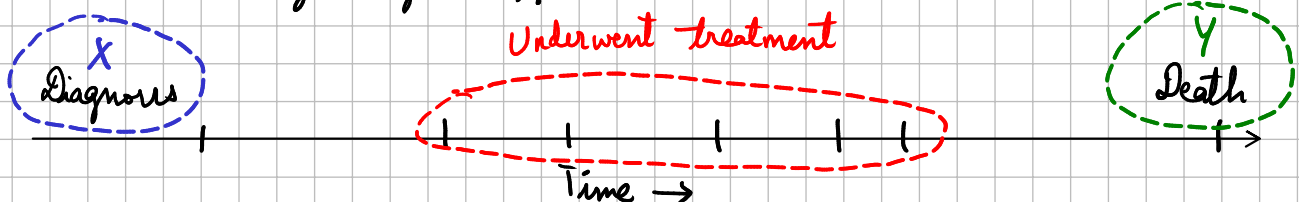


Causal Inference

- So far, purely predictive questions!
- If there are signs that there is correlation between features and target of interest, that's good enough
- Causal Directionality is irrelevant. (Not completely true)
- When there is dataset shift, ^{or non-stationary data}, causality matters. Understanding the data deeply is very helpful.
- In healthcare especially, causal questions are important to answer.
- For example, it is more important to prevent Type 2 diabetes (causal) vs early diagnosis of Type 2 diabetes (predictive).
- Naive way of inferring causality:
 - Let's say we trained a DL model to predict onset of Type 2 diabetes
 - Look at most negative feature (lowest weight), let's say it is Gastrojejunic Bypass surgery (yes or no)
 - Then does that mean that if a patient underwent this surgery, he/she won't get diabetes?
- Look at predictive weights is not enough.
- We need to come up with a mathematical model for causality
- Another example:
 - Let's say we train a DL model for predicting survival of breast cancer patient based on radiological mammogram and histopathological slides.
 - Let's say one patient diagnosed with breast cancer survived for longer than 5 years.
 - When a new patient, with similar diagnosis as the model's examples is given to the model, a higher survival is predicted.
 - Does this mean we shouldn't treat the patient?
- This is very dangerous!!



→ A longer survival time maybe because of treatment! Not solely because of diagnosis

→ But the model only learns the $X(\text{diagnosis}) \rightarrow Y(\text{survival})$ mapping

Guiding Treatment Decisions

→ Another question that needs to be answered is: How do we guide treatment decisions?

→ How do we tell who is likely to be benefitted by a given treatment?

→ But people respond differently to treatments?

→ Also, data used to guide treatments is based on existing treatment guidelines.

→ Naive way to guide treatment decisions:

Train a predictive model that learns to predict treatment decisions.

David → Treatment A

John → Treatment B

June → Treatment A

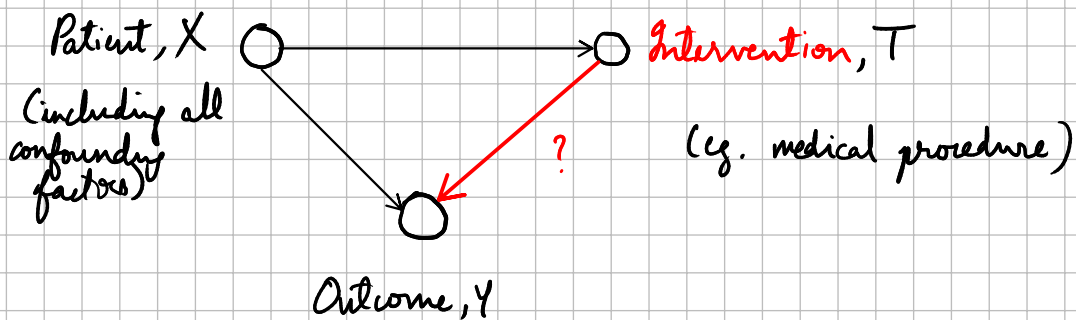
→ Best this can do is match current medical guidelines!

→ How do we go beyond this? We need to capture heterogeneity in treatment response. We need to change how we ask our question.

→ One last example:

- Traditional, does X cause Y ?
- Does smoking cause cancer?
- Doing a randomized controlled trial is unethical.
- Could we just compare $P(\text{lung cancer} | \text{smoker})$ vs $P(\text{lung cancer} | \text{non smoker})$?
- No because of confounding factors (see below)

→ To properly answer, we need to formulate as causal questions.



High dimensional

Observational data

Causal Graphs

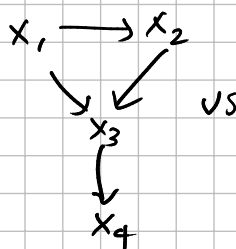
→ Instead of just $D \in \{x^{(i)}, y^{(i)}\}$, we need to think in terms of triplets: $D \in \{x^{(i)}, T^{(i)}, y^{(i)}\}$ → Interventions

Earlier
→ Causal Inference might take a form like so: (Can Skip! Won't be discussed further)

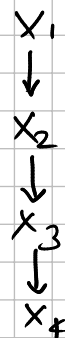
Data

x_1	x_2	x_3	x_4
0	1	0	0
0	0	1	0
...
1	1	1	0

we try to
find right
graph →



vs



What is the
underlying
causal graph?

For Two random variables and x_1 and x_2

$x_1 \rightarrow x_2$ i.e., $P(x_1) P(x_2 | x_1)$

$x_2 \rightarrow x_1$ i.e., $P(x_2) P(x_1 | x_2)$

Are indistinguishable (because of conditional probability and Bayes theorem)

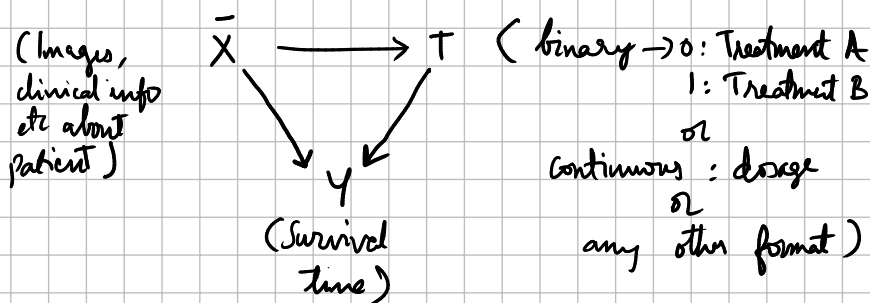
Then we might want to add interventions to $x_1 \rightarrow x_2$ and $x_2 \rightarrow x_1$ to disentangle the causalities.

→ This is the simplest possible case!

- X is highdimensional, T and Y are single random variable.
- The causal graph is assumed here.
- We 'just don't know the connection weights' strength!

Modelling Breast Cancer example as a causal graph

- X → Patient information (features) at Diagnosis
- T → Choose between treatment A or treatment B. Also treatment plans only depend on what we know about the patient at diagnosis
- Y → Survival time



- we will not change treatment in between (Dynamic treatment regimes or off policy reinforcement learning)
- Causal Inference questions have been asked for decades in political science, economics and statistics but we can't intervene (can only use observational data)
- In traditional statistics, domain knowledge was required to determine factors that determine the treatment to be taken. These factors are called confounding factors.
- Now causal inference questions are asked on high dimensional data (images) and research on leveraging machine learning algorithms, by mapping a causal inference problem down to a machine learning model.
- To formalize these notions concretely, we need a mathematical model
Potential Outcomes Framework (Rubin-Neyman Causal Model)

- Each unit (individual) x_i has two potential outcomes:
 - $Y_0(x_i)$ is the potential outcome had the unit not been treated: "control outcome"
 - $Y_1(x_i)$ is the potential outcome had the unit been treated: "treated outcome"

→ Conditional average treatment effect for unit i :

$$CATE(x_i) = E_{Y_1 \sim P(Y_1 | x_i)} [Y_1 | x_i] - E_{Y_0 \sim P(Y_0 | x_i)} [Y_0 | x_i]$$

Given x_i (such as age = 56), mean all outcomes where treatment is given

Given x_i (such as age = 56) mean all outcomes where treatment is not given.

→ also called Individual Treatment effect

→ Average treatment effect:

$$ATE : E[Y_1 - Y_0] = E_{x \sim P(x)} [CATE(x)]$$

→ Observed factual outcomes (You can only Y_0 or Y_1 at a time but not both):

$$y_i = t_i Y_1(x_i) + (1 - t_i) Y_0(x_i)$$

→ Unobserved counterfactual outcome (What if the opposite treatment was given?)

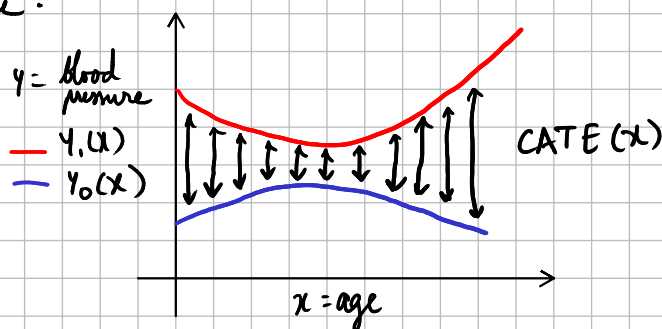
$$y_i^{CF} = (1 - t_i) Y_1(x_i) + t_i Y_0(x_i)$$

→ Since we observe only factual outcomes, we need to impute the counterfactual outcomes.

"We only ever observe one of the two outcomes"

→ Fundamental problem in Causal Inference.

Example:



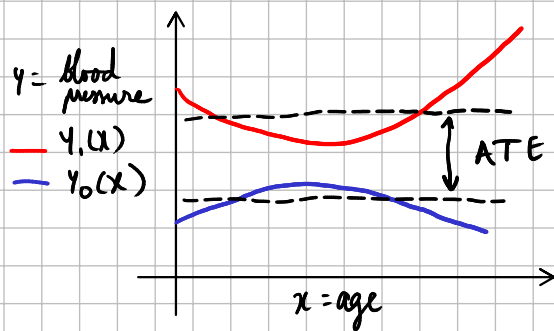
For middle aged people, difference between treatments 0 and 1 is significantly lower than younger and old aged people.

If treatment 1 is much cheaper than 0, we could use it for middle aged people.

→ We wanna predict CATE value using our data.

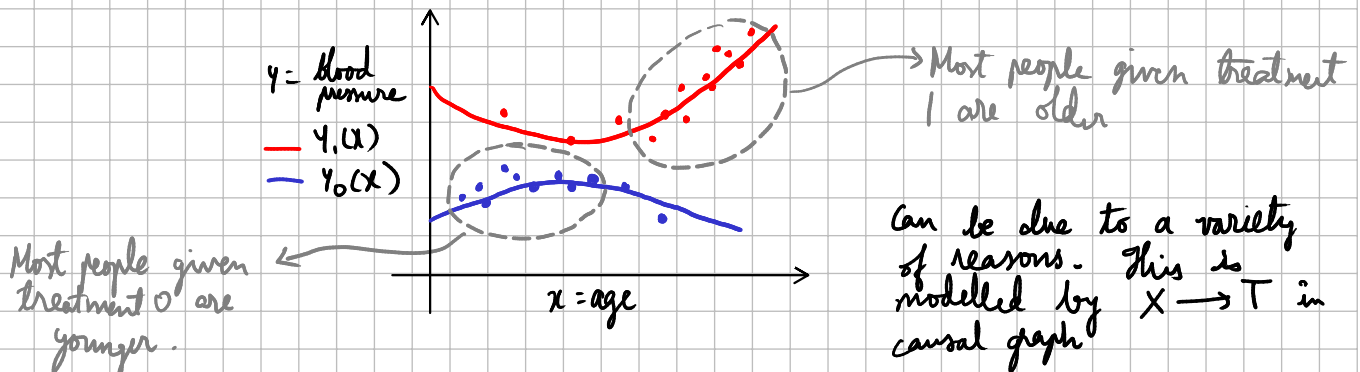
→ Sometimes we don't have the luxury of having personalized treatment effects (CATE). For example, the govt may start a policy that all men above 50 should receive prostate cancer screening.

→ Then you would wanna see the amount of decrease in prostate cancer related deaths. Such a policy is very broad. Mathematically captured by ATE.

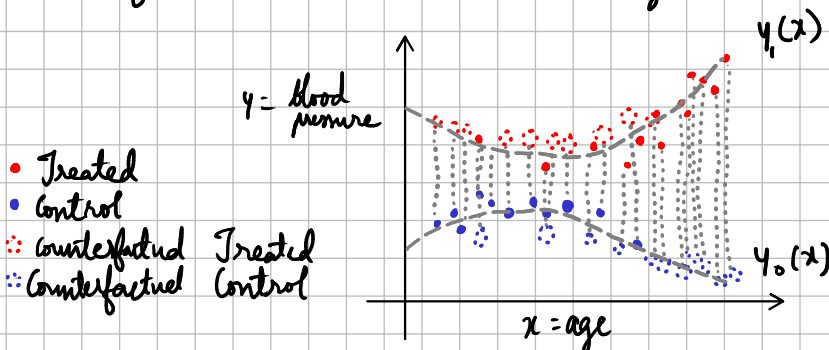


Treatment 0 on average is better than treatment 1.

→ But IRL we don't observe curves, rather we observe data points.



→ What if the other treatment had been given to the patient?



An Example:

(age, gender, exercise)	Sugar levels had they received medication A	Sugar levels had they received medication B	Observed Sugar levels
A (45, F, 0)	6	5.5	6
B (45, F, 1)	7	6.5	6.5
A (55, M, 0)	7	6	7
B (55, M, 1)	9	8	8
B (65, F, 0)	8.5	8	8
A (65, F, 1)	7.5	7	7.5
B (75, M, 0)	10	9	9
A (75, M, 1)	8	7	8

→ only observed facts.
 Medication B is worse than medication A

$$\begin{aligned} & \text{mean}(\text{sugar} | \text{medication B}) - \\ & \text{mean}(\text{sugar} | \text{medication A}) \\ & = 7.875 - 7.125 = 0.75 \end{aligned}$$

$$\begin{aligned} & \text{mean}(\text{sugar} | \text{had they received B}) - \\ & \text{mean}(\text{sugar} | \text{had they received A}) \\ & = 7.125 - 7.875 = -0.75 \end{aligned}$$

→ along with unobserved counterfactuals.
 Medication B is better than A

→ Without counterfactuals, we get the opposite reasoning for treatment B.

→ How do we solve without counterfactuals?

→ We need to make a ton of assumptions.

Assumptions in Causal Inference

No unmeasured confounders (Ignorability)

Y_0, Y_1 : potential outcomes for control and treated

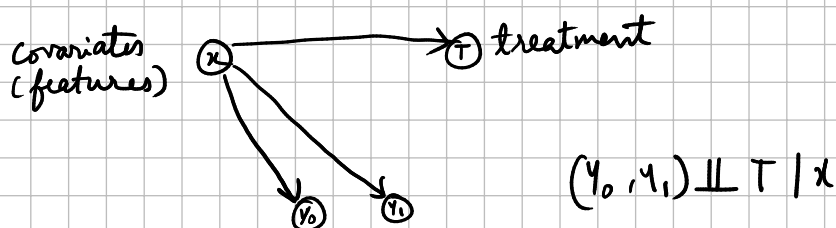
X : unit covariates (features)

T : treatment assignment.

We assume,

$$(Y_0, Y_1) \perp\!\!\!\perp T \mid X$$

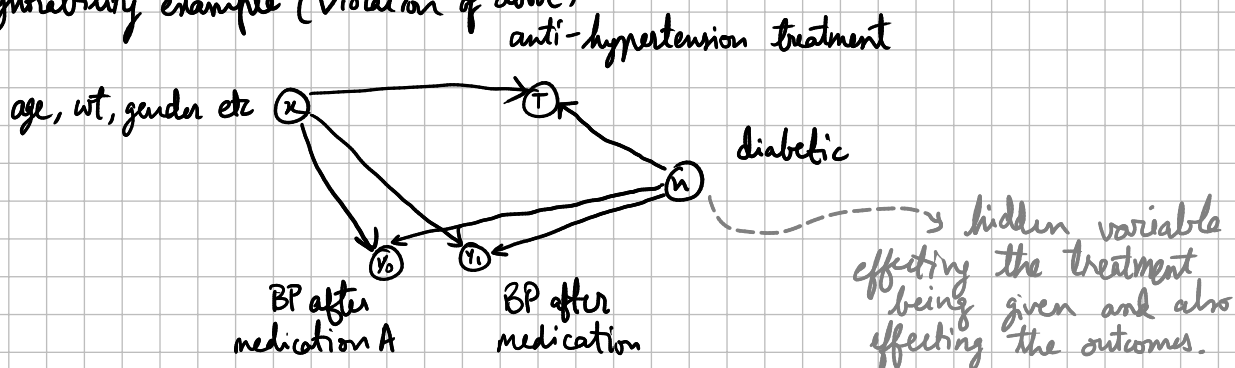
→ The potential outcomes are independent of treatment assignment, conditioned on covariates.



→ Since we already know what treatment is given to patient, it doesn't make sense to have an edge from T to Y_0 and Y_1 .

→ We are using a potential outcomes notation.

No ignorability example (Violation of above)



$$(Y_0, Y_1) \not\perp\!\!\!\perp T \mid X$$

→ h is confounding factor.

→ If h only affects treatments but not outcomes, then h is not a confounding

factor.

→ We need to observe all hidden confounding factors for causal inference to work, unlike in machine learning, where the only downside would be lower prediction accuracy.

→ To assess the robustness to violations of our assumed causal model, we can perform sensitivity analysis. "How much do my conclusions change if there were a confounding factor that is W amount strong?"

Common Support

Y_0, Y_1 : potential outcomes for control and treated

X : unit covariates (features)

T : treatment assignment

We assume,

$$P(T=t | X=x) > 0 \quad \forall t, x$$

→ It means that for a given set of patients ($X=x$), there has to be some amount of patients who have received treatment $T=t$.

→ Essentially, causal inference will be meaningless if no patients are given a certain treatment $T=t$, as we won't be able to determine the counterfactuals.

Framing the question for Causal Inference

1. Where could we go to for data to answer these questions?
2. What should X, T and Y be to satisfy ignorability? → Cannot be tested using observational data.
3. What is the specific causal inference question that we are interested in?
4. Are you worried about common support?

→ Can be tested using observational data.

→ In the table, check if (75, M, 1) B occurs, if not common support assumption is violated. But we can make approximations. For example age 74 instead of 75 exactly might be good enough.

→ Now that we know the assumptions and confirmed that they are held, how do we then do causal inference?

→ How do we actually compute CATE and ATE for observational data?

Average Treatment Effect - The adjustment formula

→ Also called G-formula.