修士論文 要旨

ダイバージェンスを用いた外れ値に頑健な因果探索手法

数理情報学専攻 48166226　野間 修平
指導教員　駒木 文保 教授

1　はじめに

本研究では外れ値が混入したデータから変数間の因果関係を推測する手法を提案する．

2　因果探索問題

因果探索問題とは，データから変数間の因果関係を推測する問題である．因果探索問題を定式化する代表的なモデルとして，Linear Non-Gaussian Acyclic Models (以下 LiNGAM) がある [4] ．これは変数間の因果関係をデータの生成過程として定義するモデルであり，効率的な推定アルゴリズムが提案されている [2, 5] ．

LiNGAM から生成される p 次元の確率ベクトル x は B を p 次の正方行列，e を p 次元の確率ベクトルとしたとき，

$$x = Bx + e \iff x = (I_p - B)^{-1} e$$

に従って生成される．ただし，e は外生変数と呼ばれ，各成分は互いに独立に連続型非ガウス分布に従う．また，B は係数行列と呼ばれ，ある置換行列 P が存在して $P^\top BP$ が対角成分が全て 0 である下三角行列となる．係数行列 B を隣接行列にもつ有向グラフを因果グラフと呼ぶ．このとき，係数行列 B の性質から，因果グラフは有向非巡回となり，トポロジカル順序（到達可能性から定義される半順序）が存在する．トポロジカル順序と矛盾しない全順序 k を LiNGAM の因果的順序と呼ぶ．LiNGAM について次の定理が成り立つ．

定理 1　p 次元の確率ベクトル x が LiNGAM に従うとする．いま，相異なる $i, j \in [p]$ 及び因果的順序 k を任意にとる．このとき，$k(i) > k(j)$ であるならば x_i を x_j で回帰したときの残差 $r_i^j = x_i - \frac{\mathrm{Cov}(x_i, x_j)}{\mathbb{V}[x_j]} x_j$ と x_j は独立である [2] ．

この定理より，任意の $i \in [p]$ に対して x_j と残差 r_i^j が独立であるならば，ある因果的順序 k が存在して $k(j) = 1$ が成り立つ．この性質を用いることで Hyvärinen らは LiNGAM の因果的順序を推定する次のようなアルゴリズムを提案している [5] ．

1. $V \leftarrow [p]$，$L \leftarrow$ (空リスト) と初期化
2. $j^* \leftarrow \left(\text{任意の } i \in V \setminus \{j\} \text{ に対し } x_j \perp\!\!\!\perp r_i^j \text{を満たす } j\right)$
3. L の末尾に j^* を追加
4. $V \leftarrow V \setminus \{j^*\}$
5. $x_i \leftarrow r_i^{j^*}$ $(i \in V)$ として 2 へ

因果的順序 k が推定されれば，係数行列 B の各成分は回帰を用いて推定することが出来る [4] ．確率変数の独立性は相互情報量を用いて評価するのが一般的である．確率変数 v, ξ の相互情報量は次のように定義される：

$$I(v, \xi) \overset{\text{def}}{=} \int p(v, \xi) \log \frac{p(v, \xi)}{p(v)\, p(\xi)} \mathrm{d}v \mathrm{d}\xi$$

相互情報量は非負であり，引数が独立であるときに限り 0 となる．相互情報量はカーネル密度推定によって推定された密度 $\hat{p}$ を用いて次のように推定される：

$$I(v, \xi) \approx \frac{1}{n} \sum_{i \in [n]} \log \frac{\hat{p}(v, \xi)}{\hat{p}(v)\, \hat{p}(\xi)}.$$

しかし，一般に 2 変数以上の同時分布を精度よく推定することは難しいという問題がある．Hyvärinen らは [5] で，相互情報量の差

$$\begin{aligned} m(x_j \to x_i) &\overset{\text{def}}{=} I\left(x_i, r_j^i\right) - I\left(x_j, r_i^j\right) \\ &= H(x_i) + H\left(r_j^i\right) - H(x_j) - H\left(r_i^j\right) \end{aligned} \tag{1}$$

を独立性の指標とすることで，同時分布の推定を回避した．

3　ロバスト推定

ロバスト推定とは外れ値に対して頑健な推定手法の総称である．いま，興味のある分布の密度を $f_\theta = f(\cdot|\theta)$，外れ値が従う密度を ρ，混入する外れ値の割合を ϵ としたとき，観測は密度

$$g = (1 - \epsilon) f_\theta + \epsilon \rho$$

に従って生成されるものとする．このとき，ロバスト推定とは g から生成された観測を用いて興味のある分布のパラメータ θ を推定することを指す．ここで，ダイバージェンスに基づくロバスト推定の手法を紹介する．g, f を任意の密度としたとき，γ-ダイバージェンス [3] 及び β-ダイバージェンス [1] はそれぞれ

$$\begin{aligned} D_\gamma[g|f] &\overset{\text{def}}{=} \frac{1}{\gamma(1+\gamma)} \log \int g(x)^{1+\gamma} \mathrm{d}x \\ &- \frac{1}{\gamma} \log \int g(x) f(x)^\gamma \mathrm{d}x + \frac{1}{1+\gamma} \log \int f(x)^{1+\gamma} \mathrm{d}x \end{aligned}$$

$$D_\beta[g|f] \stackrel{\text{def}}{=} \frac{1}{\beta(1+\beta)}\int g(x)^{1+\beta}\,\mathrm{d}x - \frac{1}{\beta}\int g(x)\,f(x)^\beta\,\mathrm{d}x + \frac{1}{1+\beta}\int f(x)^{1+\beta}\,\mathrm{d}x$$

で定義される．また，$D_\gamma[g|f_\theta]$，$D_\beta[g|f_\theta]$ の推定値を θ の関数とみたとき，その最小点を θ の推定量として定義する．

4　提案手法

本研究では LiNGAM に従う確率ベクトルが外れ値が混入するかたちで観測されたとき，LiNGAM の因果的順序を外れ値に対して頑健に推定する手法を提案する．相互情報量は同時分布から周辺分布の積への距離を Kullback-Leibler ダイバージェンスを用いて測った値である．この関係から類推し，γ-相互情報量 $I_\beta(\cdot,\cdot)$ 及び β-相互情報量 $I_\gamma(\cdot,\cdot)$ をそれぞれ $I_\gamma(v,\xi) \stackrel{\text{def}}{=} D_\gamma[p(v,\xi)|p(v)\,p(\xi)]$，$I_\beta(v,\xi) \stackrel{\text{def}}{=} D_\beta[p(v,\xi)|p(v)\,p(\xi)]$ で定義する．一方，相互情報量はエントロピー $H(\cdot)$ を用いて $I(v,\xi) = H(v)+H(\xi)-H(v,\xi)$ と書くことが出来た．右辺のエントロピーをエントロピーの拡張である Rényi エントロピー

$$H_\alpha(u) \stackrel{\text{def}}{=} \frac{1}{1-\alpha}\log\int p(u)^\alpha\,\mathrm{d}u$$

で置き換えたものを α-相互情報量と定義する．

本研究では新たな独立性の指標として，式 1 の右辺にある相互情報量 $I(\cdot,\cdot)$ を，これら 3 種類の量で置き換えることを提案する．このように定義した独立性の指標は推定に際して同時分布の推定を必要としない．これらの手法をそれぞれ G-DIV, B-DIV, A-MI と呼ぶことにする．ただし，どの手法も独立性の指標はカーネル密度推定を用いて推定し，残差 r_i^j は γ-ダイバージェンスを用いて推定する．従って，どの提案手法も独立性の指標の計算に用いる γ,β,α と，残差の推定に用いる γ_{reg} の 2 種類のハイパーパラメータを設定する必要がある．

5　数値実験

人工データに対してこれら 3 種類の手法を適用した．ただし，観測の次元及び個数はそれぞれ $p=3$, $n=200$ とし，外生変数の各成分 e_i はそれぞれ独立に分散 1 のラプラス分布に従う．外れ値は混合正規分布 (混合比 1 : 1, 平均 $-5, 5$, 分散 $1^2, 1^2$) に従い，汚染率は $\epsilon = 0.05$ とする．係数行列 B は対角成分が 0 である下三角行列で，非零成分は一様分布 $\mathcal{U}((-0.6,0.2)\cup(0.2,0.6))$ に

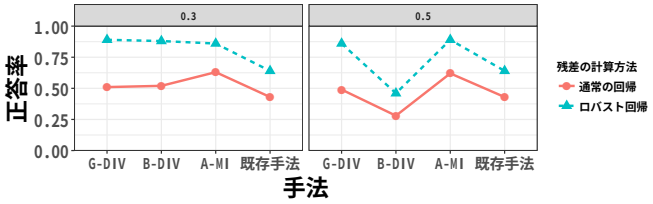

図 1. 人工データに対して各推定手法を適用した際の正答率．横軸は推定手法，縦軸は正答率を表す．左図は $\theta_{\text{MI}}=0.3$，右図は $\theta_{\text{MI}}=0.5$ とした場合に対応する．また，残差を通常の回帰と同様に計算したものを実線で，γ-ダイバージェンスを用いて推定したものを破線で表す．

従う．推定は $N=100$ 回行った．各手法の正答率 (因果的順序を正しく推定できた推定の割合) を図 1 に示した．独立性の指標の計算に用いるハイパーパラメータ $\theta_{\text{MI}} = \gamma, \beta, \alpha - 1$ は $0.3, 0.5$ の 2 通りに設定した．残差の推定に用いるハイパーパラメータ γ_{reg} は 0.7 とした．

独立性の指標をそのままに残差をロバスト推定する手法と，残差を通常と同様に計算しながら独立性の指標を頑健なものに置き換えた手法はどちらも正答率で既存手法を上回り，$\theta_{\text{MI}}=0.3$ とした実験では全ての提案手法が既存手法を優越した．しかし，B-DIV の正答率は $\theta_{\text{MI}}=0.5$ とした実験で既存手法の正答率を下回り，ハイパーパラメータの設定に強く影響されることが観察される．

参考文献

[1] A. Basu, I. Harris, N. Hjort and M. Jones: Robust and efficient estimation by minimising a density power divergence. *Biometrika*, vol. 85, (1998), pp. 549–559.

[2] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. Hoyer and K. Bollen: DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research*, vol. 12, (2011), pp. 1225–1248.

[3] H. Fujisawa and S. Eguchi: Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, vol. 99, (2008), pp. 2053–2081.

[4] S. Shimizu, P. Hoyer, A. Hyvärinen and A. Kerminen: A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, vol. 7, (2006), pp. 2003–2030.

[5] A. Hyvärinen and M. Stephen: Pairwise likelihood ratios for estimation of non-Gaussian structural equation models. *Journal of Machine Learning Research*, vol. 14, (2013), pp. 111–152.

[6] A. Rényi: On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, (1961).