

非ガウス性を用いた線形非巡回なデータ生成過程部分の発見と同定

Detection and identification of a linear acyclic data generating process part
by using non-Gaussianity of data

田代 竜也 清水 昌平 鷲尾 隆
Tatsuya Tashiro Shohei Shimizu Takashi Washio

大阪大学 産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

Recently, large amount of observed data has been accumulated in various fields and there is a growing need for estimating generating process of these data. In a recent study, it was shown that use of non-Gaussianity allows the full structure of a linear acyclic model to be identified without any prior knowledge on the structure. But, existing methods could return completely wrong estimation results if some assumptions are violated. In this talk, we propose a new method for learning the data generating process that is robust against one typical violation of the model assumptions: latent confounders. The method enable us to estimate the data generating process with more precision even so the data includes latent confounders.

1. はじめに

経験科学においてベイジアンネットワークは変数間の因果関係を分析するために広く用いられている [Spirtes 93]。しかし、一般的に線形ガウス性の仮定を用いており、ほとんどの重要なモデルにおいて、因果構造に対する事前情報なしに一意にモデルを識別することができないという、重大な識別性の問題を有している。近年、非ガウス性を利用することにより、事前情報なしに線形非巡回モデルの全ての構造が同定可能であることが示された。このモデルは LiNGAM (Linear Non-Gaussian Acyclic Model) モデル [Shimizu 06] と呼ばれ、独立成分分析 (Independent Component Analysis, ICA) [Hyvärinen 01] と密接に関係している。

LiNGAM モデルにおけるこれまでの推定手法 [Shimizu 06, Shimizu 11] は、全てのモデルの仮定が満たされていることを前提に因果的順序を学習する。故に、モデルのいずれかの仮定が満たされていない場合、これらのアルゴリズムは完全に誤った推定結果を出力することがある。そこで本稿では、仮定が破綻する典型的な原因の一つである潜在交絡変数に対してロバストに因果的順序を学習する新しいアルゴリズムを提案する。潜在交絡変数とは、観測変数に因果的影響を与えるような未観測の変数を意味する。

本稿では、第 2 章で LiNGAM モデル [Shimizu 06] 及び、LiNGAM モデルに潜在交絡変数を考慮したモデル、LvLiNGAM モデル [Hoyer 08] について述べる。次に第 3 章で LvLiNGAM モデルに対して因果的順序を学習する新たなアルゴリズムを提案し、第 4 章で人工データを用いた提案手法の評価実験について述べる。最後に、第 5 章で結論を述べる。

2. 背景 : LvLiNGAM モデル

LiNGAM (Linear Non-Gaussian Acyclic Model) モデル [Shimizu 06] は、データが非ガウス分布に従うことを仮定した上で、構造方程式モデルに線形非巡回性を持たせたモデルである。観測変数 x_i が LiNGAM モデルに従って生成されると

き、 x_i は次式で表される：

$$x_i = \sum_{k(j) < k(i)} b_{ij} x_j + e_i \quad (1)$$

$k(i)$ は変数の因果的順序、 e_i は外乱、 b_{ij} は定数の係数である。外乱 e_i は平均ゼロで分散が非ゼロの非ガウス分布に従い、互いに独立であるとする。 b_{ij} は x_j から x_i への影響の強さを表す。また、 x_i の生成過程は有向非巡回グラフ (Directed Acyclic Graph, DAG) で表現できると仮定する。式 (1) の行列表記は次式で与えられる：

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e} \quad (2)$$

\mathbf{B} は b_{ij} を集めた行列、 \mathbf{x} 、 \mathbf{e} はそれぞれ x_i 、 e_i を集めたベクトルである。今、 \mathbf{x} の生成過程が非巡回であることを仮定しているため、 \mathbf{B} は変数の因果的順序 $k(i)$ に従って行と列を同時に並び替えることで対角成分がゼロの下三角行列になる。 b_{ij} のゼロ / 非ゼロのパターンは、DAG の有向辺の有無を表す。因果的順序 $k(i)$ と係数行列 \mathbf{B} は外乱 e_i の非ガウス性を用いることで、事前情報を用いずにデータ \mathbf{x} から推定することが可能である [Shimizu 06]。これは従来、 e_i にガウス性を仮定したベイジアンネットワークよりも非常に有利な特徴である [Spirtes 93]。

次に、LiNGAM モデルに潜在交絡変数を考慮したモデル、LvLiNGAM (Latent variable LiNGAM) モデル [Hoyer 08] を次式に示す：

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{\Lambda}\mathbf{f} + \mathbf{e} \quad (3)$$

LiNGAM モデル (2) との違いは、 $\mathbf{\Lambda}\mathbf{f}$ の項が潜在交絡変数からの影響を表していることである。潜在交絡変数とは、2 つ以上の観測変数の親となるような未観測の変数である。 \mathbf{f} は潜在交絡変数 f_j ($j = 1 \dots q$) を集めたベクトルとし、各 f_j は平均ゼロで分散が非ゼロの非ガウス分布に従い、互いに独立であるとする。 $\mathbf{\Lambda}$ は λ_{ij} を集めた行列であり、 λ_{ij} は f_j から x_i への影響の強さを表す。定義より潜在交絡変数は少なくとも 2 つの子を持つため、各 j に対して少なくとも 2 つの λ_{ij} は非ゼロとなる。また [Hoyer 08] では、 x_i 、 f_i 、 e_i の相関及び条件付き相関の有無は因果グラフの構造にのみ、つまり b_{ij} と λ_{ij} のゼロ / 非ゼロのパターンによってのみ決まると仮定している。

連絡先: 田代 竜也, 大阪大学 産業科学研究所, 567-0047 大阪府茨木市美穂ヶ丘 8-1, tashiro@ar.sanken.osaka-u.ac.jp

この仮定は因果推論において、忠実性 (faithfulness) の仮定として知られている [Spirtes 93]。

LvLiNGAM モデル (3) を基にした因果推論における主な問題は、因果的順序 $k(i)$ と観測変数間の影響の強さ b_{ij} を、事前情報を用いずにデータ \mathbf{x} から可能な限り推定することである。これは多くの場合において、真のモデルと同じデータ分布を与えるようなモデルの集合のみが同定可能であるためである [Hoyer 08]。

[Hoyer 08] では過完備独立成分分析 (Overcomplete ICA) を基にした推定手法が提案されている。しかしながら、過完備独立成分分析を用いた手法は信頼性に欠ける場合や、局所解に収束してしまう場合がある。そこで [Entner 11] では、過完備独立成分分析を用いない手法が提案されている。この手法では、まず潜在交絡変数からの影響を受けていない 2 変数を探索し、次にそれら 2 変数間の因果的順序を推定する。ただし、3 変数以上の因果的順序は推定されない。

3. 提案手法

この章では、潜在交絡変数を明確にモデル化せずに、2 変数以上の因果的順序を推定する手法を提案する。最初に LvLiNGAM モデル (3) における、潜在交絡変数から影響を受けていない外生変数 (DAG における根となる変数) と終点変数 (DAG における葉となる変数) を同定するための補題を示し、次に推定アルゴリズムを示す。LiNGAM モデル (2) や LiNGAM モデルを非線形拡張したモデル [Hoyer 09] を推定する最近の手法 [Shimizu 11, Mooij 09] では、潜在交絡変数が存在しないことを仮定したうえで、上から下、もしくは下から上へ順に一つずつ因果的順序を学習していく。本稿ではこれらの発想を潜在交絡変数を考慮した場合に拡張する。

最初に、潜在交絡変数を考慮した場合の [Shimizu 11] の補題 1 を示す。

補題 1. 入力データ \mathbf{x} が厳密に LvLiNGAM モデル (3) に従うと仮定する*1。 x_i を x_j で回帰したときの残差 $r_i^{(j)}$ は、式：
$$r_i^{(j)} = x_i - \frac{\text{cov}(x_i, x_j)}{\text{var}(x_j)} x_j \quad (i \neq j)$$
 で得られる。このとき、変数 x_j が親に観測変数及び潜在交絡変数を持たず、 $i \neq j$ となる全ての残差 $r_i^{(j)}$ と独立であるならば x_j は外生変数である。また、 x_j が外生変数となるのはその場合に限る。

次に、[Mooij 09] のアイデアを潜在交絡変数のある場合に拡張した補題 2 を示す。

補題 2. 入力データ \mathbf{x} が厳密に LvLiNGAM モデル (3) に従うと仮定する*1。 x_j 以外の変数を集めたベクトルを $\mathbf{x}_{(-j)}$ としたとき、 x_j を $\mathbf{x}_{(-j)}$ で回帰したときの残差 $r_j^{(-j)}$ は式：
$$r_j^{(-j)} = x_j - \boldsymbol{\sigma}_{(-j)j}^T \boldsymbol{\Sigma}_{(-j)}^{-1} \mathbf{x}_{(-j)}$$
 で得られる。ただし、 $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_j & \boldsymbol{\sigma}_{j(-j)}^T \\ \boldsymbol{\sigma}_{j(-j)} & \boldsymbol{\Sigma}_{(-j)} \end{bmatrix}$ は $[x_j, \mathbf{x}_{(-j)}]^T$ の共分散行列である。このとき、変数 x_j が子に観測変数及び潜在交絡変数を持たず、 $\mathbf{x}_{(-j)}$ が残差 $r_j^{(-j)}$ と独立であるならば、 x_j は終点変数である。また、 x_j が潜在交絡変数を持たない終点変数となるのはその場合に限る。

補題 1, 2 の証明は [Tashiro 12] を参照されたい。

提案手法ではこの 2 つの補題を組み合わせて推定を行う。最初に補題 1 を用いて、他の変数を回帰したときの残差と最も独

*1 \mathbf{x} がモデルの全ての仮定をみたし、サンプルサイズが無限であるとき \mathbf{x} は厳密にモデルに従うという [Shimizu 11]。

立となるような変数を探索することで外生変数を同定し、その外生変数の影響を他の変数から回帰により取り除く。これを、全ての変数においてその残差のどれかとの独立性が統計的に棄却されるまで繰り返す。全ての変数がその残差のどれかと従属である場合、補題 1 における外生変数は存在しない、もしくは LvLiNGAM モデル (3) のいずれかの仮定を満たしていないことを意味する。次に補題 2 を用いて、まだ因果的順序が同定されていない残りの変数から、回帰したときのそれぞれの説明変数とその残差が最も独立となるような変数を探索することで終点変数を同定し、その終点変数を残りの変数から取り除く。これを全ての変数について独立性が統計的に棄却されるまで繰り返す。本手法では、HSIC[Gretton 08] と呼ばれるカーネル法を基にした独立性指標を用いて、変数とその残差の独立性を一对ごとに検定し、フィッシャーの方法 [Fisher 50] を用いてそれらの P 値を統合する。また、本手法では繰り返し検定を行う多重比較の問題があるため、有意水準を検定数の最大値である $p-1$ で割ることでボンフェローニ補正 [Hochberg 87] をしている。ただし、 p は観測変数の数とする。

実際の推定手順を以下に示す。

1. p 次元の変数ベクトルを \mathbf{x} 、 \mathbf{x} の要素のインデックスを表す添字の集合を U 、 $p \times n$ のデータ行列を \mathbf{X} 、検定における有意水準を α とする。初期値として変数の因果的順序リスト $K_{head} := \emptyset$ 、 $K_{tail} := \emptyset$ 及び、 $m := 1$ を与える。 K_{head} は最初から $|K_{head}|$ 番目までの因果的順序、 K_{tail} は最後から $|K_{tail}|$ 番目までの因果的順序を表す。 $|K_{head}|$ 、 $|K_{tail}|$ はそれぞれのリスト内の変数の数である。
2. $\tilde{\mathbf{x}} = \mathbf{x}$ 、 $\tilde{\mathbf{X}} = \mathbf{X}$ とし、次の手順で上から順に因果的順序を同定する：

- (a) $j \in U \setminus K_{head}$ を満たす全ての j について次の手順を行う：全ての $i \in U \setminus K_{head}$ ($i \neq j$) について \tilde{x}_i を \tilde{x}_j に最小二乗回帰し、データ行列 $\tilde{\mathbf{X}}$ から残差ベクトル $\tilde{\mathbf{r}}^{(j)}$ 、残差データ行列 $\tilde{\mathbf{R}}^{(j)}$ を計算する。次に、次式を用いて残差に対して最も独立性の高い \tilde{x}_m を探索する：

$$\tilde{x}_m = \arg \max_{j \in U \setminus K_{head}} P_{Fisher}(\tilde{x}_j, \tilde{\mathbf{r}}^{(j)}) \quad (4)$$

ただし $P_{Fisher}(\tilde{x}_j, \tilde{\mathbf{r}}^{(j)})$ は $-2 \sum_i \log\{P_H(\tilde{x}_j, \tilde{\mathbf{r}}^{(j)})\}$ で定義される検定統計量の P 値、 P_H は HSIC によって導かれる P 値である。

- (b) $P_{Fisher}(\tilde{x}_m, \tilde{\mathbf{r}}^{(m)}) < \alpha/(p-1)$ の場合、手順 3 へ移行する。
- (c) m を K_{head} の末尾に加え、 $\tilde{\mathbf{x}} := \tilde{\mathbf{r}}^{(m)}$ 、 $\tilde{\mathbf{X}} := \tilde{\mathbf{R}}^{(m)}$ とする。このとき、 $|K_{head}| = n-1$ となった場合、残った最後の変数の添字を K_{head} の末尾に加え手順 4 へ移行する。それ以外の場合は手順 (2a). に戻る。

3. $|K_{head}| < n-2$ の場合*2、 $\mathbf{x}' = \mathbf{x}$ 、 $\mathbf{X}' = \mathbf{X}$ 、 $U' := U \setminus K_{head}$ とし、次の手順で下から因果的順序を同定する。

*2 手順 2 より残りの 2 変数は潜在交絡変数の影響を受けている、もしくはモデルの仮定を満たしていないと判断する。重複を避けるため、手順 3 ではその 2 変数について探索をしない。

- (a) $j \in U' \setminus K_{tail}$ を満たす全ての j について次の手順を行う: 全ての $j \in U' \setminus K_{tail}$ について, x'_j 以外の変数を含むベクトル $\mathbf{x}'_{(-j)}$ で x'_j を最小二乗回帰し, データ行列 \mathbf{X}' から 残差 $r'_j^{(-j)}$ を計算する. 次に, 次式を用いて説明変数と残差が最も独立になるような x'_m を探索する:

$$x'_m = \arg \max_{j \in U' \setminus K_{tail}} P_{Fisher}(\mathbf{x}'_{(-j)}, r'_j^{(-j)}). \quad (5)$$

- (b) $P_{Fisher}(\mathbf{x}'_{(-m)}, r'_m^{(-m)}) < \alpha/(p-1)$ の場合, 手順 4 へ移行する.
 (c) m を K_{tail} の先頭に加え, $\mathbf{x}' = \mathbf{x}'_{(-m)}$, $\mathbf{X}' = \mathbf{X}'_{(-m)}$ とする. このとき $|U' \setminus K_{tail}| < 3$ の場合, 手順 4 へ移行する*2. それ以外の場合は手順 (3a) に戻る.

4. K_{head} と K_{tail} 内の全ての変数 x_i を, x_i の全ての非子孫 x_j ($k(j) < k(i)$) で重回帰し, 影響の強さ b_{ij} を推定する.

本アルゴリズムは補題 1, 2 における外生変数と終点変数が存在しない場合, 何も因果的順序を出力しない. しかし, この結果はアルゴリズムの出力として正しいことに注意してほしい. このような場合においてもより多くの因果的順序を学習する手法として, 変数を補題 1, 2 における外生変数や終点変数が存在するような部分集合に分割し, 部分集合ごとの結果を統合する, 分割統治アルゴリズムを導入する手法が考えられる. これは今後の重要な研究課題である.

4. 人工データによる評価実験

本章では, 提案手法の評価実験として, LiNGAM モデル (2) を潜在交絡変数を考慮せずに推定する手法, DirectLiNGAM アルゴリズム [Shimizu 11] と LvLiNGAM モデル (3) を推定する手法, Pairwise LvLiNGAM アルゴリズム [Entner 11] の 2 つの手法との人工データによる比較実験を行った. 潜在交絡変数が存在せず, 十分に大きなサンプルサイズがある場合, これら全ての手法は正しい因果的順序を推定する. 人工データは変数の数を 6, サンプルサイズを 500, 1000, 2000 とし, 図 1 のデータ生成ネットワークに従って生成した. また, e_1, e_4, f_1 は 2 つのガウス分布の多峰型非対称混合分布, e_2, e_5, f_2 は二重指数分布, e_3, e_6 は 2 つのガウス分布の多峰型対称混合分布に従って生成し, 各外乱 e_i の標準偏差は SN 比, $\text{var}(x_i)/\text{var}(e_i) - 1$ が全て 1 となるように設定した. 試行回数は 100 回, 有意水準 α は 0.05 とした.

最初に, 因果的順序 $k(i)$ の評価として, 推定した 2 変数間の因果的順序のうち正しく推定できた因果的順序の割合 (適合率) と, 真のデータ生成ネットワークで潜在交絡変数を共有していない 2 つの変数間の真の因果的順序のうち正しく推定できた因果的順序の割合 (再現率) を計算した. 提案手法や DirectLiNGAM アルゴリズムと違い, Pairwise LvLiNGAM アルゴリズムは 2 変数間の因果的順序のみを推定するため, 本実験では 2 変数間の因果的順序の評価を採用している. 表 1, 2 に結果を示す. 適合率について, 提案手法は Pairwise LvLiNGAM アルゴリズムとほぼ同等の結果となった. しかし, この 2 つの手法は全ての場合で DirectLiNGAM アルゴリズムを大きく上回った. また再現率については, 提案手法が全ての場合において DirectLiNGAM アルゴリズム, Pairwise LvLiNGAM アルゴリズムの両者を上回る結果となった.

次に, 影響の強さ b_{ij} の推定精度を評価するために, 真の b_{ij} と推定した b_{ij} の二乗平均平方根誤差を計算した. 表 3 に結果を示す. 表 3 からわかるように, 全ての場合で提案手法の方がより正確に影響の強さ b_{ij} を推定することができた. ここで, Pairwise LvLiNGAM アルゴリズムは b_{ij} を推定しないことを確認しておく.

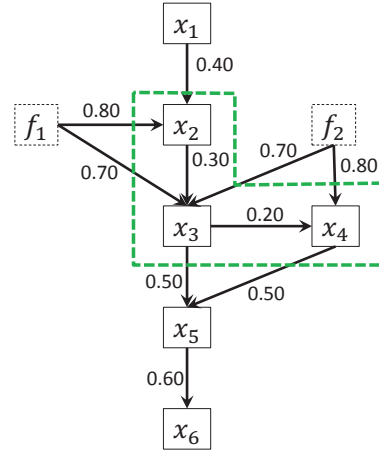


図 1: 評価実験で用いた真のデータ生成ネットワーク. 変数 f_1, f_2 は潜在交絡変数, 緑線に囲まれた変数は f_1 または f_2 を共有している変数である. 外乱 e_i は省略している.

表 1: 適合率

	サンプルサイズ		
	500	1000	2000
提案手法	0.78	0.80	0.80
DirectLiNGAM	0.65	0.64	0.64
Pairwise LvLiNGAM	0.79	0.81	0.81

表 2: 再現率

	サンプルサイズ		
	500	1000	2000
提案手法	0.97	0.99	0.99
DirectLiNGAM	0.81	0.80	0.81
Pairwise LvLiNGAM	0.86	0.89	0.90

表 3: 二乗平均平方根誤差

	サンプルサイズ		
	500	1000	2000
提案手法	0.079	0.070	0.015
DirectLiNGAM	0.090	0.079	0.057

5. 結論

本稿では, 新たに潜在交絡変数に対してロバストに因果的順序を学習する新しいアルゴリズムを提案した. 人工データによる評価実験では, 今回提案した手法が 2 つの既存手法よりも正確に因果的順序を学習した. 今後の研究として, 機能的磁気共鳴画像 (functional Magnetic Resonance Imageng, fMRI) データに本提案手法を適用し, 脳領域間の相互作用の

分析を試みる予定である。

参考文献

- [Entner 11] Entner, D. and Hoyer, P. O.: Discovering Unconfounded Causal Relationships Using Linear Non-Gaussian Models, in *New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science*, Vol. 6797, pp. 181–195 (2011)
- [Fisher 50] Fisher, R.: *Statistical methods for research workers*, Oliver and Boyd (1950)
- [Gretton 08] Gretton, A., Fukumizu, K., Teo, C., Song, L., Schölkopf, B., and Smola, A. J.: A Kernel Statistical Test of Independence, in *Advances in Neural Information Processing Systems 20*, MIT Press, Cambridge, MA (2008)
- [Hochberg 87] Hochberg, Y. and Tamhane, A. C.: *Multiple comparison procedures*, John Wiley & Sons, New York (1987)
- [Hoyer 08] Hoyer, P. O., Shimizu, S., Kerminen, A., and Palviainen, M.: Estimation of causal effects using linear non-gaussian causal models with hidden variables, *International Journal of Approximate Reasoning*, Vol. 49, No. 2, pp. 362–378 (2008)
- [Hoyer 09] Hoyer, P. O., Janzing, D., Mooij, J., Peters, J., and Schölkopf, B.: Nonlinear causal discovery with additive noise models, in *Advances in Neural Information Processing Systems 21*, pp. 689–696 (2009)
- [Hyvärinen 01] Hyvärinen, A., Karhunen, J., and Oja, E.: *Independent component analysis*, Wiley, New York (2001)
- [Mooij 09] Mooij, J., Janzing, D., Peters, J., and Schölkopf, B.: Regression by dependence minimization and its application to causal inference in additive noise models, in *Proc. the 26th Int. Conf. on Machine Learning (ICML2009)*, pp. 745–752 (2009)
- [Shimizu 06] Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A.: A linear non-Gaussian acyclic model for causal discovery, *J. Mach. Learn. Res.*, Vol. 7, pp. 2003–2030 (2006)
- [Shimizu 11] Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., Hoyer, P. O., and Bollen, K.: DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model, *J. Mach. Learn. Res.*, Vol. 12, pp. 1225–1248 (2011)
- [Spirtes 93] Spirtes, P., Glymour, C., and Scheines, R.: *Causation, Prediction, and Search*, Springer Verlag (1993), (2nd ed. MIT Press 2000)
- [Tashiro 12] Tashiro, T., Shimizu, S., Hyvärinen, A., and Washio, T.: Estimation of Causal orders in a linear non-Gaussian acyclic model: a method robust against latent confounders, *arXiv:1204.1795* (2012)