

Week 8

Wentao Gao

This month:

- Potential outcome(Rubin)
- Causal Diagram(Pearl)
- Causal inference in time series(Overview)

Potential Outcome

Overview

- Causal Effect
- Randomized Experiment
- Effect modification
- Confounding

Causal Effect

Potential Outcome

Every decision(binary) will lead to two potential outcomes. One is $Y(1)$, the other is $Y(0)$. What about counterfactual outcomes?

Individual Causal Effect

We can now provide a formal definition of a *causal effect for an individual*: The treatment A has a causal effect on an individual's outcome Y if $Y^{a=1} \neq Y^{a=0}$ for the individual. Thus, the treatment has a causal effect on Zeus's outcome because $Y^{a=1} = 1 \neq 0 = Y^{a=0}$, but not on Hera's outcome because $Y^{a=1} = 0 = Y^{a=0}$. The variables $Y^{a=1}$ and $Y^{a=0}$ are referred to as *potential outcomes* or as *counterfactual outcomes*. Some authors prefer the

Average Causal Effect

The average causal effect (ACE) is a measure that quantifies the average treatment effect in a population. It provides an estimation of the average difference in outcomes between the treated and control groups, taking into account the potential outcomes under each treatment condition.

Individual treatment effect:

$$Y(1) - Y(0)$$

Average treatment effect (ATE):

$$\begin{aligned}\tau &\triangleq \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y(1) - Y(0)], \\ \text{ATE } \mathbb{E}[Y(1) - Y(0)] &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \text{ (Binary)}\end{aligned}$$

Associational difference:

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

They are not equal due to confounding

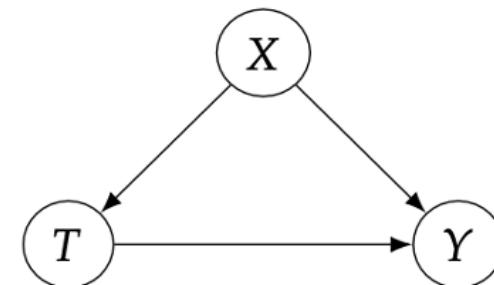


Figure 2.1: Causal structure of X confounding the effect of T on Y .

Measures of causal effect

(i) $\Pr[Y^{a=1} = 1] - \Pr[Y^{a=0} = 1] = 0$ Causal risk difference

(ii) $\frac{\Pr[Y^{a=1} = 1]}{\Pr[Y^{a=0} = 1]} = 1$ Risk ratio

The causal risk ratio (multiplicative scale) is used to compute how many times treatment, relative to no treatment, increases the disease risk.

The causal risk difference (additive scale) is used to compute the absolute number of cases of the disease attributable to the treatment.

The use of either the multiplicative or additive scale will depend on the goal of the inference.

What assumption(s) would make it so that the ATE is simply the associational difference?

i	T	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	?	?

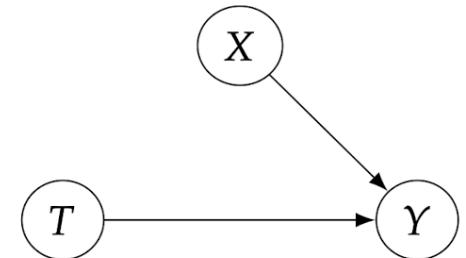


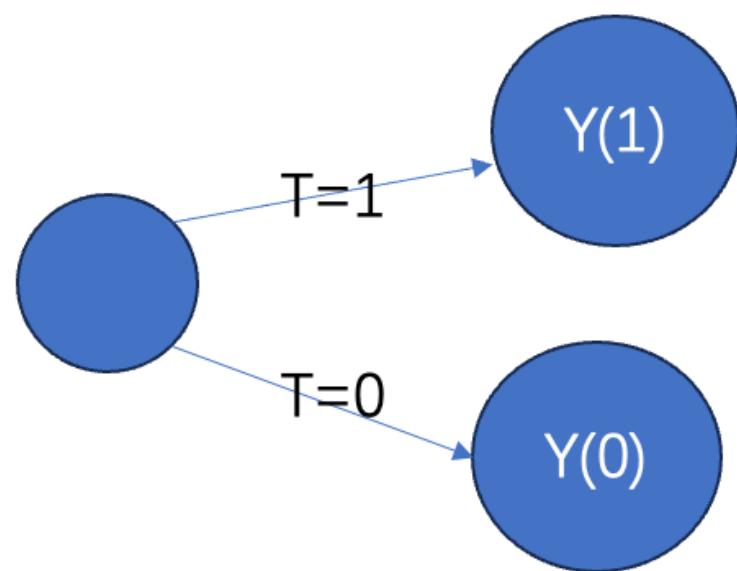
Figure 2.2: Causal structure when the treatment assignment mechanism is ignorable. Notably, this means there's no arrow from X to T , which means there is no confounding.

What makes it valid 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?

Assumption 2.1 (Ignorability / Exchangeability)

$$(Y(1), Y(0)) \perp\!\!\!\perp T$$

Another perspective on this assumption is that of exchangeability



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

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

$$\mathbb{E}[Y(1)|T = t] = \mathbb{E}[Y(1)]$$

$$\mathbb{E}[Y(0)|T = t] = \mathbb{E}[Y(0)]$$

This assumption is key to causal inference because it allows us to reduce the ATE to the associational difference:

$$\begin{aligned}\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] &= \mathbb{E}[Y(1) \mid T = 1] - \mathbb{E}[Y(0) \mid T = 0] \\ &= \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0]\end{aligned}$$

However, there is no reason to expect that the groups are the same in all relevant variables other than the treatment.

Assumption 2.2 (Conditional Exchangeability / Unconfoundedness)

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid X$$

We do not have exchangeability in the data because X is a common cause of T and Y .

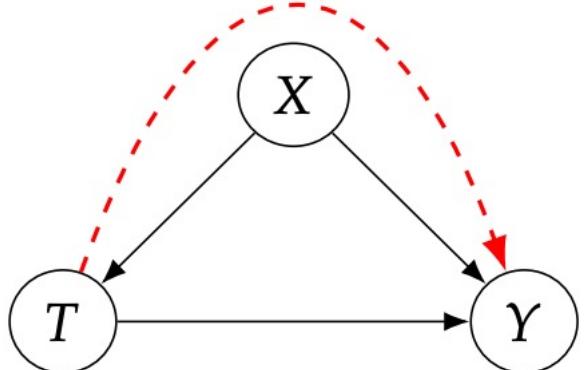


Figure 2.3: Causal structure of X **confounding** the effect of T on Y . We depict the confounding with a red dashed line.

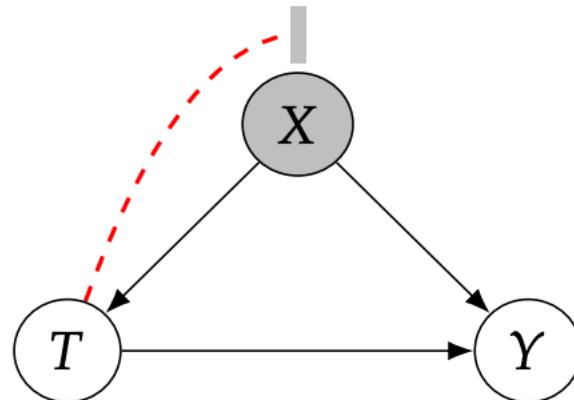


Figure 2.4: Illustration of conditioning on X leading to no confounding.

However, we do have conditional exchangeability in the data.

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

Adjustment Formula

Theorem 2.1 (Adjustment Formula) *Given the assumptions of unconfoundedness, positivity, consistency, and no interference, we can identify the average treatment effect:*

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

Definition 2.1 (Identifiability) *A causal quantity (e.g. $\mathbb{E}[Y(t)]$) is identifiable if we can compute it from a purely statistical quantity (e.g. $\mathbb{E}[Y | t]$).*

Assumption 2.3 (Positivity / Overlap / Common Support) *For all values of covariates x present in the population of interest (i.e. x such that $P(X = x) > 0$),*

$$0 < P(T = 1 \mid X = x) < 1$$

For discrete covariates and outcome, adjustment formula can be rewritten as follows:

$$\sum_x P(X = x) \left(\sum_y y P(Y = y \mid T = 1, X = x) - \sum_y y P(Y = y \mid T = 0, X = x) \right) \quad (2.10)$$

Then, applying Bayes' rule, this can be further rewritten:

$$\sum_x P(X = x) \left(\sum_y y \frac{P(Y = y, T = 1, X = x)}{P(T = 1 \mid X = x)P(X = x)} - \sum_y y \frac{P(Y = y, T = 0, X = x)}{P(T = 0 \mid X = x)P(X = x)} \right) \quad (2.11)$$

No interference means that my outcome is unaffected by anyone else's treatment

Assumption 2.4 (No Interference)

$$Y_i(t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_n) = Y_i(t_i)$$

Consistency is the assumption that the outcome we observe Y is actually the potential outcome under the observed treatment T

Assumption 2.5 (Consistency) *If the treatment is T , then the observed outcome Y is the potential outcome under treatment T . Formally,*

$$T = t \implies Y = Y(t) \tag{2.12}$$

We could write this equivalently as follow:

$$Y = Y(T) \tag{2.13}$$

Randomized Experiment: One perspective

Randomization implies covariate balance, across all covariates, even unobserved ones.

Definition 5.1 (Covariate Balance) *We have covariate balance if the distribution of covariates X is the same across treatment groups. More formally,*

$$P(X \mid T = 1) \stackrel{d}{=} P(X \mid T = 0) \quad (5.1)$$

Randomization makes causation equal to association

$$P(y \mid do(t)) = P(y \mid t)$$

Another perspective: In DAG

confounding association

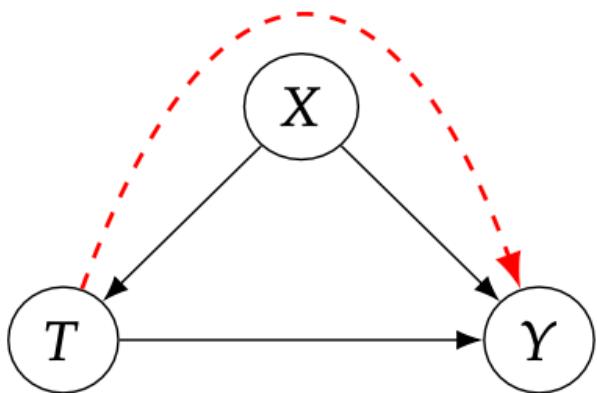


Figure 5.1: Causal structure of X confounding the effect of T on Y .

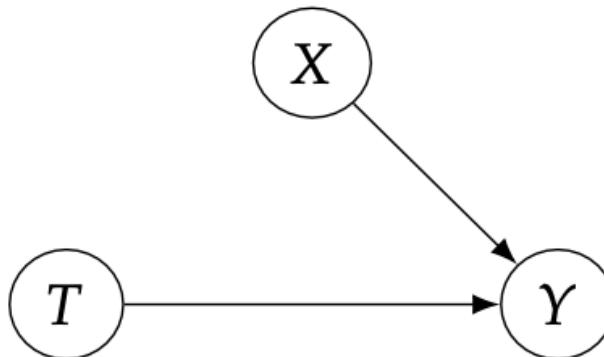


Figure 5.2: Causal structure when we randomize treatment.

X is a confounder of the effect of T on Y . Non-causal association flows along the backdoor path $T \leftarrow X \rightarrow Y$.

However, if we randomize T , something magical happens:
 T no longer has any causal parents, as we depict in Figure 5.2.

$$P(Y \mid do(T = t)) = P(Y \mid T = t)$$

Standardisation

- **Standardisation**

This method is known in epidemiology, demography, and other disciplines as *standardization*. For example, the numerator $\sum_l \Pr[Y = 1 | L = l, A = 1] \Pr[L = l]$ of the causal risk ratio is the standardized risk in the treated using the population as the standard. Under conditional exchangeability, this standardized risk can be interpreted as the (counterfactual) risk that would have been observed had all the individuals in the population been treated.

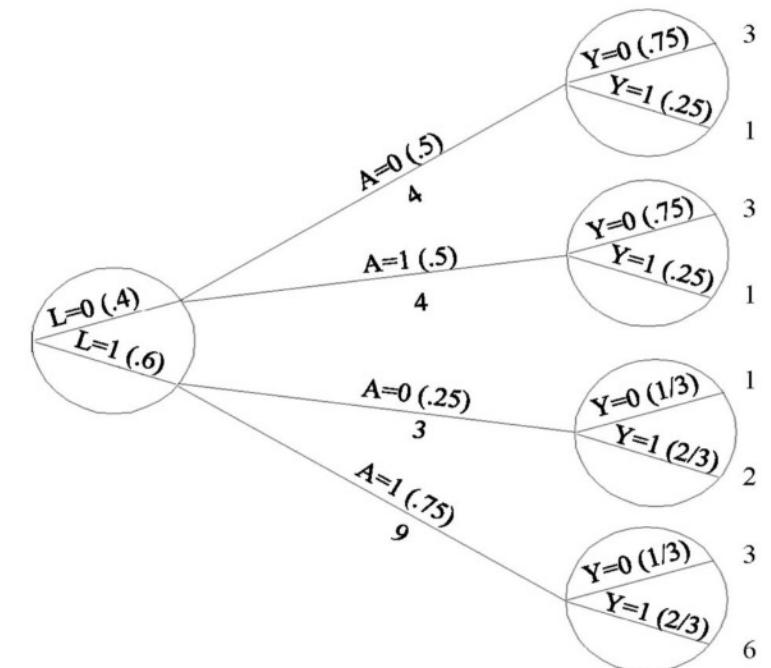
- **Standardized risk**

The standardized risks in the treated and the untreated are equal to the counterfactual risks under treatment and no treatment, respectively. Therefore, the causal risk ratio $\frac{\Pr[Y^{a=1} = 1]}{\Pr[Y^{a=0} = 1]}$ can be computed by standardization as

$$\frac{\sum_l \Pr[Y = 1 | L = l, A = 1] \Pr[L = l]}{\sum_l \Pr[Y = 1 | L = l, A = 0] \Pr[L = l]}.$$

Standardized mean

$$\sum_l E[Y | L = l, A = a] \times \Pr[L = l]$$



Inverse probability weighting

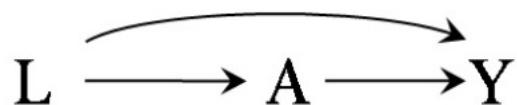
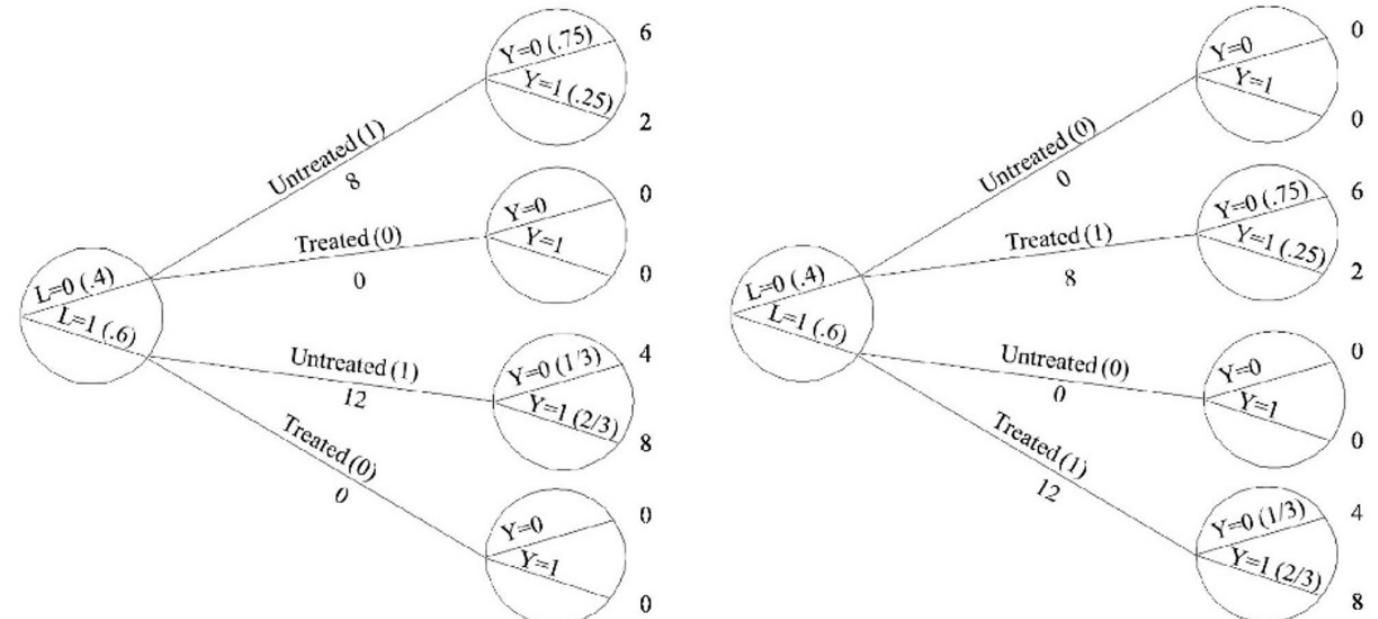


Figure 6.1

The circles contain the bifurcations defined by non treatment variables.

$$\frac{\Pr[Y^a=1 = 1]}{\Pr[Y^a=0 = 1]}$$



Effect modification

Definition:

V is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of V . Since the average causal effect can be measured using different effect measures (e.g., risk difference, risk ratio), the presence of effect modification depends on the effect measure being used.

Effect modification on additive $E[Y^{a=1} - Y^{a=0} | V = 1] \neq E[Y^{a=1} - Y^{a=0} | V = 0]$

Effect modification on multiplicative $\frac{E[Y^{a=1} | V = 1]}{E[Y^{a=0} | V = 1]} \neq \frac{E[Y^{a=1} | V = 0]}{E[Y^{a=0} | V = 0]}$)

Matching

Stratification

Stratification: the causal effect of A on Y is computed in each stratum of V . For dichotomous V , the stratified causal risk differences are:

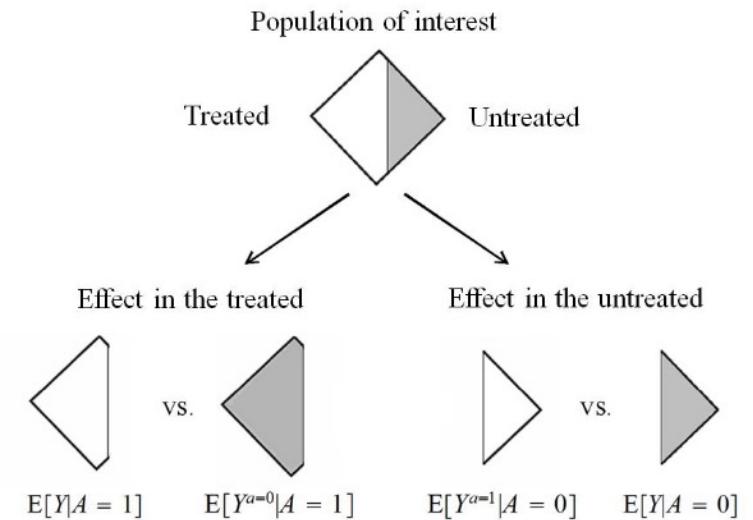
$$\Pr[Y^{a=1} = 1|V = 1] -$$

$$\Pr[Y^{a=0} = 1|V = 1]$$

and

$$\Pr[Y^{a=1} = 1|V = 0] -$$

$$\Pr[Y^{a=0} = 1|V = 0]$$



Matching is another form of adjustment used in causal inference to control for confounding variables and reduce bias. It involves creating pairs or groups of individuals with similar characteristics, based on the confounding variables, in order to compare their treatment effects.

Confounding

Definition: A confounder is a third variable that affects both the treatment and the outcome. Because of its association with both, it can create a false impression that there is a direct relationship between the treatment and the outcome when, in fact, the association is due to this third variable.

The structure of confounding

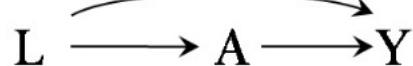


Figure 7.1

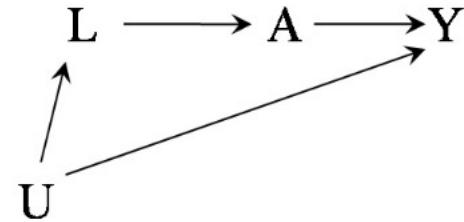


Figure 7.2

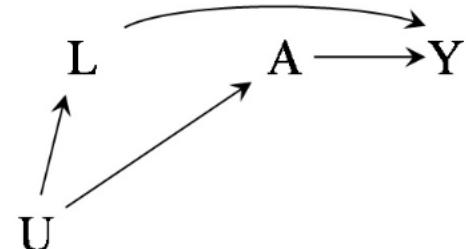


Figure 7.3

unmeasured cause U

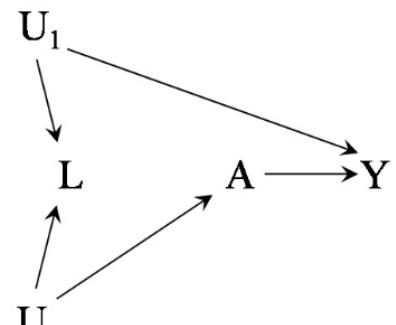


Figure 7.4

Causal Diagram(Pearl)

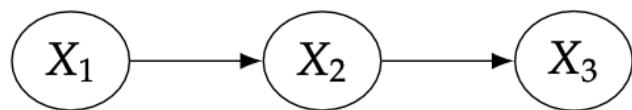
Key Assumptions

Assumption 3.2 (Minimality Assumption) *1. Given its parents in the DAG, a node X is independent of all its non-descendants (Assumption 3.1).*

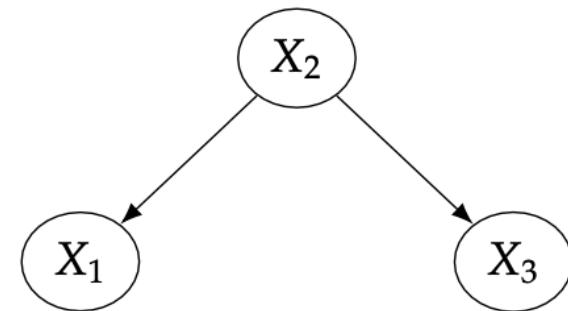
2. Adjacent nodes in the DAG are dependent.³

Assumption 3.3 ((Strict) Causal Edges Assumption) *In a directed graph, every parent is a direct cause of all its children.*

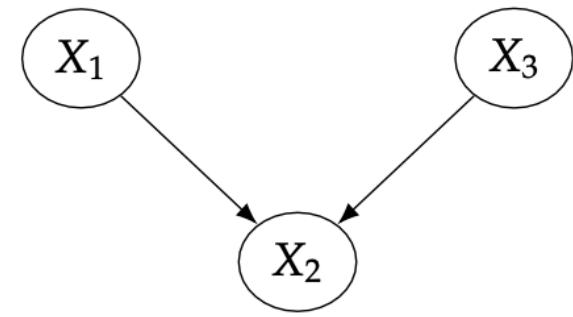
Basic Graphs



(a) Chain



(b) Fork



(c) Immorality

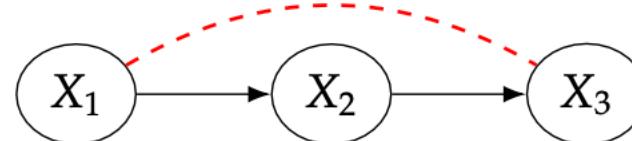


Figure 3.9: Basic graph building blocks

Figure 3.12: Chain with flow of association drawn as a dashed red arc.

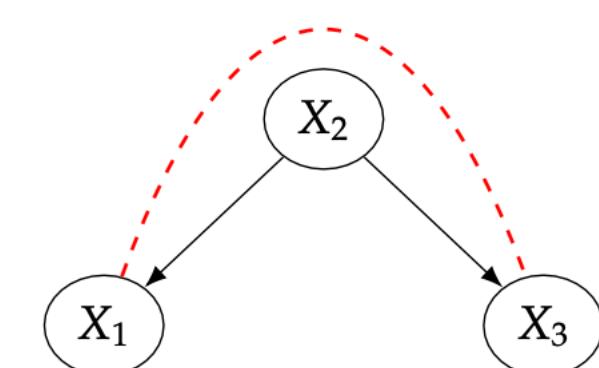


Figure 3.13: Fork with flow of association drawn as a dashed red arc.

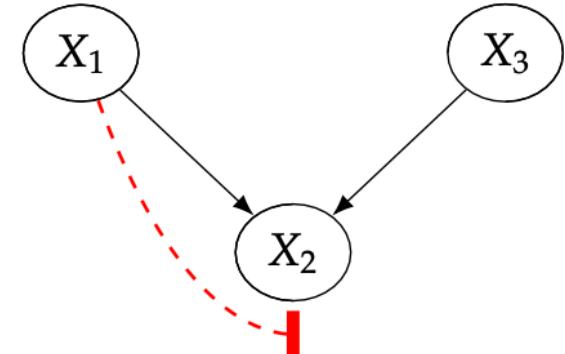


Figure 3.16: Immorality with association blocked by a collider.

Condition

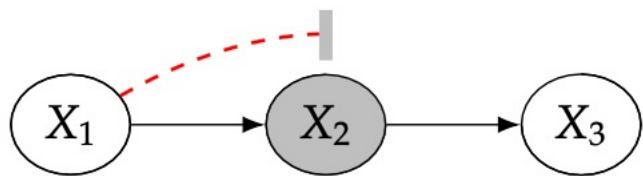


Figure 3.14: Chain with **association** blocked by conditioning on X_2 .

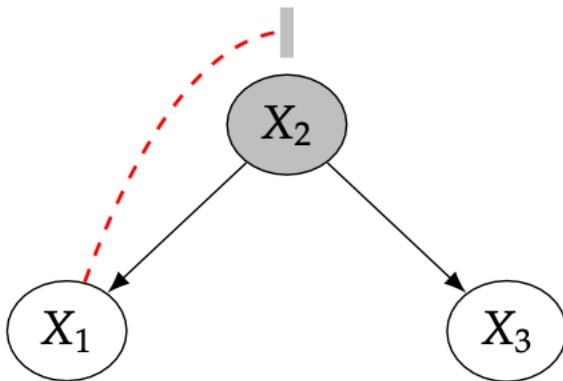


Figure 3.15: Fork with **association** blocked by conditioning on X_2 .

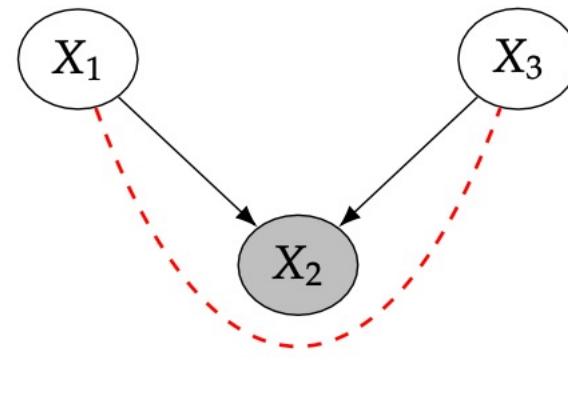
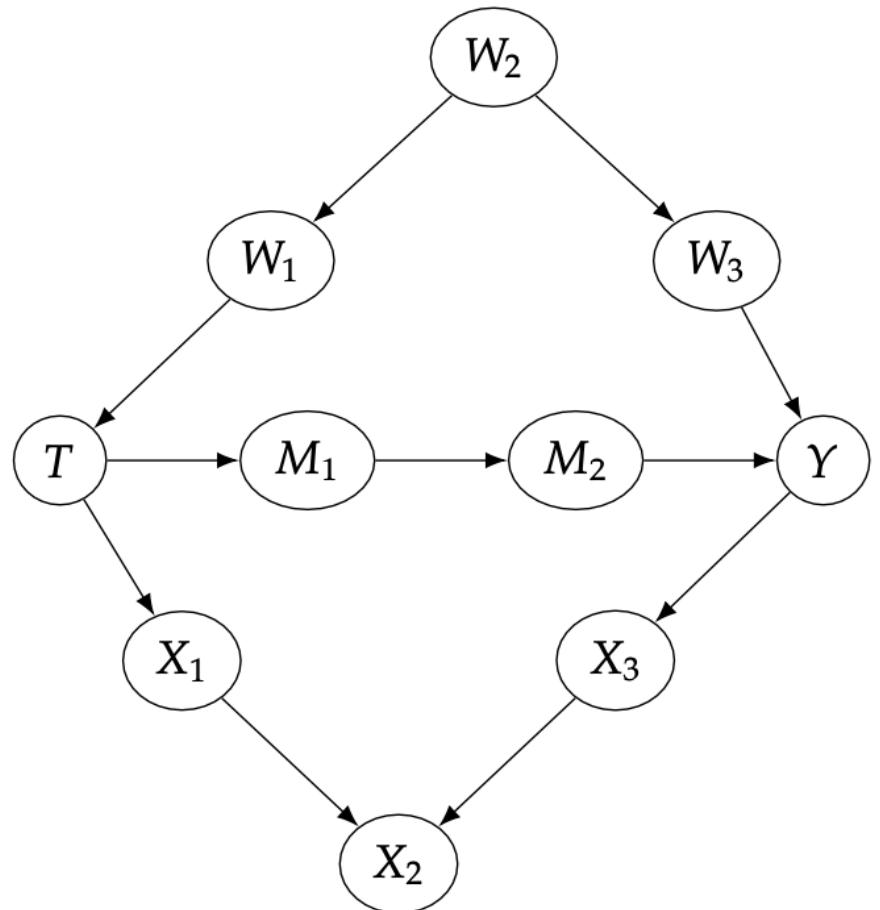


Figure 3.17: Immorality with **association** unblocked by conditioning on the collider.

Definition 3.4 (d-separation) Two (sets of) nodes X and Y are *d-separated* by a set of nodes Z if all of the paths between (any node in) X and (any node in) Y are blocked by Z [16].



(a)

1. Are T and Y d-separated by the empty set?
2. Are T and Y d-separated by W_2 ?
3. Are T and Y d-separated by $\{W_2, M_1\}$?
4. Are T and Y d-separated by $\{W_1, M_2\}$?
5. Are T and Y d-separated by $\{W_1, M_2, X_2\}$?
6. Are T and Y d-separated by $\{W_1, M_2, X_2, X_3\}$?

Flow of Association and Causation

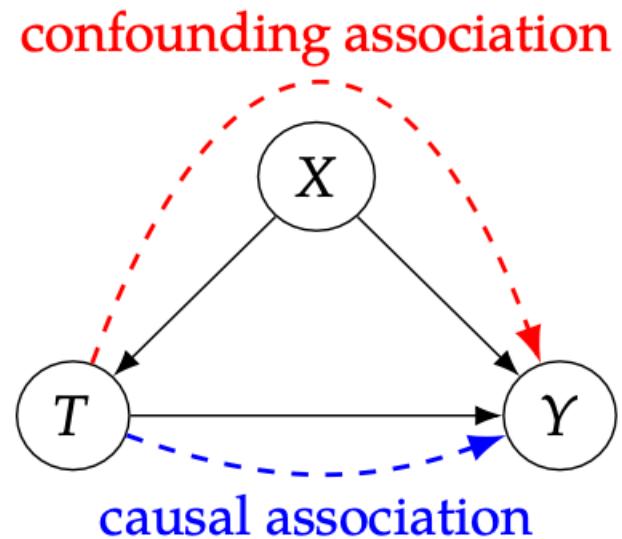


Figure 3.20: Causal graph depicting an example of how confounding association and causal association flow.

The *do*-operator and Interventional Distributions

$$P(Y(t) = y) \triangleq P(Y = y \mid do(T = t)) \triangleq P(y \mid do(t))$$

$$\text{ATE} \quad \mathbb{E}[Y \mid do(T = 1)] - \mathbb{E}[Y \mid do(T = 0)]$$

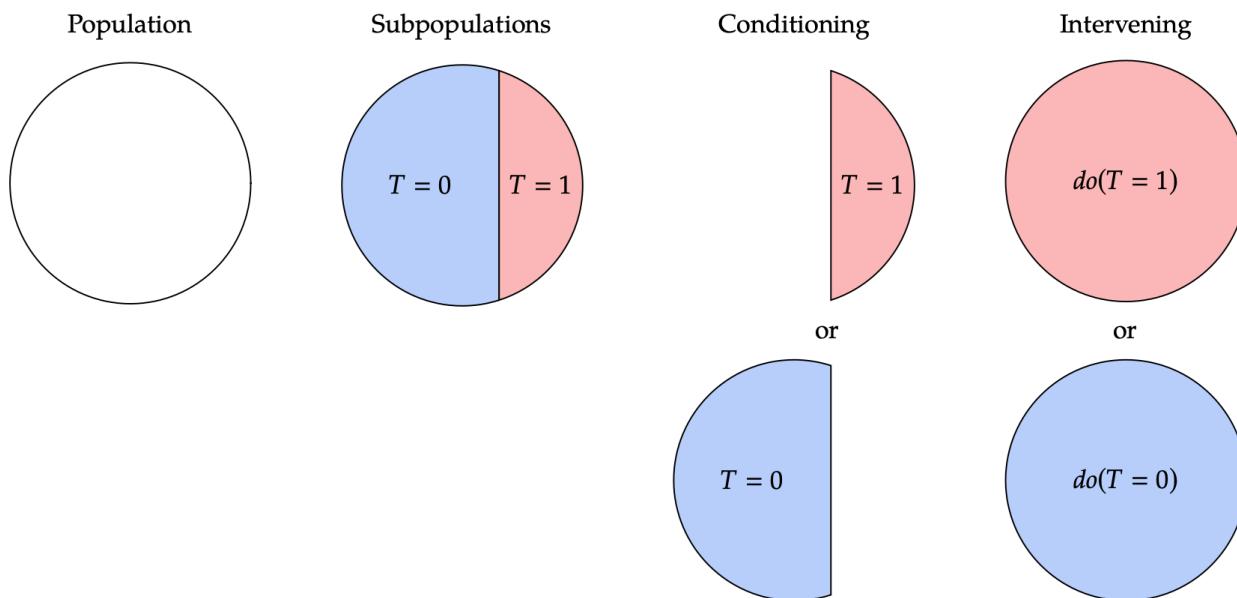


Figure 4.2: Illustration of the difference between conditioning and intervening

The Backdoor Adjustment

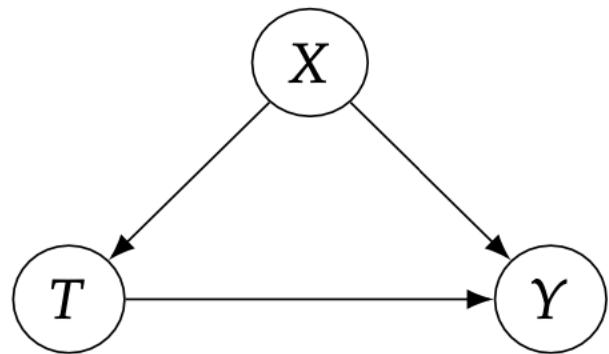


Figure 4.5: Simple causal structure where X confounds the effect of T on Y and where X is the only confounder.

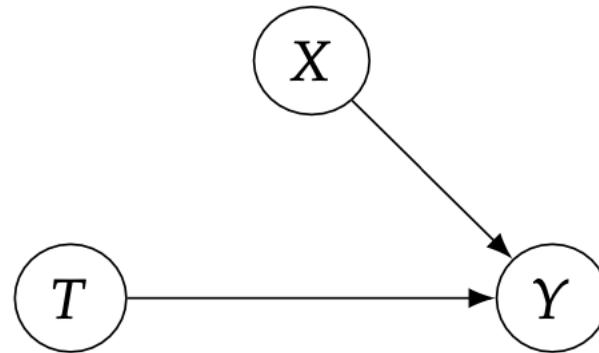


Figure 4.6: Manipulated graph that results from intervening on T , when the original graph is Figure 4.5.

Definition 4.1 (Backdoor Criterion) *A set of variables W satisfies the backdoor criterion relative to T and Y if the following are true:*

1. *W blocks all backdoor paths from T to Y .*
2. *W does not contain any descendants of T .*

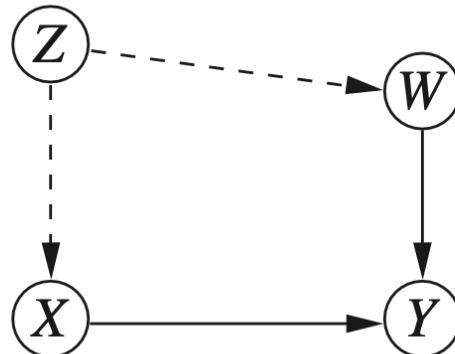


Figure 3.6 A graphical model representing the relationship between a new drug (X), recovery (Y), weight (W), and an unmeasured variable Z (socioeconomic status)

Here we are trying to gauge the effect of a drug (X) on recovery (Y). We have also measured weight (W), which has an effect on recovery. Further, we know that socioeconomic status (Z) affects both weight and the choice to receive treatment—but the study we are consulting did not record socioeconomic status.

Instead, we search for an observed variable that fits the backdoor criterion from X to Y . A brief examination of the graph shows that W , which is not a descendant of X , also blocks the backdoor path $X \leftarrow Z \rightarrow W \rightarrow Y$. Therefore, W meets the backdoor criterion. So long as the causal story conforms to the graph in Figure 3.6, adjusting for W will give us the causal effect of X on Y . Using the adjustment formula, we find

$$P(Y = y|do(X = x)) = \sum_w P(Y = y|X = x, W = w)P(W = w)$$

This sum can be estimated from our observational data, so long as W is observed.

Proof

$$\begin{aligned} P(y \mid do(t)) &= \sum_w P(y \mid do(t), w) P(w \mid do(t)) \\ &= \sum_w P(y \mid t, w) P(w \mid do(t)) \\ &= \sum_w P(y \mid t, w) P(w) \end{aligned}$$

Relationship to d-separation

W is the set we use to make t and y are d separated in another path.

Relation to Potential Outcomes

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_W [\mathbb{E}[Y | T = 1, W] - \mathbb{E}[Y | T = 0, W]] \quad \text{Potential outcome adjustment formula}$$

$$\mathbb{E}[Y | do(t)] = \sum_w \mathbb{E}[Y | t, w] P(w) \quad \text{Backdoor adjustment formula}$$

$$\mathbb{E}[Y | do(t)] = \mathbb{E}_W \mathbb{E}[Y | t, W]$$

$$\mathbb{E}[Y | do(T = 1)] - \mathbb{E}[Y | do(T = 0)] = \mathbb{E}_W [\mathbb{E}[Y | T = 1, W] - \mathbb{E}[Y | T = 0, W]]$$

$$(Y(1), Y(0)) \perp\!\!\!\perp T | W \quad \text{Conditional exchangeability}$$

using graphical causal models, we know how to choose a valid W : we simply choose W so that it satisfies the backdoor criterion.

The Front-Door Criterion

The front door criterion. The causal diagram in Figure 7.14 depicts a setting in which the treatment A and the binary outcome Y share an unmeasured cause U , and in which there is a variable M that fully mediates the effect of A on Y and that shares no unmeasured causes with either A or Y . Under this causal structure, a data analyst cannot directly use standardization (nor IP weighting) to compute the counterfactual risks $\Pr [Y^{a=1} = 1]$ and $\Pr [Y^{a=0} = 1]$ because the variable U , which is necessary to block the backdoor path between A and Y , is not available. Therefore, the average causal effect of A on Y cannot be identified using the methods described in previous chapters. However, Pearl (1995) showed that $\Pr [Y^a = 1]$ is identified by the so-called *front door formula*

$$\sum_m \Pr [M = m | A = a] \sum_{a'} \Pr [Y = 1 | M = m, A = a'] \Pr [A = a']$$

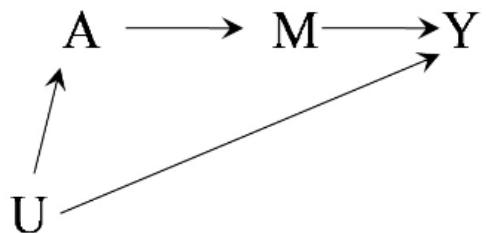


Figure 7.14

The Front-Door Criterion

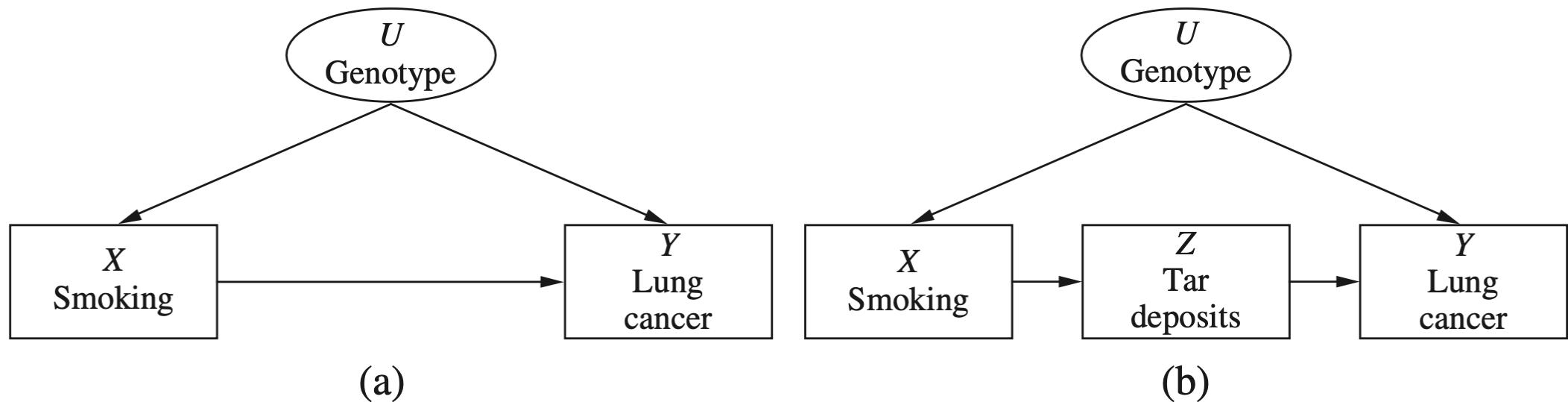


Figure 3.10 A graphical model representing the relationships between smoking (X) and lung cancer (Y), with unobserved confounder (U) and a mediating variable Z

Structural Causal Models (SCMs)

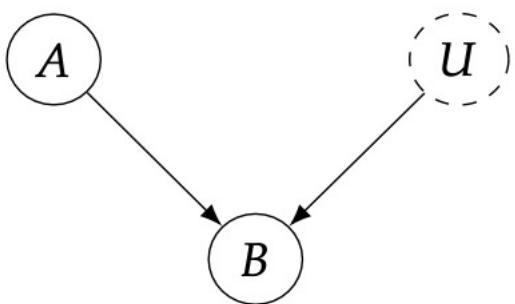


Figure 4.7: Graph for simple structural equation. The dashed node U means that U is unobserved.

$$B := f(A, U),$$

where U is some unobserved random variable. We depict this in Figure 4.7, where U is drawn inside a dashed node to indicate that it is unobserved. The unobserved U is analogous to the randomness that we would see by sampling units (individuals); it denotes all the relevant (noisy) background conditions that determine B . More concretely, there are analogs to every part of the potential outcome $Y_i(t)$: B is the analog of Y , $A = a$ is the analog of $T = t$, and U is the analog of i .

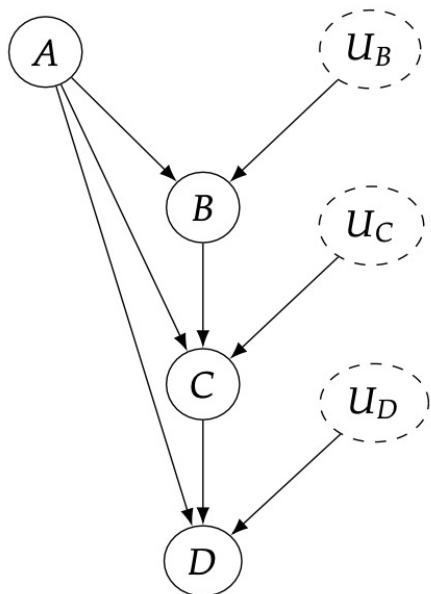


Figure 4.8: Graph for the structural equations in Equation 4.24.

$$B := f_B(A, U_B)$$

$$M : \quad C := f_C(A, B, U_C)$$

$$D := f_D(A, C, U_D)$$

Definition

Definition 4.2 (Structural Causal Model (SCM)) *A structural causal model is a tuple of the following sets:*

1. A set of endogenous variables V
2. A set of exogenous variables U
3. A set of functions f , one to generate each endogenous variable as a function of other variables

Interventions

$$M : \begin{aligned} T &:= f_T(X, U_T) \\ Y &:= f_Y(X, T, U_Y) \end{aligned}$$

$$M_t : \begin{aligned} T &:= t \\ Y &:= f_Y(X, T, U_Y) \end{aligned}$$

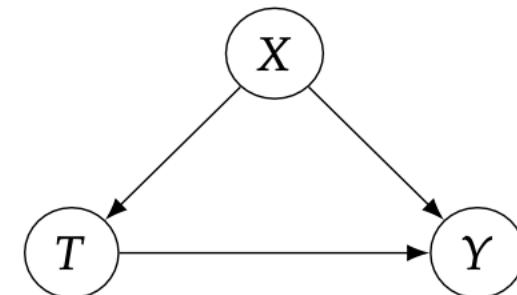


Figure 4.9: Basic causal graph

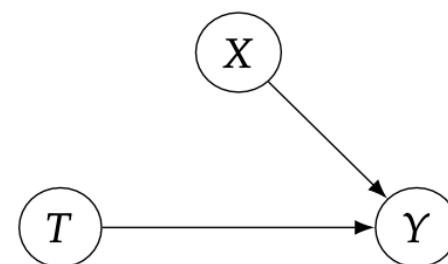


Figure 4.10: Basic causal with the incoming edges to T removed, due to the intervention $do(T = t)$.

Assumption 4.2 (Modularity Assumption for SCMs) Consider an SCM M and an interventional SCM M_t that we get by performing the intervention $\text{do}(T = t)$. The modularity assumption states that M and M_t share all of their structural equations except the structural equation for T , which is $T := t$ in M_t .

Definition 4.3 (The Law of Counterfactuals (and Interventions))

$$Y_t(u) = Y_{M_t}(u) \tag{4.27}$$

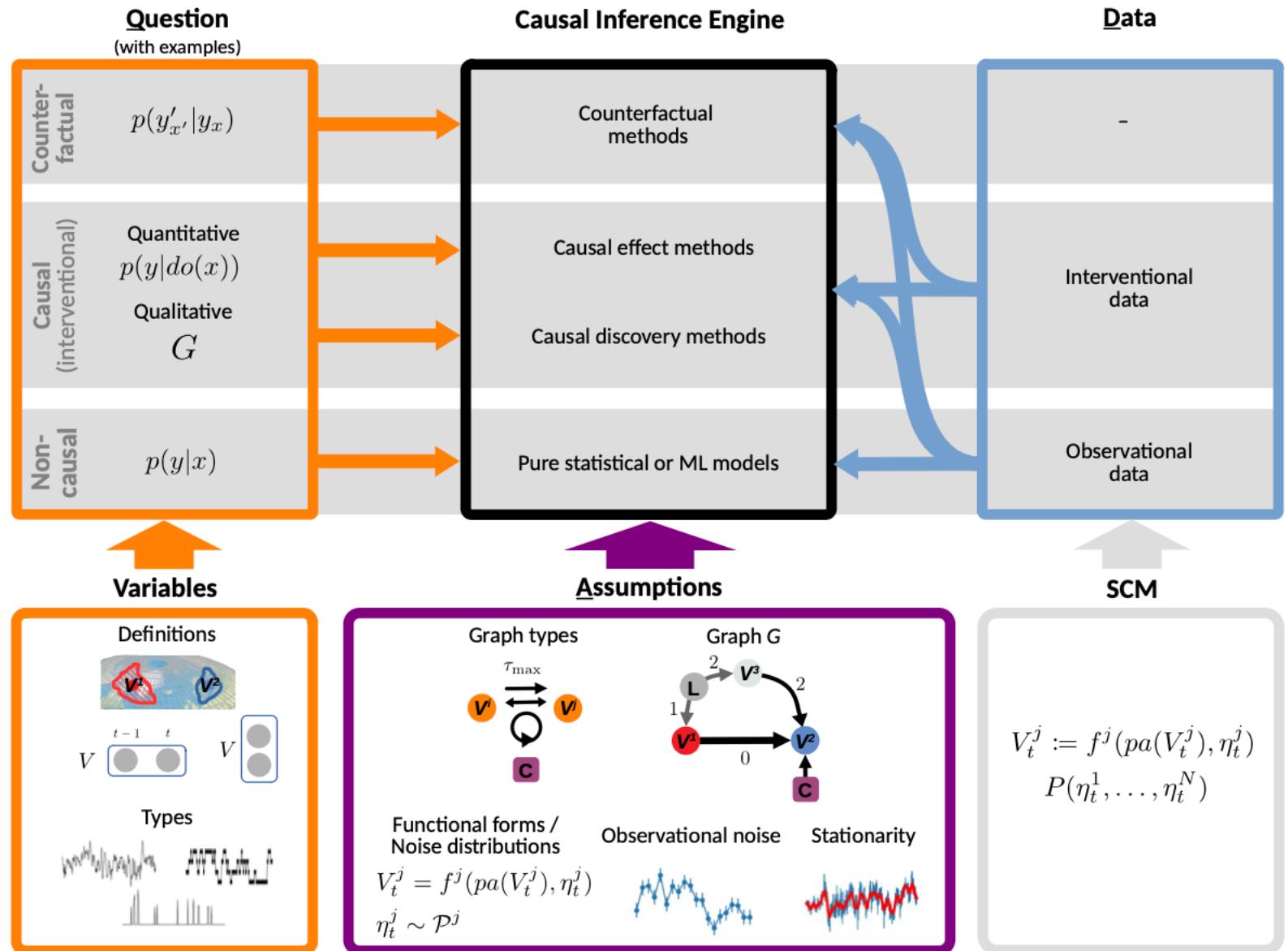
$Y_t(u)$ denotes the outcome that unit u would observe if they take treatment t , given that the SCM is M .

$Y_{M_t}(u)$ as the outcome that unit u would observe if they take treatment t , given that the SCM is M_t

Causal inference in time series(Overview)

Question-Assumptions-Data template.

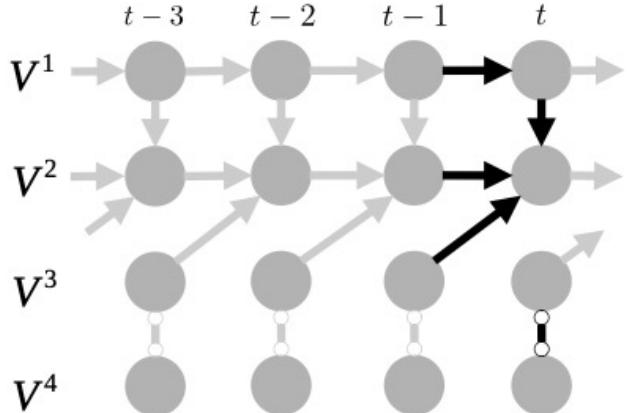
Problem defining in Causal inference



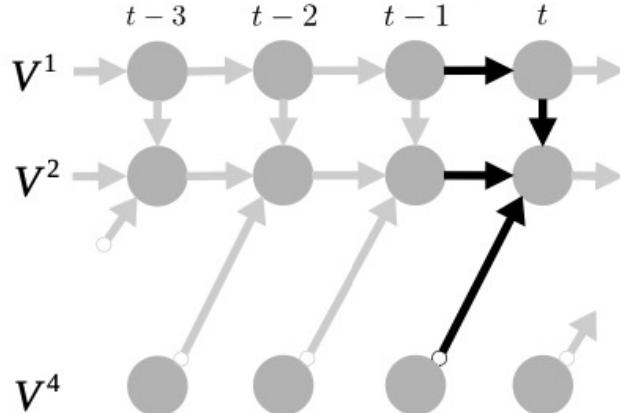
The problem transferred to Causality can be simply devided into two kinds of problem.

a Causal discovery

Assuming no hidden confounding

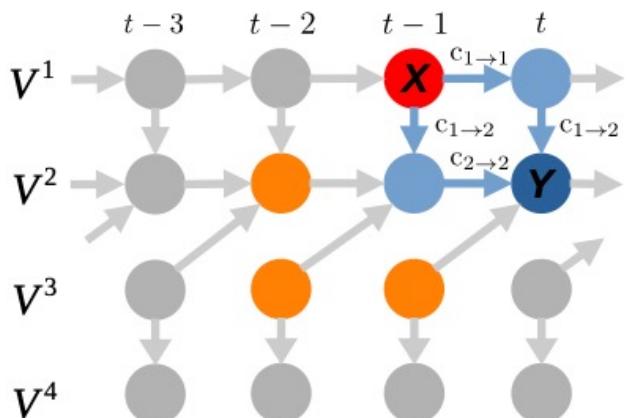


Allowing hidden confounding

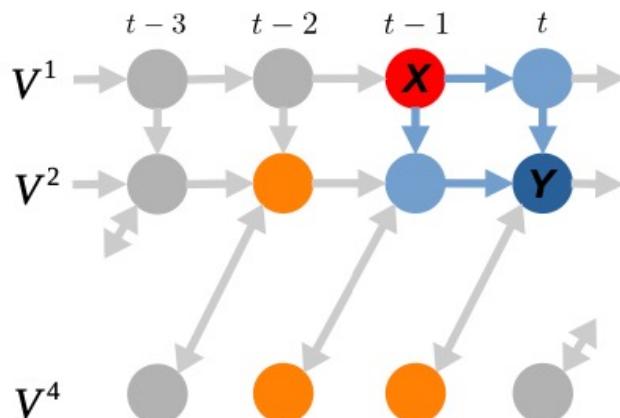


b Causal effect estimation

Graph without hidden variables



With hidden V^3



QAD-based causal inference method selector

