JMRA リサーチイノベーション委員会 2024 年度研究活動

データサイエンス研究会報告書

統計的因果推論

2025 年 9 月 20 日

編集委員　森本　修

目　次

キーワード:DAG 構造、ガウシアン・ネットワーク、共分散選択、相関と偏相関の交換性、傾向スコア、Causal effect、グラフィカルモデリング

１章　問題意識

１.１　研究の趣旨

変数間の相関関係は横断的（クロスセクショナル）な調査データからでも計量できる。しかし「相関には疑似相関がある」という Zeisel (1947)らの警告があるように観察データから因果関係を立証することは困難だった。そのためマーケティング・リサーチでは、これまで無作為割り付けが可能な CLT やオンライン AB テストなどの実験調査を用いて因果推論のリサーチ課題に応えてきた。

因果推論の研究は生物統計学の Wright (1934)によるパス解析を嚆矢として、その後 Blalock (1961)の因果連鎖グラフ、Dempster (1972)の共分散選択、Rosenbaum and Rubin (1983)の傾向スコアそして Pearl (1995)のグラフィカルモデリング（Graphical modeling: GM）という歴史をたどった。近年では機械学習にもとづく統計的因果推論の研究が進展して学術の先端的なテーマになっている。2019 年のノーベル経済学賞も因果推論にもとづく研究であった。

マーケティング・リサーチに統計的因果推論を導入する際の前提条件は何であり、その基本的なアイデアは何なのかを研究する。

１.２　プロジェクト研究の課題

今回のプロジェクト研究において、我々は次の課題意識を持った。

1. GM、とりわけベイジアンネットワーク（Bayesian network: BN）の利用価値
   - ① GM とパス解析、SEM の関係は何か
   - ② 条件付き独立の仮定とは何か
   - ③ 因果の正しい方向をデータから識別できるのか
   - ④ 探索的な目的で因果モデルを作成する方法
2. 施策効果の測定

たとえば A さんにクーポンを付与した上で、その後 A さんの購買行動を観察したとしよう。A さんはクーポンを付与しなくてもその商品を買ったかもしれない。真の効果を測定したければ、A さんにクーポンを付与しなかったときの購買行動と比較すべきだが、当然そのデータは欠測値である。これが「反実仮想」の根本問題であった。反実仮想と機械学習を融合させたモデルで実証実験を行い、機械学習によって施策効果を評価できる根拠と応用上の留意点を明らかにしたい。

２章　多項ベイジアンネットワーク

２.１　ベイジアンネットワーク

いくつかの事象の確率的な関係性をネットワーク図（グラフ構造）と確率分布でモデル化すること、端的に言えば、これがベイジアンネットワークの基本的な考え方である。

ベイジアンネットワークは因果関係や確率的相互作用を効率的に表現し、推論や意思決定に応用できる。

例えば

- 天気（晴れ or 雨）
- 渋滞（あり or なし）
- 遅刻（する or しない）

の 3 つの事象があり、事象と事象の間に

- 雨が降ると渋滞が発生しやすい。
- 渋滞があると遅刻する可能性が高くなる。

という関係があったとする。

それぞれの事象の確率が、

- 雨の確率: 30%（晴れの確率: 70%）
- 雨のとき渋滞が発生する確率: 60%（晴れのとき渋滞は 20%）
- 渋滞があると遅刻する確率: 80%（渋滞がないと遅刻は 10%）

であったとき、ベイジアンネットワークでは以下のように表現される。

この図で確率変数を示す楕円を**ノード**（node）と呼ぶ。

ノード間を繋ぐ矢印は**アーク**（arc）もしくは**エッジ**（edge）と呼ばれ、変数間の因果関係や条件付き依存関係を表す。

この時、矢印の始点となるノードを**親ノード**、矢印を受けるノードを**子ノード**と呼ぶ。

天気は渋滞の親ノードである。渋滞は天気の子ノードであり、同時に遅刻の親ノードである。

各ノードには、依存する事象がない場合は確率分布、依存する事象がある場合は**条件付確率**が紐づけられる。

条件付確率とは Y を条件として X が起きる確率である。先の例でいえば「天気が雨だった場合に渋滞する確率」が条件付確率となる。この関係は依存関係とも呼ばれ「X は Y に依存する」「渋滞は天気に依存する」と表現される。

天気は依存する事象がなく、以下の確率分布が与えられている。

天気［晴れ］	天気［雨］
70%	30%

渋滞は天気に依存しており、以下の条件付確率分布が与えられている。

	渋滞［あり］	渋滞［なし］
天気［晴れ］	20%	80%
天気［雨］	60%	40%

遅刻は渋滞に依存しており、以下の条件付確率分布が与えられている。

	遅刻［する］	遅刻［しない］
渋滞［あり］	80%	20%
渋滞［なし］	10%	90%

ベイジアンネットワークは因果関係を表現することができるが、因果関係を証明するものではない点には注意が必要である。

2.2　有向非循環グラフ（DAG）

ベイジアンネットワークは有向非循環グラフ（directed acyclic graphs、DAG）が前提となる。

DAG はグラフ理論におけるネットワーク構造の 1 つで、以下の特徴を持つ。

1. 有向：エッジに向きがあり、情報が一方通行で流れる。あるノードからあるノードへ情報が流れるが、逆方向に情報が流れることはない。
2. 非循環：始点となるノードからスタートしてエッジをたどり続けても、再び始点となるノードに戻ることはない。

まず、ベイジアンネットワークでは確率的な依存関係や因果関係を表現しようとするため、因果関係や確率的依存関係の方向性を表現するために有向である必要がある。

さらにモデルの推論や解釈を容易にするため循環を回避している。原因と結果の区別を明確にすることで、各ノードが親ノードにのみ依存すると仮定することで大

規模な問題も分解して計算できるようになる。

2.3　ベイジアンネットワークの構築

第１項　問題の定義

まず、ベイジアンネットワークによってモデル化したい課題、分析目的を明確にし、どのような問いに答えたいのかを定義する。

例えば、購買履歴やウェブサイト上の行動ログから購入される可能性が高い商品を推定し、レコメンデーションに活用する。あるいは消費者アンケートのデータから消費者の消費行動を分析し購入意向や満足度の向上につながる要素を特定する。

第２項　モデルに含める変数の選定

定義された問題に基づいて、モデルに含めるべき変数を選定する。変数は確率変数であり、問題に関連する要因や結果となる。

また、モデルに含まれる変数がすべて**離散変数**である場合は多項分布を前提とした**多項ベイジアンネットワーク**、すべて**連続変数**である場合は多変量正規分布を前提とした**ガウシアン・ベイジアンネットワーク**、それらの**混合**である場合は**条件付きガウシアン・ベイジアンネットワーク**として扱われる。

第３項　変数間の依存関係を定義

まず、変数間の依存関係や因果関係を特定し、有向グラフの矢印（エッジ）として表現する。この変数間の関係は、事前の分析や専門家のドメイン知識に基づいて定義される。

この変数間の関係はネットワーク構造として表現されるが、その際 DAG の条件を満たしている必要がある。

次に各変数の確率分布や条件付き確率分布を設定する。

第４項　構造学習

事前に設定された変数間の関係に実データを当てはめて、構造が妥当であるかを確認するのが構造学習である。

構造学習には条件付き独立性検定とネットワークスコアの大きく 2 つの方法がある。

第５項　条件付独立性検定

条件付き独立性検定は個々のエッジの有無を、子ノードの条件付確率の独立性検定によって確認するものである。

離散変数の場合、古典的ピアソンのカイ二乗検定や対数尤度比検定などが利用できる。

連続変数の場合は線形相関の検定、混合の場合は対数尤度比検定が利用できる。

条件付き独立性検定で確認できるのはエッジの有無のみであり、情報の流れる方向については検討できない。

第６項　ネットワークスコア

ネットワークスコアはグラフ構造全体のデータに対する当てはまりの良さを相対的に評価する情報量基準である。

AIC、BIC などが利用される。

これらの方法のみで探索的に構造を決めることは困難である。事前の分析や専門家のドメイン知識に基づいてネットワーク構造を定義したうえで、あるべきエッジ、あってはいけないエッジを指定して、これらの方法を使う必要がある。

第７項　パラメーター学習

構造学習によってグラフ構造（DAG）が決まったら、各ノードの真の確率分布（局所的分布）を推定する。これがパラメーター学習である。

方法として最尤推定法とベイズ推定がよく用いられる。

推定するパラメーターは離散変数であれば各カテゴリの確率値、連続変数であれば平均値と分散である。

2.4　ベイジアンネットワークの利用

構築されたベイジアンネットワークを使って推論を行う。

ここでいう推論とは、特定の条件下でのあるイベントの確率を求めたり、あるイベントの確率が最も高くなる状態を求めたりすることである。

推論には厳密推論と近似推論の 2 つの方法がある。

厳密推論は解が一意に求まるが、推論に時間がかかり大きなネットワークには不向きである。近似推論は処理が高速であり大きなネットワークにも使えるが、乱数に依存するため解が一意に求まらない。

2.5　Ｒによるベイジアンネットワーク

R 言語を用いて生データからベイジアンネットワークを構築し、利用する実際の手順を解説する。

R 言語でベイジアンネットワークを扱うパッケージは複数あるが、ここでは bnlearn パッケージを使う。

```
library(bnlearn)
```

サンプルデータとして、ggmosaic パッケージに含まれる taitanic データセットを用いる。これは 1912 年に沈没したタイタニック号の乗客および乗組員の生存状況のデータである。

このデータセットには以下の 4 つの変数が含まれる。いずれも離散変数である。

- Class: 乗客の客室のランクと乗組員の区別(1st, 2nd, 3rd, Crew)
- Sex: 性別(Male, Female)
- Age: 年齢(Child, Adult)
- Survived: 生存状況(Yes, No)

```
library(ggmosaic)
data(titanic)
```

第１項　問題の定義とモデルに含める変数の選定

ここでは生存状況を示す Survived に対して他の 3 つの変数がどう関係しているのかをモデル化することを目的とする。

第２項　構造学習

今回は事前にグラフ構造を想定せずに構造学習によって適切な DAG 構造を推定する。

まず、あり得ない関係をブラックリストとして定義する。このケースでは Survived が原因とはなりえないので、Survived が親ノードとなるエッジをすべてブラックリストに入れる。

また、Class は Sex、Age の原因にはなりえないので、これもブラックリストに入れる。さらに Sex と Age もお互いに独立した関係であるのでブラックリストに入れる。

- ブラックリスト
  - Survived → Age
  - Survived → Sex
  - Survived → Class
  - Class → Age
  - Class → Sex
  - Sex → Age
  - Age → Sex

ここでネットワークスコアを用いた構造学習を行う。

指標として BIC を使い、ネットワークスコアが最大となる DAG を探すアルゴリズムには山登り法を使う。

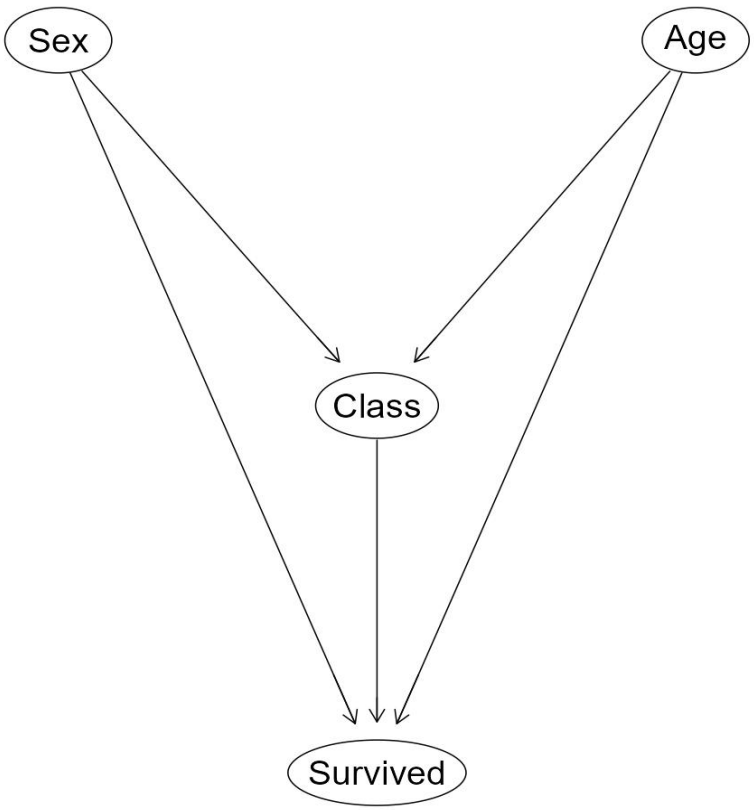
```
# ブラックリストの定義
bl <- matrix(
  c(
    "Sex", "Age",
```

```
    "Age", "Sex",
    "Survived", "Age",
    "Survived", "Sex",
    "Survived", "Class",
    "Class", "Age",
    "Class", "Sex"
  ),
  ncol = 2, byrow = TRUE,
  dimnames = list(NULL, c("from", "to"))
)

# 山登り法
learned_hc <- hc(titanic, blacklist = bl, score = "bic")
```

求められた DAG 構造を描画するには graphviz.plot()関数を用いる。

```
# 結果の描画
graphviz.plot(learned_hc, shape = "ellipse") # 楕円
```



第３項　パラメーター学習

次にこの DAG 構造に基づいて、各ノードの真の確率分布（パラメーター）を推定する。

```
# 最尤推定
bn.mle <- bn.fit(learned_hc, data = titanic, method = "mle")

# ベイズ推定
bn.bayes <- bn.fit(learned_hc, data = titanic, method = "bayes")
```

結果は以下のようになる。最尤推定の場合は観測値と一致する。ベイズ推定の場合は観測値よりも一様分布に近くなるように推定される。

```
# 観測値
prop.table(table(titanic$Age))
#
#     Child     Adult
# 0.04952294 0.95047706

# 最尤推定の結果
bn.mle$Age
#
#   Parameters of node Age (multinomial distribution)
#
# Conditional probability table:
#      Child     Adult
# 0.04952294 0.95047706

# ベイズ推定の結果
bn.bayes$Age
#
#   Parameters of node Age (multinomial distribution)
#
# Conditional probability table:
#      Child     Adult
# 0.04972752 0.95027248
```

第４項　構築されたベイジアンネットワークの利用

構築されたベイジアンネットワークは、分かっていること（エビデンス）を与えて、そのほかの変数（ノード）の同時確率を推論することで様々に利用できる。

推論の方法には大きく二つのやり方がある。

- 厳密推論(exact inference)
  - 解が一意に求まる。低速、大きなネットワークには使えない。
- 近似推論(approximate inference)
  - 高速で大きなネットワークにも使える。乱数によって解がぶれる。

近似推論を実行する例は以下である。

```
# 性別が女性の場合の生存割合
bn.bayes |>
  cpquery(
    event = (Survived == "Yes"),
    evidence = (Sex == "Female"),
    method = "ls"
  )
# [1] 0.7497666

# 成人男性の場合の生存割合
bn.bayes |>
  cpquery(
    event = (Survived == "Yes"),
    evidence = (Sex == "Male" & Age == "Adult"),
    method = "ls"
  )
# [1] 0.2114421
```

厳密推論を実行するには gRain パッケージを利用する。

```
library(gRain)
junction.bayes <- bn.bayes |>
  as.grain() |>
  compile()

# 性別が女性の場合の生存割合
junction.bayes |>
  evidence_add(list(Sex="Female")) |>
  querygrain(node = "Survived")
# $Survived
# Survived
#       No      Yes
# 0.2626725 0.7373275
```

```r
# 成人男性の場合の生存割合
junction.bayes |>
  evidence_add(list(Sex="Male", Age="Adult")) |>
  querygrain(node = "Survived")
# $Survived
# Survived
#       No      Yes
# 0.797196 0.202804
```

2.6　まとめ

本稿では、離散変数データに対する強力なモデリングツールである多項ベイジアンネットワークについて解説した。R 言語と bnlearn パッケージを用いた具体的な実装例を通じて、構造学習、パラメーター学習、そして推論といった一連のプロセスを示した。重要なポイントとして、事前の知識や仮説に基づいてネットワーク構造を定義すること、そして目的に応じて厳密推論と近似推論を使い分けることが挙げられる。これらの知識を活かし、ビジネスや研究における様々な問題解決にベイジアンネットワークが活用されることを期待する。

３章　連続量のベイジアンネットワーク

３.１　概要

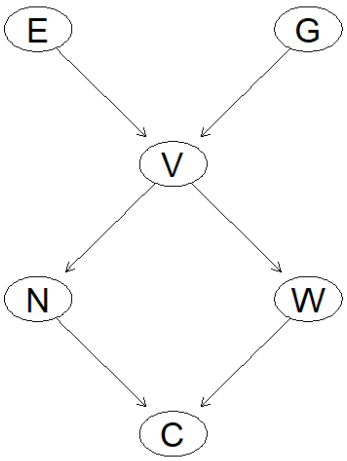
本章では、連続型データに適用されるベイジアンネットワークを扱う。多変量正規分布を想定したモデルであり、正規分布を最初に研究したカール・フリードリッヒ・ガウスに因んで、ガウシアン・ベイジアンネットワークとも呼ばれる（以降、GBNと表記する）。

構造学習→パラメータ推定→ネットワークを用いた推論、という流れは、離散型データにおける多項ベイジアンネットワークと同様である。

以下、GBN にみられる特徴と、その特徴にまつわる疑問点に絞って解説する。

３.２　ＧＢＮの特徴

GBNの主要な特徴を紹介するにあたり、具体例として次の図のような DAG を想定する（スクタリ=デニス『R と事例で学ぶベイジアンネットワーク』（金明哲監訳、財津亘訳、共立出版、2022 年）P39 の例。以下、スクタリ=デニスと表記）。



本事例は、環境的因子（E）と遺伝的因子（G）が栄養器官（V）に作用し、栄養器官（V）が種子の数（N）と重量（W）の決定要因となり、最終的に収穫量（C）が導かれる、という仮定に基づくモデルである。DAG は、変数間の条件付き独立関係をグラフで表現している（矢印が引かれていないノード間は独立のものと扱われる）。

この事例の基づき、GBN の主要な特徴を挙げる。

【特徴1】推定すべきパラメータが少なく済む。単に多変量正規分布を確定したければ、通常は「p個の平均値、p個の分散、1/2p(p−1)個の相関係数」が必要である。事例のケースは6変数なので、大域的分布を考えると、「6個の平均値、6個の分散、15個の相関係数」、つまり、27 個のパラメータが必要になりそうである。しかし、GBN

では、親ノードの値で条件づけた各ノードの局所的分布を特定すればよい。事例に当てはめると、推定すべき未知のパラメータは17個と少なくて済む[1]。

【特徴2】DAG の親ノードと子ノードで構成された部分集合をみると、交互作用項を含まない、回帰分析の理想的なモデル式となっている。 例えば DAG の変数 V の分布は $V \sim N(\alpha + \beta_1 E + \beta_2 G, \sigma^2)$ と表現される。EとG が独立なので、交互作用項は含まれていない。つまり、E と G を説明変数とし、V を目的変数とした重回帰分析と同じモデルとなっている。このことから、GBN が、単に部分ごと(重)回帰分析モデルを組み合わせたものであることがわかる。

【特徴3】パラメータ推定は、検討対象のノードとその親ノードで表される部分集合のみのデータをベースに行われている。具体例として、以下の(画像1)は、事例の DAG におけるパラメータ推定結果のうち、V の推定結果を抜粋したものである。なお、分析は R で実施した。

(画像1)

```
  Parameters of node V (Gaussian distribution)

Conditional density: V | E + G
Coefficients:
(Intercept)             E             G
-10.4547647    0.7426946    0.4552872
Standard deviation of the residuals: 5.034309
```

そして、次の(画像2)は部分集合のみをベースにした重回帰分析の結果である。単純に、V を従属変数、E と G を独立変数とした重回帰分析の結果である。

---

1 具体的に見てみると、まず、親ノードのいない外生変数 E(G も同様)は以下の数式のように平均値と分散からなる正規分布で表現されるため、2 つのパラメータが必要となる。

$$E \sim N(\alpha, \sigma^2)$$

親ノードが 2 つある V は、以下のように親ノードを説明変数とした回帰モデルで表現される(切片と回帰係数と分散)ため、4 つのパラメータが必要となる。

$$V \sim N(\alpha + \beta_1 E + \beta_2 G, \sigma^2)$$

W(N も同様)は親ノードが 1 つなので、以下の数式の通り、3 つのパラメータが必要となる。

$$W \sim N(\alpha + \beta V, \sigma^2)$$

C については、以下のように切片のない式となるので、3 つのパラメータが必要となる

$$V \sim C(\beta_1 N + \beta_2 W, \sigma^2)$$

以上から、(2×2)+4+(3×2)+3=17(個)と、少ないパラメータで済む。

なお、スクタリ=デニス P44 では、C の切片を入れて 18 個としている。しかし、種子の数と大きさがないのに収穫量が発生するはずがないので、C の切片はセロで既知の値とすべきである。

(画像2)

```
> lmV <- lm(V ~ E + G, data = cropdata200[, c("E", "G", "V")])
> coef(lmV)
(Intercept)           E           G
-10.4547647   0.7426946   0.4552872
```

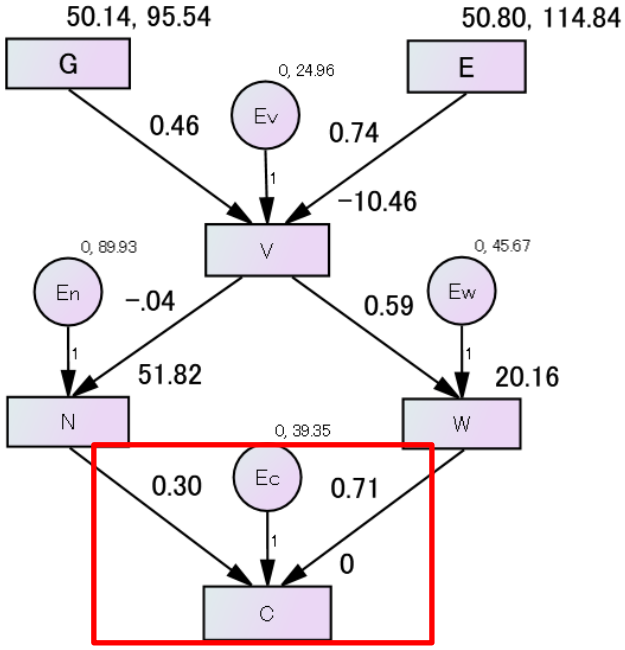
赤線部分を見ると、推定された各パラメータが完全一致していることが分かる。この結果から、GBNは、親ノードと子ノードで構成される部分集合のみで重回帰分析を行うという、シンプルな考え方で成り立っているといえる。

3.3　GBNに関する疑問点

以上のＧＢＮの特徴に関連して、次のような疑問が生じる。

【疑問点１】パラメータ推定が部分集合のみのデータをベースに行われているのであれば、GBN はパス解析とどのような違いがあるのだろうか。GBN とパス解析のパラメータ推定結果を比較し、両者の違いを考察する。

まず、以下は、本事例と同じデータを用いたパス解析の結果（非標準化推定値）である。なお、解析は AMOS で実施した。



そして、以下が GBN によって推定されたパラメータである（ノード C のみ抜粋）。

```
  Parameters of node C (Gaussian distribution)

Conditional density: C | N + W
Coefficients:
(Intercept)             N             W
  0.0000000     0.2963220     0.7105197
Standard deviation of the residuals: 6.304969
```

推定されたパラメータ（切片と回帰係数）がパス解析における推定値と一致していることが確認できる（もっとも、切片は意図的にゼロを指定している）。
では、パス解析と推定結果が同じであるならば、GBN を用いるメリットはどこにあるのだろうか。

私見では、パス解析にはない GBN のメリットとして、推定されたパラメータを付与した DAG を用いて、各ノードの分布の推論が可能な点を挙げたい。例えば、C を一定の値にした場合の V の分布や、V を任意の値にした場合の C の条件付き分布を計算することができる。このような推論により、基準変数(C)をある一定の値にしたいという目的のもと、説明変数がどのような条件を満たしていればよいかのシミュレーションが可能である。

例えば、以下は、C を仮に 80 とした場合の V がとり得る分布を求めた結果である。

```
> print8mn(condi4joint(mn.rbmn, par = "V", pour = "C", x2 = 80))
     mu  s.d.
V 80.71 8.542
```

【疑問点２】GBN は多変量正規分布を想定しているが、どの程度の正規分布であればよいのだろうか。必須の要件なのだろうか。

本事例で使用したデータについて、多変量正規性の検定(Henze-Zirkler test)を実施したところ、以下のように、p 値が 0.05 より大きく、多変量正規性を満たしていることがわかる。

```
> # 多変量正規性検定
> result <- mvn(data = cropdata200)
> print(result)
$multivariate_normality
           Test Statistic p.value     Method       MVN
1 Henze-Zirkler     1.002   0.056 asymptotic ✓ Normal
```

この時、GBN の評価指標は、以下のようになる（BIC と BGe）

```
> score(dag.bnlearn, data = cropdata200, type = "bic-g")
[1] -4201.576
>
> score(dag.bnlearn, data = cropdata200, type = "bge")
[1] -4274.598
```

次に、データを一部修正し、多変量正規分布を満たさないデータを作成した。p 値が 0.05 を下回り、多変量正規性を満たしていないことがわかる。

```
> # 多変量正規性検定
> result <- mvn(data = cropdata200)
> print(result)
$multivariate_normality
           Test Statistic p.value     Method          MVN
1 Henze-Zirkler      1.02   0.026 asymptotic X Not normal
```

この時 GBN の評価指標は、以下のようにやや悪化している（マイナスがかかっているので、プラス方向に数値が大きい方が良い）。

```
> score(dag.bnlearn, data = cropdata200, type = "bic-g")
[1] -4218.408
>
> score(dag.bnlearn, data = cropdata200, type = "bge")
[1] -4291.849
```

以上の結果より、多変量正規性が満たされていない場合は、モデル評価指標が悪化することが分かった。ただ、悪化の程度はケースによると思われ、必ずしも分析不能とは言えない。多変量正規性の検定をクリアするかを確かめ、満たされていない場合はデータ変換(Box-Cox 変換)などで正規分布に近づける方法が検討可能である。ただ、実際は、特にアンケート調査データにおいては、多変量正規性が満たされる場合はあまり多くないと想定されるため、多変量正規性を厳密に求めるならば、アンケート調査に GBN を使う場面は限定されることになろう。

3.4　まとめ

GBNの特徴をまとめる。GBN においては、①DAG によってノード間の局所的依存関係が示されており、各ノードはその親ノードに条件づけられているため、親子の依存関係のみに着目すればよい。そのため、推定すべきパラメータが少なくて済む。そして、②それぞれの局所的分布は交互作用項を含まない線形回帰モデルとして表現でき、③パラメータ推定は、検討対象のノードとその親ノードで表される部分集合のみのデータをベースに行われている(つまり、部分部分で重回帰分析を行っているだけの、簡単なモデル)、ということになる。

そして、その仕組みは基本的にパス解析と同じであるものの、シミュレーションを可能とする点に GBN のメリットがある。そして、モデルの信頼性を確保するために多変量正規性を満たしていることが望ましいが、厳密に求めるならばアンケート調査データへの適用可能性は限定される。データ変換による正規性の確保などを通じた、アンケート調査への応用可能性を探っていきたい。

４章　ベイジアンネットワークの理論

4章は、ベイジアンネットワーク（BN）の理論的基盤からビジネス、特にマーケティング領域への実践的応用までを接続することを目的とする。まず、BN を数理的に理解する上で不可欠な「条件付き独立性」の概念を解説する。次に、この概念がどのようにして複雑な「同時確率分布の因子分解」を単純化し、計算を効率化する（=さぼる）ことを可能にするかを示す。この数理的表現が、視覚的には「ネットワーク構造図」として、局所的には「局所マルコフ性」として現れることを明らかにする。

さらに、理論から実践へ移る際の最大の課題である「モデルの構造学習」に焦点を当てる。精度と計算量のトレードオフを持つ 2 つの主要なアルゴリズム系統（A. 制約ベース、B. スコアベース）を紹介し、ビジネス現場で有効なハイブリッドアプローチを考察する。最後に、具体的なマーケティング事例を想定したサンプルデータと R コードを通じて、これらの理論がどのように実装され、実用的な知見を導き出すために活用できるかを示す。

４.１　条件付き独立性と同時確率分布の単純化

複数の確率変数 $X_1, \dots, X_n$ がなす同時確率分布 $P(X_1, \dots, X_n)$ は、連鎖律によって常に式（4.1）のように分解できる。

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \dots, X_{i-1}) \qquad (4.1)$$

しかし、このままの形で確率分布を定義しようとすると、組合せ爆発という深刻な問題に直面する。例えば、各変数が「はい/いいえ」の 2 つの状態しか取らない単純なケースでも、変数が 10 個あれば状態の総組合せは 2^{10} =1,024 通りになる。この確率分布を完全に定義するには、1,023 個の独立な確率値を指定する必要があり、変数が 20 個になればその数は 100 万を超える。このように、変数の数が増えるにつれてパラメータ数が指数関数的に増加する現象は次元の呪いとも呼ばれ、現実的な計算を不可能にする。この計算量の爆発こそ、ベイジアンネットワークが条件付き独立性を用いて計算をさぼることで解決しようとする根源的な課題である。

ベイジアンネットワークは、有向非巡回グラフ（DAG）を用いて変数間の依存関係を定義し、「ある変数は、その親ノードが与えられれば、その親以外の（非子孫の）ノードとは独立である」という局所マルコフ性を仮定する。この仮定により、複雑な同時確率分布の式（4.1）は、以下の単純な形（4.2）で表現することが可能となる。

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P\big(X_i | Parents(X_i)\big) \qquad (4.2)$$

この数式で表現される「局所的な条件付き独立性の集合」が、ビジュアル的には DAG、すなわちネットワーク構造図として表現される。矢印の親子関係が、どの変数

(子)がどの変数(親)によって直接影響を受けるかを示している。この依存構造は、3つのノード間の基本的な関係性(直列、分岐、合流)に分解して理解することができる。これらの基本形における情報の伝搬(または遮断)のルール、すなわちｄ分離を理解することが、ネットワーク全体の独立性を読み解く基礎となる。

4.2　データからの構造学習

理論的な基盤を理解した上で、ベイジアンネットワークをビジネスで活用するには、データからネットワーク構造そのものを推定する「①モデルの構造学習」と、得られたモデルを用いて予測や診断を行う「②推論」のステップがある。特に、未知のデータから変数間の因果・依存関係の構造を発見する①のフェーズが、ビジネス活用の成否を分ける最も重要な課題となる。
構造学習アルゴリズムは、主に以下の 2 つのアプローチに大別される。

A. 制約ベースのアプローチ (Constraint-based)

考え方: データに対して統計的な条件付き独立性テスト(例:カイ二乗検定)を繰り返し行い、独立と判断された変数間の繋がり(エッジ)を「削ることで構造を決定する。
長所: 統計的妥当性が高く、比較的正確な因果構造を発見できる可能性がある。
短所: 変数の組み合わせを網羅的に検定するため、計算量が非常に多くなりがちで、大規模データには不向き。

B. スコアベースのアプローチ (Score-based)

考え方: ネットワーク構造全体の「良さ」を示すスコア(例:BIC、AIC などの情報量規準)を定義し、スコアが最も良くなる構造を探索的に「引く」ことで構造を決定する(例:山登り法)。
長所: 計算量が少なく高速であり、大規模なデータセットにも適用可能。
短所: 探索的なため局所最適解に陥りやすく、必ずしも真の構造を捉えられるとは限らない。

実践的なハイブリッド戦略

ビジネスの現場では、これら 2 つのアプローチを融合させることが有効である。

1. ドメイン知識の活用: まず、ビジネス担当者の知見から「ありえない因果関係(ブラックリスト)」や「確実な因果関係(ホワイトリスト)」を定義し、探索空間を限定する。
2. スコアベースでの探索: 上記の制約下で、高速なスコアベースの手法を用いて、大まかなモデル構造の候補を効率的に見つけ出す。
3. 制約ベースでの検証: 導出されたモデル構造の中で、特に解釈が重要となる部分や、統計的に曖昧な部分について、制約ベースの手法で追加検証を行い、モデルを洗練させる。

この戦略により、計算の現実性とモデルの妥当性を両立させることが可能となる。

4.3　マーケティングへの応用事例とＲ実装

ある EC サイトが、顧客のコンバージョン（商品購入）に至る要因を分析したいと考える。データとして、顧客の年齢層（Age）、居住地域（Region）、使用デバイス（Device）、流入チャネル（Channel）、リピート訪問か（RepeatVisitor）、サイト滞在時間（SessionTime）、そして購入の有無（Conversion）という 7 つの変数を考慮する（図 4.1）。これらの変数間の複雑な関係性をベイジアンネットワークでモデル化し、「構造推定の正確性」、「予測精度の検証」、「計算効率の検証（組合せ爆発の回避）」という 3 つの側面から総合的に評価することを目的とする。

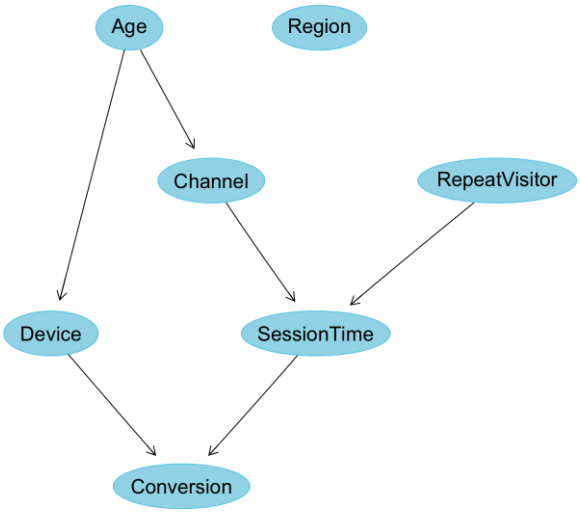


図 4.1　変数の関係（真の構造）

1. **構造推定の正確性**：データ生成時にプログラムした変数間の依存関係（真の構造）を、モデルがどの程度正確に復元できるか。
2. **予測精度の検証**：特定の条件下での購入確率を、モデルがどの程度正確に予測できるか。
3. **計算効率の検証（組合せ爆発の回避）**：モデル構築に必要なパラメータ数を、総当たりモデルと比較してどの程度削減できているか。

1. 構造推定の正確性

モデルがデータから変数間の因果・依存関係の「地図」を正しく描けているかを確認した。データ生成時に仕込んだ「真の構造」と、アルゴリズムがデータから学習した「推定構造」を比較した。

評価：

推定された構造は、真の構造と完全に一致した。 Age が Device や Channel に影響を与える関係や、最終的に Conversion に至るまでの矢印の流れ、Region が独立している点など、全ての依存・独立関係がデータから正しく復元されている。これは、モデルがデータの本質的な構造を正確に捉える能力を持つことを示している。

2. 予測精度の検証

正しく推定された構造に基づき、モデルが具体的な確率をどの程度正確に予測できるかを「真の確率」と比較した。

真の確率（答え）	0.620
モデルによる予測	0. 632
誤差	1.96 %

評価:

モデルの予測値と真の値の誤差はわずか 1.96%であり、構造だけでなく、その背後にある確率値についても非常に高い精度で学習できていることが確認できた。

3. 計算効率の検証（組合せ爆発の回避）

モデルの複雑さを、構築に必要なパラメータ（確率値）の数で評価した。

総当たりモデル（独立性を仮定しない）	431 個
ベイジアンネットワークモデル	24 個
削減率	94.43 %

評価:

ベイジアンネットワークは、正しい構造を学習することで不要な関係性を無視し、考慮すべきパラメータ数 94.43%削減した。これにより、計算コストが劇的に下がり、モデルの解釈も容易となる。

4.4 まとめ

ベイジアンネットワークは、変数間の不確実な依存関係を「条件付き独立性」という数学的基盤と、「グラフ構造」という視覚的直感性を融合させてモデル化する強力なツールである。ビジネス、特に多要因が複雑に絡み合うマーケティング分析において、その応用可能性は非常に大きい。

実践における最大の鍵は、データから適切なモデル構造を発見する構造学習にある。本稿で議論したように、計算効率に優れたスコアベースの手法と、統計的妥当性の高い制約ベースの手法、そして何よりも重要なビジネス現場のドメイン知識を組み合わせたハイブリッドアプローチが、現実的かつ効果的な解となる。

4.5　Ｒによるサンプルコードとデータ

以下に、R の bnlearn パッケージを用いて実装するコードを示す。

```r
# --- ベイジアンネットワーク R サンプルコード ---

# ----------------------------------
# 1. パッケージのインストールと読み込み
# ----------------------------------
# install.packages("bnlearn")
# if (!requireNamespace("BiocManager", quietly = TRUE))
#     install.packages("BiocManager")
# BiocManager::install("Rgraphviz")

library(bnlearn)
library(Rgraphviz)

# ----------------------------------
# 2. サンプルデータ
# ----------------------------------
set.seed(1234)
n <- 2000

# --- 変数（7 変数） ---
Age <- sample(
  c("Young", "Middle", "Senior"),
  n,
  replace = TRUE,
  prob = c(0.4, 0.4, 0.2)
  )
Region <- sample(
  c("Kanto", "Kansai", "Other"),
  n,
  replace = TRUE,
  prob = c(0.5, 0.3, 0.2)
  )
Device <- factor(
  ifelse(
    Age == "Young",
    sample(
```

```
      c("Mobile", "Desktop"),
      n,
      replace = TRUE,
      prob = c(0.8, 0.2)
      ),
    sample(
      c("Mobile", "Desktop"),
      n,
      replace = TRUE,
      prob = c(0.4, 0.6)
      )
    )
  )
RepeatVisitor <- factor(
  sample(
    c("Yes", "No"),
    n,
    replace = TRUE,
    prob = c(0.6, 0.4))
  )
Channel <- factor(
  ifelse(
    Age == "Young",
    sample(
      c("SNS", "Search", "Direct"),
      n,
      replace = TRUE,
      prob = c(0.6, 0.3, 0.1)
      ),
    sample(
      c("SNS", "Search", "Direct"),
      n,
      replace = TRUE,
      prob = c(0.2, 0.7, 0.1)
      )
    )
  )
SessionTime_prob <- ifelse(
  Channel == "Search" | RepeatVisitor == "Yes",
  0.7,
```

```
  0.4
  )
SessionTime <- factor(
  sapply(
    SessionTime_prob,
    function(p) sample(c("Long", "Short"), 1, prob = c(p, 1-p))
    )
  )
Conversion_prob <- ifelse(
  SessionTime == "Long" & Device == "Desktop",
  0.8,
  ifelse(
    SessionTime == "Long" & Device == "Mobile",
    0.6,
    0.2
    )
  )
Conversion <- factor(
  sapply(
    Conversion_prob,
    function(p) sample(c("Yes", "No"), 1, prob = c(p, 1-p))
    )
  )

# 全変数をデータフレームにまとめる
customer_data <- data.frame(
  Age, Region, Device, Channel, RepeatVisitor, SessionTime, Conversion
  )
customer_data[] <- lapply(customer_data, as.factor)

# ----------------------------------
# 3. 構造学習(スコアベース:山登り法)
# ----------------------------------
bn_model <- hc(customer_data, score = "bic")

# ----------------------------------
# 4. ネットワーク構造の可視化
# ----------------------------------
graphviz.plot(
  bn_model,
```

```
  layout = "dot",
  shape = "ellipse",
  highlight = list(
    nodes = nodes(bn_model),
    col = "skyblue",
    fill = "lightblue"
    )
  )

# ----------------------------------
# 5. パラメータ学習
# ----------------------------------
fitted_model <- bn.fit(bn_model, customer_data)

# ----------------------------------
# 6. シミュレーションと精度検証
# ----------------------------------
print("--- シミュレーション:モデル予測 vs 真の確率 ---")

# --- 検証シナリオ ---
#「関東在住の中年層(Middle)がデスクトップ経由でリピート訪問した場合の
# コンバージョン確率」
evidence_list <- list(
  Region = "Kanto",
  Age = "Middle",
  Device = "Desktop",
  RepeatVisitor = "Yes"
  )

# (A) 学習したモデルによる予測確率
model_prediction <- cpquery(
  fitted_model,
  event = (Conversion == "Yes"),
  evidence = (
    Region == "Kanto"
    & Age == "Middle"
    & Device == "Desktop"
    & RepeatVisitor == "Yes"
    )
)
```

```r
cat(
  sprintf(
    "モデルによる予測確率 P(Conversion=Yes | 証拠) = %.4f¥n",
    model_prediction
    )
  )

# (B) データ生成ルールに基づく「真の確率」の計算
# 1. 証拠(evidence)から、Conversion の親ノードである SessionTime の確率を計算
#    証拠に RepeatVisitor="Yes"が含まれるため、データ生成ルールにより
SessionTime="Long"になる確率は 0.7
p_session_long_true <- 0.7
p_session_short_true <- 1 - p_session_long_true

# 2. SessionTime の状態ごとに Conversion="Yes"になる確率を、全確率の公式を
# 使って計算
#    P(Conv=Yes) = P(Conv=Yes|Session=Long) *
#       P(Session=Long) + P(Conv=Yes|Session=Short) * P(Session=Short)
#    証拠に Device="Desktop"が含まれるため、
#    P(Conv=Yes|Session=Long, Device=Desktop)はデータ生成ルールにより 0.8
#    SessionTime が"Short"の場合、P(Conv=Yes)はデータ生成ルールにより 0.2
true_probability <- (0.8 * p_session_long_true) + (0.2 *
p_session_short_true)
cat(
  sprintf(
    "データ生成ルールに基づく真の確率 = %.4f¥n¥n",
    true_probability
    )
  )

# (C) 精度評価
error_rate <- abs(model_prediction - true_probability) /
true_probability * 100
cat(sprintf("モデルの予測と真の確率の誤差: %.2f %%¥n", error_rate))

# ------------------------------------
# 7. 組合せ爆発とパラメータ削減効果の確認
# ------------------------------------
# 組合せ爆発とパラメータ削減効果
n_levels <- sapply(customer_data, nlevels)
```

```r
full_params <- prod(n_levels) - 1
cat(sprintf("各変数のカテゴリ数: %s¥n", paste(
  names(n_levels), n_levels, sep = "=", collapse = ", ")))
cat(
  sprintf(
    "全変数を考慮した場合の総組合せ数: %s 通り¥n",
    format(prod(n_levels), big.mark = ",")
    )
  )
cat(
  sprintf(
    "=> 必要なパラメータ数 (独立な確率値): %s 個¥n¥n",
    format(full_params, big.mark = ",")
    )
  )

bn_params <- nparams(fitted_model)
cat(
  sprintf(
    "学習したベイジアンネットワークが必要とするパラメータ数: %s 個¥n¥n",
    format(bn_params, big.mark = ",")
    )
  )
reduction_rate <- (1 - bn_params / full_params) * 100
cat(
  sprintf(
    "ベイジアンネットワークによるパラメータ削減効果: %.2f %% 削減¥n",
    reduction_rate
    )
  )
```

５章　傾向スコアと二重にロバストな推定法

５.１　マーケティングにおける効果検証

マーケティング戦略の実行時のフレームワークとしてよく使われるのがマーケティングの 4P（Product：製品、Price：価格、Place：チャネル、Promotion：プロモーション）である。本節ではマーケティングの4P のなかでもプロモーション領域の効果検証において、施策効果を正しく検証するための因果推論手法として傾向スコアを利用した二重にロバストな推定法と呼ばれるバイアス補正手法について解説を行いたい。

マーケティングの4P におけるプロモーションとは、広告や広報（PR）、販売促進（セールスプロモーション）といった、消費者に対して自社のブランドや商品、サービスの認知向上、理解促進、購買促進を行う幅広い施策群を含んでいる。一般的に広告や PR は消費者の認知向上や理解促進に対して効果が高く、施策効果は長期的な効果となり、販売促進は消費者の購買促進に対して効果が高く、施策効果は短期的な効果になることが知られている。

こうした手法のなかから、特に販売促進施策において業種を問わず比較的よく利用される手法であるクーポンを使った販売促進施策について効果検証手法の検討を行いたい。伝統的な紙メディア（DM や POS クーポン、折り込みチラシチラシなど）からデジタルメディア（メールやアプリ、web サイトを通じた通知・告知など）まで経路は様々だが、クーポンを使ったインセンティブ（値引きやポイント付与など）による消費者への販売促進施策は実務の現場でよく使われる手法である。近年では企業のデジタル化の進展に伴って顧客データがデータベース化されていることも多く、顧客単位でクーポン施策を行う事が容易な環境が整っている。こうした環境を前提に顧客単位でクーポン施策の実行を行う際の効果検証方法について確認したい。

５.２　販売促進施策と効果検証

クーポンを使った販売促進施策の効果検証手法として最もよく使われる手法は A/B テストだろう。これはランダム化比較試験（RCT）のビジネス文脈での呼称であり、検証したい効果をランダムに分割した二群（一方は介入群、本節ではクーポンの送付、他方は対照群、クーポン送付せず）の比較を通じて施策効果（キャンペーン期間内の購買回数や購買金額）の差分を施策の実施効果として推定、検証する手法である。

A/B テストは、無作為割付によりクーポン送付群と非送付群の2群に分割することで潜在的交絡因子（ここではクーポン利用に影響するであろう消費者の属性や過去の商品・サービスの利用履歴など）を両群で均等に分布させることができる。このように分割された2群の間では、クーポン送付の有無以外は消費者の特性に違いがな

い状態となり、両群の差を見ることで偏りのない施策効果の推定量を得る事ができる。

クーポン施策のような販売促進施策は、実行時に何らかの条件に従ってクーポン配布を行う顧客グループを事前に決めるケースが多い（特定のターゲットセグメントやターゲティングを行わないランダムな配布など）。このように事前に対象顧客を設定できる販促施策の場合、施策実行前に対象顧客リストをランダムに分割することでA/B テストによる効果検証が可能である。

一方で事前に対象顧客を設定できない販売促進施策も存在する。参加型のキャンペーン（アプリや web サイトの告知によって顧客のエントリを促し、エントリした顧客にのみキャンペーン期間中にクーポンやポイント付与等を通じたインセンティブを付与する施策）はその一例である。参加型のキャンペーンの場合、事前にキャンペーン参加する顧客を特定してランダム分割することはできないため、効果検証時にA/B テストを使うことができない。このようなケースの場合、参加者はキャンペーン対象のサービスや商品に対するロイヤリティの高いユーザーに偏る傾向があり、処置群（キャンペーン参加者）と対照群（キャンペーン不参加者）の間で消費者特性に違いがでてしまうことが多い。

このようにキャンペーン参加の有無により、消費者間に特性の違いがある場合、A/B テストのように両群のキャンペーン期間中の利用実績の単純平均の差をキャンペーンの施策効果の推定量として捉えると問題がある事が多い。施策処置群と対照群の間に大きな特性の違いがある場合、これをバイアスとして補正し、比較可能な状態にして効果検証を実施する必要が出てくる。このようなバイアス補正の際に使われる手法の一つが二重にロバストな推定法である。

5.3　二重にロバストな推定法と傾向スコア

今回想定しているキャンペーン参加の有無による 2 群の消費者グループの違いを補正する方法について具体的に確認しよう。本来 A/B テストが使える状況であれば、キャンペーン不参加群の消費者の特性（潜在的交絡因子、以降は共変量 X として記述 ― 属性や過去の購買履歴など）について参加群の消費者特性とほぼ同等になることが期待される。しかし現実のデータではキャンペーン不参加群の消費者の特性が参加群の消費者の特性と大きく異なっているケースがしばしば発生する。このような、特性の異なる 2 群をそのまま比較することはできないため、参加群と不参加群の 2 群の消費者の特性の違いを踏まえて、データに対して補正を行う事で比較可能にしたい。この補正を行う方法として二重にロバストな推定法を紹介する。

二重にロバストな推定法は以下のようにデータに対して補正を行う。

① 処置群（参加者）の消費者に対して、処置がなかった場合（キャンペーンに参加しなかった場合）の結果（キャンペーン期間中の売上や利用率）を対照群（非参加者）のデータを使ったモデルを作成することで推定する（反実仮想）。この推定結果と実際の結果との差分を施策の実施効果とする
② 対照群（非参加者）のデータを使ったモデルの推定値と実際の対照群の消費者の結果の差分を傾向スコアを使ったウェイトで補正する
③ ①の効果を②の結果で補正する

このように補正を行って算出したキャンペーン参加による施策効果の推定量をATT（average treatment effect for the treated）推定量 と呼ぶ。この推定量を$\hat{\tau}_{(DR,ATT)}$として、二重にロバストな推定法は以下のように記述される（Moodie, Saarela & Stephens 2018）。

$$\hat{\tau}_{(DR,ATT)} = \frac{1}{N_1}\sum_{i=1}^{N_1}\left(Y_i^{\,1} - g_0(X_i)\right) - \frac{1}{N_1}\sum_{i=1}^{N_0}\frac{e(X_i)}{1-e(X_i)}\left(Y_i^{\,0} - g_0(X_i)\right)$$

ここで、

N_1 ：処置群のサンプル数
N_0 ：対照群のサンプル数
Y^1 ：処置を受けた際の結果（キャンペーン期間中の売上や利用率）
Y^0 ：処置を受けなかった際の結果（キャンペーン期間中の売上や利用率）
X ：共変量（属性や過去の購買履歴など）
$g_0(X)$ ：対照群データ（のみ）を使ったXによるYを推定するモデル
$e(X)$ ：共変量Xに基づく傾向スコア
とする。

右辺第１項が①の実際のキャンペーン期間中の結果と反実仮想の推定量との差分を表している。右辺第２項は$Y_i^{\,0} - g_0(X_i)$の部分が、対照群（非参加者）のデータを使ったモデルの推定値と実際の対照群の消費者の結果の差分の推定量である。$\frac{e(X_i)}{1-e(X_i)}$の部分は傾向スコアを使った重み付けによる補正項となる。二重頑健法は、結果変数であるＹを推定するモデルと傾向スコアを推定するモデル、２つのモデルを組み合わせることでよりロバストな（モデルの誤指定に対して頑健な特性を持つ）推定をおこなう手法である。

傾向スコア$e(X)$は処置割付け（今回の例ではキャンペーン参加）を示すダミー変数をZ、共変量（今回の例では消費者属性や過去の利用履歴）をXとしたとき個々の

データが処置（キャンペーン参加）に割り付けられる確率として表現される。

$$e(X) = P(Z = 1|X)$$

実際の傾向スコアの算出は、目的変数をZ、説明変数をXとしたロジスティック回帰などの推定法を使って行われることが多い[2]。算出されたスコアは処置への割付け確率なので、今回の例でいえばスコアが高いほどキャンペーン参加した消費者の特性に近い特性を持つことを意味する。

一方で結果変数Yを推定するモデル$g_0(X)$は、ATT 推定量の場合、対照群のデータのみを利用して目的変数をY、説明変数をXとした推定法で行う。今回の事例ではYをキャンペーン期間中の利用率とするため、ロジスティック回帰を利用したが、推定したい結果変数にあわせて種々の手法を利用する事ができる。

二重にロバストな推定法では、対照群（キャンペーン不参加）に対して傾向スコアを使って$\frac{e(X)}{1-e(X)}$という重み付け（IPW：逆確率重み付け）を行っている。傾向スコア$e(X)$は処置群の特性に近いほど高いスコアになるため、この重み付けを使うことで対照群でありながら処置群に近い特性を持つ消費者に大きな重みをつけている。この重み付けによって結果変数の推定モデル$g_0(X)$の誤指定があった場合にも、傾向スコアの推定モデルが正しければ、モデルの誤指定によるバイアスを相殺することで偏りのない推定量を得ることができる。一方で傾向スコアの推定モデルに誤指定があった場合でも、結果変数の推定モデルが正しく推定されていれば、第 2 項の$Y_i^0 - g_0(X)$が 0 に近い値となり、第 1 項の推定も精度高く行われるため、こちらも偏りのない推定量を得ることができる。

このように 2 つの推定モデルのうちどちらかが正しく推定できていれば偏りのない推定量を得ることができるのが二重にロバストな推定法の特徴である。次節で簡単なシミュレーションデータを使ってその特性を確認しよう。

5.4　二重にロバストな推定法による参加型キャンペーン施策の効果検証

参加型キャンペーン施策の効果検証を想定したデータセットを確認しよう。データセットはキャンペーン参加者（処置群）と未参加者（対照群）、それぞれ 500 サンプルで 1,000 サンプルが入ったデータになっている。データに含まれる項目は以下の通り 6 項目から構成される。

Z：処置変数（2 値）、キャンペーン参加の有無（0：未参加, 1：参加）

2　線形回帰やパラメトリックな手法に限定されるわけではなく、条件を考慮したうえで他の機械学習系の手法なども利用できる（岩崎 2015; 星野 2009）

Y:結果変数(2 値)、キャンペーン期間中の利用有無(0:未利用, 1:利用)
X_1_age:年齢(離散連続量)、消費者の年齢、20-70 歳
X_2_region:居住地域(2 値)、消費者の居住地域(0:都市部, 1:地方)
X_3_pre_usage:過去のサービス利用回数(離散連続量)
X_4_pre_campaign:過去のキャンペーン参加率(連続量)

処置変数 Z 及び結果変数 Y、4 つの共変量(X_1～X_4)を使って、キャンペーンの施策効果を推定する。データは処置に影響を与える変数(傾向スコアモデルを構成)が、X_3_pre_usage と X_4_pre_campaign の 2 変数になるように作られている。結果変数 Y に影響を与えるのは、X_1_age から X_4_pre_campaign まで 4 つの変数全てとなる。処置変数 Z 及び結果変数 Y に対して影響を与える変数が正しく指定されている場合と誤って指定されている場合での推定精度の違いを後ほど確認する。

処置群及び対照群における各データ項目の集計が以下の表である。

Z	n 数	Y	平均年齢	地方居住比率	過去平均利用回数	過去キャンペーン参加率
処置群	500	0.66	44.87	0.43	5.35	0.23
対照群	500	0.44	44.99	0.44	2.93	0.17

それぞれ 500 サンプルのデータがあり(n 数)、結果変数 Y の値(キャンペーン期間中の利用率)は処置群が 66%、対照群が 44%である。平均年齢と地方居住比率は両群であまり差がないが、過去平均利用回数(処置群平均 5.34 回、対照群平均 2.93 回)と過去キャンペーン参加率(処置群平均 23%、対照群平均 17%)と、過去の利用状況に大きな違いがある。当該サービスを過去により積極的に利用していた消費者がキャンペーン参加者に多い状況となっている。

施策効果は結果変数 Y の処置群と対照群の差となるが、共変量 X で捉えられる両群の消費者の特徴が異なる(キャンペーン参加者は過去の利用状況が非参加者に比べてより多い、いわゆるヘビーユーザー)ため、結果変数 Y の単純平均の差を施策効果とすることはできない。単純平均の差は 66%－44%＝22%となるが、両群の特性の違いを調整した真の施策効果は 17.5%となるように本データセットは設定してある。これはキャンペーン参加者の偏りから単純平均の施策効果が真の施策効果よりも高い推定量を示している状況を再現している。この状況を踏まえて、二重にロバストな推定法を使ったバイアスの補正による施策効果の推定量を確認しよう。

二重にロバストな推定法を R 言語で関数化したものが以下である。

```r
DRATT_estimator <- function(data, Z_nm, Y_nm, ps_formula, outcome_formula)
{
  Z <- data[[Z_nm]] # 処置変数
  Y <- data[[Y_nm]] # 結果変数
  N1 <- sum(Z)  # 処置群サンプルサイズ

  # 傾向スコアモデル（ロジスティック回帰）
  ps_model <- glm(ps_formula, family = binomial(link = "logit"), data =
data)
  e_X <- fitted(ps_model)

  # 結果変数モデル（対照群のみ、ロジスティック回帰）
  data_ctrl <- data[Z == 0, ]
  G_0 <- glm(outcome_formula, family = binomial(link = "logit"), data =
data_ctrl)
  Y_0 <- predict(G_0, newdata = data, type = "response")

  # DR-ATT 推定量
  tau_att <- sum(Z * (Y-Y_0) - (1-Z) * e_X / (1-e_X) * (Y-Y_0)) / N1

  # 推定結果
  return(
    list(
      tau_att = tau_att,
      ps_model = ps_model,
      g0_model = G_0,
      e_X = e_X,
      Y_0 = Y_0,
      N1 = N1
    )
  )
}
```

注）関数内の$\hat{\tau}_{(DR,ATT)}$の推定式は計算効率を考慮して以下の形式に変換してある。

$$\hat{\tau}_{(DR,ATT)} = \frac{1}{N_1}\sum_{i=1}^{N}\left[Z_i\left(Y_i^{\,1} - g_0(X_i)\right) - (1-Z_i)\frac{e(X_i)}{1-e(X_i)}\left(Y_i^{\,0} - g_0(X_i)\right)\right]$$

`DRATT_estimator`関数を使った推定は以下のように行う。

```
mkt_est1 <- DRATT_estimator(
  data = mkt_data,
  Z_nm = "Z",
  Y_nm = "Y",
  ps_formula = Z ~ X_3_pre_usage_norm + X_4_pre_campaign,
  outcome_formula = Y ~ X_1_age_norm + X_2_region + X_3_pre_usage_norm
+ X_4_pre_campaign
)
```

注）年齢（X_1_age）と過去の利用回数（X_3_pre_usage）はそれぞれ正規化（-1~1, 0~1）した上でモデルに投入、推定を行っている。

引数 data に指定している mkt_data が先ほど紹介したデータセットである。引数 Z_nm と T_nm には引数 data で与えられた mkt_data 内の処置変数名、結果変数名を指定する。引数 ps_formula 及び outcome_formula には、それぞれ傾向スコアモデルと結果変数モデルのモデル式を入力する。モデル式とは~で挟まれた左側に目的変数となる変数名を記述し、右側に説明変数となる変数名を記述する R 言語の記法である。複数の説明変数がある場合は変数間を+記号でつないで表記する。例えば、傾向スコアモデルのモデル式は引数 ps_formula に Z ~ X_3_pre_usage_norm + X_4_pre_campaign と記入しているが、これは処置変数 Z を目的変数にして、説明変数 X_3_pre_usage_norm と X_4_pre_campaign の 2 変数を説明変数としたロジスティック回帰を行う事を意味する。これらの引数の設定を行う事で DRATT_estimator 関数は二重にロバストな ATT 推定量を計算する。

今回のデータセットは、ATT による真の施策効果をデータの背後で持つ設定にしてあるので、DRATT_estimator 関数によって計算した推定量と真の施策効果を比較してみよう。

```
sprintf("真の施策効果 : %.5f", mkt_data$true_att[[1]])
## [1] "真の施策効果 : 0.17579"
sprintf("DRATT_estimatorによる推定 : %.5f", mkt_est1$tau_att)
## [1] "DRATT_estimatorによる推定 : 0.18129"
```

真の施策効果 17.5%に対して計算した推定量は 18.1%であった。この結果は処置群と対照群の単純平均の差分である偏った推定量の 22%に対して真の施策効果により近似した推定量を算出できている事がわかる。ただし、この推定で使用した傾向スコアモデル、結果変数モデルの説明変数は、今回のデータセットに合わせた「正しい」説明変数のセットである。実務の応用場面では、どのような変数が処置、結果の両変数に影響があるのかは自明ではな

い。また影響するであろう説明変数（要因）を常に的確に推定時の評価に組み込めるとも限らない。状況によっては「正しい」説明変数のセットができない場合も起こりうる。このような説明変数の「誤指定」に対しても、傾向スコアモデル、結果変数モデルのどちらかが正しければ偏りのない推定を行う特性を持つのが二重にロバストな推定法であった。以下で「誤指定」を 4 つのパターンに分類して、それぞれのケースでの推定量の確認を行う。

設定パターン	DRATT_estimator 関数の引数	モデル式の設定
①両方正しい	ps_formula	Z ~ X_3_pre_usage_norm + X_4_pre_campaign
	outcome_formula	Y ~ X_1_age_norm + X_2_region + X_3_pre_usage_norm + X_4_pre_campaign
②傾向スコアモデル誤指定	ps_formula	Z ~ X_1_age_norm + X_2_region
	outcome_formula	Y ~ X_1_age_norm + X_2_region + X_3_pre_usage_norm + X_4_pre_campaign
③結果変数モデル誤指定	ps_formula	Z ~ X_3_pre_usage_norm + X_4_pre_campaign
	outcome_formula	Y ~ X_3_pre_usage_norm + X_4_pre_campaign
④両方誤指定	ps_formula	Z ~ X_1_age_norm + X_2_region
	outcome_formula	Y ~ X_1_age_norm + X_2_region

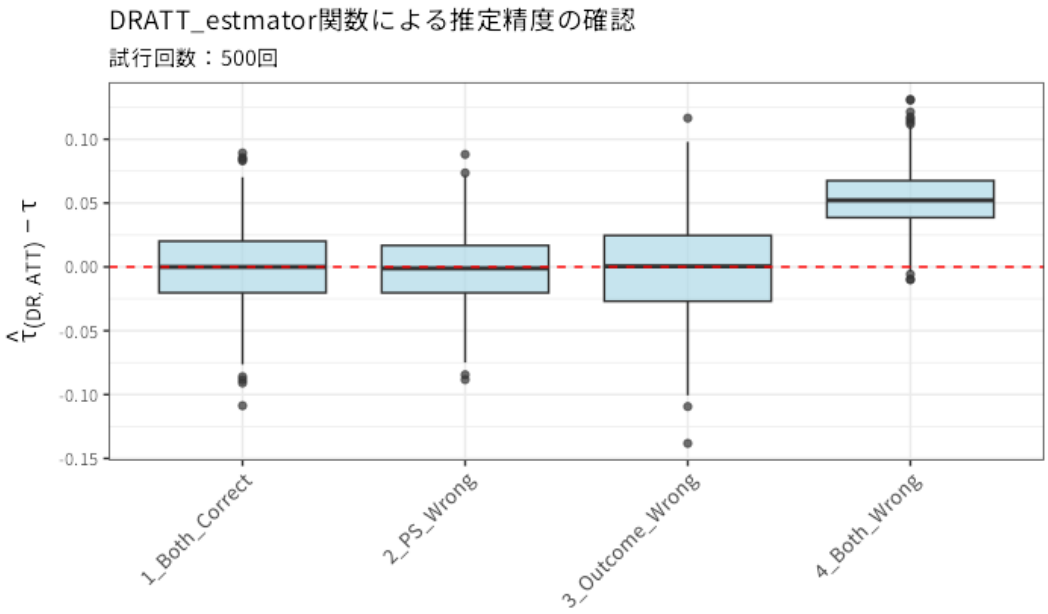
4 つのパターンは、すでに推定結果を出した①両方正しい（傾向スコアモデル、結果変数モデルともに説明変数の指定が正しい）場合に加えて、②傾向スコアモデル誤指定、③結果変数モデル誤指定、④両方誤指定の 3 パターンで構成されている。誤指定のあるパターンは、重要な説明変数が推定時に抜け落ちているパターンとなる（説明変数のパターンは上記表内のモデル式の設定参照)。それぞれのパターン別に推定した$\hat{\tau}_{(DR,ATT)}$と真の処置効果を比較してみよう。

設定パターン	真の処置効果	$\hat{\tau}_{(DR,ATT)}$	差分
①両方正しい	0.1758	0.1813	0.0055
②傾向スコアモデル誤指定	0.1758	0.1667	-0.0091
③結果変数モデル誤指定	0.1758	0.1630	-0.0128
④両方誤指定	0.1758	0.2319	0.0561

設定パターンの④両方誤指定の場合は真の処置効果から、大きく$\hat{\tau}_{(DR,ATT)}$推

定量が外れているが、それ以外のパターンでは概ね真の処置効果に近い推定をしていることがわかる。②と③のようにモデル誤指定がある場合でも真の処置効果に近い推定を行っており、二重にロバストな特性が見て取れる。

これは 1 組のデータセットについての推定結果であるが、今回分析に使用したデータセットは生成過程を R 言語で関数化してあるので（本稿末尾にサンプルデータ生成用コードを記載）、この関数を使ってランダムに生成した 500 セットのデータに対して同じ設定パターンで推定を行った結果が以下の図と表である。



ここでも同様にモデル誤指定に対しての頑健性を確認する事ができるだろう。

5.5　実務での応用にあたって

本稿では簡易なサンプルデータを利用して、二重にロバストな推定量の特性について R 言語を使った基礎的なモデリングにより確認してきた。紙幅の制約もありこれ以上記述できないが、実務での応用にあたってはいくつかの留意点が残る。

実務における効果検証の際には、処置変数と結果変数に影響するであろう様々な要因を考慮に入れて多くの説明変数を探索する事が多い。検証に当たっては影響要因の漏れがリスクとなるため、必然的に推定モデルに投入する説明変数の数は多くなる。このような場合、今回の事例で作成したようなシンプルなロジスティック回帰等の手法では推定がうまくいかない場合も多い。二重にロバストな推定法を応用・発展させた、高次元データに対応できる機械学習系

の手法（DML：Double Machine Learning）も提案されている（Chernozhukov et al. 2018; Ellickson et al. 2023 など）ため、こうした手法を取り入れることを考慮に入れた方が良い。機械学習を取り入れた手法に関しては、本ワーキングペーパーの第 8 章 「因果推論と機械学習」を参照してほしい。

また今回は紹介しきれなかったが、参加型キャンペーンの処置効果を参加/不参加で比較するだけでなく、キャンペーン開始前と後の前後の時間変化も加えて効果を推定する手法も広く利用されている。これは結果変数の処置による差分と処置前後の時間的な差分の 2 つの差分を使って推定を行う事から差分の差分法（DID：Difference-in-Differences）と呼ばれる。こうした手法と今回紹介した傾向スコアを組み合わせたり（Abadie 2005）、二重にロバストな推定法を組み合わせた手法（Sant'Anna & Zhao 2020）も提案されている。今回の事例ではキャンペーン前後の結果変数の時間変化を考慮に入れた推定にはなっていないが、実務での応用を考えるにあたってはこうした DID との組み合わせによる手法を取り入れる事でより偏りのない推定を行う事ができるだろう。

5.6 サンプルデータの生成コード

※検証時に使用したデータは seed=1234 で生成
※500 セットのデータのシミュレーション時は seed を 1～500 に設定して生成

```r
generate_mkt_data <- function(n_total = 3000,
                              n_treatment = 500,
                              n_control = 500,
                              seed = NA)  {

  # シード設定
  if(!is.na(seed)) set.seed(seed)

  # 共変量の生成
  # 年齢: Age (20-70 歳)
  age <- sample(20:70, n_total, replace = TRUE)
  age_normalized <- (age - 45) / 25  # 正規化

  # 居住地域: Region (0: 都市部, 1: 地方)
  region <- rbinom(n_total, 1, 0.4)

  # 過去の利用回数: Pre usage
  pre_usage <- rnbinom(n_total, size = 2, mu = 4)
  pre_usage_normalized <- pmin(pre_usage / 10, 1)  # 正規化
```

```r
# 過去のキャンペーン参加率: Pre campaign
pre_campaign <- rbeta(n_total, shape1 = 1.5, shape2 = 6)

# 処置変数: Z(0: 不参加, 1: 参加)
error_ps <- rnorm(n_total, mean = 0, sd = 0.3)
util_z <- -2.0 +
  2.5 * pre_usage_normalized +
  3.0 * pre_campaign +
  2.0 * pre_usage_normalized * pre_campaign +
  error_ps
true_ps <- plogis(util_z)
Z_total <- rbinom(n_total, 1, true_ps)

# 処置群/対照群の指定 n 数にサンプリング
trtm_id <- which(Z_total == 1)
ctrl_id <- which(Z_total == 0)

trtm_id_sample <- sample(trtm_id, n_treatment, replace = FALSE)
ctrl_id_sample <- sample(ctrl_id, n_control, replace = FALSE)
total_id <- c(trtm_id_sample, ctrl_id_sample)

# サンプリングした id の共変量
selected_age <- age[total_id]
selected_age_norm <- age_normalized[total_id]
selected_region <- region[total_id]
selected_pre_usage <- pre_usage[total_id]
selected_pre_usage_norm <- pre_usage_normalized[total_id]
selected_pre_campaign <- pre_campaign[total_id]
selected_ps <- true_ps[total_id]
selected_error_ps <- error_ps[total_id]
Z <- c(rep(1, n_treatment), rep(0, n_control))

# 結果変数: Y(キャンペーン期間中の利用有無)
n_selected <- length(total_id)
error_outcome <- rnorm(n_selected, mean = 0, sd = 0.4)

# Y(0): 処置を受けなかった場合の潜在アウトカム
util_y0 <- -2.0 +
  2.0 * selected_age_norm +
```

```
  2.5 * selected_region +
  1.2 * selected_pre_usage_norm +
  1.5 * selected_pre_campaign +
  error_outcome

prob_y0 <- plogis(util_y0)

# Z=1 の際の処置効果
treatment_effect_logit <- 1.0 +
  0.3 * selected_age_norm +
  0.4 * selected_region +
  0.2 * selected_pre_usage_norm +
  0.1 * error_outcome

# Y(1): 処置を受けた場合の潜在アウトカム
util_y1 <- util_y0 + treatment_effect_logit
prob_y1 <- plogis(util_y1)

# 観測された結果変数: Y = Z*Y(1) + (1-Z)*Y(0)
Y <- ifelse(Z == 1,
            rbinom(sum(Z == 1), 1, prob_y1[Z == 1]),
            rbinom(sum(Z == 0), 1, prob_y0[Z == 0]))

# 真の処置効果(ATT)
treated_id <- which(Z == 1)
true_att <- mean(prob_y1[treated_id]) - mean(prob_y0[treated_id])

# サンプルデータ用 data.frame の作成
sample_df <- data.frame(
  # 処置変数:Z と結果変数:Y
  Z = Z,
  Y = Y,
  # 共変量:X
  X_1_age = selected_age,
  X_1_age_norm = selected_age_norm,
  X_2_region = selected_region,
  X_3_pre_usage = selected_pre_usage,
  X_3_pre_usage_norm = selected_pre_usage_norm,
  X_4_pre_campaign = selected_pre_campaign,
  # 検証用データ
```

```
    true_att = true_att,
    true_ps = selected_ps,
    prob_y0 = prob_y0,
    prob_y1 = prob_y1
  )

  return(sample_df)
}
```

6章　マーケティング研究における傾向スコア・マッチングの活用とその課題

6.1　はじめに

マーケティング分析において，広告施策や販売戦略が消費者の意思決定に与える影響を因果的に推定することは極めて重要である。しかし，倫理的・実務的制約からランダム化比較実験（randomized controlled trial；以下，RCT と表記）の実施が困難な場合が多く，観察データに依拠した因果推論が求められる。その際，交絡因子によるバイアスを除去する方法としてマッチング（matching）手法が注目されている。

本章では，とくに傾向スコア（propensity score）を用いたマッチング，すなわち傾向スコア・マッチング（propensity score matching：以下，PSM と表記）に注目し，マーケティングや消費者行動研究における研究事例を取り上げ，今後の応用可能性と限界を検討する。こうした検討は，マーケティング・リサーチ実務への示唆を提供することにつながるだろう。

6.2　マッチング手法とは何か

マッチングとは，処置群（例：広告を見た群）と対照群（例：広告を見ていない群）で共変量を揃えることにより，交絡の影響を取り除く手法である。いわゆる強い無視可能性（strong ignorability）の仮定のもと，観測された共変量に基づいて処置が付与されると考える。

強い無視可能性とは，PSM が成り立つために必要な仮定のことを指す。強い無視可能性では，誰が処置を受けるか（例：広告を見るか）が，観測可能な共変量だけで決まっているものと仮定する。もし観測されていない購買意欲などの要因が処置と結果の両方に影響していると，この仮定は成立せず，バイアスが残ってしまう点には注意が必要である。

傾向スコアを用いたマッチングは，複数の共変量情報を 1 次元のスコアに集約し，実務上のマッチング操作を現実的に可能にした。処置群に属する各個体に対し，傾向スコアの近い対照群の個体を対応づけることで，処置効果の推定が行われる。

6.3　マーケティング研究におけるマッチング手法の先行研究レビュー

以下では，近年のマーケティングや消費者行動関連の実証研究において，PSM がどのように用いられているかを，最近の研究から 5 本を取り上げレビューする。それぞれの研究は，PSM の実用性，限界，施策デザインとの関連性に関して異なる視座を提供しており，マーケティングや消費者行動の領域における因果推論手法の

展開を理解するうえで重要な示唆を含んでいる。

第１項　Guenther et al. (2025)：マーケティング研究における PSM 実践ガイド

Guenther et al.（2025）は，マーケティング領域における PSM の適切な実装（correct implementation）に焦点を当てた方法論的研究を報告している。同研究は，PSM の適切な使用が因果推定の信頼性を確保する上でいかに重要であるかを，理論的枠組みと実証的検討の両面から論じている。

Guenther et al.（2025）が提示する実装手順は，①処置とアウトカムの両方に影響を与える共変量の適切な選定，②Love plot や標準化平均差（standardized mean difference：以下，SMD と表記）を用いたマッチング後のバランス検証，③共通サポートの確認と極端値の除外，④感度分析(sensitivity analysis：PSM による因果推定結果が未観測交絡の影響をどの程度受ける可能性があるかを評価する手法)による未観測の共変量への補完的検討，の 4 点から構成される。

Guenther et al.（2025）ではさらに，1 対 1 の近傍法やカリパー制約付きマッチング，層別法といった複数のマッチング手法を比較し，それぞれの特徴と適用上の留意点を丁寧に解説している。特に，マッチング後のモデル適合性を視覚的に確認するための Love plot の有効性を強調しており，可視化による診断が実務的な透明性を担保する点にも注目すべきである。

このような知見は，マーケティング・リサーチ実務において，クーポン施策の効果検証やユーザー体験（UX）改善の影響分析，チャネル施策の評価など，非実験的データを前提とする施策評価の場面で大いに活用可能である。実務の現場では，ランダム化が困難な状況が多く，観測データに基づくバイアス制御と説得力のある説明が求められるが，Guenther et al.（2025）が提示する実装ガイドは，そうした課題に対して極めて実践的な指針を提供するものである。

すなわち，Guenther et al.（2025）は，PSM を用いた因果推論の標準化された実装プロセスを提供することで，マーケティング実務におけるデータ分析の信頼性と説明責任の水準を引き上げる重要な貢献を果たしている。

第２項　Li et al. (2024)：インターネット使用と健康意識の関係性に関する PSM 適用研究

Li et al.（2024)は，中国におけるインターネットの普及が個人の健康意識に及ぼす因果的影響を検証するため，PSM を用いた実証研究を行っている。この研究は，健康関連情報へのアクセスが容易になる一方で，消費者の情報リテラシーやインフラ環境の格差が健康意識の分断を生む可能性に着目し，インターネット使用の有無を介入変数，健康意識をアウトカムとした因果推論を実施している。使用データは「中国健康と栄養調査（CHNS）」に基づき，ロジスティック回帰により傾向スコアを推定し，年齢・性別・教育水準・所得・居住地などの共変量を統制したうえで，1 対 1 マッチングを適用している。マッチング後には，SMD によるバランスチェックが実施され，共変量の分布均質性が確認された。

分析の結果，インターネット利用者は非利用者に比して健康意識スコアが有意に高く，特に教育水準の高い層においてこの傾向が顕著であった。これは，情報アク

セスの格差が，意識や行動面での不均衡を生むことを示唆しており，インターネットの利用が医療リテラシーの向上に寄与する可能性を示している。

Li et al. (2024)の意義は，PSM の実装手順を忠実に踏襲した点にあり，シンプルな 2 値介入変数の適用，ロジスティック回帰によるスコア推定，1 対 1 マッチング，SMD によるバランス評価という一連のステップが丁寧に実行されている。これは，PSM を初めて扱う研究者や実務家にとって，因果推論の導入手順を学ぶ上で非常に実践的なモデルケースといえるだろう。

このような分析枠組みは，マーケティング・リサーチの実務においても応用可能である。たとえば，広告接触の有無，クーポン受領経験，アプリ使用の有無などの 2 値的な施策を介入変数とし，購買意図やブランド認知といったアウトカムへの影響を評価する際，PSM は施策効果のより厳密な推定を可能にする。特に，ランダム化が困難なフィールド調査や観察データを用いた分析場面において，PSM を用いた因果推論のフレームは，分析の信頼性と説明責任を高める点で有効である。さらに，今後は多値 PSM や傾向スコア加重法への展開により，より複雑な介入変数や連続的アウトカムにも対応できる分析設計が求められるだろう。Li et al. (2024)は，その基礎的出発点として，マーケティング実務における因果推論の導入を後押しする貴重な実証例である。

第 3 項　Gordon et al. (2022)： Facebook 広告キャンペーンにおける PSM の検証的応用

Gordon et al.(2022)は，Facebook 広告キャンペーンの効果を，非実験的な観察データによってどの程度正確に推定できるかを大規模に検証した実証研究を行った。具体的には，RCT によって得られた真の効果を基準とし，PSM をはじめとする複数の因果推定手法によって得られた推定効果との乖離を分析することで，非実験的手法の精度と限界を評価している。

PSM の実装にあたっては，1 対 1 の近傍マッチングが用いられ，さらに傾向スコアの差が一定の閾値内に収まるようカリパー制約が導入された。この手法により，属性が大きく異なる個体同士の不適切なマッチングを防ぎ，処置群と統制群の構成を統制している。マッチング後の妥当性確認には，SMD や密度プロットが用いられ，共変量の分布の均質化が視覚的かつ数値的に評価された。

分析の結果，観測可能な共変量については PSM によってバランスが改善されたものの，広告に対する態度やオフラインでの購買傾向といった非観測交絡の影響が推定に残存しており，RCT との推定値には有意な乖離が確認された。これにより，PSM が観測データに基づく因果推論を可能にする一方で，非観測変数によるバイアスを完全に排除することは難しいという構造的限界が浮き彫りとなった。

Gordon et al.(2022)の重要な貢献は，実験的因果推論と非実験的手法との間に生じる差異を定量的に可視化した点にある。これは，マーケティング・リサーチの実務においても極めて有用な知見である。たとえば，A/B テストの実施が困難な環境において，PSM を用いて広告施策やキャンペーンの効果を準実験的に推定する際には，Gordon et al.(2022)が示すように，マッチング精度の評価，バランス診断，および非観測交絡の存在を前提とした慎重な解釈が不可欠である。したがって，マー

ケティング・リサーチにおける因果推論の導入に際しては，PSM の利点と限界の双方を理解したうえで，感度分析や補完的設計の導入を含む複線的なアプローチが望まれる。本研究は，そのような実務応用における設計指針として，高い示唆を提供している。

第４項　Langen and Huber (2023)：クーポン施策における異質性効果の分析と PSM の併用的活用

Langen and Huber（2023）は，マーケティング施策の一つであるクーポン配布が，消費者に与える効果が一様であるとは限らないという問題意識のもと，施策効果の異質性（heterogeneity）に着目した分析を行っている。分析の第一段階においてPSM を用い，クーポンを受け取った消費者における平均的処置効果（average treatment effect on the treated：以下，ATT と表記）を推定している。これは，処置群（クーポン受領者）と統制群（非受領者）の間で，年齢，購買頻度，カテゴリ嗜好などの共変量が一致するようマッチングを行い，その上で購買結果の差異を計測することにより，クーポン施策の平均的な影響を評価するものである。

第一段階で ATT を特定したのち，分析の第二段階において，消費者の属性や行動特性に応じて効果がどのように変化するかという点，すなわち，条件付き平均処置効果（conditional average treatment effect：以下，CATE と表記）の推定に移行している。CATE は，たとえば「若年層の女性」や「高頻度購買者」など，特定の条件や属性を持つサブグループにおける処置効果の平均を示すものである。この分析により，「誰にとってクーポンが有効か」「どのような属性をもつ顧客群では効果が乏しいのか」といった，より精緻なターゲティング戦略の設計が可能となる。

この CATE の推定には，因果推論に特化した機械学習手法である Causal Forest が用いられている。これは，機械学習で広く用いられる Random Forest を基礎としたアルゴリズムである。Random Forest は，複数の決定木を構築し，それらの出力の多数決により予測を行うアンサンブル学習法であり，過学習を抑えつつ高い精度を実現する点で知られている。Causal Forest はこの仕組みを応用し，処置の有無とアウトカム（購買行動）との因果関係を，個別の共変量構成に応じて学習し，サブグループや個人単位の CATE を推定できるよう設計されている。

Langen and Huber（2023）では，上記のとおり，まず PSM によって全体的な ATT を把握したうえで，Causal Forest を用いた CATE 推定により，消費者セグメントごとのクーポン施策の効果を可視化している。この二段階の分析設計により，平均的評価と個別的評価の両面から施策の効果を捉えることに成功している。特に，クーポンが「誰に効いたか」を明らかにする CATE 分析は，マーケティングの戦略的意思決定にとってきわめて有用であり，今後の施策最適化に資する知見を提供している。

第５項　Zheng and Liu (2024)：PSM とブートストラップ法を組み合わせたロバスト性検証の試み

Zheng and Liu（2024）は，非実験的データに基づく因果推論の信頼性を高める方法として，PSM とブートストラップ法（bootstrap）を併用する分析手法を提案している。

Zheng and Liu（2024）の主眼は，マーケティング領域における A/B テストに類似した環境――すなわち，施策介入の有無を恣意的に操作できない現実的な状況下――において，どの程度まで信頼性の高い因果推定が可能であるかを明らかにする点にある。

Zheng and Liu（2024）で取り上げられたブートストラップ法とは，観測データから無作為にサンプルを復元抽出し，数百～数千回の再分析を行うことにより，推定値のばらつき（不確実性）を評価する統計的手法である。Zheng and Liu は，この再サンプリングを PSM に組み込み，複数回のマッチングを通じて因果効果推定の分布を導出した上で，その安定性からロバスト性（頑健性）を評価している。すなわち，どのマッチングの回においても処置効果の推定値が一貫して安定していれば，外的撹乱に強く，再現可能性の高い推定が行われたと判断できる。

さらに Zheng and Liu（2024）では，近年注目を集める二重機械学習（Double Machine Learning：以下，DML と表記）との比較検証も行われている。DML は，機械学習により交絡因子の推定と因果効果の推定を分離することで，バイアスを低減する最新の因果推論手法である。Zheng and Liu は，PSM とブートストラップの組み合わせが，DML に匹敵する推定精度とロバスト性を備えていることを実証しており，かつ，その解釈可能性の高さからマーケティング実務における説明責任や透明性の確保において優位性を持つことを指摘している。

このように，Zheng and Liu（2024）は PSM の推定における不確実性を可視化し，従来の単一推定に比してバイアスの検出と頑健性の評価を可能にする実践的アプローチを提示している点で高く評価できる。特に，A/B テストが困難な実地調査や自然観察データに基づく施策評価において，この手法は，因果推定の信頼性を担保しつつ，推定結果の頑健性を担保する有力な選択肢となる。このことは，とりわけ，マーケティング施策の効果を部門間や経営層に説明する必要がある現場において，「なぜこの結果が導かれたのか」を追跡可能な構造を持つ PSM の価値が再評価される。

マーケティング・リサーチの実務においても，単一分析結果のみに依存せず，繰り返し検証とばらつき評価を取り入れた分析設計を採用することで，意思決定における説得力と信頼性を大きく向上させることができるだろう。Zheng and Liu（2024）の研究は，このような設計上の視点を実務に還元するうえで有益な知見を提供している。

6.4　考察：共通点と課題の整理，および今後の展望

本章で取り上げた 5 本の実証研究はいずれも比較的最近発表されたものであり，マーケティングおよび消費者行動領域における現時点での PSM の応用可能性と限界を多角的に示している。よって，PSM という因果推論手法がこの分野においていかなる意味を持つのかを立体的に理解する上でこれらの論文の報告は非常に有用である。

まず注目すべきは，PSM の導入から応用，さらには高度化までの段階的展開が，

今回のレビューを通じて明確になった点である。たとえば，Guenther et al.（2025）は，PSM の正確な実装に関する包括的なガイドラインを提示しており，共変量の選定，マッチング後のバランスチェック，共通サポートの確認，さらには感度分析といった一連のステップが，いかに因果推定の信頼性を高めるかを実証的に論じている。同様に，Li et al.（2024）の研究は，PSM の基本的構成要素を忠実に踏襲しつつ，シンプルな 2 値介入変数とロジスティック回帰に基づくマッチング手法によって，非実験的状況下における因果推論の有効性を提示している。これらは，PSM が比較的容易に実装でき，実務や初学者にとっても導入しやすい方法であることを示しており，PSM の裾野の広さを物語っている。

一方で，PSM が現実の複雑なマーケティング環境においても有効に機能するかどうかについては，より批判的な視点が必要である。Gordon et al.（2022）の研究は，RCT との比較によって，PSM が観測可能な共変量に関しては有効にバイアスを補正する一方で，非観測交絡の影響を排除することが構造的に困難であるという限界を明示している。観察データに依拠した因果推論のリスクを可視化した本研究は，PSM に対する過信を戒める重要な警鐘である。

こうした限界を補う試みとして，Langen and Huber（2023）は，まず PSM によって ATT を推定したうえで，機械学習に基づく Causal Forest を用いて，消費者セグメントごとの CATE を分析している。この二段階のアプローチは，マーケティング施策が「誰にとって効くのか」という問いに対する実務的解を提供し，PSM の基本的枠組みに依拠しながらも，個別最適化という戦略的意思決定に資する分析へと展開している。また，Zheng and Liu（2024）は，PSM とブートストラップ法を併用することで，マッチング後の推定値がサンプル特性に依存していないかを繰り返し検証し，推定のばらつきとロバスト性を数値的に評価する枠組みを提示している。さらに，DML との比較を通じて，PSM が依然として高い解釈可能性を保持している点を強調しており，実務現場における説明可能性や透明性の観点からも重要な意義を持つ。

もっとも，こうした発展的応用にもかかわらず，PSM にはいくつかの構造的な制約が存在することも否定できない。先に述べたように，非観測交絡の影響は PSM の手法そのものでは排除できず，外的妥当性や因果効果の一般化には限界がある。さらに，Liらの研究が示すように，PSMは2値処置変数への適用に最も適しているが，多値処置や連続的介入変数への適用には適しておらず，これらに対応するには傾向スコア加重法や一般化傾向スコアなどの他手法との併用が求められる。Zheng and Liu（2024）が指摘するように，カリパー制約や共通サポートの確認によってマッチング精度を高める一方で，サンプル損失や推定のばらつきが生じるリスクも無視できない。加えて，Langen and Huber（2023）が用いた Causal Forest のような機械学習ベースの手法は，分析の精緻化を可能にする一方で，アルゴリズムの透明性が損なわれ，実務上の説明責任とトレードオフの関係に立たされることになる。

以上を踏まえると，PSM は，実験的手法の導入が難しいマーケティングや社会的文脈において，準実験的な因果推論の出発点として有効なアプローチであると結論付けられる。しかし同時に，それはあくまで観測可能な共変量に基づく条件的独立性という仮定の上に成り立っており，その限界を補うためには，他の統計的手法

や機械学習技術との統合的活用が不可欠である。今後の研究においては，PSM の枠組みにとどまらず，感度分析，ブートストラップ，加重法，多値・連続介入への対応，さらにはモデル可視化や再現性の担保といった複数の視点から，より柔軟で信頼性の高い因果推論設計が求められるだろう。

本章でのレビューが示すように，PSM は単なる分析手法としての意義を超え，学術と実務をつなぐ中間領域における共通言語として機能する可能性を持っている。したがって，その限界を認識しつつ，適切な設計と補完的手法を用いた上で，より精緻な因果分析へと進化させていくことが，マーケティング・消費者行動研究の次なる課題である。

７章　マハラノビスマッチングの意味と実行コード

7.1　マハラノビスマッチングの利点

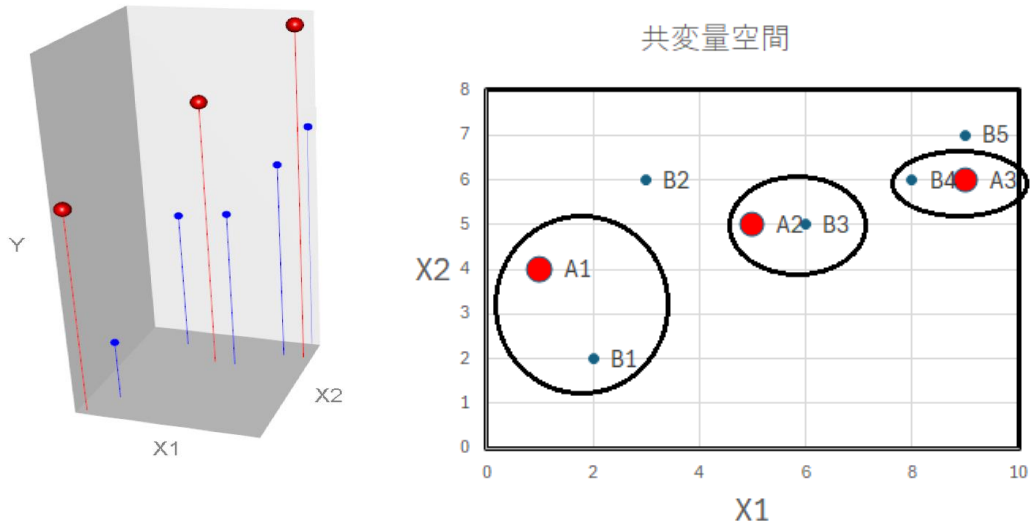


図１　マッチングによる処置効果の推定（左）とマハラノビスマッチング（右）

処置群と対照群それぞれから共変量の値が似たサンプルを選んでペアを作り、ペアごとに反応 Y の差をとれば処置の効果が分かるだろうというのがマッチングの基本的なアイデアである。図 1 左でいえば球の高さと点の高さの差が処置効果を表す。2つのサンプルが似ているかどうかは図 1 右の共変量空間における座標の位置から判定する。完全に 2 点が一致している場合を「厳密マッチング」という。

Rubin (1980)が提唱したマハラノビスマッチングは 2 点間の汎距離にもとづく推定法なのでロジックが分かりやすい。その一方で、傾向スコアは共変量の類似度を表す尺度ではないという問題がある。

なぜならロジスティック関数は共変量の線形和をベースにしているので、共変量が大きく離れたサンプルでも傾向スコアが等しくなる危険性がある。線形和が

$0.2X_1 + 0.3X_2$ だとした数値例を表 1 に示そう。

サンプル1と2は共変量の値がかけ離れているにもかかわらず、傾向スコアは同一になる。

表1　傾向スコアマッチングの問題点

サンプル	X1	X2	線形和	傾向スコア
サンプル1	0	0	0	0.5
サンプル2	3	-2	0	0.5

傾向スコア・マッチングに対する批判と反論については高橋(2022,pp169-171)が丁寧にレビューしている。

7.2　調査観察データにおける処置群と対照群のサイズ

実験研究では処置群と対照群のサイズを実験者が決めることができるが、調査観察データでは、サンプル数をコントロールできない。そして通常はマーケティング施策を受けた人は少なく、受けていない人は多い。

本節では処置群のサンプル数 n が対照群のサンプル数 m より少ない、という状況でマハラノビスマッチングの実行例を示そう。処置群を A、対照群を B と呼ぶ。共変量はいくつでも構わないが、ここでは X1,X2 とする。

■　数値例

$$\boldsymbol{A}=\begin{bmatrix}1 & 1\\ 5 & 5\\ 9 & 6\end{bmatrix},\qquad \boldsymbol{B}=\begin{bmatrix}2 & 2\\ 3 & 6\\ 6 & 5\\ 8 & 6\\ 9 & 7\end{bmatrix}$$

処置群のサンプルを A1～A3、対照群のサンプルを B1～B5 と呼ぶ。因果のアウトカムである「反応 Y」はマッチング処理には用いない。

A 群のサンプルごとに汎距離が一番小さい B 群のサンプルを探そう。これを最近距離法という。

汎距離計算に用いる分散共分散行列としては次の 3 案が候補になる。

データ A にもとづく分散共分散行列 SA

データ B にもとづく分散共分散行列 SB

SA と SB をプールした分散共分散行列 Spool、数値例の場合は

$$Spool=\frac{1}{3+5-2}\left(2\times SA+4\times SB\right)$$

SA と SB が大きく異なる場合には汎距離の値が違ってきて、その結果マハラノビスマッチングのペアが変化することがあり得る。

ATT(average treatment effect on the treated)では A 群のサンプルにフォーカスして平均処置効果を求めるのでペア数は n=3 になる。「ATT の推定には処置群の

SA を使うべきだ」というのが我々の見解である。岩崎(2015, 118 頁)には「対照群の分散共分散行列を使う」という言説があるが、それは誤りであろう。

7.3　汎距離計算と処置効果の推定

朝野(2025, 106 頁)でいうケース 6 の距離を用いる。

$$D_A^2\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)' \boldsymbol{S}_A^{-1}\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)$$

ここでiはA群、jはB群のサンプルを表す。最近距離法による結果は表 2 の通り。

表2　マハラノビス汎距離の最小値探し

	B1	B2	B3	B4	B5	Min
A1	0.25	19	2.58	3.58	5.33	B1
A2	2.25	7	0.58	1.58	1.33	B3
A3	3.25	21	1.58	0.58	1.33	B4

次に観測値である反応Yを用いて処理効果を計算した結果が表 3 である。処置によって反応は平均して 2.13 増加する。

表 3　マハラノビスマッチングによる効果測定

	2 群の反応		処置効果
ペア	YA	YB	YA−YB
A1B1	3.4	1	2.4
A2B3	4.5	2.7	1.8
A3B4	5.6	3.4	2.2
平均値			2.13

マッチングの許容限界であるカリパーの設定や、サンプルの復元・非復元など細かなバリエーションはあるものの、マハラノビスマッチングの本質的なロジックは以上の通りである。

7.4　Ｒの計算コード

```
# 処置群（A 群）のデータ
A_data <- matrix(c(1,1,5,5,9,6),
                 nrow = 3, ncol = 2, byrow = TRUE)
A <- A_data
dimnames(A) <- list(paste0("A", 1:3), c("X1", "X2"))

# 対照群（B 群）のデータ
B_data <- matrix(c(2,2,3,6,6,5,8,6,9,7),
```

```
                nrow = 5, ncol = 2, byrow = TRUE)
B <- B_data
dimnames(B) <- list(paste0("B", 1:5), c("X1", "X2"))

# -- 処置群(A)に関するオブジェクト ---
SA <- cov(A)        # 処置群の不偏分散共分散行列
SAinv <- solve(SA)  # その逆行列
n <- nrow(A)        # 距離行列の行数
m <- nrow(B)        # 距離行列の列数

# 汎距離の計算結果を格納する行列を初期化
D2 <- matrix(0, nrow = n, ncol = m)
dimnames(D2) <- list(rownames(A), rownames(B))

# 2 点間平方汎距離の計算（処置群 SA を基準にした場合）
# 処置群(A)の各メンバー(i)に対して、対照群(B)の全メンバー(j)との距離を計算
for (i in 1:n) {
  for (j in 1:m) {
    diff <- A[i, ] - B[j, ]
    # t()は転置、%*%は行列積
    D2[i, j] <- t(diff) %*% SAinv %*% diff
  }
}

# D2 をプリントして A の各要素別に一番距離が小さい B を探してマッチペアと判定する
cat("■ 処置群の共分散行列(SA)を基準としたマハラノビス距離行列¥n")
print(round(D2, 3))
cat("¥n")
```

7.5　今後の課題

共変量の数が多くて相関が高いケースでは、傾向スコア・マッチングよりもマハラノビス・マッチングの方が有効だろう。なぜなら傾向スコアではロジスティック回帰分析を用いるためマルチコ問題が考えられるからである。マーケティングの効果測定にマハラノビスのマッチングを適用することを今後の研究課題としたい。

８章　因果推論と機械学習

８.１　因果効果の異質性を捉えるメタ学習器

ビジネスにおけるデータに基づいた意思決定において、「ある施策 A を実行したら、結果 B はどれくらい変わるのか？」という因果関係を正確に把握することは極めて重要である。例えば、「テレビ CM を流せば、商品の売上は本当に伸びるのか？」といった問いに対し、明確な根拠をもって答えることは、効果的なマーケティング戦略の要となる。

しかし、因果効果を正確に測定するプロセスは単純ではない。単に CM を視聴した群と視聴しなかった群の売上を比較するだけでは、正しい効果は測定できない。これは、顧客の属性などが CM 視聴の有無と売上の両方に影響を与える「選択バイアス（セレクションバイアス）」が存在するためである。例えば、特定の居住エリア（例：郊外）の住民は CM を視聴する傾向が高く、かつ、元々の購買額の基準も異なる場合、その影響を取り除かなければ CM 本来の効果を見誤ることになる。

このような課題を解決し、選択バイアスを適切に調整した上で、個々の顧客特性に応じた因果効果の異質性を捉えるための枠組みとして機械学習手法を活用した「メタ学習器（Metalearner）」がある（Künzel (2019)）。

本章では、このメタ学習器を用いてテレビ CM が売上に与える因果効果を推定するアプローチについて、その基本的な概念と複数の手法について検討を行った結果についてまとめる。

８.２　メタ学習器とは

メタ学習器は、様々な機械学習モデルを応用して、個々の対象における処置の因果効果、すなわち「条件付き平均処置効果（CATE: Conditional Average Treatment Effect）」を推定するためのフレームワークである。ここでの「メタ」とは、単一の予測モデルを構築するのではなく、因果効果を推定するという特定の目的のために、複数の機械学習モデルを組み合わせたり、段階的に適用したりする高次の学習戦略を指す。

このフレームワークの大きな利点は、その柔軟性にある。ベース学習器（Base learner）として、線形回帰のような解釈性の高いモデルから、勾配ブースティングやニューラルネットワークといった複雑な非線形関係を捉えるモデルまで、データの特性や分析の目的に応じて自由に選択・組み込みが可能である。

主要なメタ学習器には、処置の有無を変数の一つとして単一モデルで学習する「S-Learner」、処置群と対照群で別々のモデルを学習する「T-Learner」、そして T-Learner を拡張し、群間で処置効果を相互補完する「X-Learner」などがあり、それぞれ異なる戦略で CATE の推定を行う。

■ S-Learner (Single-Learner)

S-Learner は、介入(処置)の有無を示す変数と他のすべての共変量を組み合わせて、最終的な結果変数を予測する単一の機械学習モデルで構築される。このモデルから、もし処置が行われていた場合の結果と、もし処置が行われなかった場合の結果を予測し、その差分を個別の処置効果として推定する。

数学的定義

S-Learner は、結果変数 Y の期待値(平均値)を共変量 X と処置変数 T の関数として直接モデル化する。

$$\mu(x,t) = E[Y|X = x, T = t] \qquad (7.1)$$

ここで、単一の回帰モデルfを学習させる。

$$\hat{\mu}\,(x,t) = f(x,t)$$

アルゴリズム

① **モデル学習**: 全データセットを用いて、結果変数 Yを目的変数とし、共変量 X と処置変数 T を特徴量とする単一の機械学習モデルfを学習させる。

$$Y \sim f(X,T)$$

② **CATE 推定**: 学習したモデル$\hat{\mu}$ を用いて、各個人 x について、処置を受けた場合(T =1)の予測結果と、処置を受けなかった場合(T =0)の予測結果を計算し、その差分を CATE として推定する。

$$\hat{\tau}\,(x) = \hat{\mu}\,(x,1) - \hat{\mu}\,(x,0) \qquad (7.2)$$

■ T-Learner (Two-Learner)

T-Learner は、処置群と対照群に対してそれぞれ個別の機械学習モデルを学習する点を特徴とする。この方法により、各群のデータ特性に特化した予測モデルを構築し、それらの予測結果の差分から CATE を推定する。

数学的定義

T-Learner は、処置を受けた場合の期待される結果と、処置を受けなかった場合の期待される結果を、共変量 X の条件のもとで別々にモデル化する。

- **処置群用モデル**: 処置を受けた場合の、共変量 x における結果変数 Y の期待値を推定する。

$$\mu_1(x) = E[Y|X = x, T = 1] \quad (7.3)$$

- **対照群用モデル**: 処置を受けなかった場合の、共変量 x における結果変数 Y の期待値を推定する。

$$\mu_0(x) = E[Y|X = x, T = 0] \quad (7.4)$$

アルゴリズム

① **データ分割**: まず、元のデータセットを処置変数T に基づいて二つのサブセットに分割する。

- **処置群データ**: {(Xi,Yi):Ti=1} となるデータポイントの集合でまとめる。
- **対照群データ**: {(Xi,Yi):Ti=0} となるデータポイントの集合でまとめる。

② **モデル学習**: 分割したそれぞれのサブセットに対し、独立した機械学習モデルを学習させる。

- $\hat{\mu}_1(x) = f_1(x)$ を処置群データを用いて学習させる。
- $\hat{\mu}_0(x) = f_0(x)$ を対照群データを用いて学習させる。

ここで、f_1 と f_0 は異なる、あるいは同じ種類の機械学習アルゴリズム（例: 線形回帰、決定木など）を適用する。

③ **CATE 推定**: 学習した二つのモデル $\hat{\mu}_1$ と $\hat{\mu}_0$ を用いて、各個人 x に対する CATE を推定する。これは、処置を受けた場合の予測結果と、処置を受けなかった場合の予測結果の差として計算される。

$$\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x) \quad (7.5)$$

■ X-Learner (Cross-Learner)

X-Learner は、T-Learner の拡張版であり、特に処置群と対照群のサンプルサイズ

が不均衡な状況でその真価を発揮する。この手法は、観測された結果だけでなく、「もし反対の処置を受けていたらどうなっていたか」という反事実（counterfactual）な状況を推測し、処置効果を補完すること（補完的処置効果）で、両群の情報をより効率的に相互活用し、より正確な因果効果の推定を可能とする。X-Learner は多段階のプロセスを通じて CATE を推定する。

数学的定義

X-Learnerは、まずT-Learnerと同様に各群の期待される結果をモデル化する。 第1段階で学習するモデル:

- 処置群の期待結果: $\hat{\mu}_1(x)$
- 対照群の期待結果: $\hat{\mu}_0(x)$

これらを用いて、「もし反対の処置を受けていたら」という状況をシミュレートし、補完的処置効果 (Imputed Treatment Effects) を計算する。

- **処置群における補完的処置効果**: 処置を受けた個人が、もし対照群だった場合の予測結果$\hat{\mu}_0(X_i^1)$ と、実際に観測された結果Y_iとの差分。これは、対照群のモデルを用いて処置群の欠損した対照結果を「補完する」個別処置効果と解釈される。

$$D_i^1 = Y_i^1 - \hat{\mu}_0(X_i^1) \quad (T_i = 1 \text{ のデータに対して計算})$$

- **対照群における補完的処置効果**: 対照群の個人が、もし処置群だった場合の予測結果 $\hat{\mu}_1(X_i^0)$ と、実際に観測された結果Y_iとの差分。これは、処置群のモデルを用いて対照群の欠損した処置結果を「補完する」個別処置効果と解釈される。

$$D_i^0 = \hat{\mu}_1(X_i^0) - Y_i^0 \quad (T_i = 0 \text{ のデータに対して計算})$$

アルゴリズム

X-Learner は以下の 4 つの段階を経て CATE を推定する。

第 1 段階: 結果モデルの学習

- **処置群データでモデル $\hat{\mu}_1(x)$を学習**: 処置を受けた人々の共変量 X と結果 Y の関係をモデル化する。
- **対照群データでモデル $\hat{\mu}_0(x)$を学習**: 処置を受けなかった人々の共変量 X と結果Y の関係をモデル化。

第 2 段階: 補完的処置効果の計算
第 1 段階で学習したモデルを利用し、各データポイントにおいて、実際に観測された結果と「もし反対の処置を受けていたら」という反事実的な予測結果との差を計算することで、**補完的処置効果**を作成する。

- **処置群の補完的処置効果の計算**: 処置群の各個人 i について、彼らがもし対照群だった場合の予測結果$\hat{\mu}_0(X_i)$を、実際の観測結果 Y_i から差し引く。

- **対照群の補完的処置効果の計算**: 対照群の各個人 i について、彼らがもし処置群だった場合の予測結果 $\hat{\mu}_1(X_i)$ から、実際の観測結果 Y_i を差し引く。

第 3 段階: 処置効果モデルの学習
計算された補完的処置効果を目的変数として、処置効果それ自体を予測する二つの新しいモデルを学習する。

- **処置群の処置効果モデル $\hat{\tau}_1(x)$ を学習**: 処置群のデータと、彼らの補完的処置効果 D_i^1を用いて、処置効果を予測するモデル$\hat{\tau}_1(x)$を学習する。
- **対照群の処置効果モデル $\hat{\tau}_0(x)$を学習**: 対照群のデータと、彼らの補完的処置効果 D_i^0 を用いて、処置効果を予測するモデル$\hat{\tau}_0(x)$を学習する。

第 4 段階: 最終的な CATE 推定
第 3 段階で学習した二つの処置効果モデルを統合し、傾向スコアを用いて重み付けを行うことで、最終的な CATE を推定する。

- **傾向スコア$e(x)$ の推定**: 各個人 x が処置を受ける確率 P(T=1|X=x) を推定する傾向スコアモデル（例: ロジスティック回帰）を学習する。

- **最終的な CATE の算出**: 傾向スコア $e(x)$ を重みとして、$\hat{\tau}_0(x)$と $\hat{\tau}_1(x)$を線形結合させる。これは、各個人 x がどちらの群に属する可能性が高いかに基づいて、適切な処置効果推定値を割り当てることを意味する。

$$\hat{\tau}(x) = e(x) \cdot \hat{\tau}_0(x) + (1 - e(x)) \cdot \hat{\tau}_1(x) \qquad (7.6)$$

8.3　テレビ CM の広告効果を推定する人工データ

本節では、テレビ CM の広告効果を推定する人工データを用いる。このデータセットには、CM 視聴（介入）が顧客の購入金額（結果）に与える影響を分析するための情報が含まれる。なお、この人工データの作成の設計指針については、末尾の「《参考》 人工データの生成ロジック」を参照されたい。

データセットの共変量（age、gender、residence_type、past_purchase）、介入変数、結果変数は次の通りである。

- **共変量 (X)**: gender（性別）, age（年齢）, residence_type（居住エリア）, past_purchase（過去の購入金額）。これらはCM視聴の有無と購入金額の両方に影響を与える交絡因子（選択バイアスの原因）と仮定する。
- **介入 (T)**: T_cm_watched（テレビ CM を視聴したか否か。1: 視聴, 0: 非視聴）
- **結果 (Y)**: Y_purchase（商品の購入金額）

■　データセットの探索的分析（EDA）

これらの CM 視聴の有無と結果変数である購入金額にどのように影響しているかを探索的に分析する。これにより、どのような選択バイアスが存在するのか、そしてそれが因果効果の推定にどのように影響しうるのかを具体的に確認する。以下のコードは python による。

```
#汎用ライブラリ
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

# 機械学習ライブラリ
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import StandardScaler
```

```
df = pd.read_csv('casual_sample_data.csv')
df
```

	gender	age	residence_type	past_purchase	T_cm_watched	Y_purchase	true_effect
0	1	58	0	4635.382916	0	3153.455045	1640
1	1	48	1	6295.807489	0	2260.844338	1840
2	0	34	0	4417.225966	1	5549.902117	2120
3	1	27	0	2110.317591	0	3085.970135	2260
4	1	40	1	6505.536067	0	1718.271851	2000
...	...	...	...	...	...	...	...
4995	1	38	1	3650.949080	0	436.741068	2040
4996	0	24	1	3533.925125	1	2320.000000	2320
4997	1	42	1	3267.365337	0	1482.672308	1960
4998	0	49	1	6633.326931	1	1986.164042	1820
4999	0	44	1	3748.129134	1	1920.000000	1920

5000 rows × 7 columns

■　単純比較による因果効果の「ナイーブな推定」

CM を視聴したグループ（処置群）と CM を視聴しなかったグループ（対照群）の平均購入額を比較し、その差分を算出する。　この差分はしばしば「ナイーブな効果」や「見かけ上の効果」と呼ばれている。

```
# 説明変数・原因変数・目的変数
features = ['gender', 'age', 'residence_type', 'past_purchase']

X = df[features]         #説明変数（共変数）
T = df['T_cm_watched']   #原因変数（処置変数；介入変数）
Y = df['Y_purchase']     #目的変数（結果変数）

#介入効果
true_cate = df['true_effect']
```

```
# 処置群(T=1)と対照群(T=0)のデータに分割
df_treat = df[T == 1]
df_control = df[T == 0]

X_treat, Y_treat = df_treat[X.columns], df_treat['Y_purchase']
X_control, Y_control = df_control[X.columns], df_control['Y_purchase']
```

```
#単純平均
mean_treated = Y_treat.mean()
mean_control = Y_control.mean()

naive_effect = mean_treated - mean_control

print("¥n--- [0] 単純比較 ---")
print(f"CMを見た人の平均購入額: {mean_treated:.2f}円")
print(f"CMを見ていない人の平均購入額: {mean_control:.2f}円")
print(f"単純な差分（ナイーブな効果）: {naive_effect:.2f}円")
```

```
--- [0] 単純比較 ---
CMを見た人の平均購入額: 2985.98円
CMを見ていない人の平均購入額: 2294.78円
単純な差分（ナイーブな効果）: 691.21円
```

この「単純な差分：691.21 円」は、一見すると CM を視聴することによって購入額が約 691 円増加したように見える。しかし、この数値は CM の純粋な因果効果を正確に示しているわけではない。

その理由は、CM を視聴したグループと視聴しなかったグループの間には、CM 視聴以外の特性（共変量）において既に大きな違い（選択バイアス）が存在するためである。

■　処置群と対照群の平均値比較

はじめに処置群（T=1、CM 視聴あり）と対照群（T=0、CM 視聴なし）とに分割し、それぞれの群データにおける各変数の平均値を算出する。この比較は、選

択バイアス（処置を受けるかどうかがランダムではないこと）が存在するかどうかを理解する上で必ず行うべき作業となる。選択バイアスが存在する場合は処置群と対照群の特性に偏りがあるため、単純な平均値の比較だけでは真の因果効果を正確に捉えることができない。

```
# 各グループの平均値を計算し、DataFrameにまとめる
mean_df = pd.DataFrame({
    '処置群（T=1）平均': df_treat.mean(),
    '対照群（T=0）平均': df_control.mean()
})
mean_df
```

	処置群 (T=1) 平均	対照群 (T=0) 平均
gender	0.269983	0.862972
age	38.182090	43.268514
residence_type	0.657048	0.231738
past_purchase	5049.897194	4898.836383
T_cm_watched	1.000000	0.000000
Y_purchase	2985.984367	2294.778839
true_effect	2036.358209	1934.629723

処置群（CM 視聴あり）と対照群（CM 視聴なし）の間で、各共変量の平均値を比較することで、選択バイアスの具体的な内容を確認できる。

変数	処置群 (T=1) 平均	対照群 (T=0) 平均	備考
gender	0.27	0.86	**処置群は男性（0 に近い）に大きく偏り**、対照群は女性（1 に近い）に偏っている。これは「男性ほど CM を見る確率が高い」という設定が強く反

変数	処置群 (T=1) 平均	対照群 (T=0) 平均	備考
			映されている。
age	38.2	43.3	**処置群（CM 視聴あり）の平均年齢が対照群（CM 視聴なし）よりも約 5 歳若い**。これは「若いほど CM を見る確率が高い」という設定が反映されている。
residence_type	0.657	0.232	**処置群は都心（1 に近い）に大きく偏り**、対照群は郊外（0 に近い）に偏っている。これは「都心に住んでいるほど CM を見る確率が高い」という設定が反映されている。
past_purchase	5050	4899	処置群の過去購入額が対照群よりもわずかに**高い**傾向がある。CM 視聴確率には正の影響を与えたものの、Y0 には小さな影響しか与えなかったため、バイアスとしては比較的小さい。
T_cm_watched	1	0	定義通り。
Y_purchase	2985	2294	観測された平均購買額。単純な差分（ナイーブな効果）は約 691 円ながら、これが真の因果効果ではないことは前述の通り。
true_effect	2036	1934	シミュレーションデータにおける「真の因果効果」の平均。処置群と対照群の間で真の因果効果の平均値にわずかな差があるのは、真の CATE が共変量（年齢、性別など）に依存して異質性を持つため。特に、**CM を見たグループの平均的な真の CM 効果（2036 円）と CM を見ていないグループの平均的な真の CM 効果（1934 円）は、ほとんど差がない**。

この比較結果から、**選択バイアスが非常に強く存在している**ことが明確に確認できる。 特に、gender、age、residence_type において、処置群と対照群の間で平均値に大きな違いが見られる。これは、CM を見た人と見ていない人の間に、これらの特性において顕著な偏りがあることを意味する。

8.4　メタ学習器による因果効果推定

本節では、このデータを用いて、S-Learner, T-Learner, X-Learner の各手法が CM の因果効果をどのように推定するかをベース学習器として非線形の「ランダムフォレスト」を用いて比較・検討する。

■　変数の設定

因果推論の分析を行うためのデータ準備、特に**説明変数、原因変数、目的変数の定義**を行う。

```
# 説明変数・原因変数・目的変数
features = ['gender', 'age', 'residence_type', 'past_purchase']

X = df[features]          #説明変数（共変数）
T = df['T_cm_watched']    #原因変数（処置変数；介入変数）
Y = df['Y_purchase']      #目的変数（結果変数）

#真の介入効果
true_cate = df['true_effect']
```

■　処置群(T=1)と対照群(T=0)のデータに分割

因果推論の次のステップでメタ学習器を適用するために、元のデータセットを処置群と対照群に分割する。この分割は、特に T-Learner や X-Learner といったメタ学習器で、それぞれの群に対するモデルを個別に学習させる際に不可欠な前処理となる。

```
# 処置群(T=1)と対照群(T=0)のデータに分割
df_treat = df[T == 1]
df_control = df[T == 0]

X_treat, Y_treat = df_treat[X.columns], df_treat['Y_purchase']
X_control, Y_control = df_control[X.columns], df_control['Y_purchase']
```

続いて、全員が処置された場合(T=1)とされなかった場合(T=0)を仮想的に作成する。ここでは copy()メソッドを使い、次のように treat と control の変数として定義する。

```
# 全員が処置された場合(T=1)とされなかった場合(T=0)を仮想的に作る
X_s_treat = X.copy(); X_s_treat['T'] = 1
X_s_control = X.copy(); X_s_control['T'] = 0
```

```
X_s_treat
```

	gender	age	residence_type	past_purchase	T
0	1	58	0	4635.382916	1
1	1	48	1	6295.807489	1
2	0	34	0	4417.225966	1
3	1	27	0	2110.317591	1
4	1	40	1	6505.536067	1
...	...	...	...	...	...
4995	1	38	1	3650.949080	1
4996	0	24	1	3533.925125	1
4997	1	42	1	3267.365337	1
4998	0	49	1	6633.326931	1
4999	0	44	1	3748.129134	1

5000 rows × 5 columns

第１項　S-Learner による検討

S-Learner は、CM 視聴（介入）の有無を説明変数の一部として含めた単一のモデルを構築し、そこから因果効果を導き出すアプローチを取る。S-Leaner における介入変数 T を特徴量に加えて、単一のモデルを学習させるための説明変数を次のように用意する。

```
# 介入変数Tを特徴量に加えて、単一のモデルを学習
X_s = X.copy()
X_s['T'] = T
X_s
```

	gender	age	residence_type	past_purchase	T
0	1	58	0	4635.382916	0
1	1	48	1	6295.807489	0
2	0	34	0	4417.225966	1
3	1	27	0	2110.317591	0
4	1	40	1	6505.536067	0
...	...	...	...	...	...
4995	1	38	1	3650.949080	0
4996	0	24	1	3533.925125	1
4997	1	42	1	3267.365337	0
4998	0	49	1	6633.326931	1
4999	0	44	1	3748.129134	1

5000 rows × 5 columns

次に非線形回帰の RandomForest をベース学習器として設定し、Fit メソッドによる最適化をかける。

```
# CMを見た（介入）を加えた説明変数で機械学習
model_s = RandomForestRegressor(n_estimators=100,
                                max_depth=20,
                                random_state=42)
model_s.fit(X_s, Y)
```

```
RandomForestRegressor
RandomForestRegressor(max_depth=20, random_state=42)
```

■ S-Learner による CATE の計算

S-Learner による CATE の推定は、上記で学習した `model_s` を使って行われる。

```
cate_s = model_s.predict(X_s_treat) - model_s.predict(X_s_control)
cate_s
```

```
array([ 587.52127097,  560.02247728, 2165.73067083, ..., 1725.53416474,
       1125.30377577, 1259.42305301])
```

ここで `model_s.predict(X_s_treat)` は、全顧客が CM を視聴した場合の予測購入金額を算出している。一方、`model_s.predict(X_s_control)` は、全顧客が CM を視聴しなかった場合の予測購入金額を算出している。これらの予測値の差分が、各顧客（または各行）における CM 視聴の推定される CATE となる。

■ 代表的な因果効果指標の計算

推定された CATE を用いて、S-Learner における ATE (Average Treatment Effect)、ATT (Average Treatment Effect on the Treated)、ATU (Average Treatment Effect on the Untreated)を計算する。

```
# 結果を格納する辞書
results = {}

# ATE, ATT, ATUを計算
results['S-Learner'] = {
    'ATE': cate_s.mean(),
    'ATT': cate_s[T==1].mean(),
    'ATU': cate_s[T==0].mean()
}
```

```
print("¥n--- S-Learner ---")
print(f"推定されたATE: {results['S-Learner']['ATE']:.2f}円")
print(f"推定されたATT: {results['S-Learner']['ATT']:.2f}円")
print(f"推定されたATU: {results['S-Learner']['ATU']:.2f}円")
```

```
--- S-Learner ---
推定されたATE: 1757.19円
推定されたATT: 1790.70円
推定されたATU: 1706.30円
```

■ 真の因果効果との比較

- 本来の ATE（真の平均処置効果）: 1995.97 円
- ナイーブな効果（単純比較）: 691.21 円

RandomForest をベース学習器とした S-Learner の推定 ATE（1757.19 円）は、ナイーブな効果（691.21 円）からは大幅に改善され、真の ATE（1995.97 円）に近づいている。しかし、真の ATE との差はまだ約 238 円（1995.97 − 1757.19）となっている。

これは、RandomForest が非線形な関係性を捉える能力を持っているものの、選択バイアスを効果的に調整し、データに存在する「因果効果の異質性」（CM の効果が顧客の属性によって異なること）を完全に捉えきれていない可能性を示唆している。

第２項 T-leaner による検討

T-Learner では、処置を受けたグループと受けなかったグループに対して、それぞれ独立した予測モデルを学習することで因果効果を推定する。

```
# 処置群モデルと対照群モデルをそれぞれ学習
model_1 = RandomForestRegressor(n_estimators=100,
                                max_depth=20,
                                random_state=42)
model_0 = RandomForestRegressor(n_estimators=100,
                                max_depth=20,
                                random_state=42)

model_1.fit(X_treat, Y_treat)
model_0.fit(X_control, Y_control)
```

```
RandomForestRegressor(max_depth=20, random_state=42)
```

■　T-Learner による CATE の計算

推定された CATE を用いて、T-Learner における ATE、ATT、ATU を計算する。

```
# 全データに対して、処置された場合・されなかった場合の購入額を予測
cate_t = model_1.predict(X) - model_0.predict(X)
```

```
# ATE, ATT, ATUを計算
results['T-Learner'] = {
    'ATE': cate_t.mean(),
    'ATT': cate_t[T==1].mean(),
    'ATU': cate_t[T==0].mean()
}
```

```
print("¥n--- T-Learner ---")
print(f"推定されたATE: {results['T-Learner']['ATE']:.2f}円")
print(f"推定されたATT: {results['T-Learner']['ATT']:.2f}円")
print(f"推定されたATU: {results['T-Learner']['ATU']:.2f}円")
```

```
--- T-Learner ---
推定されたATE: 1765.79円
推定されたATT: 1824.77円
推定されたATU: 1676.22円
```

■　真の因果効果との比較

S-Learnerも含めた推定値と「真の因果効果」を比較してみよう。

- 本来のATE(真の平均処置効果): 1995.97円
- S-Learnerの推定ATE: 1757.19円
- T-Learnerの推定ATE: 1765.79円

T-Learnerによる推定ATE（1765.79円）は、S-Learnerの推定値(1757.19円)とほぼ同じ値であり、真のATE(1995.97円)との間にはまだ約230円の差がある。S-Learnerよりわずかに真の値に近づいているが、劇的な改善は見られていない。これは、ベース学習器としてRandomForestを使用している場合、S-LearnerとT-Learnerの間で推定精度に大きな差が出にくい場合があることを示唆している。

第３項　X-Learner（Cross-Learner）
X-LearnerはT-Learnerを拡張した手法であり、特に処置群と対照群のサンプルサイズが不均衡な場合に有効とされる。多段階のプロセスを経てCATEを推定する。この手法は、大きく3つのステージで構成される。

ステージ1: T-Learnerと同様に、処置群と対照群それぞれで結果変数(購入金額)を予測するモデルを学習させる。

ステージ2: 各群で観測されなかった「反実仮想的な結果」をステージ1のモデルで予測し、それらを使って2つの「補完的因果効果モデル」を学習させる。これにより、データが少ない群の情報を、もう一方の群から補完する形で活用する。

ステージ3: 最後に、傾向スコア(CMを見る確率)を推定し、この傾向スコアでステージ2で学習した2つの補完的因果効果モデルを重み付けして統合することで、最終的なCATE(条件付き平均処置効果)を算出する。

Pythonのコードで記述すると以下のように計算される。

```
# 第一段階：T-Learnerと同じモデルを使用（model_1, model_0）

# 第二段階および第三段階：疑似的な効果を計算し、それを目的変数として効果モデルを学習

# 処置群の効果モデル
D_treat_imputed = Y_treat - model_0.predict(X_treat)
model_tau_1 = RandomForestRegressor(n_estimators=100,
                                    max_depth=20,
                                    random_state=42).fit(X_treat, D_treat_imputed)

# 対照群の効果モデル
D_control_imputed = model_1.predict(X_control) - Y_control
model_tau_0 = RandomForestRegressor(n_estimators=100,
                                    max_depth=20,
                                    random_state=42).fit(X_control, D_control_imputed)
```

■　X-Learner による CATE の計算
推定された CATE を用いて、X-Learner における ATE、ATT、ATU を計算する。

```
# 第四# Stage 3: 傾向スコアモデル (LogisticRegressionを使用)
prop_model = LogisticRegression(random_state=42).fit(X, T)
pscore = prop_model.predict_proba(X)[:, 1]

cate_x = pscore * model_tau_0.predict(X) + (1 - pscore) * model_tau_1.predict(X)
```

```
# ATE, ATT, ATUを計算
results['X-Learner'] = {
    'ATE': cate_x.mean(),
    'ATT': cate_x[T==1].mean(),
    'ATU': cate_x[T==0].mean()
}
```

```
print("¥n--- [3] X-Learner ---")
print(f"推定されたATE: {results['X-Learner']['ATE']:.2f}円")
print(f"推定されたATT: {results['X-Learner']['ATT']:.2f}円")
print(f"推定されたATU: {results['X-Learner']['ATU']:.2f}円")
```

```
--- [3] X-Learner ---
推定されたATE: 2080.51円
推定されたATT: 2108.45円
推定されたATU: 2038.08円
```

■　真の因果効果との比較と考察
- 本来の ATE(真の平均処置効果): 1995.97 円
- S-Learner の推定 ATE: 1757.19 円
- T-Learner の推定 ATE: 1765.79 円
- X-Learner の推定 ATE: 2080.51 円

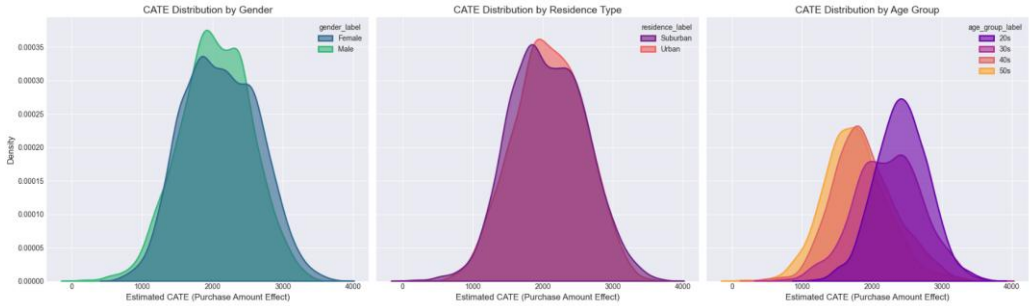
今回の結果を見ると、X-Learner の推定 ATE (2080.51 円) は、S-Learner (1757.19 円) や T-Learner (1765.79 円) の推定値と比較して真の ATE(1995.97 円)に最も近い値を推定できている。真の ATE との差はわずか約 84.54 円(2080.51 − 1995.97)であり、これは RandomForest をベース学習器としたメタ学習器の中では最も優れた推定を示している。

■　各共変量における CATE の分析
機械学習モデルから、推定された CATE を、顧客の属性(性別、居住エリア、年齢層)ごとに分析・可視化することができる。

次の 3 つのグラフは、ランダムフォレストをベース学習器とした X-Leaner 機械学習

モデルによって推定された CATE を、顧客の属性（性別、居住エリア、年齢層）別に可視化したものである。

グラフの横軸は CATE（購入金額への効果）、縦軸は密度（その効果を持つ顧客の多さ）を示しており、分布の山が右にあるほど、そのセグメントに対する広告効果が高いことを意味する。



1. 性別（Gender）による CATE 分布

女性（青緑色）の分布は、男性（明るい緑色）の分布よりもわずかに右側に位置している。女性の分布のピーク（最も多くの人が集まる効果量）は、男性のピークよりも高い値にある。これは、テレビ CM は、男性よりも女性に対して、より高い購入促進効果をもたらしたと分析できる。その差は劇的ではないが、一貫して女性の方が高い効果を示している。

2. 居住エリア（Residence Type）による CATE 分布

都心（Urban，赤色）の分布は郊外（Suburban，紫色）の分布と比較して、わずかに右側にシフトしている。郊外層の CATE のピークは二峰性を示し、約 1800 円あたりにあるのに対し、都心層のピークは 1900 円あたりと、わずかながら差が見られている。これは、テレビ CM の広告効果は、郊外在住者よりも都心在住者に対して、わずかに高かったことを示唆する。

3. 年齢層（Age Group）による CATE 分布

分布は年齢が若い順に右側に位置しており、20 代（紫色）の効果が最も高く、次いで 30 代（ピンク色）となっている。40 代（オレンジ色）と 50 代（黄色）の分布はほぼ重なっており、20 代・30 代と比較して左側に位置している。若い世代ほど分布の裾野が広く、効果のばらつきが大きい傾向が見られている。これより、広告効果は若年層ほど高く、年齢が上がるにつれて減少する傾向が明確に見て取れる。特に 20 代は、この広告キャンペーンにおける最も反応が良いコアターゲットであると言える。

4.総合的な考察とビジネスへの示唆

これらの分析結果を統合すると、今回のテレビ CM から最も大きな影響を受けた顧客像は「都心に住む 20 代の女性」であると特定できる。逆に、「郊外に住む 40 代以上の男性」は、広告効果が相対的に低いセグメントであると分析される。

8.5 まとめ

メタ学習器は、選択バイアスの存在する観測データから、施策の因果効果を推定するための強力かつ柔軟なフレームワークである。

本稿では示していないが、ベース学習器として線形重回帰の結果もあわせると、分析の結果、メタ学習器の性能は、S-Learner、T-Learner、X-Learner といった手法の選択だけでなく、その内部で用いるベース学習器（線形回帰か、ランダムフォレストかなど）との組み合わせに大きく依存することも明らかになった。

最もシンプルな S-Learner は、いずれのベース学習器でも精度に限界があった。T-Learner は状況によって高い性能を発揮するが、不安定さも併せ持つ。一方で、X-Learner は、ランダムフォレストのような表現力の高い非線形モデルと組み合わせることで、最も正確かつ頑健な因果効果の推定を実現した。

このことは、実世界の複雑なデータに対峙する際、単一の手法に固執するのではなく、データの特性を考慮しながら複数のメタ学習器とベース学習器の組み合わせを試し、その結果を慎重に比較検討することの重要性を示している。

特に、効果の異質性や選択バイアスが複雑に絡み合う状況においては、X-Learner がデータに基づいた精度の高い意思決定を導くための有力な選択肢となる。

8.6 《参考》 人工データの生成ロジック

人工データは 5000 人の仮想顧客を対象に、以下のロジックに基づいて生成している。

1. 共変量の生成:

age、gender、residence_type、past_purchase は、それぞれランダムな分布に従って生成させる。これにより、多様な顧客プロファイルを持つデータセットが構築される。

2. 選択バイアスの作り込み

CM 視聴の有無（T_cm_watched）は、ここでは意図的に「男性かつ都心に住むほどCMを見る確率が高い」という強いバイアスをかけて設定した。

具体的には、CM 視聴確率 P(T=1) は以下のロジスティック関数を用いて設計されている。

$$P(T = 1) = \frac{1}{(1 + exp(-z))}$$

$$z = -5.0 + 6.0 \cdot (1 - gender) + 5.0 \cdot residence\_type - 0.1 \cdot (age - 40) + 0.00015 \cdot past\_purchase$$

これにより、gender（男性ほど）と residence_type（都心ほど）が CM 視聴確率に強く影響を与えるように設計され、CM を見たグループと見ていないグループの間で、性別や居住エリアといった特性に大きな偏り（バイアス）を生じさせている。

3. CMを見なかった場合の購入額の設計

CM を視聴しなかった場合の購入額 Y_0 は、各顧客の共変量に依存するように設定されている。

$$Y_0 = 3500 - 1500 \cdot (1 - gender) - 1800 \cdot residence\_type - 20 \cdot age + 0.05 \cdot past\_purchase + Noise$$

ただし、

$$Noise \sim N(0, 800^2)$$

この式は、男性（1−gender が 1）であるほど、都心居住者（residence_type が 1）であるほど、そして年齢が高いほど、CM を視聴しなかった場合のベースライン購買額が低くなるという特性を組み込んでいる。

特長として、CM を見る確率が高い顧客（男性、都心居住者）ほど、CM がなかった場合の購買額 Y_0 が低いという設計としている。この相反する関係性が、選択バイアスを強く引き起こし、単純な CM 効果の比較が真の因果効果から大きく乖離する原因としている。

4. 真の因果効果（true_effect）の設計

CM が購入額に与える純粋な効果は、true_effect として別に計算する。

$$true\_effect = 2000 + (40 - age) \cdot 20$$

この式は、CM の基本的な効果が 2000 円であることに加え、「若者ほど CM の追加

効果が高い」という因果効果の異質性を加えている。（例えば、40 歳で効果 2000 円、20 歳で効果 2400 円、60 歳で効果 1600 円）。

この効果は 500 円から 3000 円の範囲に収まるようにしている。

5. 観測データの生成

最終的な購入金額 Y_purchase は、各顧客が実際に CM を視聴したかどうか（T_cm_watched）に基づいて生成される。

観測される購入額 Y は以下のロジックで計算される。

$$Y = T \cdot Y_1 + (1 - T) \cdot Y_0$$

ここで、

$$Y_1 = Y_0 + \ true\_effect$$

は CM を視聴した場合の購入額である。つまり、CM を視聴した顧客はY_1 を、視聴しなかった顧客は Y_0として観測されるものとしている。

9章　討論

9.1　本研究の要約

因果推論の研究は、パス解析から始まり、傾向スコア、グラフィカルモデリングを経て、近年では機械学習に基づく統計的因果推論が先端的なテーマとなっている。

ベイジアンネットワーク（BN）は、いくつかの事象の確率的な関係性をネットワーク図（グラフ構造）と確率分布でモデル化する手法であり、因果関係や確率的相互作用を効率的に表現できる。有向非循環グラフ（DAG）を前提とし、構造学習とパラメーター学習を経て構築される。R 言語を用いたタイタニック号の生存状況分析や EC サイトの顧客コンバージョン要因分析の事例を紹介した。連続量データにはガウシアン・ベイジアンネットワーク（GBN）が適用され、パス解析にはない推論・シミュレーション機能が利点とされる。

ランダム割付ができないキャンペーンのような販売促進施策では、参加者と不参加者の間で消費者特性に違い（選択バイアス）が生じるため、単純な比較では施策効果を正しく測定できない。二重にロバストな推定法は、処置群の反実仮想結果の推定と、対照群の傾向スコアを使った重み付け補正を組み合わせることで、結果変数モデルと傾向スコアモデルのどちらか一方が正しく指定できていれば、偏りのない推定量を得られるという頑健な特性を持つ。

マハラノビスマッチングでは、傾向スコアが持つ「共変量が大きく離れていても傾向スコアが等しくなる危険性」という課題に対し、2 点間の汎距離に基づく類似度評価がより直接的であると指摘した。一方で、傾向スコアマッチング（PSM）には「強い無視可能性」の仮定、未観測交絡の影響、2 値処置変数への適用範囲の限定といった構造的課題も存在する。

さらに、因果推論と機械学習を融合したメタ学習器は、選択バイアスのある観察データから、個々の顧客特性に応じた因果効果の異質性（CATE）を正確に把握するための柔軟なフレームワークである。S-Learner、T-Learner、X-Learner が主要手法であり、テレビ CM 広告効果推定の人工データ分析事例では、X-Learner が真の平均処置効果（ATE）に最も近い推定値を示すことが明らかにされた。メタ学習器の性能は、手法とベース学習器の組み合わせに大きく依存することも強調している。

第 1 章において、以下の研究課題を今回の研究の課題意識として提示した。

1. GM、とりわけベイジアンネットワーク（Bayesian network: BN）の利用価値
   ①　GM とパス解析、SEM の関係は何か
   ②　条件付き独立の仮定とは何か
   ③　因果の正しい方向をデータから識別できるのか

④　探索的な目的で因果モデルを作成する方法

2.　施策効果の測定

これらに対して本研究では以下の回答を示すことができた。

1.GM、とりわけベイジアンネットワーク（Bayesian network: BN）の利用価値

第 2 章、第 3 章、第 4 章において、ベイジアンネットワークが確率的関係性のモデル化に活用できることが明らかにされ、マーケティングでの応用事例も示された 。

①GM とパス解析、SEM の関係は何か

第 3 章において、連続量のベイジアンネットワーク（GBN）とパス解析のパラメータ推定値が一致する一方、GBN はシミュレーションが可能な点で優位性がある明らかになった 。

②条件付き独立の仮定とは何か

第 4 章において、この仮定がベイジアンネットワークの数理的基盤であり、同時確率分布の計算を単純化する上で不可欠であることが解説された

③探索的な目的で因果モデルを作成する方法

第 2 章および第 4 章で、スコアベースや制約ベースのアプローチを、専門家の知見（ブラックリスト等）と組み合わせるハイブリッド戦略が有効な方法として提示された 。

2.施策効果の測定

第 5 章の「二重にロバストな推定法」や第 8 章の「メタ学習器」など、複数の章でセレクションバイアスを補正し、施策効果を測定するための具体的な手法が詳述された 。

9.2　今後の研究課題

データからの因果の正しい方向の識別法：

当初の課題意識に含まれていたものの、積み残しになったのが「因果の正しい方向をデータから識別できるのか」という課題である。今回の研究では専門家のドメイン知識に基づいて「ありえない因果関係（ブラックリスト）」などを定義し、探索空間を限定する必要性を強調するに留まった。データから因果の方向を識別する研究は方法としては進んでいて、まだ実社会への実践導入が遅れているのが実情である。何が実践導入の障害になっているのかを含め

ての検討が必要になっている。

実データを用いた複数手法の比較検証：

ベイジアンネットワーク、傾向スコア、Causal Forest など多様な手法を紹介したが、今後の課題として、同一のマーケティング課題（例：特定のキャンペーン効果測定）に対してこれらの手法を横断的に適用し、各手法の予測精度や解釈のしやすさ、計算コストなどを比較検証することが考えられる。これにより、どのような状況でどの手法が最も有効であるかという実践的な知見が得られる。

異質な介入効果の深層的分析：

Causal Forest によって顧客ごとの「異質な介入効果」が推定できるが、なぜそのような異質性が生まれるのか、その要因を深掘りする研究が次のステップとなりえる。例えば、効果が高かった顧客セグメントの属性や行動データをさらに分析し、より効果的なターゲティング戦略の立案に繋げる研究が考えられる。

時系列データを組み込んだ因果推論：

本稿で紹介されている手法の多くは、ある一時点での効果測定に焦点を当てている。しかし、マーケティング施策の効果は時間と共に変化することが少なくない。今後の研究課題として、時系列分析の視点を加え、施策の長期的な効果や、効果が減衰するパターンなどをモデル化することが挙げられる。

付録 A 相関と偏相関の相互交換性

Dempster (1972)は偏相関を用いて変数間のアローをカットする共分散選択を提唱した。因果モデルをシンプルにする方法であった。小島（2008）はこの偏相関こそがグラフィカルモデリングの根幹的な概念であると指摘している。

この付録ではまず偏相関の導出過程を示そう。すでに稲垣（2003,213 頁）などで専門的な解説はなされているが、本稿では前提知識を要さない導出法を示す。次に相関と偏相関が同一の変換で交換可能なことを指摘する。我々の知る限りではこれは新しい指摘である。なお標準的な分析ソフトでは共分散選択後の相関係数を計算できない。なぜなら共分散選択後のデータ行列は不明だからである。そのため共分散選択後の偏相関行列にもとづいて相関行列を導く変換が必要になる。

A.1 偏相関の導出過程

■ 偏相関の幾何学的定義

サンプル数を n 人、変数を p 個、偏相関を計算したい変数の添字を $j,k=1,2,\cdots,p,(j\neq k)$ としよう。変数の配置を並べ替えても変数の情報は変わらないので、関心のある変数 y,z を $j=1,k=2$ の位置に移動させ、残りの変数をまとめて rest と呼ぶことにしよう。次に p 個の変数を $U=(y,z),\ T=(U,X)$ の 2 群に分割する。Tの 2 番目の Xが rest 変数群を指す。y,z を一つずつ目的変数に選び Xを説明変数にして重回帰分析をしたとして、残差ベクトルを $\boldsymbol{e}$ で表す。下記の $\prod_X^{\perp}$ は X の直交補空間への射影子であり、rest 変数の影響を取り除く作用をする[3]。

$$\boldsymbol{e}_y=\prod_X^{\perp}\boldsymbol{y}\ ,\boldsymbol{e}_z=\prod_X^{\perp}\boldsymbol{z}$$

偏相関（partial correlation,ここでは $Pcor$ と略す）は、この２つの残差ベクトルの相関係数として定義される。

$$Pcor_{yz}=\cos\theta=\frac{(\boldsymbol{e}_y,\boldsymbol{e}_z)}{\|\boldsymbol{e}_y\|\|\boldsymbol{e}_z\|}=\frac{\boldsymbol{y}'\prod_X^{\perp}\boldsymbol{z}}{\sqrt{\boldsymbol{y}'\prod_X^{\perp}\boldsymbol{y}}\sqrt{\boldsymbol{z}'\prod_X^{\perp}\boldsymbol{z}}}\quad\cdots\cdots\cdots\cdots\cdots\cdots①$$

[3] $\prod_X^{\perp}=\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

偏相関の意味を図解したのが図１である。２つの残差ベクトルの角度を θ とすれば $\cos\theta$ が偏相関である。

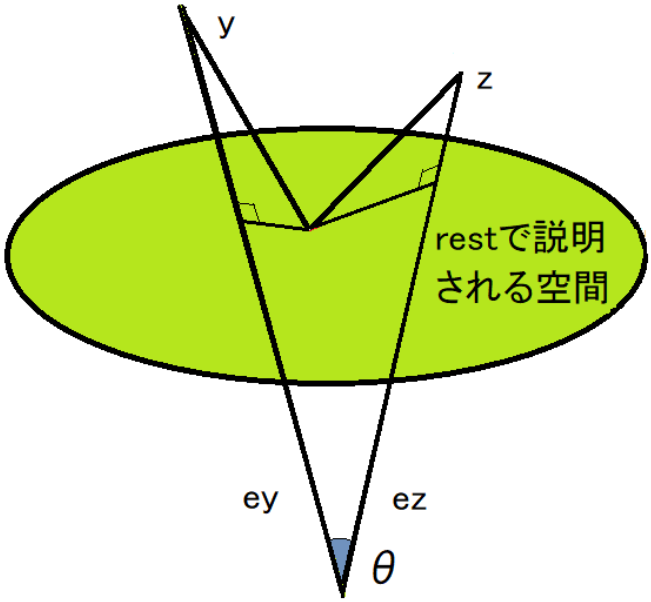


図１　偏相関の幾何学的定義

実際に重回帰分析を実行して残差ベクトルどうしの相関を計算すれば偏相関が得られる。しかし通常の統計ソフトでは、このような煩雑な計算はしない。後の⑤式の計算式で済ませている。この計算式を導出する準備として、まず必要な記法について説明しよう。

■ 相関行列とその逆行列

p 個の変数を平均 0，分散 1 に規準化しておこう。全体の相関行列を U と X の 2 群に分割した $p \times p$　次の相関行列は②の通りになる。これを分割行列あるいはブロック行列と呼ぶ。

$$\boldsymbol{R} = \frac{1}{n}\boldsymbol{T}'\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & \boldsymbol{C} \end{bmatrix} \quad \cdots\cdots\cdots\cdots\cdots\cdots ②$$

p 個の変数が一次独立だと仮定すれば相関行列は逆行列を持つ。相関行列の逆行列 $\boldsymbol{R}^{-1}$ も②と同じ次数で４分割する。逆行列で関心があるのは次式の部分行列 $\boldsymbol{E}^{-1}$ なので、残りの部分行列は―で示した。また逆行列の部分行列であることを上付き添字で示した。

$$\boldsymbol{R}^{-1}=\begin{bmatrix}\boldsymbol{R}^{11} & \boldsymbol{R}^{12}\\ \boldsymbol{R}^{21} & \boldsymbol{R}^{22}\end{bmatrix}=\begin{bmatrix}\boldsymbol{E}^{-1} & -\\ - & -\end{bmatrix}$$

ここで $\boldsymbol{E}=\boldsymbol{A}-\boldsymbol{B}\boldsymbol{C}^{-1}\boldsymbol{B}'$ という対称行列の関係式を用いる[4]。$\boldsymbol{E}$ が $\boldsymbol{A}$ のみならず分割行列全体の情報に依存して定まることに注意したい。

$T=(U,X)$ であるから 2 行 2 列の $\boldsymbol{R}^{11}$ の要素は、

$$\begin{aligned}\boldsymbol{R}^{11}&=\left(\boldsymbol{R}_{11}-\boldsymbol{R}_{12}\boldsymbol{R}_{22}^{-1}\boldsymbol{R}_{21}\right)^{-1}=n\left(\boldsymbol{U}'\boldsymbol{U}-\boldsymbol{U}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{U}\right)^{-1}\\ &=n\left(\boldsymbol{U}'\prod\nolimits_{X}^{\perp}\boldsymbol{U}\right)^{-1}=\begin{bmatrix}r^{11} & r^{12}\\ r^{21} & r^{22}\end{bmatrix}\end{aligned}\qquad\cdots\cdots\cdots\cdots③$$

相関行列の逆行列を③で示した。以上で必要な準備を終える。

■ ③を偏相関の定義式と関係づける

③の４つの要素をベクトル $\boldsymbol{y},\boldsymbol{z}$ を使って書くと④になる。

$$\begin{bmatrix}r^{11} & r^{12}\\ r^{21} & r^{22}\end{bmatrix}=n\begin{bmatrix}\boldsymbol{y}'\prod\nolimits_{X}^{\perp}\boldsymbol{y} & \boldsymbol{y}'\prod\nolimits_{X}^{\perp}\boldsymbol{z}\\ \boldsymbol{z}'\prod\nolimits_{X}^{\perp}\boldsymbol{y} & \boldsymbol{z}'\prod\nolimits_{X}^{\perp}\boldsymbol{z}\end{bmatrix}^{-1}=\frac{1}{\Delta}\begin{bmatrix}\frac{1}{n}\boldsymbol{z}'\prod\nolimits_{X}^{\perp}\boldsymbol{z} & -\frac{1}{n}\boldsymbol{z}'\prod\nolimits_{X}^{\perp}\boldsymbol{y}\\ -\frac{1}{n}\boldsymbol{y}'\prod\nolimits_{X}^{\perp}\boldsymbol{z} & \frac{1}{n}\boldsymbol{y}'\prod\nolimits_{X}^{\perp}\boldsymbol{y}\end{bmatrix}$$

$$\Delta=\left\|\prod\nolimits_{X}^{\perp}\boldsymbol{y}\right\|^{2}+\left\|\prod\nolimits_{X}^{\perp}\boldsymbol{z}\right\|^{2}-\left(\boldsymbol{y}'\prod\nolimits_{X}^{\perp}\boldsymbol{z}\right)^{2}$$

……④

④では 2 次の行列の行列式Δ を用いて逆行列を示した[5]。ここで行列の三つ組みの要素からなる次の関数を再表現しよう。すると定数の n とΔは分母子の割り算でキャンセルされ、またマイナスの符号もキャンセルされるため

$$-\frac{r^{12}}{\sqrt{r^{11}}\sqrt{r^{22}}}=-\frac{-\boldsymbol{z}'\prod\nolimits_{X}^{\perp}\boldsymbol{y}}{\sqrt{\boldsymbol{z}'\prod\nolimits_{X}^{\perp}\boldsymbol{z}}\sqrt{\boldsymbol{y}'\prod\nolimits_{X}^{\perp}\boldsymbol{y}}}=\frac{\boldsymbol{y}'\prod\nolimits_{X}^{\perp}\boldsymbol{z}}{\sqrt{\boldsymbol{y}'\prod\nolimits_{X}^{\perp}\boldsymbol{y}}\sqrt{\boldsymbol{z}'\prod\nolimits_{X}^{\perp}\boldsymbol{z}}}\qquad\cdots\cdots⑤$$

この⑤の右辺が①で示した偏相関の定義に一致する。以上から⑤式左辺の変換によって偏相関が導けることが証明できた。ここで変数番号を表す 1,2

---

[4] 線形数学のテキストを参照のこと。たとえば永田（2005）152 頁にも記述がある。

[5] $\boldsymbol{A}=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$ とすれば $\boldsymbol{A}^{-1}=\dfrac{1}{ad-bc}\begin{bmatrix}d & -c\\ -b & a\end{bmatrix}$

は、任意の $j,k=1,2,\cdots,p \quad (j\neq k)$ に置き換えて構わないので、データ行列のどの変数の組み合わせについても同じ変換式で偏相関が求められる。

A.2　無向独立グラフ

■ 行列計算

⑤の計算式に従って相関行列 $\boldsymbol{R}$ から出発して偏相関行列 $\boldsymbol{Pcor}$ を求めるアルゴリズムは次の通り。

$$\boldsymbol{D}=diag\left(\boldsymbol{R}^{-1}\right)=diag\left(d_{jj}\right)$$

$$\boldsymbol{D}^{-\frac{1}{2}}=diag\left(\frac{1}{\sqrt{d_{jj}}}\right)$$

$$\boldsymbol{Pcor}=-\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{R}^{-1}\boldsymbol{D}^{-\frac{1}{2}} \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots⑥$$

$diag(\ )$ は行列からその主対角要素からなる対角行列を作る演算子である。

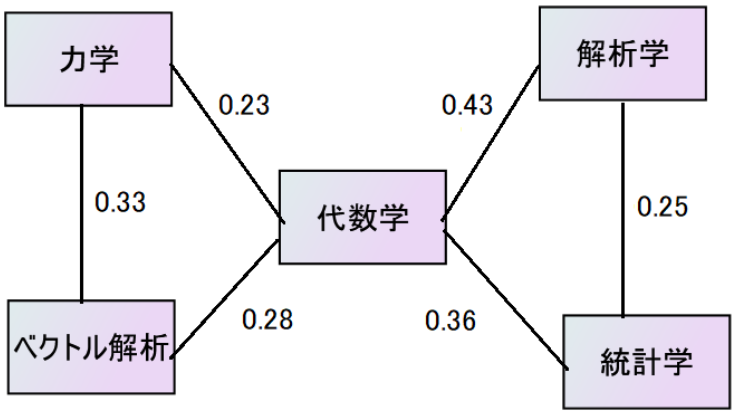
Mardia ら(1979)の数学系5教科の成績データを用いて⑥式の計算を例示しよう。科目名は配列順に X1 力学、X2 ベクトル解析、X3 代数、X4 解析、X5 統計であり n=88 であった。

$$\boldsymbol{R}=\begin{bmatrix} 1 & & & & \\ 0.553 & 1 & & & \\ 0.547 & 0.610 & 1 & & \\ 0.409 & 0.485 & 0.711 & 1 & \\ 0.389 & 0.436 & 0.665 & 0.607 & 1 \end{bmatrix},\quad \boldsymbol{R}^{-1}=\begin{bmatrix} 1.604 & & & & \\ -0.560 & 1.802 & & & \\ -0.509 & -0.658 & 3.043 & & \\ 0.003 & -0.155 & -1.112 & 2.178 & \\ -0.043 & -0.037 & -0.863 & -0.517 & 1.921 \end{bmatrix}$$

$$\boldsymbol{D}^{-\frac{1}{2}}=\begin{bmatrix} 0.79 & & & & \\ & 0.745 & & 0 & \\ & & 0.573 & & \\ & 0 & & 0.678 & \\ & & & & 0.722 \end{bmatrix},\quad \boldsymbol{Pcor}=\begin{bmatrix} -1 & & & & \\ 0.329 & -1 & & & \\ 0.230 & 0.281 & -1 & & \\ -0.002 & 0.078 & 0.432 & -1 & \\ 0.025 & 0.020 & 0.357 & 0.253 & -1 \end{bmatrix}$$

相関行列 $\boldsymbol{R}$ 、その逆行列 $\boldsymbol{R}^{-1}$ および偏相関行列 $\boldsymbol{Pcor}$ はいずれも対称行列なので上三角行列の要素の記述を省いた。

図 2 には偏相関係数を利用してアークを減らした無向独立グラフを示す。この図は日本品質管理学会(1999)の 88 頁に準拠したものである。ただし同書では偏回帰係数の値がベクトル解析と代数学で 0.33、代数学と解析学で 0.45, 解析学と統計学



で 0.26 となっていた。計算の誤りあるいは誤植と考えられるので、ここでは数値を修正している。偏相関係数が小さい場合に 0 と置き換えたので 4 通りの組み合わせについてはアークを描かなかった。

図2　数学系 5 教科の無向独立グラフと偏相関

ここで偏相関の性質を示すために、図 2 から統計学をカットしよう。X1 力学、X2 ベクトル解析、X3 代数、X4 解析の偏相関は次の通りになる。

$$\boldsymbol{Pcor}4 = \begin{bmatrix} -1 & & & \\ 0.330 & -1 & & \\ 0.256 & 0.308 & -1 & \\ 0.005 & 0.086 & 0.578 & -1 \end{bmatrix}$$

この 4 変数で測った偏相関は 5 変数の ***Pcor*** とは異なる。相関係数ならば変数群の構成を追加・削除しても、削除されなかった 2 変数の相関は変化しなかった。

しかし分析者が変数群をどう指定するかで偏相関係数は変わってしまう。この変数群依存性は偏相関に特有の性質を示すものであり、この本質を理解しないと偏相関の解釈を誤る。分析結果を読むだけのユーザーでも統計学の基礎を理解することは大事である。

A.3　相関と偏相関は同一の変換で交換可能である

■ 相関行列の再生

前節では相関行列 ***R*** から偏相関行列 ***Pcor*** に変換した。そのための操作は、

{逆行列⇒三つ組み(triad)を使った規準化} という 1 連の変換だった。

この変換を要約すればば図 3 の通りである。三つ組みの対象要素を太字で示した。

$$\text{逆行列}=\begin{bmatrix} 11 & 1k & \cdots & 1j & 1p \\ k1 & \boldsymbol{kk} & \cdots & kj & kp \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ j1 & \boldsymbol{jk} & \cdots & \boldsymbol{jj} & jp \\ p1 & pk & \cdots & pj & pp \end{bmatrix} \Rightarrow -\frac{jk\text{の要素}}{\sqrt{|jj\text{の要素}|}\sqrt{|kk\text{の要素}|}} \quad \cdots\cdots\cdots ⑦$$

図 3　具体的な変換

さて逆行列の逆行列が元の行列に戻ることから、偏相関への変換も同じ逆変換が成り立つと予想される。つまり図 4 のループである。

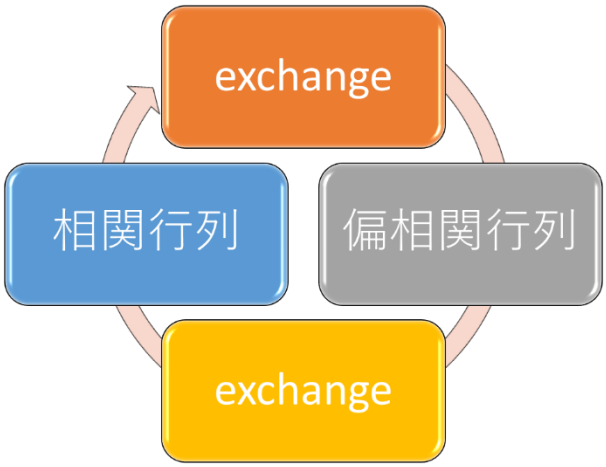


図4　交換の無限ループ

Mardia らのデータで実際に確認すると、下記の通り偏相関から相関が再生される。本稿では再生された相関行列を $\boldsymbol{R}_{rep}$ で区別した。

$$\boldsymbol{Pcol}^{-1}=\begin{bmatrix} -1.604 & & & & \\ -0.941 & -1.802 & & & \\ -1.208 & -1.428 & -3.043 & & \\ -0.765 & -0.961 & -1.83 & -2.178 & \\ -0.683 & -0.812 & -1.61 & -1.242 & -1.921 \end{bmatrix}, \boldsymbol{D}_p=\begin{bmatrix} 1.604 & & & & \\ & 1.802 & & 0 & \\ & & 3.043 & & \\ & 0 & & 2.178 & \\ & & & & 1.921 \end{bmatrix}$$

$$\boldsymbol{D}_p^{-\frac{1}{2}}=\begin{bmatrix} 0.79 & & & & \\ & 0.745 & & 0 & \\ & & 0.573 & & \\ & 0 & & 0.678 & \\ & & & & 0.722 \end{bmatrix}, \quad \boldsymbol{R}_{rep}=\begin{bmatrix} 1 & & & & \\ 0.553 & 1 & & & \\ 0.547 & 0.610 & 1 & & \\ 0.409 & 0.485 & 0.711 & 1 & \\ 0.389 & 0.436 & 0.665 & 0.607 & 1 \end{bmatrix}$$

⑦では行列計算の都合から $\boldsymbol{Pcol}^{-1}$ の主対角要素の絶対値をとった。たとえば変数1と2の変換式の分母を計算すれば規準化項は次のようになる。

$\sqrt{|-1.604|}\sqrt{|-1.802|} = 1.2665 \times 1.3424 = 1.70$。もし 2 項をまとめて計算するならば $\sqrt{(-1.604)(-1.802)} = \sqrt{2.890} = 1.70$ であり、この式なら絶対値をとる必要はない。また上記の $\boldsymbol{D}_p^{-\frac{1}{2}}$ から計算すれば規準化項は $(0.79 \times 0.745)^{-1} = 1.70$ である。

本項で述べた「相関と偏相関が同一の関数で交換できる」という指摘は、著者の知る範囲では新しい指摘である。

A.4　飽和モデルからの乖離を発見する

Pcor から R を再生させることに何の必要があるのかが疑問になるかもしれないので、ここで説明しよう。グラフィカルモデリングでは偏相関の値が小さいアークをカットする処理を行う。偏相関を 0 に修正した影響で相関も変化する。そこで元の $\boldsymbol{R}$ と $\boldsymbol{R}_{rep}$ の乖離を比べれば、どの変数ペアで大きな乖離が生じたかを調べることができる。

朝野ら（2005，122 頁、表 7.2）は、独立グラフが飽和モデルと適合しているかをモデル全体で評価する適合度指標を整理した[6]。モデル全体の評価はそれでよいが、個々の相関関係で大きく飽和モデルから逸脱したのはどの変数の組なのかを知るには、相関どうしで比較しなければならない。

実際にデータで確認してみよう。図 2 の無向独立グラフでは 5 教科から 2 教科のペアをつくる 10 組のペアのうち 4 つのペアについてアークをカットしていた。

[6] 全変数をパスでつないだモデルを「飽和モデル」と呼ぶ。観測データへの適合はよいが、できるだけシンプルなモデルで現象の本質を表現しようという「オッカムの剃刀」の趣旨からすれば望ましくないモデルである。

$$\boldsymbol{Pcor} = \begin{bmatrix} -1 & & & & \\ 0.329 & -1 & & & \\ 0.230 & 0.281 & -1 & & \\ -0.002 & 0.078 & 0.432 & -1 & \\ 0.025 & 0.020 & 0.357 & 0.253 & -1 \end{bmatrix} \quad \boldsymbol{Pcor}^{*} = \begin{bmatrix} -1 & & & & \\ 0.329 & -1 & & & \\ 0.230 & 0.281 & -1 & & \\ 0 & 0 & 0.432 & -1 & \\ 0 & 0 & 0.357 & 0.253 & -1 \end{bmatrix}$$

$$\boldsymbol{R}_{rep} = \begin{bmatrix} 1 & & & & \\ 0.497 & 1 & & & \\ 0.482 & 0.520 & 1 & & \\ 0.316 & 0.341 & 0.656 & 1 & \\ 0.296 & 0.318 & 0.613 & 0.553 & 1 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 1 & & & & \\ 0.553 & 1 & & & \\ 0.547 & 0.610 & 1 & & \\ 0.409 & 0.485 & 0.711 & 1 & \\ 0.389 & 0.436 & 0.665 & 0.607 & 1 \end{bmatrix}$$

このことは $\boldsymbol{P}^{*}_{col}$ で囲んだ４つの偏相関を０に置き換えたことと等しい。この行列から相関係数を再生すると原データの相関係数とは値が異なってくる。$\boldsymbol{R}_{rep}$ と $\boldsymbol{R}$ を比べることでモデルと観測データの乖離を知ることができる。

$\boldsymbol{R}_{rep}$ に対応したデータ行列は存在しないのだから、偏相関を相関に変換する関数が必要になる。

■ 交換関数の実装

```
# R のコード
# 相関行列Rはすでに出来ているとして
# exchange()の引数に一方の行列を入力すれば他方の行列に変換できる

exchange <- function(X){
   Xinv <- solve(X)                          #逆行列
   D <- diag(diag(abs(Xinv)))                #主対角要素を正に
   Dsq <- solve(sqrt(D))                     #規準化処理
   return ( (-1)*Dsq %*% Xinv %*% Dsq )
}
# 使用例
Pcor <- exchange(R)                  #相関行列 R を偏相関化
Rrep <- exchange(Pcor)               #この Rrep と R を比較する

# Python のコード
import pandas
import numpy as np
```

```
df = pandas.read_csv('math5test.csv', sep=',')
X = df.to_numpy()
R = np.corrcoef(X.T)  # NumPy で相関行列 R を作成する

def exchange (X):
     Xinv = np.linalg.inv(X)
     D = np.diag(np.diag(np.abs(Xinv)))
     Dsq = np.linalg.inv(np.sqrt(D))
     return ( -1*(Dsq @ Xinv @ Dsq))

# 使用例
Pcor = exchange(R)
Rrep = exchange(Pcor)
```

引用文献

日本語文献(五十音順)

1. 朝野熙彦・鈴木督久・小島隆矢(2005)「入門共分散構造分析の実際」講談社
2. 朝野熙彦編著(2025)「データから戦略を導く理論と実践入門」東京図書
3. 稲垣宣生(2003)「数理統計学改訂版」裳華房
4. 岩崎学(2015)「統計的因果推論」朝倉書店
5. 小島隆矢(2008) 入門 構造方程式モデリングの実際,日本行動計量学会第 11 回春の合宿セミナー講演資料集,19-59.
6. 高橋将宜(2022)「統計的因果推論の理論と実装」共立出版
7. 永田靖 (2005)「統計学のための数学入門 30 講」朝倉書店
8. 日本品質管理学会編(1999)「グラフィカルモデリングの実際」日科技連出版
9. 星野崇宏(2009)「調査観察データの統計科学」岩波書店

欧文文献(アルファベット順)

1. Abadie, Alberto (2005) Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, **72**(1), 1-19.
2. Blalock, H.M., Jr. (1961) "*Causal Inferences in Nonexperimental Research*". Chapel Hill, Univ. North Carolina Press.
3. Chernozhukov, V., et al. (2018) Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal,* **21**(1), 1-68.
4. Dempster, A.P. (1972) Covariance selection. *Biometrics*, **28**,157-175.
5. Ellickson, P. B., Kar,W. and Reeder, J. C. III. (2023) Estimating marketing component effects: Double machine learning from targeted digital promotions. *Marketing Science,* **42**(4), 704-728.
6. Gordon, B. R., Moakler, R., and Zettelmeyer, F. (2022) Close enough: A large-scale exploration of non-experimental approaches to advertising measurement. arXiv preprint arXiv, 220107055.
7. Guenther, P., Guenther, M., Misra, S., Koval, M., and Zaefarian, G. (2025) Propensity score modeling for business marketing research. *Industrial Marketing Management*, **127**, 14-28.
8. Künzel. S.R., Sekhon. J.S., Bickel. P.J., and Yu. B. (2019) Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4156-4165.
9. Langen, H., and Huber, M. (2023) How causal machine learning can leverage marketing strategies: Assessing and improving the performance of a coupon campaign. *Plos one*, **18**(1), e0278937.

10. Li, X., Zhang, S., and Song, X. (2024) The impact of internet use and involvement on residents' attitudes to healthcare in China: A propensity score matching analysis. *Plos one*, **19**(8), e0305664.
11. Marco Scutari and Jean-Baptiste Denis (2022) "*Bayesian Networks: With Examples in R Second Edition*") (邦訳:『Rと事例で学ぶベイジアンネットワーク』(金明哲監訳、財津亘訳、共立出版、2022年))
12. Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) "*Multivariate Analysis*". Academic Press.
13. Moodie EEM, Saarela O. and Stephens DA. (2018) A doubly robust weighting estimator of the average treatment effect on the treated. *Stat*, **7**(1)
14. Pearl, J.(1995) Causal diagrams for empirical research. *Biometrika*, **82**, 669-709.
15. Rosenbaum, P.R. and Rubin, D.R.(1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55.
16. Rubin D.B. (1980) Bias reduction using Maharanobis-metric matching. *Biometrics*, **36**, 293-298.
17. Sant'Anna, Pedro H.C. and Zhao (2020) Doubly robust difference-in-differences estimators. *Journal of Econometrics*, **219**(1), 101-122.
18. Wright, S (1934) The method of path coefficients. *Annals of Math. Stat.*, **5**,161-215.
19. Zeisel,H. (1957) "*Say it with Figures*". New York: Harper & Brothers.
20. Zheng, Z., and Liu, C. (2024) Bootstrap matching: A robust and efficient correction for non-random A/B test, and its applications. arXiv preprint arXiv:2408.05297.

【研究会の実施概容】
産学協同研究会のメンバーは次の通りである。

研究代表者　朝野熙彦　東京都立大学元教授
研究メンバー
奥瀬喜之　専修大学
水師　裕　国士舘大学
松波成行　国立研究開発法人物質・材料研究機構
後藤太郎　CCCMK ホールディングス株式会社
松本　健　株式会社ジンズ
石原聖子　石原事務所
河原達也　株式会社ビデオリサーチ
嶋田佑児　株式会社クロス・マーケティング
梅山貴彦　株式会社クロス・マーケティング
JMRA リサーチイノベーション委員会委員長
* 森本　修　株式会社ディー・エヌ・エー
JMRA リサーチイノベーション委員会委員
* 編集委員
研究期間：2024 年 12 月 9 日～2025 年 9 月 8 日（月 1 回）
研究会場：専修大学神田キャンパス