Causal Inference using Causal Graphs in Epidemiologic Context

Ehsan Karim
ehsan@stat.ubc.ca

Department of Statistics

December 11, 2010

Motivation >

- Statistical Model
- Causation and Association
- Epidemiologic context $\Rightarrow$ Confounding

Objective >

- Gentle introduction to causal graphs
- Identifying confounding

Outline

- Background
- Causal Diagrams (DAGs)
- Identifying Confounding using DAGs
- DAGs in Longitudinal Settings
- Miscellaneous Issues

Background > Causation >

Ways to define causation

1. Philosophical Definition
2. Counterfactual Definition

Background > Causation > Philosophical

Philosopher Hume presented two useful definitions:

1. Cause is an event followed by another (effect).
   - both $\mathcal{E}$ and $\mathcal{D}$ occurs
   - the cause $\mathcal{E}$ occurs before effect $\mathcal{D}$
   - two events should be closely connected (through time or space)
2. Without the first event (cause), the second (effect) would never happen. i.e,
   - if $\mathcal{E}$ had not occurred, then $\mathcal{D}$ would not have occurred as well

Background > Causation > Philosophical

Philosopher Hume presented two useful definitions:

1. Cause is an event followed by another (effect).
   - both $\mathcal{E}$ and $\mathcal{D}$ occurs
   - the cause $\mathcal{E}$ occurs before effect $\mathcal{D}$
   - two events should be closely connected (through time or space)
2. Without the first event (cause), the second (effect) would never happen. i.e,
   - if $\mathcal{E}$ had not occurred, then $\mathcal{D}$ would not have occurred as well

Background > Causation > Counterfactual

- General Idea
  1. with $\mathcal{E}$, record $\mathcal{D}$ status
  2. do the same for $\bar{\mathcal{E}}$
  3. repeat
  4. find the average difference
- Relation with Philosophical Definition (Hume's 2nd)
- Probabilistic causality (smoking causes cancer)

Background > Causation > Counterfactual > Criticism

- Unobservable objects
- Formulation appropriate for randomized trials only, observational studies will require ignorability

Background > Confounding >

Three ways to define:

1. Conventional
2. Counterfactual

Background > Confounding > Conventional

Outcome Variable Confounding variable must be related with the outcome variable.

Exposure Variable Confounding variable is associated with the exposure.

External Variable Confounding variable must not be affected by exposure or outcome.

Background > Confounding > Conventional > Illustration

- Say, *male subjects in the sample are more likely to drink coffee than the female subjects* $\Rightarrow$ exposure to coffee, $\mathcal{E}$ depends on gender, $\mathcal{C}$
- Say, *gender* $\mathcal{C}$ *is associated with pancreatic cancer,* $\mathcal{D}$
- But gender $\mathcal{C}$ is neither determined by coffee drinking habit $\mathcal{E}$, nor the pancreatic cancer $\mathcal{E} \Rightarrow \mathcal{C}$ is a confounder

Background > Confounding > Counterfactual

Table: Counterfactual Items

Population	Ω		$\bar{\Omega}$	
Treatment	x_1	x_0	x_1	x_0
Outcome	y_1	y_0	y_1	y_0
Marginal Distribution	$F_\Omega(y_1)$	$F_\Omega(y_0)$	$F_{\bar{\Omega}}(y_1)$	$F_{\bar{\Omega}}(y_0)$
Summary parameter	$\mu_{\Omega 1}$	$\mu_{\Omega 0}$	$\mu_{\bar{\Omega} 1}$	$\mu_{\bar{\Omega} 0}$

$\mu_{\Omega 1} - \mu_{\Omega 0}$ measures the effect of treatment x_1 in the population Ω.

Background > Confounding > Counterfactual

Table: Observed Items

Population	Ω		$\bar{\Omega}$	
Treatment	x_1	$\times$	$\times$	x_0
Outcome	y_1	$\times$	$\times$	y_0
Marginal Distribution	$F_{\Omega}(y_1)$	$\times$	$\times$	$F_{\bar{\Omega}}(y_0)$
Summary parameter	$\mu_{\Omega 1}$	$\times$	$\times$	$\mu_{\bar{\Omega} 0}$

$\mu_{\Omega 1} - \mu_{\bar{\Omega} 0}$ measures the association of treatments with outcomes across the populations. If $\mu_{\Omega 0} \neq \mu_{\bar{\Omega} 0}$, then the confounding exists.

Background > Control >

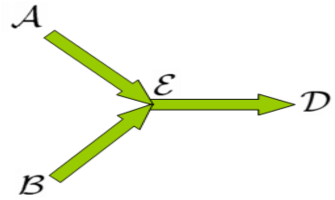
1. Design Stage
   - Randomization
   - Exchangeability
   - Matching
2. Analysis Stage
   - Stratification
   - Regression Modelling
   - Propensity Score

Background > Control >

1. Design Stage
   - Randomization
   - Exchangeability
   - Matching
2. Analysis Stage
   - Stratification
   - Regression Modelling
   - Propensity Score

Causal Graphs > Basic Terminologies

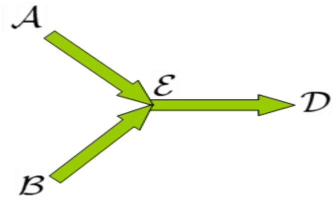
- Node, Edge, Path
- Direct Cause, Indirect Cause, Directed acyclic graph
- Child / descendant, Parent / ancestor
- Exogenous, endogenous, terminal, essential
- Directed path, Directed Acyclic Graph
- Collider
- Open path, closed path or block



Causal Graphs > d-separation Rules

d-separation and d-connection rule

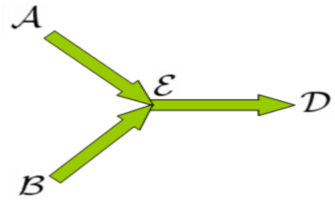
- Unconditional $\mathcal{A}$ and $\mathcal{B}$ are d-separated $\Leftrightarrow$ no open path
- Conditional
  1. conditioning on a collider $\mathcal{E}$ in $\mathcal{A} \to \mathcal{E} \leftarrow \mathcal{B} \Rightarrow (\mathcal{A}$ and $\mathcal{B})$ d-connected
  2. conditioning on a non-collider $\mathcal{E}$ in $\mathcal{A} \to \mathcal{E} \to \mathcal{D} \Rightarrow (\mathcal{A}$ and $\mathcal{D})$ d-seperated



Causal Graphs > d-separation Rules

d-separation and d-connection rule

- Unconditional $\mathcal{A}$ and $\mathcal{B}$ are d-separated $\Leftrightarrow$ no open path
- Conditional
  1. conditioning on a collider $\mathcal{E}$ in $\mathcal{A} \rightarrow \mathcal{E} \leftarrow \mathcal{B} \Rightarrow (\mathcal{A}$ and $\mathcal{B})$ d-connected
  2. conditioning on a non-collider $\mathcal{E}$ in $\mathcal{A} \rightarrow \mathcal{E} \rightarrow \mathcal{D} \Rightarrow (\mathcal{A}$ and $\mathcal{D})$ d-seperated



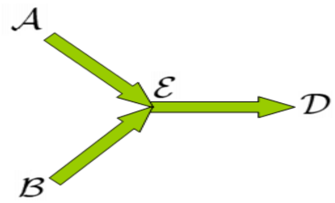
Causal Graphs > d-seperation & Independence

Compatibility Rule

($\mathcal{A}$ and $\mathcal{B}$) d-separated $\Leftrightarrow$ ($\mathcal{A}$ and $\mathcal{B}$) statistically independent

Weak Faithfulness Rule

no d-connecting path between ($\mathcal{A}$ and $\mathcal{B}$) conditional on $\mathcal{E}$ $\Leftrightarrow$ independence of $(\mathcal{A}, \mathcal{B} | \mathcal{E})$



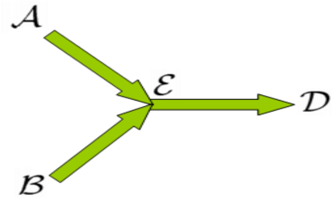
Causal Graphs > d-seperation & Independence

Compatibility Rule

($\mathcal{A}$ and $\mathcal{B}$) d-separated $\Leftrightarrow$ ($\mathcal{A}$ and $\mathcal{B}$) statistically independent

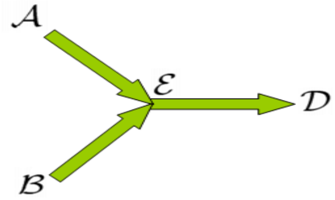
Weak Faithfulness Rule

no d-connecting path between ($\mathcal{A}$ and $\mathcal{B}$) conditional on $\mathcal{E}$ $\Leftrightarrow$ independence of $(\mathcal{A}, \mathcal{B}|\mathcal{E})$



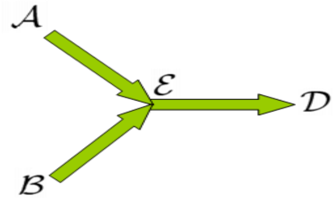
Causal Graphs > Causation & Association

- $\mathcal{E}$ is a direct cause of $\mathcal{D}$, and $\mathcal{A}$ and $\mathcal{B}$ are indirect causes.
- $\mathcal{A}$ and $\mathcal{B}$ both directly causes $\mathcal{E}$
- $\mathcal{A}$ and $\mathcal{B}$ have no causal relationship between themselves.



Causal Graphs > Causation & Association

- $(\mathcal{E}, \mathcal{D})$, $(\mathcal{A}, \mathcal{D})$, $(\mathcal{B}, \mathcal{D})$, $(\mathcal{A}, \mathcal{E})$, $(\mathcal{B}, \mathcal{E})$ are statistically dependent.
- $(\mathcal{A}, \mathcal{B})$ are statistically independent.
- $(\mathcal{A}, \mathcal{D})$ are statistically independent, conditional on $\mathcal{E}$.
- $(\mathcal{A}, \mathcal{B})$ are statistically dependent, conditional on $\mathcal{E}$.

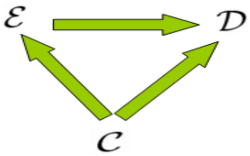


Confounding Identification >

Identifying the confounding conditions is the most lucrative feature of graphical approach ⇒ 'back-door criterion'.

Confounding Identification > Back–door Criterion

Back–door criterion basically says that if there exists an undirected path between ($\mathcal{E}$, $\mathcal{D}$), then confounding exists.

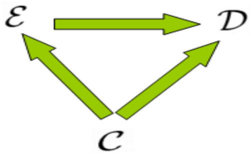


Confounding Identification > Detection steps

We determine the existence of confounding in the following steps:

1. first delete all the arrows originating from $\mathcal{E}$
2. try to find any unblocked backdoor path from $\mathcal{E}$ to $\mathcal{D}$ in the second stage.

If no such path exists, there exists no confounding.

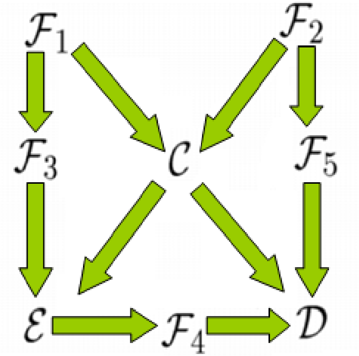


Confounding Identification > Caution

Back-door criterion does not define a confounder, rather describes the confounding situation.

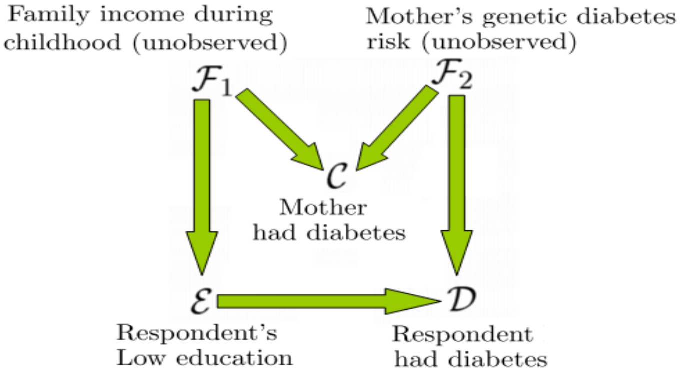
Confounding Identification > Illustration

Figure: Application of Back–door Criterion



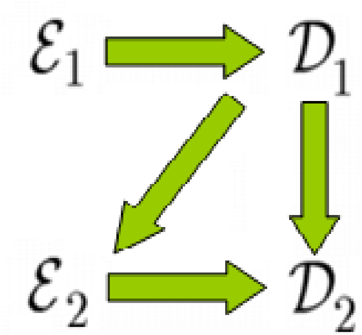
Confounding Identification > Compare

Figure: Conventional vs. Graphical Method



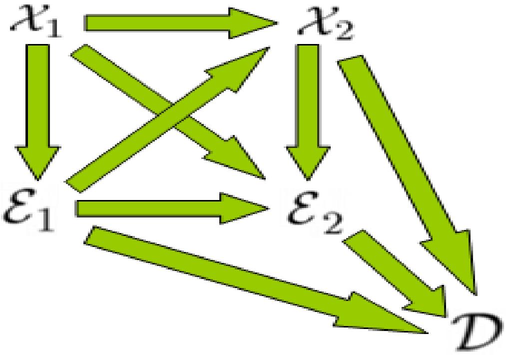
Longitudinal > Longitudinal Settings > Example I

Figure: Time dependent treatment and outcome for $t = 1, 2$



Longitudinal > Longitudinal Settings > Example II

Figure: Time dependent treatment, covariate for $t = 1, 2$ & outcome



Miscellaneous Issues >

- Residual Confounding
- Over-adjustment
- Sufficient set of Confounders

Thank You!