

A Study on Approximation Algorithm of Bayesian Optimal Estimation for Average Treatment Effect

INOUE Kazuma

1. 研究背景と目的

各変数間に因果関係を仮定し、統計モデルを用いて観測データから因果的効果を推定する因果分析は社会科学・工学・医学といった学術的分野に留まらず、マーケティングなどのビジネス現場でも用いられている。因果分析を行う手法の一つにランダム化実験があげられる。ランダム化実験では、立案した施策を行う処理群と対照群の2つに無作為で母集団を分割し、これらの群間の平均値の差によって介入の効果を測定し、施策の評価を行う。このとき用いられる尺度を平均介入効果 [1] といい、介入を行ったことによって変化した目的変数の確率分布の平均値で評価する。しかし、ランダム化実験の実施に莫大な費用を要する場合や倫理的な問題が伴う場合には、実験自体を行うことが難しくなってしまう。そのため観測データから平均介入効果を推定する手法のニーズが増しており、定量的に因果効果を推定する統計的因果推論がそのための手法として注目されている。こうした分析を行う具体的な手法として傾向スコア法やパス解析といった方法が挙げられ、様々な分野で応用事例が報告されている [2]。こうした手法で推定した因果効果はある症状に対する治療法の評価や政策の評価などの重要な場面で役立てられることが考えられるため、介入効果を一層正しく推定する方法が求められている。

以上の背景のもと、堀井ら [3] は、介入効果を推定する問題を統計的決定理論 [4] の枠組みを用いて定式化を行い、変数間の因果関係を表現した因果ダイアグラムが既知の場合と一意に定まらない場合のそれぞれについてベイズ最適な介入効果の推定量を導いている。また、ベイズ基準のもとで最適となる決定関数を導出した。この手法では、モデルや構造方程式のパラメータを一意に定めることなくモデルの事後確率やパラメータの事後分布で重み付けを行うことで、ベイズ最適な介入効果の推定式が示されている。堀井らの従来手法は、因果ダイアグラムが既知もしくは、事前知識や観測データから候補となるモデルが分かっている場合にベイズ最適な介入効果を求めることが可能な方法である。しかし、観測データには複数の確率変数が存在し、データの生成過程が複雑と考えられる状況で因果ダイアグラムの候補を列挙するのは困難である。

そこで本研究では、予め変数間の因果関係が向きを含めて事前知識や観測データによって分かっているという仮定の下で、ベイズ最適な平均介入効果の近似推定を可能にする実用的なアルゴリズムの提案を目的とする。具体的には、全ての因果関係を考慮したフルモデルから、一部の矢線を取り除くことで得られるモデルによって構成されるモデルの集合から

無作為に任意の数だけモデルをサンプリングし、サンプリングされた少数のモデルで混合操作を取り、ベイズ最適な平均介入効果の近似計算を行うアルゴリズムを提案する。数値実験では、提案アルゴリズムによって得られる推定値とベイズ最適な平均介入効果に対する近似精度の評価を行い、結果と考察を示す。

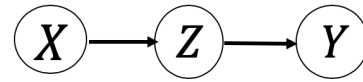


図 1: 因果ダイアグラム  $G$  の例

2. 準備

この節では、因果ダイアグラムと構造方程式の定義について述べた後、平均介入効果の定義について述べる。

2.1. 因果ダイアグラムと構造方程式

[定義 1] 非巡回的有向グラフ  $G$  とその頂点に対応する確率変数の集合  $V = \{X_1, X_2, \dots, X_P\}$  が与えられている。グラフ  $G$  が確率変数間の因果的関係を

$$X_i = g_i(\text{pa}(X_i), \epsilon_i) \quad (i = 1, \dots, P), \quad (1)$$

の形に規定し、確率変数がこの因果的関係に従って生成されるとき、このグラフを因果ダイアグラムという。また、式 (1) を  $X_1, X_2, \dots, X_P$  に対する構造方程式という。関数  $g_i$  は任意の非線形関数を表す。ここで、 $\text{pa}(X_i) \subset V$  は  $X_i$  へ矢線を持つ変数の集合である。また、誤差変数  $\epsilon_1, \dots, \epsilon_P$  は互いに独立であるとする。

2.2. 平均介入効果

因果ダイアグラムが与えられているとき、Pearl[5] はある処置変数  $X$  の値を他の変数に関係なく一定の値に固定する処理を介入と定義し、このときの別の変数  $Y$  の分布を介入効果として以下のように定義した。

[定義 2] 非巡回的有向グラフ  $G$  とその頂点に対応する確率変数の頂点集合  $V = \{X, Y, Z_1, \dots, Z_p\}$  とするとき、

$$p(y|\text{do}(X = x)) = \int \dots \int \frac{p(x, y, z_1, \dots, z_p)}{p(x|\text{pa}(x))} dz_1 \dots dz_p, \quad (2)$$

を  $X$  の  $Y$  への介入効果という。ここで、 $\text{do}(X = x)$  は介入したことによって  $X$  の値を  $x$  に固定する処理のことを表している。また、 $p(x|\text{pa}(x))$  は  $\text{pa}(x)$  を与えたときの  $X$  の条件付き分布である。

式 (2) は因果ダイアグラムが一意に定まり、各確率変数間の確率分布が推定できて初めて計算可能なものになっている。

ここで、堀井らの研究では因果ダイアグラムを表す変数を  $m$  とおき、確率分布がパラメータ  $\theta_m$  により規定されるパラメトリックな分布で表現されるとした。  $y$  が連続値であるときの平均介入効果 [6] は、式 (2) を用いると以下ようになる。

$$\bar{y}_x(m, \theta_m) = \int y \cdot p(y|\text{do}(X = x), m, \theta_m) dy. \quad (3)$$

[例 1] 確率変数の集合  $V = \{X, Y, Z\}$  の因果ダイアグラムが図 1 のように与えられ、そのときの構造方程式が線形構造方程式であるとする、

$$Z = \theta_{Z|X}X + \epsilon_z, \quad \epsilon_z \sim \mathcal{N}(0, 1^2), \quad (4)$$

$$Y = \theta_{Y|Z}Z + \epsilon_y, \quad \epsilon_y \sim \mathcal{N}(0, 1^2), \quad (5)$$

で与えられる。このとき、 $\theta_m = (\theta_{Z|X}, \theta_{Y|Z})$  である。ここで  $\mathcal{N}(\mu, \sigma^2)$  は平均  $\mu$ 、分散が  $\sigma^2$  の正規分布を表すとする。このモデルの平均介入効果は、

$$\bar{y}_x(m = G, \theta_{m=G}) = \theta_{Y|Z}\theta_{Z|X}x, \quad (6)$$

で与えられる。

### 3. 決定理論に基づく介入効果の推定

この節では堀井らが提案した決定理論に基づいて介入効果を推定する手法について説明する。特に、平均介入効果に対してベイズ基準のもとで最適となる決定関数を導出し、具体例を用いて説明を行う。まず、決定理論によって介入効果を推定する問題の定式化とベイズ最適な平均介入効果の導出 [7] を行う。最初に従来研究の仮定について説明する。因果ダイアグラムのモデルを表すパラメータ  $m$  は未知であり、そのもとで定まる  $\theta_m$  も未知である。そのため、これらを得られたデータから推定する必要がある。各変数  $(X, Y, Z_1, \dots, Z_p)$  のサンプルとして、 $D^N = \{(x_n, y_n, z_{1n}, \dots, z_{pn})\}_{n=1, \dots, N}$  が与えられているとき、決定関数  $Ay(x, D^N)$  は  $D^N$  を入力として、介入効果の推定値を指す。決定関数の誤差を測るために損失関数を設定する必要があり、平均介入効果を求める際には、二乗誤差損失を採用する。このときの損失関数を以下に示す。

$$\begin{aligned} \text{loss}(m, \theta_m, Ay(x, D^N)) = \\ (\bar{y}_x(m, \theta_m) - Ay(x, D^N))^2. \end{aligned} \quad (7)$$

リスク関数は式 (7) の損失関数を  $D^N$  に関して期待値をとって定義されるため、

$$\begin{aligned} \text{Risk}(m, \theta_m, Ay(x)) = \\ E_{D^N|\theta_m}[\text{loss}(m, \theta_m, Ay(x, D^N))], \end{aligned} \quad (8)$$

リスク関数はパラメータ  $m, \theta_m$  の関数となる。しかし、任意のパラメータに関してリスク関数を最小化する決定関数は存在しない。堀井らの研究では因果ダイアグラムの候補集合  $\mathcal{M}$  が与えられたもとで  $m \in \mathcal{M}$  に対して事前分布  $p(m)$  を

因果ダイアグラム  $m$  のもとで定まるパラメータ  $\theta_m$  の事前分布として  $p(\theta_m)$  を仮定できるとし、ベイズリスク関数を考える。

$$BR(Ay(x)) = E_m[E_{\theta_m|m}[\text{Risk}(m, \theta_m, Ay(x))]]. \quad (9)$$

このとき、以下の定理が成り立つ。

[定理 1] 式 (9) を最小化するベイズ最適な決定関数は以下で与えられる。

$$Ay^*(x, D^N) = E[y|\text{do}(X = x), D^N]. \quad (10)$$

ただし、

$$\begin{aligned} E[y|\text{do}(X = x), D^N] = \\ \sum_{m \in \mathcal{M}} p(m|D^N)E[y|\text{do}(X = x), m, D^N], \end{aligned} \quad (11)$$

とおいた。ここで、

$$\begin{aligned} E[y|\text{do}(X = x), m, D^N] = \\ \int \bar{y}_x(m, \theta_m)p(\theta_m|m, D^N)d\theta_m, \end{aligned} \quad (12)$$

となり、式 (11) を計算することによって因果ダイアグラムが一意に定まらない場合にベイズ最適な平均介入効果を推定することが可能になる。

[例 2] 確率変数の集合  $V = \{X, Y, Z\}$  の因果ダイアグラムが図 1 のように与えられ、その時の構造方程式が式 (4), (5) で与えられるとする。各パラメータ  $\theta_{Z|X}, \theta_{Y|Z}$  の事前分布として  $\theta_{Z|X}, \theta_{Y|Z} \sim \mathcal{N}(0, \alpha^{-1})$  を仮定する。このときのベイズ最適な平均介入効果は、以下のように与えられる。

$$\begin{aligned} E[y|\text{do}(X = x), m = G, D^N] = \\ \int \int \theta_{Y|Z}\theta_{Z|X}x \cdot N(\theta_{Y|Z}; \mu_{Y|Z}, s_{Y|Z}^{-1}) \times \\ N(\theta_{Z|X}; \mu_{Z|X}, s_{Z|X}^{-1})d\theta_{Y|Z}d\theta_{Z|X}, \end{aligned} \quad (13)$$

$$\mu_{Y|Z} = s_{Y|Z}^{-1}\mathbf{z}^T\mathbf{y}, \quad (14)$$

$$s_{Y|Z} = \alpha + \mathbf{z}^T\mathbf{z}, \quad (15)$$

$$\mu_{Z|X} = s_{Z|X}^{-1}\mathbf{x}^T\mathbf{z}, \quad (16)$$

$$s_{Z|X} = \alpha + \mathbf{x}^T\mathbf{x}. \quad (17)$$

ここで、 $\mathbf{x} = (x_1, \dots, x_N)^T, \mathbf{y} = (y_1, \dots, y_N)^T, \mathbf{z} = (z_1, \dots, z_N)^T$  とした。

### 4. 提案アルゴリズム

この節では提案アルゴリズムについて説明する。最初に本研究の仮定について説明する。本研究では、因果ダイアグラムが未知であり、因果ダイアグラムに基づいて定まるパラメータも同様に未知である。その一方で、予め専門的な事象に対

する知識や観測データから各変数間の  $l$  個の因果関係は向きを含めて既知であるが、どの変数間の因果関係が生じているかは分からないとする。このとき、 $l$  個の因果関係が全て存在していると見なしたモデルをフルモデルと定義する。フルモデルから一部の因果関係を取り除いて考えられる  $2^l$  個のモデルをそれぞれ  $m_1, m_2, \dots, m_{2^l}$  と記述する。これらのモデルの集合を  $\mathcal{M} = \{m_1, m_2, \dots, m_{2^l}\}$  と表すと、モデルの集合  $\mathcal{M}$  の中には正しい因果関係だけを取り入れた真のモデルが含まれている。各モデル  $m_i \in \mathcal{M}$  に対して事前分布  $p(m_i)$  を仮定する。また、因果ダイアグラム  $m_i$  が定まったもとでパラメータ  $\theta_{m_i}$  が定まるものとし、各パラメータ  $\theta_{m_i}$  に対して事前分布  $p(\theta_{m_i})$  が仮定できるとする。

#### 4.1. 提案へのアプローチ

このような状況において、ベイズ最適な平均介入効果を求めるには  $|\mathcal{M}| = 2^l$  個の各モデル  $m_i \in \mathcal{M}$  に対して混合操作をとる必要が生じる。このとき、因果関係の数が増加するにつれて指数関数的にモデルの候補は増加するため従来手法の適用は困難である。また、各モデルの平均介入効果を求める際にはパラメータの事後分布で重み付けを行うため数値積分を要し、パラメータの数が増大するにつれて更に計算が困難になってしまう。そのため、簡単な計算で平均介入効果の推定をしながら、全てのモデルを混合せずにベイズ最適な平均介入効果の近似が可能になる実用的な方法が求められている。そこで本研究では、事後分布での重み付けをする代わりに最大事後確率推定 (MAP 推定)[8] で求めたパラメータで重み付けを行い、計算量を削減する方法を提案する。また、少数の因果ダイアグラムをサンプリングし、これらに対して混合操作を取り、ベイズ最適な介入効果を近似するアルゴリズムを提案する。

#### 4.2. MAP 推定値による平均介入効果の近似

ここでは、パラメータの事後分布による重み付けをする代わりに MAP 推定値で重み付けし、近似する方法について説明する。式 (12) ではパラメータの事後分布で重み付けを行うため数値積分を必要とし、パラメータ数が増えるにつれて計算が困難になる。そこで、事後分布の最頻値で与えられる MAP 推定値で重み付けを行うことを考える。以下に MAP 推定値を与える式を示す。

$$\theta_m^{MAP} = \arg \max_{\theta_m} p(D^N | m, \theta_m) p(\theta_m). \quad (18)$$

MAP 推定値による重み付けを行うときの平均介入効果は、

$$E[y | \text{do}(X = x), m, D^N] \approx \bar{y}_x(m, \theta_m^{MAP}) p(\theta_m^{MAP} | m, D^N), \quad (19)$$

のように与えられる。このように計算を行うことで数値積分が省略され、パラメータの重み付けに要する計算量が軽減さ

れる。ここで、MAP 推定値で近似した平均介入効果がベイズ最適な平均介入効果を近似できているか検証するために式 (12) と式 (19) で推定した平均介入効果を比較する実験を行った。予め任意の因果ダイアグラムを設定し、そのモデルに対応する構造方程式にパラメータを与えたもとでデータを生成した。生成したデータを用いて平均介入効果の推定を行ったところ、両者の誤差は  $10^{-9}$  程度の非常に小さい差であった。そのためパラメータの事後分布で重み付け積分を計算する代わりに MAP 推定値で重み付けを行っても推定精度への影響は非常に小さいと見なし、以降の計算では、式 (19) を用いて各モデルの平均介入効果の近似計算を行う。

#### 4.3. 提案近似アルゴリズム

次に、ベイズ最適な平均介入効果の近似を行うアルゴリズムについて説明する。従来手法を用いてベイズ最適な平均介入効果を推定するには  $|\mathcal{M}| = 2^l$  個のモデル全てを混合して推定を行う必要があり、因果関係の数が増えるにつれて従来手法の適用は困難になる。そこで、本研究では、全てのモデルではなく、モデル  $m_i \in \mathcal{M}$  を無作為に任意の数だけサンプリングして、サンプリングした少数のモデルを混合し、ベイズ最適な平均介入効果を近似することを考える。以下に提案近似アルゴリズムを示す。

Step1) 各モデル  $m_i \in \mathcal{M}$  を任意の個数だけ無作為にサンプリングする

Step2) サンプリングされたモデルを用いて式 (11) に従って平均介入効果の計算を行う

以上のようにサンプリングした少数のモデルを混合する操作を取ることにより、ベイズ最適な平均介入効果の近似を行う。これらの方法を用いることによって、従来手法に比べて手軽な計算量でベイズ最適な平均介入効果の近似推定が可能になると考えられる。次の節では提案アルゴリズムを用いてベイズ最適な平均介入効果の近似精度を検証する実験を行う。

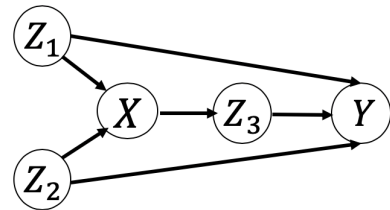


図 2: 数値実験で扱う因果ダイアグラム

### 5. 数値実験

この節では、提案アルゴリズムで与えられる推定値のベイズ最適な平均介入効果に対する近似精度を検証するために評価実験を行う。

#### 5.1. 実験条件

本研究では図 2 の因果ダイアグラムを用いて実験を行う。このとき、予め分かっている各変数間の因果関係の数  $l = 6$  である。因果ダイアグラムの候補となるモデ

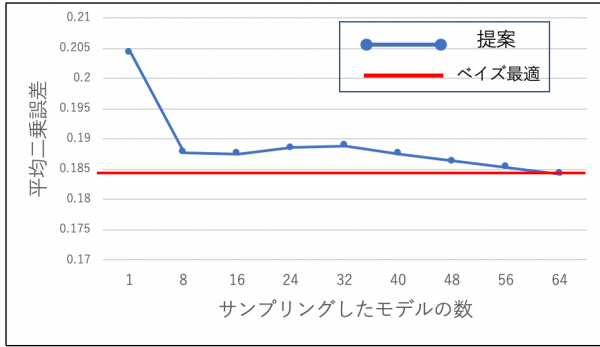


図 3: サンプル数に対する MSE の平均値推移

表 1: 真の因果ダイアグラムにおける  $l$  の数ごとの MSE

因果関係の数 $l$	MSE
0	0.0001
1	0.0004
2	0.0412
3	0.1408
4	0.3069
5	0.5408
6	0.7588

ルの集合は  $\mathcal{M} = \{m_1, m_2, \dots, m_{2^6}\}$  となる. 各モデル  $m_i \in \mathcal{M}$  の事前分布は  $p(m_1) = p(m_2) = \dots = p(m_{2^6}) = 1/2^6$  と設定した. 各モデルのパラメータの事前分布  $p(\theta_{m_1}), p(\theta_{m_2}), \dots, p(\theta_{m_{2^6}})$  は  $\mathcal{N}(0, 1^2)$  に従うものとする. データを生成するにあたって, 確率変数  $Z_1, Z_2$  は  $\mathcal{N}(0, 1^2)$  の分布からデータを生成し, それぞれサンプルサイズは  $N = 50$  とした.

### 5.2. 実験手順

まず, ランダムに真のモデル  $m_i \in \mathcal{M}$  をサンプリングする. 次に, このモデルに対応する各構造方程式に真のパラメータを設定し, データの生成を行う. 生成された各変数のサンプル  $D^N$  を用いて提案アルゴリズムによりパラメータを推定する. 予め設定した真の因果ダイアグラムから定められる平均介入効果と提案アルゴリズムによって推定した平均介入効果の二乗誤差を計測する. この手順を単一のモデルから  $|\mathcal{M}| = 2^6$  個のモデルがサンプリングされるまで行う. 以上の実験を各サンプル数に対して 100,000 回ずつ繰り返し行い, その都度, 真の因果ダイアグラムのサンプリングも行う. 最後に平均二乗誤差 (MSE) によって評価を与える. その結果を図 3 に示す.

### 5.3. 結果と考察

図 3 と表 1 を元に実験結果の分析と考察を行う. 図 3 より, 8 個のモデルをサンプリングしたところで MSE が大幅に改善し, 下限であるベイズ最適な平均介入効果の値に近い推定ができていたことを確認した. それ以降はサンプリング数が増加するにつれて少しずつ誤差が小さくなっていくことが伺える. 表 3 より, 因果関係の数が少ない場合には MSE

が小さい一方で, 因果関係の数が増加すると MSE は徐々に大きくなる傾向がある. このことから, 真の因果ダイアグラムに存在する因果関係が少数の場合と多数の場合これら全てを考慮したことにより, 全体的な推定精度が良好になっていることがわかる.

## 6. まとめと今後の課題

本研究では, ベイズ最適な平均介入効果を求める際に必要となる数値積分の計算を回避するためにベイズ最適な平均介入効果は構造方程式のパラメータを MAP 推定した値で重み付けを行うことで近似できることを示した. また, 各モデル  $m_i \in \mathcal{M}$  を任意の個数だけサンプリングし, これらに対して混合操作を行い, ベイズ最適な平均介入効果の近似を行うアルゴリズムを提案した. 実験の結果,  $l = 6$  の場合にはモデル候補の総数の約 8 分の 1 をサンプリングして混合するだけで, ベイズ最適な平均介入効果のある程度近似できることが分かった. 提案アルゴリズムにより, 推定精度を大きく劣化させずに計算量を低減させることが可能になる.

今後の課題としては, より多くの変数や因果関係を考慮した場合に同様の実験を行い, 提案アルゴリズムによってベイズ最適な介入効果の近似精度の検証やモデルの総数に対してどれだけの割合のモデル数が必要か考察することが考えられる. また, 実際に介入を行ったデータに対して提案アルゴリズムを適用し, 平均介入効果の推定を行い, 手法の推定精度や計算時間の観点から実用性を検討するといったことが考えられる.

## 参考文献

- [1] Holland, Paul W. "Statistics and causal inference." Journal of the American statistical Association, Vol.81.No.396,pp.945–960,1986.
- [2] 星野崇宏, 調査観察データの統計科学 因果推論・選択バイアス・データ融合., 岩波書店,2009.
- [3] 堀井俊佑, 須子統太, "統計的決定理論に基づいた因果効果の推定法に関する一考察", 信学技法, IBISML2018-97, pp. 397–402, 2018.
- [4] Berger, James O., "Statistical decision theory and Bayesian analysis.", Springer Science Business Media, 2013.
- [5] Pearl,Judea, "Graphs, causality, and structural equation models." Sociological Methods Research, Vol.27.2, pp.226–284, 1998.
- [6] 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, 2004.
- [7] Horii,Shunsuke, "A Note on the Bayes Optimal Estimator of the Expected Intervention Effect", Proceedings of Workshop AEW11: Concepts in Information Theory and Communications,978–90–74249–30–0,2019.
- [8] Bishop, Christopher M., パターン認識と機械学習 上 ベイズ理論による統計的予測., 丸善出版, 2008.