

マーケティング・リサーチにおける 統計的因果探索を用いた因果仮説構築に関する研究

データサイエンス研究科, 株式会社マクロミル

2020 年 11 月 30 日

概要

本研究では,

目次

1	序論	2
1.1	研究背景	2
1.2	既存手法・研究	2
1.3	研究目的	2
1.4	本論文の構成	2
2	モデルと識別可能性	3
2.1	数学的準備	3
2.2	LiNGAM	4
2.3	2 次分散関数 (QVF) DAG モデル	5
2.4	QVF-DAG モデルの識別可能性	6
2.5	提案モデル	8
2.6	提案モデルの識別可能性	8
	謝辞	9

1 序論

あああ

1.1 研究背景

あああいいいあああ

1.2 既存手法・研究

aaaaaa

1.2.1 マーケティング・リサーチにおける既存手法

aaa

1.2.2 統計的因果探索

aaa

1.3 研究目的

あああ aaa

1.4 本論文の構成

ああああああ

2 モデルと識別可能性

本章ではまず、本論文で用いる数学記号を導入し、非巡回有向グラフ (Directed Acyclic Graph, DAG) モデルとその識別可能性を定義する。その後、本論文の提案モデルを構成する 2 つの既存モデルについて概説する。既存モデルの 1 つ目は、連続変数を扱う DAG モデルで、Shimizu *et al.*(2006)[5] によって提案された線形非ガウス非巡回モデル (Linear Non-Gaussian Acyclic Model; LiNGAM) である。2 つ目は、主に離散変数を扱う DAG モデルで、Park and Raskutti(2017)[3] によって提案された 2 次分散関数 (Quadratic Variance Function, QVF) DAG モデルである。QVF-DAG モデルの識別可能性は、Park and Raskutti(2017)[3] によって過分散スコアを用いて証明されたが、本論文では Park and Park(2019)[1] が提案したモーメント比スコアを拡張することで、識別可能条件の緩和を行う。最後に、連続変数と離散変数が混合したデータにおける DAG モデルを提案し、その識別可能性について議論する。

2.1 数学的準備

グラフは頂点 (node) の集合 $V = \{1, 2, \dots, p\}$ と、頂点同士をつなぐ辺 (edge) の集合 $E \subset V \times V$ によって、 $G = (V, E)$ と表現される。グラフの辺は有向辺 (矢線) と無向辺 (双方向矢線) に分けることができ、2 つの頂点 $j, k \in V$ において、 $(j, k) \in E$ かつ $(k, j) \notin E$ のとき、 j から k への矢線があるという。これを $j \rightarrow k$ と表現することもある。一方で、 $(j, k) \in E$ かつ $(k, j) \in E$ のとき、 j と k の間に双方向矢線があるという。すべての辺が有向辺であるグラフを有向グラフ (directed graph) という。本論文では、特に断りのない限り、頂点 j から k への矢線がある場合、 j が k の原因であるといった因果関係があることを表すとする。つまり、本論文で扱うグラフにおける矢線の有無は因果関係の有無を表しており、矢線の始点が原因で、矢線の終点が結果である。このような定性的な因果関係を表すグラフを因果グラフ (causal graph) という。また、グラフ G からすべての矢印を取り除くことによって得られるグラフを G のスケルトンという。

頂点の系列 $\alpha_1, \alpha_2, \dots, \alpha_{n+1}$ について、すべての $i = 1, 2, \dots, n$ で、 $\alpha_i \rightarrow \alpha_{i+1}$ 、または $\alpha_{i+1} \rightarrow \alpha_i$ となる矢線がある時、長さ n の道 (path) という。特に、すべての $i = 1, 2, \dots, n$ で、 $\alpha_i \rightarrow \alpha_{i+1}$ となる矢線がある時、長さ n の有向道 (directed path) という。また、長さ n の有向道で、 $\alpha_1 = \alpha_{n+1}$ となるものを巡回閉路 (cycle) という。一方で、巡回閉路のない有向グラフは非巡回的 (acyclic) であるという。本論文では、非巡回有向グラフ (Directed Acyclic Graph; DAG) のみを扱う。

頂点 j から k への矢線がある時、 j を k の親 (parent) といい、 k を j の子 (child) という。また、 $(j, k) \in E$ であるすべての頂点 j からなる集合を $Pa(k)$ と表記する。頂点 j から k への有向道がある時、 j を k の祖先 (ancestor)、 k を j の子孫 (descendant) という。頂点 k のすべての祖先からなる集合を $An(k)$ 、すべての子孫からなる集合を $De(k)$ と表記する。また、すべての頂点から k と k の子孫を除いたものを、 k の非子孫 (non-descendant) といい、その集合を $Nd(k) \equiv V \setminus (\{k\} \cup De(k))$ と表記する。さらに、因果的順序 (causal ordering) について定義する。因果的順序とは、その順序に従って変数を並び替えると、すべての矢線 $(j, k) \in E$ について、 k が j の原因になることがない順序のことであり、 $\pi = (\pi_1, \dots, \pi_p)$ と表記する。DAG で表現される因果グラフには、このような順序が (一意とは限らないが) 存在するという特徴がある。つまり、因果グラフを同定することは、因果的順序を同定することとスケルトンを同定することという 2 つの工程に分解することができる。

有向グラフ G における頂点上の標本空間 \mathcal{X}_V の確率分布に従う確率変数の集合 $X \equiv (X_j)_{j \in V}$ について考

える。ここで、確率変数ベクトル X は、同時確率密度関数 $f_G(X) = f_G(X_1, X_2, \dots, X_p)$ で与えられていると仮定する。 V の任意の部分集合 S について、 $X_S \equiv \{X_j : j \in S \subset V\}$ と $\mathcal{X}_S \equiv \times_{j \in S} \mathcal{X}_j$ を定義する。ただし、 \mathcal{X}_j は X_j の確率空間である。また、任意の頂点 $j \in V$ について、確率変数ベクトル X_S を与えたときの変数 X_j の条件付き確率を $f_j(X_j|X_S)$ と表記する。すると、DAG G によるモデルは以下のように因数分解することができる [4]。

$$f_G(X) = f_G(X_1, X_2, \dots, X_p) = \prod_{j=1}^p f_j(X_j|X_{Pa(j)}) \quad (1)$$

ここで、 $f_j(X_j|X_{Pa(j)})$ は、 X_j の親変数 $X_{Pa(j)} \equiv \{X_k : k \in Pa(j) \subset V\}$ を与えた条件付き確率である。

また、本論文では観察データから因果グラフを同定するという問題を扱うため、因果グラフの識別可能性について定義する。識別可能性を直感的に説明すると、条件付き確率分布 $f_j(X_j|X_{Pa(j)})$ に対してある仮定を置くと、同時確率密度関数 $f_G(X)$ を与えた DAG G の構造を一意に決定付けることができるということである。

識別可能性について詳細に定義するために、すべての $j \in V$ に関する条件付き確率分布 $f_j(X_j|X_{Pa(j)})$ の集合を \mathcal{P} と表記する。また、グラフ $G = (V, E)$ について、グラフ G に関する同時分布のクラスと、分布 \mathcal{P} のクラスを以下で定義する。

$$\mathcal{F}(G; \mathcal{P}) \equiv \{f_G(X) = \prod_{j \in V} f_j(X_j|X_{Pa(j)}); \text{ where } f_j(X_j|X_{Pa(j)}) \in \mathcal{P} \quad \forall j \in V\} \quad (2)$$

続いて、 p 個の変数からなる非巡回的有向グラフの集合を \mathcal{G}_p と表記する。そこで、DAG \mathcal{G}_p の空間上の確率分布のクラス \mathcal{P} における識別可能性を以下のように定義する。

定義 2.1 (識別可能性). 条件付き分布のクラス \mathcal{P} が \mathcal{G}_p において識別可能であるとは、 $G, G' \in \mathcal{G}_p$ において $G \neq G'$ であるならば、 $f_G = f_{G'}$ を満たすような $f_G \in \mathcal{F}(G; \mathcal{P})$ と $f_{G'} \in \mathcal{F}(G'; \mathcal{P})$ が存在しないことである。

2.2 LiNGAM

本節では、連続変数データから因果構造を推定するモデルとして、LiNGAM[5] について概説する。

LiNGAM は、観測データが DAG によって表現されるデータ生成過程から生成されるものと仮定する。 p 個の観測変数 $X = \{X_1, \dots, X_p\}$ に対する LiNGAM は以下のように書ける。

$$X_j = e_j + b_{j0} + \sum_{k \in Pa(j)} b_{jk} X_k \quad \text{with } e_j \sim \text{non-Gaussian} \quad (3)$$

各観測変数 X_j は、その変数以外の観測変数 X_k とその誤差変数 e_j の線形和である。また、それぞれの係数 b_{jk} は、変数 X_k から変数 X_j への直接的な因果効果の大きさを表す。加えて、誤差変数 e_j はすべて平均 0、分散非ゼロの非ガウス分布に従う連続確率変数であり、お互いに独立であることを仮定する。つまり、潜在交絡変数が存在しないことを仮定する [6]。

LiNGAM は Shimizu *et al.*(2006)[5] によって、因果グラフとグラフを構成するパラメータが一意に識別可能であることが証明されている。

2.3 2 次分散関数 (QVF) DAG モデル

本節では、主に離散変数を扱う DAG モデルとして、Park and Raskutti(2017)[3] によって提案された 2 次分散関数 (QVF) DAG モデルについて概説する。QVF-DAG モデルは、各頂点の親による条件付き分布 \mathcal{P} の分散が、平均の 2 次式で与えられているというモデルであり、以下のように定義される。

定義 2.2 (QVF-DAG モデル [3]). 2 次分散関数 (Quadratic variance function, QVF) DAG モデルは、各頂点の親による条件付き確率分布が、以下で表現される 2 次分散関数性 (quadratic variance function property) を満たすような DAG モデルである。

すべての $j \in V$ について、以下を満たすような $\beta_{j0}, \beta_{j1} \in \mathbb{R}$ が存在する。

$$\text{Var}(X_j | X_{Pa(j)}) = \beta_{j0} E(X_j | X_{Pa(j)}) + \beta_{j1} E(X_j | X_{Pa(j)})^2 \quad (4)$$

QVF 確率分布のクラスには、ポアソン分布、二項分布、負の二項分布、ガンマ分布などが含まれることが知られている。DAG モデルにおいては、各頂点の分布がその頂点の親変数の影響を受けており、各頂点の条件付き期待値は、任意の単調で微分可能なリンク関数 $g_j: \mathcal{X}_{Pa(j)} \rightarrow \mathbb{R}^+$ によって、 $E(X_j | X_{Pa(j)}) = g_j(X_{Pa(j)})$ で表現される。

また、本論文の後半では、各頂点間の関係について線形性を仮定するため、QVF-DAG モデルの特殊系として、QVF 構造方程式モデル (structural equation model, SEM) を導入する。QVF-SEM は、リンク関数 g_j がパラメータに関して線形であることを仮定したものである。

$$g_j(X_{Pa(j)}) = \theta_j + \sum_{k \in Pa(j)} \theta_{jk} X_k \quad (5)$$

ここで、 $(\theta_{jk})_{k \in Pa(j)}$ は親変数の重み付け係数である。例えば、ある頂点の条件付き確率分布がポアソン分布の場合、 $g_j(X_{Pa(j)}) = \exp(\theta_j + \sum_{k \in Pa(j)} \theta_{jk} X_k)$ となる。

より一般的には、指数分布族の定義を用いて、以下のように表現することができる。

$$P(X_j | X_{Pa(j)}) = \exp \left(\theta_{jj} X_j + \sum_{(k,j) \in E} \theta_{jk} X_k X_j - B_j(X_j) - A_j \left(\theta_{jj} + \sum_{(k,j) \in E} \theta_{jk} X_k \right) \right) \quad (6)$$

ここで、 $A_j(\cdot)$ は対数分配関数 (log-partition function)、 $B_j(\cdot)$ は指数分布族によって決まる関数、 $\theta_{jk} \in \mathbb{R}$ は頂点 j に対応するパラメータである。DAG モデルの因数分解 (1) 式により、QVF-DAG モデルの同時確率分布は、以下のように記述することができる。

$$P(X) = \exp \left(\sum_{j \in V} \theta_{jj} X_j + \sum_{(k,j) \in E} \theta_{jk} X_k X_j - \sum_{j \in V} B_j(X_j) - \sum_{j \in V} A_j \left(\theta_{jj} + \sum_{(k,j) \in E} \theta_{jk} X_k \right) \right) \quad (7)$$

このモデルは、各条件付き分布がそれぞれ異なる分布に従っているような混合 DAG モデルにも拡張することが可能である。また、各頂点の分布 \mathcal{P} が (4) 式で定義される 2 次分散関数性を満たす場合、非線形モデルやノンパラメトリックモデルに拡張することも可能である。

2.4 QVF-DAG モデルの識別可能性

本節では QVF-DAG モデルが識別可能であることを証明する。QVF-DAG モデルの識別可能性は Park and Raskutti(2017)[3] によって初めて証明されたが、本論文では Park and Park(2019)[1] のアイデアを用いることにより、識別可能条件の緩和も行う。

まず初めに、QVF-DAG モデルにおけるモーメント (積率) について以下のような関係性が成立していることを示し、識別可能性の証明に利用する。

命題 2.3. リンク関数 $(g_j(X_{Pa(j)}))_{j \in V}$ が非退化である QVF-DAG モデル (4) において、任意の頂点 $j \in V$ 、任意の集合 $S_j \subset Nd(j)$ に関して、以下のモーメント関係が成立している。

$$\frac{E(X_j^2)}{E[\beta_0 E(X_j|X_{S_j}) + (\beta_1 + 1)E(X_j|X_{S_j})^2]} \geq 1 \quad (8)$$

同様に、

$$E(\text{Var}(E(X_j|X_{Pa(j)}|X_{S_j})) \geq 0 \quad (9)$$

等号成立は、 S_j が頂点 j の親変数すべてを含むとき ($Pa(j) \subset S_j$) である。

証明. 分散とモーメントの関係性と、2 次分散関数性の定義を利用すると、2 次分散関数性を満たす確率変数 X のモーメントについて、以下の関係性が成り立つ。

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 && \text{分散の公式より} \\ &= \beta_0 E(X) + \beta_1 E(X)^2 && \text{2 次分散関数性の定義より} \end{aligned}$$

よって、

$$E(X^2) = \beta_0 E(X) + (\beta_1 + 1)E(X)^2$$

ここで、記号の簡単のために、 $f(\mu) = \beta_0 \mu + (\beta_1 + 1)\mu^2$ と関数を定義する。すると、任意の頂点 $j \in V$ 、任意の空でない集合 $S_j \subset Nd(j)$ について、以下のように書ける。

$$\begin{aligned} E(X_j^2|S_j) &= E(E(X_j^2|X_{Pa(j)})|S_j) \\ &= E(f(E(X_j|X_{Pa(j)}))|S_j) \end{aligned} \quad (10)$$

イェンセンの不等式と関数 $f(\cdot)$ が凸であることを利用すると、以下が導ける。

$$\begin{aligned} E(f(E(X_j|X_{Pa(j)}))|S_j) &\geq f(E(E(X_j|X_{Pa(j)}|S_j)) \\ &= f(E(X_j|S_j)) \end{aligned} \quad (11)$$

ここで、モデルの定義より、 $E(X_j|X_{Pa(j)}) = g_j(X_{Pa(j)})$ であり、関数 $g_j(\cdot)$ は非退化であることを利用すると、等号は S_j が頂点 j の親変数すべてを含むとき ($Pa(j) \subset S_j \subset (Nd)(j)$) のみ成立する。

式 (10) と式 (11) を整理すると、

$$\begin{aligned} E(X_j^2|S_j) - f(E(X_j|S_j)) &\geq 0 \\ E(X_j^2|S_j) - (\beta_0 E(X_j|S_j) + (\beta_1 + 1)E(X_j|S_j)^2) &\geq 0 \end{aligned}$$

となり、さらに期待値を取ること、

$$E(X_j^2) - E(\beta_0 E(X_j|S_j) + (\beta_1 + 1)E(X_j|S_j)^2) \geq 0$$

が得られる。よって、以下が成り立つ。

$$\frac{E(X_j^2)}{E(\beta_0 E(X_j|S_j) + (\beta_1 + 1)E(X_j|S_j)^2)} \geq 1$$

ここからは、 $E(X_j^2) \geq E(\beta_0 E(X_j|S_j) + (\beta_1 + 1)E(X_j|S_j)^2)$ が、 $E(\text{Var}(E(X_j|X_{Pa(j)})|X_{S_j})) \geq 0$ と同値であることを証明する。分散の公式を用いると… ここ書く… \square

直感的な理解を得るために、各頂点の親変数による条件付き確率分布がポアソン分布である 2 変数 DAG モデルを例にその識別可能性を証明する。そこで、図 1 のような DAG モデルを考える。

- $G_1: X_1 \sim \text{Poisson}(\lambda_1), \quad X_2 \sim \text{Poisson}(\lambda_2)$ ただし、 X_1 と X_2 は独立
 - $G_2: X_1 \sim \text{Poisson}(\lambda_1), \quad X_2|X_1 \sim \text{Poisson}(g_2(X_1))$
 - $G_3: X_2 \sim \text{Poisson}(\lambda_2), \quad X_1|X_2 \sim \text{Poisson}(g_1(X_2))$
- ただし、 g_1 と g_2 は非退化な任意の関数である。 $(g_1, g_2: \mathbb{N} \cup \{0\} \rightarrow \mathbb{R}^+)$

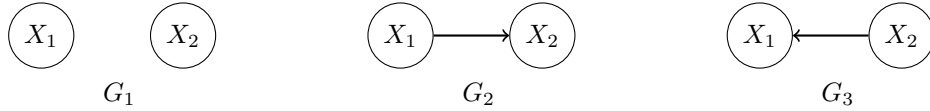


図 1 2 変数の DAG モデル

命題 2.3 より、 G_1 におけるすべての頂点 $j \in \{1, 2\}$ について、 $E(X_j^2) = E(X_j) + E(X_j)^2$ である。 G_2 においては、以下が成り立つ。

$$E(X_1^2) = E(X_1) + E(X_1)^2, \quad \text{and} \quad E(X_2^2) > E(X_2) + E(X_2)^2$$

同様に、 G_3 においては、以下が成り立つ。

$$E(X_1^2) > E(X_1) + E(X_1)^2, \quad \text{and} \quad E(X_2^2) = E(X_2) + E(X_2)^2$$

つまり、モーメント比 $E(X_j^2)/(E(X_j) + E(X_j)^2)$ によって、真のグラフ構造を同定することが可能である。

命題 2.3 のモーメント比を用いる方法は、一般的な p 変数の QVF-DAG モデルにも適用することが可能であり、モーメント比 (8) が 1 か 1 以上かを確かめることで識別可能性を証明することができる。

定理 2.4 (QVF-DAG モデルの識別可能性). 2 次分散関数性を満たす係数 $(\beta_{j0}, \beta_{j1})_{j=1}^p$ が存在し、QVF-DAG モデル (1) のクラスについて考える。任意の頂点 $j \in V$ について、 $\beta_{j1} > -1$ であり、リンク関数 $g_j(\cdot)$ が非退化であるならば、QVF-DAG モデルは識別可能である。

証明. 一般性を失わずに、真の因果順序が一意であり、 $\pi = (\pi_1, \dots, \pi_p)$ であると仮定する。また、簡単のために、 $X_{1:j} = (X_{\pi_1}, X_{\pi_2}, \dots, X_{\pi_j})$ 、 $X_{1:0} = \emptyset$ と定義する。加えて、モーメント関連関数 $f(\mu) = \beta_0 \mu, (\beta_1 + 1)\mu^2$ を定義する。

ここからは数学的帰納法を用いて QVF-DAG モデルの識別可能性を証明する。

Step(1) 因果順序が最初である π_1 について、命題 2.3 を用いると、 $E(X_{\pi_1}^2) = E(f(E(X_{\pi_1})))$ であるのに対し、任意の頂点 $j \in V \setminus \{\pi_1\}$ については、 $E(X_j^2) > E(f(E(X_j)))$ である。よって、因果順序が 1 番目の要素 π_1 を特定することができる。

Step(m-1) 因果順序が $(m-1)$ 番目の要素について、因果順序が先の $m-1$ 個の要素とその親が正しく推定されていると仮定する。

Step(m) 因果順序が m 番目の要素とその親について考える。命題 2.3 より、 π_m は、 $E(X_{\pi_m}^2) = E(f(E(X_{\pi_m}|X_{1:(m-1)})))$ である。

一方で、 $j \in \{\pi_{m+1}, \dots, \pi_p\}$ については、 $E(X_j^2) > E(f(E(X_j|X_{1:(m-1)})))$ である。よって、因果順序が m 番目の要素 π_m を推定することができる。

親変数に関しては、 $P(G)$ の因数分解 (1) による以下の条件付き独立関係より導くことができる。

$$\begin{aligned} E(X_{\pi_m}^2) &= E(f(E(X_{\pi_m}|X_{1:(m-1)}))) \\ &= E(f(E(X_{\pi_m}|X_{Pa(\pi_m)}))) \end{aligned}$$

つまり、上記の関係が成立するような最小の集合を $X_{1:(m-1)}$ の中から π_m の親として選択することができる。

□

Park and Raskutti(2017)[3] によって証明された QVF-DAG モデルの識別可能条件には、 $Pa(j) \not\subseteq S_j$ のとき、すべての $x \in \mathcal{X}_{S_j}$ について、 $Var(E(X_j|X_{Pa(j)})|X_{S_j} = x) > 0$ という仮定が含まれていた。しかし、本論文における識別可能条件にはそのような仮定は含まれていない。つまり、従来の識別可能条件 [3] を緩和している。この識別可能条件の緩和によって Poisson-SEM の学習が容易になることが、Park and Park(2019)[1] の 3.2 節において議論されている。

2.5 提案モデル

本節では LiNGAM [5] と QVF-DAG モデル [3] を組み合わせることによって、連続変数と離散変数の両方から構成される DAG モデルを提案する。

2.6 提案モデルの識別可能性

本節では、前節で提案した DAG モデルの識別可能性を証明する

参考文献

- [1] Gunwoong Park and Sion Park. High-Dimensional poisson structural equation model learning via ℓ_1 -Regularized regression. *J. Mach. Learn. Res.*, Vol. 20, No. 95, pp. 1–41, 2019.
- [2] Gunwoong Park and Garvesh Raskutti. Learning Large-Scale poisson DAG models based on OverDispersion scoring. In C Cortes, N D Lawrence, D D Lee, M Sugiyama, and R Garnett, editors, *Advances in Neural Information Processing Systems 28*, pp. 631–639. Curran Associates, Inc., 2015.
- [3] Gunwoong Park and Garvesh Raskutti. Learning quadratic variance function (QVF) DAG models via overdispersion scoring (ODS). *J. Mach. Learn. Res.*, Vol. 18, No. 1, pp. 8300–8342, January 2017.
- [4] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009.
- [5] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear Non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.*, Vol. 7, No. Oct, pp. 2003–2030, 2006.
- [6] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, Prediction, and Search*. MIT Press, 2000.

謝辞

ありがとうございました.