

統計的因果推論の基礎と LiNGAM について

多田駿介

東北大学 MathCCS

shunsuke.tada.e6@tohoku.ac.jp

Abstract このノートでは統計的因果推論の基礎と LiNGAM(Linear Non-Gaussian Acyclic Model)の推定アルゴリズムの一つである Direct LiNGAM について概説する。これは Python の lingam ライブリに実装されており、それを用いた LiNGAM モデルの推定デモンストレーションも行う。さらに Direct LiNGAM で得られた結果の信頼性評価に関する手法も紹介する。本ノートの内容は主に [1] および [2]に基づいている。

Contents

1. Preliminaries	1
1.1. 因果推論のモチベーション	1
1.2. 構造方程式モデル	2
1.3. 因果グラフ	3
1.4. 平均因果効果	4
1.5. 統計的因果推論の種類	7
1.5.1. 因果関係が既知の場合	7
1.5.2. 因果関係が未知の場合	9
2. Direct LiNGAM	10
2.1. Direct LiNGAM 推定アルゴリズム	10
2.1.1. Step1 について	12
2.1.2. Step 2 について	14
2.2. デモンストレーション	14
2.3. ブートストラップ法による信頼度の評価	14
2.4. データが LiNGAM の仮定を満たすかの確認方法	14
Bibliography	16

Notation. 行列 A に対し、その転置行列を A^T とかく。ここで扱う行列は実数体上の行列である。確率変数は特に指定しない限り、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, Y に対して X の期待値を $\mathbb{E}(X)$ 、 X の分散を $\text{Var}(X)$ とし、 X, Y の共分散を $\text{Cov}(X, Y)$ で表すとする。

1. Preliminaries

この節では、まず因果推論のモチベーションや、統計的因果推論に関する基本事項として、構造方程式モデルやその因果グラフ、平均因果効果を定める。そして、統計的因果推論の手法をいくつか紹介する。

1.1. 因果推論のモチベーション

物事の因果関係や因果の強さを推測し意思決定に役立てることが因果推論の主なモチベーションである。

例えば、アイスクリームの売上と水難事故の件数には相関関係があるとする。このとき、アイスクリームが売れるに水難事故が増えるという推論や、水難事故が増えるとアイスクリームが売れるという推論は不合理であろう。にもかかわらず、誤った推論を行い、水難事故を減らすためにアイスクリームの販売を中止する決定を下せば、水難事故の件数も変わらないばかりか、アイスクリームの業者が不利益を被るおそれがある。このような誤った意思決定を避けるためにも、因果関係を正しく認識することは重要である。なお、アイスクリームの売上と水難事故の件数との間に見られる相関関係は、第三の変数である「最高気温」が影響していると考えられるだろう。

また、例えはある病気に対して、2種類の薬のどちらがより効くのかを知りたいとする。このとき、両者の治療効果を定量的に評価できれば、より効果の高い薬のみを使用することで、より多くの人命を救える可能性がある。このような観点からも、因果の強さを評価することは、より良い選択を行う上で極めて重要である。

補足. 統計的因果推論が実際の意思決定で役立った例を挙げる。

- エボラウイルス病に対する4種類の実験的治療薬について、治療が死亡率を低下させる因果効果（平均因果効果）を比較するため、因果推論の手法であるランダム化実験が実施された。その結果、治療薬 MAb114 と REGN-EB3 は ZMapp と比べて死亡率を低下させることができた[3]。
- 神奈川県葉山町のゴミ捨て場では、ゴミ収集後の後出しが問題となっていた。この改善案として、「ゴミ収集終了」の看板の設置する案が挙げられた。その改善案の効果をランダム化実験により検証した。結果、無対策の場合に比べて問題改善への効果が確認されたため、町内のゴミ捨て場に全てに看板を設置することになった(cf. <https://www.town.hayama.lg.jp/soshiki/seisaku/osirase/6870.html>).

1.2. 構造方程式モデル

因果関係や因果の効果を統計的手法を用いて記述するための枠組みを与える。

Definition 1.2.1 (構造方程式モデル): 実数値確率変数 $X_1, \dots, X_n, E_1, \dots, E_n$ が、

$$F_{X_i} : X_i = f_i(X_1, \dots, X_n, E_i) \quad (i = 1, \dots, n) \tag{1}$$

を満たすとする。このとき、 $(F_{X_i})_i$ を構造方程式モデルという。また、 F_{X_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、

$$F_{X_i} : X_i = \sum_{j=1}^n b_{i,j} X_j + E_i \quad (b_{j,i} \in \mathbb{R}) \tag{2}$$

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。

線形構造方程式モデルは行列を用いて、

$$X = BX + E \tag{3}$$

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。

Remark 1.2.2: 構造方程式モデルを定義するために内生変数や外生変数は確率変数である必要はない。しかし、後に紹介する平均因果効果を定義するためにそれらは確率変数であるとしている。

Example 1.2.3:

1. $X, Y, Z : \{1, \dots, 12\} \rightarrow \mathbb{R}$ を月ごとのアイスクリームの売上, 水難事故件数, 平均気温を表すとして,

$$F_X : X = f(Z) + E_X, \quad F_Y : Y = g(Z) + E_Y, \quad F_Z : Z = E_Z \quad (4)$$

と書けるとき, $M := (F_X, F_Y, F_Z)$ は構造方程式モデルである. ただし, E_X, E_Y, E_Z はノイズを表す外生変数とする.

2. 上の例において $f(Z) = aZ, g(Z) = bZ, a, b \in \mathbb{R}$ とすると M は線形構造方程式モデルである. 行列表示すると

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} E_X \\ E_Y \\ E_Z \end{pmatrix} \quad (5)$$

となる.

1.3. 因果グラフ

構造方程式を因果グラフで可視化することにより因果関係を直感的に捉えることができるようになる. この節では構造方程式から因果グラフを構成する方法を述べる.

M を内生変数 $(X_i)_{i=1}^n$, 外生変数 $(E_i)_{i=1}^n$ 及び構造方程式 $(F_{X_i} : X_i = f_i)_{i=1}^n$ からなる構造方程式モデルとする. 有向グラフ $G_M = (V, E)$ を

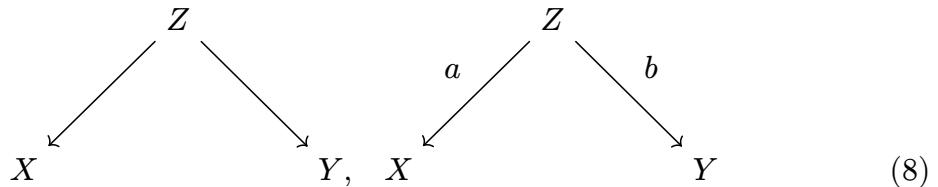
$$\begin{aligned} V &:= \{X_1, \dots, X_n\}, \\ E &:= \{(X_i, X_j) \in V^2 \mid f_j \text{ が } X_i \text{ の値に依存する}\} \end{aligned} \quad (6)$$

から定まる有向グラフとする. この有向グラフ G_M を構造方程式モデル M の因果グラフと呼ぶ. 因果グラフが acyclic なとき, つまり, 任意の頂点 X から有向辺をたどり X に戻ってこれないとき, この構造方程式モデルは acyclic であるという. M が線形構造方程式モデル

$$X = BX + E \quad (7)$$

のとき, $w : E \rightarrow \mathbb{R}$ を $w(X_i, X_j) := b_{j,i}$ で定めることで重み付き有向グラフ $G_M = (V, E, w)$ を得る. 線形構造方程式モデルの因果グラフは以下重み付き有向グラフであるとする.

Example 1.3.1: Example 1.2.3 の 1,2 の構造方程式モデルの因果グラフは



と表示できる.

Remark 1.3.2: 因果グラフの頂点に外生変数を入れることもあるが本ノートでは省略する.

1.4. 平均因果効果

因果の大きさの定量化をする方法の一つに平均因果効果がある. この章では平均因果効果を紹介する. M を内生変数 $(X_i)_{i=1}^n$, 外生変数 $(E_i)_{i=1}^n$ 及び構造方程式 $(F_{X_i} : X_i = f_i)_{i=1}^n$ からなる構造方程式モデルとする.

まずは介入と呼ばれる構造方程式モデルから新たな構造方程式モデルを作る操作を導入する. 介入の操作が well-defined となるために, 本ノートでは構造方程式モデルに acyclic 性を課す.

Definition 1.4.1 (介入): Acyclic な構造方程式モデル M と $i \in \{1, \dots, n\}$, $a \in \mathbb{R}$ に対して,

$$\begin{aligned} X_{i, X_i=a} &:= a \\ X_{j, X_i=a} &:= f_j(X_1, \dots, a, \dots, X_n, E_j) \quad (j \neq i) \end{aligned} \tag{9}$$

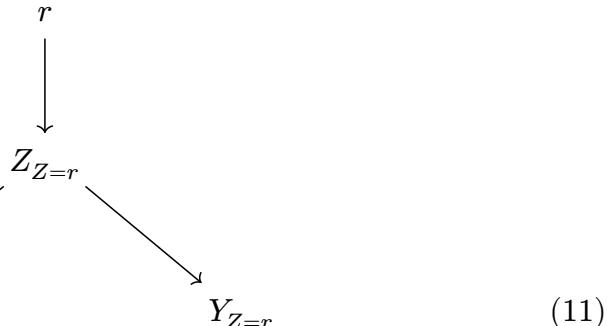
を定める(ことができる). このとき $M_{X_i=a}$ を 内生変数が

$$(a, X_{1, X_i=a}, \dots, X_{n, X_i=a}) \tag{10}$$

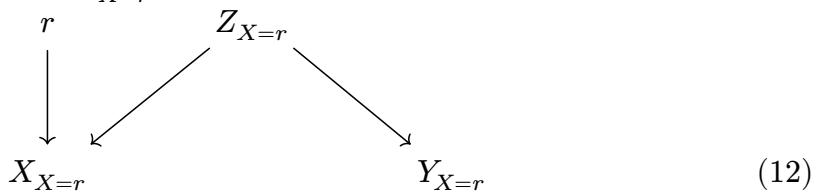
で外生変数が (E_1, \dots, E_n) , 構造方程式が Equation 9 で与えられる構造方程式モデルと定める.

構造方程式モデル M から新たな構造方程式モデル $M_{X_i=a}$ を作る操作を X_i に介入するという.

Example 1.4.2: Example 1.2.3, 1 の構造方程式モデル M に対して, $M_{Z=r}$ は因果グラフを用いて

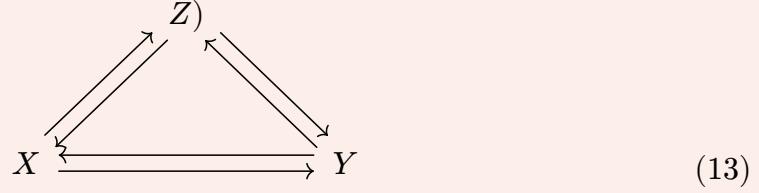


のように表示できる. また, $M_{X=r}$ は因果グラフを用いて



のように表示できる.

Remark 1.4.3: 次のような因果グラフを持つ構造方程式モデルでは介入がうまく定義できなかったので acyclic を課した(外生変数をいじればよい?).



介入を通して平均因果効果を定めることができる.

Definition 1.4.4 (平均因果効果; Average Causal Effect [4]): 確率変数 X, Y を構造方程式モデル M の内生変数とする. $a, b \in \mathbb{R}$ に対して,

$$\text{ACE}(X, Y; a, b) := \mathbb{E}[Y_{X=a}] - \mathbb{E}[Y_{X=b}] \quad (14)$$

を X から Y への平均因果効果と呼ぶ.

Example 1.4.5: Example 1.2.3 の構造方程式モデルにおいて X から Z への因果効果は $c, d \in \mathbb{R}$ として,

$$\begin{aligned} \text{ACE}(X, Z; c, d) &= \mathbb{E}[Z_{X=c}] - \mathbb{E}[Z_{X=d}] \\ &= f(c, Y) + E_3 - (f(d, Y) + E_3) \\ &= f(c, Y) - f(d, Y) \end{aligned} \quad (15)$$

と書ける. もし $f(X, Y) = aX + bY$ ならば

$$\begin{aligned} \text{ACE}(X, Z; c, d) &= f(c, Y) - f(d, Y) \\ &= (ac + bY) - (ad + bY) \\ &= a(c - d) \end{aligned} \quad (16)$$

となる.

一方で Z から X への因果効果は

$$\begin{aligned} \text{ACE}(Z, X; c, d) &= \mathbb{E}[X_{Z=c}] - \mathbb{E}[X_{Z=d}] \\ &= E_1 - E_1 \\ &= 0 \end{aligned} \quad (17)$$

となる.

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

$$p : x_0 = x \rightarrow x_1 \rightarrow \cdots \rightarrow x_{n-1} \rightarrow x_n = y \quad (18)$$

に対して,

$$w(p) := \prod_{i=0}^{n-1} w(x_i, x_{i+1}) \quad (19)$$

で定める.

線形構造方程式モデルにおいては平均因果効果は次のようにして計算できる.

Proposition 1.4.6 ([5], p.82): 線形構造方程式モデル M の因果グラフが acyclic であるとする. このとき, M の内生変数 X, Y に対して,

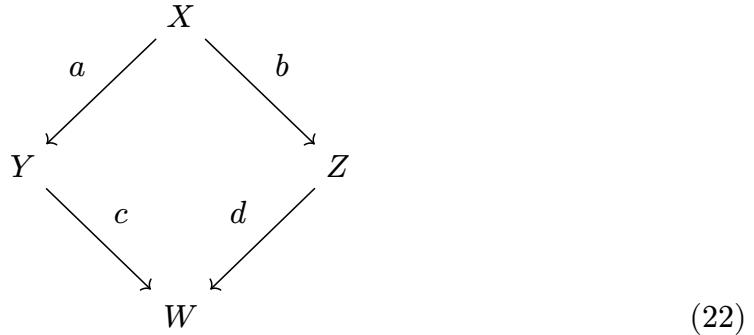
$$\text{ACE } (X, Y; a, b) = \left(\sum_{p: X \rightarrow \dots \rightarrow Y} w(p) \right) \cdot (a - b) \quad (20)$$

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.

Example 1.4.7: X, Y, Z, W を内生変数, E_X, E_Y, E_Z, E_W を外生変数,

$$\begin{aligned} X &= E_X, \\ Y &= aX + E_Y, \\ Z &= bX + E_Z, \\ W &= cY + dZ + E_W \end{aligned} \quad (21)$$

を構造方程式とする線形構造方程式モデルの X から W への平均因果効果を求める.



$$\begin{aligned} \text{ACE } (X, W; e, f) &= \mathbb{E}[W_{X=e}] - \mathbb{E}[W_{X=f}] \\ &= \mathbb{E}[cY_{X=e} + dZ_{X=e} + E_W] - \mathbb{E}[cY_{X=f} + dZ_{X=f} + E_W] \\ &= c\mathbb{E}[Y_{X=e} - Y_{X=f}] + d\mathbb{E}[Z_{X=e} - Z_{X=f}] \\ &= c(ae - af) + d(be - bf) \\ &= ac(e - f) + bd(e - f) \\ &= (ac + bd)(e - f) \end{aligned} \quad (23)$$

補足(因果の定義)

- 本ノートでは使わない概念だが次の言葉を定義しておく. X, Y をある構造方程式モデルの内生変数とする. X が Y の原因である (Y が X の結果である) とは, ある $a, b \in \mathbb{R}$ が存在して $Y_{X=a}$ と $Y_{X=b}$ の確率分布が異なることを指す.
- X, Y を内生変数とする. $P(X = a) \neq 0$ なる $a \in \mathbb{R}$ を取る (条件付き期待値を考えるために). 一般に $\mathbb{E}[Y_{X=a}]$ と条件付き期待値 $\mathbb{E}[Y|(X = a)]$ は成立するとは限らない. 実際に, 例えば, X をコイン投げの施行に対応するような確率変数 (表 $\mapsto 1$, 裏 $\mapsto 0$) とし, $X = E_X = E_Y, Y = X + E_Y$ で定める. このとき,

$$\mathbb{E}[Y_{X=0}] = 0 \neq \frac{1}{2} = \mathbb{E}[Y|X = 0]. \quad (24)$$

$\mathbb{E}[Y_{X=a}]$ と条件付き期待値 $\mathbb{E}[Y|(X = a)]$ が成立する十分条件も知られており, それはランダム化実験のところで紹介する.

1.5. 統計的因果推論の種類

統計的因果推論では、因果関係を既知としてデータから因果の大きさを定量的に評価することや、因果関係を未知としてデータからその因果関係を推測することが主眼になる。特に後者を統計的因果探索という。後で紹介する Direct LiNGAM は未知の因果関係および因果効果の推定の両方を行う。

1.5.1. 因果関係が既知の場合

因果関係を既知としたとき、因果効果の推定手法としては、例えば次のものが挙げられる。

- ランダム化実験 (Randomized Controlled Trial, RCT)
- 操作変数法 (Method of Instrumental Variables, IV)
- 回帰不連続デザイン (Regression Discontinuity Design, RDD)
- 傾向スコアマッチング(Propensity Score Matching, PSM)
- 差分の差分法(Difference-in-Differences, DiD)

などが挙げられる[6]。

上の手法の中でも RCT は確率変数 X から Y への平均因果効果を推定する手法の golden standard [5]といわれる。これは変数 X の値を(コイン投げなどで)ランダムに割り当てるができるときに行われる手法である。この手法は信頼性が高い一方で倫理的・実務的な制約からランダムに値を割り当てることが困難な場面も多い。例えば、喫煙の健康影響を評価するために喫煙の有無を無作為に割り当てるることは倫理上許されないし、他にも学習時間の成績への影響を評価するために学習時間をランダムに割り当てることも現実的ではない。こうした場合には、IV, RDD, 傾向スコア法, DiD などの手法を用い、それぞれの仮定の下で因果効果を推定する。それぞれの手法の関係性やデータに課す仮定に関しては[6]を見よ。

以下ランダム化実験を紹介する。他の手法については [6]を見てください。

ランダム化実験.ここでは Corollary 1.5.1.4 を示し、ランダム化実験により平均因果効果を推定できることを具体例 Example 1.5.1.5 を通して確認する。

Corollary 1.5.1.4 を示すための補題を準備する。

Lemma 1.5.1.1: 独立な実数値確率変数 X, Y と任意の可測関数 $f, g : \mathbb{R} \rightarrow \mathbb{R}$ に対して、 $f \circ X, g \circ Y$ は独立である。

Proof: 省略。 □

Lemma 1.5.1.2: 独立な実数値確率変数 X, Y と $P(X = a) \neq 0$ を満たす任意の $a \in \mathbb{R}$ に対して、条件付き期待値 $\mathbb{E}[Y|X = a]$ は $\mathbb{E}[Y]$ に等しい。

Proof: 省略。 □

Proposition 1.5.1.3: 実数値確率変数 X, Y を独立とする。 $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ を可測関数とする。このとき $P(X = a) \neq 0$ を満たす任意の $a \in \mathbb{R}$ に対して $\mathbb{E}[f(a, Y)] = \mathbb{E}[f(X, Y) | X = a]$ が成立する。

Proof: 以下の計算から従う.

$$\begin{aligned}
 \mathbb{E}[f(a, Y)] &= \mathbb{E}[f(a, Y)|X = a] \\
 &= \lim_{n \rightarrow \infty} \sum_{i=0}^{n2^n} \frac{P\left(\left[\frac{i}{2^n} \leq f(a, Y) < \frac{i+1}{2^n}\right] \cap (X = a)\right)}{P(X = a)} \\
 &= \lim_{n \rightarrow \infty} \sum_{i=0}^{n2^n} \frac{P\left(\left[\frac{i}{2^n} \leq f(X, Y) < \frac{i+1}{2^n}\right] \cap (X = a)\right)}{P(X = a)} \\
 &= \mathbb{E}[f(X, Y)|X = a].
 \end{aligned} \tag{25}$$

ただし, 一つ目の等号は Lemma 1.5.1.1, Lemma 1.5.1.2 を用いた. \square

上の命題からただちに次が従う.

Corollary 1.5.1.4: X, Y を内生変数とし, E_X, E_Y を外生変数, $X = f(E_X), Y = g(X, E_Y)$ を構造方程式とする 構造方程式モデルを考える. 外生変数 E_X, E_Y が独立のとき, $P(X = a) \neq 0$ を満たす任意の $a \in \mathbb{R}$ に対して

$$\mathbb{E}[Y_{X=a}] = \mathbb{E}[Y|X = a] \quad (a \in \mathbb{R}) \tag{26}$$

が成立する. 特に

$$\text{ACE}(X, Y; a, b) = \mathbb{E}[Y|X = a] - \mathbb{E}[Y|X = b] \quad (a, b \in \mathbb{R}) \tag{27}$$

が成立する.

Proof: Proposition 1.5.1.3 から従う. \square

上の主張の要点は, 外生変数の独立性の仮定の下では, X から Y への平均因果効果は条件付き確率で表せる点にある. 特に, 条件付き期待値はデータから推定できるため, 平均因果効果も推定可能である. ランダム化実験では, この独立性の仮定を満たすために, コイン投げなどによる X の値の無作為に割り当てる.

Example 1.5.1.5: $\Omega := \{a_1, \dots, a_n\}$ をある病気を患っている人の集合とし, 一様確率測度を入れた確率空間とみなす. いま, ある薬がこの病気にどの程度効果があるのかを知りたいとする. その効果を平均因果効果によって定量化する.

まず内生変数 X, Y と外生変数 E_X, E_Y を次のように定める:

- $E_X : \Omega \rightarrow \{0, 1\}$ は a_i に薬を与えるならば 1 を, 与えないならば 0 をとる関数とする. これはコイン投げでランダムに決めるとする.
 - $X := E_X$.
 - $E_Y : \Omega \rightarrow \mathbb{R}$ はノイズを表す外生変数とする.
 - $Y =: \Omega \rightarrow \{0, 1\}$ は a_i が完治すれば 1, そうでないなら 0 をとる関数とする. このとき, E_X, E_Y は独立であると仮定する (コイン投げの結果と).
- 上記の関数は次のような構造方程式で関連付けられているとする:
- $X = E_X, Y = g(X, E_Y)$.

$$\begin{array}{ccc} X & \xleftarrow{\hspace{1cm}} & E_X \\ \downarrow & & \\ Y & \xleftarrow{\hspace{1cm}} & E_Y \end{array} \quad (28)$$

このとき標本サイズ n が十分大きければ

$$\begin{aligned} \mathbb{E}[Y|X=1] &\approx \frac{\#\{\text{治った人で薬を服用した人}\}}{\#\{\text{薬を服用した人}\}} \\ \mathbb{E}[Y|X=0] &\approx \frac{\#\{\text{治った人で薬を服用しなかった人}\}}{\#\{\text{薬を服用しなかった人}\}} \end{aligned} \quad (29)$$

と推定でき, さらに Corollary 1.5.1.4 より

$$\text{ACE}(X, Y; 1, 0) = \mathbb{E}[Y|X=1] - \mathbb{E}[Y|X=0] \quad (30)$$

がなり立つので, $\text{ACE}(X, Y; 1, 0)$ はデータから推定できる.

1.5.2. 因果関係が未知の場合

因果関係が未知の状態からその因果グラフを求める手法は構造方程式と外生変数に課す仮定で 3 つに大別される [1] :

1. ノンパラメトリック・アプローチ : 構造方程式と外生変数に条件を課さない.
2. パラメトリック・アプローチ (線形ガウス) : 構造方程式は線形でかつ, 外生変数の分布はガウス分布に従うと仮定する.
3. セミパラメトリック・アプローチ (線形・非ガウス) : 構造方程式は線形でかつ, 外生変数の分布はガウス分布と仮定する.

ノンパラメトリック・アプローチやパラメトリック・アプローチでは, 一般にデータのみから因果グラフを一意には推定できない. この手法ではさらに因果マルコフ性・忠実性・などの仮定を課すことでの, マルコフ同値類まで因果グラフを復元できる (cf. [1]). 一方 LiNGAM に代表されるセミパラメトリック・アプローチの枠組みでは, 適切な仮定の下で因果グラフを一意に推定できる場合がある.

Memo ノンパラメトリック・アプローチの代表的な手法として PC アルゴリズム (Peter-Clark アルゴリズム)[7] が知られている. (なお出力は変数の入力の順番に依存

する。もしも因果探索でノンパラメトリック・アプローチを取るなら順序に依存しない PC-stable アルゴリズム[8]がよさそう。)

2. Direct LiNGAM

ここでは LiNGAM(Linear Non-Gaussian Acyclic Model)モデルを推定する手法である Direct LiNGAM を紹介する。そのために必要な用語を確認しておく。確率変数 X が正規分布(normal distribution)に従うまたは Gaussian とは、ある $\mu \in \mathbb{R}^2, \sigma^2 > 0$ が存在して

$$P(X = x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (x \in \mathbb{R}) \quad (31)$$

となることをいう。ただし $(X = x) := \{\omega \in \Omega \mid X(\omega) = x\}$ 。これに対し、この条件を満たさないとき X は非ガウス (non-Gaussian) であるという。有向グラフに対して、ある頂点が source であるとはどの頂点からも辺が入ってこないことをいう。

Definition 2.1 (LiNGAM): 内生変数 $(X_i)_{i=1}^n$, 外生変数 $(E_i)_{i=1}^n$ 及び構造方程式 $(F_{X_i} : X_i = f_i)_{i=1}^n$ からなる構造方程式モデルが以下を満たすとき LiNGAM モデルという:

- 線形構造方程式モデルである。つまり、実行列 B が存在して

$$X = BX + E \quad (32)$$

が成立する。ただし、 $X := (X_1, \dots, X_n)^T$ で $E := (E_1, \dots, E_n)^T$ とする。

- 上の実行列 B はべき零行列 ($\exists m \in \mathbb{N}$ s.t. $B^m = 0$)。
- E_1, \dots, E_n は Non-Gaussian。

Remark 2.2: 上の定義において二つ目の条件は因果グラフが acyclic であることと同値である。

以下さらに次を仮定する:

Assumption 2.3: LiNGAM は次を満たすと仮定する :

- E_1, \dots, E_n は独立(未確認の共通原因が無い)。
- $X_i (i = 1, \dots, n)$ の分散はゼロでない。

以上の仮定のもと、観測データが LiNGAM モデルに従うとき、行列 B が推定できることが知られている。推定方法は ICA(独立成分分析)を用いたもの[9]や、回帰分析と独立の評価を用いたものがある[10] (cf. [1])。以下では後者の方法とその計算アルゴリズムである Direct LiNGAM を紹介する。

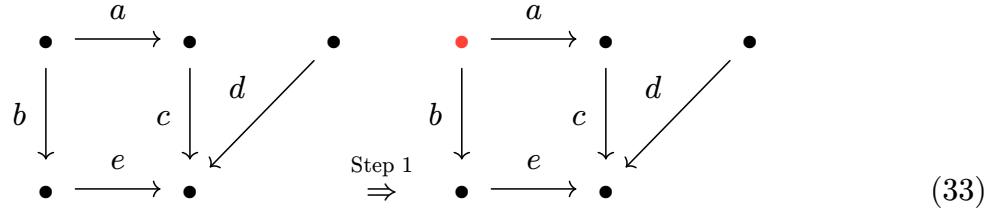
2.1. Direct LiNGAM 推定アルゴリズム

この節の主な内容は[10]による。この節では LiNGAM モデル $X = BX + E$ が Assumption 2.3 を満たすとき、内生変数から行列 B (つまり因果グラフ) を一意的に復元する Direct LiNGAM を紹介する。なおこのアルゴリズムは python で実装されて

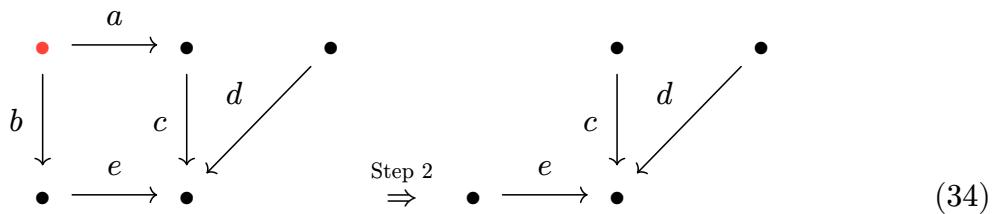
いて誰でも簡単に使えるようになっている.(<https://lingam.readthedocs.io/en/latest/tutorial/lingam.html>)

推定アルゴリズムは次の4つのステップからなる.

Step 1. 因果グラフの source を特定 (Proposition 2.1.1.2).



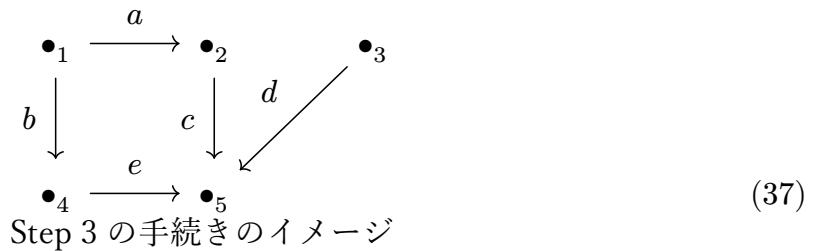
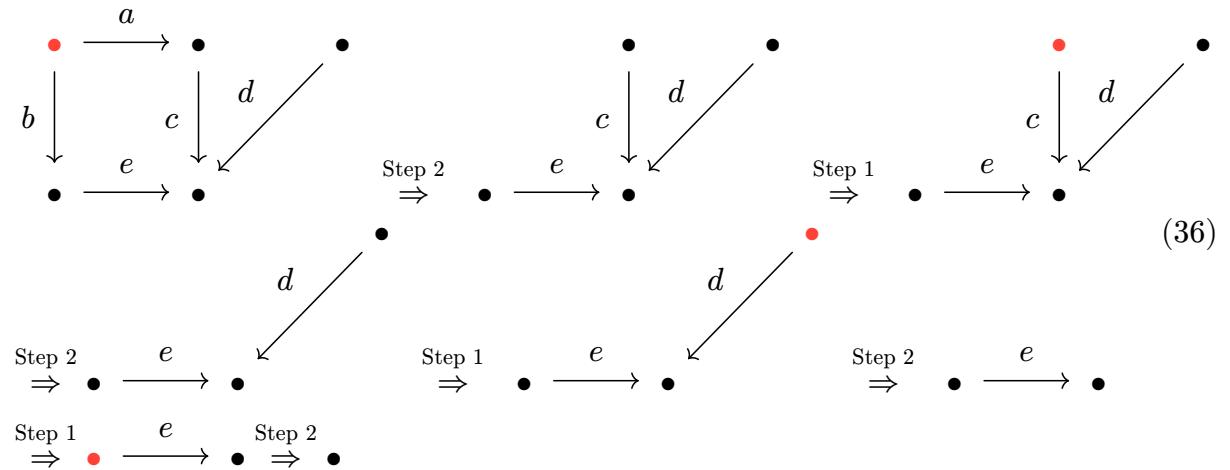
Step 2. 新たな構造方程式モデルを構成する. ただし, その因果グラフは Step1 で考えていた因果グラフから Step1 で特定した頂点を抜いて得られた重み付き有向グラフである (Proposition 2.1.2.1)



Step 3. Step 1 と Step 2 を繰り返し, 順々に特定した頂点を $1, \dots, n$ と対応付けることで, 最初に与えられた LiNGAM モデルの因果順序 K が得られる. ただし因果順序とは, 全単射

$$L : \{X_1, \dots, X_n\} \xrightarrow{1:1} \{1, \dots, n\} \quad (35)$$

で因果グラフの任意の有向辺 $Y \rightarrow Z$ に対して $L(Y) < L(Z)$ を満たすものをいう.



Step 4. Step3 で得られた因果順序 K に基づいて,

$$X_i = \sum_{K(X_j) > K(X_i)} b_{j,i} X_j + E_i \quad (38)$$

の係数を最小二乗法を用いて B を復元する.

以下, 非自明な Step である Step1,2 のみを見る.

2.1.1. Step1 について

まず次の定理を紹介する.

Theorem 2.1.1.1 (Darmois-Skitovitch の定理): 独立な確率変数 E_1, \dots, E_n に対して, それらの線形結合として

$$Y_1 = \sum_{i=1}^n \alpha_i E_i, \quad Y_2 = \sum_{i=1}^n \beta_i E_i, \quad (39)$$

と定義される 2 つの確率変数 Y_1, Y_2 を考える. このとき, もし Y_1, Y_2 が独立であるならば, $\alpha_j \neq 0 \neq \beta_j$ となるとき E_j は正規分布に従う.

この定理を用いて次の主張を示すことができる (Non-Gaussian の条件が本質的に使われる部分).

Proposition 2.1.1.2 ([10], Lemma 1): X_i を LiNGAM モデルの内生変数とする. 内生変数 X_j を X_i に対して線形回帰したときの残差 $R_j^{(i)}$ を次で定義する:

$$R_j^{(i)} := X_j - \frac{\text{Cov}(X_i, X_j)}{\text{Var}(X_i)} X_i. \quad (40)$$

このとき, 変数 X_i が因果グラフにおいて source であることと, X_i がすべての $j (\neq i)$ に対して $R_j^{(i)}$ と独立であることとは同値である.

上の主張から残差と確立変数の独立性を確認することで LiNGAM モデルの因果グラフの source が特定できる.

Remark 2.1.1.3 ([10]): 上の主張において

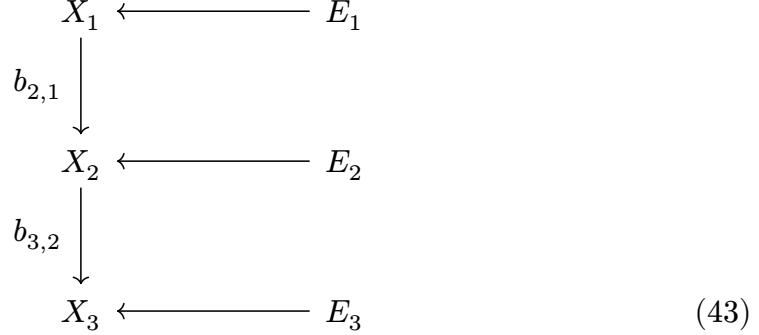
$$T(X_i) := \sum_{j(\neq i)} MI\left(X_i, R_i^{(j)}\right) \quad (41)$$

で定める. ただし, $MI\left(X_i, R_i^{(j)}\right)$ は相互情報量 (mutual information) とする (相互情報量は独立性を測る指標である. 詳しくは [10] を見てください). このとき, X_j が因果グラフの source であることは

$$j = \arg \min_{i \in \{1, \dots, n\}} T(X_i) \quad (42)$$

となることで判断される.

Example 2.1.1.4: 次のような因果グラフを持つ LiNGAM モデルを考える(外生変数も表示している).



R_1^2, R_1^3 を求め、それぞれ X_1 と独立であることを確認し、Proposition 2.1.1.2 を適用することで X_1 が因果グラフの source になっていることを示す。まず

$$\begin{aligned}
 \text{Cov}(X_1, X_2) &= \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1] \mathbb{E}[X_2] \\
 &= \mathbb{E}[X_1(b_{2,1}X_1 + E_2)] - \mathbb{E}[X_1] \mathbb{E}[b_{2,1}X_1 + E_2] \\
 &= b_{2,1}(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) \\
 &= b_{2,1} \text{Var}(X_1)
 \end{aligned}$$

$$\begin{aligned}
 \text{Cov}(X_1, X_3) &= \mathbb{E}[X_1 X_3] - \mathbb{E}[X_1] \mathbb{E}[X_3] \tag{44} \\
 &= \mathbb{E}[X_1(b_{3,2}X_2 + E_3)] - \mathbb{E}[X_1] \mathbb{E}[b_{3,2}X_2 + E_3] \\
 &= \mathbb{E}[X_1(b_{3,2}(b_{2,1}X_1 + E_2) + E_3)] - \mathbb{E}[X_1] \mathbb{E}[b_{3,2}(b_{2,1}X_1 + E_2) + E_3] \\
 &= b_{3,2}b_{2,1}(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) \\
 &= b_{3,2}b_{2,1} \text{Var}(X_1)
 \end{aligned}$$

より

$$\begin{aligned}
 R_1^{(2)} &= X_2 - \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)} X_1 = X_2 - b_{2,1}X_1 = E_2 \\
 R_1^{(3)} &= X_3 - \frac{\text{Cov}(X_1, X_3)}{\text{Var}(X_1)} X_1 = X_3 - b_{3,2}b_{2,1}X_1 = b_{3,2}E_2 + E_3
 \end{aligned} \tag{45}$$

を得る。それぞれ $X_1 = E_1$ と独立なので因果グラフの source になる。

一方で、

$$\begin{aligned}
 R_2^{(1)} &= X_1 - \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_2)} X_2 \\
 &= X_1 - \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_2)} (b_{2,1}X_1 + E_2) \\
 &= \left(1 - \frac{b_{2,1}^2 \text{Var}(X_1)}{\text{Var}(X_2)}\right) E_1 + \frac{b_{2,1} \text{Var}(X_1)}{\text{Var}(X_2)} E_2 \\
 X_2 &= b_{2,1}E_1 + E_2
 \end{aligned} \tag{46}$$

は独立でない(もしも独立なら Darmois-Skitovitch の定理より E_2 はガウス分布に従うことになるが LiNGAM の仮定に反する)。よって X_2 は因果グラフの source にはならない。 X_3 についても同様なことが言える。

2.1.2. Step 2 について

次の主張は, 与えられた LiNGAM モデルから, その因果グラフの source を抜いて得られる重み付き有向グラフと同型になるような因果グラフをもつ LiNGAM モデルを新たに構成できることを主張する.

Proposition 2.1.2.1 ([10], Lemma 2, Proof of Corollary 1): 内生変数 X_j が因果グラフにおいて source であると仮定する. このとき, $X_i (i \neq j)$ を X_j に対して回帰したときの残差に対して, $R^{(j)} := (R_1^{(j)}, \dots, R_{j-1}^{(j)}, R_{j+1}^{(j)}, \dots, R_n^{(j)})$ とする. また, $E^{(j)} := (E_1, \dots, E_{j-1}, E_{j+1}, \dots, E_n)$ とする. するとある $n - 1$ 次の行列が一意的の存在して, 次のような LiNGAM モデルが得られる:

$$R^{(j)} = B^{(j)} R^{(j)} + E^{(j)}. \quad (47)$$

ただし, $B^{(j)}$ は行列 B の j 行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_j を除いて得られるグラフと(重み付きグラフとして)同型.

Example 2.1.2.2: Example 2.1.1.4 の例において

$$\begin{aligned} R_1^{(2)} &= E_2 \\ R_1^{(3)} &= b_{3,2} E_2 + E_3 \end{aligned} \quad (48)$$

は LiNGAM の仮定を満たし, その因果グラフは

$$\begin{array}{ccc} R_1^{(2)} & \xleftarrow{\hspace{1cm}} & E_2 \\ \downarrow b_{3,2} & & \\ R_1^{(3)} & \xleftarrow{\hspace{1cm}} & E_3 \end{array} \quad (49)$$

であり, 元々の因果グラフから X_1 を除いたものと同型.

2.2. デモンストレーション

Direct LiNGAM による因果推論をデモンストレーションする. ここから試せる.
<https://colab.research.google.com/drive/1I0IHHXh40tK2FpBy1sY7Sp-8uztIr-UM?usp=sharing>

2.3. ブートストラップ法による信頼度の評価

ここでは, lingam ライブリに実装されているブートストラップ法を実際に試す. ここから試せる. <https://colab.research.google.com/drive/1d6AEKjsRzz33RdlJDUGOJ5NzlUTot6to?usp=sharing>

2.4. データが LiNGAM の仮定を満たすかの確認方法

Direct LiNGAM を適用する上でデータは以下の仮定を満たさなければならなかつた:

1. 線形性
2. 外生変数の分布が 非ガウス分布 である
3. 因果グラフが acyclic である

4. 外生変数は独立である

5. 内生変数の分散が 0 でない

データが以上の仮定を満たすかを評価することは出力された因果グラフの信頼できるかを知る上で重要である。(データが仮定を満たさない場合、因果推論がうまくいかないので試してみてください <https://colab.research.google.com/drive/1d6AEKjsRzz33RdlJDUGOJ5NzlUTot6to?usp=sharing>) 以下では、それぞれの仮定を評価するための方法を紹介する。なお、それらのうちの一部の評価方法は上の共有ファイルで実際に試すことができる。

1. 線形性の評価

線形性の確認する方法としては

- 各組 (X_i, X_j) を(散布図などを用いて)可視化.

• Todo: add examples

2. 非ガウス性の評価

非ガウス性の評価には例えば、次のようなものがある[6].

- ヒストグラムによる可視化
- 尖度と歪度の計算(ガウス分布の尖度と歪度はそれぞれ 3,0 なのでこの値と離れているかで評価する)
- Shapiro-Wilk 検定
- コルモゴロフ・スマルノフ検定

などがある。外生変数の推定値をこれらで評価することで非ガウス性の評価ができる。

3. acyclic 性の確認

データから構造方程式の acyclic 性を評価する指標があるのか不明でした。もしも線形性や非ガウス性を満たしていないのにおかしな因果グラフが出力されたときに acyclic 性を疑う？

- Todo: find methods

4. 外生変数の独立性の評価

独立性の評価の手法は以下のものがある。

- Hilbert-Schnidt 独立性基準 (HSIC)

なお lingam ライブラリには、HSIC に基づく外生変数の独立性を確認する機能がある。

- Todo: add examples

5. 分散の値がノンゼロ

これは直接計算すればよい。

データから構造方程式の acyclic 性を評価する指標があるのか不明でした。もしも線形性や非ガウス性を満たしていないのにおかしな因果グラフが出力されたときに acyclic 性を疑う？

- Todo: find methods

Bibliography

- [1] 清水昌平 and 講談社サイエンティフィク, 統計的因果探索 = Causal discovery, no. . in MLP 機械学習プロフェッショナルシリーズ. 講談社, 2017. [Online]. Available: <https://ci.nii.ac.jp/ncid/BB23682920>
- [2] N. コミュニケーションズ イノベーションセンター, “ごちきか,” [Online]. Available: https://gochikika.ntt.com/Modeling/causal_VARLiNGAM.html
- [3] S. Mulangu et al., “A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics,” *The New England Journal of Medicine*, vol. 381, 2019, doi: 10.1056/NEJMoa1910993.
- [4] D. B. Rubin, “Estimating causal effects of treatments in randomized and nonrandomized studies.,” *Journal of educational Psychology*, vol. 66, no. 5, p. 688, 1974.
- [5] J. Pearl, M. Glymour, and N. Jewell, *Causal Inference in Statistics: A Primer*. Wiley, 2016.
- [6] 金本拓, 因果推論: 基礎から機械学習・時系列解析・因果探索を用いた意思決定のアプローチ. オーム社, 2024.
- [7] P. Spirtes and C. Glymour, “An algorithm for fast recovery of sparse causal graphs,” *Social science computer review*, vol. 9, no. 1, pp. 62–72, 1991.
- [8] D. Colombo and M. H. Maathuis, “Order-independent constraint-based causal structure learning.” [Online]. Available: <https://arxiv.org/abs/1211.3295>
- [9] S. Shimizu, P. O. Hoyer, A. Hyvärinen;inen, and A. Kerminen, “A Linear Non-Gaussian Acyclic Model for Causal Discovery,” *Journal of Machine Learning Research*, vol. 7, no. 72, pp. 2003–2030, 2006, [Online]. Available: <http://jmlr.org/papers/v7/shimizu06a.html>
- [10] S. Shimizu et al., “DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model.” [Online]. Available: <https://arxiv.org/abs/1101.2489>