

Causal Inference: The Mixtape Chapter 3

Directed Acyclic Graph

佐々木毬乃 石井颯太

早稲田大学

2022 年 4 月 18 日

本日のポイント

- ① 因果推論の目的と論点について理解する。
- ② 既知の概念を DAG に結びつけて理解する。
- ③ バックドアとその閉じ方について理解する。
- ④ DAG を用いて適切な回帰モデルのデザインを立てられるようになる。
- ⑤ データ自体に存在する可能性のある、合流点バイアスについて理解する。

- 1 Causal Inference
- 2 Direct Acyclic Graph (DAG)
- 3 Causal Inference with DAG
- 4 Understanding Biased Data with DAG
- 5 例題

Causal Inference

Why Do We Need to Learn Causal Inference?

「因果」関係を「類推」することは人間だれしも日常的に行っている。

ではなぜわざわざ経済学の訓練として学ぶのか？

それは

- ① 我々が日頃感じる因果関係は、一時的な相関関係であることが多い。
- ② 自分以外の人間が関わる時、「肌感覚」では正確なことがわからない。

から。

The Research-wise Need for Causal Inference

人間は色々な理由で、世の中の全体像など把握できないなりにどうにか自分の狙った効果の実現を目指す。

⇒ 結果として、よくわからない相関関係が見つかり、因果関係ではないかと言われる。

例) 電気をつけたまま寝かせると子供が近視になる。

⇒ 近視の親は寝る時に電気をつけたまま寝ることが多く、子供は遺伝的に近視になりやすかった。¹

自分の狙った効果を実現するためには、どの相関関係が因果関係で、どれがそうでないのか見分けなければならない。

¹伊藤公一郎 (2017) 『データ分析の力 因果関係に迫る思考法』 光文社新書

因果推論とは？

⇒ データから因果関係を導き出すこと

因果関係がわかると何が嬉しいのか？

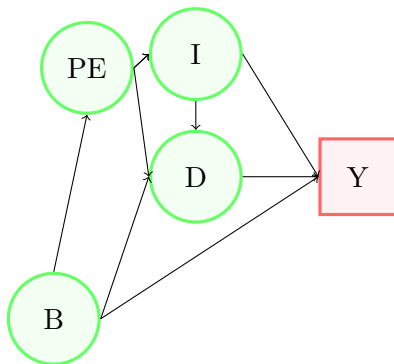
⇒ 自分の狙った効果を実現できる。

因果関係を立証するのが難しい理由（伊藤, 2017）¹

- ① 他の要因が関係していた可能性がある
- ② 逆の因果関係だった可能性がある

Direct Acyclic Graph (DAG)

DAG: The Appearance



PE: 親による子供の教育、I: 親の収入、D: 子供の大学教育、B: 子供の大学教育に影響を与える様々な要素、Y: 子供の収入

非循環有向グラフ：

想定される * 因果関係を、現象 ** 間に因果を表す矢印 *** を引くことで示したグラフ。

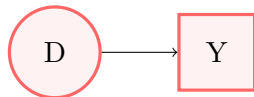
* 「想定される」：理論上想定されるということ。これは先行研究、自分の思考から導く。

** 「現象」：観察された現象。具体的には、分析者によって把握されているデータ生成過程によって生み出された、確立変数。

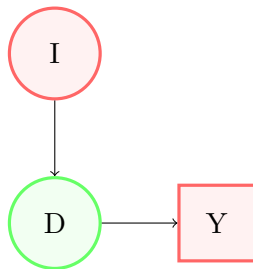
*** 「因果を表す矢印」：時間的に先行する事象から、続く事象に向かって出ている。

⇒DAG は反事実と比較しての効果を考えるためのものであり、同時性や逆向きの因果関係について考えるためには用いない。

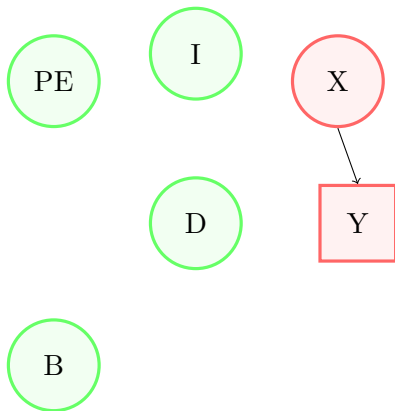
DAG representation: Direct Effect



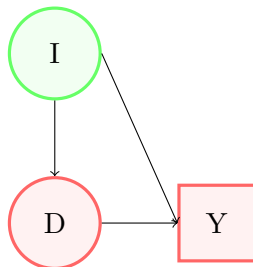
DAG representation: Mediation



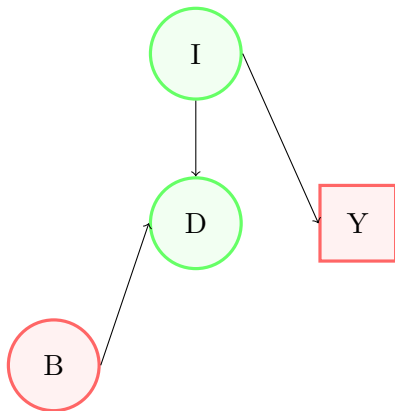
DAG representation: Error term



DAG representation: Confound



DAG representation: Collide



バックドアとは、直感的に関心のある変数のノードをつなぐ、直接でない矢印の連鎖のこと。

あるリサーチデザインが求めたい因果効果を独立して計測できるためには、バックドアが全て閉じていることの必要十分条件である以下の基準を満たしている必要がある。

「変数の集合 X が DAG においてバックドア基準を満たすことの必要十分条件は、 X が D から Y への矢印を含むすべての交絡変数間の経路を塞いであること」

経路の塞ぎ方

- ① 統制する（詳細は今後）
回帰分析において統制するほか、下位分類法、マッチング、回帰不連続デザインなどの手法が存在する（詳細は今後の発表に譲る）。
- ② 経路中に合流点が存在する
 $I \Rightarrow D \Leftarrow B \Rightarrow Y$

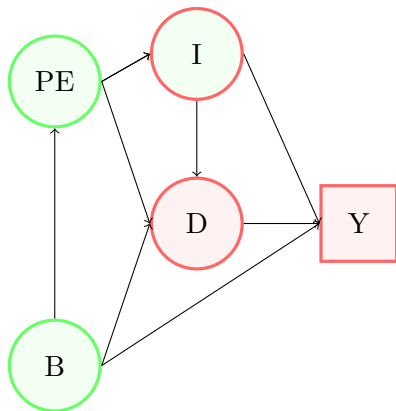
* 合流点を統制すると経路が開く

$$I \Rightarrow D \Leftarrow B \Rightarrow Y$$

合流点の後には必ず交絡変数がある。合流点について統制した時、交絡変数について統制しないと第4節のような問題が起きる。

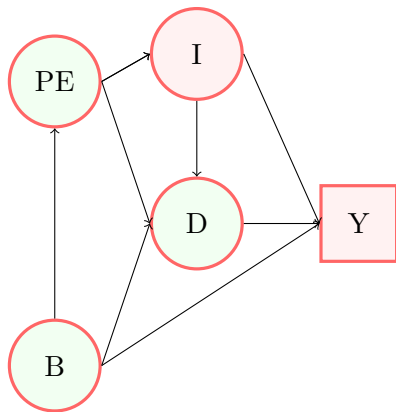
Causal Inference with DAG

When you want to know the effect of D on Y



$$Y_i = \alpha + \delta D_i + \beta I_i + \epsilon_i$$

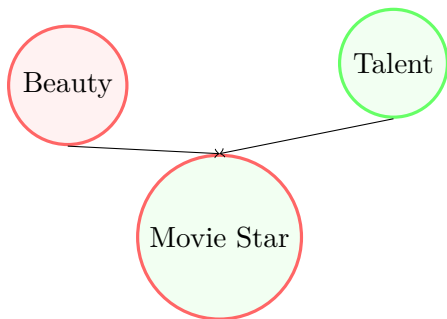
When you want to know the effect of I on Y



$$Y_i = \alpha + \delta I_i + \beta PE_i + \zeta B_i + \epsilon_i$$

Understanding Biased Data with DAG

Collider Bias 1



Collider Bias 1

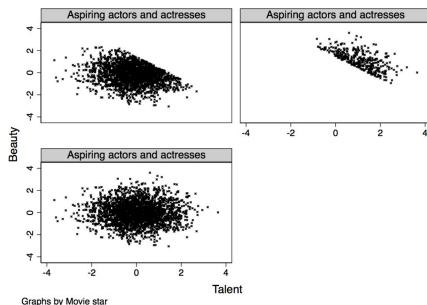
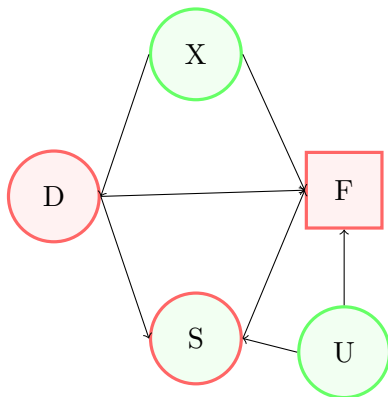


Figure 1: 合流点 Movie Star を統制することによって生じるバイアス



D: マイノリティ、X: さまざまな統制変数、S: 職務質問を受ける、U: 怪しさ、F: 実力行使

ここで、記録されているデータは職務質問を受けたサンプルに限られる。すなわちデータ自体がSについて統制されている。

⇒D から S への効果がわからないため、D の Y に対する効果も不明。

例題

例題 1

大学での成績の決定要因について考えてみましょう。

- ① 関連する変数を挙げ、
 - ② 現状提示されている DAG にその変数を加えてください
- (答えはオープンです)

例題 2²

重症度 =I, ドクターヘリによる搬送 =H, 搬送時間 =T, 合併症 =C, 年齢 =A, 30 日後の生存 =S とする。

ここで、以下が成り立つとする。

- $C \leftarrow I \Rightarrow H$
- $C \leftarrow A \Rightarrow S$
- $H \Rightarrow T \Rightarrow C \Rightarrow S$

上記全ての変数のデータが揃っているとし、以下の問いについて答えよ。

Q1) 変数 I, H, T, C, A, S 間の関係を DAG で表せ。

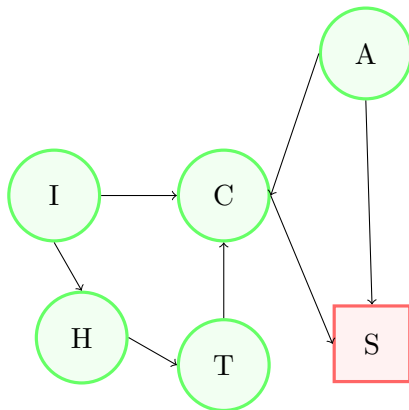
Q2) I が S にもたらす効果を考える時、統制が必要な変数が存在すれば全て挙げよ。存在しない場合は「なし」と答えよ。

Q3) C が S にもたらす効果を考える時、統制が必要な変数が存在すれば全て挙げよ。存在しない場合は「なし」と答えよ。

²小松裕和、鈴木越治、土居弘幸（2009）「研究仮説の立て方と Directed Acyclic Graph (疫学各論 2)」『日救急医学会誌』 vol. 20, pp. 397-403. を参考に作成

解答 2

(1)



(2) なし

(3) A

END OF FILE