

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
小西 伶児

2021 年 1 月 4 日

概要

本研究では,

目次

1	序論	3
1.1	研究背景	3
1.2	既存手法・研究	4
1.3	研究目的	5
1.4	本論文の構成	6
2	既存モデル	7
2.1	数学的準備	7
2.2	Additive Noise Model の識別可能条件	8
2.3	2 次分散関数 (QVF) DAG モデル	10
3	QVF-DAG モデルの識別可能性	12
4	提案モデル	15
4.1	提案モデル	15
4.2	提案モデルの識別可能性	16
5	推定アルゴリズム	21
6	数値実験	24
6.1	設定	24
6.2	結果	25
6.3	閾値の設定について	25

7	結論	28
	参考文献	30
	謝辞	30

1 序論

1.1 研究背景

企業は自社の商品やサービスを顧客に提供するために、様々なマーケティング活動を行っている。近年では消費者の嗜好が多様化したり、新型コロナウイルス感染症が流行したりするなど、企業活動を取り囲む環境が日々大きく変化しており、企業はその環境変化に適応する必要がある。それ故、商品・サービスの開発や消費者とのコミュニケーションなどのマーケティング活動を適切に実行するためには、消費者の行動について深く理解することがより一層重要となっている。

アメリカ・マーケティング協会 (AMA) によると、マーケティングの定義は「顧客、クライアント、パートナー、社会のための価値の創造、伝達、提供、交換という全体の活動」であり、マーケティング・リサーチは「消費者、顧客、公衆とマーケターが情報を介してつながる機能」であると定義している [1]。また、その具体的な業務として、「必要な情報を特定し、情報収集のための方法を設計し、データ収集プロセスを管理・実施し、結果を分析し、分析結果と結果から得られる示唆伝えること」としている。つまり、上記の定義を合わせて考えると、マーケティング・リサーチの目的は「マーケティング課題の発見や施策の実行に必要なデータを適切に収集し、得られたデータを分析することで、企業のマーケティング活動を支援すること」と理解することができる。

企業のマーケティング活動は、主に以下の4つのフェーズに分類することが可能で、商品・サービスのカテゴリにも依存するが、約1～数年程度の期間で繰り返されることが一般的である。この繰り返しのことをマーケティング・サイクルと呼ぶ。

- 市場機会の発見
- コンセプト開発
- コミュニケーション内容・販売施策の策定
- 施策後の効果検証

マーケティング・リサーチは、各フェーズにおけるマーケティング担当者の関心事に対して、適切な示唆を与えることが求められている。「市場機会の発見」では、「市場にある商品で満たされていないニーズは何か？」や「この商品を購入している人はどのような特徴があるのか？」といったことがマーケティング担当者の関心事となり、未充足ニーズの探索や、消費者セグメントの整理などがリサーチの役割となる。次に「コンセプト開発」では、「どのような価値を提供すると売れるのか？」や「ターゲットとなる消費者の規模はどのくらいか？」といったことが関心事となり、コンセプトの受容性確認や、消費者セグメントの規模感把握などがリサーチに求められる。また「コミュニケーション内容・販売施策の策定」では、「商品の特長をどのように伝えと購買に結びつくのか？」や「商品パッケージや価格はどのようにすればよいのか？」といったことが主な関心事となり、広告内容の精査や、商品の改善点抽出などが行われる。最後に「施策後の効果検証」においては、市場浸透度の確認や、広告量などの投資に対する売上の費用対効果などが行われる。各フェーズでのマーケティング担当者の関心事を俯瞰すると、多くは「マーケティング活動と消費者行動の因果関係」にあると言える。

「マーケティング」の他に「ブランディング」も類似の意味を持つ言葉として頻繁に用いられるが、それぞれの言葉の意味には異なる部分が存在する。音部 (2019)[30] によると、マーケティングは属性の順位を変換し

て市場を創造することを目指し、結果的にニーズを作り出すことにつながっており、ブランディングはブランドの意味の確立を目指し、結果的にベネフィットを作り出すことにつながっている。つまり、マーケティングは消費者行動の因果構造そのものを変化させることによって、自社の商品が有利に購買されるような状況を作り出す活動であることに對し、ブランディングは現在の消費者行動の因果構造はそのままに、自社の商品に対する消費者の認識を変化させることによって、他社の商品ではなく自社の商品を購入するように行動を変化させる活動であると捉えることができる。

このように企業は、自社の商品やサービスを顧客に効率的に提供するために、マーケティング・リサーチを通じて消費者に関する情報を収集・分析し、消費者行動の因果関係について仮説を立てたり解釈を行ったりしている。そのため、マーケティング・リサーチにおいて、消費者行動の因果関係に関する情報を得る手段は、非常に重要な役割を担っている。

1.2 既存手法・研究

因果関係に関する考察は、科学の基本的な問いとして因果推論という文脈で研究されてきた。近年では計算機科学や統計科学の発展により、経済学やマーケティングの分野でも注目されている [18]。因果推論は大まかには、(i) 因果構造の同定、と (ii) 因果構造を既知としたその因果関係の大きさの推定、という 2 つの問題に分類できる。(i) 因果構造の同定 は、非常に困難な問題であるとされているが、いくつかの仮定の上で識別可能なモデルが提案されている。(ii) 因果構造を既知としたその因果関係の大きさの推定 は、ランダム化比較試験を中心とした実験研究と、特に実験を行わない観察研究の双方において統計科学の文脈で研究されている。

本節では、マーケティング・リサーチにおいて消費者行動の因果関係に関する情報を得る方法として従来より広く用いられてきた一般的な既存手法について述べた後、統計的因果探索と呼ばれる因果構造の同定に関する近年の研究について俯瞰する。

1.2.1 マーケティング・リサーチにおける既存手法

マーケティング・リサーチにおいて消費者行動の因果関係に関する情報を得る方法は、大きく定性調査と定量調査に分類される。

定性調査は一般的に、消費者の内部にある意見や態度を理解することで、マーケティング課題の詳細な定義、仮説の設定、定量調査における調査項目の優先順位の決定、消費者独自の考え方や表現の理解、企業のマーケティング担当者の不足している知識の吸収、定量調査における最重要な項目に関する示唆を得るなどの目的で実施される [26]。主に深層面接法 (デプス・インタビュー) や集団面接法 (グループ・インタビュー) といった、インタビュアーが回答者との対話を通じて質問をし回答を得る方法が一般的である。そのため、調査対象者が日常生活であまり気に留めていないことや、誰かに問いかけられて初めて気づくことなどを収集することができ [26]、マーケティング担当者の周辺知識では想定しきれなかった因果関係を発見できる可能性がある。一方で、定性調査は定量調査と比べて一般的に時間や費用が多くかかることや、定量的な評価ができず、得られた意見や行動の一般性・代表性に関する議論ができないことなどのデメリットがある。

定量調査では、主にアンケート調査に代表される意識データや、POS データ、ID 付き POS データ、位置情報などの行動データが用いられる。これらのデータの分析手法は非常に多岐に渡るが、消費者行動の因果関係を評価する手法としては、一般化線形モデル (generalized linear model, GLM) や構造方程式モデル (structural equation model, SEM) を用いることが多い [24][28]。GLM では、目的変数と説明変数の関係を定式化し、各係数を目的変数に対する説明変数の因果関係の大きさとして解釈を行うことが慣例となってい

る。しかし、宮川 (2004)[25] で述べられている通り、GLM は説明変数を与えたときの目的変数の条件付き確率分布に関するモデルである。つまり、GLM による分析の目的は、説明変数を観測したときの目的変数の予測であり、説明変数に外的操作を行ったときの目的変数の因果効果の定量化ではない。そのため、推定されたパラメータに対して因果的な解釈を行うことは誤りとなる可能性がある。一方で、SEM はデータの生成過程を記述した統計的因果モデルであり、その係数は単なる相関関係の尺度ではなく、因果的な解釈を行うことができる [25]。ただし、SEM は分析者の事前知識を積極的に利用することで因果構造の仮説を有向グラフで表現した上で、その構造に対してモデリングを行う手法である。つまり、SEM は上述の (ii) 因果構造を既知としたその因果関係の大きさの推定を行う手法であると言える。そのため、因果構造の仮説構築は分析者側の事前知識の質や量に依存し、妥当だと思われるモデルを得るまでに長い時間を要したりしているという現状がある。また、近年の社会の発展により消費者の行動は複雑化しているため、因果構造に関する仮説を構築することが難しい場合も少なくない。

1.2.2 統計的因果探索

統計的因果探索とは、因果構造が未知である際に、どのような条件の下で観測データから因果構造が復元することが可能であるかを明らかにし、観測データがその条件を満たしているという仮定の下で、因果構造を推定するアルゴリズムを研究する学問分野である。

以下の内容あたりをまとめる。

- 忠実性を仮定すると、条件付き独立関係からマルコフ同値類まで探索することができる
- 関数形や誤差分布などに仮定を置くと、因果構造が一意に復元可能なモデルが提案されている (LiNGAM などの ANM)
- ほとんどの既存手法は、データがすべて離散かすべて連続を仮定している
- カウントデータを仮定したモデルとして Park and Raskutti(2015)[10]
- それを拡張する方法などとして、Park and Raskutti(2017)[11] や、Park and Park(2019)[9]
- 連続・離散混在のモデルとして、Wenjuan *et al.*(2018)[19] や、Li and Shimizu(2018)[7] や、Yamayoshi *et al.*(2020)[20] など

1.3 研究目的

本研究の目的は、統計的因果探索の手法を応用し、マーケティング・リサーチにおいて消費者行動の因果構造に関する仮説を構築する手法の開発を行うことである。マーケティング・リサーチで扱う定量データは、先述のようにアンケート調査データや、POS データ、ID 付き POS データなどが中心である。特にアンケート調査データには、聴取内容に応じて様々な尺度のデータが含まれており、離散変数と連続変数の両方が混在している。

以下のような例を表にまとめる

- 知ってる/知らない などの 01 データ
- 最もよく買うブランドはどれか? などの名義尺度 (ダミー変数に変換して分析)
- あてはまる～あてはまらない のようなリッカート尺度 (連続変数とみなすことが多い)
- 購買頻度 (毎日買う/週 1 以上/月 1 以上) のような順序尺度 (もともとはカウントデータ)

そこで、本論文では離散変数と連続変数の両方が含まれる構造的因果モデルを提案し、その因果構造の識別可能性を議論する。

1.4 本論文の構成

最後にまとめて書く

2 既存モデル

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 モデルとして広く用いられている Additive Noise Model [15] [5] [13] [14] [8] 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)[11] 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。

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、因果順序 (causal ordering) について定義する。因果順序とは、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 によるモデルは因果マルコフ条件によ

り以下のように因数分解することができる [12]。

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

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、本論文で扱う因果モデルには、因果極小性 (causal minimality) を仮定する。因果極小性とは、DAG G で表現される因果構造に従って生成された分布 $f_G(X)$ は、 G の部分グラフにおいては因果マルコフ条件を満たさないことを言う [21]。つまり、全ての頂点 $j \in V$ とその親の 1 つ $k \in Pa(j)$ について、以下が成り立つ。

$$\forall Pa(j) \setminus \{k\} \subset S \subset Nd(k) \setminus \{k\}, \quad X_j \not\perp\!\!\!\perp X_k | X_S \quad (2)$$

最後に、本論文では観測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。

$$\mathcal{F}(G; \mathcal{P}) \equiv \{f_G(X) = \prod_{j \in V} f_j(X_j | X_{Pa(j)}); \text{ここで、} f_j(X_j | X_{Pa(j)}) \in \mathcal{P} \quad \forall j \in V\} \quad (3)$$

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 Additive Noise Model の識別可能条件

本節では、連続変数データを扱う DAG モデルとして、広く用いられている Additive Noise Model (ANM) [15] [5] [13] [14] [8] とその識別可能性について概説する。

ANM は、観測変数の同時分布が、以下の構造方程式と誤差から生成される DAG モデルである。

$$X_j = f_j(X_{Pa(j)}) + e_j, \quad e_j \sim (0, \sigma_j^2) \quad (4)$$

$(f_j)_{j \in V}$ は任意の関数であり、 $(e_j)_{j \in V}$ は平均ゼロでそれぞれ異なる分散 $(\sigma_j^2)_{j \in V}$ に従う互いに独立な確率変数である。ここで、 $(e_j)_{j \in V}$ の分布の形は任意である。つまりガウス分布など特定の分布であるとは限らない。

また、ANM の特殊形として、 $(f_j)_{j \in V}$ が全て線形な関数で記述されるモデルを線形構造方程式モデル (SEM) という。つまり、観測変数の同時分布が以下の線形方程式で定義される DAG モデルである。

$$X_j = \theta_j + \sum_{k \in Pa(j)} \theta_{jk} X_k + e_j, \quad e_j \sim (0, \sigma_j^2) \quad (5)$$

ここで、それぞれの係数 θ_{jk} は、DAG G における頂点 k から頂点 j の直接的な因果効果の大きさを表す。つまり、頂点 k が頂点 j の親であるときは $\theta_{jk} \neq 0$ であり、それ以外の場合は $\theta_{jk} = 0$ である。

これらの DAG モデルは、関数形や誤差変数の分布についていくつかの制約を課すと、識別可能であることが証明されている。代表的な識別可能条件について以下に簡単にまとめる。

- 全ての関数 $(f_j)_{j \in V}$ が非線形である非線形 ANM[5]
- 全ての関数 $(f_j)_{j \in V}$ が線形であり、観測変数 $(X_j)_{j \in V}$ または誤差変数 $(e_j)_{j \in V}$ のいずれかの確率分布が、非ガウス分布に従う線形非ガウス非巡回モデル (LiNGAM)[15]
- 全ての関数 $(f_j)_{j \in V}$ が線形であり、誤差変数 e_j の分散が全て等しい、または全て既知である線形 ANM[13]

このように関数形や誤差分布に関する仮定を置くことで、様々な識別可能な ANM が提案されてきた。一方で、全ての関数形が非線形であることや、誤差分布が非ガウス分布に従うこと、誤差分布の分散が全て等しいことなどの仮定はあまり現実的でないといった批判もある。そこで、誤差変数の分散の大きさだけでなく、誤差変数に対する親変数の影響の大きさも加味することで、関数形や誤差変数の分布の制約を受けない以下のような識別可能条件が示されている [8]。

定理 2.2 (ANM の識別可能条件 [8]). 同時確率 $f(X)$ が DAG G の ANM(4) から生成されているとする。このとき、任意の頂点 $j = \pi_m \in V, k \in De(j), l \in An(j)$ に関して、以下の 2 つの条件のいずれかが満たされているならば、DAG G は一意に識別可能である。ここで DAG G における因果順序を π で表す。

$$\begin{aligned} \text{(A)} \quad & \sigma_j^2 < \sigma_k^2 + E(\text{Var}(E(X_k | X_{Pa(k)}) | X_{\pi_1}, \dots, X_{\pi_{m-1}})) \\ \text{(B)} \quad & \sigma_j^2 > \sigma_l^2 - E(\text{Var}(E(X_l | X_{\pi_1}, \dots, X_{\pi_m} \setminus X_l) | X_{Pa(l)})) \end{aligned}$$

条件 (A) は、頂点 j の条件付き分散が、非子孫 $Nd(j)$ で条件づけたときの $De(j)$ の条件付き分散より小さいときに、ANM は識別可能であることを表現している。また、条件 (B) は、頂点 j の条件付き分散が、祖先 $An(j)$ の親と子孫の和集合で条件づけたときの $An(j)$ の条件付き分散より大きいときに、ANM は識別可能であることを表現している。

ここでは条件 (A) について直感的な理解を得るために、図 1 の DAG で表される 2 変数の正規線形構造方程式モデルを用いて、その識別可能性を証明する。

- $G_1: X_1 = e_1, \quad X_2 = \theta_1 X_1 + e_2$
- $G_2: X_1 = \theta_2 X_2 + e_1, \quad X_2 = e_2$
- ただし、全ての $j \in \{1, 2\}$ について、 $e_j \sim N(0, \sigma_j^2)$ である。



図 1 2 変数の正規線形構造方程式モデル

G_1 において、もし誤差変数の分散について条件 (A) $\sigma_1^2 < \sigma_2^2 + \beta_1^2 \sigma_1^2$ が成立しているならば、全分散の公式より X_1 と X_2 の分散について以下の関係が導かれる。

$$\begin{aligned} \text{Var}(X_2) &= E(\text{Var}(X_2|X_1)) + \text{Var}(E(X_2|X_1)) \\ &= \sigma_2^2 + \beta_1^2 \sigma_1^2 \\ &> \sigma_1^2 \\ &= \text{Var}(X_1) \end{aligned}$$

この関係について直感的に述べると、 X_1 における確率変動は e_1 のみによって起こるが、 X_2 における確率変動は X_1 と e_2 によって起こるため、 X_2 の不確実性 (分散) のほうが X_1 の不確実性より大きくなると理解することができる。よって、条件 (A) $\sigma_1^2 < \sigma_2^2 + \beta_1^2 \sigma_1^2$ が成立しているならば、 G_1 における真の因果順序が $\pi = (1, 2)$ という順であることを観測変数より特定することができる。

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)[11] 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 2.3 (QVF-DAG モデル [11]). 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。

$$\text{Var}(X_j|X_{Pa(j)}) = \beta_{j0}E(X_j|X_{Pa(j)}) + \beta_{j1}E(X_j|X_{Pa(j)})^2 \quad (6)$$

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 QVF 構造方程式モデル (structural equation model, SEM) を導入する。

$$g_j(X_{Pa(j)}) = g_j \left(\theta_j + \sum_{k \in Pa(j)} \theta_{jk} X_k \right) \quad (7)$$

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。

$$P(X_j|X_{Pa(j)}) = \exp \left(\theta_j X_j + \sum_{(k,j) \in E} \theta_{jk} X_k X_j - B_j(X_j) - A_j \left(\theta_j + \sum_{(k,j) \in E} \theta_{jk} X_k \right) \right) \quad (8)$$

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。

$$P(X) = \exp \left(\sum_{j \in V} \theta_j X_j + \sum_{(k,j) \in E} \theta_{jk} X_k X_j - \sum_{j \in V} B_j(X_j) - \sum_{j \in V} A_j \left(\theta_j + \sum_{(k,j) \in E} \theta_{jk} X_k \right) \right) \quad (9)$$

このモデルは、関数 $A_j(\cdot)$ や $B_j(\cdot)$ が頂点 j によって異なることを許容するため、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 モデルを表現することも可能である。また、各頂点の分布 \mathcal{P} が (6) 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。

3 QVF-DAG モデルの識別可能性

本章では、前章で導入した QVF-DAG モデル [11] 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)[11] によって初めて証明されたが、本論文では Park and Park(2019)[9] 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。

Park and Park 2019 の方針で証明する意義付けを明確にする必要あり…

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 3.1. 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 (6) 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。

$$\frac{E(X_j^2)}{E[\beta_0 E(X_j|X_{S_j}) + (\beta_1 + 1)E(X_j|X_{S_j})^2]} \geq 1 \quad (10)$$

同様に、

$$E(\text{Var}(E(X_j|X_{Pa(j)}|X_{S_j})) \geq 0 \quad (11)$$

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、関数 $f(\mu) = \beta_0 \mu + (\beta_1 + 1)\mu^2$ 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。

$$\begin{aligned} E(X_j^2|S_j) &= E(E(X_j^2|X_{Pa(j)}|S_j)) \\ &= E(f(E(X_j|X_{Pa(j)}|S_j))) \end{aligned} \quad (12)$$

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。

$$\begin{aligned} E(f(E(X_j|X_{Pa(j)}|S_j))) &\geq f(E(E(X_j|X_{Pa(j)}|S_j))) \\ &= f(E(X_j|S_j)) \end{aligned} \quad (13)$$

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 (12) と式 (13) 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で、

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

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… 後ほど書く… \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 2 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。 $(g_1, g_2: \mathbb{N} \cup \{0\} \rightarrow \mathbb{R}^+)$

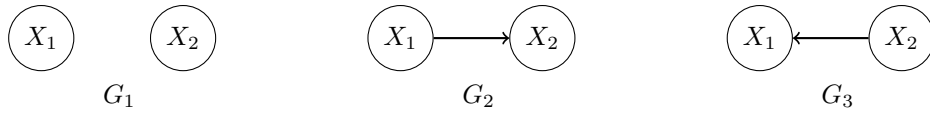


図 2 2 変数の Poisson-DAG モデル

命題 3.1 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 3.1 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 (10) 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 3.2 (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 3.1 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。つまり、因果順序が $(m - 1)$ 番目の要素については、 $E(X_{\pi_{m-1}}^2) = E(f(E(X_{\pi_{m-1}}|X_{1:(m-2)})))$ が成立していると仮定する。一方で、任意の頂点 $k \in \{\pi_m, \dots, \pi_p\}$ については以下が成立していると仮定する。

$$E(X_j^2) > E(f(E(X_j|X_{1:(m-2)})))$$

Step(m)

因果順序が m 番目の要素とその親について考える。帰納法の仮定より、 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 を特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)[11] 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 [11] 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)[9] の 3.2 節において議論されている。

4 提案モデル

本章では前章で俯瞰した Additive Noise Model[8] と QVF-DAG モデル [11] を用いることによって、連続変数と離散変数が混在するデータにおける DAG モデルを提案し、その識別可能性を証明する。

4.1 提案モデル

提案モデルにおける変数は、離散変数と連続変数に分けられ、離散変数は 0 以上の整数を取る確率変数であると仮定する。そこで提案モデルを以下のように定義する。

定義 4.1 (提案モデル). p 個の観測変数 $X = \{X_1, \dots, X_p\}$ は DAG G によって表現されるデータ生成過程から生成されており、各変数の親変数がその変数の直接的な原因である。また、 p 個の観測変数 $X = \{X_1, \dots, X_p\}$ は、連続変数 ($X_j \in \mathbb{R}$) か離散変数 ($X_j \in \{0, \mathbb{N}\}$) のいずれかに割り当てられ、それぞれ以下のデータ生成過程により生成されている。

1. 連続変数に割り当てられた変数 $X_j (j \in C)$ は、その親変数 $Pa(j)$ と誤差変数 e_j の線形和である。

$$X_j = \theta_j + \sum_{k \in Pa(j)} \theta_{jk} X_k + e_j, \quad e_j \sim (0, \sigma_j^2) \quad (14)$$

それぞれの係数 θ_{jk} は、DAG G における変数 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。つまり、頂点 k が頂点 j の親であるときは $\theta_{jk} \neq 0$ であり、それ以外のときは $\theta_{jk} = 0$ である。また、誤差変数 e_j は平均ゼロでそれぞれ異なる分散 σ_j^2 に従う互いに独立な確率変数である。

2. 離散変数に割り当てられた変数 $X_j (j \in D)$ は、その親変数 $Pa(j)$ による条件付き確率が、2 次分散関数性を満たす。つまり、以下を満たすような $\beta_{j0}, \beta_{j1} \in \mathbb{R}$ が存在する。

$$Var(X_j | X_{Pa(j)}) = \beta_{j0} E(X_j | X_{Pa(j)}) + \beta_{j1} E(X_j | X_{Pa(j)})^2 \quad (15)$$

また、各変数の条件付き期待値は、その変数の親変数 $Pa(j)$ と任意の単調で微分可能なリンク関数 $g_j: \mathcal{X}_{Pa(j)} \rightarrow \mathbb{R}^+$ によって以下のように記述される。ここで、それぞれの係数 θ_{jk} は、DAG G における変数 X_k から変数 X_j への直接的な関係性の強さを表す。つまり、頂点 k が頂点 j の親であるときは $\theta_{jk} \neq 0$ であり、それ以外のときは $\theta_{jk} = 0$ である。

$$E(X_j | X_{Pa(j)}) = g_j(X_{Pa(j)}) = g_j \left(\theta_j + \sum_{k \in Pa(j)} \theta_{jk} X_k \right) \quad (16)$$

本モデルは解釈性が重視されるマーケティング・リサーチで用いることを想定しているため、連続変数のデータ生成過程や、離散変数のデータ生成過程におけるリンク関数 g_i は、パラメータ θ に関して線形な関数であることを仮定しているが、次節の識別可能性の議論においてはそのような仮定は必要ではない。そのため、この提案モデルは非線形モデルへと簡単に拡張することが可能である。

マーケティング・リサーチの文脈において、上記のような因果モデルが有用であると考えられるかについて述べる。まず第一に、消費者行動の研究のうち、「消費者の知覚、選好および選択」を扱った分野との親和性である。消費者の知覚・選好・選択モデルは、消費者がどのようにブランドを理解し、そのブランドに対してど

のような態度形成を経て購買に至るかをモデル化したものであり、具体的には「選好回帰分析」や「コンジョイント分析」といった手法によって分析が行われる [29]。これらの分析の基礎的な考え方に「多属性態度モデル」が採用されており、それは以下のように表される。

$$U_j = a_1 z_{1j} + a_2 z_{2j} + \cdots + a_l z_{lj} \quad (17)$$

ここで、 U_j は対象 j に対する態度、 z_{ij} を対象 j の第 i 属性の水準、 a_i を第 i 属性に対する消費者の重視度である。例えば、ペットボトル入りの緑茶飲料に対する購入意向 (態度) を、苦み・渋み・甘みなどの消費者が感じる味覚の良し悪し (属性の水準) と、各属性に対する消費者の重視度の線形加重和で表そうとするモデルである。本論文の提案モデルにおける連続変数のデータ生成過程の定義である式 (14) は、選好回帰分析やコンジョイント分析が前提としている考え方を表現していると言える。通常、 U_j や z_{ij} のデータは、2～7 段階尺度 (「あてはまる」～「あてはまらない」など) で評価したデータが用いられる [29][26]。これらのデータは本来連続量である態度や評価を便宜的にいくつかの段階に分割した順序付きカテゴリカルデータであり、通常、各カテゴリに等間隔の数値を割り当て、割り当てられた値自体をデータとして回帰分析や因子分析を行うことが多い。この方法には批判もあるが、相関係数の推定精度の観点などから、5 段階以上の尺度であれば連続変数と同様に扱っても大きな問題とはならないという結果も見られる [23]。そのため、本論文では 5 段階以上の尺度で得られたデータは連続変数として扱う。

第二に、カウント (計数) データを扱うことである。マーケティング・リサーチにおいては、商品の購買個数や Web ページへの訪問回数、インターネット広告への接触回数など、ある事象が発生した回数を記録したカウントデータを扱うことが多い。これらのデータに対しては、ポアソン回帰モデルを用いた分析などが用いられる。例えば、ある商品の販売点数を目的変数とし、その商品価格や小売店における山積み陳列実施の有無などを説明変数としたモデルによって市場反応分析を行うことなどが挙げられる [24]。また、非耐久消費財の購入に関するデータは負の二項分布を用いてモデル化することができることが示されており [3]、負の二項分布モデルを拡張したディリクレモデルを用いてブランドの購入率や購入回数を分析したりする事例も見られる [27]。本論文の提案モデルにおける離散変数のデータ生成過程の性質を満たす分布には、ポアソン分布や二項分布、負の二項回帰などが含まれており、マーケティング・リサーチの分野において活用しやすいと考えられる。

4.2 提案モデルの識別可能性

本節では、前節で定義した DAG モデルの識別可能性を証明する。提案モデルは、連続変数と離散変数とが混在することを許容する DAG モデルであるため、その特殊形として、全てが連続変数であるモデルや全てが離散変数であるモデルを考えることも可能である。全てが連続変数である場合は、Additive Noise Model となり、モデルの識別可能条件が複数証明されている [15] [5] [13] [14] [8]。また、全てが離散変数である場合は、QVF-DAG モデル [11] となり、識別可能性が既に証明されている [11]。そこで以下では、観測変数集合に連続変数と離散変数の両方が含まれる場合に関する識別可能条件について議論する。まず、証明の方針について直感的な理解を得るために、図 3 のような 3 変数モデルを用いてその識別可能性を示す。ここで X, Z は連続変数、 Y は離散変数であるとする。図 3 の 3 つの因果グラフから生成される分布は、いずれも $X \perp Z | Y$ という条件付き独立関係のみが成立しており、因果マルコフ条件のみでは識別できない例である。しかし、以下で示すように、提案モデルの特徴を利用すると識別することが可能である。

$$\begin{aligned}
G_1: X &= \theta_X + e_X, \quad e_X \sim N(0, \sigma_X^2) \\
Y|X &\sim \text{Poisson}(\lambda), \quad \log(\lambda) = \theta_Y + \theta_{YX}X \\
Z &= \theta_Z + \theta_{ZY}Y + e_Z, \quad e_Z \sim N(0, \sigma_Z^2)
\end{aligned}$$

$$\begin{aligned}
G_2: X &= \theta_X + \theta_{XY}Y + e_X, \quad e_X \sim N(0, \sigma_X^2) \\
Y|Z &\sim \text{Poisson}(\lambda), \quad \log(\lambda) = \theta_Y + \theta_{YZ}Z \\
Z &= \theta_Z + e_Z, \quad e_Z \sim N(0, \sigma_Z^2)
\end{aligned}$$

$$\begin{aligned}
G_3: X &= \theta_X + \theta_{XY}Y + e_X, \quad e_X \sim N(0, \sigma_X^2) \\
Y &\sim \text{Poisson}(\lambda) \\
Z &= \theta_Z + e_Z, \quad e_Z \sim N(0, \sigma_Z^2)
\end{aligned}$$

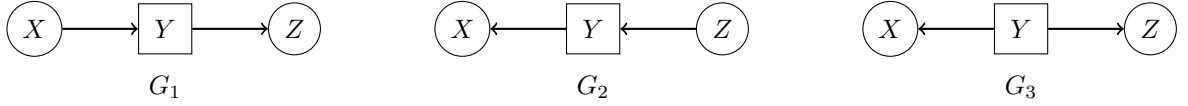


図3 3変数の DAG モデル

命題 3.1 より、 G_1, G_2 においては

$$E(Y^2) > E(Y) + E(Y)^2$$

である一方で、 G_3 においては

$$E(Y^2) = E(Y) + E(Y)^2$$

となる。よって、離散変数のモーメント比 (10) が 1 か 1 以上かを確かめることで G_1, G_2 と G_3 は識別可能である。

次に G_1 について、もし連続変数 X, Z の誤差変数の分散が $\sigma_X^2 < \sigma_Z^2 + \text{Var}(E(Z|Y))$ を満たすならば、全分散の公式を用いて以下が成り立つ。

$$\begin{aligned}
\text{Var}(Z) &= E(\text{Var}(Z|Y)) + \text{Var}(E(Z|Y)) \\
&= \sigma_Z^2 + \text{Var}(E(Z|Y)) \\
&> \sigma_X^2 \\
&= \text{Var}(X)
\end{aligned}$$

よって、 X のほうが因果順序が早いことが分かる。つまり、誤差変数の分散が $\sigma_X^2 < \sigma_Z^2 + \text{Var}(E(Z|Y))$ を満たすならば、 G_1 の因果順序を特定することが可能である。

G_2 についても同様に、連続変数 X, Z の誤差変数の分散が、 $\sigma_Z^2 < \sigma_X^2 + \text{Var}(E(X|Y))$ を満たすならば、真の因果順序 $\pi = (Z, Y, X)$ を特定することが可能である。

ここからは、上記の3変数モデルでの証明の方針を拡張し、提案モデルが一般的な p 変数の場合においても識別可能であることを証明する。

まず初めに、提案モデルにおける離散変数に関して、命題 3.1 と同様の関係が成立していることを示す。

補題 4.2. 提案モデルにおいて、離散変数が割り当てられた任意の頂点 $j \in D$ 、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。

$$\frac{E(X_j^2)}{E[\beta_0 E(X_j|X_{S_j}) + (\beta_1 + 1)E(X_j|X_{S_j})^2]} \geq 1 \quad (18)$$

等号成立は、 S_j が頂点 $j \in D$ の親変数全てを含むとき ($Pa(j) \subset S_j$) である。

証明. 提案モデルにおいて、離散変数が割り当てられた頂点は、式 (15) を満たすため、任意の変数 $X_j \in X_D$ の2次モーメントは以下のように表現できる。

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

ここで、記号の簡単な簡単のために、関数 $f(\mu) = \beta\mu + (\beta_1 + 1)\mu^2$ を定義する。すると、任意の頂点 $j \in D$ 、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。

$$\begin{aligned} E(X_j^2|S_j) &= E(E(X_j^2|X_{Pa(j)})|S_j) \\ &= E(f(E(X_j|X_{Pa(j)}))|S_j) \end{aligned} \quad (19)$$

提案モデルにおいては連続変数と離散変数が混在するモデルを考えているため、離散変数が割り当てられた頂点 $j \in D$ の非子孫の集合 $Nd(j)$ には、連続変数と離散変数の両方が含まれている可能性がある。つまり、式 (19) は S_j に連続変数と離散変数のどちらが含まれていても成立する。

以降の証明は、命題 3.1 の証明と同様である。

□

補題 4.2 で証明したモーメント関係は p 変数の提案モデルでも利用することができる。つまり、式 (18) が 1 に等しくなる $X_j \in X_D$ が存在するかどうかを確認することによって、因果順序の1番目の変数が離散変数か否かを判断することができ、離散変数の場合はその変数を特定することができる。

定理 4.3 (提案モデルの識別可能性). 定義 4.1 によって定義される DAG モデルは、以下の仮定を満たすとき識別可能である。ここで、 π は DAG G における因果順序を表す。

- (A) 連続変数が割り当てられた任意の頂点 $j = \pi_m \in C, k \in De(j) \subset C$ のデータ生成過程における誤差変数の分散について、以下が満たされている。

$$\sigma_j^2 < \sigma_k^2 + E(\text{Var}(E(X_k|X_{Pa(k)})|X_{\pi_1}, \dots, X_{\pi_{m-1}}))$$

- (B) 離散変数が割り当てられた任意の頂点 $j \in D$ について、 $\beta_{j1} > -1$ が満たされている。

仮定 (B) は、ベルヌーイ分布や多項分布による DAG モデルを除外するための仮定である。なぜなら、ベルヌーイ分布や多項分布による DAG モデルは識別不能であることが知られているためである [4]。

以下では、定理 4.3 を証明する。

証明. 一般性を失わずに、DAG G における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。DAG G において、連続変数に割り

当てられた変数からなる頂点の集合を C 、離散変数に割り当てられた変数からなる頂点の集合を D とする。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。ここから数学的帰納法を用いて提案モデルの識別可能性を証明する。

Step(1)

(i) $\pi_1 = j \in D$ の場合

補題 4.2 より、 $E(X_{\pi_1}^2) = E(f(E(X_{\pi_1})))$ が成立する。一方で、頂点 $j \in D \setminus \{\pi_1\}$ では、 $E(X_j^2) > E(f(E(X_j)))$ となる。そのため、因果順序が 1 番目の要素 π_1 は、 $E(X_j^2) = E(f(E(X_j)))$ となるような $j \in D$ である。もし、そのような変数が存在しなければ、 X_{π_1} は連続変数である。

(ii) $\pi_1 = j \in C$ の場合

定理 4.3 の仮定 (A) より、任意の頂点 $k \in C \setminus \{\pi_1\}$ について、以下が成立する。

$$\begin{aligned} \text{Var}(X_{\pi_1}) &= \sigma_{\pi_1}^2 \\ &< \sigma_k^2 + \text{Var}(E(X_k|X_{Pa(k)})) \\ &= E(\text{Var}(X_k|X_{Pa(k)})) + \text{Var}(E(X_k|X_{Pa(k)})) \\ &= \text{Var}(X_k) \end{aligned}$$

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、因果順序が早い $(m-1)$ 個の要素とその親が正しく推定されていると仮定する。つまり、以下の 2 つの関係が成立していると仮定する。

- 因果順序が $(m-1)$ 番目の要素については、 $E(X_{\pi_{m-1}}^2) = E(f(E(X_{\pi_{m-1}}|X_{1:(m-2)})))$ が成立している。一方で、任意の頂点 $k \in \{\{\pi_m, \dots, \pi_p\} \cap D\}$ については以下が成立している。

$$E(X_j^2) > E(f(E(X_j|X_{1:(m-2)})))$$

- 任意の頂点 $k \in \{\{\pi_m, \dots, \pi_p\} \cap C\}$ のデータ生成過程における誤差変数の分散について、以下が成立している。

$$\sigma_{\pi_{m-1}}^2 < \sigma_k^2 + E(\text{Var}(E(X_k|X_{Pa(k)}|X_{1:(m-2)})))$$

Step(m)

因果順序が m 番目の要素とその親について考える。

(i) $\pi_m = j \in D$ の場合

帰納法の仮定より、 $E(X_{\pi_m}^2) = E(f(E(X_{\pi_m}|X_{1:(m-1)})))$ が成立する。一方で、頂点 $j \in \{\{\pi_{m+1}, \dots, \pi_p\} \cap D\}$ では、 $E(X_j^2) > E(f(E(X_j|X_{1:(m-1)})))$ となる。そのため、因果順序が m 番目の要素 π_m は、 $E(X_j^2) = E(f(E(X_j|X_{1:(m-1)})))$ となるような $j \in D$ である。もし、そのような変数が存在しなければ、 X_{π_m} は連続変数である。

(ii) $\pi_m = j \in C$ の場合

帰納法の仮定より、任意の頂点 $k \in \{\{\pi_{m+1}, \dots, \pi_p\} \cap C\}$ について、以下が成立する。

$$\begin{aligned} E(\text{Var}(X_{\pi_m}|X_{1:(m-1)})) &= \sigma_{\pi_m}^2 \\ &< \sigma_k^2 + E(\text{Var}(E(X_k|X_{Pa(k)}|X_{1:(m-1)}))) \\ &= E(E(\text{Var}(X_k|X_{Pa(k)}|X_{1:(m-1)}))) + E(\text{Var}(E(X_k|X_{Pa(k)}|X_{1:(m-1)}))) \\ &= E(\text{Var}(X_k|X_{1:(m-1)})) \end{aligned}$$

よって、因果順序が m 番目の要素 π_m を特定することができる。

各頂点の親に関しては、因果マルコフ条件に基づく $P(G)$ の因数分解(1) によって表現される条件付き独立関係と、因果極小性 (2) により導くことができる。つまり、以下を満たす頂点 k を π_m の親として特定することができる。

$$Pa(k) := \{k \in \{\pi_1, \dots, \pi_{m-1}\} \mid X_k \not\perp\!\!\!\perp X_{\pi_m} \mid X_{1:(m-1)} \setminus X_k\}$$

よって、数学的帰納法により定理 4.3 の証明を完了する。

□

5 推定アルゴリズム

本章では、定義 4.1 による提案モデルが定理 4.3 の識別可能条件を満たす時に、因果順序と DAG における変数間の関係性の強さを推定するアルゴリズム (Algorithm 1) を提案する。推定アルゴリズムは、(i) 各変数の因果順序を推定するステップ、と (ii) 各変数の親変数との関係性の強さを推定するステップ、という 2 つのステップによって構成される。提案モデルには異なるデータ生成過程に従う連続変数と離散変数が混在しているため、(i) 因果順序を推定するステップは、連続変数と離散変数を分けて逐次的に求める。以下では、連続変数と離散変数の因果順序の推定法についてそれぞれ述べる。

まず、離散変数の因果順序の推定法について述べる。離散変数の因果順序は、以下で定義される各頂点のモーメント比スコアを比較することで推定する。

$$\begin{aligned}\hat{S}(1, j) &\equiv \frac{\hat{E}(X_j^2)}{\beta_0 \hat{E}(X_j) + (\beta_1 + 1) \hat{E}(X_j)^2} \\ \hat{S}(m, j) &\equiv \frac{\hat{E}(X_j^2)}{\hat{E}(\beta_0 \hat{E}(X_j | X_{\hat{\pi}_{1:(m-1)}}) + (\beta_1 + 1) \hat{E}(X_j | X_{\hat{\pi}_{1:(m-1)}})^2)}\end{aligned}\tag{20}$$

ここで、各推定量は以下のように求める。

$$\begin{aligned}\hat{E}(X_j) &= \frac{1}{n} \sum_{i=1}^n X_j^{(i)} \\ \hat{E}(\hat{E}(X_j | X_S)) &= \frac{1}{n} \sum_{i=1}^n g_j(\hat{\theta}_j^S + \sum_{k \in S} \hat{\theta}_{jk}^S X_k^{(i)}) \\ \hat{E}(\hat{E}(X_j | X_S)^2) &= \frac{1}{n} \sum_{i=1}^n \{g_j(\hat{\theta}_j^S + \sum_{k \in S} \hat{\theta}_{jk}^S X_k^{(i)})\}^2\end{aligned}$$

ここで $(\hat{\theta}_j^S, \hat{\theta}_{jk}^S)$ は、 X_j を目的変数、 X_S を説明変数にとった一般化線形モデルのパラメータを最尤推定した値である。つまり、離散変数の因果順序を推定する際は、事前に各離散変数の条件付き確率分布を仮定する必要がある。

式 (20) によるモーメント比スコアは、式 (18) の推定量であるため、正しい因果順序の要素のスコアは 1 に等しくなり、それ以外の場合は 1 より大きくなる。つまり、モーメント比スコアが最小となる頂点のスコアが 1 に等しい場合は、因果順序が m 番目の要素は離散変数であり、その変数を特定することができる。一方で、全てのモーメント比スコアが 1 より大きい場合は、因果順序が m 番目の要素は連続変数であることが分かる。ただし、推定誤差があるためモーメント比スコアが厳密に 1 となることはない。そのため、本論文では閾値を設定することでモーメント比スコアが 1 に等しいかどうかを判断する。

次に、連続変数の因果順序の推定法について述べる。因果順序 m 番目の頂点が連続変数である場合、定理 4.3 の条件 (A) により、因果順序 m 番目の要素の条件付き分散は、連続変数が割り当てられた他のどの頂点の条件付き分散よりも厳密に小さい。つまり、因果順序 m 番目の要素が連続変数である場合は、 $S = \{\hat{\pi}_1, \dots, \hat{\pi}_{m-1}\}$ を与えたときの各頂点の条件付き分散 $\hat{\sigma}_{j|S}$ が最小となる頂点 $j \in C$ が因果順序 m 番目の要素であると推定できる。条件付き分散の一致推定量は、一般化線形モデルや一般化加法モデルなどの回帰モデルによって得ることができる。提案モデルにおける連続変数のデータ生成過程が親変数と誤差変数との線形和であるため、本論文では、 X_j を目的変数に、 X_S を説明変数にとり、最小二乗法を用いて線形回帰分析を行ったときの残差の分散で条件付き分散 $\text{Var}(X_j | X_S)$ を推定する。

続いて、(ii) 各変数の親変数との関係性の強さを推定するステップについて述べる。上述の方法により因果

順序 $\hat{\pi}$ を推定すると、因果順序に従って回帰分析を行うことにより、変数間の関係性の強さを表すパラメータ θ_{jk} を推定する。具体的には、 X_j を目的変数に、 $X_S = \{X_{\hat{\pi}_1}, \dots, X_{\hat{\pi}_{j-1}}\}$ を説明変数にとった一般化線形モデルのパラメータを最尤法によって推定する。ただし、パラメータ θ_{jk} の値が実際には 0 であっても、最尤法による推定値は厳密には 0 にならず、本来は存在しない冗長な有向辺が残ってしまう。そこで、Shimizu *et al.*(2011)[16] のように adaptive Lasso[22] を用いることで冗長な有向辺を削除する。adaptive Lasso ではサンプルサイズが十分大きければ、親変数候補 $X_S = \{X_{\hat{\pi}_1}, \dots, X_{\hat{\pi}_{j-1}}\}$ の中から正しい親変数を選択することができる。つまり、 $\theta_{jk} \neq 0$ となるような変数の集合 $X_K \subset X_S$ を見つけることができる。

提案モデルには連続変数と離散変数が混在している。そのため、adaptive Lasso による推定値は、それぞれ以下の目的関数を最小化することによって求める。

- 連続変数

$$\left\| X_j - \sum_{k \in S} \theta_{jk} X_k \right\|^2 + \lambda \sum_{k \in S} \frac{|\theta_{jk}|}{|\hat{\theta}_{jk}|^\gamma} \quad (21)$$

- 離散変数

$$\sum_{i=1}^n \left(-X_j^{(i)} \left(\theta_j + \sum_{k \in S} \theta_{jk} X_k^{(i)} \right) + g_j \left(\theta_j + \sum_{k \in S} \theta_{jk} X_k^{(i)} \right) \right) + \lambda \sum_{k \in S} \frac{|\theta_{jk}|}{|\hat{\theta}_{jk}|^\gamma} \quad (22)$$

ここで λ と γ は調整パラメータであり、 $\hat{\theta}_{jk}$ は θ_{jk} の一致推定量による推定値である。Zou(2006)[22] では、調整パラメータは 5 分割交差検証 (5-fold cross-validation) によって選択し、 $\hat{\theta}_{jk}$ は最小二乗法または最尤法によって推定することが提案されている。

Algorithm 1 提案モデルの推定アルゴリズム

Input: 提案モデルから生成された n 個の i.i.d 標本 $X^{1:n}$ 、各離散変数の条件付き確率分布の仮定

Output: 因果順序 $\hat{\pi} = (\hat{\pi}_1, \dots, \hat{\pi}_p)$ 、グラフ構造 $\hat{E} \in \{0, 1\}^{p \times p}$ 、頂点間の関係性の強さ $\hat{\theta} \in \mathbb{R}^{p \times p}$

Step1: 因果順序の推定

初期化 $\hat{\pi}_0 = \emptyset$

for $m = \{1, 2, \dots, p\}$ **do**

$S = \{\hat{\pi}_1, \dots, \hat{\pi}_{m-1}\}$

for $j \in \{\{V \setminus S\} \cap D\}$ **do**

式 (20) を用いて $\hat{S}(m, j)$ を計算する

end for

if $\min \hat{S}(m, j) < \text{閾値}$ **then**

因果順序 m 番目の要素を推定 $\hat{\pi}_m = \arg \min_j \hat{S}(m, j)$

else

for $j \in \{\{V \setminus S\} \cap C\}$ **do**

X_S による X_j の条件付き分散 $\hat{\sigma}_{j|S}^2$ を計算する

end for

因果順序 m 番目の要素を推定 $\hat{\pi}_m = \arg \min_j \hat{\sigma}_{j|S}^2$

end if

end for

Step2: 変数間の関係性の強さの推定

for $m = \{2, 3, \dots, p\}$ **do**

$K = \{\hat{\pi}_1, \dots, \hat{\pi}_{m-1}\}$

目的変数 $X_{\hat{\pi}_m}$ 、説明変数 X_K とし、adaptive Lasso(21)(22) を用いて $\hat{\theta}_{mk}$ を求める

因果順序 m 番目の要素の親変数を推定 $\widehat{Pa}(\hat{\pi}_m) = \{k \in K \mid \hat{\theta}_{mk} \neq 0\}$

end for

グラフ構造を推定 $\hat{E} = \bigcup_{m \in V} \{(k, \hat{\pi}_m) \mid k \in \widehat{Pa}(\hat{\pi}_m)\}$

6 数値実験

本章では数値実験を行い、定義 4.1 による提案モデルが前章で提案した Algorithm 1 によって推定可能であることを示す。また、Algorithm 1 による提案モデルの推定精度が、既存の DAG 推定アルゴリズムより高いことを示す。最後に、離散変数の因果順序を推定する際の閾値の設定に関しても考察する。

6.1 設定

本論文では、サンプルサイズ $n = \{100, 250, 500, 1000, 1500, 2000, 2500\}$ 、頂点数 $p = \{5, 10\}$ 、連続変数の誤差変数の分布 {正規分布, 一様分布} であるデータセットを、提案モデルに従ってそれぞれ 100 個ずつ生成した。その他のデータ生成や推定における設定の詳細は以下とした。

1. p 個の変数をランダムに連続変数と離散変数に割り当てた。連続変数に割り当てられる確率は 0.6、離散変数に割り当てられる確率は 0.4 とした。また、離散変数はポアソン分布と二項分布のいずれかに確率 0.5 で割り当てた。ただし、二項分布のパラメータは Park and Raskutti(2017)[11] と同様に $N_j = 4$ に固定した。
2. グラフの構造をランダムに生成するために、まず全ての要素が 0 である $p \times p$ 隣接行列を生成した。次に、Kalisch and Bühlmann(2007)[6] と同様に、成功確率 s に従う独立なベルヌーイ試行に基づいて隣接行列の下三角成分を 1 に置き換えた。成功確率 s は、各変数と隣接する変数の個数の期待値が 2 または 5 になるようにランダムに設定した^{*1}。隣接行列のゼロ・非ゼロパターンは、提案モデルにおける変数間の関係性の強さを表すパラメータ θ_{jk} のゼロ・非ゼロパターンと同じである。
3. 提案モデルにおける非ゼロのパラメータ $\theta_{jk} \in \mathbb{R}$ は、範囲 $\theta_{jk} \in [-0.5, -0.2] \cup [0.2, 0.5]$ からランダムに選択した。また、定数項 θ_j は、範囲 $\theta_j \in [-0.5, 0.5]$ からランダムに選択した。
4. 連続変数のデータ生成過程における誤差変数の分散 σ_j^2 は、範囲 $\sigma_j^2 \in [0.7, 0.8]$ からランダムに選択した。
5. 離散変数の因果順序を推定する際の閾値は 1.01 とした。
6. 各変数の親変数との関係性の強さを推定する際の adaptive Lasso の調整パラメータについて、 λ は 10 分割交差検証によって選択した。具体的には、「最小の逸脱度 (deviance) + 標準誤差」を超えない最大の λ を選択した。また、 γ は 1 に固定した。

既存の DAG 推定アルゴリズムには、Greedy Equivalent Search (GES)[2]、Max-Min Hill-Climbing (MMHC)[17]、DirectLiNGAM[16]、を用いた。DAG を推定するアルゴリズムは、制約に基づくアプローチ (constraint-based approach) とスコアに基づくアプローチ (score-based approach) に大別される。GES はスコアに基づくアプローチであり、ベイズ情報量規準 (Bayesian Information Criterion; BIC) を用いてモデルの当てはまりが最も良いマルコフ同値類を推定する方法である。MMHC は制約に基づくアプローチとスコアに基づくアプローチの両方を組み合わせたハイブリッドなアルゴリズムである。具体的には、まず制約に基

^{*1} 各変数と隣接する変数の個数の期待値は $s(p-1)$ である [6]

づくアプローチによって各変数の親候補と子候補を求め、次にスコアに基づくアプローチによってモデルの当てはまりが最も良いマルコフ同値類を推定する。1 段階目として制約に基づくアプローチによって親候補と子候補を求めることによって、2 段階目のスコアに基づくアプローチの際の探索空間を小さくしていることが特徴である。DirectLiNGAM は LiNGAM[15] に特化したアルゴリズムであり、回帰分析と独立性の評価を繰り返すことによって因果順序を推定する。

DAG の推定精度は、適合率 (precision) と再現率 (recall) によって測定した。適合率と再現率の算出方法は以下の通りである。

- 適合率 (precision)

$$\frac{\text{正しく推定された辺の個数}}{\text{推定された DAG における辺の個数}}$$

- 再現率 (recall)

$$\frac{\text{正しく推定された辺の個数}}{\text{真の DAG における辺の個数}}$$

ただし、GES と MMHC はマルコフ同値類を推定するアルゴリズムであるため、推定されたグラフの一部の辺は無向辺のままである。そのため、本論文では Park and Park(2019)[9] と同様に、推定されたマルコフ同値類すべてにおける適合率と再現率の平均値で評価する。

6.2 結果

結果だよ

6.3 閾値の設定について

閾値の設定について語るよ～

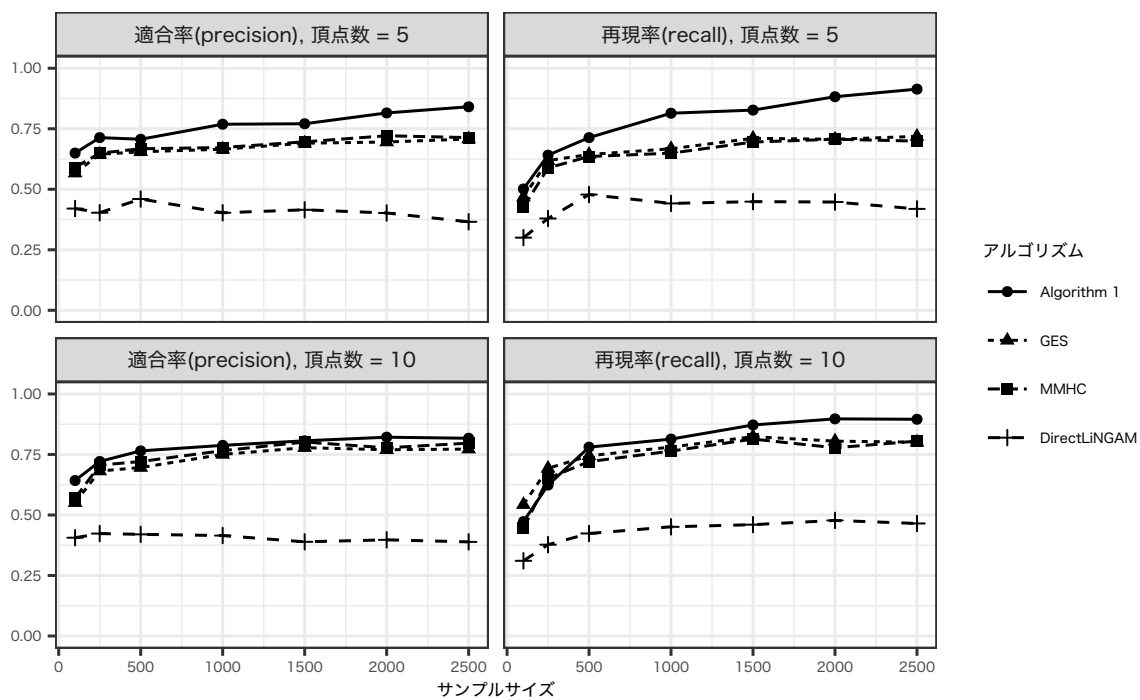


図4 連続変数の誤差変数が正規分布に従う提案モデルにおける各アルゴリズムの精度比較

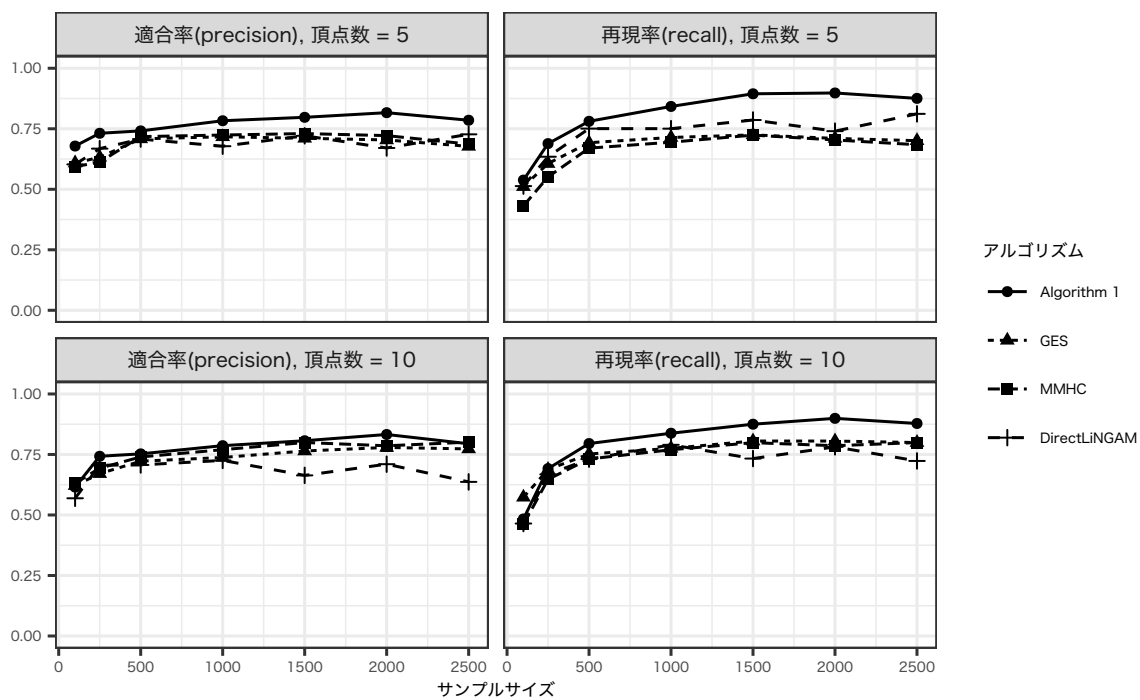


図5 連続変数の誤差変数が一様分布に従う提案モデルにおける各アルゴリズムの精度比較

7 結論

結論だよ～

参考文献

- [1] American marketing association. what is marketing? — the definition of marketing —. <https://www.ama.org/the-definition-of-marketing-what-is-marketing/>. Accessed: 2020-12-21.
- [2] D. M. Chickering. Optimal structure identification with greedy search. *J. Mach. Learn. Res.*, 3(Nov):507–554, 2002.
- [3] A. S. C. Ehrenberg. The pattern of consumer purchases. *J. R. Stat. Soc. Ser. C Appl. Stat.*, 8(1):26–41, Mar. 1959.
- [4] D. Heckerman, D. Geiger, and D. M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Mach. Learn.*, 20(3):197–243, Sept. 1995.
- [5] P. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and others. Nonlinear causal discovery with additive noise models. *Adv. Neural Inf. Process. Syst.*, 2008.
- [6] M. Kalisch and P. Bühlmann. Estimating High-Dimensional directed acyclic graphs with the PC-Algorithm. *J. Mach. Learn. Res.*, 8(22):613–636, 2007.
- [7] C. Li and S. Shimizu. Combining linear Non-Gaussian acyclic model with logistic regression model for estimating causal structure from mixed continuous and discrete data. Feb. 2018.
- [8] G. Park. Identifiability of additive noise models using conditional variances. *J. Mach. Learn. Res.*, 21(75):1–34, 2020.
- [9] G. Park and S. Park. High-Dimensional poisson structural equation model learning via ℓ_1 -Regularized regression. *J. Mach. Learn. Res.*, 20(95):1–41, 2019.
- [10] G. Park and G. 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 eds., *Advances in Neural Information Processing Systems 28*, pp. 631–639. Curran Associates, Inc., 2015.
- [11] G. Park and G. Raskutti. Learning quadratic variance function (QVF) DAG models via overdispersion scoring (ODS). *J. Mach. Learn. Res.*, 18(1):8300–8342, Jan. 2017.
- [12] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009.
- [13] J. Peters and P. Bühlmann. Identifiability of gaussian structural equation models with equal error variances. *Biometrika*, 101(1):219–228, Nov. 2013.
- [14] J. Peters, J. M. Mooij, D. Janzing, and B. Schölkopf. Causal discovery with continuous additive noise models. *J. Mach. Learn. Res.*, 15(58):2009–2053, 2014.
- [15] S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. A linear Non-Gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.*, 7(Oct):2003–2030, 2006.
- [16] S. Shimizu, T. Inazumi, Y. Sogawa, A. Hyvärinen, Y. Kawahara, T. Washio, P. O. Hoyer, and K. Bollen. DirectLiNGAM: A direct method for learning a linear non-gaussian structural equation model. *J. Mach. Learn. Res.*, 12:1225–1248, 2011.
- [17] I. Tsamardinos, L. E. Brown, and C. F. Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Mach. Learn.*, 65(1):31–78, Oct. 2006.
- [18] H. R. Varian. Causal inference in economics and marketing. *Proc. Natl. Acad. Sci. U. S. A.*,

- 113(27):7310–7315, July 2016.
- [19] W. Wenjuan, F. Lu, and L. Chunchen. Mixed causal structure discovery with application to prescriptive pricing. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 5126–5134. ijcai.org, 2018.
 - [20] M. Yamayoshi, J. Tsuchida, and H. Yadohisa. An estimation of causal structure based on latent LiNGAM for mixed data. *Behaviormetrika*, 47(1):105–121, Jan. 2020.
 - [21] J. Zhang and P. Spirtes. Intervention, determinism, and the causal minimality condition. *Synthese*, 182(3):335–347, Oct. 2011.
 - [22] H. Zou. The adaptive lasso and its oracle properties. *J. Am. Stat. Assoc.*, 101(476):1418–1429, Dec. 2006.
 - [23] 萩生田, 繁榊. 順序付きカテゴリカルデータへの因子分析の適用に関するいくつかの注意点. *心理学研究*, 67(1):1–8, 1996.
 - [24] 佐藤忠彦. マーケティングの統計モデル. 朝倉書店, Aug. 2015.
 - [25] 宮川雅巳. 統計的因果推論: 回帰分析の新しい枠組み. 朝倉書店, Mar. 2004.
 - [26] 星野崇宏, 上田雅夫. マーケティング・リサーチ入門. 有斐閣, Dec. 2018.
 - [27] 森岡毅, 今西聖貴. 確率思考の戦略論: USJ でも実証された数学マーケティングの力. KADOKAWA, May 2016.
 - [28] 照井伸彦, ウィラワン・ドニ・ダハナ, 伴正隆. マーケティングの統計分析. 朝倉書店, Sept. 2009.
 - [29] 片平秀貴. マーケティング・サイエンス. 東京大学出版会, 1987.
 - [30] 音部大輔. マーケティングプロフェッショナルの視点. 日経 BP, Apr. 2019.

謝辞

ありがとうございました.