統計的因果探索の出発点
（第1章）
因果探索における最大の困難

擬似相関
同じ相関関係でも異なる因果関係

チョコレートをたくさん食べている国ほど受賞者が多い（相関関係）

相関係数 0.79
p値 < 0.0001

複数の因果関係が
同じ相関関係を与える

チョコレートをたくさん食べさせれば受賞者が増えるのか？
（因果関係）
用語

未観測共通原因

定性的な因果関係:
因果グラフ

実線
観測されるデータあり

点線
未観測データなし

結果 ← 原因

共通原因
記号

$Z$ (GDP)

$X$ (輸入)

$Y$ (販売)

エラー変数

$E_X$、$E_Y$

$x$の値は、$Z, E_X$の値から求まる

$x = f_x (Z, E_X)$

$y$の値は、

$x, Z, E_Y$の値から求まる

$y = f_y (x, Z, E_Y)$
デコーダ生成過程

\[ x = f_x(z, e_x) \]
\[ y = f_y(x, z, e_y) \]

さらに線形性を仮定すると、

\[ x = \frac{\lambda x z}{Z} + e_x \]
\[ y = \frac{\lambda y z}{Z} x + e_y \]

定数
データ生成過程。

\[ x = 0.3z + e_x \]
\[ y = 0.7x + 0.3z + e_y \]

\[
\begin{bmatrix}
  z \\
  e_x \\
  e_y
\end{bmatrix}
\sim
N
\left(\begin{bmatrix}
  0 \\
  0 \\
  0
\end{bmatrix},
\begin{bmatrix}
  1 & 0.91 & 0 \\
  0.91 & 0 & 0 \\
  0 & 0 & 0.29
\end{bmatrix}\right)
\]

データ生成における、\(z, e_x, e_y\)の値を上のガウス白帯から生成。

\[ z^{(i)} = 0.47 \]
\[ e_x^{(i)} = 0.74 \]
\[ e_y^{(i)} = -0.19 \]
これら \( z^{(i)} = 0.47 \) を用い、\( x^{(i)} \) と \( y^{(i)} \) を生成する。

\[
e_{x}^{(i)} = 0.74
\]
\[
e_{y}^{(i)} = -0.19
\]

7 まり、

\[
x^{(i)} = 0.3 z^{(i)} + e_{x}^{(i)}
\]
\[
= 0.3 \times 0.47 + 0.74
\]
\[
= 0.88
\]

\[
y^{(i)} = 0.7 x^{(i)} + 0.3 z^{(i)} + e_{y}^{(i)}
\]
\[
= 0.7 \times 0.88 + 0.3 \times 0.47 - 0.19
\]
\[
= 0.57
\]

これで 100 回によりも 100 個の観測値で生成（100 個のデータ）
3つのデータ生成過程の散布図

未観測共通原因

\[ x = 0.3z + e_x \]
\[ y = 0.7x + 0.3z + e_y \]

\( z, e_x, e_y \) はガウス分布

未観測共通原因

\[ x = 0.7y + 0.3z + e_x \]
\[ y = 0.3z + e_y \]

\( z, e_x, e_y \) はガウス分布

未観測共通原因

\[ x = 0.89z + e_x \]
\[ y = 0.89z + e_y \]

\( z, e_x, e_y \) はガウス分布

\( x, y, z \) の平均 0

\( z \) の分散 1

\( x, y \) の分散が 1になるように \( e_x \) と \( e_y \) の分散を設定

どれも相関係数は 0.79
実際 に生成 する と

\[ x = 0.3 \, z + \varepsilon_x \]
\[ y = 0.7 \, x + 0.3 \, z + \varepsilon_y \]

\[
\begin{bmatrix}
  z \\
  e_x \\
  e_y \\
\end{bmatrix}
\sim
N
\begin{bmatrix}
  0 \\
  0 \\
  0 \\
\end{bmatrix},
\begin{bmatrix}
  1 & 0 & 0 \\
  0 & 0.91 & 0 \\
  0 & 0 & 0.29 \\
\end{bmatrix}
\]

Excel や スペードシート

1. \[ y^{(1)} = 0.7 \, x^{(1)} + \varepsilon_x^{(1)} \]  \( x^{(0)} = 0.3 \, z + \varepsilon_x \)
2. ...
3. ...
4. ...
散布図

t=[12713]

同様に、以下の生成元散布図

\[ x = 0.7y + 0.3z + e_x \]
\[ y = 0.3z + e_y \]

\[ x = 0.89z + e_x \]
\[ y = 0.89z + e_y \]
因果推論のフレームワーク
（第2章）
反実仮想モデルによる
因果の定義
個体レベルの因果
(Neyman, 1923; Rubin, 1974; cf. Hernan, 2004)

• 患者ゼウスにとって、投薬は治癒の原因か？
  - 「もしも薬を飲ませたとしたら、治癒するか」
  ≠ 「もしも飲ませなかったとしたら、治癒するか」
  ⇒ 「ゼウスにとって、投薬は治癒するかどうかの原因(の1つ)」
  - もしも○○だとしたら: 反事実モデル

• 投薬という変化を起こすと、治癒するという変化が起きる

3日後
治癒
≠
治癒せず
因果推論の基本問題
(Holland, 1986)

- 個体レベルの因果は、その個体のデータだけからは同定できない
  - 事実は1つ: 薬を飲ませてしまったら、薬を飲ませなかった場合にどうなるかは不明

ゼウス

薬

なし

3日後
治癒
≠
?

(Holland, 1986)
集団レベルの因果
(Neyman, 1923; Rubin, 1974)

• 患者集団(ゼウス、ヘラ、アテナ、アポロン…たくさん)
  - 「もしも全員に薬を飲ませた場合の治癒する割合」
  ≠「もしも全員に飲ませなかった場合の治癒する割合」
  ⇒「この集団において、投薬は治癒の原因」

• 集団における因果は同定できる場合がある(以降のスライド)
データ生成過程のモデル：構造方程式モデル
構造方程式モデル (Bollen, 1989; Pearl, 2000)

- データ生成過程のモデル
  - 変数の「値」が、どういう過程を経て生成されるか

- 構造方程式: 変数の「値」の決定関係を表す
  - \( y = g_y(x, f, e_y) \)
    - 「単なる等式ではない (:= や ← と解釈)」
    - \( e_y: y\) の値を決定するために必要な要因全て(\( x\)以外)

\[
\begin{align*}
  x &= g_x(f, e_x) \\
  y &= g_y(x, f, e_y)
\end{align*}
\]

構造方程式

\( x \) (薬) \hspace{2cm} (f \) (重症度)
\( y \) (治癒)

図果グラフ
パス図のルール

片方向矢印(有向辺)

- 左辺を計算するのに必要かも
  - 必要ない: 右辺の当該引数の値を、どの違う値にしたとしても、他の引数の値を変えなければ、左辺の値が変わらない

両方向矢印付き円弧(有向円弧)

- 未観測共通原因が存在するかも
  - 複数の変数の親(祖先)となる潜在変数
  - $e_x$ と $e_y$ が従属に

\[
x = e_x
\]
\[
y = f_y(x, e_y)
\]

構造方程式

パス図
(ノンパラ) 構造方程式モデル: 一般に

・ 以下の四つ組(Pearl, 2000):
  • \( \mathbf{v} = [v_1, \ldots, v_p] \): 内生変数
  • \( \mathbf{u} = [u_1, \ldots, u_q] \): 外生変数
  • \( \mathbf{f} = \{f_1, \ldots, f_p\} \): (決定的)関数
  \[ v_i = f_i(v, u) \quad -- \text{構造方程式} \]
  • \( p(u) \): 外生変数の分布

・ 関数\( f \)と分布\( p(u) \)から分布\( p(v) \)が決まる

・ \( \mathbf{v} \)のうちの観測変数\( \mathbf{o} \)の分布\( p(\mathbf{o}) \)から
  統計的推測を行う
構造方程式モデルを用いて
集団レベルの因果を記述
介入 (Pearl, 2000)

- 介入: 変数の値を(他の変数に依らず=強制で)固定する
  - 薬を飲ませる: do(薬=飲む) or do(x=1)
  - xの構造方程式を「x=1」に取り換える

- 介入前のデータ生成過程 (観察データ: 自然におまかせ)

\[
\begin{align*}
x &= g_x(f, e_x) \\
y &= g_y(x, f, e_y)
\end{align*}
\]

構造方程式

- 介入後のデータ生成過程 $M_{x=1}$

\[
\begin{align*}
x &= g_x(f, e_x) \\
y &= g_y(x, f, e_y)
\end{align*}
\]

自律性の仮定: 他の関数は変わらない
介入後の分布 (Pearl, 2000)

・介入後のyの分布 := 介入後のモデル $M_{x=1}$ での分布

$$p(y \mid do(x = 1)) := p_{M_{x=1}}(y)$$

介入後のモデル $M_{x=1}$

構造方程式

x (薬) ← 1

y (治癒) ← $e_y$

また、もし介入後のyの分布が違うxの値cとdがあれば、「この集団において、xはyの原因(のひとつ)」と言う

$$p(y \mid do(x = c)) \neq p(y \mid do(x = d))$$
定量化: 因果効果
(Rubin, 1974; Pearl, 2000)

(平均)因果効果 := \( E(y \mid do(x = d)) - E(y \mid do(x = c)) \)

・変数xの値をcからdに変化させた時に、変数yの値が平均的にどのくらい変化するか

・変化させる: do(x=c)をした後、cをdに変える

・分散で測る: \( Var(y \mid do(x = d)) - Var(y \mid do(x = c)) \)
例1

- $x$を定数$c$から$d$へ変化させたときの$y$への因果効果
  
  \[
  = E(\ y\ |\ 母集団全員のxをdにする\ )
  - E(\ y\ |\ \cdots\ \ xをcにする\ )
  \]

  \[
  = E(y \mid do(x = d)) - E(y \mid do(x = c))
  = E(b_{yx} \ d + e_y) - E(b_{yx} \ c + e_y)
  = b_{yx}(d - c)
  \]

モデル1:
\[
\begin{align*}
  x &= e_x \\
  y &= b_{yx}x + e_y
\end{align*}
\]

モデル1’:
\[
\begin{align*}
  x &= d \\
  y &= b_{yx}x + e_y
\end{align*}
\]
例2

yを定数 c から d へ変化させたときの x への因果効果

= \( E(x \mid \text{母集団全員の} y = d) - E(x \mid \text{母集団全員の} y = c) \)

= \( E(e_x) - E(e_x) \)

= 0

モデル1:

\[
x = e_x
\]

\[
y = b_{yx} x + e_y
\]

モデル1”:

\[
x = e_x
\]

\[
y = d
\]
補足：個体レベルの因果 (Pearl, 2000)

• do(x=d)のモデルでのゼウスのyと
do(x=c)のモデルでのゼウスのyを比較

\[ y_{x=d} - y_{x=c} = (b_{yx}d + e_{y}^{\text{ゼウス}}) - (b_{yx}c + e_{y}^{\text{ゼウス}}) \]
\[ = b_{yx}(d - c) \]
\[ = f_{y}(d, e_{y}^{\text{ゼウス}}) - f_{y}(c, e_{y}^{\text{ゼウス}}) \]

同じ \( e_{y}^{\text{ゼウス}} \) の値 \( e_{y}^{\text{ゼウス}} \) : 同じ個体（x以外の条件は同じ）

モデル1:

\[ x = e_{x} \]
\[ y = b_{yx}x + e_{y} \]

\[ \text{モデル1’ do(x=d)}: \]

\[ x = d \]
\[ y = b_{yx}x + e_{y} \]
$E(y \mid do(x = c))$ の計算

介入前のデータ生成過程 (自然におまかせ)

\[ x = \lambda_x f + e_x \]
\[ y = bx + \lambda_y f + e_y \]

xの値を強制的にcにする $do(x = c)$

\[ x = c \]
\[ y = bx + \lambda_y f + e_y \]

下のモデルでの $E(y)$ が

\[ E(y \mid do(x = c)) = bc \]
逆に、yの値を変化させたら？

試しに計算してみましょう
総合効果・直接効果・間接効果
因果効果=総合効果

- 因果効果(総合効果)
  - 性別を男から女に変えると、雇用される確率はどのくらい変わるか？
  - 保育士？

ここで、この因果グラフが正しいとしよう (Pearl, 2000)
構造方程式モデルで表現

・xを男から女に変えたときのz(雇用)への総合効果:

\[ E(z \mid do(x = \text{女})) - E(z \mid do(x = \text{男})) \]

別の表現:

\[ E(z_{x=\text{女}}) - E(z_{x=\text{男}}) \]

強制的にx=女の仮想集団

\[
\begin{array}{cccc}
& & & \\
& & e_y & \\
y(適性) & & \\
& x (性別) & & \\
& & z(雇用) & \\
& & & e_z \\
\end{array}
\]

vs.

強制的にx=男の仮想集団

\[
\begin{array}{cccc}
& & & \\
& & e_y & \\
y(適性) & & \\
& x (性別) & & \\
& & z(雇用) & \\
& & & e_z \\
\end{array}
\]
直接効果

- 直接効果: 性別は男から女に変えるが、適性は変えないとき、
  雇用される確率はどのくらい変わるか？
  - これが大きいと、性差別がある
構造方程式モデルで表現
(Robins & Greenland, 1992; Pearl, 2001)

・xを男から女に変えたときのz(雇用)への直接効果:

\[ E(z_{x=女, y=y_{x=男}}) - E(z_{x=男}) \]

強制的にx=女, y = y_{x=男}の仮想集団

強制的にx=男の仮想集団

vs.
間接効果 (Pearl, 2001)

- 間接効果: 性別は男のまま、
  適性を性別を女に変えた場合と同じにしたとしたら、
  雇用される確率はどのくらい変わるか？
  - 性差別を取り除いた時の、性別から雇用への効果
構造方程式モデルで表現 (Pearl, 2001)

・xを男から女に変えたときのz(雇用)への間接効果:

\[ E(z_{x=男, y=y_{x=女}}) - E(z_{x=男}) \]

強制的にx=男, y = y_{x=女} の仮想集団

強制的にx=男の仮想集団

\[ x \quad \text{(性別)} \]

\[ y_{x=女} \quad \text{(適性)} \]

\[ z(雇用) \]

\[ e_z \]

vs.

\[ x \quad \text{(性別)} \]

\[ y(適性) \]

\[ z(雇用) \]

\[ e_z \]
（仮に）線形なら

- 性別xを男(0)から女(1)へ変化させた時の雇用zへの
  - 総合効果=a+bc
  - 直接効果=a
  - 間接効果=bc

性別 = e_x
適性 = b × 性別 + e_y
雇用 = a × 性別 + c × 適性 + e_z
因果効果の識別性 (推定可能性)
因果構造（因果グラフ・パス図）が既知の場合
因果効果の識別性：doのない形にかけるか？

• 非巡回で未観測共通原因がない場合(Pearl, 1995):

\[
E(y \mid do(x = d)) = E(y \mid x = d)
\]

モデル1’ do(x=d):

\[
\begin{align*}
x &= d \\
y &= f_y(x, e_y)
\end{align*}
\]

証明:

\[
\begin{align*}
E[y \mid do(x = d)] &= E[f_y(d, e_y)] \\
&= E[f_y(d, e_y)] \\
&= E[y \mid x = d]
\end{align*}
\]

\[
E[y \mid x = d] = E[f_y(x, e_y) \mid x = d]
\]

\[
= E[f_y(d, e_y) \mid x = d]
\]

\[
= E[f_y(d, e_y)]
\]

\[
p(e_y \mid x) = p(e_y)
\]

\[
x(= e_x)とe_yが独立だから
\]
因果効果の識別性：doのない形にかけるか？

・非巡回で交絡変数がある場合:
  - xの親を観測して調整 (十分条件(Pearl, 1995)):

\[
E(y \mid do(x)) = E_{x\text{'s parents}}[E(y \mid x, x\text{'s parents})]
\]

- 線形なら

\[
E[y \mid do(x = d)] - E[y \mid do(x = c)]
= x\text{'s partial regression coefficient} \times (d - c)
\]

・より詳しくは Shpitser and Pearl (2006, 2008)
  - 因果構造(因果グラフ)が分れば、識別可能か判定可能
  - 因果構造に関する知識が必要
ランダム化実験: グラフが既知に

• 仮定1: ランダム化
  ⇒ 外生変数が独立に=交絡変数がない

• 仮定2(事前知識): 時間情報
  ⇒ ありうる向きが決まる(非巡回)

\[
E(y \mid do(x)) = E(y \mid x)
\]

観察データの因果グラフ

ランダム化実験の因果グラフ
因果効果の推定には
因果構造に関する情報が必要

• いつもランダム化できるとは限らない
  - 倫理
  - コスト

• ランダム化のないデータ=観察データでは、
  例えば親変数がどれか知る必要がある

• 因果構造自体も興味の対象
やってみよう

- 次は、集団Aの構造方程式モデルである。以下、集団Aについて答えよ。

\[ x = e_x \]
\[ y = 0.5x + e_y \]
\[ z = 0.3x - 0.8y + e_z \]

\[ e_x, e_y, e_z \]は独立な確率変数: \( p(e_x, e_y, e_z) = p(e_x)p(e_y)p(e_z) \)

- 問題1: この構造方程式モデルのパス図をかけ。
- 問題2: xをcからdに変えたときのzへの総合効果を求めよ。
- 問題3: xをcからdに変えたときのzへの直接効果を求めよ。
- 問題4: xをcからdに変えたときのzへの間接効果を求めよ。