

非ガウスモデルにおける統計的因果構造のベイズ的探索

Bayesian discovery of statistical causal structure in non-Gaussian model

数学専攻 辻 響実

TSUJI, Kyomi

1 はじめに

従来からある統計的な方法として、因果関係を見つけ出す「統計的因果推論」とよばれる技術がある。統計的因果推論は、変数間の因果関係をデータから推測するための方法論であり、とくに、変数間の因果構造を未知としてどのような条件で因果構造が推測可能なのかを明らかにする技術を統計的因果探索とよぶ。LiNGAM (linear non-Gaussian acyclic model) は, Shimizu (2006) において提案された因果探索における代表的なモデルである。本稿では、これらの LiNGAM を用いた因果構造の推測、及びその推測にベイズ的アプローチを用いる BayesLiNGAM について述べる。

2 統計的因果探索

統計的因果探索は、統計的因果推論の技術の 1 つであり、代表的なモデルである LiNGAM には、因果構造を一意に推測できるという特長がある。

2.1 LiNGAM

観測変数 $x_i (i = 1, \dots, p)$ の LiNGAM モデル (linear non-Gaussian acyclic model) は、前の変数に割り当てられている値の線形関数に誤差変数 e_i を加えて

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

で表される。観測変数 $x_i (i = 1, \dots, p)$ は、順番が後の変数が前の変数を引き起こさないように、因果的順序を配置でき、このような因果的順序を $k(i)$ で示す。つまり、生成過程は再帰的であり、これは有向非巡回グラフ (DAG) によってグラフィカルに表すことができる。さらに、誤差変数 e_i はすべて非ゼロ分散の非ガウス分布をもつ連続確率変数であり、互いに独立している。つまり、

$$p(e_1, \dots, e_p) = \prod_i p_i(e_i)$$

である。

また、この LiNGAM モデルは行列を用いて

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e} \tag{2.1}$$

と表せる。観測変数ベクトル \mathbf{x} と誤差変数ベクトル \mathbf{e} はそれぞれ、観測変数 x_i と誤差変数 $e_i (i = 1, \dots, p)$ を、正方行列 \mathbf{B} は係数 $b_{ij} (i, j = 1, \dots, p)$ をまとめて表している。そして、係数行列 \mathbf{B} は観測変数の順序 $k(i)$ にしたがって行と列を並び替えると、下三角行列になる (ただし、対角成分も 0 になる)。

2.2 独立成分分析を用いたモデル推定

式 (2.1) の LiNGAM モデルを観測変数ベクトル \mathbf{x} について解くと,

$$\begin{aligned} (\mathbf{I} - \mathbf{B})\mathbf{x} &= \mathbf{e} \\ \iff \mathbf{x} &= (\mathbf{I} - \mathbf{B})^{-1}\mathbf{e} \\ \iff \mathbf{x} &= \mathbf{A}\mathbf{e} \end{aligned} \quad (2.2)$$

となる. ここで, $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}$ である. LiNGAM モデルの仮定より, これは独立成分分析モデル (Hyvärinen & Oja, 2000) と解釈できる. 行列 \mathbf{A} は独立成分分析モデルの混合行列, 行列 \mathbf{A} の逆行列 $\mathbf{W} (= \mathbf{A}^{-1})$ は独立成分分析モデルの復元行列にあたる. そして, 係数行列 \mathbf{B} とは

$$\mathbf{W} = \mathbf{I} - \mathbf{B}$$

という関係が成り立つ.

LiNGAM モデルにおいては, 係数 b_{ij} の適切な解釈のために, 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る必要があるが, 独立成分分析では, 行の順序と尺度の不定性があるため, 復元行列は

$$\mathbf{W}_{\text{ICA}} = \mathbf{P}\mathbf{D}\mathbf{W} \quad (2.3)$$

と推定される. ここで, 行列 \mathbf{P} は行の順序を表す置換行列, 行列 \mathbf{D} は行の尺度を表す対角行列である. 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. そこで, 誤差変数 e_i (\mathbf{W}_{ICA} の行) の正しい順序を見つけるには, 行列 $\tilde{\mathbf{P}}\mathbf{W}_{\text{ICA}}$ の対角成分に絶対値の小さな値がこないような置換行列 $\tilde{\mathbf{P}}$, つまり

$$\sum_i \frac{1}{|\tilde{\mathbf{W}}_{ii}|}$$

を最小にするような置換行列 $\tilde{\mathbf{P}}$ を探す. ただし, $\tilde{\mathbf{W}} = \tilde{\mathbf{P}}\mathbf{W}_{\text{ICA}}$ とし, $\tilde{\mathbf{W}}_{ii}$ は $\tilde{\mathbf{W}}$ の第 i 対角成分を表す.

誤差変数の正しい順序がわかれば,

$$\begin{aligned} \tilde{\mathbf{P}}\mathbf{W}_{\text{ICA}} &= \tilde{\mathbf{P}}\mathbf{P}\mathbf{D}\mathbf{W} \\ &= \mathbf{D}\mathbf{W} \end{aligned} \quad (2.4)$$

が得られる. 独立成分分析モデルにおける推定では通常, 尺度の不定性に対処するために独立成分の分散を 1 に固定するため, 行列 \mathbf{D} 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, $\tilde{\mathbf{W}}$ の各行を対応する対角成分で割って基準化し,

$$\tilde{\mathbf{W}}' = \text{diag}(\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}$$

で表す. したがって,

$$\hat{\mathbf{B}} = \mathbf{I} - \tilde{\mathbf{W}}'$$

によって, 係数行列 \mathbf{B} の推定値 $\hat{\mathbf{B}}$ が求まる.

次に, 推定した係数行列 \mathbf{B} が厳密な下三角行列になるような因果的順序の推定である. 式 (2.1) の LiNGAM モデルにおいて, 置換行列 $\bar{\mathbf{P}}$ を用いて

$$\begin{aligned} \bar{\mathbf{P}}\mathbf{x} &= \bar{\mathbf{P}}\mathbf{B}\mathbf{x} + \bar{\mathbf{P}}\mathbf{e} \\ &= (\bar{\mathbf{P}}\mathbf{B}\bar{\mathbf{P}}^T)\bar{\mathbf{P}}\mathbf{x} + \bar{\mathbf{P}}\mathbf{e} \end{aligned}$$

と式変形すると、これは置換行列 $\bar{\mathbf{P}}$ によって並び替えた観測変数ベクトル $\bar{\mathbf{P}}\mathbf{x}$ の LiNGAM モデルとして解釈できる。そこで、 $\bar{\mathbf{P}}\hat{\mathbf{B}}\bar{\mathbf{P}}^T$ が下三角行列にできるだけ近くなるような置換行列 $\bar{\mathbf{P}}$ 、つまり

$$\hat{\mathbf{P}} = \min_{\bar{\mathbf{P}}} \sum_{i \leq j} \tilde{\mathbf{B}}_{ij}^2$$

となるような置換行列 $\hat{\mathbf{P}}$ を探す。ここで、 $\tilde{\mathbf{B}} = \bar{\mathbf{P}}\hat{\mathbf{B}}\bar{\mathbf{P}}^T$ である。

こうして因果的順序を推測できたら、次はその順序に従って回帰分析を行い、係数行列 \mathbf{B} を推定する。まず、各観測変数 x_i を目的変数に、その親変数候補 x_j を説明変数にとり、線形回帰分析をする。ここで、変数 x_i の親変数候補とは、因果的順序が $k(j) < k(i)$ となるような変数 x_j の集合のことである。すると、各変数 x_i の構造方程式

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

の右辺にある係数 b_{ij} は、 x_i から x_j への偏回帰係数によって推定できる。

3 BayesLiNGAM

BayesLiNGAM(Hoyer & Hyttinen, 2009) とは、LiNGAM モデルの推定においてベイズ的アプローチを用いる推定法であり、データが与えられた際に事後確率が最大となる DAG を選択し、出力する。

3.1 LiNGAM におけるベイズモデル

いま、観測データセットを $D = \{\mathbf{x}^1, \dots, \mathbf{x}^N\}$ 、考えられ得る DAG を $G_k (k = \{1, \dots, N_g\})$ とする。 N_g は n 個の変数上の異なる DAG の数を表す。また、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。このとき、ベイズの定理より、事後確率は

$$P(G_k|D) = \frac{p(D|G_k)P(G_k)}{p(D)} \quad (3.1)$$

と表される。

まず、事前分布 $P(G_k)$ は事前情報を表す。ここでは一様分布を仮定し

$$P(G_k) = \frac{1}{N_g}$$

とする。また、 $p(D)$ は事後確率を正規化する定数であり、

$$p(D) = \sum_{m=1}^{N_g} p(D|G_m)P(G_m)$$

と計算できる。そして、尤度は

$$p(D|G_k) = \int p(D|\theta, G_k)p(\theta|G_k)d\theta \quad (3.2)$$

と書ける。ここで、 θ はすべてのパラメータ（すなわち係数 b_{ij} 、および外乱密度 $p_i(e_i)$ ）をまとめた変数である。

3.2 ラプラス近似による事後分布の近似

式 (3.2) を近似するにあたり, 本稿ではその積分の近似にラプラス近似を用いる. そこで, 式 (3.2) において積分の中身を $\theta = \tilde{\theta}$ (モード) のまわりで Taylor 展開すると,

$$p(D|\theta, G_k)p(\theta|G_k) \approx \exp \left\{ \log p(D|\tilde{\theta}, G_k)p(\tilde{\theta}|G_k) + (\theta - \tilde{\theta})^T b + \frac{1}{2}(\theta - \tilde{\theta})^T B(\theta - \tilde{\theta}) + \dots \right\}$$

となる. ただし, $p(D|\theta, G_k)p(\theta|G_k) = f(\theta)$ としたとき

$$b_i = \frac{\partial}{\partial \theta_i} \log f(\theta) \Big|_{\theta=\tilde{\theta}}, \quad B_{i,j} = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(\theta) \Big|_{\theta=\tilde{\theta}}$$

とする. これを整理すると, 式 (3.2) は

$$p(D|G_k) = \int C \cdot \exp \left\{ \frac{1}{2}(\theta - \tilde{\theta})^T B(\theta - \tilde{\theta}) \right\} d\theta$$

と書ける. ここで, ガウス分布の確率密度関数の性質を用いて, 式 (3.2) のラプラス近似は

$$p(D|G_k) \approx p(D|\tilde{\theta}, G_k)p(\tilde{\theta}|G_k) \sqrt{\frac{(2\pi)^{\dim \theta}}{|-B|}}$$

となる.

3.3 推定

尤度 $p(D|G_k)$, 事前分布 $P(G_k)$, 正規化定数 $p(D)$ の値をそれぞれ式 (3.1) に代入して, 事後確率 $P(G_k|D)$ は

$$P(G_k|D) = \frac{p(D|\tilde{\theta}, G_k)p(\tilde{\theta}|G_k) \sqrt{\frac{(2\pi)^{\dim \theta}}{|-B|}}}{N_g \sum_{m=1}^{N_g} p(D|G_m)P(G_m)}$$

と計算できる.

4 おわりに

本研究では, 統計的因果探索において因果構造を一意に推測可能な LiNGAM モデル, およびベイズ推論を使用した BayesLiNGAM について述べた.

今後の研究課題として, BayesLiNGAM における θ に関する事前分布 $p(\theta|G_k)$ の妥当な設定について, 及びマルコフ連鎖モンテカルロ (MCMC) 法など他の手法での事後確率の近似の検証があげられる.

参考文献

- [1] Hoyer, P.O. and Hyttinen, A. (2009). Bayesian discovery of linear acyclic causal models. In *Proc. 25th Conference on Uncertainty in Artificial Intelligence*, pp.240-248.
- [2] Hyvärinen, A., Oja, E. (2000). Independent Component Analysis : Algorithms and Applications. *Neural Networks*, 13(4-5), pp.411-430.
- [3] Shimizu, S., Hoyer, P.O., Hyvärinen, A. and Kerminen, A. (2006). A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7, pp.2003-2030.