Methods of estimating Causal Effect

Abhinek Shukla

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

August 31, 2020

Outline of the talk

- Specification of the problem.

Outline of the talk

- Specification of the problem.
- Difficulties associated in estimation.

Outline of the talk

- Specification of the problem.
- Difficulties associated in estimation.
- Assumptions required to identify causal effect from data.

Outline of the talk

- Specification of the problem.
- Difficulties associated in estimation.
- Assumptions required to identify causal effect from data.
- Role of causal graphs, especially DAGs.

Outline of the talk

- Specification of the problem.
- Difficulties associated in estimation.
- Assumptions required to identify causal effect from data.
- Role of causal graphs, especially DAGs.
- Methods to control for confounding (observed or unobserved).

Notations

Suppose we want to study causal effect of some treatment A on some outcome Y (e.g. recover from a disease or develop it).

Notations

Suppose we want to study causal effect of some treatment A on some outcome Y (e.g. recover from a disease or develop it).

Potential outcomes can be thought of as the outcomes we would see under each possible treatment option. Let Y^a denote the outcome that would be observed under treatment $A = a$.

Notations

Suppose we want to study causal effect of some treatment A on some outcome Y (e.g. recover from a disease or develop it).

Potential outcomes can be thought of as the outcomes we would see under each possible treatment option. Let Y^a denote the outcome that would be observed under treatment $A = a$.

Then potential outcomes will be denoted by collection of values Y^a for each possible a considered under study.

- We will consider only causal effects of treatments that we can imagine being randomized or manipulated in a hypothetical trial.

- We will consider only causal effects of treatments that we can imagine being randomized or manipulated in a hypothetical trial.

- Causal effects of variables like age, race, gender etc. do not fit so cleanly in the potential outcome framework.

Questions that can be answered

Suppose treatment is binary, i.e. A takes values 0 and 1. Then in general, we can observe causal effect if the potential outcomes are not equal ($Y^0 \neq Y^1$).

Questions that can be answered

Suppose treatment is binary, i.e. A takes values 0 and 1. Then in general, we can observe causal effect if the potential outcomes are not equal ($Y^0 \neq Y^1$).

We can observe only one potential outcome for each unit !

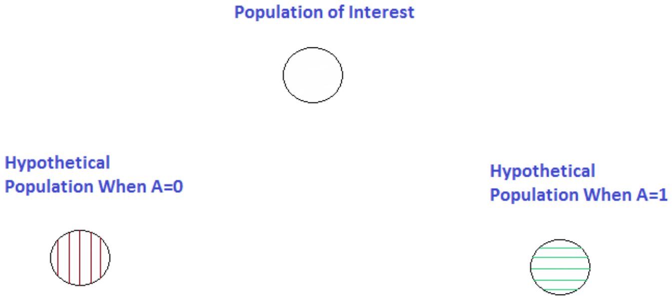
Questions that can be answered

Suppose treatment is binary, i.e. A takes values 0 and 1. Then in general, we can observe causal effect if the potential outcomes are not equal ($Y^0 \neq Y^1$).

We can observe only one potential outcome for each unit !

Do the outcomes on average differ if we treat whole population with treatment $A = 0$ versus $A = 1$?

Average Causal Effect



Difference of means $\mathbb{E}(Y^1) - \mathbb{E}(Y^0)$ will be the Average Causal Effect.

Subset versus Hypothetical

In practice, sub-populations might differ from population and what we can estimate from data is $\mathbb{E}(Y \big| A = 1) - \mathbb{E}(Y \big| A = 0)$

Subset versus Hypothetical

In practice, sub-populations might differ from population and what we can estimate from data is $\mathbb{E}(Y \big| A = 1) - \mathbb{E}(Y \big| A = 0)$

which is generally not equal to $\mathbb{E}(Y^1) - \mathbb{E}(Y^0)$.

Assumptions for estimation

- Stable Unit Treatment Value Assumption (SUTVA) : No interference

Assumptions for estimation

- Stable Unit Treatment Value Assumption (SUTVA) :
  No interference and one version of treatment;

Assumptions for estimation

- Stable Unit Treatment Value Assumption (SUTVA) :
  No interference and one version of treatment;

- Consistency : The potential outcome Y^a under treatment $A = a$, is equal to the observed outcome Y if the actual treatment received is $A = a$;

Assumptions for estimation

- Stable Unit Treatment Value Assumption (SUTVA) :
  No interference and one version of treatment;

- Consistency : The potential outcome Y^a under treatment $A = a$, is equal to the observed outcome Y if the actual treatment received is $A = a$;

- Positivity : For every set of covariates X, treatment assignment is such that $\mathbb{P}(A = a \big| X = x) > 0, \ \forall x$;

Assumptions for estimation

- Stable Unit Treatment Value Assumption (SUTVA) :
No interference and one version of treatment;

- Consistency : The potential outcome Y^a under treatment $A = a$, is equal to the observed outcome Y if the actual treatment received is $A = a$;

- Positivity : For every set of covariates X, treatment assignment is such that $\mathbb{P}(A = a \big| X = x) > 0, \ \forall x$;

- Ignorability : For particular $X = x$, treatment assignment is independent from potential outcomes, i.e.,

$$(Y^0, Y^1) \perp\!\!\!\perp A \big| X$$

Observed Data and Potential Outcomes

Due to consistency, we have :

$$\mathbb{E}(Y \big| A = a, X = x) = \mathbb{E}(Y^a \big| A = a, X = x)$$

$$= \mathbb{E}(Y^a \big| X = x) \quad \text{by ignorability}$$

Observed Data and Potential Outcomes

Due to consistency, we have :

$$\mathbb{E}(Y \big| A = a, X = x) = \mathbb{E}(Y^a \big| A = a, X = x)$$

$$= \mathbb{E}(Y^a \big| X = x) \quad \text{by ignorability}$$

Suppose X is categorical, then standardization gives :

$$\mathbb{E}(Y^a) = \sum_x \mathbb{E}(Y \big| A = a, X = x)\mathbb{P}(X = x) \tag{1}$$

Difficulties with Standardization

- In practice, there will be a lots of X variables required to achieve ignorability.
- Stratification would lead to no data case for many combinations of levels of covariates.

Difficulties with Standardization

- In practice, there will be a lots of X variables required to achieve ignorability.
- Stratification would lead to no data case for many combinations of levels of covariates.
- Some alternatives to standardization for estimating causal effects : matching, propensity score and inverse probability of treatment weighting.

Why Ignorability ?

We try to eliminate effect of treatment assignment mechanism on the potential outcomes so that we can only focus on direct effect of treatment.

Why Ignorability ?

We try to eliminate effect of treatment assignment mechanism on the potential outcomes so that we can only focus on direct effect of treatment.

But why do we care about the mechanism ?

Why Ignorability ?

We try to eliminate effect of treatment assignment mechanism on the potential outcomes so that we can only focus on direct effect of treatment.

But why do we care about the mechanism ?

In many cases, 'sicker' patients are more likely to be treated or treated patients might be at higher risk of some bad outcome.

Why Ignorability ?

We try to eliminate effect of treatment assignment mechanism on the potential outcomes so that we can only focus on direct effect of treatment.

But why do we care about the mechanism ?

In many cases, 'sicker' patients are more likely to be treated or treated patients might be at higher risk of some bad outcome.

Within levels of covariates (e.g., people of same age, with same co-morbid conditions etc.), treatment assignment would not be dependent on potential outcomes and hence, through assignment mechanism, no bias would arise.

Identify Covariates

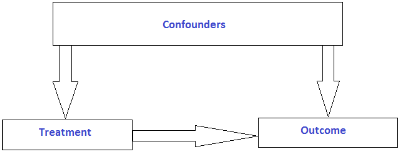


Figure: Real Situation

Identify Covariates

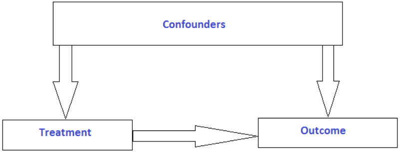


Figure: Real Situation

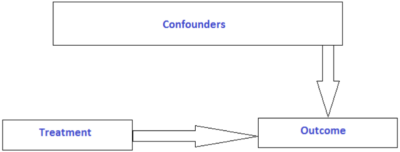


Figure: For Ignorability

Identify Covariates

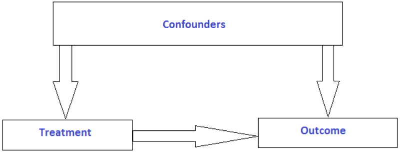


Figure: Real Situation

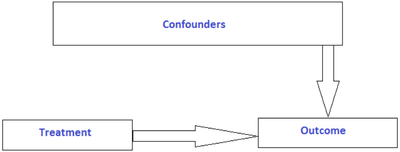


Figure: For Ignorability

If we clearly identify the variables we need to consider so as to achieve ignorability, then we can estimate average causal effect $\mathbb{E}(Y^1) - \mathbb{E}(Y^0)$ using equation (1).

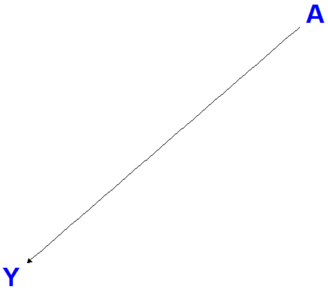
Causal Graphs

Causal Graphs will help us in determining the variables to control for.

Causal graphs encode assumptions about relationship among variables (i.e., which are independent, dependent and conditionally independent) along-with displaying explicit direction of effect from one to other variable.

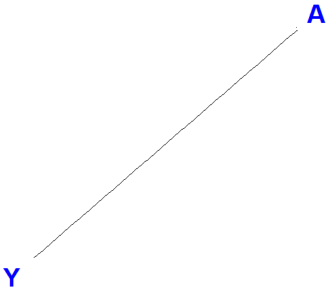
Directed vs. Undirected

Directed :



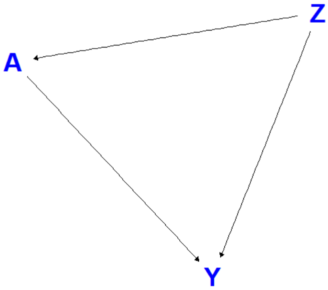
(a) A affects Y

Undirected Graph :

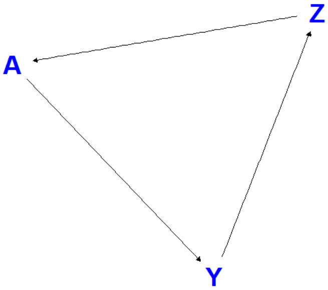


(b) A and Y are associated

Acyclic vs. Cyclic

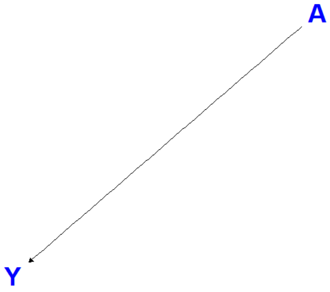


(a) Acyclic Graph



(b) Cyclic Graph

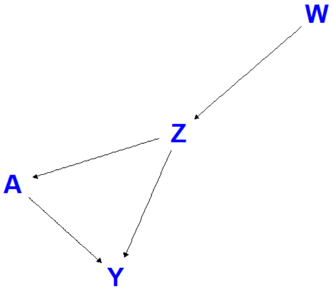
Terminology



A affects Y

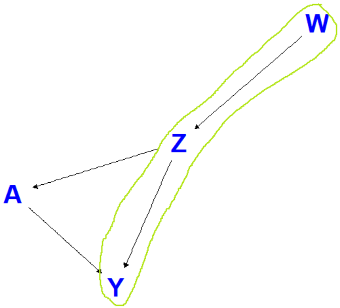
- Variables A and Y are known as nodes or vertices;
- Arrow form A to Y is an edge;
- Variables connected by an edge are adjacent.

Paths

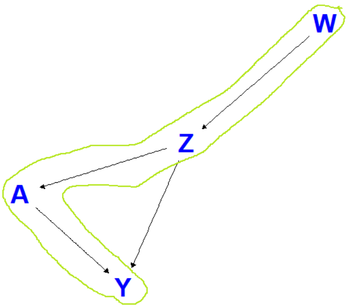


- A path is a way to get from vertex to another along edges;
- There are two paths from W to Y.

Paths

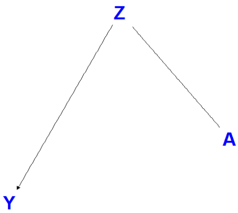


(a) $W \to Z \to Y$

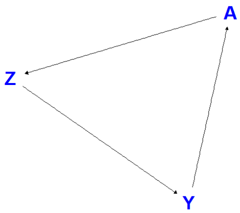


(b) $W \to Z \to A \to Y$

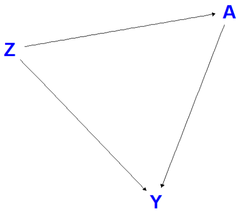
Directed Acyclic Graphs (DAGs)



(a) Undirected Acyclic

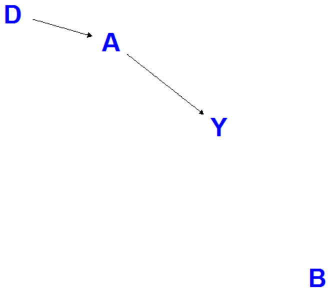


(b) Directed Cyclic



(c) DAG

Probabilistic interpretation of DAGs



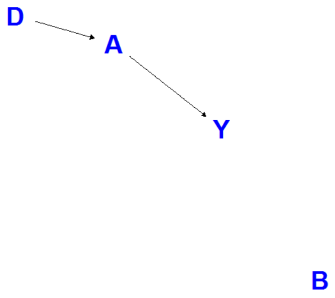
- $\mathbb{P}(B \big| A, D, Y) = \mathbb{P}(B)$
- $\mathbb{P}(Y \big| A, D, B) = \mathbb{P}(Y \big| A)$

$$\implies Y \perp\!\!\!\perp (D, B) \big| A$$

- $\mathbb{P}(D \big| A, B, Y) = \mathbb{P}(D \big| A)$

Example 1

Probabilistic interpretation of DAGs

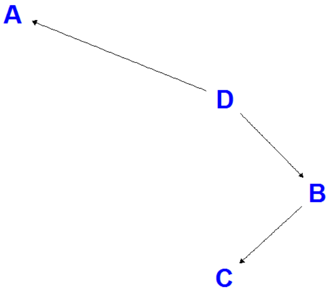


Example 1

- $\mathbb{P}(B \big| A, D, Y) = \mathbb{P}(B)$
- $\mathbb{P}(Y \big| A, D, B) = \mathbb{P}(Y \big| A)$

  $\implies Y \perp\!\!\!\perp (D, B) \big| A$
- $\mathbb{P}(D \big| A, B, Y) = \mathbb{P}(D \big| A)$
- No more independence ?

More examples on interpretation



Example 2

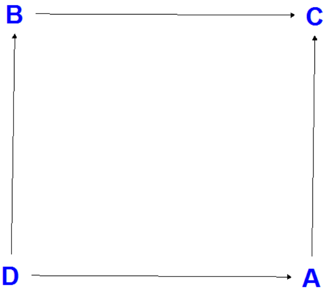
- $\mathbb{P}(D \big| A, C, B) = \mathbb{P}(D \big| A, B)$

  $\implies D \perp\!\!\!\perp C \big| B$

- $\mathbb{P}(A \big| C, D, B) = \mathbb{P}(A \big| D)$

  $\implies A \perp\!\!\!\perp (B, C) \big| D$

Examples Continued



Example 3

- $\mathbb{P}(D \big| A, C, B) = \mathbb{P}(D \big| A, B)$

  $\implies D \perp\!\!\!\perp C \big| (A, B)$

- $\mathbb{P}(A \big| C, D, B) = \mathbb{P}(A \big| D)$

  $\implies A \perp\!\!\!\perp B \big| (C, D)$

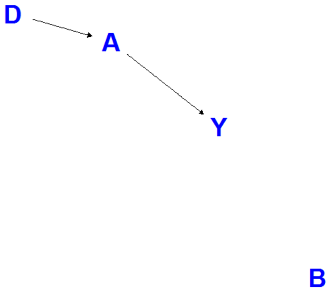
Joint Distribution

Suppose $X_1, X_2, ..., X_n$ are jointly distributed random variables, then by multiplication rule of conditional probability, we know that :

$$\mathbb{P}(X_1, X_2, X_3, ..., X_n) = \mathbb{P}(X_1) \times \mathbb{P}(X_2 \big| X_1) \times \mathbb{P}(X_3 \big| X_1, X_2) \times$$

$$... \times \mathbb{P}(X_n \big| X_1, X_2, X_3, ..., X_{n-1})$$

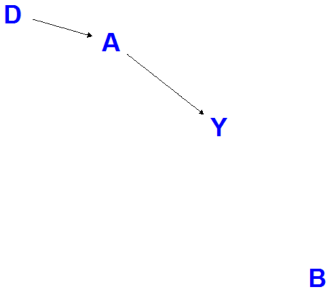
Decomposition Example 1



Example 1

$$\mathbb{P}(A, B, Y, D) =$$

Decomposition Example 1

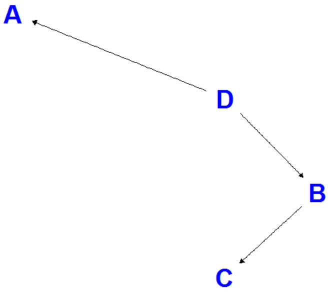


Example 1

$$\mathbb{P}(A, B, Y, D) =$$

$$\mathbb{P}(B) \times \mathbb{P}(D) \times \mathbb{P}(A \middle| D) \times \mathbb{P}(Y \middle| A)$$

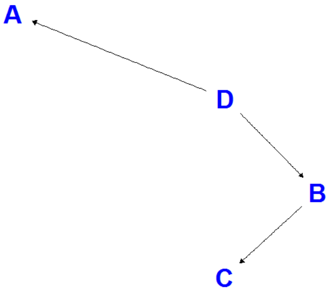
Decomposition Example 2



$$\mathbb{P}(A, B, C, D) =$$

Example 2

Decomposition Example 2

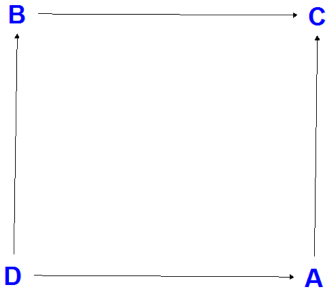


$$\mathbb{P}(A, B, C, D) =$$

$$\mathbb{P}(D) \times \mathbb{P}(B \middle| D) \times \mathbb{P}(A \middle| D) \times \mathbb{P}(C \middle| B)$$

Example 2

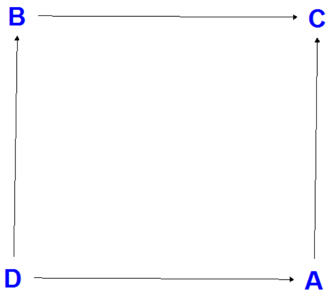
Decomposition Example 3



Example 3

$$\mathbb{P}(A, B, C, D) =$$

Decomposition Example 3



Example 3

$$\mathbb{P}(A, B, C, D) =$$

$$\mathbb{P}(D) \times \mathbb{P}(B \middle| D) \times \mathbb{P}(A \middle| D) \times \mathbb{P}(C \middle| A, B)$$

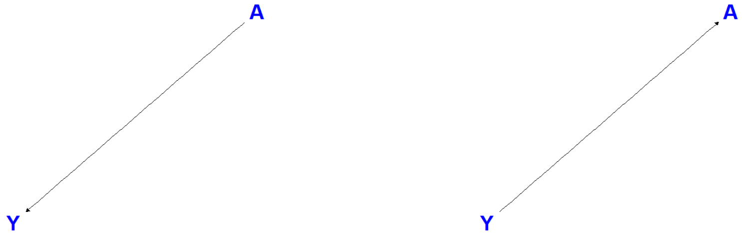
DAGs and Distributions

For a given DAG, there is a unique decomposition of joint probability distribution.

DAGs and Distributions

For a given DAG, there is a unique decomposition of joint probability distribution.

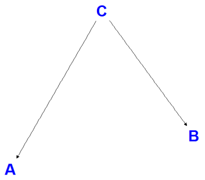
What about converse ?



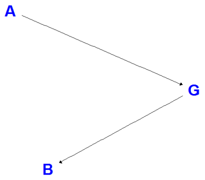
(a) A affects Y (b) Y affects A

Association through Paths

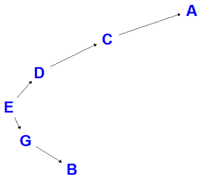
Nodes A and B are associated (via particular path) if :



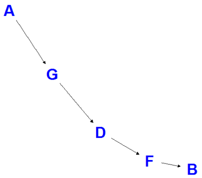
(a) Fork



(b) chain

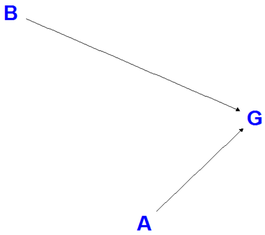


(c) Long Fork

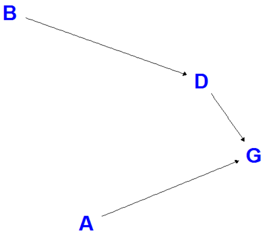


(d) Long Chain

Paths that do not induce Association

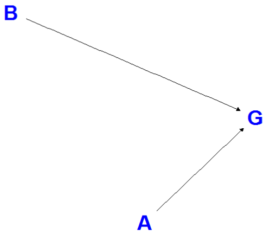


(a) Inverted Fork

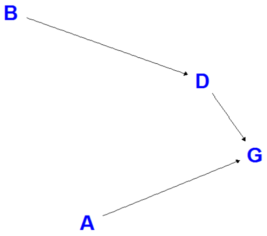


(b) Inverted Fork

Paths that do not induce Association



(a) Inverted Fork



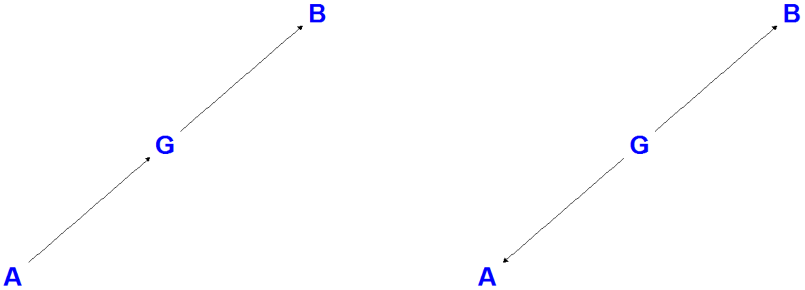
(b) Inverted Fork

Information flows from nodes A and B to G but collide and do not flow to each other which implies A and B are independent through the path.

G is known as collider

Blocking

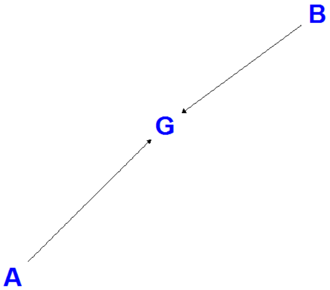
We can block the path by conditioning on nodes when nodes on the end are associated through this path.



(a) (b)

Blocking is not always a solution !

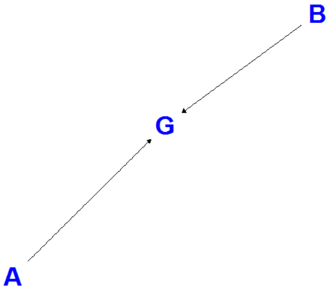
A and B are not associated marginally via this path.



G is a collider

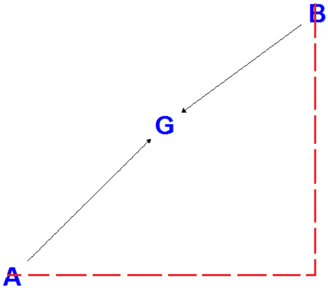
Blocking is not always a solution !

A and B are not associated marginally via this path.



G is a collider

But, conditioning on G induces association between A and B.



Conditioned on G

path d-separated

A path p is d-separated by a set of nodes S if :

path d-separated

A path p is d-separated by a set of nodes S if :

- p contains a chain $(D \rightarrow E \rightarrow F)$ and the middle part of chain is in S, or;

path d-separated

A path p is d-separated by a set of nodes S if :

- p contains a chain ($D \rightarrow E \rightarrow F$) and the middle part of chain is in S, or;
- p contains a fork ($D \leftarrow E \rightarrow F$) and the middle part of fork is in S, or;

path d-separated

A path p is d-separated by a set of nodes S if :

- p contains a chain $(D \rightarrow E \rightarrow F)$ and the middle part of chain is in S, or;
- p contains a fork $(D \leftarrow E \rightarrow F)$ and the middle part of fork is in S, or;
- p contains an inverted fork $(D \rightarrow E \leftarrow F)$ and middle part is not in S, nor any of its descendants are in S.

d-separation

Two nodes, A and B are d-separated by a set of nodes S if it blocks every path from A to B.

Above implies, $A \perp\!\!\!\perp B \big| S$

d-separation

Two nodes, A and B are d-separated by a set of nodes S if it blocks every path from A to B.

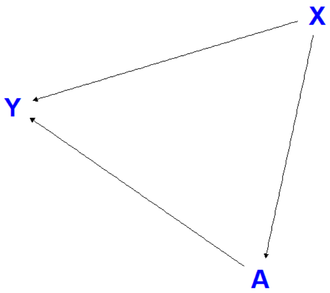
Above implies, $A \perp\!\!\!\perp B \big| S$

Ignorability satisfied ?

$(Y^0, Y^1) \perp\!\!\!\perp A \big| X$

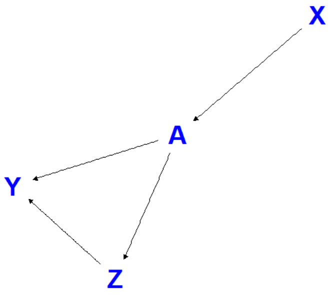
Frontdoor Path

Frontdoor path from A (treatment or exposure) to Y (outcome) is the path that begins with an arrow emanating out of A.



- $A \rightarrow Y$ is a frontdoor path from A to Y.
- Here, A affects Y directly.

Effect through Frontdoor



- $A \rightarrow Z \rightarrow Y$ is a frontdoor path from A to Y.
- Here, A affects Y indirectly through its effect on Z.

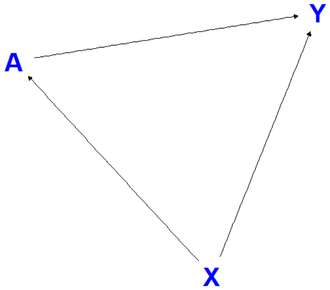
We care about how Y is affected by manipulating A !

Controlling for Z would be controlling for some portion of affect of treatment which is not desired.

Backdoor paths

Causal Mediation Analysis focuses on quantifying effect of treatment to outcome through frontdoor paths.

Backdoor paths from A to Y are paths from A to Y that travel through arrows going into A.

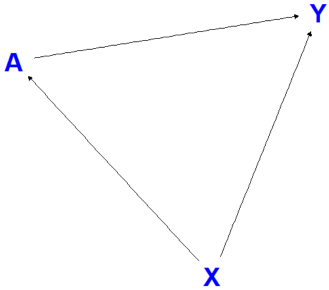


$A \leftarrow X \rightarrow Y$ is backdoor path from A to Y

Backdoor paths

Causal Mediation Analysis focuses on quantifying effect of treatment to outcome through frontdoor paths.

Backdoor paths from A to Y are paths from A to Y that travel through arrows going into A.



Backdoor paths confound the relationship between A and Y.

$A \leftarrow X \rightarrow Y$ is backdoor path from A to Y

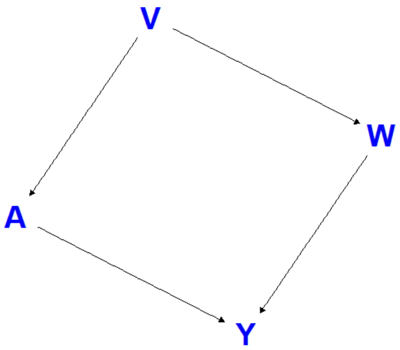
Criterion

A set of variables X is sufficient to control for confounding if:

- it blocks all backdoor paths from treatment to the outcome.
- it does not include any descendants of the treatment.

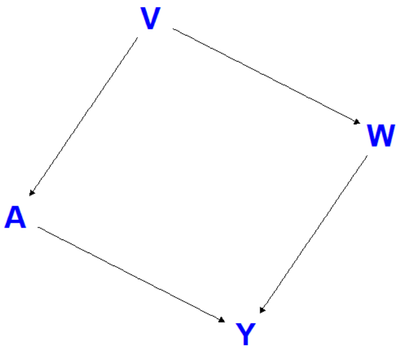
This is the backdoor path criterion.

Example 1



$A \leftarrow V \rightarrow W \rightarrow Y$ is backdoor path from A to Y

Example 1



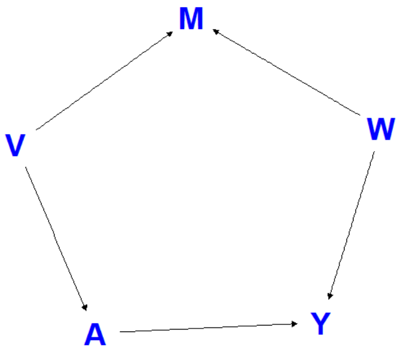
$A \leftarrow V \rightarrow W \rightarrow Y$ is backdoor path from A to Y

Not blocked by a collider.

Sets of variables:

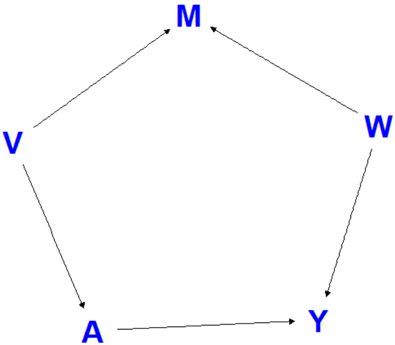
- {V};
- {W};
- {V,W}

Example 2



$A \leftarrow V \rightarrow M \rightarrow W \rightarrow Y$ is backdoor path from A to Y

Example 2

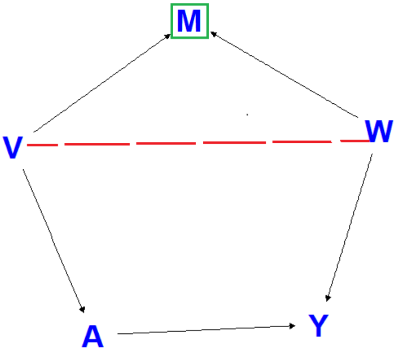


$A \leftarrow V \rightarrow M \rightarrow W \rightarrow Y$ is backdoor path from A to Y

Backdoor Path is blocked by a collider M.

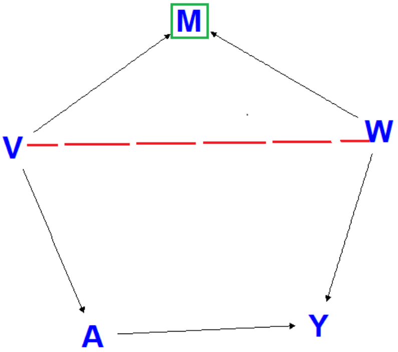
No need to control any variable.

Example 2 Continued



$A \leftarrow V \rightarrow M \rightarrow W \rightarrow Y$ is backdoor path from A to Y

Example 2 Continued

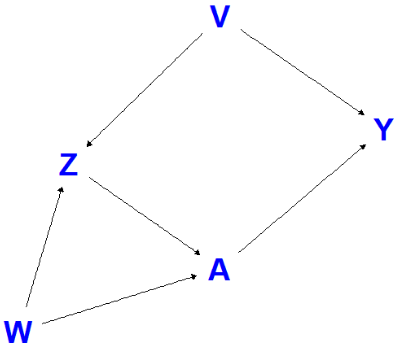


$A \leftarrow V \rightarrow M \rightarrow W \rightarrow Y$ is backdoor path from A to Y

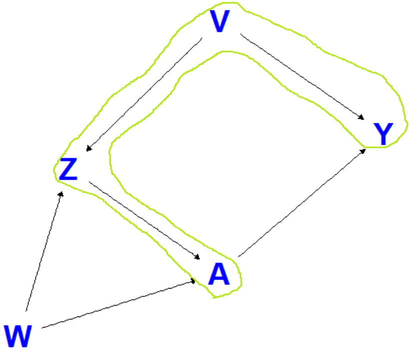
Sets of variables:

- {}, {V}, {W};
- {V,W}, {M,V},{M,W};
- {M,V,W}

Example 3



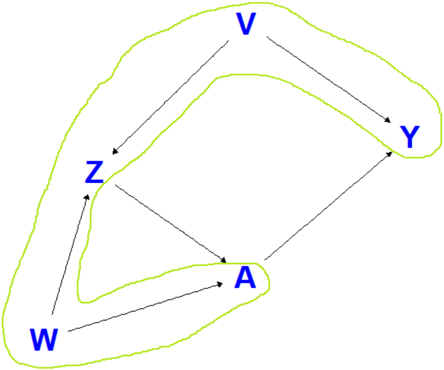
Example 3



First Backdoor Path:

$A \leftarrow Z \leftarrow V \rightarrow Y$

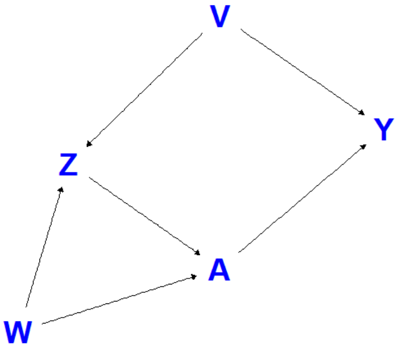
Example 3



Second Backdoor Path:

$A \leftarrow W \rightarrow Z \leftarrow V \rightarrow Y$

Example 3 Continued

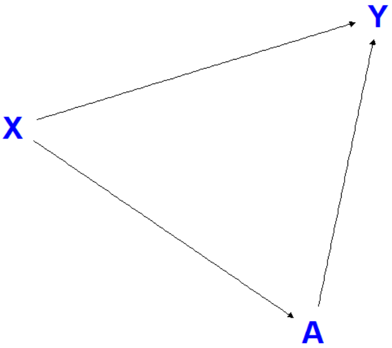


Sets of variables:

- {V};
- {V,Z},{V,W},{Z,W};
- {V,Z,W}.

Not included {Z} or {W}

How to Control ?



If X is observed:

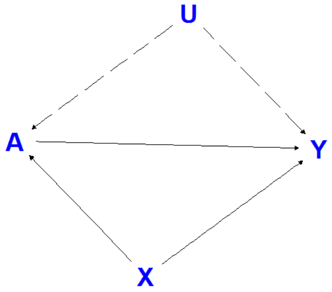
- Matching;
- Propensity Score Matching;
- Inverse Probability Treatment Weighting.

Unmeasured Confounding

- Suppose U is unobserved variable that affects A and Y, then we have unmeasured confounding.
- Ignorability assumption is violated.
- Cannot control for confounders if we do not observe them all by using previous methods.

Unmeasured Confounding

- Suppose U is unobserved variable that affects A and Y, then we have unmeasured confounding.
- Ignorability assumption is violated.
- Cannot control for confounders if we do not observe them all by using previous methods.
- Instrumental Variables (IV) is an alternative causal inference method that does not rely on ignorability assumption.



Literature

- Textbooks : Judea 2010; Spirtes, C. N. Glymour, et al. 2000; Pearl et al. 2016
- Multiple treatment effects : Lopez, Gutman, et al. 2017
- Incomplete data : Gelman and Meng 2004
- Critics on subject received : Humphreys and Freedman 1996; Spirtes, C. Glymour, et al. 1997
- Comparison of observational and experimental : Rosenbaum 2017
- Causal Graphs : Scheines 1997; Greenland and Pearl 2014

Bibliography I

- Gelman, Andrew and Xiao-Li Meng (2004). *Applied Bayesian modeling and causal inference from incomplete-data perspectives*. John Wiley & Sons.
- Greenland, Sander and Judea Pearl (2014). "Causal diagrams". In: *Wiley StatsRef: Statistics Reference Online*.
- Humphreys, Paul and David Freedman (1996). *The grand leap*.
- Judea, Pearl (2010). "An introduction to causal inference". In: *The International Journal of Biostatistics* 6.2, pp. 1–62.
- Lopez, Michael J, Roee Gutman, et al. (2017). "Estimation of causal effects with multiple treatments: a review and new ideas". In: *Statistical Science* 32.3, pp. 432–454.
- Pearl, Judea, Madelyn Glymour, and Nicholas P Jewell (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.
- Rosenbaum, Paul R (2017). *Observation and experiment*. Harvard University Press.
- Scheines, Richard (1997). "An introduction to causal inference". In:

Bibliography II

Spirtes, Peter, Clark Glymour, and Richard Scheines (1997). "Reply to humphreys and freedman's review of causation, prediction, and search". In: *The British journal for the philosophy of science* 48.4, pp. 555–568.

Spirtes, Peter, Clark N Glymour, et al. (2000). *Causation, prediction, and search*. MIT press.


THANK
YOU