



THE UNIVERSITY
of EDINBURGH

Methods for Causal Inference

Lecture 5: Rubin's framework, propensity score, IPTW

Ava Khamseh

School of Informatics
2023-2024

Potential Outcomes: Assumptions

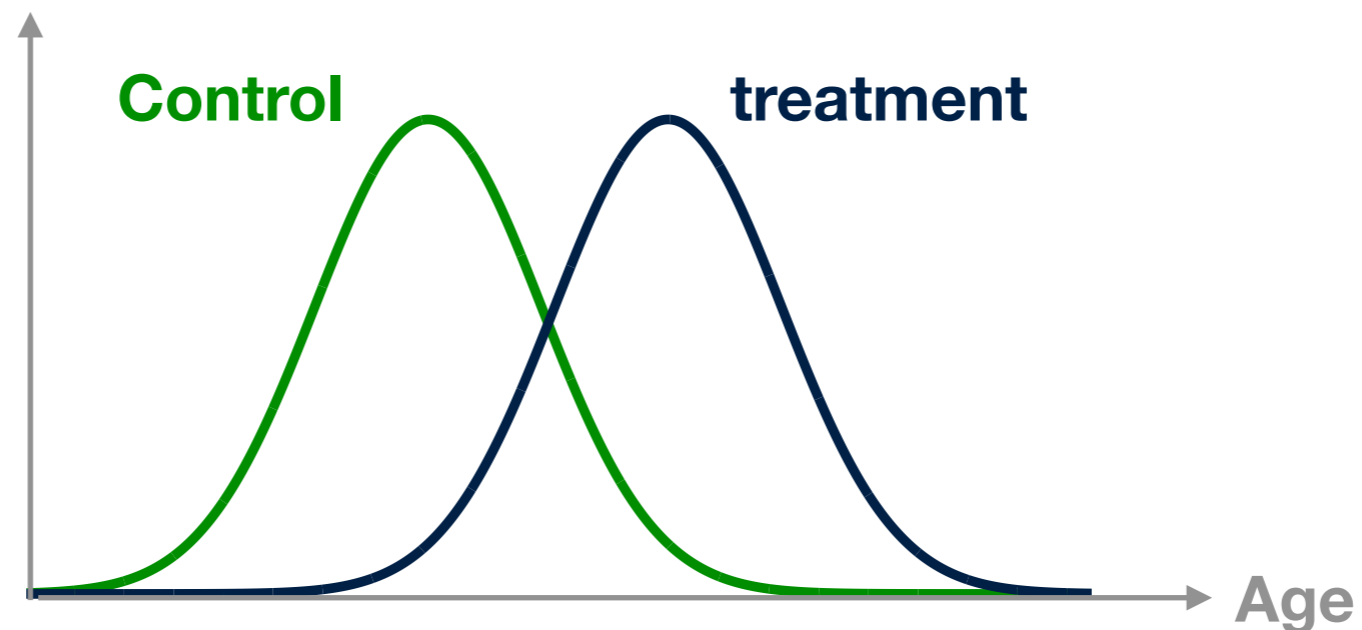
- **SUTVA: Stable Unit Treatment Value Assumption**
 - **Consistency:** Well-defined treatment (no different versions) potential outcome is independent of how the treatment is assigned
 - **No interference:** Different individuals (units) within a population do not influence each other (e.g. does not work in social behavioural studies, care must be taken for time series data when defining the units)

Potential Outcomes: Assumptions

- **SUTVA: Stable Unit Treatment Value Assumption**
 - **Consistency:** Well-defined treatment (no different versions) potential outcome is independent of how the treatment is assigned
 - **No interference:** Different individuals (units) within a population do not influence each other (e.g. does not work in social behavioural studies, care must be taken for time series data when defining the units)
- **Positivity:** Every individual has a non-zero chance of receiving the treatment/control:
$$p(t = 1|x) \in (0, 1) \text{ if } P(x) > 0$$
- **Unconfoundedness (ignorability/exchangeability):** Treatment assignment is random, given confounding features X

Observational data: What goes wrong?

$$p(x|t = 1) \neq p(x|t = 0)$$



$$\left(\int y_1(x)p(x|t = 1)dx - \int y_0(x)p(x|t = 0)dx \right) \neq \int (y_1(x) - y_0(x))p(x)dx$$

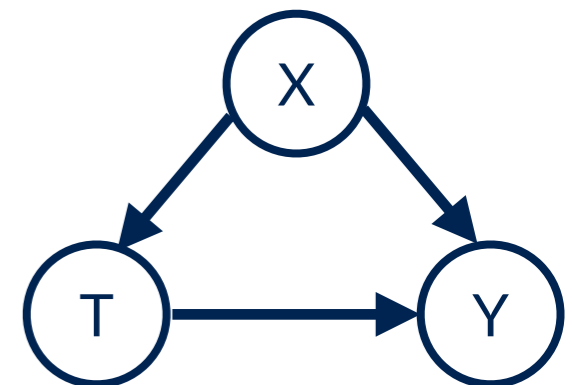
Adjustment formula (will be revisited later)

$$\mathbb{E}[Y_1 - Y_0 | X] = \mathbb{E}[Y_1 | X] - \mathbb{E}[Y_0 | X]$$

$$= \mathbb{E}[Y_1 | T = 1, X] - \mathbb{E}[Y_0 | T = 0, X] \quad \text{By Unconfoundedness: } Y_1, Y_0 \perp\!\!\!\perp T \mid X$$

$$= \mathbb{E}[Y | T = 1, X] - \mathbb{E}[Y | T = 0, X] \quad \text{By construction: } Y = TY_1 + (1 - T)Y_0$$

Also need positivity



Adjustment formula (will be revisited later)

$$\begin{aligned}\mathbb{E}[Y_1 - Y_0|X] &= \mathbb{E}[Y_1|X] - \mathbb{E}[Y_0|X] \\ &= \mathbb{E}[Y_1|T = 1, X] - \mathbb{E}[Y_0|T = 0, X] && \text{By Unconfoundedness: } Y_1, Y_0 \perp\!\!\!\perp T \mid X \\ &= \mathbb{E}[Y|T = 1, X] - \mathbb{E}[Y|T = 0, X] && \text{By construction: } Y = TY_1 + (1 - T)Y_0\end{aligned}$$

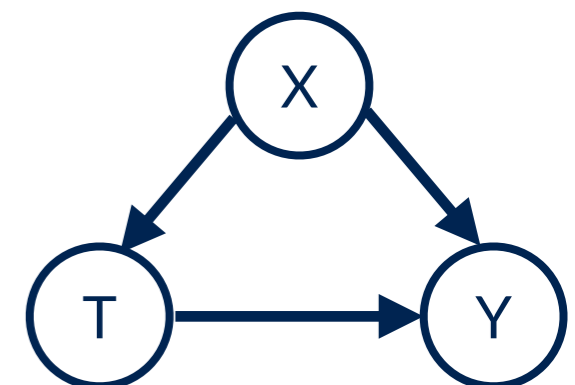
Also need positivity

$$\mathbb{E}[Y_1 - Y_0] = \mathbb{E}_X \left[\mathbb{E}[Y_1 - Y_0|X] \right]$$

ATE

$$= \mathbb{E}_X \left[\mathbb{E}[Y|T = 1, X] - \mathbb{E}[Y|T = 0, X] \right]$$

The adjustment formula



Adjustment formula (will be revisited later)

$$\begin{aligned}\mathbb{E}[Y_1 - Y_0|X] &= \mathbb{E}[Y_1|X] - \mathbb{E}[Y_0|X] \\ &= \mathbb{E}[Y_1|T = 1, X] - \mathbb{E}[Y_0|T = 0, X] && \text{By Unconfoundedness: } Y_1, Y_0 \perp\!\!\!\perp T \mid X \\ &= \mathbb{E}[Y|T = 1, X] - \mathbb{E}[Y|T = 0, X] && \text{By construction: } Y = TY_1 + (1 - T)Y_0\end{aligned}$$

Also need positivity

$$\begin{aligned}\mathbb{E}[Y_1 - Y_0] &= \mathbb{E}_X \left[\mathbb{E}[Y_1 - Y_0|X] \right] \\ &= \mathbb{E}_X \left[\underbrace{\mathbb{E}[Y|T = 1, X] - \mathbb{E}[Y|T = 0, X]} \right]\end{aligned}$$

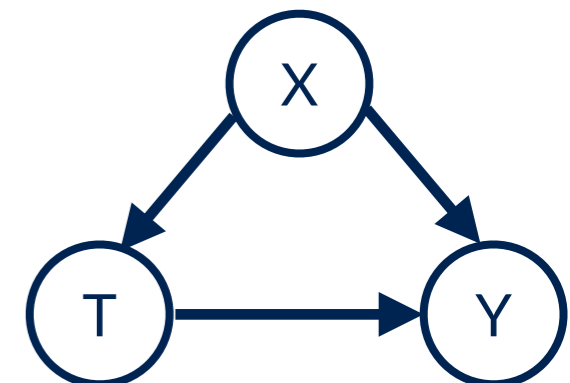
The adjustment formula

Hypothetical world

Real world

i.e., can be estimated from observational data

Causal identifiability



Regression Adjustment: Another perspective

Fit a model for $Q(T, X) = \mathbb{E}[Y|T, X]$

(last time we substituted $T=1$ and $T=0$ into individual treatment effect $= Q(1, x^{(i)}) - Q(0, x^{(i)})$, then took average over all individuals i , via linear regression). Under the linearity assumption:

$$\mathbb{E}[Y|T, X] = \alpha_0 + \beta_x X + \beta_t T + \epsilon, \quad \mathbb{E}[\epsilon] = 0$$

Regression Adjustment: Another perspective

Fit a model for $Q(T, X) = \mathbb{E}[Y|T, X]$

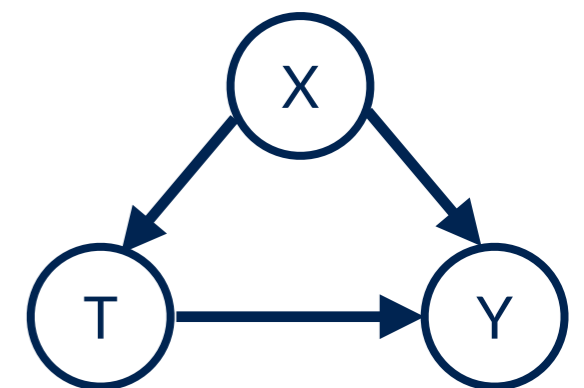
(last time we substituted $T=1$ and $T=0$ into individual treatment effect $= Q(1, x^{(i)}) - Q(0, x^{(i)})$, then took average over all individuals i , via linear regression). Under the linearity assumption:

$$\mathbb{E}[Y|T, X] = \alpha_0 + \beta_x X + \beta_t T + \epsilon, \quad \mathbb{E}[\epsilon] = 0$$

$$\begin{aligned} ATE &= \mathbb{E}_X \left[\mathbb{E}[Y|T = 1, X] - \mathbb{E}[Y|T = 0, X] \right] \\ &= \left(\alpha_0 + \beta_x \mathbb{E}[X] + \beta_t \right) - \left(\alpha_0 + \beta_x \mathbb{E}[X] \right) \\ &= \beta_t \end{aligned}$$

Important remarks about the previous form:

- 1) Depends on the structure of the causal graph of interest
- 2) Data need not be linear
model-misspecification -> statistical bias



Important remarks about the previous form:

2) Data need not be linear, example:

Say we fitted $\mathbb{E}[Y|T, X] = \alpha_0 + \beta_x X + \beta_t T + \epsilon$, $\mathbb{E}[\epsilon] = 0$

And obtained β_t for the causal effect,

BUT, in reality the true data generating distribution is e.g.

$$\mathbb{E}[Y|T, X] = \alpha_0 + \beta_x X + \beta_t T + \gamma X.T + \epsilon, \mathbb{E}[\epsilon] = 0$$

Or e.g. non-linear:

$$\mathbb{E}[Y|T, X] = e^{\alpha_0 + \beta_x X + \beta_t T + \gamma X.T}$$



Important remarks about the previous form:

2) Data need not be linear, example:

Say we fitted $\mathbb{E}[Y|T, X] = \alpha_0 + \beta_x X + \beta_t T + \epsilon$, $\mathbb{E}[\epsilon] = 0$

And obtained β_t for the causal effect,

BUT, in reality the true data generating distribution is e.g.

$$\mathbb{E}[Y|T, X] = \alpha_0 + \beta_x X + \beta_t T + \gamma X.T + \epsilon, \mathbb{E}[\epsilon] = 0$$

Or e.g. non-linear:

$$\mathbb{E}[Y|T, X] = e^{\alpha_0 + \beta_x X + \beta_t T + \gamma X.T}$$

Then $ATE = \mathbb{E}_X \left[\mathbb{E}[Y|T = 1, X] - \mathbb{E}[Y|T = 0, X] \right]$

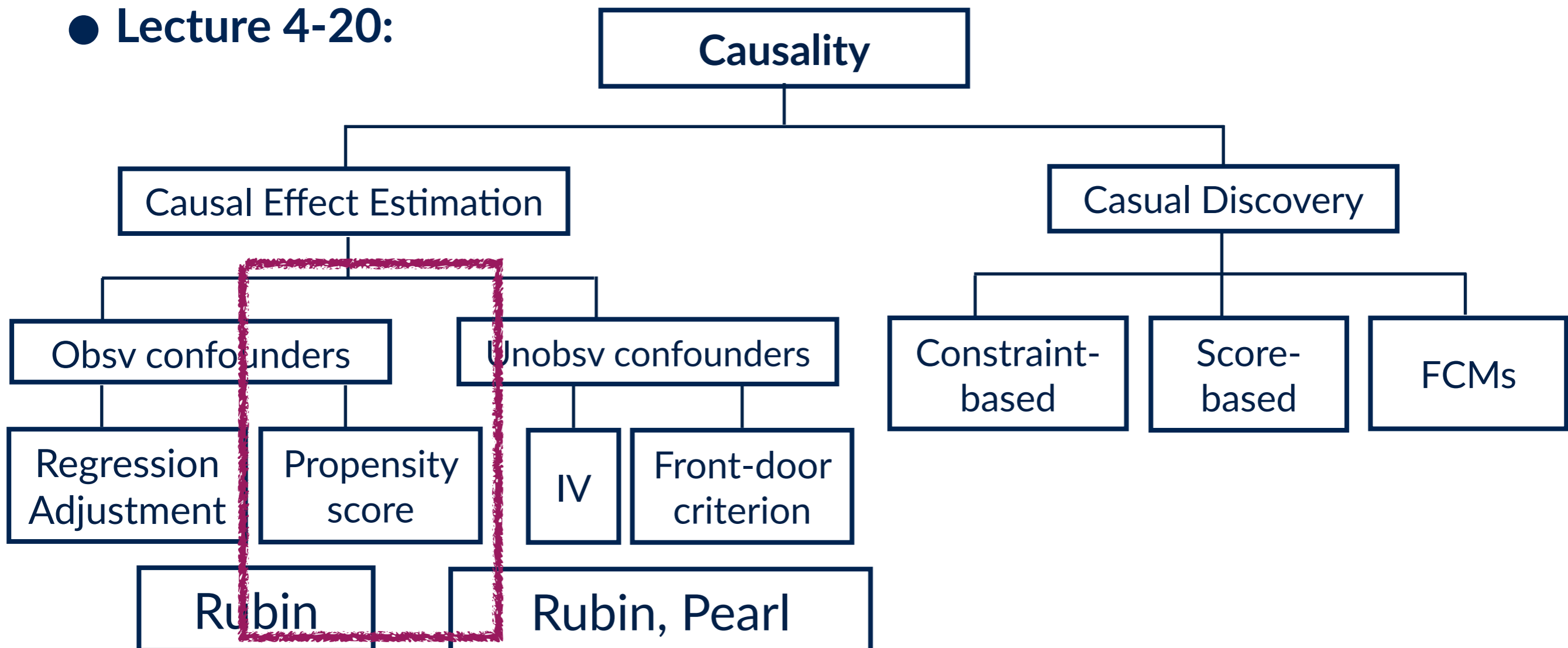
is **not** simply β_t !!

Valid causal inference requires correctly-specified models and mathematical guarantees!



Overview of the course

- **Lecture 1:** Introduction & Motivation, why do we care about causality? Why deriving causality from observational data is non-trivial.
- **Lecture 2:** Recap of probability theory, variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3:** Recap of regression, multiple regression, graphs, SCM
- **Lecture 4-20:**

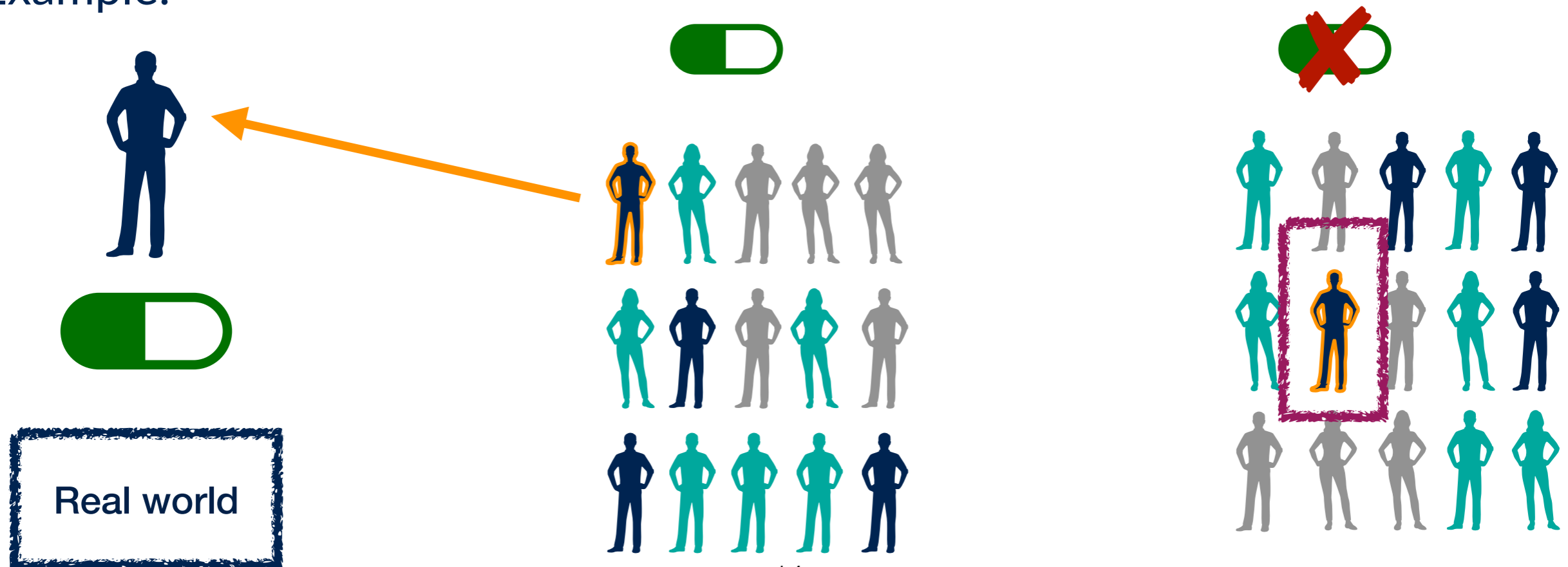


Matching

Idea: Create a 'clone/twin' for each individual (in terms of X)
i.e. if individual 1 has $t = 1$, then their 'clone/twin' has $t = 0$.

Blind ourselves to the outcomes, try to get as similar to a randomised experiment as possible ('correct for confounding')

Example:



Balancing Score

- In a perfect randomised trial: $p(t=1|x)=p(t=1)$
- In an **observational study**, $p(t=1|x)$ can be **estimated**, since it involves **observational data** at a t and x (hence identifiable).

- A **balancing score** is any function $b(x)$ such that:

$$x \perp\!\!\!\perp t | b(x)$$

- i.e., distribution of confounders is independent of treatment given $b(x)$:

$$p(X = x | b(x), t = 1) = p(X = x | b(x), t = 0)$$

Balancing Score: Proof 1 [non-examinable]

Unconfoundedness given a balancing score. Suppose we have unconfoundedness, i.e., $Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid X^{(i)}$. Then for a balancing score $b(x)$ we have:

$$Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid b(X^{(i)})$$

Balancing Score: Proof 1 [non-examinable]

Unconfoundedness given a balancing score. Suppose we have unconfoundedness, i.e., $Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid X^{(i)}$. Then for a balancing score $b(x)$ we have:

$$Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid b(X^{(i)})$$

Proof: Need to show $p_T(T = 1 \mid Y_1, Y_0, b(X)) = p_T(T = 1 \mid b(X))$

Balancing Score: Proof 1 [non-examinable]

Unconfoundedness given a balancing score. Suppose we have unconfoundedness, i.e., $Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid X^{(i)}$. Then for a balancing score $b(x)$ we have:

$$Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid b(X^{(i)})$$

Proof: Need to show $p_T(T = 1 \mid Y_1, Y_0, b(X)) = p_T(T = 1 \mid b(X))$

Maths aside: Iterated expectations $\mathbb{E}_Z[Z] = \mathbb{E}_W[\mathbb{E}_{Z|W}[Z|W]]$

Balancing Score: Proof 1 [non-examinable]

Unconfoundedness given a balancing score. Suppose we have unconfoundedness, i.e., $Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid X^{(i)}$. Then for a balancing score $b(x)$ we have:

$$Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid b(X^{(i)})$$

Proof: Need to show $p_T(T = 1 \mid Y_1, Y_0, b(X)) = p_T(T = 1 \mid b(X))$

Maths aside: Iterated expectations $\mathbb{E}_Z[Z] = \mathbb{E}_W[\mathbb{E}_{Z|W}[Z|W]]$

$$\begin{aligned} \mathbb{E}_W[\mathbb{E}_{Z|W}[Z|W]] &= \sum_w \sum_z p(Z = z \mid W = w) z p(W = w) \\ &= \sum_{w,z} \frac{p(z, w)}{p(w)} z p(w) = \sum_z p(z) z = \mathbb{E}_Z[Z] \end{aligned}$$

Balancing Score: Proof 1 [non-examinable]

Unconfoundedness given a balancing score. Suppose we have unconfoundedness, i.e., $Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid X^{(i)}$. Then for a balancing score $b(x)$ we have:

$$Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid b(X^{(i)})$$

Proof: Need to show $p_T(T = 1 \mid Y_1, Y_0, b(X)) = p_T(T = 1 \mid b(X))$

$$p_T(T = 1 \mid Y_1, Y_0, b(X)) = \mathbb{E}_T [T \mid Y_1, Y_0, b(X)]$$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid Y_1, Y_0, b(X), X] \mid Y_1, Y_0, b(X) \right]$$

$$\mathbb{E}[Z \mid W] = \mathbb{E}_{V \mid W} [\mathbb{E}[Z \mid W, V] \mid W]$$

Balancing Score: Proof 1 [non-examinable]

Unconfoundedness given a balancing score. Suppose we have unconfoundedness, i.e., $Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid X^{(i)}$. Then for a balancing score $b(x)$ we have:

$$Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid b(X^{(i)})$$

Proof: Need to show $p_T(T = 1 \mid Y_1, Y_0, b(X)) = p_T(T = 1 \mid b(X))$

$$p_T(T = 1 \mid Y_1, Y_0, b(X)) = \mathbb{E}_T [T \mid Y_1, Y_0, b(X)]$$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid Y_1, Y_0, b(X), X] \mid Y_1, Y_0, b(X) \right]$$

$$\mathbb{E}[Z \mid W] = \mathbb{E}_{V \mid W} [\mathbb{E}[Z \mid W, V] \mid W]$$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid b(X), X] \mid Y_1, Y_0, b(X) \right]$$

By Unconfoundedness: $Y_1, Y_0 \perp\!\!\!\perp T \mid X$

Balancing Score: Proof 1 [non-examinable]

Unconfoundedness given a balancing score. Suppose we have unconfoundedness, i.e., $Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid X^{(i)}$. Then for a balancing score $b(x)$ we have:

$$Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid b(X^{(i)})$$

Proof: Need to show $p_T(T = 1 \mid Y_1, Y_0, b(X)) = p_T(T = 1 \mid b(X))$

$$p_T(T = 1 \mid Y_1, Y_0, b(X)) = \mathbb{E}_T [T \mid Y_1, Y_0, b(X)]$$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid Y_1, Y_0, b(X), X] \mid Y_1, Y_0, b(X) \right]$$

$$\mathbb{E}[Z \mid W] = \mathbb{E}_{V \mid W} [\mathbb{E}[Z \mid W, V] \mid W]$$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid b(X), X] \mid Y_1, Y_0, b(X) \right]$$

By Unconfoundedness: $Y_1, Y_0 \perp\!\!\!\perp T \mid X$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid b(X)] \mid Y_1, Y_0, b(X) \right]$$

By definition of balancing score:

$$X \perp\!\!\!\perp T \mid b(X)$$

Balancing Score: Proof 1 [non-examinable]

Unconfoundedness given a balancing score. Suppose we have unconfoundedness, i.e., $Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid X^{(i)}$. Then for a balancing score $b(x)$ we have:

$$Y_1^{(i)}, Y_0^{(i)} \perp\!\!\!\perp T^{(i)} \mid b(X^{(i)})$$

Proof: Need to show $p_T(T = 1 \mid Y_1, Y_0, b(X)) = p_T(T = 1 \mid b(X))$

$$p_T(T = 1 \mid Y_1, Y_0, b(X)) = \mathbb{E}_T [T \mid Y_1, Y_0, b(X)]$$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid Y_1, Y_0, b(X), X] \mid Y_1, Y_0, b(X) \right]$$

$$\mathbb{E}[Z \mid W] = \mathbb{E}_{V \mid W} [\mathbb{E}[Z \mid W, V] \mid W]$$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid b(X), X] \mid Y_1, Y_0, b(X) \right]$$

By Unconfoundedness: $Y_1, Y_0 \perp\!\!\!\perp T \mid X$

$$= \mathbb{E}_{X \mid Y_1, Y_0, b(X)} \left[\mathbb{E} [T \mid b(X)] \mid Y_1, Y_0, b(X) \right]$$

By definition of balancing score:

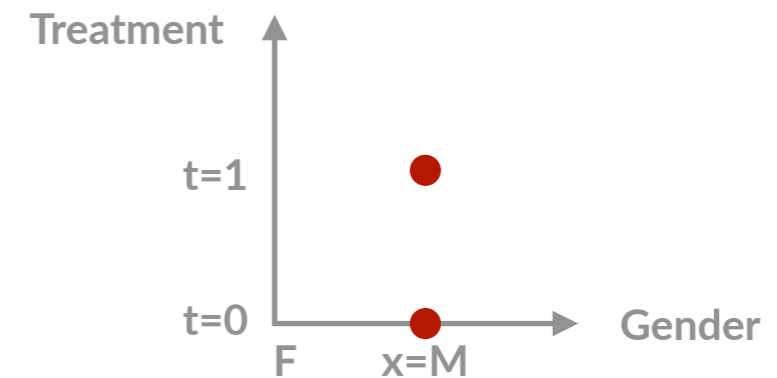
$$X \perp\!\!\!\perp T \mid b(X)$$

$$= \mathbb{E}_{T \mid b(X)} [T \mid b(X)] = p_T(T = 1 \mid b(X))$$

Propensity Score

- Candidate $b(x) = x$, trivially satisfies:

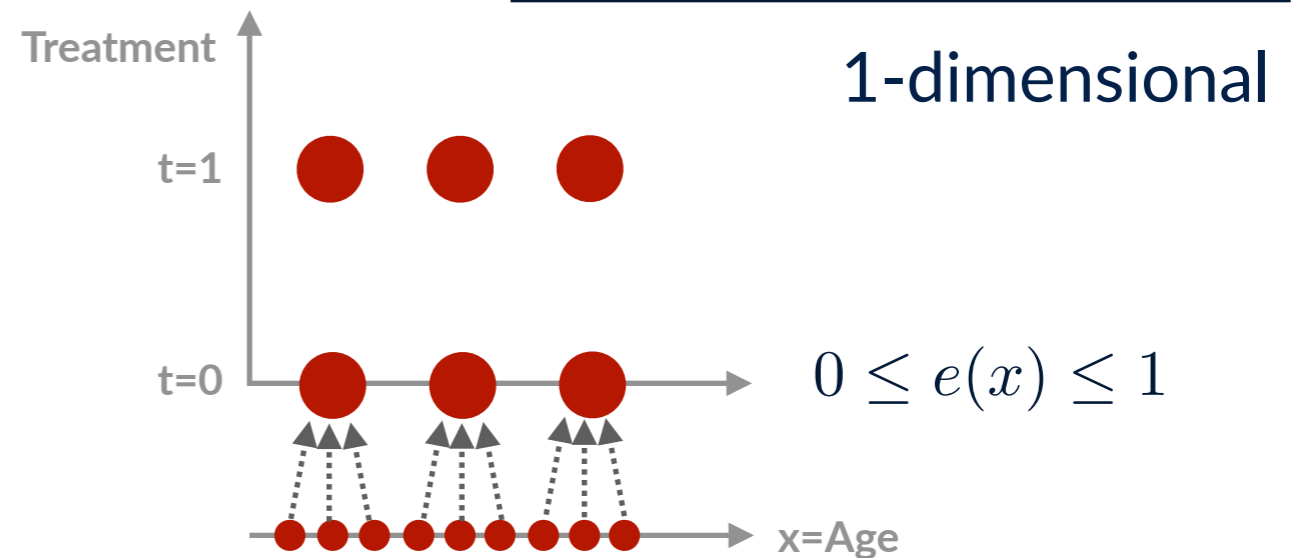
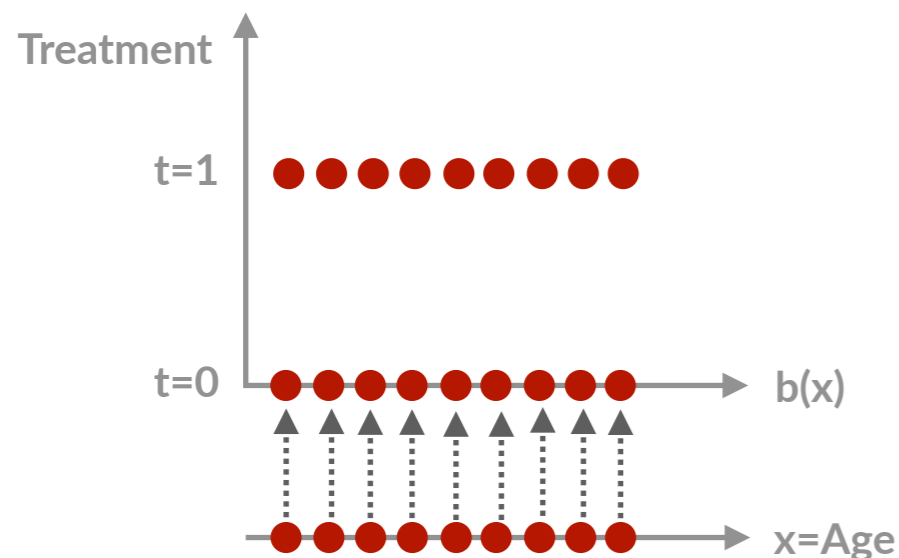
$$p(X = x|x, t = 1) = p(X = x|x, t = 0) = 1$$



- $b(x) = x$ is the **finest** such function: OK for e.g. binary confounders, but only gives point estimates for (almost) continuous confounders!

- **Propensity score** is the **coarsest** such function (i.e. more data points, leading to better estimates):

$$e(x) = p(t = 1|x)$$



Propensity is balancing: Proof 2 [non-examinable]

The propensity score is a balancing score: $X \perp\!\!\!\perp T | e(X)$

Proof: Need to show $p_T(T = 1 | X, e(X)) = p_T(T = 1 | e(X))$

$$\text{LHS: } p_T(T = 1 | X, e(X)) = p_T(T = 1 | X) = e(X)$$

↑
Propensity score
is a function of X

↑
Propensity score
definition

Propensity is balancing: Proof 2 [non-examinable]

The propensity score is a balancing score: $X \perp\!\!\!\perp T | e(X)$

Proof: Need to show $p_T(T = 1 | X, e(X)) = p_T(T = 1 | e(X))$

LHS: $p_T(T = 1 | X, e(X)) = p_T(T = 1 | X) = e(X)$

RHS:

$$\mathbb{E}[Z|W] = \mathbb{E}_{V|W}[\mathbb{E}[Z|W, V]|W]$$

$$\begin{aligned} p_T(T = 1 | e(X)) &= \mathbb{E}[T | e(X)] = \mathbb{E}_{X|e(X)} \left[\underbrace{\mathbb{E}[T | e(X), X]}_{e(X)} | e(X) \right] \\ &= \mathbb{E}[e(X) | e(X)] = e(X) \end{aligned}$$

Propensity is balancing: Proof 3 [non-examinable]

The propensity score is the coarsest balancing score, i.e., it is a function of every balancing score $b(x) : e(x) = f(b(x))$

Propensity is balancing: Proof 3 [non-examinable]

The propensity score is the coarsest balancing score, i.e., it is a function of every balancing score $b(x) : e(x) = f(b(x))$

Proof: Let $b(x)$ be a balancing score. Suppose we **cannot** write the propensity score $e(x)$ as $e(x) = f(b(x))$. Therefore, there must be a case where :
 $b(x) = b(x') = b^*$ while $e(x) \neq e(x')$. Then,

$$p(t = 1 | \underline{x}, b(x)) = p(t = 1 | x) = e(x) \neq e(x') = p(t = 1 | x') = p(t = 1 | \underline{x'}, b(x'))$$

Propensity is balancing: Proof 3 [non-examinable]

The propensity score is the coarsest balancing score, i.e., it is a function of every balancing score $b(x) : e(x) = f(b(x))$

Proof: Let $b(x)$ be a balancing score. Suppose we **cannot** write the propensity score $e(x)$ as $e(x) = f(b(x))$. Therefore, there must be a case where : $b(x) = b(x') = b^*$ while $e(x) \neq e(x')$. Then,

$$p(t = 1 | x, \cancel{b(x)}) \neq p(t = 1 | x', \cancel{b(x')})$$

b^* b^*

Recall definition of balancing score:

$$x \perp\!\!\!\perp t | b(x)$$

Propensity is balancing: Proof 3 [non-examinable]

The propensity score is the coarsest balancing score, i.e., it is a function of every balancing score $b(x) : e(x) = f(b(x))$

Proof: Let $b(x)$ be a balancing score. Suppose we **cannot** write the propensity score $e(x)$ as $e(x) = f(b(x))$. Therefore, there must be a case where : $b(x) = b(x') = b^*$ while $e(x) \neq e(x')$. Then,

$$p(t = 1|x, b(x)) \neq p(t = 1|x', b(x'))$$

i.e., probability of treatment changes depending on value of x **despite** b^* :

$$x \not\perp t | b(x)$$

This violates the definition of a balancing score.

Proof by contradiction.

Propensity Score Matching Algorithms

- Match control and treatment individuals based on their propensity score
- Greedy matching:
 - Randomly order list of control and treated.
 - Start with the first individual from e.g. treated and match to control with the smallest distance (i.e. obtains the **local** minimum)
 - Remove individuals from control and matched treated
 - Move to the next treated subject

Treatment	Control
40	50
65	25

Propensity Score Matching Algorithms

- Match control and treatment individuals based on their propensity score
- Greedy matching:
 - Randomly order list of control and treated.
 - Start with the first individual from e.g. treated and match to control with the smallest distance (i.e. obtains the **local** minimum)
 - Remove individuals from control and matched treated
 - Move to the next treated subject



Propensity Score Matching Algorithms

- Match control and treatment individuals based on their propensity score
- Greedy matching:
 - Randomly order list of control and treated.
 - Start with the first individual from e.g. treated and match to control with the smallest distance (i.e. obtains the **local** minimum)
 - Remove individuals from control and matched treated
 - Move to the next treated subject



Propensity Score Matching Algorithms

- Match control and treatment individuals based on their propensity score
- Greedy matching:
 - Randomly order list of control and treated.
 - Start with the first individual from e.g. treated and match to control with the smallest distance (i.e. obtains the **local** minimum)
 - Remove individuals from control and matched treated
 - Move to the next treated subject
- Optimal matching: Minimises the **global** distance, computationally demanding

- **ATE:** $\tau = \hat{\mathbb{E}}[\tau^{(i)}] = \hat{\mathbb{E}}[y_1^{(i)} - y_0^{(i)}] = \frac{1}{N} \sum_{i=0}^N (y_1^{(i)} - y_0^{(i)})$

Inverse Probability of Treatment Weighting (IPTW)

- Inflate the weight for under represented-subjects due to missing data
- Based on propensity score
- Weight: inverse probability of receiving observed treatment, for individual i with covariate x :

$$w_i = \begin{cases} \frac{1}{e(x_i)} & \text{if } t_i = 1 \\ \frac{1}{1-e(x_i)} & \text{if } t_i = 0 \end{cases}$$

$$e(x) = p(t = 1|x)$$

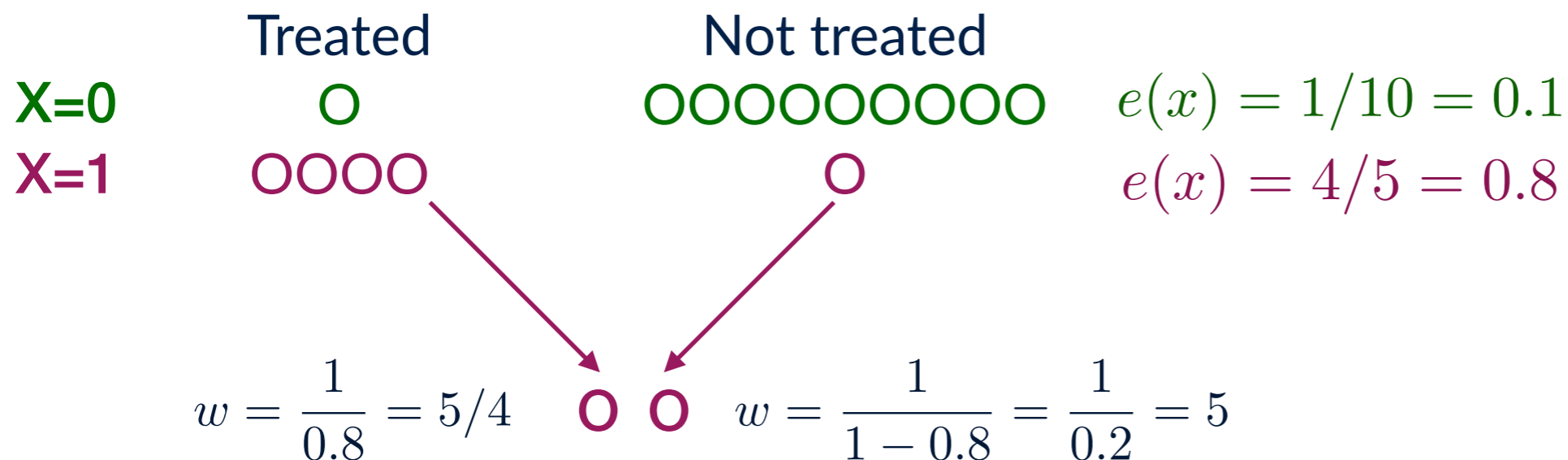
- Example: Suppose individual (i) has a large $e(x)$, i.e., their probability of receiving treatment is high.
 - If $t_i = 1$ then $w_i \approx 1$ (typical behaviour: most with x_i are treated)
 - If $t_i = 0$ then $w_i \gg 1$ (underrepresented: boost weight for rare event)

Inverse Probability of Treatment Weighting (IPTW)

- Inflate the weight for under represented-subjects due to missing data
- Based on propensity score
- Weight: inverse probability of receiving observed treatment, for individual i with covariate x :

$$w_i = \begin{cases} \frac{1}{e(x_i)} & \text{if } t_i = 1 \\ \frac{1}{1-e(x_i)} & \text{if } t_i = 0 \end{cases}$$

$$e(x) = p(t = 1|x)$$



Inverse Probability of Treatment Weighting (IPTW)

$$e(x) = p(t = 1|x)$$

$$\frac{1}{N_1} \sum_{\text{treated}} y_{\underline{1}}^{(i)} \frac{1}{e(x_i)} - \frac{1}{N_0} \sum_{\text{not treated}} y_{\underline{0}}^{(i)} \frac{1}{1 - e(x_i)} \quad N = N_1 + N_0$$

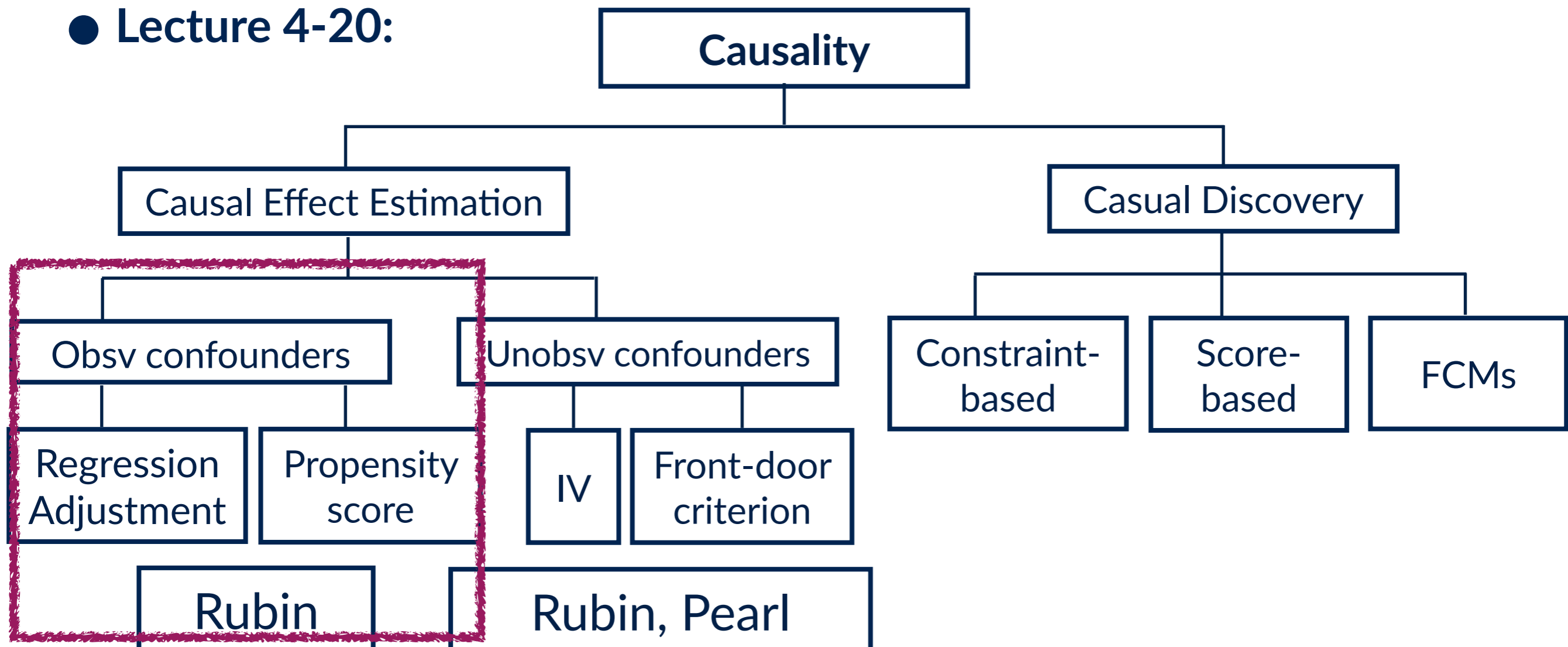
Weights may be inaccurate/unstable for subjects with a very low probability of receiving the observed treatment (other estimators exist)

In a randomised control trial (RCT) limit, $p(t = 1|x) = p(t = 0|x)$
above reduces to:

$$\frac{1}{N_1} \sum_{\text{treated}} y_{\underline{1}}^{(i)} - \frac{1}{N_0} \sum_{\text{not treated}} y_{\underline{0}}^{(i)}$$

Overview of the course

- **Lecture 1:** Introduction & Motivation, why do we care about causality? Why deriving causality from observational data is non-trivial.
- **Lecture 2:** Recap of probability theory, variables, events, conditional probabilities, independence, law of total probability, Bayes' rule
- **Lecture 3:** Recap of regression, multiple regression, graphs, SCM
- **Lecture 4-20:**





THE UNIVERSITY
of EDINBURGH

Methods for Causal Inference

Lecture 5: Rubin's framework, propensity score, IPTW

Ava Khamseh

School of Informatics
2023-2024