

統計的因果推論入門

東北大学 MathCSS
多田駿介

決定論的支配方程式による非線形・散逸・偶然性の動力学
因果解析班ミーティング
2025/12/08

目的

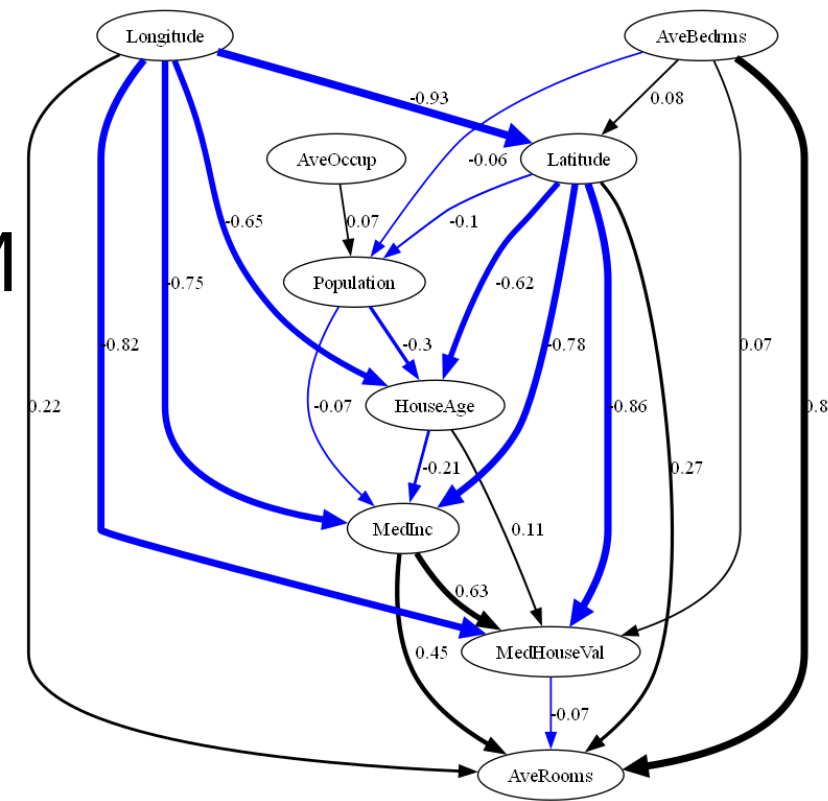
- DirectLiNGAM の出力(因果グラフ)をちゃんと理解する.
- DirectLiNGAM を動かせるようになる(+出力結果の信頼性評価)

データ

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	MedHouseVal
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422
...
20635	1.5603	25.0	5.045455	1.133333	845.0	2.560606	39.48	-121.09	0.781
20636	2.5568	18.0	6.114035	1.315789	356.0	3.122807	39.49	-121.21	0.771
20637	1.7000	17.0	5.205543	1.120092	1007.0	2.325635	39.43	-121.22	0.923
20638	1.8672	18.0	5.329513	1.171920	741.0	2.123209	39.43	-121.32	0.847
20639	2.3886	16.0	5.254717	1.162264	1387.0	2.616981	39.37	-121.24	0.894

DirectLiNGAM

重み付き有向グラフ



内容

(0)統計的因果推論とは

(1)統計的因果推論基礎

- 構造方程式モデル(SEM)
- SEMの因果グラフ
- 平均因果効果

(2)DirectLiNGAM

- LiNGAMモデル
- DirectLiNGAMによる因果探索のデモンストレーション
- 結果の信頼性評価とそのデモンストレーション

(3)流体データに適用

内容

(0)統計的因果推論とは

(1)統計的因果推論基礎

- 構造方程式モデル(SEM)
- SEMの因果グラフ
- 平均因果効果

(2)DirectLiNGAM

- LiNGAMモデル
- DirectLiNGAMによる因果探索のデモンストレーション
- 結果の信頼性評価とそのデモンストレーション

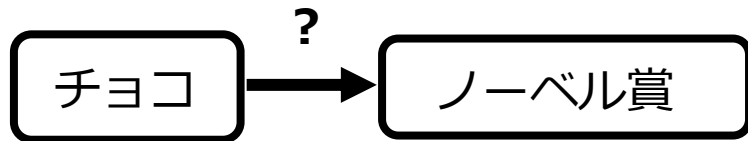
(3)流体データに適用

(0)統計的因果推論とは

統計的因果推論：データを物事の**因果関係**や**因果の強さ**を統計的手法を用いて調べる分野

必要とされる例

- ・チョコは脳の効率的な活用に寄与する[1].
- ・国ごとのチョコの消費量とノーベル賞の受賞者数に相関あり[2].



[1]The effects of dark chocolate on cognitive performance during cognitively demanding tasks: A randomized, single-blinded, crossover, dose-comparison study, *Heliyon*, 2024, 10 (2)

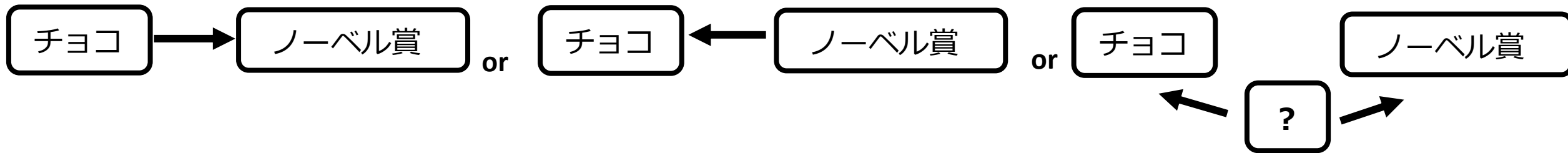
[2]Franz H. Messerli, "Chocolate Consumption, Cognitive Function, and Nobel Laureates." *New England Journal of Medicine* (367), 1563, Massachusetts Medical Society, 2012.

(0)統計的因果推論とは

統計的因果推論：データを物事の**因果関係**や**因果の強さ**を統計的手法を用いて調べる分野

必要とされる例

- ・チョコは脳の効率的な活用に寄与する[1].
- ・国ごとのチョコの消費量とノーベル賞の受賞者数に相関あり[2].



→**因果関係**や**因果の強さ**をちゃんと知りたい！

[1]The effects of dark chocolate on cognitive performance during cognitively demanding tasks: A randomized, single-blinded, crossover, dose-comparison study, *Heliyon*, 2024, 10 (2)

[2]Franz H. Messerli, "Chocolate Consumption, Cognitive Function, and Nobel Laureates." *New England Journal of Medicine* (367), 1563, Massachusetts Medical Society, 2012.

(0)統計的因果推論とは

統計的因果推論：データを物事の**因果関係**や**因果の強さ**を統計的手法を用いて調べる分野

因果関係を探索

- ・ PC アルゴリズム
- ・ . . .

- ・ DirectLiNGAM
- ・ ICA-LiNGAM
- ・ VAR-LiNGAM
- ・ . . .

因果の強さ(ex. 平均因果効果)を測る

- ・ ランダム化実験
(Golden standard)
- ・ 操作変数法
- ・ 回帰不連続デザイン
- ・ . . .

->事例

(0)統計的因果推論とは：因果の強さを測る

事例 1 (PALM試験)

エボラウイルス熱に対して、どの薬がより治療効果が大きいか？

-> MAb114 と REGN-EB3 は ZMappよりも効果が大きい

S. Mulangu et al, A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics, The New England Journal of Medicine, vol. 381, 2019.

事例 2 (きれいな資源ステーション協働プロジェクト)

ゴミ捨て場の不法投棄問題の対策として「チラシの投函」と「ゴミ収集終了の看板」の設置はどちらが効果的？

-> ゴミ収集終了の看板が効果的

-> 町の政策に反映された

<https://www.town.hayama.lg.jp/soshiki/seisaku/osirase/6870.html>

(0)統計的因果推論とは：因果探索

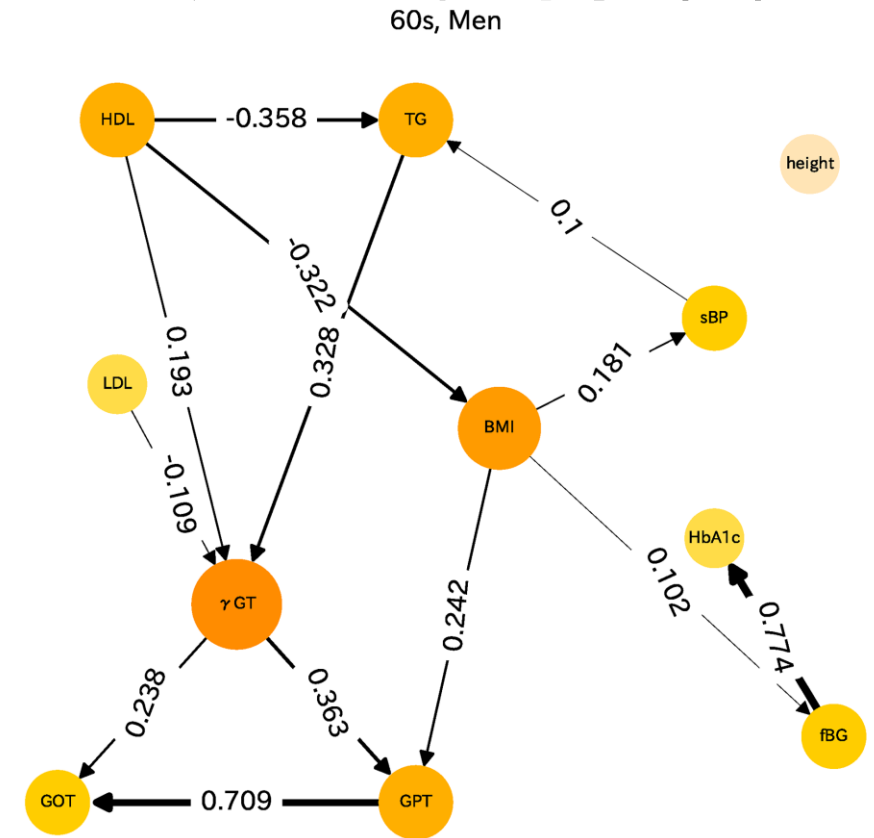
例 3(健康指標間の因果関係)

大阪府の健康診断データにDirectLiNGAMを適用して健康指標間の因果関係を調べた

-> 新たな知見
(HDLがBMIとTGに及ぼす影響は重要[3])

HDL: high-density lipoprotein cholesterol
(高比重リポタンパクコレステロール)

TG : triglyceride (中性脂肪)



[3] J. Kotoku et al, Causal relations of health indices inferred statistically using the DirectLiNGAM algorithm from big data of Osaka prefecture health checkups. Plos one, 15(12), 2020, Fig4

(0)統計的因果推論とは

統計的因果推論：データを物事の**因果関係**や**因果の強さ**を統計的手法を用いて調べる分野

因果関係を探索

- ・ PC アルゴリズム
- ・ . . .

- ・ **DirectLiNGAM**
- ・ ICA-LiNGAM
- ・ VAR-LiNGAM
- ・ . . .

因果の強さ(ex. 平均因果効果)を測る

- ・ ランダム化実験
(Golden standard)
- ・ 操作変数法
- ・ 回帰不連続デザイン
- ・ . . .

内容

(0)統計的因果推論とは

(1)統計的因果推論基礎

- 構造方程式モデル(SEM)
- SEMの因果グラフ
- 平均因果効果

(2)DirectLiNGAM

- LiNGAMモデル
- DirectLiNGAMによる因果探索のデモンストレーション
- 結果の信頼性評価とそのデモンストレーション

(3)流体データに適用

(1)統計的因果推論基礎

Setting

- ・ 行列 A に対し, その転置行列を A^T とかく.
- ・ 確率変数はある確率空間 (Ω, \mathcal{F}, P) 上の実数値確率変数とする.
- ・ 確率変数 X に対して X の期待値を $\mathbb{E}[X]$, 分散を $\text{Var}(X)$ とかく.
- ・ 確率変数 X, Y に対してそれらの共分散を $\text{Cov}(X, Y)$ と書く.

(1)統計的因果推論基礎

定義 (構造方程式モデル)

確率変数 $X_1, \dots, X_n, E_1, \dots, E_n$ と $f_1, \dots, f_n: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ が

$$F_i: X_i = f_i(X_1, \dots, X_n, E_i) \quad (i = 1, \dots, n)$$

を満たすとする. このとき, $(X_i)_i, (E_i)_i, (F_i)_i$ からなる三つ組を**構造方程式モデル**という. また, F_i, X_i, E_i をそれぞれ**構造方程式**, **内生変数**, **外生変数**という. また,

$$F_i: X_i = \sum_{j=1}^n b_{i,j} X_j + E_i \quad (b_{i,j} \in \mathbb{R})$$

と書けるときこの構造方程式モデルを**線形構造方程式モデル**という.

- X_1, \dots, X_n が観測データを表し, E_1, \dots, E_n がノイズなどを表すイメージ.
- 上の線形構造方程式モデルは行列を用いて,

$$X = B X + E$$

と集約して書ける. ただし, $X := (X_1, \dots, X_n)^T$ $E := (E_1, \dots, E_n)^T$ で B は (i, j) 成分が $b_{i,j}$ となる実行列である.

(1)統計的因果推論基礎

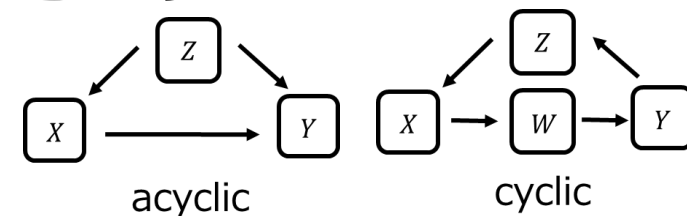
内生変数 $X_i (i = 1, \dots, n)$ が次で与えられる線形構造方程式モデル

$$X_i = \sum_{j=1}^n b_{i,j} X_j + E_i \quad (b_{i,j} \in \mathbb{R})$$

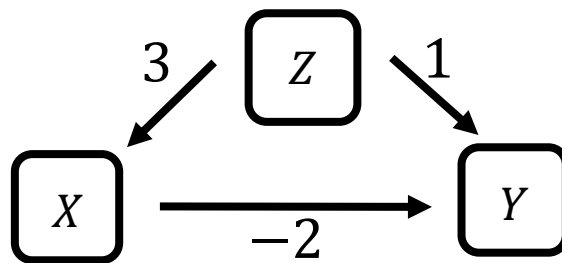
を考える． 次で定まる重み付き有向グラフを**因果グラフ**と呼ぶ：

頂点集合： $\{X_1, \dots, X_n\}$, 辺集合： $\{(X_i, X_j) | b_{j,i} \neq 0\}$, 辺の重み： $w(X_i, X_j) = b_{j,i}$ ．

因果グラフがacyclicのときその構造方程式モデルは**acyclic**という．



例 $X = 3Z + E_X$, $Y = -2X + Z + E_Y$, $Z = E_Z$ からなる線形構造方程式モデルの因果グラフは次で与えられる．



(1)統計的因果推論基礎

定義 (介入)

M を内生変数 X_1, \dots, X_n , 外生変数 E_1, \dots, E_n と構造方程式

$$X_i = f_i(X_1, \dots, X_n, E_i) \quad (i = 1, \dots, n)$$

からなるacyclicな構造方程式モデルとする. $i \in \{1, \dots, n\}, a \in \mathbb{R}$ に対して

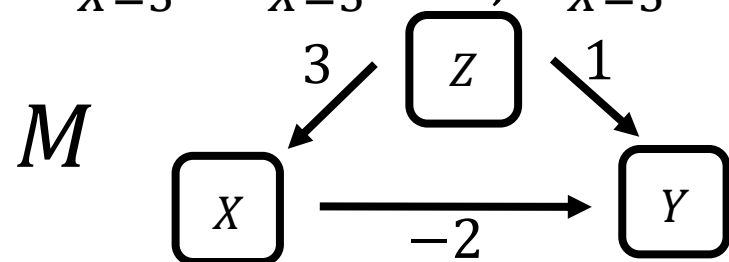
$$X_{i, X_i=a} := a,$$

$$X_{j, X_i=a} := f_j(X_{1, X_i=a}, \dots, X_{n, X_i=a}, E_j) \quad (j \neq i) \quad \dots (1)$$

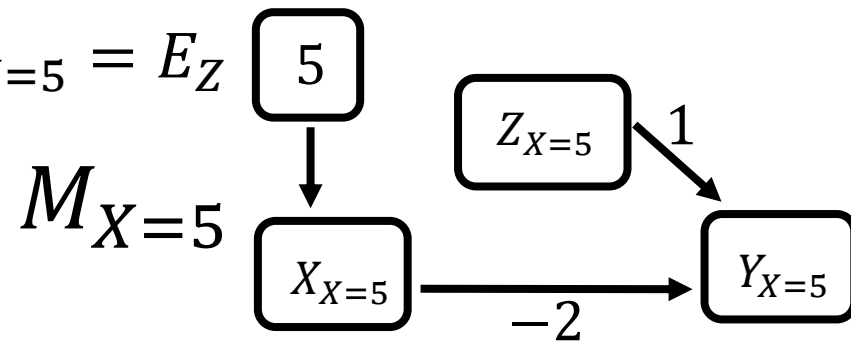
を定める. このとき, 内生変数 $a, X_{i, X_i=a}, \dots, X_{n, X_i=a}$, 外生変数 E_1, \dots, E_n , 構造方程式が(1)となる構造方程式モデルを $M_{X_i=a}$ とかく. この M から $M_{X_i=a}$ を作る操作を X_i に**介入**するという.

例 $M: X = 3Z + E_X, Y = -2X + Z + E_Y, Z = E_Z$

$\rightarrow M_{X=5}: X_{X=5} = 5, Y_{X=5} = -2X_{X=5} + Z_{X=5} + E_Y, Z_{X=5} = E_Z$



X に介入
 \longrightarrow



(1)統計的

このノートでは介入がうまく定義されるために**acyclic**性を課す.
介入に関して清水先生の「統計的因果探索入門」などを参照

定義 (介入)

M を内生変数 X_1, \dots, X_n , 外生変数 E_1, \dots, E_n と構造方程式

$$X_i = f_i(X_1, \dots, X_n, E_i) \quad (i = 1, \dots, n)$$

からなるacyclicな構造方程式モデルとする. $i \in \{1, \dots, n\}, a \in \mathbb{R}$ に対して

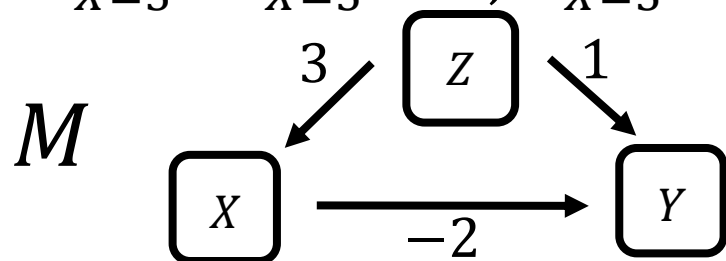
$$X_{i, X_i=a} := a,$$

$$X_{j, X_i=a} := f_j(X_{1, X_i=a}, \dots, X_{n, X_i=a}, E_j) \quad (j \neq i) \quad \dots (1)$$

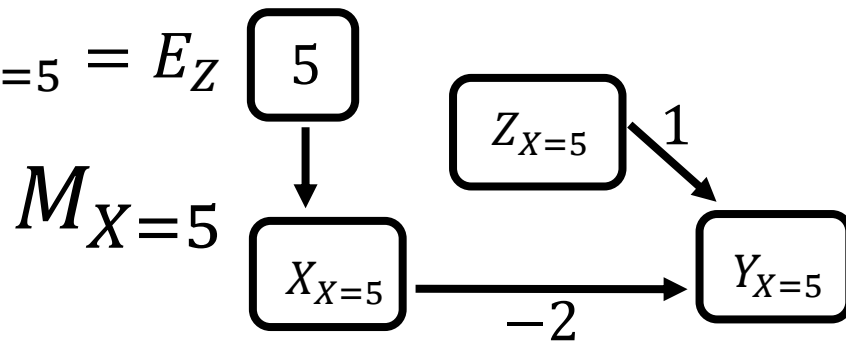
を定める. このとき, 内生変数 $a, X_{i, X_i=a}, \dots, X_{n, X_i=a}$, 外生変数 E_1, \dots, E_n , 構造方程式が(1)となる構造方程式モデルを $M_{X_i=a}$ とかく. この M から $M_{X_i=a}$ を作る操作を X_i に**介入**するという.

例 $M: X = 3Z + E_X, Y = -2X + Z + E_Y, Z = E_Z$

$\rightarrow M_{X=5}: X_{X=5} = 5, Y_{X=5} = -2X_{X=5} + Z_{X=5} + E_Y, Z_{X=5} = E_Z$



X に介入
→



(1)統計的因果推論基礎

定義 (介入)

M を内生変数 X_1, \dots, X_n , 外生変数 E_1, \dots, E_n と構造方程式

$$X_i = f_i(X_1, \dots, X_n, E_i) \quad (i = 1, \dots, n)$$

からなるacyclicな構造方程式モデルとする. $i \in \{1, \dots, n\}, a \in \mathbb{R}$ に対して

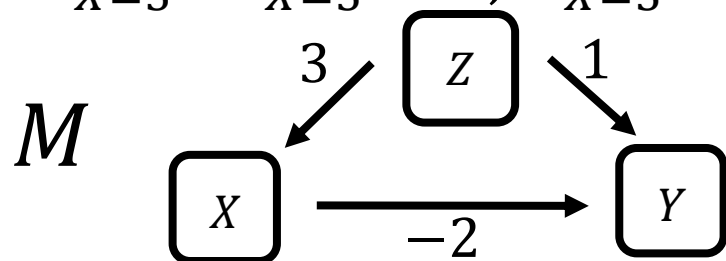
$$X_{i, X_i=a} := a,$$

$$X_{j, X_i=a} := f_j(X_{1, X_i=a}, \dots, X_{n, X_i=a}, E_j) \quad (j \neq i) \quad \dots (1)$$

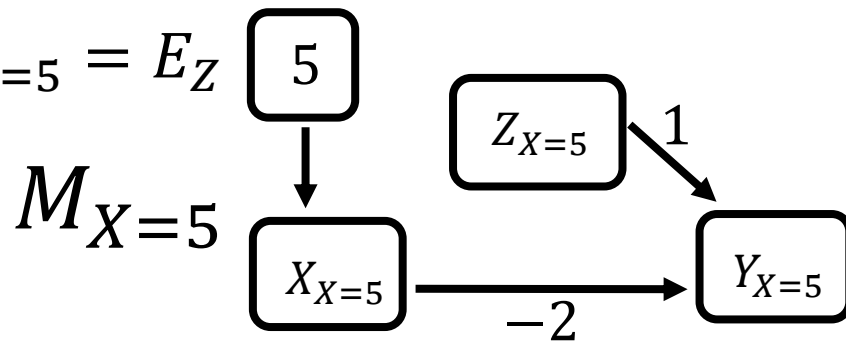
を定める. このとき, 内生変数 $a, X_{i, X_i=a}, \dots, X_{n, X_i=a}$, 外生変数 E_1, \dots, E_n , 構造方程式が(1)となる構造方程式モデルを $M_{X_i=a}$ とかく. この M から $M_{X_i=a}$ を作る操作を X_i に**介入**するという.

例 $M: X = 3Z + E_X, Y = -2X + Z + E_Y, Z = E_Z$

$\rightarrow M_{X=5}: X_{X=5} = 5, Y_{X=5} = -2X_{X=5} + Z_{X=5} + E_Y, Z_{X=5} = E_Z$



X に介入
 \longrightarrow



(1)統計的因果推論基礎

定義 (介入)

M を内生変数 X_1, \dots, X_n , 外生変数 E_1, \dots, E_n と構造方程式

$$X_i = f_i(X_1, \dots, X_n, E_i) \quad (i = 1, \dots, n)$$

からなるacyclicな構造方程式モデルとする. $i \in \{1, \dots, n\}, a \in \mathbb{R}$ に対して

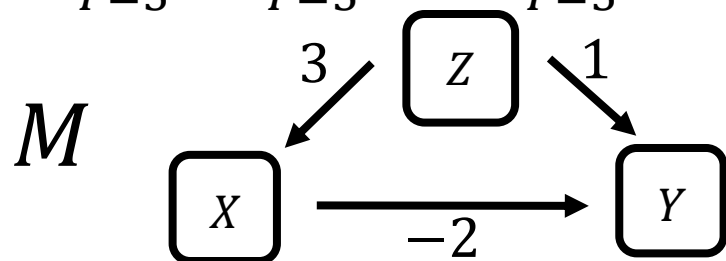
$$X_{i, X_i=a} := a,$$

$$X_{j, X_i=a} := f_j(X_{1, X_i=a}, \dots, X_{n, X_i=a}, E_j) \quad (j \neq i) \quad \dots (1)$$

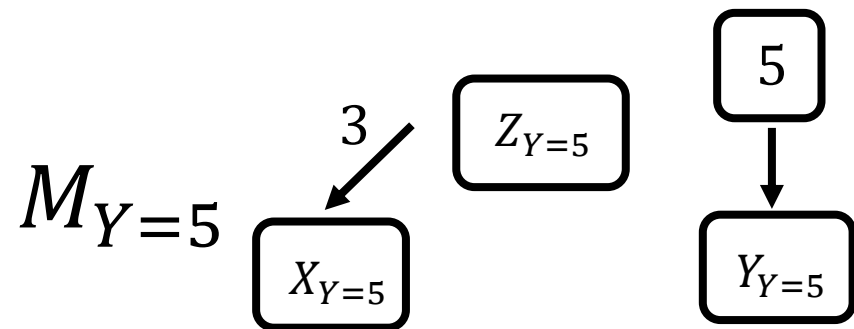
を定める. このとき, 内生変数 $a, X_{i, X_i=a}, \dots, X_{n, X_i=a}$, 外生変数 E_1, \dots, E_n , 構造方程式が(1)となる構造方程式モデルを $M_{X_i=a}$ とかく. この M から $M_{X_i=a}$ を作る操作を X_i に**介入**するという.

例 $M: X = 3Z + E_X, Y = -2X + Z + E_Y, Z = E_Z$

$\rightarrow M_{Y=5}: X_{Y=5} = 3Z_{Y=5} + E_X, Y_{Y=5} = 5, Z_{Y=5} = E_Z$



Y に介入
→



(1)統計的因果推論基礎

定義 (介入)

M を内生変数 X_1, \dots, X_n , 外生変数 E_1, \dots, E_n と構造方程式

$$X_i = f_i(X_1, \dots, X_n, E_i) \quad (i = 1, \dots, n)$$

からなるacyclicな構造方程式モデルとする. $i \in \{1, \dots, n\}, a \in \mathbb{R}$ に対して

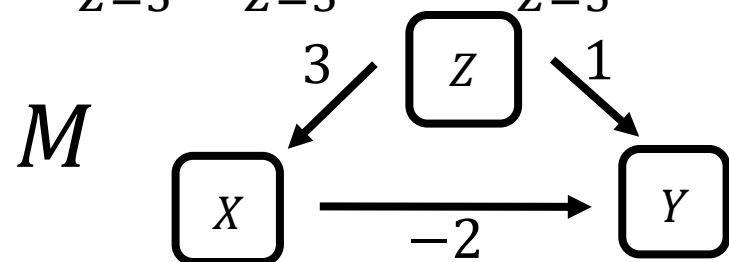
$$X_{i, X_i=a} := a,$$

$$X_{j, X_i=a} := f_j(X_{1, X_i=a}, \dots, X_{n, X_i=a}, E_j) \quad (j \neq i) \quad \dots (1)$$

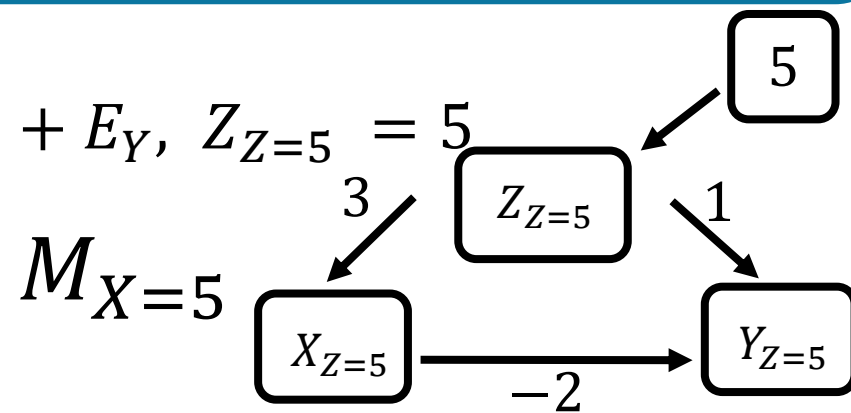
を定める. このとき, 内生変数 $a, X_{i, X_i=a}, \dots, X_{n, X_i=a}$, 外生変数 E_1, \dots, E_n , 構造方程式が(1)となる構造方程式モデルを $M_{X_i=a}$ とかく. この M から $M_{X_i=a}$ を作る操作を X_i に**介入**するという.

例 $M: X = 3Z + E_X, Y = -2X + Z + E_Y, Z = E_Z$

$$\rightarrow M_{Z=5}: X_{Z=5} = 3Z_{Z=5} + E_X, Y_{Z=5} = -2X_{Z=5} + Z_{Z=5} + E_Y, Z_{Z=5} = 5$$



Z に介入
→



(1)統計的因果推論基礎

定義 [Rubin-1974](平均因果効果;Average Causal Effect)

確率変数 X, Y を構造方程式モデルの内生変数とする. $a, b \in \mathbb{R}$ に対して,

$$\text{ACE}(X, Y; a, b) := \mathbb{E}[Y_{X=a}] - \mathbb{E}[Y_{X=b}]$$

を X から Y への**平均因果効果(ACE)**と呼ぶ.

補足 X, Y : 内生変数, E_X, E_Y : 外生変数からなる構造方程式モデルを考える. このとき,

- ・ 一般に $\mathbb{E}[Y_{X=a}]$ は条件付き期待値 $\mathbb{E}[Y | X = a]$ とは異なる.

Ex. $X = E_X, Y = X + E_Y, E_X = E_Y$ で与えられる構造方程式モデルを考える. ただし $X = E_X = E_Y$ はコイン投げの施行に対応する確率変数とする(表 $\mapsto 1$, 裏 $\mapsto 0$). すると

$$\mathbb{E}[Y_{X=0}] = \mathbb{E}[E_Y] = 1/2$$

$$\mathbb{E}[Y | X = 0] = \mathbb{E}[X + E_Y | X = 0] = 2\mathbb{E}[E_Y | E_Y = 0] = 0$$

より $\mathbb{E}[Y_{X=0}] \neq \mathbb{E}[Y | X = 0]$.

(1)統計的因果推論基礎

定義 [Rubin-1974](平均因果効果;Average Causal Effect)

確率変数 X, Y を構造方程式モデルの内生変数とする. $a, b \in \mathbb{R}$ に対して,

$$\text{ACE}(X, Y; a, b) := \mathbb{E}[Y_{X=a}] - \mathbb{E}[Y_{X=b}]$$

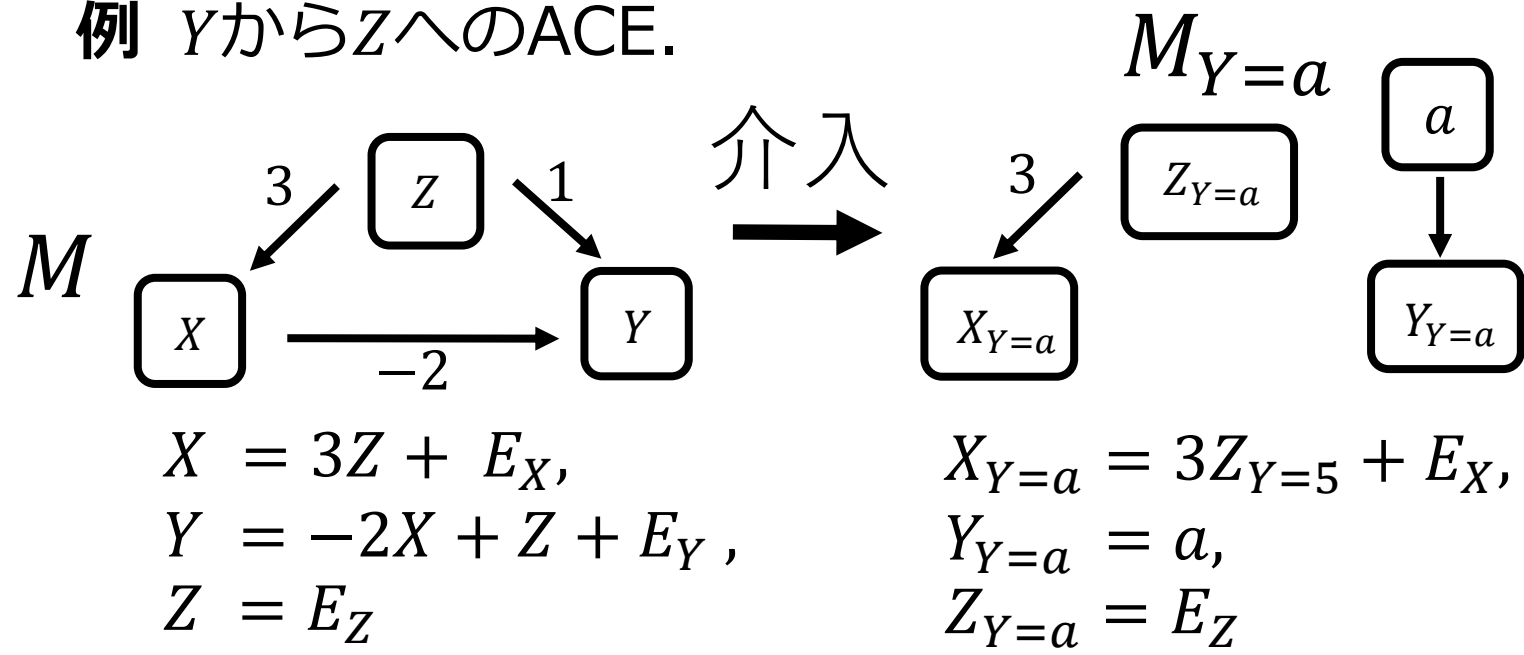
を X から Y への**平均因果効果(ACE)**と呼ぶ.

補足 X, Y : 内生変数, E_X, E_Y : 外生変数からなる構造方程式モデルを考える. また実数 a で $P(X = a) \neq 0$ を満たすものを取る(条件付き期待値を考えるため).このとき,

- 一般に $\mathbb{E}[Y_{X=a}]$ は条件付き期待値 $\mathbb{E}[Y | X = a]$ とは異なる.
- さらに E_X, E_Y が独立で構造方程式が $Y = f(X, E_Y), X = E_X$ で与えられるとき $\mathbb{E}[Y_{X=a}] = \mathbb{E}[Y | X = a]$ が成立(cf. ノートProposition 1.5.1.3, 「統計的因果探索入門」 p.43, 清水昌平). ランダム化実験では E_X, E_Y が独立となるような状況を作り, ACEを求めることを条件付き期待値を求めることに帰着する. 条件付き期待値はデータから推定するのでACEが推定できる

(1)統計的因果推論基礎

例 YからZへのACE.



ACE($Y, Z; a, b$)

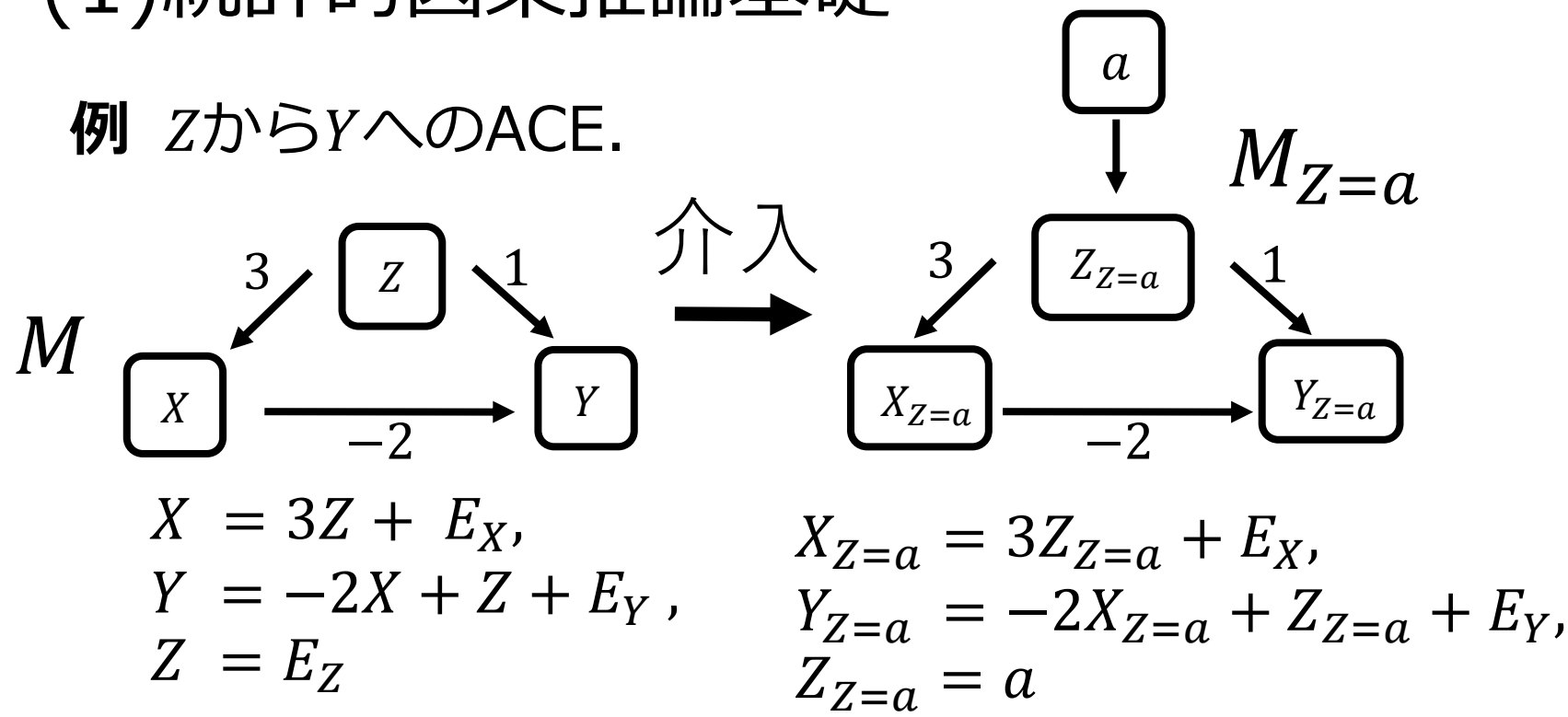
$$\begin{aligned} &= \mathbb{E}[Z_{Y=a}] - \mathbb{E}[Z_{Y=b}] \\ &= \mathbb{E}[E_Z] - \mathbb{E}[E_Z] \\ &= 0 \end{aligned}$$

ACE(列, 行; a, b)

	X	Y	Z
X	0	0	$3(a - b)$
Y	$-2(a - b)$	0	$-5(a - b)$
Z	0	0	0

(1)統計的因果推論基礎

例 ZからYへのACE.



ACE(列, 行; a, b)

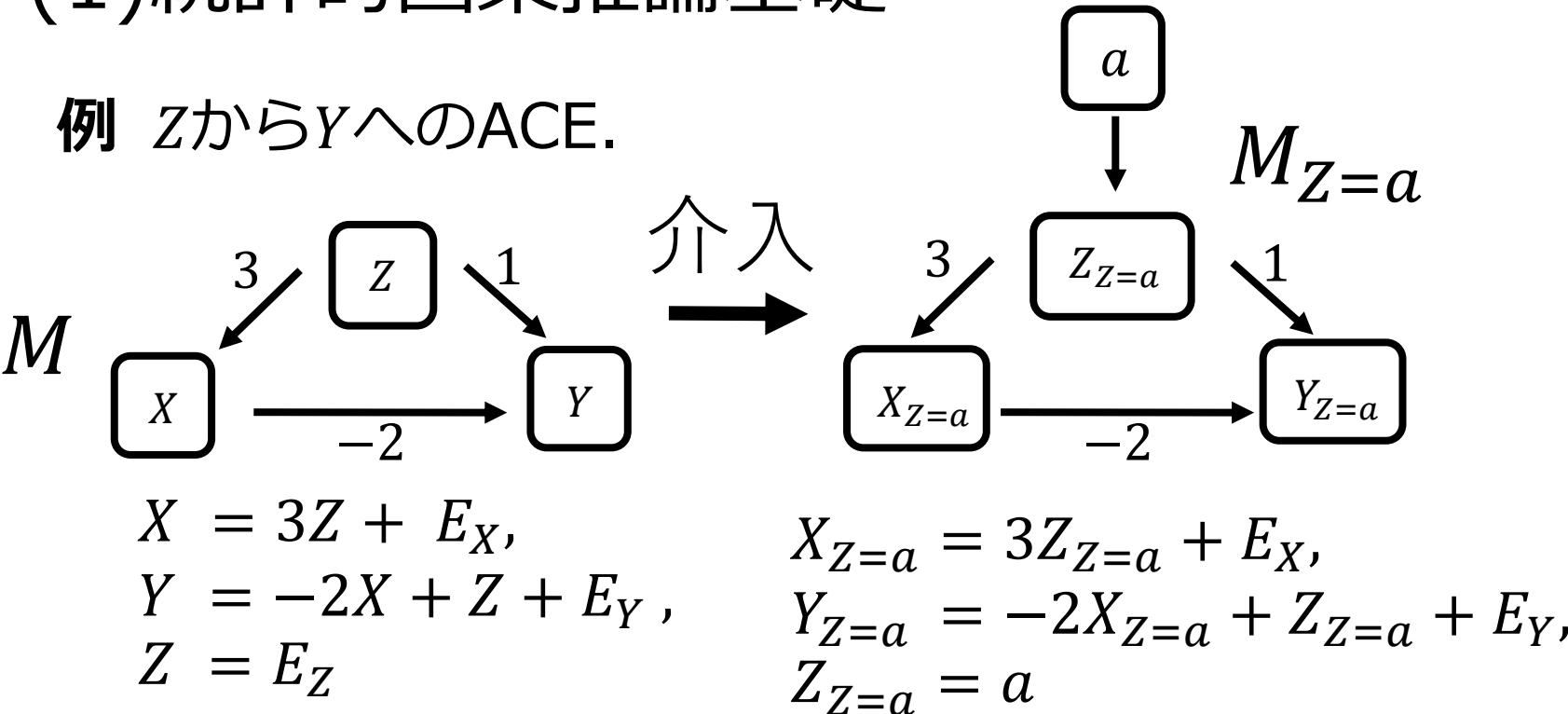
	X	Y	Z
X	0	0	$3(a - b)$
Y	$-2(a - b)$	0	$-5(a - b)$
Z	0	0	0

ACE($Z, Y; a, b$)

$$\begin{aligned} &= \mathbb{E}[Y_{Z=a}] - \mathbb{E}[Y_{Z=b}] \\ &= \mathbb{E}[-2X_{Z=a} + Z_{Z=a} + E_Y] - \mathbb{E}[-2X_{Z=b} + Z_{Z=b} + E_Y] \\ &= -2\mathbb{E}[X_{Z=a} - X_{Z=b}] + \mathbb{E}[Z_{Z=a} - Z_{Z=b}] \\ &= -2\mathbb{E}[3Z_{Z=a} + E_X - (3Z_{Z=b} + E_X)] + (a - b) \\ &= -6(\mathbb{E}[Z_{Z=a}] - \mathbb{E}[Z_{Z=b}]) + (a - b) = -5(a - b) \end{aligned}$$

(1)統計的因果推論基礎

例 ZからYへのACE.



ACE(列, 行; a, b) の計算式:

$$\begin{aligned} \text{ACE}(Z, Y; a, b) &= \mathbb{E}[Y_{Z=a}] - \mathbb{E}[Y_{Z=b}] \\ &= \mathbb{E}[-2X_{Z=a} + Z_{Z=a} + E_Y] - \mathbb{E}[-2X_{Z=b} + Z_{Z=b} + E_Y] \\ &= -2\mathbb{E}[X_{Z=a} - X_{Z=b}] + \mathbb{E}[Z_{Z=a} - Z_{Z=b}] \\ &= -2\mathbb{E}[3Z_{Z=a} + E_X - (3Z_{Z=b} + E_X)] + (a - b) \\ &= -6(\mathbb{E}[Z_{Z=a}] - \mathbb{E}[Z_{Z=b}]) + (a - b) = -5(a - b) \end{aligned}$$

ACE(列, 行; a, b)

	X	Y	Z
X	0	0	$3(a - b)$
Y	$-2(a - b)$	0	$-5(a - b)$
Z	0	0	0

=>計算公式

(1)統計的因果推論基礎

計算公式の準備

$G = (V, E, w)$ を重み付きグラフとする. 頂点 $x \in V$ から $y \in V$ へのpath

$$p : x_0 = x \rightarrow x_1 \rightarrow \cdots \rightarrow x_{n-1} \rightarrow x_n = y$$

に対してpathの重みを

$$w(p) := \prod_{i=0}^{n-1} w(x_i, x_{i+1})$$

で定める.

(1)統計的因果推論基礎

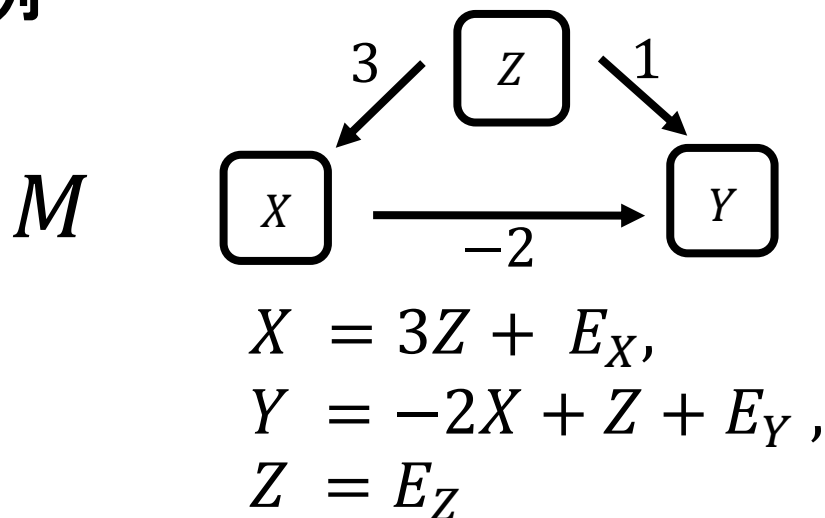
命題[cf. Pearl-Jewell, p82]

線形構造方程式モデル M をacyclicとする. このとき, 内生変数 X, Y に対して,

$$ACE(X, Y; a, b) = \left(\sum_{p: X \rightarrow \dots \rightarrow Y} w(p) \right) (a - b)$$

が成立する. ただし, $p: X \rightarrow \dots \rightarrow Y$ は M の因果グラフのpathを表す.

例



ACE(列, 行; a, b)

	X	Y	Z
X	0	0	$3(a - b)$
Y	$-2(a - b)$	0	$-5(a - b)$
Z	0	0	0

内容

(0)統計的因果推論とは

(1)統計的因果推論基礎

- 構造方程式モデル(SEM)
- SEMの因果グラフ
- 平均因果効果

(2)DirectLiNGAM

- LiNGAMモデル
- DirectLiNGAMによる因果探索のデモンストレーション
- 結果の信頼性評価とそのデモンストレーション

(3)流体データに適用

(2)DirectLiNGAM

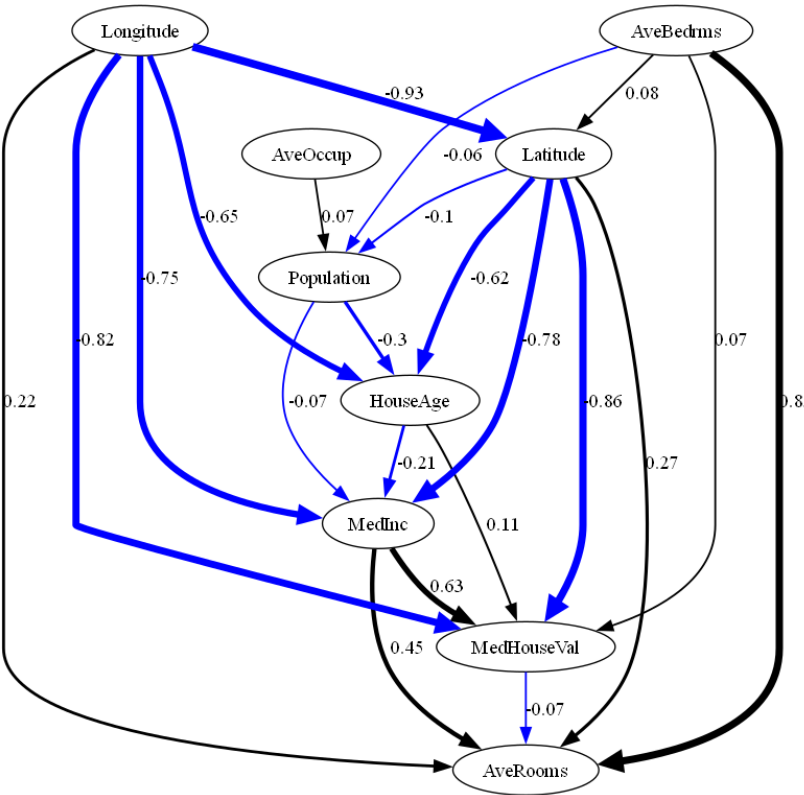
データ

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	MedHouseVal
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422
...
20635	1.5603	25.0	5.045455	1.133333	845.0	2.560606	39.48	-121.09	0.781
20636	2.5568	18.0	6.114035	1.315789	356.0	3.122807	39.49	-121.21	0.771
20637	1.7000	17.0	5.205543	1.120092	1007.0	2.325635	39.43	-121.22	0.923
20638	1.8672	18.0	5.329513	1.171920	741.0	2.123209	39.43	-121.32	0.847
20639	2.3886	16.0	5.254717	1.162264	1387.0	2.616981	39.37	-121.24	0.894



DirectLiNGAM

因果グラフ



(2)DirectLiNGAM

定義 (LiNGAMモデル)

構造方程式モデルが以下を満たすとき**LiNGAM**(Linear Non-Gaussian Acyclic Model)**モデル**という:

- ・ 線形構造方程式モデルである
- ・ 構造方程式モデルがAcyclic
- ・ 外生変数はNon-Gaussian

仮定

LiNGAMモデルは次を満たす:

- ・ 外生変数の独立性は独立(未確認の共通原因が無い).
- ・ 内生変数の分散はゼロでない.

以下LiNGAMは上の仮定を満たすとする. このとき内生変数から因果グラフが得られる=>DirectLiNGAM.

(2)DirectLiNGAM

DirectLiNGAM

LiNGAMモデルの因果グラフを内生変数だけから復元するアルゴリズムを紹介する.

Step 1. 因果グラフのsourceを特定する.

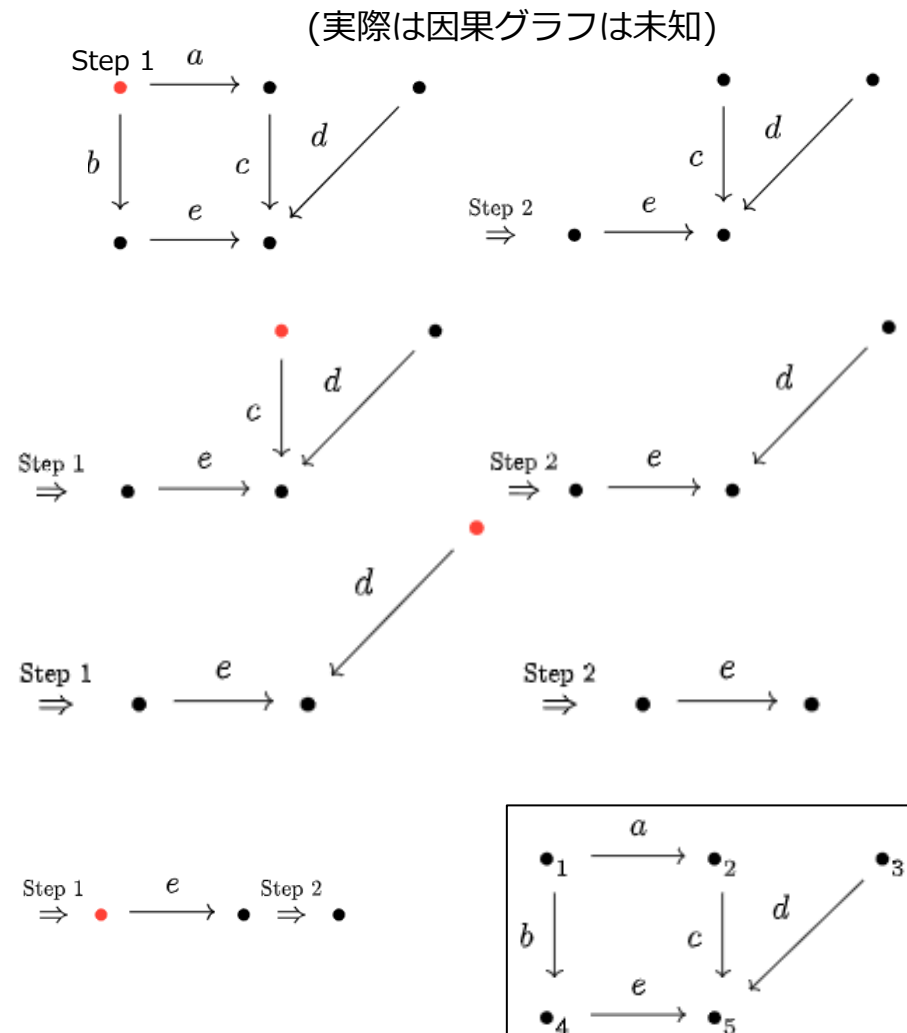
Step 2. Step1の因果グラフから特定したsourceを抜いて得られる因果グラフを持つLiNGAMモデルを構成する.

Step 3. Step 1 と Step 2 を繰り返し, 順々に特定した内生変数を $1, \dots, n$ と対応付ける. つまり**因果順序**(全単射 $K: \{\text{内生変数}\} \rightarrow \{1, \dots, n\}$)を作る.

Step 4. Step3で得られた因果順序 K に基づいて,

$$X_i = \sum_{K(X_j) < K(X_i)} b_{i,j} X_j + E_i$$

の係数を最小二乗法により復元する



(2)DirectLiNGAM

Step 1.

命題 (Shimizu-Inazumi-Hyvarinen-Kawahara-Washio-Hoyer-Bollen, Lemma 1)

X_i をLiNGAMモデルの内生変数とする. 各 X_j に対して残差 $R_i^{(j)}$ を次で定義する:

$$R_j^{(i)} := X_j - \frac{\text{Cov}(X_i, X_j)}{\text{Var}(X_i)} X_i.$$

このとき, X_i がsource $\Leftrightarrow \forall j (\neq i), X_i$ が $R_j^{(i)}$ と独立.

\Leftarrow) 非ガウス性を用いる.

- どのようにして非ガウス性が使われているのかを実感したい人はノートのExample2.1.1.4参照
- 分散がノンゼロの条件もここで使われる.
- 独立性の評価は相互情報量(mutual information)というものを用いる

(2)DirectLiNGAM

Step 2.

命題 (Shimizu-Inazumi-Hyvarien-Kawahara-Washio-Hoyer-Bollen, Lemma 2, Proof of Corollary 1)
LiNGAMモデル

$$X = B X + E$$

の内生変数 X_i が因果グラフにおいてsourceであると仮定する. このとき

$$R^{(i)} := (R_1^{(i)}, \dots, R_{i-1}^{(i)}, R_{i+1}^{(i)}, \dots, R_n^{(i)})$$

$$E^{(i)} := (E_1, \dots, E_{i-1}, E_{i+1}, \dots, E_n)$$

とおくと次のようなLiNGAMモデルが得られる:

$$R^{(i)} = B^{(i)} R^{(i)} + E^{(i)}.$$

ただし, $B^{(i)}$ は行列 B の i 行 i 列を抜いて得られる正方行列である. 特にこの新たなLiNGAMモデルの因果グラフはもともとの因果グラフの頂点 X_i を除いて得られるグラフと(重み付きグラフとして)同型.

(2)DirectLiNGAM

デモンストレーション・信頼性評価

(1)実際にDirectLiNGAMを動かしてみる

(共有ファイル: LinGAM_practice_1

<https://colab.research.google.com/drive/1I0IHHXh40tK2FpBy1sY7Sp-8uztIr-UM?usp=sharing>

<https://colab.research.google.com/drive/1d6AEKjsRzz33RdIJDUGOJ5NzlUTot6to?usp=sharing>)

(2) データがDirectLiNGAMの仮定を満たさない場合, うまく推論できないことをみる(共有ファイル: LinGAM_practice_2)

(3) データがDirectLiNGAMの仮定を満たすかの評価

(共有ファイル: LinGAM_practice_2, LinGAM_practice_1)

内容

(0)統計的因果推論とは

(1)統計的因果推論基礎

- 構造方程式モデル(SEM)
- SEMの因果グラフ
- 平均因果効果

(2)LiNGAM モデル

- Direct LiNGAM
- 因果探索のデモンストレーション
- 結果の信頼性評価とそのデモンストレーション

(3)流体データに適用

まとめ

- ・ 構造方程式モデル(SEM), 因果グラフ, 平均因果効果(ACE)を学んだ. 線形SEMの場合ACEはグラフのパスから計算できた.
- ・ DirectLiNGAMを動かせるようになった + 信頼性評価
- ・ 流体データの因果グラフを眺めた.

データ

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	MedHouseVal
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422
...
20635	1.5603	25.0	5.045455	1.133333	845.0	2.560606	39.48	-121.09	0.781
20636	2.5568	18.0	6.114035	1.315789	356.0	3.122807	39.49	-121.21	0.771
20637	1.7000	17.0	5.205543	1.120092	1007.0	2.325635	39.43	-121.22	0.923
20638	1.8672	18.0	5.329513	1.171920	741.0	2.123209	39.43	-121.32	0.847
20639	2.3886	16.0	5.254717	1.162264	1387.0	2.616981	39.37	-121.24	0.894



DirectLiNGAM

