

## 1. イントロダクション

例：「抑うつ気分」と「睡眠障害」には強い相関がある。

このとき因果はどちらの方向？あるいは因果なし？

→ この問題に答えるために最も効果的なのが**ランダム化実験**である。

つまり睡眠障害の度合いの大小でランダムに集団を形成しデータを比較する。

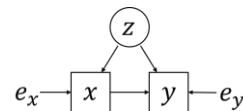
しかし、この例の状況においては倫理的にも技術的にもランダム化実験が難しい。

以降ではランダム化実験が行えない場合の次善策を提案する。

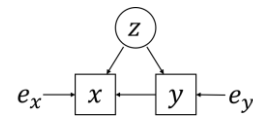
## 2. 因果探索の基本問題

$x, y$  を観測変数、 $z$  を未観測な交絡変数、 $e_x, e_y$  を誤差変数として  $x$  と  $y$  に関する次の 3 つのモデルをデータ  $(x, y)$  から識別する問題を因果探索の基本問題とする。[2]

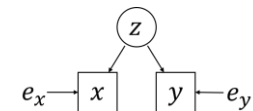
$$\text{モデル 1} \quad \begin{cases} x = f_x(z, e_x) \\ y = f_y(x, z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$



$$\text{モデル 2} \quad \begin{cases} x = f_x(y, z, e_x) \\ y = f_y(z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$



$$\text{モデル 3} \quad \begin{cases} x = f_x(z, e_x) \\ y = f_y(z, e_y) \\ p(z, e_x, e_y) = p(z)p(e_x)p(e_y) \end{cases}$$



この構造的因果モデルによって次の 2 つの性質が表現されている。

- ・ 自律性（構造方程式のどれに介入してもそれ以外の関数形や分布は変わらない）
  - $x$  に介入（ $x = c$ ）しても  $f_y$  や  $p(z, e_x, e_y)$  は変わらない
- ・ 内生変数の関数形（ $f_x, f_y$ ）と外生変数の分布（ $p(z, e_x, e_y)$ ）
  - 両者が決まれば内生変数の分布が決まる

統計的因果探索ではどのような仮定を置けば因果グラフを推測できるかを問う。

### 3. 因果探索の基本問題への 3 つのアプローチ

#### ① ノンパラメトリック・アプローチ [2, 3]

関数形：仮定なし                      外生変数の分布：仮定なし

(ただし因果関係は一方向であり双方向ではない)

→ 出来る限り仮定を置かずにデータのみからどの程度因果を推測できるか、を明らかにする

#### ② パラメトリック・アプローチ

関数形：線形                      外生変数の分布：ガウス分布

→ 単純だが、それでも**識別性**がないため因果を推定できない (次章で説明)

#### ③ セミパラメトリック・アプローチ [4]

関数形：線形                      外生変数の分布：非ガウス (連続) 分布

**線形性非ガウス非巡回モデル** (Linear Non-Gaussian Acyclic Model, **LiNGAM**) アプローチとも言う。

→ この仮定の下では因果を推定できる

ガウス分布：平均と分散共分散行列が定まれば分布形は一意に定まる

非ガウス分布：平均と分散以外にも尖度や歪度などのパラメータを持つ

→ 多くの情報量を持つ

### 4. 3 つのアプローチと識別可能性

これまでは未観測な交絡変数 $z$ の存在を仮定していたが、一般化するために

「未観測な交絡変数は存在せず (i.e. 因果に関するすべての変数が観測されている)、変数間のすべての因果関係が未知である」

とすると、各変数 $x_i (i = 1, \dots, p)$ の生成過程は次の式で表される。

$$x_i = f_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p, e_i)$$

ここから因果関係の**非巡回性**・誤差変数の独立性の仮定を追加すると次のようになる。

$$x_i = f_i(pa(x_i), e_i)$$

ただし $pa(x_i)$ は因果グラフにおいて変数 $x_i$ の**親**にあたる変数の集合を表す。

さらに議論を単純化するために関数形を線形に、誤差関数を連続変数とすると

$$x_i = \sum_{x_j \in pa(x_i)} b_{ij}x_j + e_i$$

となる。

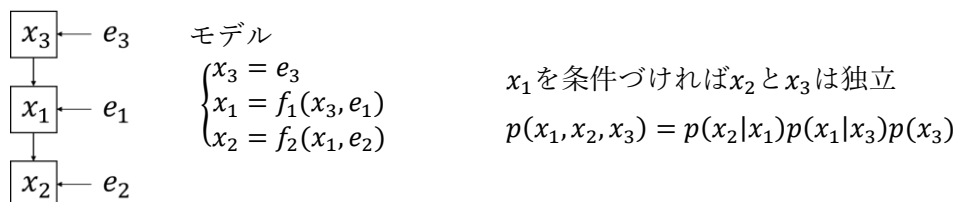
ここからはこの式で生成された変数を例にとり、先に挙げた 3 つのアプローチがどのように因果構造を推定するか見ていく。

# ① ノンパラメトリック・アプローチと識別性

このアプローチでは因果グラフを推測するために2つの仮定を置く

## ・ 因果的マルコフ条件

各変数 $x_i$ はその親で条件づけると非子孫と独立となる



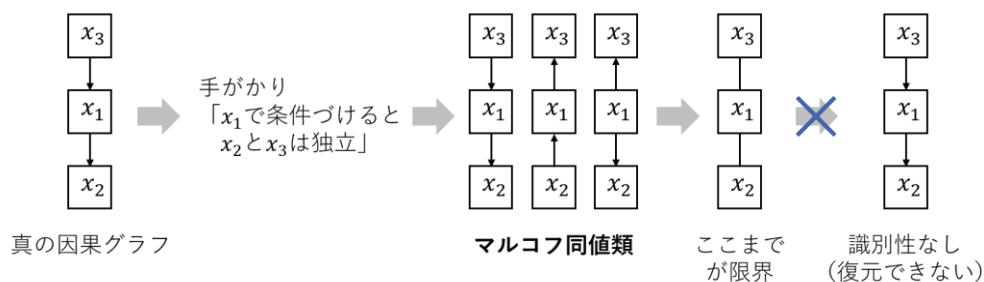
## ・ 忠実性

変数間に成り立つ条件付独立性は因果的マルコフ条件からのみ導かれる

$$\begin{aligned} \begin{cases} x = e_x \\ y = -x + e_y \\ z = x + y + e_z \end{cases} & \quad \text{Cov}(x, z) = E(xz) - E(x)E(z) \\ & = E(e_x(e_y + e_z)) - E(e_x)E(e_y + e_z) \\ & = 0 \end{aligned}$$

因果グラフに従えば $x$ と $z$ は従属であるはずだが、共分散を計算すると独立となる。これはガウス分布において独立と無相関が同値であることに起因する。

これらの仮定を用いて例えば次のように因果グラフを推測する。



このように関数形と分布の仮定を使わずに出来る限り因果構造を推測する。

→ 因果グラフを一意に推測できる場合もあれば、出来ない場合(上の例)もある。

ノンパラメトリック・アプローチは大きく2つの推測方法に分けられる

## 1. 制約に基づくアプローチ [5, 6]

仮説検定による条件付独立性を基にそれを満たす因果グラフを探索する。

## 2. スコアに基づくアプローチ

各マルコフ同値類を BIC などのスコアによって比較する。

② パラメトリック・アプローチと識別性

因果的マルコフ条件は関数形と誤差変数の分布に仮定を置かずとも使える原理であったが、それだけでは因果グラフを識別できなかった。

そこで関数形に**線形性**を、誤差変数の分布に**ガウス分布**の仮定を置く。

以下では例として次の2つの構造方程式モデルを考える。

$$\begin{array}{ll} \text{モデル 1} & \begin{cases} x_1 = e_1 \\ x_2 = 0.8x_1 + e_2 \end{cases} & \text{モデル 2} & \begin{cases} x_1 = 0.8x_2 + e_1 \\ x_2 = e_2 \end{cases} \end{array}$$

ただしモデル1においては、 $E(e_1) = E(e_2) = 0, V(e_1) = 1, V(e_2) = 0.6^2, e_1 \perp e_2$

モデル2においては、 $E(e_1) = E(e_2) = 0, V(e_1) = 0.6^2, V(e_2) = 1, e_1 \perp e_2$ とする。

まずノンパラメトリック・アプローチを考える。

このとき $x_1$ と $x_2$ の共分散を計算すると、両モデルとも $Cov(x_1, x_2) = 0.8$ となる。

因果グラフが異なるにもかかわらず成り立つ条件付独立性が同じであるため、ノンパラメトリック・アプローチでは因果グラフを識別できない。

そこで、関数形が線形であり、誤差変数がガウス分布に従うとする。

しかし、ガウス分布の**再生性**の性質（ガウス分布に従う変数の線形和で表される変数もまたガウス分布に従う）により、

$$x_1, x_2 \sim N(0, 1), \quad Cov(x_1, x_2) = 0.8$$

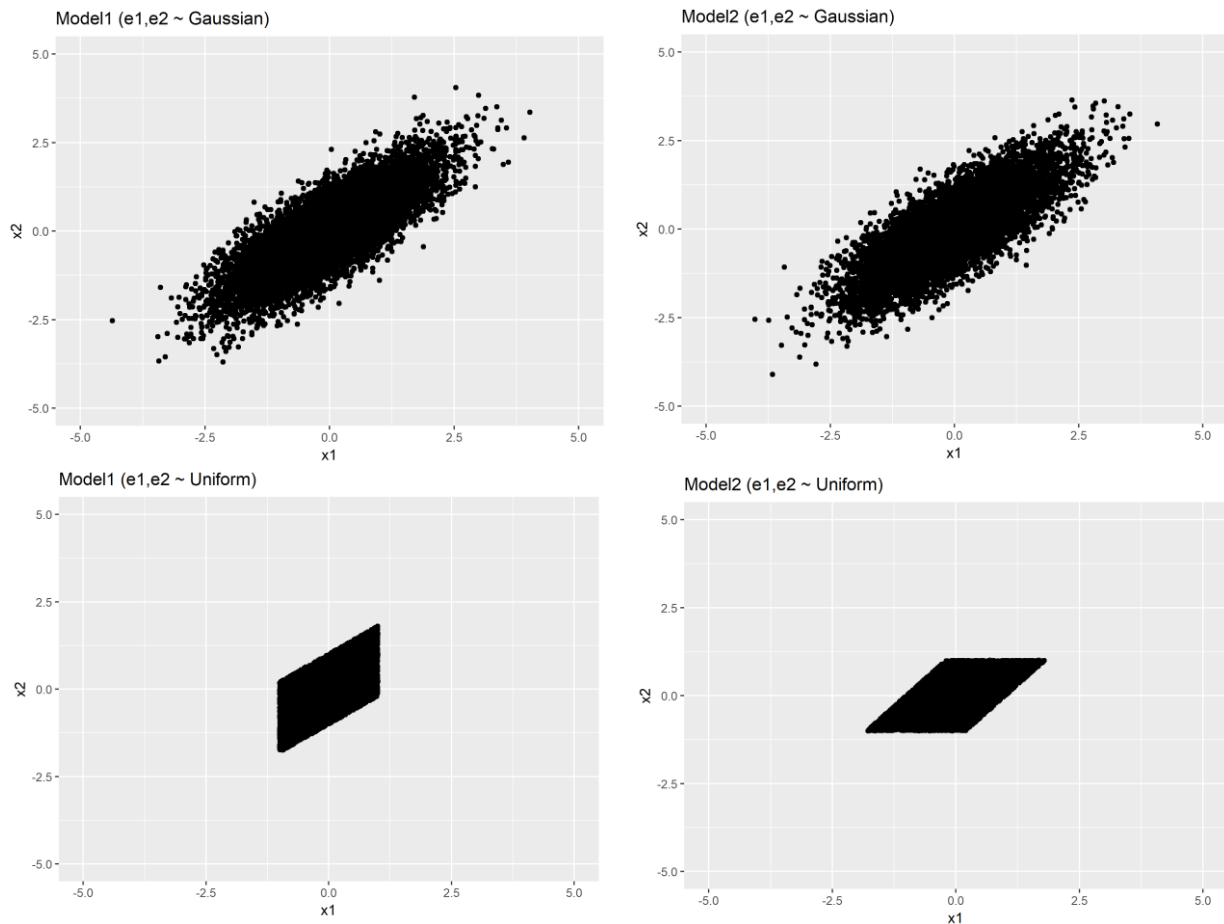
なので、観測変数 $x_1, x_2$ の分布 $p(x_1, x_2)$ が両モデルで同じガウス分布となってしまう。つまり、パラメトリック・アプローチでも依然識別できない因果グラフがあることが分かる。

③ セミパラメトリック・アプローチと識別性

関数形に線形性の仮定を置き、誤差変数の分布は非ガウス分布であればよいとする LiNGAM を考える。

誤差変数 $e_1, e_2$ にガウス分布の仮定を置くと $p(x_1, x_2)$ は同じ分布となってしまうが、非ガウス分布に従う場合、両モデルで $p(x_1, x_2)$ は異なる分布となることが証明されている。[2]

ここでは以下の図を用いて異なる分布となることを確認するにとどめる（LiNGAMの詳細は次回）。



#### 参考文献

- [1] 清水昌平. (2017). *統計の因果探索*. 講談社.
- [2] Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge press.
- [3] Spirtes, P., Glymour, C. N., Scheines, R., Heckerman, D., Meek, C., Cooper, G., & Richardson, T. (2000). *Causation, prediction, and search*. MIT press.
- [4] Shimizu, S. (2014). Non-Gaussian Structural Equation Models for Causal Discovery. In *Statistics and Causality: Methods for Applied Empirical Research*, 153-184, Wiley.
- [5] Pearl, J., & Verma, T. (1991). A theory of inferred causation. *KR 91*, 441-452.
- [6] Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. *Social science computer review*, 9(1), 62-72.
- [7] Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov), 507-554.
- [8] Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- [9] Dodge, Y., & Rousson, V. (2001). On asymmetric properties of the correlation coefficient in the regression setting. *The American Statistician*, 55(1), 51-54.