1A5-3

データの非正規性を活用する因果構造探索法と事前情報の利用

Use of prior knowledge in a non-Gaussian method for causal structure learning

稲積 孝紀 Takanori Inazumi　十河 泰弘 Yasuhiro Sogawa　清水 昌平 Shohei Shimizu　河原 吉伸 Yoshinobu Kawahara　鷲尾 隆 Takashi Washio

大阪大学 産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University

We discuss causal structure learning based on linear structural equation models. Most of conventional learning methods implicitly assume Gaussianity and often create many indistinguishable models. Therefore, in many cases it is difficult to obtain much information on the structure. Recently, a non-Gaussian learning method called LiNGAM has been proposed to identify the model structure without using prior knowledge on the structure. However, more efficient learning can be achieved if some prior knowledge on a part of the structure is available. In this paper, we propose to use prior knowledge to improve the performance of a state-of-art non-Gaussian method. Experiments on artificial data show that the accuracy and computational time are significantly improved even if the amount of prior knowledge is not so large.

1. はじめに

データマイニングの分野では，変数間の因果関係を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．統計的因果推論の分野では，(線形) 構造方程式モデル (Structural Equation Model, SEM) [Bollen 89] と呼ばれる統計モデルが主に用いられる．構造方程式モデルは，ある変数が別の変数に影響を与えその変数がさらに別の変数に影響を与えるという関係を，それら変数から構成される方程式の組として表し，データがその方程式に従って生成されていると仮定するモデルである．そしてそのモデルの元で，データから方程式がどのようなものであるかを推定する．

従来は，データの分布が正規分布に従うと仮定する手法が標準的であった [Bollen 89]．しかし，正規分布を仮定する手法では多くの場合因果関係の向きを推定することができず，また，現実には正規分布に従わないデータも多数存在する．そこで近年，データの分布に非正規分布を仮定することで，正規分布を仮定する従来手法の問題点を解消し，因果関係の強さと向きの両方を推定できる手法が提案されている [Shimizu 06]．この手法は，因果関係の推定に実際に蓄積されたデータのみを利用し，因果の向きに関する事前情報は利用しない．しかし，現実には推定を行う前の段階で全てではなくともある程度の情報を事前に知っている場合がある．この事前情報を推定に生かすことができれば，より効率的な推定を行うことができると考えられる．本稿では，既存の非正規性に基づく手法の中で最新鋭の方法 [Shimizu 09] を事前情報も利用して推定を行うように改良する．そして，人工的に生成されたデータに対し性能評価を行い，その結果について考察する．

2. 先行研究

2.1 LiNGAM モデル

LiNGAM モデル (Linear Non-Gaussian Acyclic Model) [Shimizu 06] は構造方程式モデルの特別な場合である線形非巡回モデルのもとで，データの生成過程に非正規分布を仮定するモデルである．

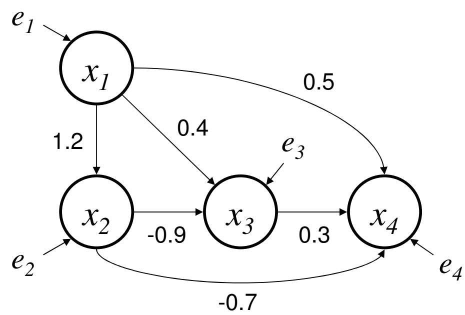

図 1　LiNGAM モデルを表す有向非巡回グラフ．式 $x_1 = e_1$, $x_2 = 1.2x_1 + e_2$, $x_3 = 0.4x_1 - 0.9x_2 + e_3$, $x_4 = 0.5x_1 - 0.7x_2 + 0.3x_3 + e_4$ で表されるモデルに対応する．

まず，観測変数の関係が図 1 のような有向非巡回グラフ (Directed Acyclic Graph, DAG) により表されるものと仮定する．変数の数を p としたとき，この有向非巡回グラフを $p \times p$ 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 $k(i)$ を，有向非巡回グラフにおいて順序が後の変数から前の変数に向かう有向パスが存在しないような順序とする．言い換えれば，$k(i) < k(j)$ のとき x_i から x_j への有向パスは存在し得るが，x_j から x_i への有向パスは存在しない．加えて，変数の間の関係が線形であると仮定する．また，一般性を失うことなく，観測変数それぞれの平均が 0 であると仮定する．以上より，e_i を外的影響として，LiNGAM モデルは次の式で表される．

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

外的影響 e_i は全て平均 0，分散 1 の非正規分布に従う連続変数であり，ランダムな値をもつ．また，推定に必要な変数が全て観測されているものと仮定し，これにより e_i は互いに独立である [Spirtes 93].

連絡先: 稲積 孝紀，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，inazumi@ar.sanken.osaka-u.ac.jp

式 (1) は行列を用いて次のように表せる．

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e} \tag{2}$$

ここで $\mathbf{x} = [x_1, \ldots, x_p]^T$ は p 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]．式 (1) は，$\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$ を用いて次のようにも表せる．

$$\mathbf{x} = \mathbf{A}\mathbf{e} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{e} \tag{3}$$

ここで行列 $\mathbf{A}$ の各要素 a_{ij} は変数 x_j から変数 x_i への総合効果を表す [Hoyer 08]．因果効果は普通，総合効果を指す [Pearl 00]．

変数 x_i について，モデル中の他の全ての変数 $x_j\ (j \neq i)$ から x_i への有向パスが存在しないとき，変数 x_i は外的影響 e_i と等しい．このとき，外的影響 e_i は変数 x_i として観測される．このような変数 x_i を**外生変数**と呼ぶ．また，これ以外の場合の e_i を**誤差**と呼ぶ．非巡回の仮定および推定に必要な変数が全て観測されているとする仮定により，どのようなモデルの場合についても，外生変数は少なくとも 1 つ存在する [Shimizu 09]．

2.2 DirectLiNGAM アルゴリズム

LiNGAM モデルを探索するアルゴリズムの 1 つとして，DirectLiNGAM [Shimizu 09] が提案されている．まず最初に，correlation-faithfulness と呼ばれる概念の定義を引用する．

定義 1 (correlation-faithfulness)**.** 変数ベクトル $\mathbf{x}$ について，x_i の相関および条件付き相関が，x_i を生成するグラフ構造すなわち b_{ij} の値がゼロか非ゼロかによってのみ規定されるとき，$\mathbf{x}$ の分布はそのグラフに対して correlation-faithful であると言う．

さらに，文献 [Shimizu 09] で証明された 2 つの補題と 1 つの系を引用する．DirectLiNGAM アルゴリズムの概要を以下に述べる．まず，次の補題 1 を適用し，因果的順序が 1 番目となる変数である外生変数を探す．

補題 1. 入力データ $\mathbf{x}$ が式 (2) の LiNGAM モデルに従い，さらに $\mathbf{x}$ の分布が correlation-faithful であると仮定する．変数 x_i を x_j に回帰したとき，その残差は次式で得られる: $r_i^{(j)} = x_i - \frac{\mathrm{cov}(x_i,x_j)}{\mathrm{var}(x_j)} x_j\ (i \neq j)$．ここで，変数 x_j が $i \neq j$ となる全ての残差 $r_i^{(j)}$ と独立ならば x_j は外生変数であり，また，外生変数となるのはその場合に限る．

次に，最小自乗回帰により外生変数の影響を他の変数から取り除き，その残差を得る．そして，その残差に対しても式 (2) の LiNGAM モデルが成り立つことが，次の補題 2 により示される．さらに，残差間の因果的順序が対応する元の観測変数の因果的順序と等しいことが，次の系 1 により示される．

補題 2. 補題 1 の仮定を満たすものとし，さらに変数 x_j が外生変数であると仮定する．$i \neq j$ となる全ての変数 x_i を x_j に回帰した残差 $r_i^{(j)}$ を並べた $p-1$ 次元ベクトルを $\mathbf{r}^{(j)}$ とする．このとき，残差ベクトル $\mathbf{r}^{(j)}$ は LiNGAM モデルに従う: $\mathbf{r}^{(j)} = \mathbf{B}^{(j)}\mathbf{r}^{(j)} + \mathbf{e}^{(j)}$．$\mathbf{B}^{(j)}$ は行と列を同時に並び換えることで厳密な下三角行列に変形できる．また，$\mathbf{e}^{(j)}$ の各要素は非正規分布に従い，互いに独立である．

系 1. 補題 2 の仮定を満たすものとする．変数 x_i の因果的順序を $k(i)$ としたように，残差 $r_i^{(j)}$ の因果的順序を $k_{r^{(j)}}(i)$ とする．このとき，残差の因果的順序は元となった観測変数の因果的順序と同じである: $k_{r^{(j)}}(l) < k_{r^{(j)}}(m) \Leftrightarrow k(l) < k(m)$．

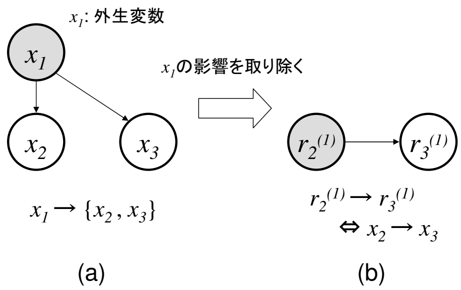

図 2 (a) は観測変数 $\mathbf{x}$ についての関係を表し，外生変数が x_1 であることがわかったとする．そして，残りの変数を外生変数 x_1 に回帰した残差 $\mathbf{r}^{(1)}$ を計算し，その関係が (b) のように表せたとする．このとき，残差モデル (b) の変数間の因果的順序は元の観測変数モデル (a) における残差の元となった変数間の順序と同じである．

以上より，残差の LiNGAM モデルの解析，すなわち残差に対して補題 1 を適用し「外生」残差を探すことで，元の観測変数において因果的順序が 2 番目となる変数を探すことができる．この操作の流れを図 2 に示す．この手順を繰り返すことにより，元の観測変数における全ての因果的順序が得られる．

3. 提案手法

3.1 DirectLiNGAM アルゴリズムにおける事前情報の利用

本節では，第 3.3 節で提案する DirectLiNGAM アルゴリズムの変種の正当性を保証する 3 つの補題を示す．まず，補題 1 の仮定のもとで，事前情報を表す行列 $\mathbf{A}^{pk} = \left[a_{ji}^{pk}\right]$ を次のように定義する．

$$a_{ji}^{pk} = \begin{cases} 0 & x_i \text{ から } x_j \text{ への有向パスがない，} \\ & \text{すなわち } a_{ji} = 0 \\ 1 & x_i \text{ から } x_j \text{ への有向パスがある，} \\ & \text{すなわち } a_{ji} \neq 0 \\ -1 & \text{上の 2 つのどちらが正しいかを} \\ & \text{判断できる事前情報がない} \end{cases} \tag{4}$$

外生変数と事前情報行列 $\mathbf{A}^{pk}$ の定義より，直ちに次の 3 つの補題が導かれる．

補題 3. 補題 1 の仮定を満たすものとする．全ての $i \neq j$ について a_{ji}^{pk} が 0 ならば，観測変数 x_j は外生変数である．

補題 4. 補題 1 の仮定を満たすものとする．a_{ji}^{pk} が 1 となる $i \neq j$ が存在するならば，観測変数 x_j は内生変数である，すなわち外生変数ではない．

補題 5. 補題 1 の仮定を満たすものとする．a_{ji}^{pk} が 0 ならば，観測変数 x_j は x_i による影響を含まない．

第 3.3 節に示すアルゴリズムの概要を以下に述べる．まず，外生変数が事前情報から特定できる場合は，補題 1 の代わりに補題 3 を適用し，因果的順序が 1 番目である外生変数を探す．この場合は，最小自乗回帰による残差の計算や観測変数と残差の間の独立性の評価をしない．事前情報から外生変数が特定で

きない場合は，次に，補題 4 を適用し内生変数，すなわち外生変数でない変数を探す．内生変数は外生変数にならないため，補題 1 における外生変数の探索空間を小さくすることができる．さらに補題 5 より，ある目的変数がある説明変数の影響を含まない場合は，その場合における残差の計算を省略する．以上より，推定する必要がある因果的順序と結合の強さの数を減らすことができ，第 4 章で示すように推定精度の向上と計算時間の短縮を得ることができる．

3.2 独立性の評価

DirectLiNGAM アルゴリズムおよび本稿におけるその変種では，ある観測変数とその残差の間の独立性（単なる無相関ではない）の指標を定義する必要がある．独立性の指標として，文献 [Shimizu 09] に示される以下の統計量を用いる．変数 x_i の添字の集合を $U = \{1, \ldots, p\}$ とする．次の統計量は，変数 x_i を x_j に回帰したときの，x_j とその残差 $r_i^{(j)} = x_i - \frac{\mathrm{cov}(x_i, x_j)}{\mathrm{var}(x_j)} x_j$ の間の非線形相関を評価する．

$$T(x_j; U) = \sum_{i \in U, i \neq j} \left[|\mathrm{corr}\{g(r_i^{(j)}), x_j\}| + |\mathrm{corr}\{r_i^{(j)}, g(x_j)\}| \right] \tag{5}$$

ここで，g は非線形かつ非二次関数となる関数，たとえば $g(\cdot) = \tanh(\cdot)$ である．変数 x_j とその残差 $r_i^{(j)}$ が独立であれば，式 (5) の統計量 T は 0 となる．厳密には，独立であることは統計量 T が 0 になることよりもはるかに厳しい条件である．しかし，独立成分分析に関する文献 [Hyvärinen 01] にあるように，式 (5) のような非線形相関の評価を用いることは，実際上は，多くの場合において十分に良い方法といえる．必要であれば，独立性を評価するために式 (5) の統計量 T を用いる代わりに，より洗練されたノンパラメトリックな独立性指標 [Bach 02, Gretton 05] を用いることができる．

3.3 pk-DirectLiNGAM アルゴリズム

以下に DirectLiNGAM アルゴリズムの変種として，correlation-faithfulness の仮定のもと，事前情報を利用する新しいアルゴリズムを示す．これを pk-DirectLiNGAM アルゴリズムと呼ぶ．

---

pk-DirectLiNGAM アルゴリズム

1. p 次元の変数ベクトルを $\mathbf{x}$，その要素を表す添字の集合を U，変数ベクトルからなる $p \times n$ のデータ行列を $\mathbf{X}$，$p \times p$ 次元の事前情報行列を $\mathbf{A}^{pk}$ とし，初期値として変数の因果的順序 $K := \emptyset$ とする．
2. $p-1$ 個の添字が K に追加されるまで，以下を繰り返す．
   (a) $\mathbf{A}^{pk}$ の j 行目 $(j \in U - K)$ について，$i \in U - K \, (i \neq j)$ となる i 列目の全てが 0 となるような変数 x_j を探し，その集合を U_{exo} とする．U_{exo} が空集合でないならば，$U_c = U_{exo}$ とする．U_{exo} が空集合ならば，$\mathbf{A}^{pk}$ の j 行目 $(j \in U - K)$ について，$i \in U - K \, (i \neq j)$ となる i 列目の少なくとも 1 つが 1 であるような変数 x_j を探し，その集合を U_{end} として，$U_c = U - K - U_{end}$ とする．
   (b) 各 $j \in U_c$ について，$a_{ij}^{pk} = 0$ となるような添字 $i \in U - K \, (i \in j)$ の集合を $V^{(j)}$ で表す．まず $i \in V^{(j)}$ となる全ての i について $\mathbf{r}_i^{(j)} = \mathbf{x}_i$ とし，次に $i \in U - K - V^{(j)}$ となる全ての i について x_i を目的変数，x_j を説明変数として最小自乗回帰を行い，全ての $j \in U_c$ についてデータ行列 $\mathbf{X}$ から残差ベクトル $\mathbf{r}^{(j)}$ と残差データ行列 $\mathbf{R}^{(j)}$ を得る．U_c に含まれる変数が 1 つならば，その変数を x_m とする．U_c に含まれる変数が 2 つ以上ならば，U_c の中から残差に対して最も独立性の高い変数 x_m を探す．

   $$x_m = \arg\min_{j \in U_c} T(x_j; U - K) \tag{6}$$

   T は式 (5) で定義される独立性指標である．
   (c) m を K の末尾に加える．
   (d) $\mathbf{x} := \mathbf{r}^{(m)}$，$\mathbf{X} := \mathbf{R}^{(m)}$ とする．
3. 残った最後の変数の添字を K の末尾に加える．
4. 得られた順序 K に従い厳密な下三角行列 $\mathbf{B}$ を作り，元の変数ベクトル $\mathbf{x}$ とデータ行列 $\mathbf{X}$ に対して最小二乗法や最尤法など既存の共分散に基づく回帰手法を適用し，結合の強さ b_{ij} を推定する．

事前情報を全く与えない場合，pk-DirectLiNGAM アルゴリズムは元の DirectLiNGAM アルゴリズムと同じである．このアルゴリズムを実装した MATLAB コードは Web 上にあり，次の URL から入手できる: `http://www.ar.sanken.osaka-u.ac.jp/~inazumi/`.

4. 評価実験

4.1 実験条件

変数の数 $p = 5, 10, 30, 50$，サンプルサイズ $n = 100, 300, 500, 1000$ のそれぞれの組み合わせについて，51 個のデータセットを次の手順でランダムに生成した．

1. 親変数による変数 x_i の標準偏差が $[0.5, 1.5]$ の範囲におさまるように，$p \times p$ 次元の厳密な下三角行列 $\mathbf{B}$ をランダムに生成した．このときの変数ネットワークの構造は，変数間が完全に繋がっているものか疎なもののどちらかとした．そして，外的影響 e_i の標準偏差を $[0.5, 1.5]$ の範囲からランダムに選択した．
2. 種々の平均 0，分散 1 の非正規分布に従う外的影響 e_i をそれぞれ独立に選び，サンプルサイズ n のデータを生成した．まず最初に，平均 0，分散 1 の正規分布に従う変数 z_i を生成し，式 $e_i = \mathrm{sign}(z_i)|z_i|^{q_i}$ により非正規分布に従う変数に変換した．非正規化のための指数 q_i は $[0.5, 0.8] \cup [1.2, 2.0]$ の範囲からランダムに選択した．$[0.5, 0.8]$ の範囲にある q_i はサブガウシアン分布を生成し，$[1.2, 2.0]$ の範囲にある q_i はスーパーガウシアン分布を生成する．最後に，変換された変数 e_i を平均 0，分散 1 となるように正規化した．
3. 式 (2) の LiNGAM モデルに従い，観測変数 x_i を生成した．その後，x_i の順序をランダムに並び換えた．
4. $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$ を計算し，非ゼロとなる要素の値を全て 1 で置き換え，さらに対角要素を全て 0 とした．次に，この行列の対角部分を除く各要素に対し，確率 $1 - prop$ でその値を -1 とし，事前情報行列 $\mathbf{A}^{pk}$ を生成した．ここで，$prop = 0.0, 0.02, \ldots, 0.18, 0.2, 0.3, \ldots, 0.9, 1.0$ である．確率 $prop$ は隠されない要素の割合を表す．

観測変数 x_i の生成手順は文献 [Shimizu 06] のものと同じであり，correlation-faithful の仮定を満たすようにパラメータの値を選ぶことはしていない．

その後，生成したデータセットに pk-DirectLiNGAM アルゴリズムを適用した．各試行ごとに，まず pk-DirectLiNGAM アルゴリズムによって推定された因果的順序 K を用いて，真の

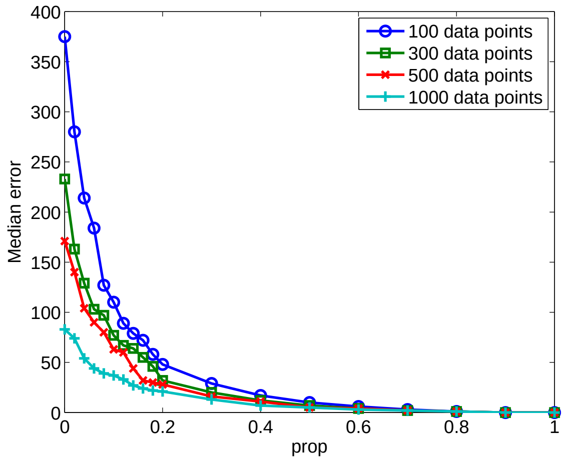

図 3　50 変数の場合のエラー数の中央値．横軸: 与えた事前情報の割合 *prop*．縦軸: エラー数の中央値．

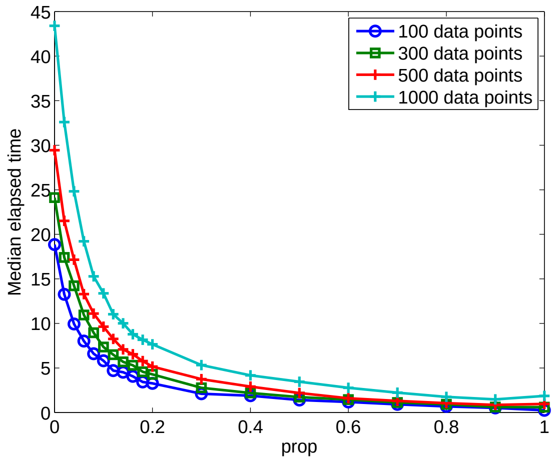

図 4　50 変数の場合の計算時間の中央値．横軸: 与えた事前情報の割合 *prop*．縦軸: 計算時間の中央値．

関係の強さを表す行列 $\mathbf{B}$ を並び換えた．そして厳密な上三角部分，すなわち対角要素を除いた上三角部分に，値が非ゼロとなる要素がいくつ存在するかを調べ，それをエラーの数とした．順序 K が正しく推定されていれば，厳密な上三角部分の要素の値は全てゼロである．また，計算時間として推定に要した時間を計測した．

4.2　結果

変数の数 $p = 50$ の場合の，エラー数の中央値を図 3 に示す．図 3 から，与えた事前情報の割合が増加するにつれて，エラー数の中央値が大きく減少していることがわかる．例えば，モデルの全構造の 10% のみを事前情報として与えた場合でも，エラー数の中央値は 50% 以上減少している．

次に，変数の数 $p = 50$ の場合の，計算時間の中央値を図 4 に示す．図 4 から，与えた事前情報の割合がそれほど大きくない場合でも，アルゴリズムの計算時間が大幅に短縮されていることがわかる．

4.3　考察

本実験ではランダムにどの変数に関する事前情報を与えるかを決定したため，例えば 10% の情報を与えた際，グラフ構造の仮定から 10% 以上の情報が与えられているという場合が起こり得る．例えば，$x_1 \to x_2$，$x_2 \to x_3$ の情報が与えられている場合，非巡回の仮定より x_1 と x_3 について $x_3 \to x_1$ の関係は成り立たない．さらに，$x_3 \to x_4$ の情報が与えられれば，$x_4 \to x_1$，$x_4 \to x_2$ の関係も成り立たない．このように，与える情報が増えれば，間接的に得られる情報も増える．仮に性能の向上がこのような間接的に得られる情報によるものだとすると，与える事前情報の割合が増えるにつれエラー数や計算時間は大きく減少するはずである．しかし，実験結果では与える事前情報が少ない場合のほうが減少の度合いは大きく，その度合いは徐々に小さくなっている．このことから，性能の向上は間接的に得られた情報によるものだけではなく，推定する必要がある関係が減ったことによるものとの両方による結果であると考えられる．

5.　まとめ

本稿では，因果推論分野における既存の非正規性に基づく手法に事前情報の利用を可能とする改良を加え，その性能を評価した．従来の非正規性に基づく手法では，いくらかの事前情報が存在する場合でも，その事前情報を推論に利用しなかった．本稿の成果により，事前情報がない場合は既存手法と同じ方法で推定を行い，事前情報がある場合はその情報も利用することで，より高精度かつ高速に推定することが可能となった．そして，実験結果により比較的わずかな情報を与えただけでも，推定精度および処理速度が十分に向上することを確認できた．

参考文献

[Bach 02] Bach, F. R. and Jordan, M. I.: Kernel independent component analysis, *J. Machine Learning Research*, Vol. 3, pp. 1–48 (2002)

[Bollen 89] Bollen, K. A.: *Structural Equations with Latent Variables*, John Wiley & Sons (1989)

[Gretton 05] Gretton, A., Bousquet, O., Smola, A. J., and Schölkopf, B.: Measuring Statistical Dependence with Hilbert-Schmidt Norms, in *Algorithmic Learning Theory: 16th Int. Conf. (ALT2005)*, pp. 63–77 (2005)

[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)

[Hyvärinen 01] Hyvärinen, A., Karhunen, J., and Oja, E.: *Independent component analysis*, Wiley, New York (2001)

[Pearl 00] Pearl, J.: *Causality: Models, Reasoning, and Inference*, Cambridge University Press (2000)

[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 09] Shimizu, S., Hyvärinen, A., Kawahara, Y., and Washio, T.: A direct method for estimating a causal ordering in a linear non-Gaussian acyclic model, in *Proc. 25th Conf. on Uncertainty in Artificial Intelligence (UAI2009)* (2009)

[Spirtes 93] Spirtes, P., Glymour, C., and Scheines, R.: *Causation, Prediction, and Search*, Springer Verlag (1993), (2nd ed. MIT Press 2000)