日本統計学会誌
第37巻，第2号，2008年3月
223頁～237頁

独立成分分析による線形逐次モデルの探索

清　水　昌　平*

Learning linear acyclic models using independent component analysis

Shohei Shimizu*

構造方程式モデルは，無作為割付を伴わない非実験データからの因果分析に最も頻繁に用いられている多変量解析の一つである．構造方程式モデルによる因果分析では通常，観測変数間の因果関係が線形逐次モデルで記述できると仮定する．そして線形逐次モデルのパラメータを推定することで，因果に関する情報を得ようとする．しかし，構造方程式モデルを含む伝統的な多変量解析は陰に陽に正規性の仮定に基づいており，モデル識別に共分散行列の情報しか用いないことが多い．そのため従来の線形逐次モデルの探索法は，多くの場合にモデルを一意に識別できない．一方，信号処理の分野で比較的最近提案された独立成分分析は，観測変数に非正規性が認められる場合に，その非正規性をモデル識別に積極的に利用する．本稿では，この独立成分分析のアイデアを線形逐次モデルの探索に適用した新しい探索方法について解説する．従来法と違い，この非正規性を利用する方法は観測変数間の線形逐次モデルを一意に識別できる．

Developing statistical methods to estimate causal structures of observed variables based on non-experimental (observational) data is an important topic of current research and is useful for many empirical studies including social sciences, bioinformatics and neuroinformatics. However, most of existing methods including structural equation models explicitly or implicitly assume a Gaussian assumption and utilize covariances alone for model identification, which leads to a number of indistinguishable models. We have recently proposed that a non-Gaussian assumption actually allows linear acyclic models for observed variables uniquely identified. The solution relies on the use of a statistical method known as independent component analysis. In this paper, we first review briefly structural equation models and independent component analysis. Then we describe the new non-Gaussian method for estimating linear acyclic models.

Key Words and Phrases: Independent component analysis, structural equation models, non-Gaussianity

1. はじめに

因果を吟味するための効果的な方法は，無作為割付を伴う実験（以下単に「実験」という）を行うことである（Holland, 1986）．しかし実際には，実験を行うことが困難な状況がある．というのは，実験を行うためには，因果に関する事前情報（例えば，因果の向きに関する情報）が必要であるが，そのような事前情報が十分にないことがあるからである．さらに，応用研究の性格上，コストや倫理的な観点から，実験が行えない事も多い．例えば，心理学や社会学などの社会科学（Bollen, 1989），バイオインフォマティクスにおける遺伝子ネットワーク推定（Imoto, Goto and Miyano, 2002），ニューロインフォマティクスにおける脳内部位結合性解析（Brain connectivity analysis）（Kim et al., 2007）等である．そこで，無作為割付を伴わない非

* 日本学術振興会特別研究員 PD（統計数理研究所）：〒106-8569 東京都港区南麻布 4-6-7

実験データ（観察データ）を用いて，因果に関する仮説を探索することが行われている．

構造方程式モデル（Bollen, 1989）は，観察データからの因果分析に最も頻繁に用いられている多変量解析の一つである．構造方程式モデルによる因果分析では通常，観測変数間の因果関係が線形逐次モデルで記述できると仮定する．そして線形逐次モデルのパラメータを推定することで，因果に関する情報を得ようとする．しかし，構造方程式モデルを含む伝統的な多変量解析は陰に陽に正規性の仮定に基づいており，モデルの識別に共分散行列の情報しか用いないことが多い．しかし，これが原因で識別可能なモデルが限られてしまうため，従来の線形逐次モデルの探索法（Chickering, 2002; Marcoulides and Drezner, 2001; Pearl, 2000; Spirtes, Glymour and Scheines, 2000）は，多くの場合にモデルを一意に識別できない．

一方，独立成分分析（Hyvärinen, Karhunen and Oja, 2001）は，信号処理の分野で提案された比較的新しい統計解析法である．独立成分分析は因子分析の拡張の一つと捉えることもできる．しかし通常の因子分析と異なり，観測変数に（強い）非正規性が認められる場合に，その非正規性をモデル識別に積極的に利用するところに特徴がある[1]．非正規性を利用することで，潜在因子が非正規分布に従いかつ互いに独立である場合に，因子回転の不定性を取り除くことができる．つまり因子負荷行列を本質的に一意に識別できる（Comon, 1994）．

本稿では，この独立成分分析のアイデアを線形逐次モデルの探索に適用した新しい探索方法（Shimizu, Hoyer, Hyvärinen and Kerminen, 2006）について解説する．この新しい方法は，従来法と異なり，観測変数間の線形逐次モデルを一意に識別できる．

もちろんこの非正規性を利用する方法は，観測変数が正規分布に従うか非常に近い場合には機能しない．しかし近年，ニューロインフォマティクスにおける磁気共鳴画像（fMRI）のような（強い）非正規性が認められるようなデータが増えつつある（Hyvärinen et al., 2001）．また，正規性の仮定に基づく構造方程式モデルが適用されることの多い社会科学の分野でも，正規分布では上手く近似できないデータが少なくないことが指摘されている（Micceri, 1989）．このような非正規性の認められるデータに対して，本稿で紹介する新しい探索法は重要な役割を果たすと思われる．

本稿の構成は次の通りである．まず第2節と第3節では，構造方程式モデルと独立成分分析の基本的事項を述べる．第4節では，従来の線形逐次モデルの探索手法とその限界について簡単に述べる．そして第5節において，観測変数の非正規性を利用する線形逐次モデルの探索法を解説する．第6節で本稿をまとめる．

2. 構造方程式モデル

この節では，構造方程式モデルについて簡単にまとめる．構造方程式モデルの重要な下位モデルには，線形逐次モデルや因子分析モデルがある．ここでは，本稿の主題である線形逐次モデルについてのみ述べる．より詳しくは，Bollen（1989）等を参照されたい．

2.1 線形逐次モデル

線形逐次モデルとは，データの発生機構に線形性と逐次性（ループがない）を仮定するモデルである．まず，m 次元の観測変数ベクトルを x，m 次元の誤差変数ベクトルを e と書く．観測変数 x_i の（因果的）順番を $k(i)$ で表す．例えば，2変数 x_i と x_j に対して $k(i)<k(j)$ なら，x_i は x_j から（因果的）影響を受けないとする[2]．このとき観測変数 x_i は，それより順番

[1] ここでの「非正規」とは「離散」変数のことではなく，「非正規」分布に従う「連続」変数のことを指す．

[2] 互いに影響を受けない場合は，$k(i)<k(j)$ も $k(i)>k(j)$ のどちらも成り立つが，どちらの順序を用いても本質的な違いはない．

が早い変数と誤差変数の線形和で書ける：

$$x_i=\sum_{k(j)<k(i)} b_{ij}x_j+e_i+c_i \tag{2.1}$$

ここで b_{ij} は x_j から x_i への影響の大きさを表す係数（パス係数）を表し，定数項 c_i は x_i の切片項を表す．誤差変数 e_i の平均はゼロとする．さらに，誤差変数 e_i は互いに独立であるとする．これは未観測交絡変数が存在しないことを意味する（Bollen, 1989）．

以下では（表記の）簡単のため，観測変数 x_i の平均はゼロとする，つまり $c_i=0$ とする．もし x_i がモデル内の他の変数から影響を受けない外生変数[3]であれば，$x_i=e_i$ となることに注意．

さて，観測変数 x_i，誤差変数 e_i，パス係数 b_{ij} をそれぞれ，ベクトル $\boldsymbol{x}$，$\boldsymbol{e}$，行列 $\mathbf{B}$ でまとめて表すことにする．すると，上述のモデル（2.1）は次のように書ける：

$$\boldsymbol{x}=\mathbf{B}\boldsymbol{x}+\boldsymbol{e} \tag{2.2}$$

ここで $\mathbf{B}$ は，$m\times m$ のパス係数行列（対角成分はすべてゼロ）である．逐次性の仮定より，パス係数行列 $\mathbf{B}$ は，観測変数の順序 $k(i)$ に従って行と列を並び替えると，下三角行列（かつ対角成分もゼロ）になる（Bollen, 1989）．線形逐次モデルを図で表現した例（パス図と言う）をいくつか図1に挙げる．

この線形逐次モデル（2.2）は，観測変数の順序 $k(i)$ が既知であれば，観測変数の共分散行列を用いて識別可能である（Bollen, 1989）．しかし，順序が全く未知であれば，識別可能ではない．というのは，観測変数の（重複を除いた）分散・共分散の数 $m(m+1)/2$ より，推定すべき $\mathbf{B}$ の非対角成分と e_i の分散の数 m^2 の方が大きくなるからである．

2.2 推定と適合度

構造方程式モデルにおける統計的推測は観測変数の共分散構造に基づいて行われる．そこでまず，線形逐次モデルの共分散構造を導く．式（2.2）を $\boldsymbol{x}$ について解くと次のように書ける[4]：

$$\boldsymbol{x}=(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{e} \tag{2.3}$$

そして，推定すべき未知パラメータであるパス係数 b_{ij} と誤差変数 e_i の分散をまとめて，ベクトル $\boldsymbol{\theta}$ で表そう．すると，観測変数の共分散構造 $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ は次のように書ける：

$$\begin{aligned}\boldsymbol{\Sigma}(\boldsymbol{\theta})&=\mathrm{cov}(\boldsymbol{x})\\&=(\mathbf{I}-\mathbf{B})^{-1}\mathrm{cov}(\boldsymbol{e})(\mathbf{I}-\mathbf{B}^T)^{-1}\end{aligned} \tag{2.4}$$

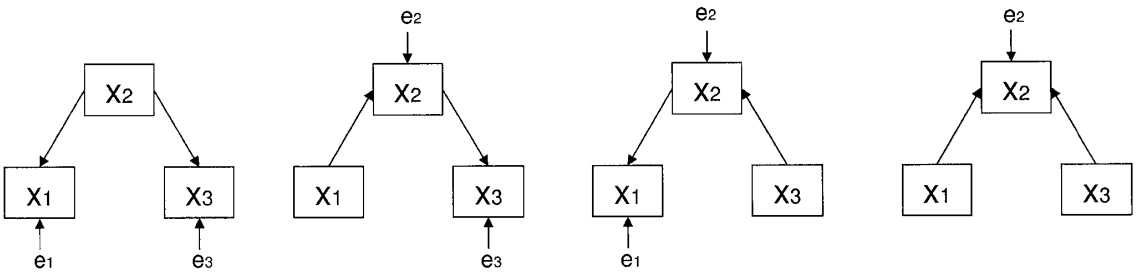

図1　線形逐次モデルの例

[3] モデル内の他の変数から影響を受ける変数を内生変数と言う．

[4] パス係数行列 $\mathbf{B}$ は（行と列を同時に並び替えて）下三角行列に並び替え可能な行列であり，$\mathbf{I}-\mathbf{B}$ の対角成分はすべて1だから，$\mathbf{I}-\mathbf{B}$ の逆行列は存在する．

この共分散構造はモデルのパラメータ $\boldsymbol{\theta}$ のみに依存しており，誤差変数 e_i の従う分布とは関係ない．

構造方程式モデルの推定には，観測変数に多変量正規性を仮定した最尤推定が用いられることが多い（Bollen, 1989）．標本サイズ n の無作為標本を $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ で表す．そして観測変数の共分散構造の標本推定値を $\mathbf{S}=1/(n-1)\sum_{k=1}^{n}(\boldsymbol{x}_k-\bar{\boldsymbol{x}})(\boldsymbol{x}_k-\bar{\boldsymbol{x}})^T$ とする．（ここで $\bar{\boldsymbol{x}}=1/n\sum_{k=1}^{n}\boldsymbol{x}_k$ は標本平均である．）このとき $\boldsymbol{\theta}$ の最尤推定量 $\hat{\boldsymbol{\theta}}_{ML}$ は次式：

$$F_{ML}(\boldsymbol{\theta})=\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})|-\log|\mathbf{S}|-\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\mathbf{S}-\boldsymbol{\Sigma}(\boldsymbol{\theta}))]) \tag{2.5}$$

を最小化する解である．なお，$\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})=\mathbf{S}$ のとき，$F_{ML}=0$ となる．

またモデルの適合度を評価することは構造方程式モデルの重要なポイントである．例えば，次の仮説を統計的に検定することによって適合度を評価できる（帰無仮説を H_0，対立仮説を H_1 と表す）：

$$H_0:\ \mathrm{cov}(\boldsymbol{x})=\boldsymbol{\Sigma}(\boldsymbol{\theta})\quad \text{versus}\quad H_1:\ \text{not } H_0 \tag{2.6}$$

そのための検定統計量には例えば $T_{ML}=n\times F_{ML}(\hat{\boldsymbol{\theta}}_{ML})$ が使える．この統計量は帰無仮説の下で漸近的に自由度 $u-v$ のカイ自乗分布で近似できる．ここで u は推定に利用される観測変数の分散・共分散の個数であり，v は未知パラメータの個数である．これ以外にも，情報量基準や記述的指標など数多くの適合度指標が提案されている（Bollen, 1989; Bollen and Long, 1993）．

2.3 飽和モデルと同値モデル

この節では，構造方程式モデルにおけるモデル比較に関する重要な概念である飽和モデルと同値モデルについて説明する．

飽和モデルとは，未知パラメータの数と観測変数の分散・共分散の数が等しいために常にデータに完全に適合するモデルのことを言う．つまりモデルの適否に関わらず，$\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})=\mathbf{S}$ となる．そのためモデルがデータに適合しているのかどうかに関する情報は何も得られない．

同値モデルとは，（モデルが異なるにも関わらず）同一の共分散構造を持つモデルのことである：

$$\boldsymbol{\Sigma}_1(\boldsymbol{\theta}_1)=\boldsymbol{\Sigma}_2(\boldsymbol{\theta}_2) \tag{2.7}$$

共分散構造が同一であるから，適合度指標も同一の値となり，どちらのモデルの方がデータに適合しているかはわからない．

最もシンプルな例は，次の2変数間の線形逐次モデルである：

$$\text{モデル1}：x_1=b_{12}x_2+e_1 \tag{2.8}$$

$$\text{モデル2}：x_2=b_{21}x_1+e_2 \tag{2.9}$$

この2つのモデルは，飽和モデルかつ同値モデルである（Shimizu and Kano, 2007）．したがって，これら2つの（構造の全く異なる）モデルを適合度の観点から区別することはできない．飽和モデルと同値モデルについて詳しくは，Bekker, Merckens and Wansbeek（1994）; Bollen（1989）を参照されたい．

2.4 構造方程式モデルにおける非正規性の利用

構造方程式モデルは通常，観測変数の共分散構造に基づいて統計的推測を行うが，非正規性（高次モーメント）を利用する研究も多くはないが存在する．

例えば，非正規母集団における統計的推測である．Browne（1982, 1984）は，標本分散・共分散の共分散行列（4次モーメントが含まれる）を重み行列とする一般化最小自乗法を提案し

た．この推定法はADF（asymptotically distribution-free）法と呼ばれ，観測変数の4次モーメントが有界という条件の下で，標本共分散行列に基づく推定量の中で漸近分散が最小となる．

しかし，この種の研究は，非正規性をモデルの識別に利用しているわけではない．共分散構造を用いて識別できないモデルは多い．Bentler（1983）は，観測変数が非正規分布に従っている場合に，その非正規性を利用することで，多くのモデルが識別可能になるのではないかと指摘した．続いて，Mooijaart（1985）はBentlerのアイデアを因子分析に適用し，「潜在因子が独立であり，その3次モーメントが非ゼロかつ互いに異なれば」因子回転が一意に定まることを示した．つまり，非正規性をモデル識別に利用することで，因子回転の不定性を解消できることを指摘したのである．実はこの方法は，次節で紹介する独立成分分析と同じアイデアであると言えるのだが，これ以上は発展しなかった．その理由はおそらく，構造方程式モデルの主たる応用分野である社会科学では，興味の対象である潜在因子は相関を持つのが普通だからではないかと思われる（Jennrich and Trendafilov, 2005）．一方，これとは独立に信号処理の分野で提案された独立成分分析は，音声分離や自然画像の分解など多くの分野に応用され，大きく発展することとなった（Hyvärinen et al., 2001）．

3. 独立成分分析

この節では，信号処理の分野で比較的最近提案された独立成分分析（Independent Component Analysis, ICA）について簡単にまとめる．詳細は，Hyvärinen et al.（2001）等を参照されたい．

3.1 モデルと識別性

独立成分分析（ICA）モデル（Comon, 1994; Jutten and Hérault, 1991）は次のように書ける：

$$\boldsymbol{x}=\mathbf{A}\boldsymbol{s} \tag{3.1}$$

ここで$\boldsymbol{x}$はm次元の観測変数ベクトル，$\boldsymbol{s}$はn次元の潜在変数ベクトル，$\mathbf{A}$は$m\times n$の定数行列を表す（混合行列と呼ばれる）．さらにs_iは独立な連続変数であるとする（独立成分と呼ばれる）．つまり，独立な源信号s_iが混合行列$\mathbf{A}$によって線形混合され，観測信号$\boldsymbol{x}$として測定されるという設定である．ICAの特徴は，観測変数の共分散行列だけでなく非正規性をモデルの識別に利用することにある．

Comon（1994）は，次の条件が成り立てば，ICAモデル（3.1）の混合行列$\mathbf{A}$が列の順番とスケールの不定性を除いて識別されることを示した：(1) s_iは独立であり，正規分布に従う成分は高々1つである．(2) s_iは退化していない．(3) 観測変数x_iの数は潜在変数s_iの数以上である（$m\geq n$）．(4) 混合行列$\mathbf{A}$は列フルランクである．

これは次のことを意味する．仮に$\boldsymbol{x}=\mathbf{A}\boldsymbol{s}=\mathbf{H}\boldsymbol{t}$としよう．ここで$\mathbf{A}$，$\mathbf{H}$，$\boldsymbol{s}$，$\boldsymbol{t}$は上記の条件を満たすとする．このとき$\mathbf{H}=\mathbf{APD}$（$\mathbf{P}$は置換行列，$\mathbf{D}$は対角行列）となり，混合行列$\mathbf{A}$は列の順番とスケールの不定性を除いて識別される．共分散行列のみをモデル識別に使う通常の因子分析と異なり，因子回転の不定性はない．

この結果は，独立成分の数が観測変数の数よりも多い場合（つまり$m<n$）（Eriksson and Koivunen, 2004）や，正規分布に従う誤差変数が存在する場合（Davies, 2004）に拡張されている．それらの場合も基本的に，s_iが非正規分布に従うなら，$\mathbf{A}$を列の順番とスケールの不定性を除いて識別できる．

以下では，一般性を失うことなく，独立成分の平均をゼロ，分散を1とする．また，観測変数の数は独立成分の数と等しい（$m=n$）とする．というのは，ICAの推定法のほとんどは，この$\mathbf{A}$が正方行列の場合を扱っており（Hyvärinen et al., 2001），さらに本稿で解説する探索

法（第5節）においても，この典型的な場合のICAが重要な役割を果たすからである．

3.2 推　定

ICAモデルの推定には数多くの提案がある（Cichocki and Amari, 2002; Hyvärinen et al., 2001）．多くの場合，混合行列 $\mathbf{A}$ を直接推定するのではなく，その逆行列 $\mathbf{W}=\mathbf{A}^{-1}$（復元行列と呼ばれる）を推定する．この流儀で言えば，ICAは復元信号ベクトル $\boldsymbol{y}=\mathbf{W}\boldsymbol{x}$ の成分が独立となるような復元行列 $\mathbf{W}$ を探すとも言える．ここでは主要なアプローチをいくつか簡単に紹介する．

まずは（擬）最尤法（Pham and Garrat, 1997）である．独立性の仮定より，復元信号ベクトル $\boldsymbol{y}=\mathbf{W}\boldsymbol{x}$ の密度関数は $p(\boldsymbol{y})=\prod_{i=1}^{m}p_i(y_i)$ と書けるから，観測ベクトル $\boldsymbol{x}$ の対数尤度 L は次のように書ける：

$$L(\mathbf{W})=\sum_{t=1}^{T}\sum_{i=1}^{m}\log p_i(\boldsymbol{w}_i^T\boldsymbol{x}_t)+T\log|\det\mathbf{W}| \tag{3.2}$$

ここで T は標本サイズ，$\boldsymbol{w}_i^T$ は $\mathbf{W}$ の第 i 行ベクトルを表す．最適化は自然勾配法（Amari, 1998; Cardoso and Laheld, 1996）がよく用いられる．実際には p_i は未知であるが，正確な分布形がわからず，適当な方法で決め打ったとしても一致性がある（最尤法と区別して擬最尤法と言う）．例えば，独立成分 s_i の尖度の正負に応じて，その符号と合うように密度関数 p_i を選択すればよい（Amari, Chen and Cichocki, 1997; Cardoso and Laheld, 1996）．

もう1つは，非正規性最大化法（Comon, 1994; Hyvärinen, 1999a）である．この方法では，復元信号 $y_i=\boldsymbol{w}_i^T\boldsymbol{x}$ の非正規性が最大になるような $\boldsymbol{w}_i$ を探す．直感的には次の様に意味づけされる．いくつもの非正規な独立成分が足し合わされると，中心極限定理によりその分布は正規分布に近づいていく．そこで，逆に出来るだけ単一の成分を抽出するには，非正規性を最大にするような変換を探せばよいという様にである（Hyvärinen et al., 2001）．

復元信号 $y_i=\boldsymbol{w}_i^T\boldsymbol{x}$ の非正規性を測るための基準として，次のクラスが提案されている（Hyvärinen, 1999a）：

$$J_G(\boldsymbol{w}_i)=[E\{G(y_i)\}-E\{G(\nu)\}]^2=[E\{G(\boldsymbol{w}_i^T\boldsymbol{x})\}-E\{G(\nu)\}]^2 \tag{3.3}$$

ここで $G(\cdot)$ は非線形かつ非2次の関数であり，ν は標準正規分布に従う変数を表す．仮に $G(y_i)=y_i^4$ とすれば，これは尖度になる．しかし4次モーメントは，はずれ値に対して敏感であることが知られている．そのため，はずれ値への頑健性を重視し，y_i の値が大きくなっても，あまり急激に大きくならないような $G(\cdot)$ がよく使われる．例えば，$G(y_i)=\log\cosh(y_i)$ である．

Hyvärinen（1999a）は，この非正規性基準を用いた推定法を提案した（FastICAと呼ばれる）．まず最初に，主成分分析などを用いて，観測変数を無相関化する．つまり，$\boldsymbol{z}=\mathbf{R}\boldsymbol{x}$ の共分散行列が単位行列となるような行列 $\mathbf{R}$ を定める（$\mathrm{cov}(z)=\mathbf{I}$）．そして復元信号ベクトル $\boldsymbol{y}=\mathbf{Q}\boldsymbol{z}$（$=\mathbf{QR}\boldsymbol{x}=\mathbf{W}\boldsymbol{x}$）の成分の非正規性の和を最大にする直交行列 $\mathbf{Q}$ を求める：

$$\hat{\mathbf{Q}}=\arg\max_{\mathbf{Q}\in O(m)}\sum_{i=1}^{m}J_G(\boldsymbol{q}_i) \tag{3.4}$$

ここで $\boldsymbol{q}_i^T$ は $\mathbf{Q}$ の第 i 行ベクトル，$O(m)$ は m 次元の直交行列のクラスを表す．この $\mathbf{R}$ と $\mathbf{Q}$ から，復元行列 $\mathbf{W}$ は $\mathbf{W}=\mathbf{QR}$ と推定できる．

この非線形関数 $G(\cdot)$ が次の条件を満たす十分滑らかな偶関数であるとき，この推定法は一致性がある：

$$E\{s_i g(s_i) - g'(s_i)\}[E\{G(s_i)\} - E\{G(\nu)\}] > 0 \tag{3.5}$$

ここで$g(\cdot)$は$G(\cdot)$の導関数，$g'(\cdot)$は$g(\cdot)$の導関数，νは標準正規分布に従う変数を表す．この条件は，かなり多くの$G(\cdot)$とs_iの分布について成り立つと言われる（Hyvärinen, 1999a）．

この非正規性最大化法は最尤法と深く関連している．例えば，独立成分s_iが同一の分布に従うと仮定し，$G(\cdot)=\log p(\cdot)$とすれば，最尤法とほぼ同じになる（Hyvärinen, 1999b）．また，このとき漸近分散が最小になる（Tichavský, Koldovský and Oja, 2006）．ただ，非正規性最大化法では，擬最尤法と違い，独立成分の尖度の正負に応じて非線形関数$G(\cdot)$を選ぶというようなことをしなくてよい．これは実用上，大きな利点である．

3.3 局所解，標本サイズ，非正規性

実際にICAの最適化を行う際にやっかいなのは，局所解の問題である．上述の非正規性最大化法では，非正規性を測る非線形関数を$G(y)=y^4$とすると，局所解がないことが証明される（Delfosse and Loubaton, 1995; Hyvärinen and Oja, 1997）．しかし，それ以外の非線形関数については必ずしもそうではなく，異なる初期値を用いると，異なる推定値を得る可能性がある．そのため，局所解の回避を試みる方法がいくつか提案されている（Himberg, Hyvärinen and Esposito, 2004; Tichavský et al., 2006）．例えばHimberg et al.（2004）は，ブートストラップ法（Efron and Tibshirani, 1993）を用いた対処法を提案している．その基本的なアイデアは次の通りである．まず複数の初期値を用意するとともに，リサンプリングを行うことで目的関数の形状を少し変化させ，複数のデータセットに対してICAの解を求める．そして，局所解であれば，リサンプリングすることでその解は消失しうるが，大域解であれば消失しないであろうと考え，安定して何度も推定される解を大域解と考える．

次に，標本サイズが少ないことによる過学習の問題がある（Hyvärinen, Särelä and Vigário, 1999）．例えば極端な場合，標本サイズと観測変数の数そして独立成分の数が等しければ（$T=m=n$），データ行列$\mathbf{X}$，復元信号の行列$\mathbf{Y}$，混合行列$\mathbf{A}$は3つとも同じ次元の正方行列になる．そのため，$\mathbf{A}$を変化させることで$\mathbf{Y}(=\mathbf{A}^{-1}\mathbf{X})$はどのような値にもなる．つまり，$\mathbf{Y}$はデータ行列$\mathbf{X}$に関係なく，目的関数に何を選ぶかによって決まってしまう（Särelä and Vigário, 2003）．このような過学習を避けるためには，十分大きい標本サイズを集める必要がある．はっきりとした基準を示すことは難しいが，Särelä and Vigário（2003）は経験上，推定するパラメータ数の少なくとも5倍（$T>5m^2$）は必要ではないかとしている．

非正規性を利用するICAの性能は，従来法と違い，標本サイズだけでなく非正規性の強さにも依存している．つまり，非正規性が強いほど推定は正確になる（Hyvärinen, 1998）と考えられるし，また実用上，正規分布に近い場合は機能しない（Pham and Garrat, 1997）．しかし，非正規性の強さと性能の関係についての理論的研究は，あまり進んでいないと思われる．

この正規分布に近い場合の悪さの原因の1つは，非正規性（独立性）を評価するとき，単一の非線形関数Gしか用いないことにあるかもしれない．大まかにはではあるが，上述の擬最尤法（独立成分が全て同一の分布に従う場合）と非正規性最大化法はどちらも，復元信号の非線形相関$\mathrm{cov}(g(y_i), y_i)$（$g$は$G$の導関数）ができるだけゼロにちかくなるような$\mathbf{W}$を探すと言える（Hyvärinen, 1999b）．しかし，これはもちろん独立性の十分条件ではない．そこでBach and Jordan（2002）は，より広いクラスの非線形相関$\mathrm{corr}(f(y_i), h(y_i))$（$f$, hは再生核ヒルベルト空間から選ぶ）を用いて独立性を評価する方法を提案した．この非線形相関が（絶対値の意味で）最大になるようなf, hに対して，その非線形相関ができるだけゼロに近くなるような復元行列$\mathbf{W}$を探すのである．これは，この関数のクラスにおける最悪評価を行うということであり，また復元信号の独立性を評価するために可能な限り適切な非線形関数を選択

しているとも言えるだろう．実際，正規分布に近いデータに対して，擬最尤法や非正規性最大化法などの従来法よりもかなり頑健であるという数値実験結果が報告されている（Bach and Jordan, 2002）．

そういうわけで実データの解析においては，標本サイズの小ささや正規分布に近いこと等が原因となり推定が不安定になっていないかを，例えばブートストラップ法に基づく方法（Himberg et al., 2004; Meinecke, Ziehe, Kawanabe and Müller, 2002）を用いて検討する必要があるだろう．

4. 従来の線形逐次モデルの探索法

本稿の主題である線形逐次モデルの探索とは，線形逐次モデル（2.2）のパス係数行列 $\mathbf{B}$ と観測変数の順序 $k(i)$ を「それらの値を知らずに」推定することを指す．この節では，従来の探索法とその限界について簡単にまとめる．多くの従来法（Neapolitan, 2003; Pearl, 2000; Spirtes et al., 2000）は，連続変数に正規性を仮定する．あるいは，非正規性分布を想定しても，やはり共分散行列のみをモデル識別に用いることが多い．

主要な方法の1つは，条件付独立性を利用する方法である（Spirtes et al., 2000）．まず，観測変数の中で，他の変数すべてを与えたとき条件付独立（条件付無相関）となる変数の組を探す．そして，線形逐次モデル（2.2）の構造に由来する独立性・条件付独立性以外は，観測変数の同時分布においても成り立たない（忠実性）と仮定し，観測変数間に成り立つ条件付独立性に矛盾するモデルを候補から除いていくのである．しかし，同値モデル（第2.3節）の問題があり，線形逐次モデルを一意に識別できないことが多い．例えば，図1の左3つのモデル $x_1 \leftarrow x_2 \rightarrow x_3$，$x_1 \rightarrow x_2 \rightarrow x_3$，$x_1 \leftarrow x_2 \leftarrow x_3$（$x_i \rightarrow x_j$ は x_i から x_j にパスがあることを表す）は同値モデルである．

もう1つは，適合度を比較する方法がある（Neapolitan, 2003）．基本的なアイデアは，すべての可能なモデルについて，つまり順序 $k(i)$ とパス係数 b_{ij} のゼロ非ゼロのパターンの全ての組み合わせについて適合度を求め，最良のモデルを選択するというものである．適合度には，モデルの自由度を考慮するためにAICやBIC等の情報量基準（正規分布が仮定される）が用いられることが多い．第2.1節で述べたように，順序 $k(i)$ が与えられれば，観測変数の共分散行列を用いて線形逐次モデル（2.2）は識別可能である．しかし，変数の数が増えるにしたがい，候補のモデルの数は爆発的に増加するため，総当りで適合度を比較するわけにはいかなくなる．そのため，計算量減少の観点から，様々な提案がある（Chickering, 2002; Marcoulides and Drezner, 2001; Tsamardinos, Brown and Aliferis, 2006）．しかしながら，共分散行列の情報のみを使う限り，やはり同値モデルの問題がある．例えば第2.3節の2変数間の線形逐次モデル1と2を適合度の観点からは区別できない．したがって，真のモデルがモデル1（$x_1 \leftarrow x_2$）であれば，モデル2（$x_1 \rightarrow x_2$）も最良のモデルの1つとして，選択されてしまう．

この2変数間の線形逐次モデル1と2の同値性を解消するために，Shimizu and Kano（2007）は，観測変数が非正規な場合に，その共分散行列だけでなく高次モーメントも用いて適合度を評価することを提案した．仮にモデル1が真のモデルとしよう．もし外生変数 x_2 あるいは誤差変数 e_1 の歪度あるいは尖度がゼロでなければ（正規分布に従う変数と異なれば），モデル1と2の3次モーメント構造あるいは4次モーメント構造は異なることが示される．したがって，この2つのモデルを適合度の観点から区別できる．この方法を3変数以上へ拡張するための1つの考え方は，総当りで適合度を比較することである．しかし，やはり変数の数が増えるにしたがって，候補のモデルの数は膨大になる．また，3変数以上の場合に，同値モデルの問題がどの程度解消されるのかは，まだよくわかっていない．おそらく，非正規変数の歪度・尖度の

値によって，高次（3次，4次）モーメント構造も同じになるモデル群があると思われる．

5. 非正規性を利用する探索法

この節では，観測変数の非正規性をモデル識別に利用し，線形逐次モデルを探索する方法を解説する．上述の従来法が，複数のモデルを比較しなければならない理由の一つは，観測変数の順序 $k(i)$ が全く未知であれば，共分散構造を用いるだけでは線形逐次モデル（2.2）を識別できないことにある．しかし，仮にパス係数行列 $\mathbf{B}$ が直接推定可能であるとすれば，そのようなモデル比較の過程を経る必要はない．例えば，$\mathbf{B}$ の第 i 行が全てゼロであれば，それは観測変数 x_i が他のどの変数からもパスを受けないことを意味するから，この x_i は外生的である．また b_{ij} のみが非ゼロであるなら，x_i は x_j からのみパスを受ける内生変数である．以下では，観測変数の非正規性をモデル識別に利用し，パス係数行列 $\mathbf{B}$ を直接推定する方法（Shimizu, Hoyer, et al., 2006）を紹介する．

5.1 LiNGAM モデル

やや重複するが，ここで考えるモデルを確認しよう．まず，m 次元の観測変数ベクトルを $\boldsymbol{x}$，m 次元の誤差変数ベクトルを $\boldsymbol{e}$ と書く．観測変数 x_i の（因果的）順番を $k(i)$ で表す．観測変数 x_i と誤差変数 e_i の平均はゼロとする．このとき，観測変数間の線形逐次モデルは次のように書ける：

$$\boldsymbol{x}=\mathbf{B}\boldsymbol{x}+\boldsymbol{e} \tag{5.1}$$

ここで $\mathbf{B}$ は，$m\times m$ のパス係数行列（対角成分はすべてゼロ）である．逐次性の仮定より，パス係数行列 $\mathbf{B}$ は，観測変数の順序 $k(i)$ に従って行と列を（同時に）並び替えると，下三角行列（かつ対角成分もゼロ）になる．さらに，誤差変数 e_i は互いに（無相関なだけでなく）独立であるとする．この独立性の仮定は，未観測交絡変数が存在しないことを意味する（詳しくは第 5.2 節）．

従来法では多くの場合，誤差変数 e_i（外生的観測変数が含まれていることに注意）に正規性を仮定する．しかし，これから紹介する探索法では，誤差変数 e_i は（未知の）非正規分布に従うと仮定する[5]．

このような線形性，非正規性，逐次性の 3 つの性質をもつモデルを Linear, Non-Gaussian, Acyclic Model（LiNGAM モデル）と呼ぶことにする．ここでの目的は，このモデルの無作為標本から，パス係数行列 $\mathbf{B}$ と観測変数の順序 $k(i)$ を「それらの値を知らずに」推定することである．

5.2 なぜ無相関でなく独立か？

この節では，線形逐次モデルにおいて「共分散行列」と「無相関」だけでなく「非正規性」と「独立性」が重要になる理由を簡単に説明しよう．観察データによる因果分析で最もやっかいなのは未観測交絡変数の存在である．未観測交絡変数を z で表し，

$$\begin{aligned} x_2&=b_{21}x_1+g_{23}z+e_2\\ x_1&=\qquad\quad g_{13}z+e_1 \end{aligned} \tag{5.2}$$

としよう．このとき，観測変数 x_1 と x_2 の共分散は

$$\mathrm{cov}(x_1, x_2)=b_{21}\mathrm{var}(x_1)+g_{23}g_{13}\mathrm{var}(z) \tag{5.3}$$

---

[5] 適当な次元のモーメントは存在すると仮定する．

となる．未観測交絡変数zからx_1, x_2へのパス係数g_{13}, g_{23}の値によっては，たとえb_{21}がゼロでも，この$\mathrm{cov}(x_1, x_2)$は非ゼロになりうるし，またb_{21}が十分大きな値でもゼロになりうる．このことが，x_1, x_2の関係について分析者を誤った結論に導いてしまうことがある．これが未観測交絡変数が問題になる理由である．

もう少し詳しく説明する．もし未観測交絡変数zに気づかずに分析をするとしよう．このとき

$$\begin{aligned} x_2 &= \beta x_1 + e_2^* \\ e_2^* &= (b_{21} - \beta) x_1 + g_{23} z + e_2 \end{aligned} \tag{5.4}$$

とモデル（5.2）を書き直し，$\beta = \mathrm{cov}(x_1, x_2)/\mathrm{var}(x_1)$とすれば，説明変数$x_1$と誤差変数$e_2^*$を「無相関」にすることができてしまう：

$$\begin{aligned} \mathrm{cov}(x_1, e_2^*) &= (b_{21} - \beta)\mathrm{var}(x_1) + g_{23}\mathrm{cov}(x_1, z) + \mathrm{cov}(x_1, e_2) \\ &= b_{21}\mathrm{var}(x_1) - \mathrm{cov}(x_1, x_2) + g_{23}g_{13}\mathrm{var}(z) + 0 \\ &= 0 \quad （式（5.3）より） \end{aligned} \tag{5.5}$$

しかし，e_2^*はx_1の関数であるから，x_1とe_2^*は「独立」でない．未観測交絡変数が存在すれば，説明変数と誤差変数の間に従属性が生まれる．したがって説明変数と誤差変数が「独立」であれば，それは未観測交絡変数が存在しないことを示唆する．しかし「無相関」だけでは十分でないことが上記の議論からわかる（Shimizu, Hyvärinen, et al., 2006）．単に無相関なだけなのか，それとも独立なのか．これを検討するには，非正規性を利用する必要がある．そういうわけで，「非正規性」と「独立性」が鍵となるのである．

5.3 独立成分分析によるモデルの識別

非正規性を用いて線形逐次モデルを復元するための鍵は，観測変数x_iが非正規独立な誤差変数e_iの線形混合であると認識することである．LiNGAMモデル（5.1）を$\boldsymbol{x}$について解くと，次のモデルが得られる：

$$\mathbf{x} = \mathbf{A}\mathbf{e} \tag{5.6}$$

ここで$\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$である．この$\mathbf{A}$も（$\mathbf{B}$と同様に）下三角行列に並び替えることができる．ただしこの場合，並び替えられた行列の対角成分は全て非ゼロとなる．このモデル（5.6）は誤差変数$\mathbf{e}$の成分が非正規かつ独立であることに着目すると，独立成分分析（ICA）モデルである．したがって，$\mathbf{W}(= \mathbf{A}^{-1} = \mathbf{I} - \mathbf{B})$を用いて

$$\boldsymbol{e} = \mathbf{W}\boldsymbol{x} \tag{5.7}$$

とも書ける．この場合も，$\mathbf{W}$は下三角行列に並び替えることができ，そのときの対角成分は全て非ゼロとなる．以下では，この$\mathbf{W}$を使って話を進めよう．

ICAは，復元行列$\mathbf{W}$を本質的に識別可能であるが，行の順序とスケールの不定性がある．ICAのほとんどの応用（例えば音声分離や自然画像の分解）では，行の（独立成分の）順序の不定性は重要でなく無視してもよい．しかしLiNGAMモデルでは，独立成分（ここでは誤差変数e_i）の正しい順序を見つける必要がある．というのは，パス係数b_{ij}を適切に解釈するためには，観測変数x_iと誤差変数e_i（$\mathbf{W}$の列と行）の順序が正しく対応していなければならないからである．しかし，行の順序とスケールの不定性のため，復元行列は$\mathbf{W}_{ica} = \mathbf{P}_{ica}\mathbf{D}_{ica}\mathbf{W}$（$\mathbf{P}_{ica}$は（未知の）置換行列，$\mathbf{D}_{ica}$は（未知の）対角行列）と推定される．したがって，一般に観測変数x_iと誤差変数e_iの順序が正しく対応しない．誤差変数e_i（$\mathbf{W}_{ica}$の行）の正しい順序

を見つけるには，$\tilde{\mathbf{P}}\mathbf{P}_{ica}=\mathbf{I}$ となるような置換行列 $\tilde{\mathbf{P}}$ を求める必要がある．そのような置換行列は，$\tilde{\mathbf{P}}\mathbf{W}_{ica}$ の対角成分にゼロが来ないような $\tilde{\mathbf{P}}$ であることが証明される（Shimizu, Hoyer, et al., 2006）．つまり，並べ替えられた復元行列の対角成分に1つでもゼロがあれば，それは観測変数と誤差変数の順序の対応が正しく取れていないことを意味する（図2）．これは，逐次性を仮定していることが効いている．

誤差変数（$\mathbf{W}_{ica}$ の行）の正しい順序がわかれば，$\mathbf{D}_{ica}\mathbf{W}$ が得られる．ICA の推定では通常，スケールの不定性に対処するため，独立成分の分散を1に固定する．そのため $\mathbf{D}_{ica}$ は独立成分の分散を1にするような対角行列である．ここで $\mathbf{D}_{ica}\mathbf{W}$ の各行を対応する対角成分で割ってやれば，$\mathbf{W}(=\mathbf{I}-\mathbf{B})$ が得られる．パス係数行列 $\mathbf{B}$ は，$\mathbf{B}=\mathbf{I}-\mathbf{W}$ で推定できる．

最後に，観測変数の順序 $k(i)$ の推定である．逐次性の仮定より，順序 $k(i)$ に従って $\mathbf{B}$ を（行と列を同時に）並び替えると下三角行列（かつ対角もゼロ）になる．そこで逆に，$\tilde{\mathbf{P}}\mathbf{B}\tilde{\mathbf{P}}^T$ が下三角行列（かつ対角成分もゼロ）になるような $\tilde{\mathbf{P}}$ を求めれば，そのときの $\tilde{\mathbf{P}}\mathbf{x}$ の成分の順序が $k(i)$ となる[6]．

5.4 LiNGAM アルゴリズム

前節ではパス係数行列 $\mathbf{B}$ と順序 $k(i)$ が識別可能であることを説明したが，実際にデータから推定するためには，もういくつか述べておかなくてはならないことがある．

まずは，ICA の推定を実行するアルゴリズムの選択である．基本的に，標準的なアルゴリズム（Amari, 1998; Hyvärinen, 1999a）なら，どれでも使うことができる．しかし，できるだけ広いクラスの非正規分布に対応するアルゴリズムを選択する必要がある．誤差変数 e_i の分布を事前に知ることは難しいからである．例えば，上述（第3節）の FastICA（Hyvärinen, 1999a）は，擬最尤法と違い，独立成分の尖度の正負を知らずに ICA モデルを推定することができる．

次は，復元行列 $\mathbf{W}_{ica}$ の行の並び替えについてである．仮に復元行列が推定誤差なしに推定できれば，対角成分が全て非ゼロとなるような行の順序を見つけることは容易い．しかし，実際には推定誤差があるので，その値がゼロであるパス係数でも推定値はゼロにはならない．したがって実際には，対角成分に絶対値の小さな値が来ないような行の順序を探す．例えば，$\sum_i 1/|\tilde{\mathbf{W}}_{ii}|$（$\tilde{\mathbf{W}}=\tilde{\mathbf{P}}\mathbf{W}_{ica}$, $\tilde{\mathbf{W}}_{ii}$ は $\tilde{\mathbf{W}}$ の第 i 対角成分）が最小になるような置換行列 $\tilde{\mathbf{P}}$ を探す．この最適化は，典型的な線形割当て問題（linear assignment problem）（Burkard and Cela, 1999）

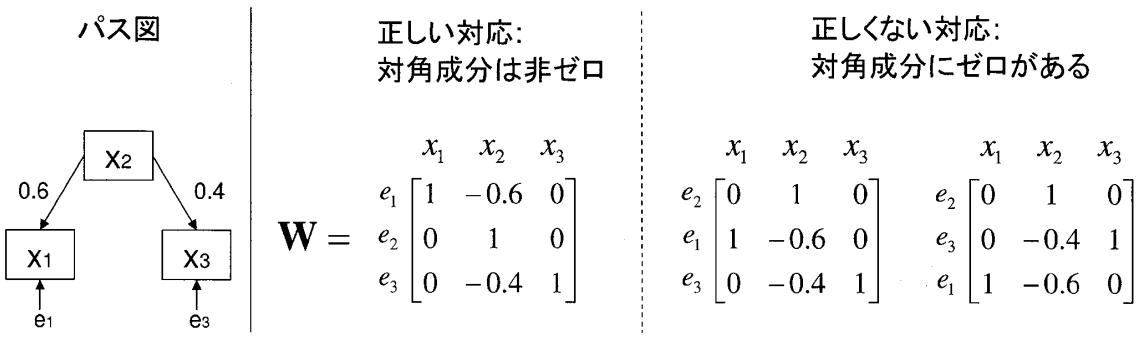

図2　観測変数（$\mathbf{W}$ の列）と誤差変数（$\mathbf{W}$ の行）の順序の対応の例

[6] 互いに影響を受けない変数がある場合は，観測変数の順序は一意ではない．例えば，図2のモデルの x_1 と x_3 については，$k(1)<k(3)$ も $k(1)>k(3)$ のどちらも成り立つ．しかし，どちらの順序を用いても，x_1 と x_3 の間のパス係数（b_{31} あるいは b_{13}）は結局ゼロになるから，本質的な違いはない．この意味で順序 $k(i)$ は本質的に一意に定まる．

である．この目的関数は，この直感的な意味づけに加えて，最尤法の観点からも意味づけされる（Shimizu, Hoyer et al., 2006）．

さて次に，順序 $k(i)$ を推定するためのパス係数行列 $\mathbf{B}$ の下三角化についてである．先ほどと同様に，もしも推定誤差がないなら，下三角化するための並び替えは自明であるが，実際には推定誤差があるので，下三角行列（かつ対角成分もゼロ）にできるだけ近くなるように並び替えることになる．目的関数としては例えば，上三角部分と対角成分の二乗和 $\sum_{i\leq j}\tilde{\mathbf{B}}_{ij}^2$（ここで $\tilde{\mathbf{B}}=\bar{\mathbf{P}}\mathbf{B}\bar{\mathbf{P}}^T$）が使えるだろう．この目的関数の最小化については，観測変数の数が比較的小さいとき（おおよそ8以下）は，単純に総当りで解を求めることができる．しかし，より高次元の場合は別の適切な方法（Hoyer, Shimizu, Hyvärinen et al., 2006）が必要になる．

以上の考察に基づいて，Shimizu, Hoyer et al.（2006）は，ICAを用いてパス係数行列 $\mathbf{B}$ を推定するためのアルゴリズム（LiNGAM アルゴリズム）を提案した：

LiNGAM アルゴリズム

1．まず，ICA（例えば，FastICA）を用いて，復元行列 $\mathbf{W}$ を推定する．
2．推定された復元行列 $\mathbf{W}_{ica}$ の行を並び替えて，対角成分が非ゼロとなるような行列 $\tilde{\mathbf{W}}$ を見つける．ここでは $\sum_i 1/|\tilde{\mathbf{W}}_{ii}|$（$\tilde{\mathbf{W}}=\tilde{\mathbf{P}}\mathbf{W}_{ica}$）を最小にする置換行列 $\tilde{\mathbf{P}}$ を探す．
3．$\tilde{\mathbf{W}}$ の各行を対応する対角成分で割り，$\tilde{\mathbf{W}}$ を基準化する．この基準化された行列（対角成分が全て1）を $\tilde{\mathbf{W}}'$ で表す．つまり，$\tilde{\mathbf{W}}'=diag(\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}$ である．
4．$\hat{\mathbf{B}}=\mathbf{I}-\tilde{\mathbf{W}}'$ によって $\mathbf{B}$ の推定値 $\hat{\mathbf{B}}$ を計算する．
5．最後に，観測変数の順番 $k(i)$ を見つけるために $\hat{\mathbf{B}}$ の置換行列 $\bar{\mathbf{P}}$ を探す．この置換行列 $\bar{\mathbf{P}}$ は $\hat{\mathbf{B}}=\bar{\mathbf{P}}\hat{\mathbf{B}}\bar{\mathbf{P}}^T$ が下三角行列（対角成分もゼロ）にできるだけ近づくように定める．ここでは，目的関数として $\sum_{i\leq j}\tilde{\mathbf{B}}_{ij}^2$ を使う．

このアルゴリズムを実装するMatlabとOctaveのコードが次のWebページからダウンロードできる：http://www.cs.helsinki.fi/group/neuroinf/lingam/

このコードを用いた数値実験の結果を簡単に紹介する．詳細はShimizu, Hoyer et al.（2006）を参照されたい．観測変数の数は20，標本サイズは200，1000，2000とした．各誤差変数には，それぞれ異なる非正規分布（尖度が正の分布と負の分布の両方を含む）を用いた．パス係数行列 $\mathbf{B}$ の構造（ゼロ非ゼロのパターン）とその値および順序 $k(i)$ はランダムに生成した．LiNGAM アルゴリズムによる推定結果を示した散布図が図3である．それぞれの散布図について，上述のようにして5種類のデータセットを発生させた．散布図の横軸がパス係数 b_{ij} の

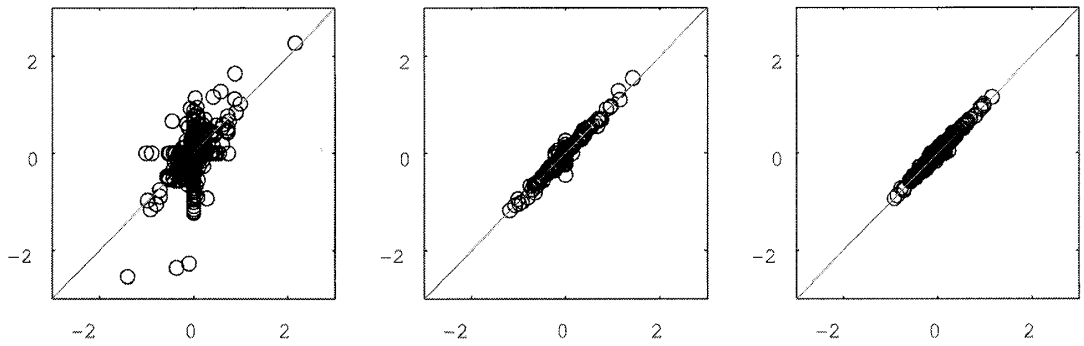

図3　左から順に，標本サイズ200，1000，2000．横軸がパス係数 b_{ij} の真値，縦軸がその推定値を表す．散布図を明瞭にするため，1000個のデータ点を無作為に選び，プロットした．

真値，縦軸がその推定値を表しているので，各データ点が主対角に集まれば集まるほど推定が正確であることを示している．この20変数の例では，標本サイズ200では不十分だが，1000以上あれば十分ではないかと思われる．

さて，上述の復元行列の並び替えについては，依然としてかなり発見的な方法であるため，改良の必要があるだろう．また，復元行列の推定値は観測変数のスケールの影響を受けるから，あらかじめ観測変数を標準化しておくことも，尺度不変になりよいと思われる．

最後に，余分なパスの枝刈りについて述べる．観測変数の順序 $k(i)$ を推定できれば，逐次性の仮定から，$\mathbf{B}$ を下三角に並び替えた時の上三角部分と対角部分（つまり $\tilde{\mathbf{B}}$ の上三角部分と対角部分）のパス係数はゼロとおくことができる．しかし，それ以外の下三角部分のパス係数は（通常）非ゼロである．その中には，真の値がゼロであるパス係数も含まれているだろう．そのようなパス係数は枝刈りしたい（ゼロとおきたい），あるいは枝刈りする必要がある．第2.1節で述べたように，観測変数の順序 $k(i)$ がわかれば，線形逐次モデルは共分散構造のみを用いて識別可能である．そのため，このような枝刈りを実行するために，既に多くの方法が利用可能である．例えば，ブートストラップ法，各種の適合度指標やパスの有意性検定（Bollen, 1989）を用いることができる．また，パス係数にスパース性を課す回帰分析 Lasso（Tibshirani, 1996）のアイデアを用いることもできるだろう．例えば，観測変数 x_i を目的変数，それに対してパスを持つ観測変数を説明変数とする Lasso を，各観測変数に対して繰り返し実行するのである．また，そもそも最初の ICA の推定の段階からスパース性を課すということも提案されている（Zhang and Chan, 2006b）．

6. おわりに

無作為割付を伴わない非実験データ（観察データ）に基づく因果分析法を開発することは，多くの応用分野が考えられ，非常に重要である．例えば，社会科学（Bollen, 1989）やバイオインフォマティクス（Imoto et al., 2002），そしてニューロインフォマティクス（Kim et al., 2007）等の応用分野がある．もちろん，推定される因果モデルの妥当性（例えば線形逐次モデルが本当に因果を表しているかどうか）を観察データのみから検証することはできない．しかし，コストや倫理的な観点から実験を行うことが難しい状況では，観察データに基づく分析法は有益な情報を与えうると思われる．

連続変数の線形逐次モデルを探索するための従来法（Bollen, 1989; Spirtes et al., 2000; Pearl, 2000）は，陰に陽に正規性の仮定に基づいていており，共分散行列の情報しかモデル識別に用いないことが多い．そのため，データから区別できない同値モデルの問題を抱えている．本稿で紹介した探索法（LiNGAM アルゴリズム）は，観測変数に非正規性が認められる場合に，その非正規性をモデル識別に利用することで，従来法の欠点を克服し線形逐次モデルを一意に識別できる．さらに，この方法の拡張もいくつか議論され始めている．例えば，未観測交絡変数（Hoyer, Shimizu and Kerminen, 2006）や潜在クラス（Shimizu and Hyvärinen, 2006）のある場合，そして非線形（Zhang and Chan, 2006a, 2007）への拡張である．

この LiNGAM アルゴリズムの価値は，より大規模な数値実験や実際の応用研究において，従来の分析法と比較することで評価する必要がある．また，多くの応用研究では，様々な背景知識が利用可能であることが多い．そういった追加の情報をどのように取り込むか（例えば事前分布を導入する（Zhang and Chan, 2006b）等）ということも興味深い話題である．これらは，これからの重要な課題である．

謝　辞

原稿の段階でコメントをいただいた南美穂子先生，藤澤洋徳先生に厚くお礼申し上げます．

また，貴重なコメントをいただいた査読者に感謝いたします．この研究の一部は，新領域融合研究センターおよびHelsinki Institute for Information Technology（フィンランド）で行われました．

参考文献

Amari, S. (1998). Natural gradient learning works efficiently in learning, *Neural Computation*, **10**, 251–276.

Amari, S., Chen, T.-P. and Cichocki, A. (1997). Stability analysis of learning algorithms for blind source separation, *Neural Networks*, **10** (8), 1345–1351.

Bach, F. R. and Jordan, M. I. (2002). Kernel independent component analysis, *J. of Machine Learning Research*, **3**, 1–48.

Bekker, P. A., Merckens, A. and Wansbeek, T. (1994). *Identification, equivalent models and computer algebra*. Boston: Academic Press.

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: specification and estimation of moment structures, *Psychometrika*, **48**, 493–517.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley and Sons.

Bollen, K. A. and Long, J. S. (1993). *Testing Structural Equation Models*. Sage.

Browne, M. W. (1982). Covariance structures, In D. M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis* (p. 72–141). Cambridge: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures, *British Journal of Mathematical and Statistical Psychology*, **9**, 665–672.

Burkard, R. E. and Cela, E. (1999). Linear assignment problems and extensions, in P. M. Pardalos and D. Z. Du (Eds.), *Handbook of Combinatorial Optimization - Supplement Volume A* (pp. 75–149). Kluwer.

Cardoso, J.-F. and Laheld, B. H. (1996). Equivariant adaptive source separation, *IEEE Trans. on Signal Processing*, **44**, 3017–3030.

Chickering, D. (2002). Optimal structure identification with greedy search, *J. of Machine Learning Research*, **3**, 507–554.

Cichocki, A. and Amari, S. (2002). *Adaptive Blind Signal and Image Processing: Learning algorithms and applications.* New York, NY, USA: John Wiley and Sons, Inc.

Comon, P. (1994). Independent component analysis. a new concept?, *Signal Processing*, **36**, 62–83.

Davies, M. (2004). Identifiability issues in noisy ICA, *IEEE Signal Processing Letters*, **11**(5), 470–473.

Delfosse, N. and Loubaton, P. (1995). Adaptive blind separation of independent sources: a deflation approach, *Signal Processing*, **45**, 59–83.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Eriksson, J. and Koivunen, V. (2004). Identifiability, separability and uniqueness of linear ICA models, *IEEE Signal Processing Letters*, **11**, 601–604.

Himberg, J., Hyvärinen, A. and Esposito, F. (2004). Validating the independent components of neuroimaging time-series via clustering and visualization, *Neuroimage*, **22**, 1214–1222.

Holland, P. W. (1986). Statistics and causal inference (with discussion), *J. of the American Statistical Association*, **81**, 945–970.

Hoyer, P. O., Shimizu, S., Hyvärinen, A., Kano, Y. and Kerminen, A. (2006). New permutation algorithms for causal discovery using ICA, in *Proc. Int. Conf. on Independent Component Analysis and Blind Signal Separation, Charleston, SC, USA* (pp. 115–122).

Hoyer, P. O., Shimizu, S. and Kerminen, A. (2006). Estimation of linear, non-gaussian causal models in the presence of confounding latent variables, in *Proc. the third European Workshop on Probabilistic Graphical Models (PGM2006)* (pp. 155–162).

Hyvärinen, A. (1998). Independent component analysis for time-dependent stochastic processes, in *Proc. Int. Conf. on Artificial Neural Networks (ICANN98), Skövde, Sweden* (pp. 541–546).

Hyvärinen, A. (1999a). Fast and robust fixed-point algorithms for independent component analysis, *IEEE Trans. on Neural Networks*, **10**, 626–634.

Hyvärinen, A. (1999b). The fixed-point algorithm and maximum likelihood estimation for independent component analysis, *Neural Processing Letters*, **10**, 1–5.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*. New York: Wiley.

Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis, *Neural Computation*, **9**, 1483-1492.

Hyvärinen, A., Särelä, J. and Vigário, R. (1999). Bumps and spikes: Artifacts generated by independent component analysis with insufficient sample size, in *Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA99), Aussois, France* (pp. 425-429).

Imoto, S., Goto, T. and Miyano, S. (2002). Estimation of genetic networks and functional structures between genes by using Bayesian network and nonparametric regression, in *Proc. Pacific Symposium on Biocomputing* (Vol. 7, pp. 175-186).

Jennrich, R. I. and Trendafilov, N. T. (2005). Independent component analysis as a rotation method: A very different solution to Thurston's box problem, *British Journal of Mathematical and Statistical Psychology*, **58**, 199-208.

Jutten, C. and Hérault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing*, **24**(1), 1-10.

Kim, J., Zhu, W., Chang, L., Bentler, P. M. and Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data, *Human Brain Mapping*, **28**, 85-93.

Marcoulides, G. A. and Drezner, Z. (2001). Specification searches in structural equation modeling with a genetic algorithm, in G. A. Marcoulides and R. E. Schumaker (Eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 247-268). Lawrence Erlbaum Associates.

Meinecke, F., Ziehe, A., Kawanabe, M. and Müller, K.-R. (2002). A resampling approach to estimate the stability of one-dimensional or multidimensional independent components, *IEEE Trans. on Biomedical Engineering*, **49** (12), 1514-1525.

Micceri, T. (1989). The unicorn, the normal curve and other improbable creatures, *Psychological Bulletin*, **105**(1), 156-166.

Mooijaart, A. (1985). Factor analysis for non-normal variables, *Psychometrika*, **50**, 323-342.

Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Prentice Hall.

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

Pham, D. T. and Garrat, P. (1997). Blind separation of mixture of independent sources through a quasi-maximum likelihood approach, *Signal Processing*, **45**, 1457-1482.

Särelä, J. and Vigário, R. (2003). Overlearning in marginal distribution-based ICA: Analysis and solutions, *J. of Machine Learning Research*, **4**, 1447-1469.

Shimizu, S., Hoyer, P. O., Hyvärinen, A. and Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery, *J. of Machine Learning Research*, **7**, 2003-2030.

Shimizu, S. and Hyvärinen, A. (2006). Discovery of linear acyclic models in the presence of latent classes using ICA mixtures, *in NIPS* 2006 *Workshop on Causality and Feature Selection, Whistler, Canada*.

Shimizu, S., Hyvärinen, A., Hoyer, P. O. and Kano, Y. (2006). Finding a causal ordering via independent component analysis, *Computational Statistics and Data Analysis*, **50**(11), 3278-3293.

Shimizu, S. and Kano, Y. (2007). Use of non-normality in structural equation modeling: Application to direction of causation, *J. of Statistical Planning and Inference*. (In press)

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction and Search, 2nd ed.* MIT Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *J. of the Royal Statistical Society: Series B*, **58** (1), 267-288.

Tichavský, P., Koldovský, Z. and Oja, E. (2006). Performance analysis of the FastICA algorithm and Cramèr-Rao bounds for linear independent component analysis, *IEEE Trans. on Signal Processing*, **54**(4), 1189-1203.

Tsamardinos, I., Brown, L. E. and Aliferis, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm, *Machine Learning*, **65**, 31-78.

Zhang, K. and Chan, L.-W. (2006a). Extensions of ICA for causality discovery in the Hong Kong stock market, in *Proc. 13th Int. Conf. on Neural Information Processing (ICONIP 2006), Hong Kong*.

Zhang, K. and Chan, L.-W. (2006b). ICA with sparse connections, in *Proc. 7th Int. Conf. on Intelligent Data Engineering and Automated Learning (IDEAL 2006), Burgos, Spain*.

Zhang, K. and Chan, L.-W. (2007). Nonlinear independent component analysis with minimal nonlinear distortion, in *Proc. 24th Int. Conf. on Machine learning (ICML2007), Corvalis, Oregon* (pp. 1127-1134).