

潜在交絡変数が存在する場合の ベイズ的アプローチによる非ガウス因果構造推定法

A Bayesian estimation approach for analyzing non-Gaussian data generating processes when there are latent confounding variables

田中 直樹 清水 昌平 鷲尾 隆
Naoki Tanaka 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. It has been considered to estimate the data generating processes of variables using a linear, acyclic model based on non-Gaussianity of external influences (LiNGAM). However, results of the estimation can be biased if there are latent confounding variables. In this paper, we first review LiNGAM, its extended model and estimation procedure for LiNGAM in a Bayesian framework. Then we propose a new Bayesian estimation procedure that solves the problem.

1. はじめに

データマイニングの分野では、データの生成モデルに非ガウス分布を仮定することで変数間の因果関係を向きも含めて導出する因果推論に関する手法が盛んに研究されている。しかし、潜在交絡変数と呼ばれる、図1のような未観測かつ複数の観測変数に影響を与える変数が存在する場合、推定結果が歪められる可能性がある。近年、潜在交絡変数が存在する場合の観測変数間の因果関係を推定する方法が研究されているが、局所解に収束する場合があります安定した推定ができない等の問題がある [Hoyer 08]。そこで本稿では非ガウス性に基づく既存手法を、潜在交絡変数が存在する場合でも因果構造を正しく推定できるように改良することを試みる。そして、人工的に生成したデータを用いて評価実験を行い、その結果について考察する。

潜在交絡変数

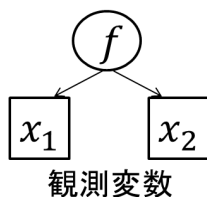


図1: 二変数間に潜在交絡変数が存在する場合の有向非巡回グラフ。 f は潜在交絡変数, x_1, x_2 は観測変数を表す。

2. 既存手法

2.1 LiNGAM モデル

まず、観測されるデータが図2のような有向非巡回グラフ (Directed Acyclic Graph, DAG) により生成されるものと仮定する。変数の数を n としたとき、この有向非巡回グラフを $n \times n$ の隣接行列 $\mathbf{B} = \{b_{ij}\}$ で表す。ここで b_{ij} は有向非巡回グラフにおける、変数 x_j から変数 x_i への結合の強さを表すものとする。さらに変数 x_i の因果的順序を $k(i)$ で表すこと

連絡先: 田中 直樹, 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, tanaka@ar.sanken.osaka-u.ac.jp

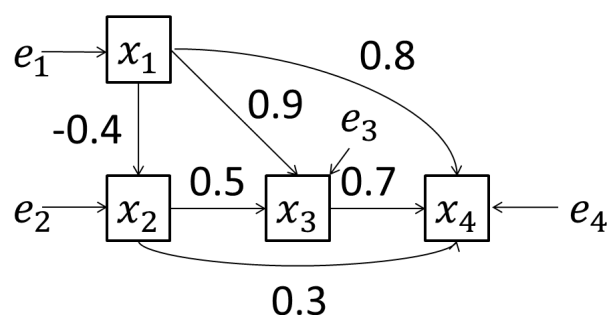


図2: LiNGAM モデルを表す有向非巡回グラフ。ノードは変数を表し、辺は変数間の因果関係の向きとその結合の強さを表す。

とする。加えて、変数間の関係が線形であると仮定する。以上より、 e_i を外的影響、 μ_i を定数として、LiNGAM モデル [Shimizu 06] は次の式で表される。

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

外的影響 e_i は全て平均 0, 分散非ゼロの非ガウス分布に従う連続な確率変数である。ただし、平均や分散等の適当な次数のモーメントの存在を仮定する。また、潜在交絡変数は存在しないと仮定する。もし潜在交絡変数が存在する場合、潜在交絡変数からの影響が外的影響 e_i に含まれるため e_i は互いに独立ではない。そのため、式 (1) によって因果構造を正しく表現することができない。逆に、もし潜在交絡変数が存在しなければ e_i は互いに独立である [Spirtes 93]。

式 (1) は次の形の行列で表せる。

$$\mathbf{x} = \mathbf{B}\mathbf{x} + (\mathbf{I} - \mathbf{B})\boldsymbol{\mu} + \mathbf{e} \quad (2)$$

ここで $\mathbf{x} = [x_1, \dots, x_n]^T$ は n 次元ベクトルであり、 \mathbf{B} は非巡回の仮定より、行と列を同時に並び換えることで**厳密な下三角行列**、すなわち対角要素が全て 0 の下三角行列に変形できる [Bollen 89]。

2.2 LiNGAM 混合モデル

この節では、 l 個の異なる構造の LiNGAM モデルに従って生成されたデータが混在する場合を表現した LiNGAM 混合モデル [Shimizu 08] について述べる。ここで構造とは、データの変数間の結合の強さ b_{ij} 、外的影響の確率密度 p_i 、定数 μ_i のことを指す。LiNGAM モデルの構造が同じであるデータの集合をクラスとし、クラス c に属するデータの変数間の結合の強さを $b_{ij}^{(c)}$ 、外的影響を $e_i^{(c)}$ 、定数を $\mu_i^{(c)}$ とすると、LiNGAM 混合モデルによって、クラス c に属するデータは行列の形で以下のように表される。

$$\mathbf{x} = \mathbf{B}^{(c)} \mathbf{x} + (\mathbf{I} - \mathbf{B}^{(c)}) \boldsymbol{\mu}^{(c)} + \mathbf{e}^{(c)} \quad (3)$$

また、 \mathbf{x} の要素 x_i については、

$$x_i = \sum_{k(j) < k(i)} b_{ij}^{(c)} (x_j - \mu_j^{(c)}) + \mu_i^{(c)} + e_i^{(c)} \quad (4)$$

と表される。クラスの数 l が 1 ならば LiNGAM 混合モデルは LiNGAM モデルと等しい。

さらに、以下の式に従って各クラスに属する観測データの確率密度が混合されるとする。

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^l p(\mathbf{x}|\boldsymbol{\mu}^{(c)}, \mathbf{B}^{(c)}) p(c=k) \quad (5)$$

ただし、 $\boldsymbol{\theta} = [\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(l)}]$ 、 $\boldsymbol{\theta}^{(c)} = [(\boldsymbol{\mu}^{(c)})^T, \text{vec}(\mathbf{B}^{(c)})^T]^T$ であり、 $\text{vec}(\cdot)$ は行列を列ごとに分解した後、一列目から順に上から並べて列ベクトルにする操作を表す。また、 $p(c=k)$ は各確率密度の重み係数である。

2.3 BayesLiNGAM 法

本稿では 2.2 節で述べた LiNGAM 混合モデルに従ってデータが生成すると仮定して、ベイズ的アプローチにより推定を行うが、そのための準備として、この節ではデータは 2.1 節で述べた LiNGAM モデルに従って生成される、すなわち全て同じ構造 (同じ b_{ij}, p_i, μ_i) の LiNGAM モデルに従うと仮定し、そのデータの因果関係を推定する際にベイズ推定を行う BayesLiNGAM 法 [Hoyer 09] について述べる。ベイズ推定は、ある複数の仮説の中からそれらの事後確率が最大である仮説を選択する推定法である。よって、BayesLiNGAM においては二変数間の因果関係の有無と向きを仮定すると、三種類の DAG ($G_1: x_1 \rightarrow x_2$ or $G_2: x_1 \leftarrow x_2$ or $G_3: x_1 \perp x_2$) が存在するので、 G_1, G_2, G_3 それぞれの事後確率を計算し、その値が最も大きい DAG を出力する。

ある因果構造を $G_k (k = 1, 2, 3)$ 、観測データセットを $D (D = \{\mathbf{x}^1, \dots, \mathbf{x}^N\}, N$ はサンプル数) とする。ここで、データの各サンプルが互いに独立であるという仮定より $p(\{\mathbf{x}^1, \dots, \mathbf{x}^N\}) = \prod_{m=1}^N p(\mathbf{x}^m)$ である。ベイズの定理より事後確率 $P(G_k|D)$ は次式で表される：

$$P(G_k|D) = \frac{p(D|G_k)P(G_k)}{p(D)} \quad (6)$$

よって式 (6) において、尤度 $p(D|G_k)$ 、事前確率 $P(G_k)$ 、事後確率を正規化する定数 $p(D)$ を求めれば事後確率が算出できる。

事前確率 $P(G_k)$ は事前情報を表す。もし事前情報が何もなければ、因果構造 G_k が三通りであることから $P(G_k) = \frac{1}{3}$ となる。

また、 $p(D)$ は事後確率を正規化する定数であり、事前確率で評価したデータの実現確率 (尤度) の期待値であるので、以下の式で求められる：

$$P(D) = \sum_{k=1}^3 p(D|G_k)P(G_k) \quad (7)$$

最後に、尤度 $p(D|G_k)$ は、ある因果構造 G_k を仮定した時にデータが D である確率を表し、以下の式で求められる：

$$P(D|G_k) = \int p(D|\boldsymbol{\theta}, G_k)p(\boldsymbol{\theta}|G_k)d\boldsymbol{\theta} \quad (8)$$

ここで、 $\boldsymbol{\theta}$ は LiNGAM モデルの式 (1) における b_{ij}, μ_i 、そして外的影響 e_i の確率密度 p_i を一つのパラメータにまとめた変数である。 $p(\boldsymbol{\theta}|G_k)$ については、 b_{ij} は標準ガウス分布に従うと仮定すること、そして以下のように p_i をモデリングすることで特定できる。

e_i の非ガウス性を表現するために、本稿ではガウス分布に形状 (shape) パラメータを加えた一般化ガウス分布 (generalized gaussian distribution) を用いる。一般化ガウス分布は対称であり、あらゆるパラメータのガウス分布とラプラス分布、そして有界な区間の一様分布を含む。また、確率密度は尺度 (scale) パラメータ $\alpha_i (> 0)$ 、形状パラメータ $\beta_i (> 0)$ を用いて以下の式で表される：

$$p_i(e_i) = \frac{\beta_i \exp(-(|e_i|/\alpha_i)^{\beta_i})}{2\alpha_i \Gamma(1/\beta_i)} \quad (9)$$

$\Gamma(\cdot)$ はガンマ関数であり、定義式は実部が正である複素数 z について、

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \quad (Rz > 0) \quad (10)$$

である。また、一般化ガウス分布の分散 σ_i^2 は以下の式で表される：

$$\sigma_i^2 = \frac{\alpha_i^2 \Gamma(3/\beta_i)}{\Gamma(1/\beta_i)} \quad (11)$$

さらに、 $p(D|\boldsymbol{\theta}, G_k)$ については、サンプルが互いに独立であるという仮定より $p(\{\mathbf{x}^1, \dots, \mathbf{x}^N\}) = \prod_{m=1}^N p(\mathbf{x}^m)$ であるので、 $p(\mathbf{x}|\boldsymbol{\theta}, G_k)$ を全てのサンプルについて掛け合わせればよい。そこで、 $p(\mathbf{x}|\boldsymbol{\theta}, G_k)$ は LiNGAM モデルより導出すると、

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}, G_k) &= \prod_{i=1}^n p_i(e_i) \\ &= \prod_{i=1}^n p_i(x_i - \mu_i - \sum_{k(j) < k(i)} b_{ij}(x_j - \mu_j)) \end{aligned} \quad (12)$$

となり、 $\boldsymbol{\theta}, G_k$ が与えられた時の観測データ \mathbf{x} の確率密度が求められる。

3. 提案手法

3.1 潜在交絡変数を含む場合のデータ生成モデル

LiNGAM モデルにおいて潜在交絡変数が存在するならば、それが外的影響に含まれ e_i は互いに独立ではなく、推定結果

を歪める原因となる．そこで本稿では， e_i は互いに独立であることを前提とする LiNGAM モデルに潜在交絡変数を組み込むために，潜在交絡変数を変数ではなく定数として扱うことを考える．そのために潜在交絡変数の値を一定以上の精度で幾つか（比較的少数）の定数値に離散化することができると仮定する．もし，この仮定が満たされるならば，離散化した潜在交絡変数の値が同じデータをまとめてクラスとし，それぞれのクラス内では潜在交絡変数の値は一定となり，定数として扱うことができる．そうすれば e_i は互いに独立であり，それぞれのクラスに属するデータは LiNGAM モデルに従って生成すると考えることができる．ただし，クラス毎に定数 μ_i が異なるので，2.2 節で述べた LiNGAM 混合モデルに従ってデータが生成されると考える．また， b_{ij}, e_i は潜在交絡変数の影響を受けないため，全てのクラスの b_{ij} ，そして e_i の確率密度 p_i は同じである．

以上より，データが潜在交絡変数を含む場合のクラス c のデータ生成モデルは式 (3) を用いて以下のように表される．

$$\mathbf{x} = \mathbf{B}\mathbf{x} + (\mathbf{I} - \mathbf{B})\boldsymbol{\mu}^{(c)} + \mathbf{e} \quad (13)$$

また， \mathbf{x} の要素 x_i については，

$$x_i = \sum_{k(j) < k(i)} b_{ij}(x_j - \mu_j^{(c)}) + \mu_i^{(c)} + e_i \quad (14)$$

と表される．確率密度の混合も，式 (5) と同様に，

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^l p(\mathbf{x}|\boldsymbol{\mu}^{(c)}, \mathbf{B})p(c = k) \quad (15)$$

と表される．

よって，外的影響 e_i の確率密度 p_i を式 (9) でモデリングした場合のクラス c に属するデータ \mathbf{x} の確率密度は， $\boldsymbol{\theta}^{(c)} = [b_{ij}, \mu_i^{(c)}, p_i]$ とすると式 (12)，式 (14) より，

$$p(\mathbf{x}|\boldsymbol{\theta}^{(c)}) = \prod_{i=1}^n p_i(x_i - \mu_i^{(c)} - \sum_{k(j) < k(i)} b_{ij}(x_j - \mu_j^{(c)}))$$

と表され，これを式 (15) により全クラスの \mathbf{x} の確率密度を混合すると，

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}) &= \sum_{k=1}^l p(\mathbf{x}|\boldsymbol{\theta}^{(c)})p(c = k) \\ &= \sum_{k=1}^l \left\{ \prod_{i=1}^n p_i(x_i - \mu_i^{(c)}) \right. \\ &\quad \left. - \sum_{k(j) < k(i)} b_{ij}(x_j - \mu_j^{(c)}) \right\} p(c = k) \end{aligned} \quad (16)$$

となり， $\boldsymbol{\theta}^{(c)}, G_k$ が与えられた時の観測データ \mathbf{x} の確率密度を求めることができる．

3.2 ディリクレ分布と多項分布

前節でモデリングした，クラス毎に異なるデータの確率密度を混合する際の各確率密度の重み係数 $p(c = k)$ を決定するために本研究ではディリクレ分布を用いる．ディリクレ分布からサンプルを採るには，ガンマ分布に従う独立なサンプル $\gamma_1, \dots, \gamma_Q$ を発生させて，それらの和が 1 になるように，

$$p_r = \frac{\gamma_r}{\sum_{r=1}^Q \gamma_r} \quad (17)$$

と正規化すればよい．ここで，ガンマ分布の確率密度関数は形状パラメータ $u > 0$ ，尺度パラメータ $v > 0$ の二つのパラメータを用いて次のように表される：

$$f(x) = x^{u-1} \frac{\exp(-x/v)}{\Gamma(u)v^u} \quad (x > 0) \quad (18)$$

$\Gamma(\cdot)$ はガンマ関数であり，前述の式 (10) で表される．

さらに，ディリクレ分布から得た重み係数 $p(c = k)$ をパラメータとして，多項分布によりデータがどのクラスに属するのかが決定する．多項分布は一般に，試行回数を N ，事象 X_k が起こる確率を $p_k (k = 1, \dots, Q)$ とすると以下の式で表される：

$$\begin{aligned} P(X_1 = n_1, X_2 = n_2, \dots, X_Q = n_Q) \\ = \frac{N!}{n_1!n_2! \dots n_Q!} p_1^{n_1} p_2^{n_2} \dots p_Q^{n_Q} \end{aligned} \quad (19)$$

ここで， p_k はディリクレ分布から得た重み係数 $p(c = k)$ と対応している．また， $p_k > 0, p_1 + p_2 + \dots + p_Q = 1, n_1 + n_2 + \dots + n_Q = N$ である．

3.3 階層ベイズ法

本稿では観測変数間の結合の強さ b_{ij} ，定数 μ_i の事前分布としてガウス分布を用いる．その際ガウス分布のパラメータ（平均と分散）を決めなければならないが，平均は 0 とし，分散にはハイパーパラメータを用いる．ハイパーパラメータとは，分布のパラメータを規定するパラメータである．すなわち，分散を定数にするのではなく，さらに事前分布を仮定する．そして，最も事後確率を高くするパラメータを推定に利用する．このように，事前分布にさらに事前分布を仮定するようなモデルは階層ベイズ法 [伊庭 04] と呼ばれ，観測データから適切だと思われるパラメータを推定し，利用することができる．なお，本稿ではハイパーパラメータは逆ガンマ分布に従うとする．

4. 評価実験

本稿ではサンプルサイズ $N = 50, 100, 200$ ，真のグラフが $G_1 : x_1 \rightarrow x_2, G_2 : x_1 \rightarrow x_2, G_3 : x_1 \leftarrow x_2$ である，潜在交絡変数を含むデータセットをそれぞれ 100 個ずつ生成した．潜在交絡変数は 3 個存在するとし，潜在交絡変数と外的影響 e_i 合わせて 5 個の変数はそれぞれラプラス分布，混合ガウス分布（混合数 2，多峰型かつ対称），混合ガウス分布（混合数 2，多峰型かつ非対称）のうちランダムでいずれかに従って生成させた．その後，潜在交絡変数と外的影響それぞれを正規化（平均を 0，分散を 1 に）した．また，データは LiNGAM モデルに従って生成させた．具体的には，LiNGAM モデルの式 (2) において潜在交絡変数と観測変数を分けた以下の式に従って生成させた（ただし $\boldsymbol{\mu} = \mathbf{0}$ とした）．

$$\mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\lambda}\mathbf{f} + \mathbf{e}) \quad (20)$$

\mathbf{f} は潜在交絡変数を， $\boldsymbol{\lambda}$ は潜在交絡変数から観測変数への影響の強さを表す．行列 $\boldsymbol{\lambda}$ は，一行目の要素を (0.3, 0.9, -1)，二行目の要素を (0.7, 0.6, 0.8) とした．また，真のグラフが G_2, G_3 の時の観測変数間の結合の強さはそれぞれ $b_{21} = 0.8, b_{12} = 0.8$ とした．さらに，外的影響の確率密度 p_i に用いる一般化正規分布のパラメータ（式 (9) の α_i, β_i ）と， b_{ij}, μ_i の事前分布（ガウス分布）の分散に，形状・尺度パラメータともに 3 である逆ガンマ分布を用いた．そしてデータセットそれぞれについて提案手法と DirectLiNGAM [Shimizu 11]，

表 1: 真のグラフが G_1 の時の実験結果. 数値は 100 回中正しく推定した回数を示す. PairwiseLvLiNGAM については, 「正しく推定した回数/潜在交絡変数は存在しないと推定した回数」という形で表記している.

G_1	サンプル数		
	50	100	200
提案手法	84	94	97
DirectLiNGAM	80	77	69
PairwiseLvLiNGAM	64/64	52/52	42/42

表 2: 真のグラフが G_2 の時の実験結果. 数値は 100 回中正しく推定した回数を示す. PairwiseLvLiNGAM については, 「正しく推定した回数/潜在交絡変数は存在しないと推定した回数」という形で表記している.

G_2	サンプル数		
	50	100	200
提案手法	93	98	100
DirectLiNGAM	47	44	31
PairwiseLvLiNGAM	6/12	7/20	0/14

表 3: 真のグラフが G_3 の時の実験結果. 数値は 100 回中正しく推定した回数を示す. PairwiseLvLiNGAM については, 「正しく推定した回数/潜在交絡変数は存在しないと推定した回数」という形で表記している.

G_3	サンプル数		
	50	100	200
提案手法	93	99	100
DirectLiNGAM	41	35	38
PairwiseLvLiNGAM	6/16	5/24	2/14

PairwiseLvLiNGAM[Entner 11] アルゴリズムにより推定を行い, 正しく推定できたかどうかを判定し, 両者の性能を比較した. ただし, PairwiseLvLiNGAM は潜在交絡変数が存在すると推定した場合, 結果を出力しないアルゴリズムであるので, 結果を出力した場合のみの比較となる. また, DirectLiNGAM は式 (2) の LiNGAM モデルの標準的な推定アルゴリズムであり, 潜在交絡変数は存在しないと仮定している. 提案手法について, 潜在交絡変数を離散化した時のクラスの数 $c = 1, \dots, 2\log N$ (小数点以下切り捨て) それぞれについて推定を行った. そして各クラス数において事後確率が最も高かったグラフを決定し, さらにその中でも一番事後確率の高かったグラフを推定結果として出力させた.

実験結果を表 1~表 3 に示す. 実験結果より提案手法は, 潜在交絡変数が存在しないと仮定する DirectLiNGAM とは対照的に, サンプルサイズの増加に伴って推定精度も向上していた. このことから, 提案手法は推定結果が歪められていないことがわかった. また, PairwiseLvLiNGAM は推定により潜在交絡変数が存在するとして結果を出力しない場合があったのに対して, 提案手法は結果を出力し, なおかつ正しく推定する割合も PairwiseLvLiNGAM より高かった.

5. 結論

本稿では, 因果推論分野における既存の非ガウス性に基づく手法に潜在交絡変数が存在しても正しい推定を可能にする改良を加え, その性能を評価した. 従来の非ガウス性に基づく手法では, 潜在交絡変数が存在する場合に結果を出力しない, 安定した推定ができない等の問題点がある. 本稿の成果により, 潜在交絡変数が存在する場合でも結果を出力し, より安定した推定をすることが可能となった. そして, 実験結果により潜在交絡変数を無視する, あるいは潜在交絡変数が存在する場合に

結果を出力しない従来のアルゴリズムに比べて推定精度が向上することを確認できた.

今後の課題として二点挙げられる. 一点目は, 人工的に生成したデータではなく, 現実に蓄積されたデータに対して提案手法を適用し性能評価を行うことである. 二点目は, 潜在交絡変数を離散化した時の最適なクラスの数を観測データから推定する手法 (例えばディリクレ過程 [上田 07] など) の導入である.

参考文献

- [Bollen 89] Bollen, K. A.: *Structural Equations with Latent Variables*, John Wiley & Sons (1989)
- [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*, 6797, pp. 181–195 (2011)
- [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, *Int. J. Approximate Reasoning*, Vol. 49, No. 2, pp. 362–378 (2008)
- [Hoyer 09] Hoyer, P. O. and Hyttinen, A.: Bayesian Discovery of Linear Acyclic Causal Models, in *Proc. 25th Conf. on Uncertainty in Artificial Intelligence (UAI2009)* (2009)
- [Shimizu 06] Shimizu, S., Hoyer, P. O., Hyvärinen, A., and Kerminen, A.: A linear non-gaussian acyclic model for causal discovery, *J. Machine Learning Research*, Vol. 7, pp. 2003–2030 (2006)
- [Shimizu 08] Shimizu, S. and Hyvärinen, A.: Discovery of linear non-gaussian acyclic models in the presence of latent classes, in *Proc. 14th Int. Conf. on Neural Information Processing (ICONIP2007)*, Kitakyushu, Japan, pp. 752–761 (2008)
- [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)
- [伊庭 04] 伊庭 幸人, 石黒 真木夫, 松本 隆, 乾 敏郎, 田邊 國士: 階層ベイズモデルとその周辺, 岩波書店 (2004)
- [上田 07] 上田 修功, 山田 武士: Introduction to Nonparametric Bayesian Models, *応用数理* 17(3), pp. 196–214 (2007)