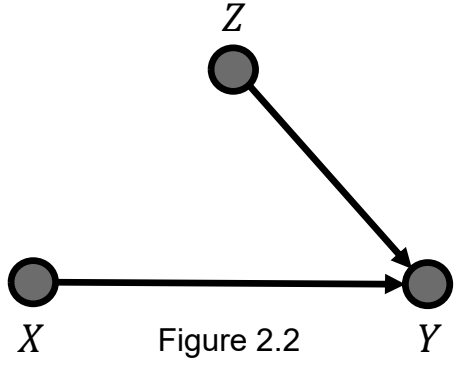


Figure 2.2

**IDENTIFIABILITY**

A causal quantity (e.g. $\mathbb{E}[Y(x)]$) is identifiable if we can compute it from a purely statistical quantity (e.g. $\mathbb{E}[Y|X = x]$).

**IGNORABILITY – EXCHANGEABILITY** is extremely important, but how realistic of an assumption is it?

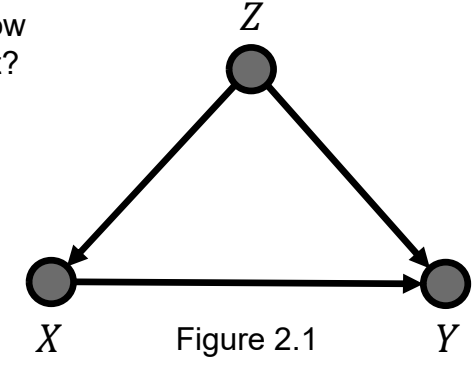


Figure 2.1

**IGNORABILITY – EXCHANGEABILITY**

$$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X$$

**COMPLETELY UNREALISTIC**, confounding is likely to happen in most data we observe.

In **OBSERVATIONAL DATA**, it is unrealistic to assume that the groups (treatment and control) are exchangeable. In other words, there is no reason to expect that the groups (treatment and control) are the same in all relevant variables $Z \in \mathbf{Z}$ other than the treatment $X$.

However, if we control for relevant variables by conditioning, then maybe the groups will be exchangeable.

- What do we mean by "relevant variables"?
- For now, let's just say they are all of the covariate variables $\mathbf{Z}$.

**CONDITIONAL EXCHANGEABILITY – UNCONFOUNDEDNESS**

$$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X \mid \mathbf{Z}$$ where $\mathbf{Z}$ are the covariate variables.

Although the treatment $X$ and potential outcomes $Y(1)$ and $Y(0)$ may be unconditionally associated (due to confounding), within levels of $Z$, they are not associated.

No confounding within levels of $Z$ because **CONTROLLING FOR** $Z$ makes the treatment group $(X = 1)$ and the control group $(X = 0)$ comparable.

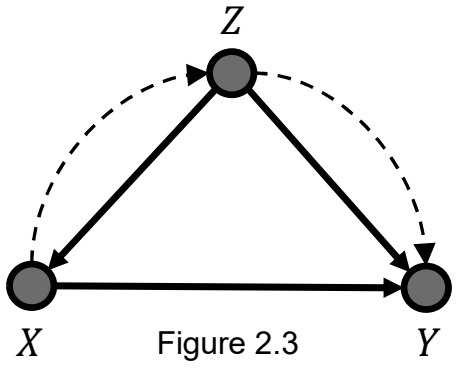


Figure 2.3

We do not have **IGNORABILITY – EXCHANGEABILITY** in the data because $Z$ is a common cause of $X$ and $Y$.

**NON-CAUSAL ASSOCIATION** between $X$ and $Y$, flows along the following path:

$$X \leftarrow Z \rightarrow Y$$

However, **CONDITIONAL EXCHANGEABILITY – UNCONFOUNDEDNESS** holds in the data.

Indeed, when conditioning on $Z$, non-causal association between $X$ and $Y$ no longer exists.

Non-causal association is "**BLOCKED**" at $Z$ by conditioning on $Z$.

**CONDITIONAL EXCHANGEABILITY – UNCONFOUNDEDNESS** is the main assumption necessary for causal inference.

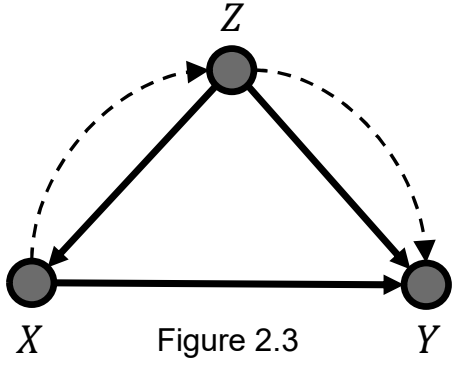


Figure 2.3

**CONDITIONAL EXCHANGEABILITY – UNCONFOUNDEDNESS**

$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X \mid \mathbf{Z}$ where $\mathbf{Z}$ are the covariate variables.

We can now identify the causal effect within levels of $\mathbf{Z}$, just like we did with (unconditional) ignorability – exchangeability:

$$\begin{aligned}\mathbb{E}[Y(1) - Y(0) \mid \mathbf{Z}] &= \mathbb{E}[Y(1)|\mathbf{Z}] - \mathbb{E}[Y(0)|\mathbf{Z}] \\ &= \mathbb{E}[Y(1)|X = 1, \mathbf{Z}] - \mathbb{E}[Y(0)|X = 0, \mathbf{Z}] \\ &= \mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\end{aligned}$$

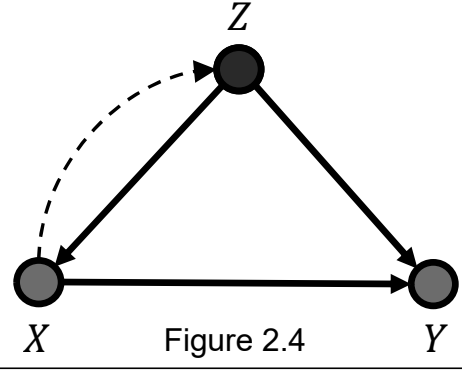


Figure 2.4

However, **CONDITIONAL EXCHANGEABILITY – UNCONFOUNDEDNESS** holds in the data.

Indeed, when conditioning on $Z$, non-causal association between $X$ and $Y$ no longer exists.

Non-causal association is "**BLOCKED**" at $Z$ by conditioning on $Z$.

**CONDITIONAL EXCHANGEABILITY – UNCONFOUNDEDNESS** is the main assumption necessary for causal inference.

**CONDITIONAL EXCHANGEABILITY – UNCONFOUNDEDNESS**

$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X \mid \mathbf{Z}$ where $\mathbf{Z}$ are the covariate variables.

We can now identify the causal effect within levels of $\mathbf{Z}$, just like we did with (unconditional) ignorability – exchangeability:

$$\begin{aligned}\mathbb{E}[Y(1) - Y(0) \mid \mathbf{Z}] &= \mathbb{E}[Y(1)|\mathbf{Z}] - \mathbb{E}[Y(0)|\mathbf{Z}] \\ &= \mathbb{E}[Y(1)|X = 1, \mathbf{Z}] - \mathbb{E}[Y(0)|X = 0, \mathbf{Z}] \\ &= \mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\end{aligned}$$

Conditional exchangeability is a core assumption for causal inference and goes by many names:

- unconfoundedness,
- conditional ignorability,
- no unobserved confounding,
- selection on observables,
- no omitted variable bias
- etc…

I will use the term

**UNCONFOUNDEDNESS**

If we want the marginal effect that we had before when assuming (unconditional) ignorability – exchangeability, we can get that by simply marginalizing out $\mathbf{Z}$ as follows

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{\mathbf{Z}}\mathbb{E}[Y(1) - Y(0)|\mathbf{Z}]$$
$$= \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\big]$$

We can now introduce an important result for causal inference, that we will formally prove in next lectures.

**ADJUSTMENT FORMULA**

Given the assumptions of unconfoundedness, positivity, consistency, and no interference, we can identify the ATE:

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\big]$$

We moved from the assumption of ignorability – exchangeability to that of unconfoundedness because it seems more realistic.

However, we often cannot know for certain if unconfoundedness holds!!!

Conditional exchangeability is a core assumption for causal inference and goes by many names:

- unconfoundedness,
- conditional ignorability,
- no unobserved confounding,
- selection on observables,
- no omitted variable bias
- etc…

I will use the term

**UNCONFOUNDEDNESS**

There may be some **UNOBSERVED CONFOUNDERS** ($W$ in Figure 2.3B) that are not part of $\mathbf{Z} = \{\boldsymbol{M}\}$, meaning **UNCONFOUNDEDNESS IS VIOLATED**.

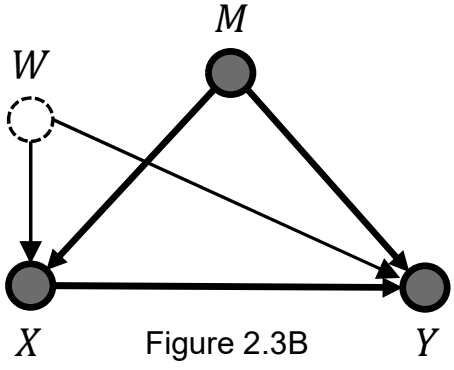


Figure 2.3B

Fortunately, that is not a problem in **RANDOMIZED EXPERIMENTS**.

Unfortunately, it is something that we must always be conscious of in **OBSERVATIONAL DATA**.

Intuitively, the best thing we can do is to observe and fit (adjust for) as many covariates into $\mathbf{Z}$ as possible to try to ensure unconfoundedness.

However, we will see in next lectures, that it is not necessarily true that conditioning on more covariates $\mathbf{Z}$ always helps our causal estimates to be less biased.

Indeed, it can be the case we obtain more biased estimates when adjusting for the "wrong covariates" in $\mathbf{Z} = \{\boldsymbol{M}, \boldsymbol{N}\}$.

In Figure 2.3C, adjusting for $M$ provides unbiased estimates, while adjusting for $N$ results in biased estimates.

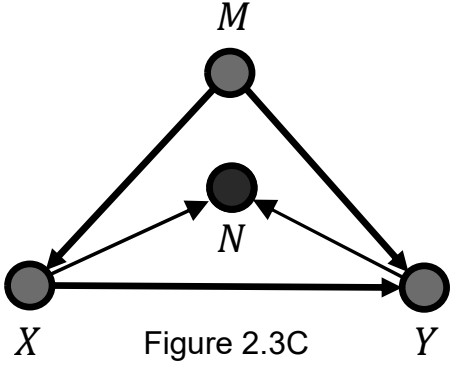


Figure 2.3C

Conditioning on many covariates is attractive for achieving unconfoundedness, but it can be detrimental for another reason that has to do with **POSITIVITY**.

Positivity is the condition that all subgroups of the data with different value $\boldsymbol{z}$ for covariates $\mathbf{Z}$ have some probability of receiving any value of treatment $X$.

**POSITIVITY – OVERLAP – COMMON SUPPORT**

For all values $\boldsymbol{z}$ of covariates $\mathbf{Z}$ present in the population of interest (i.e., $\boldsymbol{z}$ such that $\boldsymbol{P}(\mathbf{Z} = \boldsymbol{z}) > 0$)

$$0 < P(X = 1|\mathbf{Z} = \boldsymbol{z}) < 1$$

If we have a **POSITIVITY VIOLATION**, then we will be conditioning on a zero probability event.

To clearly see how a positivity violation translates to division by zero, let's rewrite the right-hand side of

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\big]$$

in the case of discrete covariate variables $\mathbf{Z}$, to obtain

**POSITIVITY IS ESSENTIAL TO DEFINE CAUSAL EFFECT!!!**

$$= \sum_{\boldsymbol{z}} P(\mathbf{Z} = \boldsymbol{z}) \left( \sum_{y} y\, P(Y = y|X = 1, \mathbf{Z} = \boldsymbol{z}) - \sum_{y} y\, P(Y = y|X = 0, \mathbf{Z} = \boldsymbol{z}) \right)$$

$$= \sum_{\boldsymbol{z}} P(\mathbf{Z} = \boldsymbol{z}) \left( \sum_{y} y \frac{P(Y = y, X = 1, \mathbf{Z} = \boldsymbol{z})}{P(X = 1|\mathbf{Z} = \boldsymbol{z})P(\mathbf{Z} = \boldsymbol{z})} - \sum_{y} y \frac{P(Y = y, X = 0, \mathbf{Z} = \boldsymbol{z})}{P(X = 0|\mathbf{Z} = \boldsymbol{z})P(\mathbf{Z} = \boldsymbol{z})} \right)$$

**POSITIVITY – OVERLAP – COMMON SUPPORT**

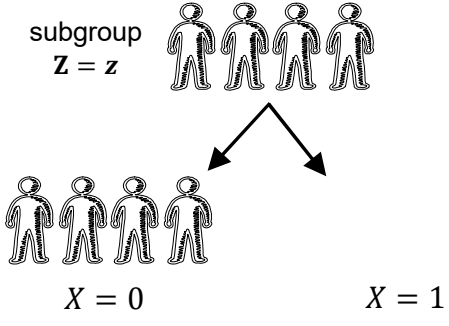
It wouldn't make any sense to be able to estimate the causal effect of treatment $(X = 1)$ vs. control $(X = 0)$ in subgroup $\mathbf{Z} = \boldsymbol{z}$, since we see only treatment $(X = 1)$ or only control $(X = 0)$, i.e., we never see the alternative in subgroup $\mathbf{Z} = \boldsymbol{z}$.

For all values $\boldsymbol{z}$ of covariates $\mathbf{Z}$ present in the population of interest (i.e., $\boldsymbol{z}$ such that $\boldsymbol{P}(\mathbf{Z} = \boldsymbol{z}) > 0$)

$$0 < P(X = 1|\mathbf{Z} = \boldsymbol{z}) < 1$$

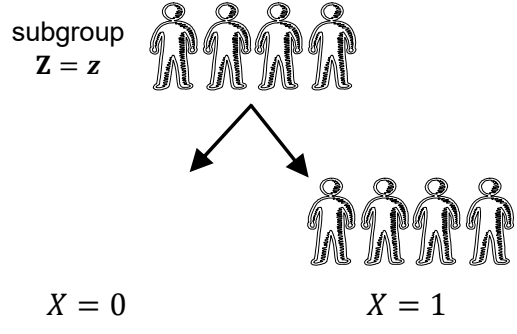
- $if\ \exists\ \boldsymbol{z} : P(X = 1|\mathbf{Z} = \boldsymbol{z}) = 0$

  All units belonging to subgroup $\boldsymbol{Z} = \boldsymbol{z}$ do not receive the treatment $(X = 0)$.




- $if\ \exists\ \boldsymbol{z} : P(X = 1|\mathbf{Z} = \boldsymbol{z}) = 1$

  All units belonging to subgroup $\mathbf{Z} = \boldsymbol{z}$ do receive the treatment $(X = 1)$.



Positivity is also referred to as **OVERLAP**, in the sense we want the covariate distribution of the treatment group $(X = 1)$

$$P(\mathbf{Z}|X = 1)$$

to overlap with the covariate distribution of the control group $(X = 0)$

$$P(\mathbf{Z}|X = 0)$$

This is why another common alias for positivity is **COMMON SUPPORT**.

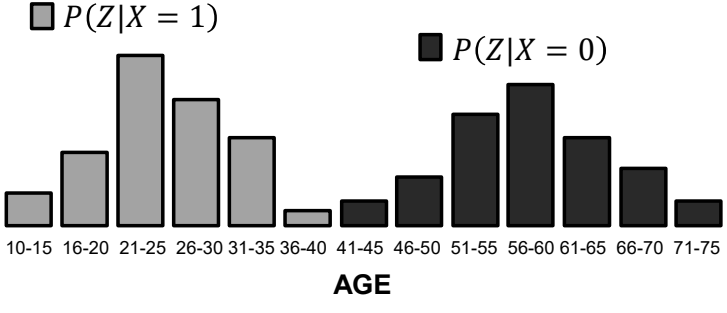
**POSITIVITY – OVERLAP – COMMON SUPPORT**

For all values $\boldsymbol{z}$ of covariates $\mathbf{Z}$ present in the population of interest (i.e., $\boldsymbol{z}$ such that $\boldsymbol{P}(\mathbf{Z} = \boldsymbol{z}) > 0$)

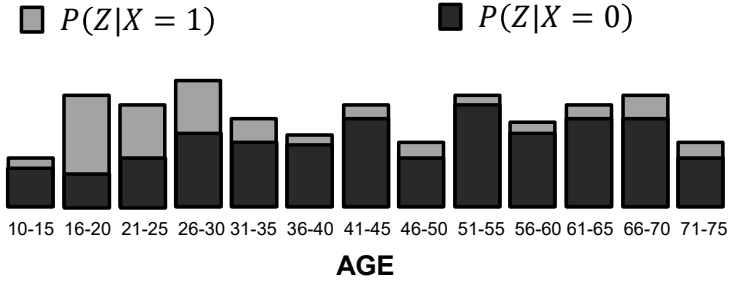
$$0 < P(X = 1|\mathbf{Z} = \boldsymbol{z}) < 1$$

In the case where the covariates $\mathbf{Z}$ consist of a single variable $Z$ (i.e., Age) we have the following graphical representation of positivity, overlap, and common support:

**NO POSITIVITY – NO OVERLAP – NO COMMON SUPPORT**




**POSITIVITY – OVERLAP – COMMON SUPPORT**



adjusting (conditioning) on more covariates $\mathbf{Z}$

could lead to ⇘ higher chance of satisfying unconfoundedness

could lead to ⇙ higher chance of violating positivity

**POSITIVITY – OVERLAP – COMMON SUPPORT**

For all values $\boldsymbol{z}$ of covariates $\mathbf{Z}$ present in the population of interest (i.e., $\boldsymbol{z}$ such that $\boldsymbol{P}(\mathbf{Z} = \boldsymbol{z}) > 0$)

$$0 < P(X = 1|\mathbf{Z} = \boldsymbol{z}) < 1$$

increase the "*dimension*" of the covariates $\mathbf{Z}$ ⇨ makes the subgroups for any level $\boldsymbol{z}$ of the covariates $\mathbf{Z}$ smaller ⇨ **CURSE OF DIMENSIONALITY**

As each subgroup $\mathbf{Z} = \boldsymbol{z}$ gets smaller, there is a higher and higher chance that either the whole subgroup $\mathbf{Z} = \boldsymbol{z}$ will have treatment $(X = 1)$ or the whole subgroup $\mathbf{Z} = \boldsymbol{z}$ will have control $(X = 0)$.

size of any subgroup $\mathbf{Z} = \boldsymbol{z}$ equal to 1 ⇨ positivity is guaranteed to not hold

adjusting (conditioning) on more covariates $\mathbf{Z}$

could lead to

higher chance of satisfying unconfoundedness

could lead to

higher chance of violating positivity

demanding too much from models and getting very bad behavior in return

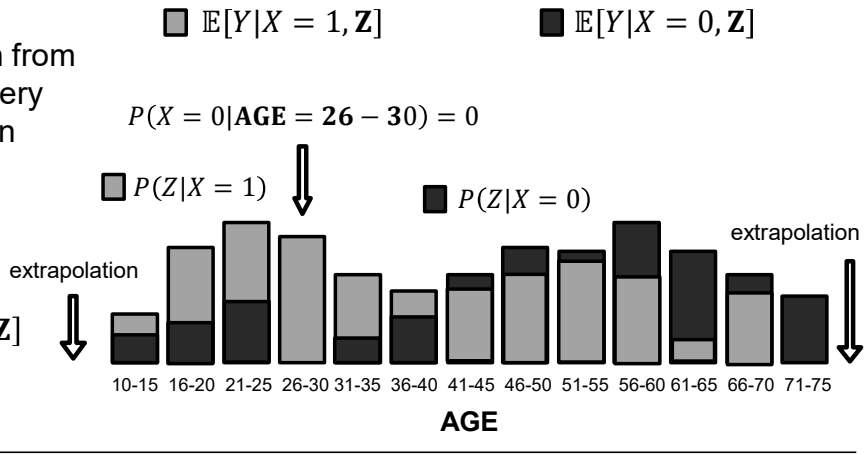
fit a model to $\mathbb{E}[Y|X,\mathbf{Z}]$ using the available data $(x, y, \boldsymbol{z})$

**POSITIVITY – OVERLAP – COMMON SUPPORT**

For all values $\boldsymbol{z}$ of covariates $\mathbf{Z}$ present in the population of interest (i.e., $\boldsymbol{z}$ such that $\boldsymbol{P}(\mathbf{Z} = \boldsymbol{z}) > 0$)

$$0 < P(X = 1|\mathbf{Z} = \boldsymbol{z}) < 1$$

inputs to the model $\mathbb{E}[Y|X,\mathbf{Z}]$ are $(x, \boldsymbol{z})$ pairs, while the output is the outcome $y$.



Another assumption is that of **NO INTERFERENCE**.

**NO INTERFERENCE**

The outcome $Y_i$ of each unit "$i$" is unaffected by anyone else's treatment $X_j, j \neq i$.

$$Y_i(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{n-1}, x_n) = Y_i(x_i)$$

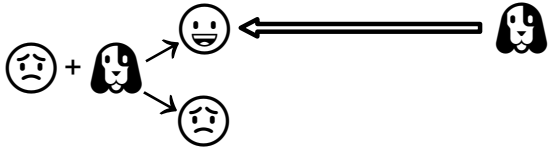
Rather, the outcome $Y_i$ of each unit "$i$" is only a function of treatment $X_i$.

We have implicitly made this assumption till now.

This assumption could be violated.

**TREATMENT** $X =$ "**GET A DOG**"

**OUTCOME** $Y =$ "**MY HAPPYNESS**"

Violations of the no interference assumption are almost sure in network data.

The last assumption is **CONSISTENCY**.

**CONSISTENCY**

If the treatment is $X$, then the observed outcome $Y$ is the potential outcome under treatment $X$. Formally,

$$X = x \Longrightarrow Y = Y(x)$$

We could write this also as follows:

$$Y = Y(X)$$

**SUTVA**

The **S**table **U**nit-**T**reatment **V**alue **A**ssumption (**SUTVA**) is satisfied if unit (individual) $i$'s outcome $Y_i$ is simply a function of unit $i$'s treatment $X_i$.

SUTVA is a combination of consistency and no interference (and also deterministic potential outcomes).

It might seem like consistency is obviously true, but that is not always the case.

$X = \begin{cases} 1 & \text{"get a dog"} \\ 0 & \text{"do not get a dog"} \end{cases}$

$Y = 1$ because I needed an energetic friend

$Y = 0$ because I got an old and low energy dog

"no multiple versions of treatment."

$X = 1$ $Y(1)$ is not well defined, since it will be 1 or 0, depending on something that is not captured by the treatment $X$ specification.

# PART III

# TYING IT ALL TOGETHER

The following assumptions are all needed for solving the problem of causal inference:

**CONDITIONAL EXCHANGEABILITY – UNCONFOUNDEDNESS**

$\big(Y(1) - Y(0)\big) \perp\!\!\!\perp X \mid \mathbf{Z}$ where $\mathbf{Z}$ are the covariate variables.

**POSITIVITY – OVERLAP – COMMON SUPPORT**

For all values $\boldsymbol{z}$ of covariates $\mathbf{Z}$ present in the population of interest (i.e., $\boldsymbol{z}$ such that $\boldsymbol{P}(\mathbf{Z} = \boldsymbol{z}) > 0$)

$$0 < P(X = 1 | \mathbf{Z} = \boldsymbol{z}) < 1$$

**SUTVA**

**NO INTERFERENCE**

The outcome $Y_i$ of each unit "$i$" is unaffected by anyone else's treatment $X_j, j \neq i$.

$$Y_i(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{n-1}, x_n) = Y_i(x_i)$$

**CONSISTENCY**

If the treatment is $X$, then the observed outcome $Y$ is the potential outcome under treatment $X$. Formally,

$$X = x \Longrightarrow Y = Y(x)$$

We could write this also as follows:

$$Y = Y(X)$$

**ADJUSTMENT FORMULA**

Given the assumptions of unconfoundedness, positivity, consistency, and no interference, we can identify the ATE:

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\big]$$

**NO INTERFERENCE** justifies that the quantity we should be looking at for causal inference is

We now come back to give a formal proof of the **ADJUSTMENT FORMULA**.

**AVERAGE TREATMENT EFFECT - ATE**

The average treatment effect (ATE) is obtained by taking an average over the ITEs:

$$\tau \triangleq \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y(1) - Y(0)]$$

where we recall that the average is over the individuals "$i$" if $Y_i(x)$ is deterministic.

All of these assumptions tie together give us identifiability of the ATE.

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \qquad \text{(linearity of expectations)}$$

$$= \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y(1)|\mathbf{Z}] - \mathbb{E}[Y(0)|\mathbf{Z}]\big] \qquad \text{(law of iterated expectations)}$$

$$= \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y(1)|X = 1, \mathbf{Z}] - \mathbb{E}[Y(0)|X = 0, \mathbf{Z}]\big] \qquad \text{(unconfoundedness and positivity)}$$

$$= \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\big] \qquad \text{(consistency)}$$

We need to introduce some terminology that will help clarify the discussion.

**ESTIMAND**

An estimand is a quantity that we want to estimate.

$$\mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X=1,\mathbf{Z}]-\mathbb{E}[Y|X=0,\mathbf{Z}]\big]$$

estimand we care about for estimating the ATE

**ESTIMATE**

An approximation of some estimand, which we get using data.

Given an estimand $\alpha$, we let $\hat{\alpha}$ be its estimate.

**ESTIMATOR**

A function that maps a dataset to an estimate of the estimand.

**ESTIMATION**

The process that we use to go from data + estimand to a concrete number is known as estimation.

- **CAUSAL ESTIMAND** refers to any estimand that contains a potential outcome in it.
- **STATISTICAL ESTIMAND** refers to any estimand that does not contain a potential outcome in it.

In the following formula

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{\mathbf{Z}}\Big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\Big]$$

$\mathbb{E}[Y(1) - Y(0)]$ is the **CAUSAL ESTIMAND** that we are interested in

To actually estimate this causal estimand, we must translate it into a **STATISTICAL ESTIMAND**

$$\mathbb{E}_{\mathbf{Z}}\Big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\Big]$$

**IDENTIFICATION**

The process of moving from a causal estimand to an equivalent statistical estimand.

**ESTIMATION**

The process of moving from a statistical estimand to an estimate.

How do we do when we go to actually estimate quantities such as

$$\mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X=1,\mathbf{Z}]-\mathbb{E}[Y|X=0,\mathbf{Z}]\big]$$

We will often use a model (e.g., linear regression or some more flexible predictor from machine learning) in place of the conditional expectations

$$\mathbb{E}[Y|X=x,\mathbf{Z}=\boldsymbol{z}]$$

We will refer to estimators that use models like this as **MODEL-ASSISTED ESTIMATORS**.

We now need to discuss estimation.

We now give an example complete with estimation(*).

$Y$

"systolic blood pressure"

$X$

"daily sodium intake"

Sodium intake is a continuous variable, so to easily apply

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\big]$$

which is specified for binary treatment, we binarize $Y$

$$Y = 1 \; if \; daily \; sodium \; intake \geq 3.5 \; gr.$$

$$Y = 0 \; if \; daily \; sodium \; intake < 3.5 \; gr.$$

We estimate the causal effect of sodium intake on blood pressure.

The data also include the following covariates $\mathbf{Z}$ for each individual:

- Age
- Amount of protein in the urine

Because we are using data from a simulation, we know that the true ATE of sodium on blood pressure is 1.05.

$$\mathbb{E}[Y(1) - Y(0)] = 1.05$$

* Miguel Angel Luque-Fernandez, Michael Schomaker, Daniel Redondo-Sanchez, Maria Jose Sanchez Perez, Anand Vaidya, and Mireille E Schnitzer. 'Educational Note: Paradoxical collider effect in the analysis of non-communicable disease epidemiological data: a reproducible illustration and web application'. In: International Journal of Epidemiology 48.2 (Dec. 2018), pp. 640–653. doi: 10.1093/ije/dyy275 (cited on pages 16, 45).

How do we actually estimate the ATE?

1) We assume consistency, positivity, and unconfoundedness given $\mathbf{Z}$. This means that ATE is identified by

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\big]$$

2) We then take that outer expectation over $\mathbf{Z} = \{\text{Age}, \text{Amount of proteine}\}$ and replace it with an empirical mean over the data, giving us the following:

$$\frac{1}{n}\sum_{i=1}^{n}\big[\mathbb{E}[Y|X = 1, \mathbf{Z} = \boldsymbol{z_i}] - \mathbb{E}[Y|X = 0, \mathbf{Z} = \boldsymbol{z_i}]\big]$$

3) To complete our estimator, we then fit a machine learning model to the conditional expectation $\mathbb{E}[Y|x, \boldsymbol{z}]$.

We can plug in any machine learning model for $\mathbb{E}[Y|x, \boldsymbol{z}]$, which gives us a **MODEL-ASSISTED ESTIMATOR**.

We use linear regression, which works out nicely since blood pressure is generated as a linear combination of other variables (daily sodium intake, age and amount of protein in the urine).

$$\mathbb{E}[Y(1) - Y(0)] = 0.856$$

$$\%error = \frac{|0.856 - 1.05|}{1.05} \times 100\% = 18\%$$

So, if we use linear regression, which works out nicely since blood pressure is generated as a linear combination of other variables (daily sodium intake, age and amount of protein in the urine).

$$\mathbb{E}[Y(1) - Y(0)] = 0.856 \qquad \%error = \frac{|0.856 - 1.05|}{1.05} \times 100\% = 18\%$$

However, if we were to naively regress $Y$ on only $X$ (daily sodium intake) we would get

$$\mathbb{E}[Y(1) - Y(0)] = 5.37 \qquad \%error = \frac{|5.37 - 1.05|}{1.05} \times 100\% = 411\%$$

All of the above is obtained using the **ADJUSTMENT FORMULA** with **MODEL-ASSISTED ESTIMATION**, where:

1) we fit a model $\mathbf{M}$ for the conditional expectation $\mathbb{E}[Y|x, \boldsymbol{z}]$
2) we take an empirical mean over $\mathbf{Z}$ using model $\mathbf{M}$

**ADJUSTMENT FORMULA**

Given the assumptions of unconfoundedness, positivity, consistency, and no interference, we can identify the ATE:

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_{\mathbf{Z}}\big[\mathbb{E}[Y|X = 1, \mathbf{Z}] - \mathbb{E}[Y|X = 0, \mathbf{Z}]\big]$$

# Potential Outcomes

Alessio Zanga, Fabio Stella

March 27, 2021

[1]:
```
%%capture
!pip install networkx numpy pandas statsmodels
```

[2]:
```
import networkx as nx
import numpy as np
import pandas as pd
import statsmodels.api as sm
from typing import Set
```

## 1 Potential Outcomes

By Alessio Zanga and Fabio Stella

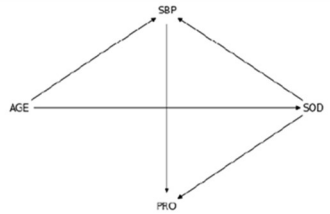
### 1.1 Abstract

This notebook illustrates a step-by-step example on estimating the average causal effect (ACE) of sodium on blood pressure following Luque-Fernandez et al. (2018). The main focus of this work is to highlight the differences between a naive estimate and the adjusted estimate.

### 1.2 Introduction

Exceeding the recommendations for 24-h dietary **sodium** (SOD) intake is associated with increased levels of **systolic blood pressure** (SBP). Furthermore, with advancing **age**, the adaptive mechanism responsible for maintaining the composition and volume of the extracellular fluid is compromised. Age is a common cause of both high systolic blood pressure and impaired sodium homeostasis, acting as a confounder. However, high levels of 24-h excretion of **urinary protein** (PRO) are caused by sustained high SBP and increased dietary SOD, acting as a collider.

[3]:
```
G = nx.DiGraph()
G.add_edges_from([("SOD", "SBP"), ("AGE", "SOD"), ("AGE", "SBP"), ("SOD",␣
↪"PRO"), ("SBP", "PRO")])
nx.draw_circular(G, node_size=1000, node_color="white", with_labels=True)
```




Assuming **linear relationships** between the variables and **gaussian noise**, a data generation process that is consistent with the represented causal graph follows directly.

[4]:
```
def sample_data(size: int = int(1e6), seed: int = 31):
    # Set random generator seed for results reproducibility
    np.random.seed(seed)
    # Sample age with mean 65 and std 5
    age = np.random.normal(65, 5, size)
    # Sample sodium with additive noise
    sod = 0.056 * age + np.random.normal(0, 1, size)
    # Binarize sodium following cutoff
    sod = (sod > 3.5).astype(int)
    # Sample systolic blood pressure
    sbp = 1.05 * sod + 2 * age + np.random.normal(0, 1, size)
    # Sample urinary protein
    pro = 0.4 * sod + 0.3 * sbp + np.random.normal(0, 1, size)
    # Create a dataframe from sampled variables
    return pd.DataFrame({"AGE": age, "SOD": sod, "SBP": sbp, "PRO": pro})
```

Here, the true ACE is given by the coefficient 1.05 that assign SBP a value given SOD.

[5]:
```
data = sample_data()
data.describe()
```

[5]:

| | AGE | SOD | SBP | PRO |
|---|---|---|---|---|
| count | 1000000.000000 | 1000000.000000 | 1000000.000000 | 1000000.000000 |
| mean | 65.001134 | 0.552995 | 130.583161 | 39.396720 |
| std | 4.993518 | 0.497184 | 10.161539 | 3.263477 |
| min | 40.964418 | 0.000000 | 82.816817 | 24.421988 |
| 25% | 61.627703 | 0.000000 | 123.720103 | 37.190745 |
| 50% | 65.009236 | 1.000000 | 130.597297 | 39.402344 |
| 75% | 68.372794 | 1.000000 | 137.445514 | 41.602773 |
| max | 87.717317 | 1.000000 | 176.708462 | 54.573313 |

### 1.3 Methods

Assuming consistency, positivity, and unconfoundedness, the average causal effect (ACE) is defined:

$\tau = \mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_Z[\mathbb{E}[Y(1) - Y(0)|Z]] = \mathbb{E}_Z[\mathbb{E}[Y|X = 1, Z] - \mathbb{E}[Y|X = 0, Z]]$

Replacing expectations with empirical mean:

$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_Z[\mathbb{E}[Y|X = 1, Z = z_i] - \mathbb{E}[Y|X = 0, Z = z_i]]$

It is possible to choose any machine learning model for $\mathbb{E}[Y|X, Z]$, such as a linear regression model, which works out nicely since blood pressure is generated as a linear combination of other variables.

[6]:
```
def ACE(data: pd.DataFrame, X: str, Y: str, Z: Set[str]):
    # Define the regresion model formula
    formula = f"{Y} ~ {X}"
    if len(Z) != 0: formula += "+" + "+".join(Z)
    # Fit Ordinary Least Square regression model
    estimator = sm.OLS.from_formula(formula, data).fit()
    # Compute potential outcomes by fixing X
    Y1 = estimator.predict(data.assign(**{X: 1}))
    Y0 = estimator.predict(data.assign(**{X: 0}))
    # Compute average causal effect
    return np.mean(Y1 - Y0)
```

### 1.4 Results

The true ACE is:

[7]:
```
ace = 1.05
```

The estimated ACE adjusting for AGE and PRO is:

[8]:
```
t = ACE(data, X = "SOD", Y = "SBP", Z = ["AGE", "PRO"])
f"Estimated ACE: {t:.3}, Relative Error: {(np.abs((t-ace)/ace*100)):.4}%"
```

[8]:
```
'Estimated ACE: 0.856, Relative Error: 18.46%'
```

While the naive estimated ACE without adjustment is:

[9]:
```
t = ACE(data, X = "SOD", Y = "SBP", Z = [])
f"Estimated ACE: {t:.3}, Relative Error: {(np.abs((t-ace)/ace*100)):.4}%"
```

[9]:
```
'Estimated ACE: 5.37, Relative Error: 411.5%'
```

### 1.5 Conclusions

Applying a naive regression model without adjustment leads to an estimated ACE which is four times off. A regression model with a valid adjustment set reduce the relative error to only 18%.

The associated notebook is available here

https://colab.research.google.com/github/AlessioZanga/CaMo/blob/develop/examples/potential_outcomes.ipynb

**COLAB EXECUTABLE**