Casual Inference

The Science of WHY

Qiunuo Chen

Fuzhou University

March, 2025

1. Graph
2. Structed Causal Model(SCM)
3. Intransitive case
4. V-structure
5. D-separation
6. References

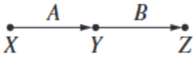
1. Graph
2. Structed Causal Model(SCM)
3. Intransitive case
4. V-structure
5. D-separation
6. References

Graph

- Simple graphical diagram can be useful in describing casual relatioship between variables.

Definition

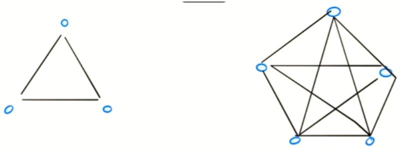
A graph is a collection of vertices or nodes and edges.The nodes are connected by the edges.



- Two nodes are adjacent if there is an edge between them.

Definition

A graph is a collection of vertices or nodes and edges. The nodes are connected by the edges.



- The graph is a complete graph if there is an edge between every pair of nodes in the graph.

1 Graph

2 Structed Causal Model(SCM)

3 Intransitive case

4 V-structure

5 D-separation

6 References

Defintion of Causation

用简化的模型研究自然界中的因果关系;(fire...)
利用抽象的数学语言刻画两个变量的关系。
A variable X is a direct cause(Treatment) of a Variable Y if X appears in the functions that determines Y's value.

$$Y = f(X)$$

Sometimes X is not the only cause.

$$Y = f(X, Z, \ldots)$$

Defintion of Causation

X might be an indirect cause .

$$Z = g(X)$$

$$Y = f(Z)$$

$$Y = f(g(X))$$

X is a cause of Y if it is a direct cause of Y, or of any cause of Y.

Association between Causation and Graphs

If Y is the child of X, then X is the direct cause of Y.
If Y is the descendant of X, then X is a potential cause of Y.

1. Y is caused by X

$$Y = f(X)$$

$$X - - - - - - -- > Y$$

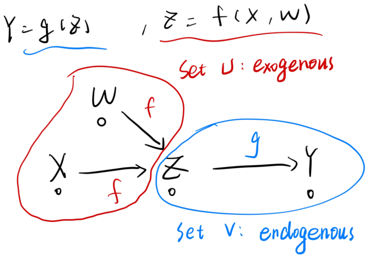
(Y is the child of X)

2. Y is caused by X and Z

$$Y = f(X, Z)$$

$$X/Z - - - - - - -- > Y$$

Association between Causation and Graphs



- A SCM consists of two sets of variables U and V, and a set of functions f
- The variables in U are called exogenous variables (external to the model)
- The variables in V are called endogenous variables. (internal)
- Every endogenous variable is a descendant of at least one exogenous variable

Association between Causation and Graphs

- Root nodes $\Leftrightarrow$ Exogenous variables U
- Descendant nodes of root nodes $\Leftrightarrow$ Endogenous variables V
- Edges $\Leftrightarrow$ The functions f.

Example:Association between Causation and Graphs

$$V = \{Height, Gender, Performance\}$$

$$U = \{u_1, u_2, u_3\} \quad f = \{f_1, f_2, f_3\}$$

$$Gender = f_1(u_1)$$

$$Height = f_2(U_2, Gender)$$

$$Performance = f_3(Height, Gender, U_3)$$

Example:Association between Causation and Graphs

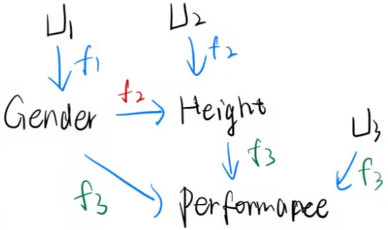
$V = \{Height, Gender, Performance\}$

$U = \{u_1, u_2, u_3\}$ $f = \{f_1, f_2, f_3\}$

$Gender = f_1(u_1)$

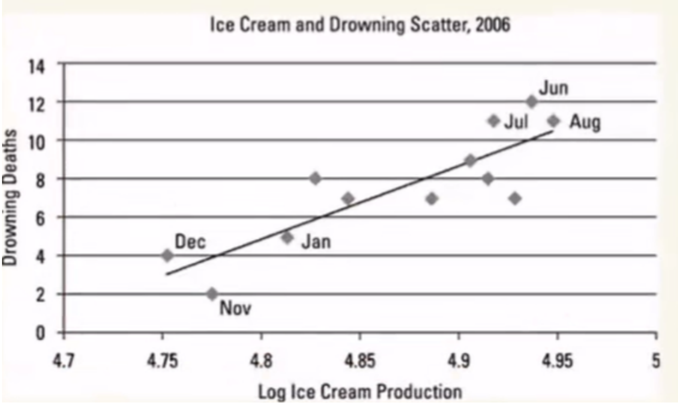
$Height = f_2(U_2, Gender)$

$Performance = f_3(Height, Gender, U_3)$



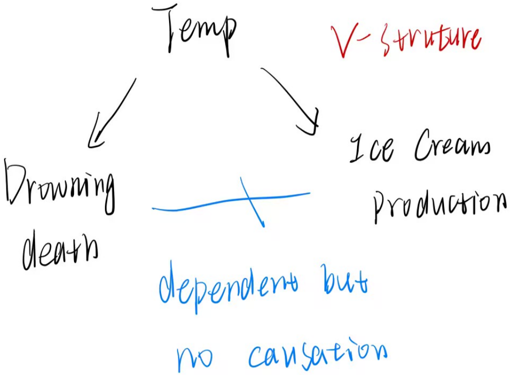
Example:Association between Causation and Graphs

- Statistical dependence doesn't necessarily imply causation.



Example:Association between Causation and Graphs

- Statistical dependence doesn't necessarily imply causation.



1 Graph

2 Structed Causal Model(SCM)

3 Intransitive case

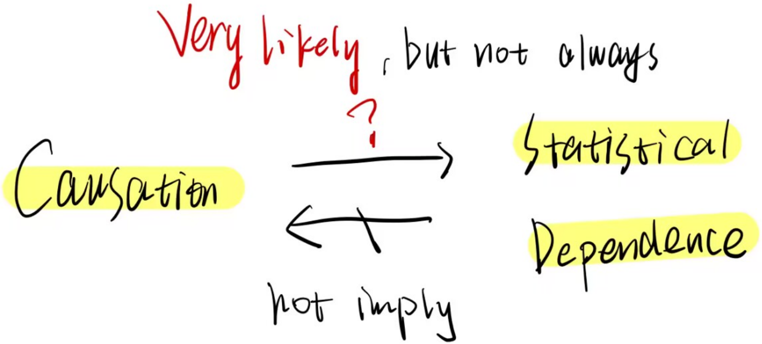
4 V-structure

5 D-separation

6 References

- If X and Y are statistically dependent, X does not necessarily cause Y (or Y causes X).
- But, on the other hand, if X causes Y, are X and Y statistically dependent?
- The answer is: very likely X and Y are dependent, but not always.
  for example,Gene1 causes disease. However, if gene2 exists, it will antagonize the effect of Gene1. 看起来不相关，但是有因果性。

Conclusion


Very likely, but not always
?
Causation
Statistical
Dependence
not imply

1 Graph

2 Structed Causal Model(SCM)

3 Intransitive case

4 V-structure

5 D-separation

6 References

Intro

chain ,fork ,collider 是构造更复杂图的基本组件。

Chain

SCM

$V = \{X, Y, Z\}$

working hours training performance

$U = \{U_X, U_Y, U_Z\}$

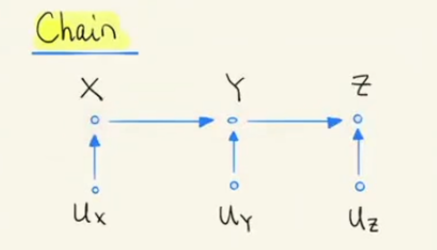
$F = \{f_X, f_Y, f_Z\}$

$f_X : X = U_X$

$f_Y : Y = 84 - X + U_Y$

$f_Z : Z = 100Y + U_Z$

Chain



1. Z and Y are **likely** dependent
2. Y and X are **likely** dependent
3. Z and X are **likely** dependent
4. Z and X are independent, **conditional on** Y
   i.e.

$$P(Z = z \mid X = x, \ \mathbf{Y} = \mathbf{c}) = P(Z = z \mid \mathbf{Y} = \mathbf{c})$$

Chain

Z and X are independent, **conditional on** Y
i.e.

$$P(Z = z \mid X = x,\ \mathbf{Y} = \mathbf{c}) = P(Z = z \mid \mathbf{Y} = \mathbf{c})$$

$$f_X : X = U_X$$

$$f_Y : Y = 84 - X + U_Y = C$$

$$f_Z : Z = 100Y + U_Z$$

Chain-Rule1

Rule 1 (Conditional independence in Chains)

If there is only one chain between X and Z, and Y is any set of variables that intercept that chain, then(相当于y从中间切断了)

$$X \perp\!\!\!\perp Z \mid Y$$

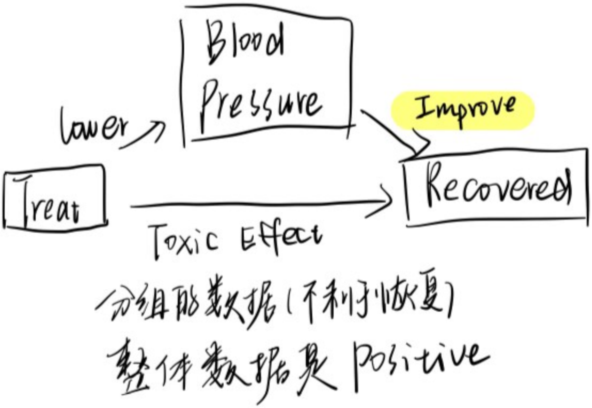
$$\begin{tikzcd}
X \arrow[r] \arrow[dr] \arrow[ddr] & Y_1 \arrow[r] & Z \\
& Y_2 \arrow[ur] & \\
& Y_3 \arrow[uur] &
\end{tikzcd}$$

$$X \equiv Z; \quad '\equiv' \ means \quad dependent$$

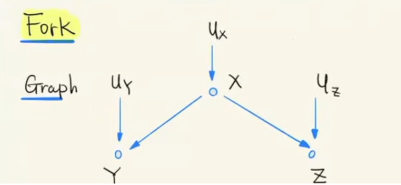
$$X \perp\!\!\!\perp Z \mid \{Y_1, Y_2, Y_3, \ldots\}$$

Explanation - Causal Diagram

我们可以用因果图来解释辛普森悖论现象。(看整体数据)



Fork

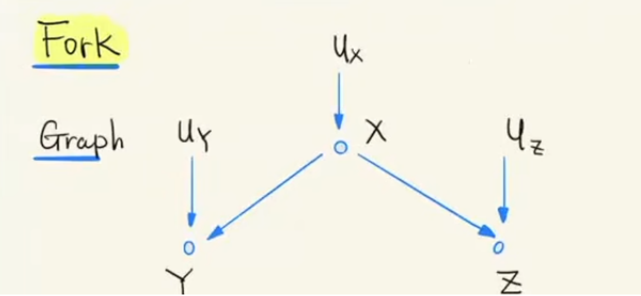


$$U = \{U_X, U_Y, U_Z\}$$
$$F = \{f_X, f_Y, f_Z\}$$
$$V = \{X, Y, Z\}$$

crowd on beach　ice cream sales　drowning

$$f_X : X = U_X = C$$
$$f_Y : Y = 4X + U_Y$$
$$f_Z : Z = \frac{X}{10} + U_Z$$

Fork

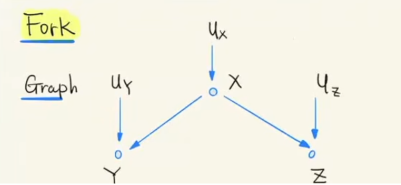


1. X and Y are **likely** dependent
2. X and Z are **likely** dependent
3. Y and Z are **likely** dependent
4. Y and Z are independent conditional on X

i.e.

$$P(Z = z \mid Y = y, X = c) = P(Z = z \mid X = c)$$

Fork



$$U = \{U_X, U_Y, U_Z\}$$
$$F = \{f_X, f_Y, f_Z\}$$
$$V = \{X, Y, Z\}$$

crowd on beach ice cream sales drowning

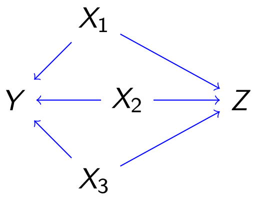
$$f_X : X = U_X = C$$
$$f_Y : Y = 4X + U_Y$$
$$f_Z : Z = \frac{X}{10} + U_Z$$

Fork-Rule2

Rule 2 (Conditional independence in Forks)

If X is a common cause of Y and Z, and there is only one path between Y and Z, then

$$Y \perp\!\!\!\perp Z \mid X$$



$$Y \perp\!\!\!\perp Z \mid X_1 \quad ? \quad \textit{No}$$

$$Y \perp\!\!\!\perp Z \mid X_1, X_2, X_3 \quad ? \quad \textit{YES}$$

$$Y \longleftarrow X_1 \longrightarrow X_2 \longrightarrow X_3 \longrightarrow Z$$

$$Y \equiv Z \quad ? \quad \textit{Yes}$$

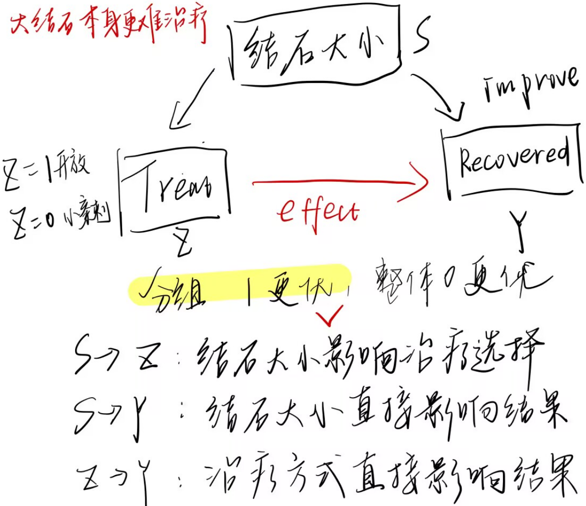
$$Y \perp\!\!\!\perp Z \mid X_2 \quad ? \quad \textit{Yes}$$

$$Y \perp\!\!\!\perp Z \mid X_1 \quad ? \quad \textit{Yes}$$

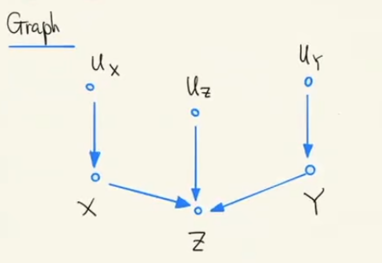
$$Y \perp\!\!\!\perp Z \mid X_3 \quad ? \quad \textit{Yes}$$

Explanation - Causal Diagram

我们可以用因果图来解释辛普森悖论现象。（看分组数据）

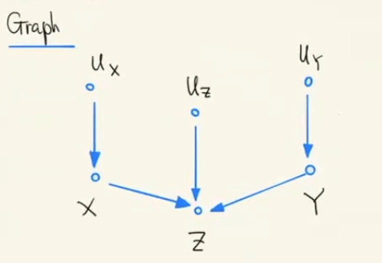


Collider



$$U = \{U_X, U_Y, U_Z\}$$
$$F = \{f_X, f_Y, f_Z\}$$
$$V = \{X, Y, Z\}$$
$$f_X : X = U_X$$
$$f_Y : Y = U_Y$$
$$f_Z : Z = X + Y + U_Z$$

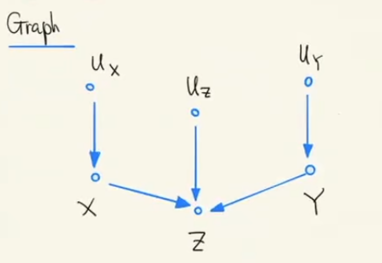
Collider



1. Z and Z are **likely** dependent
2. Y and Z are **likely** dependent
3. X and Y are **likely** independent
4. X and Y are dependent, **conditional on** Z
   i.e.

$$P(X = x \mid Y = y,\ \mathbf{Z} = \mathbf{c}) \neq P(X = x \mid \mathbf{Z} = \mathbf{c})$$

Collider

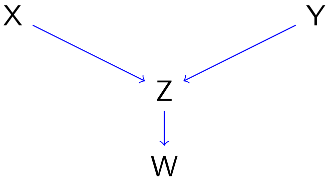


$$U = \{U_X, U_Y, U_Z\}$$
$$F = \{f_X, f_Y, f_Z\}$$
$$V = \{X, Y, Z\}$$
$$f_X : X = U_X$$
$$f_Y : Y = U_Y$$
$$f_Z : Z = X + Y + U_Z = C$$

Collider-Rule3

Rule 3 (Conditional independence in Colliders)

If Z is a collision of X and Y and there is only one path between X and Y, then X and Y are unconditionally independent but are dependent conditional on Z, or any descendent of Z.



$$X \perp\!\!\!\perp Y$$

$$X \perp\!\!\!\perp Y \mid Z \quad ? \quad (Rule3)$$

$$X \perp\!\!\!\perp Y \mid W \quad ? \quad (Generalization)$$

1. Graph
2. Structed Causal Model(SCM)
3. Intransitive case
4. V-structure
5. D-separation
6. References

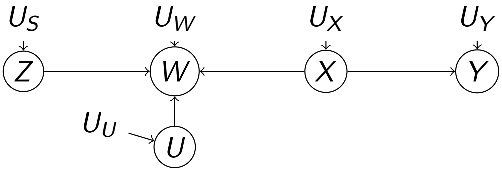
Basic Concept

总结之前的三个模型：

V - structure	Uncondition	Condition
Fork	unblock	block
Chain	unblock	block
Collider	block	unblock
(or descendants)		

- $$X \text{ and } Y \text{ d - separated} \Leftrightarrow X, Y \text{ independent.}$$
- $$X \text{ and } Y \text{ d - sep condition } Z \Leftrightarrow X, Y \text{ indep. } |Z$$
- X and Y are d - separated:
  If every path between X and Y is blocked, or d - separated.
- X and Y are d - connected:
  If there exists an unblocked path.

Example



1. $Z \perp Y$ Yes
2. $Z \equiv Y \mid U$ Yes
3. $Z \perp Y \mid \{W, X\}$ Yes

1 Graph

2 Structed Causal Model(SCM)

3 Intransitive case

4 V-structure

5 D-separation

6 References

[1] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*.
Basic Books, May 2018.

[2] J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer*.
John Wiley & Sons, reissue ed., 2016.

Thank You