

因果性の学習と推論における因果ベイズネットについて

斎藤 元幸

Causal knowledge enables us to explain past events, to control present environment, and to predict future outcomes. Over the last decade, causal Bayes nets have been recognized as a normative framework for causality and used as a psychological model to account for human causal learning and inference. This article provides an introduction to causal Bayes nets. According to causal Bayes nets, causal inference can be divided into three processes: (a) learning the structure of the causal network, (b) learning the strength of the causal relations, and (c) inferring the effect from the cause or the cause from the effect. For each process, I describe the predictions of causal Bayes nets, review experimental results, and suggest future directions. Although there are a few exceptions (e.g., Markov violation), most of the results are consistent with the predictions of causal Bayes nets. The current problems of the Bayesian approach and its future perspective are discussed.

Keywords: causal reasoning (因果推論), Bayesian modeling (ベイズモデリング), causal Bayes nets (因果ベイズネット), causal Markov condition (因果マルコフ条件), intervention (介入)

1. はじめに

環境内に存在する因果関係についての情報は、過去の説明および現在の制御ならびに未来の予測を可能にする手段を我々に提供する。適応的に生活するためには、因果の知識を利用して望ましいものを獲得し、望ましくないものを避けなければならない。因果関係をどのように認識するかという問題は、HumeやKantを初めとする数多くの哲学者の論考を背景に、心理学においても様々なアプローチから解決が試みられてきた(Shanks, 2007 for a review). 近年、新たなアプローチとして因果ベイズネット(causal Bayes nets)を用いて因果性の学習と推論を説明する試みがなされている(Holyoak & Cheng, 2011; Sloman & Lagnado, 2015 for reviews). 因果ベイズネットとは、既存の情報から因果関係を適

切に導き出し、与えられた情報から推測を行うために哲学や統計学や計算機科学といった分野で開発された統計的因果推論の手法である(Pearl, 2000). 因果ベイズネットの根幹には、Thomas Bayesが発見したベイズの定理(Bayes, 1763)に基づく計算がある。

データ d を生成する仮説 h について考える事態において、データを得る前の仮説についての確信の度合いを $P(h)$ という主観確率で表す。ベイズの定理はデータを得た後に仮説に対する確信の度合いをどのように変化させるべきかを定めるものであり、以下の式で表現される。

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)} \quad (1)$$

ここで $P(d|h)$ は仮説 h のもとでデータ d が発生する確率を表すが、データが発生した後では仮説の尤もらしさを表すため、仮説 h についての尤度と呼ばれる。また、分母の $P(d)$ は考えられるすべて

の仮説のもとでデータが得られる確率を表す。この数値は分子の計算可能なすべての組合せについて、合計が1になるように正規化する項であると考えられることから、正規化定数と呼ばれる。ベイズの定理はデータを得る前の仮説の確からしさである事前確率 $P(h)$ から、データを得た後の仮説の確からしさである事後確率 $P(h|d)$ への確信の度合いの変化を記述する。結果であるデータに基づき、それを生成する原因である仮説を推測していることから、逆確率の定理と言われることもある。

ベイズの定理は、事前の知識を新たに獲得したデータに基づいて更新するという学習過程を表現したものと解釈可能であり、その汎用性の高さから様々な場面に応用されている。母数の推定や仮説の比較にベイズの定理を用いることはベイズ統計と呼ばれ、従来の伝統的な統計学では扱うことができなかった問題に解決手段を提供している (cf. 豊田, 2015)。ベイジアンネットワークとは、変数間の関係性をグラフと確率で表現し、グラフ構造の学習や確率推論を行う数理的手法であり、機械学習などの領域で盛んに研究されている (Pearl, 1985; cf. 植野, 2013)。また、因果関係に基づいてベイジアンネットワークを構築した場合は、因果ベイズネットと言われる (Glymour, 2001)。ベイズ的手法を駆使して認知モデルを作成することはベイズモデリング¹⁾と呼ばれ、条件づけ (Courville, Daw, & Touretzky, 2006)、視知覚 (Kersten, Mamassian, & Yuille, 2004; Knill & Richards, 1996)、感覚運動学習 (Körding & Wolpert, 2004)、語彙学習 (Xu & Tenenbaum, 2007)、言語処理 (Chater & Manning, 2006)、意味表象 (Griffiths, Steyvers, & Tenenbaum, 2007)、帰納推論 (Kemp & Tenenbaum, 2009)、因果推論 (Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004; Griffiths & Tenenbaum, 2009) など様々な領域に適用されている²⁾ (see Tenenbaum, Kemp, Griffiths, & Goodman, 2011 for a review)。

本論文では、因果ベイズネットと関連する実験結

果を概観しながら、その心理学的妥当性について考察を進める。はじめに、因果ベイズネットとその哲学的背景について説明する。因果性の学習と推論を因果ベイズネットの観点から考えた場合、何が原因で何が結果であるか、原因は結果にどのくらい影響するか、原因や結果の状態からどのような推測が可能かという3つの問題に分けられる。それぞれの問題に対して因果ベイズネットがどのような予測を行うか示し、予測と一致した実験結果が得られているか概観し、今後どのような研究が必要とされるか考察する。おわりに、ベイズ的アプローチに対する批判を踏まえ、今後の展望について述べる。

2. 因果ベイズネット

2.1 因果ベイズネットとは

因果ベイズネットは事象間の因果関係を非循環有向グラフで、その強さを条件付き確率で表現する。グラフ内のノードは変数を、エッジ (矢印) は変数間の因果関係を表しており、非循環とはノードからエッジを辿っても元のノードに戻らないことを意味している。変数は2値変数のような離散変数の場合もあれば、連続変数の場合もあり、また、顕在変数や潜在変数のどちらでも組み込むことができる。因果関係は、決定的あるいは確率的、発生的あるいは抑制的、線形もしくは非線形といったように様々な形態をとる。

因果ベイズネットのグラフ構造は変数の同時確率分布を定義するために使用される。同時確率分布に基づき、グラフ構造が明らかな場合は変数の予測を行い、グラフ構造が明らかでない場合はその学習を行う。例えば、喫煙 (S) が歯の黄ばみ (Y) と肺がん (L) を引き起こす関係は図1左側の因果モデルで表現される。これら3つの変数の同時確率分布は連鎖公式³⁾を用いて以下のように表現される。

$$P(S, Y, L) = P(S)P(Y|S)P(L|Y, S) \quad (2)$$

また、連鎖公式においてこれらの変数は任意に並び替えることが可能であり、以下のように記述することも可能である。

$$P(Y, L, S) = P(Y)P(L|Y)P(S|L, Y) \quad (3)$$

1) ベイズ的手法を用いて統計モデルを作成することもベイズモデリングと呼ばれるが、認知モデルでは認知主体がベイズ推定を行うことが想定されている。

2) ベイズモデリングについては、*Trends in Cognitive Sciences* 誌 (Chater, Tenenbaum, & Yuille, 2006; Kousta, 2010) や *Developmental Science* 誌 (Gopnik & Tenenbaum, 2007) や *Cognition* 誌 (Xu & Griffiths, 2011) の特集号なども参照されたい。

3) 連鎖公式によると、 n 個の離散型確率変数の分布は n 個の条件付き確率の積として分解することができる (e.g., $P(X_1, X_2, \dots, X_n) = P(X_n|X_{n-1}, \dots, X_2, X_1) \dots P(X_2|X_1)P(X_1)$)。

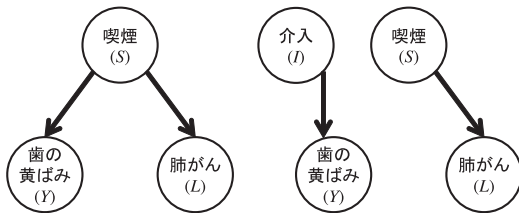


図1 喫煙 (S) と歯の黄ばみ (Y) と肺がん (L) の因果モデル (左図). 歯の黄ばみに対する介入によって喫煙から歯の黄ばみへのエッジが取り除かれた因果モデル (右図).

しかしながら, これら (2) 式や (3) 式は実際のグラフ構造を反映した形にはなっておらず, 変数の同時確率分布を一意かつ節約的に表現するために, 因果ベイズネットの核となる仮定が設けられている.

2.2 因果マルコフ条件

因果マルコフ条件 (causal Markov condition) によると, 変数はその直接の原因事象を固定した場合, その結果事象以外の変数と独立になる. 上述の例では, ある人が喫煙していると知ってしまうと, その人の歯が黄色いか否かは肺がんの予測とは無関係となる. このように, ある変数を所与として変数間に独立関係が生じることは条件付き独立と言われる. 因果マルコフ条件を設けることによって (2) 式や (3) 式は以下のように書き換えられる.

$$P(S, Y, L) = P(S)P(Y|S)P(L|S) \quad (4)$$

(2) 式における $P(L|Y, S)$ が (4) 式では $P(L|S)$ に単純化されており, 同時確率分布が変数の順序に沿った形で分解されていることが見て取れる. この例における因果マルコフ条件は喫煙と歯の黄ばみに基づく肺がんの予測と喫煙のみに基づく肺がんの予測が等しい (i.e., $P(L|Y, S) = P(L|S)$) という仮定に他ならない. 推論に関連する変数と関連しない変数を明確に区別することで計算を大幅に単純化するため, 複雑なグラフ構造における推論において因果マルコフ条件は極めて重要な役割を果たす.

因果マルコフ条件の背景には, ある変数の原因となるすべての変数がネットワークに含まれていることが仮定されている. また, 因果のない相関はないという Reichenbach (1956) の共通原因に関する仮定を満たしていることも想定されている. つまり,

変数間に相関関係が生じている時は, どちらか一方の変数が他方の変数の原因となっているか, 両方の変数を引き起こす第三の変数が存在するというわけである. これら2つの仮定が満たされるとき, 因果マルコフ条件は正当化される (Pearl, 2000).

2.3 介入

因果ベイズネットの特徴の1つとして変数を測定した場合の推論と変数に操作を加えた場合の推論を明確に区別することが挙げられる. 変数を強制的にある値に設定することは介入 (intervention) と呼ばれ, 介入によって制御された変数は本来の原因とは独立になる. このことは, 外生変数として介入の変数を加え, 介入された事象に対するエッジを取り除くことで表現され, グラフ切除 (graph surgery) や無効化 (undoing) と言われる (Pearl, 2000; Sloman, 2005)⁴⁾. 上記の例においてホワイトニングによって歯の黄ばみを除去した場合, 喫煙から歯の黄ばみへのエッジが除去され (図1右側の因果モデル), 歯の状態から喫煙の有無や肺がんの状態について予測することはできなくなる. このことは当然のように思われるかもしれないが, 観察と介入の違いを表現し, 因果関係を適切に記述できるようになったのは Pearl (2000) によって因果ベイズネットが開発されてからのことである.

3. 因果ベイズネットの哲学的背景

3.1 確率

不確実性を表現する方法として確率を用いることは多いが, その解釈には様々な立場が存在する (Hájek, 2003). 最も一般的なのは頻度主義と呼ばれる立場であり, この立場では同一条件下における無限回の試行の結果, 相対頻度が収束したものとして確率を定義する. 例えば, コインを投げた際に表が出る確率は, コインを繰り返し投げても, 何回表が出たか数えることで求められる. 一方, ベイズ主義の立場では, ある事象についての個人的な信念の度合いとして確率が定義される⁵⁾. このような確率は

4) 対象となる変数の値を固定する介入が強い介入 (strong intervention) と呼ばれるのに対して, 変数に影響を及ぼすが, その値を固定しないような操作は弱い介入 (weak intervention) と呼ばれ, この場合は他の変数からのエッジが取り除かれることはなく, 依然として本来の原因からの影響も受ける (Eberhardt & Scheines, 2007; Meder, Gerstenberg, Hagmayer, & Waldmann, 2010; Woodward, 2003).

主観確率と呼ばれ、測定には様々な方法が存在する (e.g., 戸田, 1982). 確信の程度として確率を捉えることによって、明日雨が降る確率といった未知の事象や火星に生命が存在する確率といった無知に起因する不確定な事象に対しても確率を定義することができる。

主観確率が確率法則に従うことは様々な形で示されてきた (Fishburn, 1986). 例えば, Cox (1946) は信念の度合いに関するいくつかの合理的仮定から確率法則が導出されること示している. また, de Finetti (1937) は信念の度合いに対する合理的な制限として確率の公理を提唱しており, 確率の公理が満たされない場合には不公平な賭けが生じることを指摘している (cf. 園, 2014). 例えば, コイントスの結果を予測し, 予測が的中すれば 1000 円貰える賭けに参加したとする. コインの表が出ることに對してある人が 0.6 の信念を抱いている場合, その人は 600 円までならお金を払って賭けを行うと考えられる. また同時に, コインの裏が出ることに對しても 0.6 の信念を抱いていたとする. この場合, 賭けの主催者はこれら 2 つの賭けを合わせて 1200 円で売れば, 結果の如何に関わらず必ず 200 円儲けることができる. 確率の公理が満たされていない事態では, 相手が必ず負ける賭けを仕掛けることが可能になる. 主観確率が上記のような不公平な賭けを生み出さない場合には確率の公理が満たされることが知られており, このような正当化はダッチブック論証と言われる. また, 確率の公理から不公平な賭けが導かれることはない.

ベイズ統計では主観確率の主観性が問題になることもあるが, 認知モデルにおいては主観確率を使用することは個人差を扱いやすくなる利点と考えられる. 数理モデルの多くは平均的な反応を予測の対象としているが, 最終的には個人差を含めた反応が予測対象となるだろう. 個々人の反応を予測する際には主観確率に基づいて個人ごとに異なる事前分布を想定することが必要になるかもしれない.

5) ベイズ主義の中には相対頻度に関する信念の度合いを確率と定義し, 客観的な確率解釈に留まろうとする考えもある (e.g., Neapolitan, 2004). なお, ベイズ統計における主観ベイズと客観ベイズの違いは, 確率解釈に基づくものではなく, 事前確率分布の設定方法の違いに基づくものである (Berger, 2006; Dey & Rao, 2005 繁樹他訳 2011).

3.2 因果性

確率の概念は不確実性を表現するためだけではなく, 因果性の定義⁶⁾にも使用されてきた (e.g., Eells, 1991; 大塚, 2010). Suppes (1970) は原因事象 (C) が結果事象 (E) の生起確率を上昇させるものとして因果関係を捉えた.

$$P(E|C) > P(E) \quad (5)$$

しかしながら, この定義だけでは因果関係と相関関係を区別することはできない. 例えば, 歯の黄ばみが存在している場合は, そうでない場合よりも肺がんの確率が高くなるが, 歯の黄ばみが肺がんを引き起こしているわけではない. 両者の関係は喫煙という共通原因によってもたらされる擬似相関だからである (図 1 参照). 擬似相関を排除するには, 他の要因 (F) を統制した上で (5) 式の関係が維持されるか確認する必要がある.

$$P(E|C, F) > P(E|F) \quad (6)$$

ただし, この式には時間情報が含まれておらず, 因果の方向性を決定することはできない. 例えば, 喫煙が肺がんを引き起こすか ($S \rightarrow L$), 肺がんになりやすい人が喫煙習慣を持ちやすいか ($S \leftarrow L$) 判断することはできない. また, 擬似相関の可能性を取り去るためには他の事象との因果的関連を考慮に入れなければならないが, それを判断するにはまた別の事象について考慮する必要が生じ, 無限後退に陥ってしまう.

因果ベイズネットは変数間の確率的関係を要約しているが, 確率のみによって因果性を定義することは稀であり, 変数の操作である介入と組み合わせて因果性を定義することが多い. このような立場は介入主義 (interventionism) と呼ばれている. 介入主義によれば, 変数 X への介入が変数 Y の値に影響を及ぼす時, 変数 X は変数 Y の原因である. さらに, 他の変数を一定の値に統制した時に介入が影響する場合, 変数 X は変数 Y の直接の原因であると言える (Woodward, 2003).

$$P(E|do(C=c)) > P(E|do(C=\neg c)) \quad (7)$$

事象に対して何らかの介入を行うことは $do(\cdot)$ と記述され, $do(C=c)$ は変数 C を値 c に固定することを表している. Pearl (2000) によって導入された

6) 因果性の定義に関する様々な議論については Psillos (2005) を参照されたい.

このような表記法は *do* オペレータと呼ばれ、観察や介入に基づく確率推論が定式化された。このような介入による因果性の定義は、「もし原因事象が存在しなければ結果事象は存在しなかった」といった反事実による因果性の定義 (e.g., Lewis, 1973) との整合性が認められている。

4. 因果ベイズネットによる因果構造の学習

事象間の因果関係は、その有無と方向性を表す因果構造と、関係の強さを表す因果強度に二分される。因果構造の学習では変数間の因果の方向性が未知であり、学習者は様々な手がかりに基づいて因果構造を推測するように求められる。一般的な実験事態では3つ以上の2値変数が繰り返し提示され、実験参加者はそれらの生起・非生起といった共変情報や事象の提示順序といった時間情報に基づいて因果構造を推測する。例えば、Steyvers, Tenenbaum, Wagenmakers, & Blum (2003) の実験では、心を読む3人の宇宙人の思考伝播を観察して、どのような方向性で思考伝播が生じているか判断することが実験課題として用いられている。ヒトが因果の方向性をどのように判断しているかという問題は、因果ベイズネットが認知モデルとして適用されてから検討され始めた問題であるため、どのような手がかりがどのように利用されているかについてはまだ分かっていないことが多い (Lagnado, Waldmann, Hagmayer, & Sloman, 2007)。

因果ベイズネットでは構造学習について様々なアルゴリズムが考案されている (Neapolitan, 2004; Pearl, 2000)。ここでは制約ベース法 (constraint-based method) とベイズ法 (Bayesian method) という2つの代表的な手法について述べる。制約ベース法とベイズ法は、共変関係から因果関係を導くために定常性 (stability) もしくは忠実性 (faithfulness) と呼ばれる仮定を設けている。定常性とは、任意のパラメータによって変数間の独立・従属関係が変化しないことである (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993)。このことは、変数間の独立関係は因果マルコフ条件によってのみ生じ、特定のパラメータから生じることはないことを意味している。例えば、変数 X が変数 Y に直接的だけでなく、媒介変数 Z を通して間接的にも影響を及ぼしているとする。このような関係は、 $Z = aX$ と $Y = bX + cZ$ という2つの構造方程式で記述で

きる。このような因果モデルでは、ほとんど全てのパラメータにおいて変数 X と変数 Y は従属関係になる。しかしながら、 $b = -ac$ という条件を満たすパラメータでは、直接効果と間接効果が互いを完全に相殺するため、変数 X と変数 Y は独立関係になってしまい、共変関係から正しい因果関係が導かれることは無い。このような問題は理論的には可能であっても、現実的にはほとんど生じないと考えられている。

4.1 制約ベース法

制約ベース法は事象間の共変情報を基にボトムアップで因果構造を学習するアルゴリズムであり、2変数間の独立性の検定と因果関係の方向付けという2段階の処理で構成されている (e.g., PC アルゴリズム, Spirtes et al., 1993)。PC アルゴリズムは IC アルゴリズム (Verma & Pearl, 1990) を改良して開発されたものであり、最も代表的な手法である。PC アルゴリズムによると、最初のステップでは全ての変数を無向辺で結び、無向グラフを構築する。次のステップでは2つの変数を取り出して χ^2 検定などを用いて独立性を分析し、独立関係にある2変数間のエッジを取り除く。続いて、他の変数を固定した上での条件付き独立性を調べる。因果構造が異なれば独立・従属関係も変化するため、これらの検定結果は因果構造に制約をもたらす。

方向付けを行うステップにおいても、独立・従属関係は重要となる。変数 X と変数 Y がそれぞれ変数 Z と結びついており (i.e., $X-Z-Y$)、 Z を所与として X と Y に従属関係が生じる場合、 Z が共通結果となるようにエッジを付ける (i.e., $X \rightarrow Z \leftarrow Y$)。上記以外の無向辺については、共通結果を形成しないようにエッジの方向付けを行う。例えば、変数 X が変数 Z に影響しており、変数 Y が変数 Z と結びついている場合 (i.e., $X \rightarrow Z-Y$)、 Z から Y へエッジを付ける (i.e., $X \rightarrow Z \rightarrow Y$)。制約ベース法は事象間の共変情報から因果構造を推測するため領域固有の知識などは必要としない。Gopnik et al. (2004) はいくつかの実験結果 (e.g., Gopnik, Sobel, Schulz, & Glymour, 2001) を制約ベース法で説明しているが、それらは次に述べるベイズ法を用いても説明することができる。

4.2 ベイズ法

ベイズ法 (e.g., Cooper & Herskovitz, 1992; Heckerman, Geiger, & Chickering, 1995) は共変動や他の手がかり (e.g., 時間順序, 先行知識) を利用してトップダウンで因果構造を学習するアルゴリズムであり, 候補となる因果構造に事前確率を割り当て, ベイズの定理を使用して因果構造の仮説を改訂していく. 事前確率は一様に与えられることもあるが, 先行知識に基づいて特定の因果構造の事前確率が高くなることもある. 事前確率の設定後, それぞれの因果構造において特定の生起パターンに対する尤度が算出される. 例えば, 共通結果の因果モデル (e.g., $X \rightarrow Z \leftarrow Y$) では原因が独立に生起するため, 3つの事象が共生起する確率は共通原因の因果モデル (e.g., $X \leftarrow Z \rightarrow Y$) よりも低くなる. 利用可能なデータにベイズの定理を適用して事後確率が計算され, 事後確率の最も高い因果構造がデータを生成した因果モデルとして選択される. Griffiths & Tenenbaum (2009) は, 様々な先行知識を因果モデルの仮説空間の制約として使用することによって多様な実験結果を統一的に説明することに成功している.

4.3 因果構造の学習における介入

定常性を仮定しても共変情報だけでは因果構造を特定できないことがある. 例えば, 変数 X が変数 Y と変数 Z に影響している因果モデル (i.e., $Y \leftarrow X \rightarrow Z$) の確率分布を因果マルコフ条件に基づいて分解した場合, $P(X, Y, Z) = P(X)P(Y|X)P(Z|X)$ のように記述される. また, 変数 Y が変数 X に影響し, 変数 X が変数 Z に影響している因果モデル (i.e., $Y \rightarrow X \rightarrow Z$) の場合, その確率分布は $P(X, Y, Z) = P(Y)P(X|Y)P(Z|X)$ のように記述される. ここで $P(X)P(Y|X) = P(Y)P(X|Y) = P(X, Y)$ となるため, これら2つの因果モデルの確率分布は同一となる. このような因果モデルはマルコフ等価モデル (Markov equivalent model) と呼ばれており, 観察によって得られた共変情報からは因果構造を一意に定めることができない.

介入によって変数の状態を操作することは, マルコフ等価モデルの区別を可能にする. 変数 X が変数 Y と変数 Z に影響する共通原因モデル (i.e., $Y \leftarrow X \rightarrow Z$) において X を発生させる場合, その

結果である Y と Z も生起することが予測される. 一方, 変数 Y が変数 X を媒介して変数 Z に影響する因果連鎖モデル (i.e., $Y \rightarrow X \rightarrow Z$) において X を発生させる場合, X の結果である Z は生起するが, X の原因である Y が生起することは無い. 介入を行うことによって共通原因モデルと因果連鎖モデルでは異なる共変情報が生じるようになり, マルコフ等価モデルを区別することが可能になる. このことは, 調査研究や観察研究から因果関係を実証することは困難であるが, 独立変数の操作を伴う実験研究では可能になるという研究法の違いと対応している.

介入による因果構造の学習の促進効果は多くの研究で認められている (e.g., Steyvers et al., 2003; Lagnado & Sloman, 2004). Steyvers et al. (2003) は, 観察による学習に比べて介入による学習では因果構造の判断の正答率が上昇することを示した. また Lagnado & Sloman (2004, 2006) は, 介入による因果構造の学習の促進効果には時間順序手がかりの効果が含まれており, 両者が加算的に作用することを明らかにしている. さらに, 介入によって因果構造を学習する際にはどの事象に介入するかという意思決定も重要であることが Sobel & Kushnir (2006) によって示されている. 近年では, 介入対象の選択を獲得情報量や正事例検証方略の観点から説明する試みがなされている (Coenen, Rehder, & Gureckis, 2015; Bramley, Lagnado, & Speekenbrink, 2015). 介入による促進効果は就学前児を対象とした研究においても確認されており (Burns & McCormack, 2009; but see also Kushnir, Wellman, & Gelman, 2009), 関連する様々な要因について実験的検討が行われている (e.g., Sobel & Sommerville 2009).

介入によって因果構造の学習が促進されることは因果ベイズネットの予測と一致している. しかしながら, ベイズ法と制約ベース法の両者に共通する問題は3つ以上の変数で構成される因果構造を学習する場合に計算コストが高くなることである. ベイズ法では複数の因果モデルを逐次更新することが, 制約ベース法では多くのデータを用いて複数の変数間の独立性を検討することがそれぞれ求められており, 心理学的妥当性という観点からは疑問が残る. 実際, 因果構造の判断は共変情報よりも時間情報に影響を受けることが明らかにされており (McCor-

mack, Frosch, Patrick, & Lagnado, 2015; Sloman & Lagnado, 2006), 課題が複雑になればなるほどその傾向は顕著になる (Saito & Shimazaki, 2012). このような背景のもとに近年では様々なヒューリスティックが提唱されている (e.g., Fernbach & Sloman, 2009; Lagnado & Sloman, 2004; Mayrhofer & Waldmann, in press). 計算コストという点ではヒューリスティックが優位であるが, モデルの柔軟性という点ではベイズモデルに強みがある. 一部のヒューリスティックについては事前分布に対称性への選好を組み込んだベイズモデルと一致することが報告されており (Devereett & Kemp, 2012), 両者は相互に排他的な関係とは限らない. ヒューリスティックが重視する時間情報や計算コストが共変情報とどのように統合されるかについては今後の検討課題と言える.

5. 因果ベイズネットによる因果強度の学習

因果構造が因果の方向性を表すのに対して, 因果強度は原因と結果の関係の強さを表す. 因果強度の学習では, 原因候補と結果が予め決められており, 学習者はそれらの因果関係の強さを判断することが求められる. なお, 原因候補と表現するのは, 実際には無関係である可能性も含まれているからである. 一般的な実験事態では原因候補と結果の2値2変数が使用され, これらの共変情報は図2に示すような 2×2 の分割表で表現することができる. 実験参加者はこれらの情報を継続的もしくは要約的に取得した後, 0 から 100 の間で因果強度を評価する. 例えば, Buehner, Cheng, & Clifford (2003) の実験では, 放射線被曝がウイルスの突然変異を引き起こすか調べるというカバーストーリーの下で, 被曝の有無と突然変異の有無に関する情報が与えられ, 放射線がどのくらい変異を引き起こすか判断することが求められた.

原因候補と結果の共変関係の強さは, 以下に示す ΔP で記述されることが多い (Jenkins & Ward, 1965).

$$\Delta P = P(e|c) - P(e|\neg c) \quad (8)$$

ここで $P(e|c)$ は原因候補が生起した時に結果が生起する条件付き確率を, $P(e|\neg c)$ は原因候補が生起しなかった時に結果が生起する条件付き確率を表す. ΔP は -1 から $+1$ までの値を取り, その値によ

		Event E	
		e	$\neg e$
Event C	c	a	b
	$\neg c$	c	d

図2 原因事象(C)と結果事象(E)の分割表. アルファベット a, b, c, d はそれぞれの共生起の頻度を表す.

て異なった意味を持つ. 原因候補の生起によって結果の生起確率が上昇する場合 (i.e., $\Delta P > 0$), 原因と結果の間には発生的因果関係が想定される. 一方, 原因候補の生起によって結果の生起確率が低下する場合 (i.e., $\Delta P < 0$), 原因と結果の間には抑制的因果関係が想定される. 原因候補の有無によって結果の生起確率が変化しない場合 (i.e., $\Delta P = 0$), 原因候補と結果は無関係となる.

因果強度の評定値と ΔP にはおおまかな対応関係がみられたことから (e.g., Dickinson, Shanks, & Evenden, 1984), 説明モデルとして ΔP ルールが使用されるようになった. しかし, ΔP から逸脱した結果も報告されており (e.g., Wasserman, Elek, Chatlosh, & Baker, 1993), その後, 様々な説明モデルが提唱されてきた (see Hattori & Oaksford, 2007; Perales & Shanks, 2007).

5.1 パワー PC 理論

Cheng (1997) は ΔP ルールの主要な問題点として, 以下の2点を指摘している. 原因の有無に関わらず結果が常に生起する天井効果の事態 (i.e., $P(e|c) = P(e|\neg c) = 1$) では, 実際には因果効果を判断することはできないが, ΔP ルールは効果がないと判断してしまう. また, ΔP は他の原因事象を考慮していないため, 因果効果の適切な推定値とはならない. 彼女はこれらの問題点を克服する新たな理論として, パワー PC 理論 (Cheng, 1997; Novick & Cheng, 2004) を提唱し, 共変動と因果を区別する因果パワー (causal power) を定式化した. 因果パワーとは, 他の原因が存在しない事態において原因事象が結果事象を発生させる確率である. パワー PC 理論は因果ベイズネットとは独立に提唱されたが, 特定の状況下では両者が一致するこ

とが後に指摘されている (Glymour, 2001, 2003).

パワー PC 理論では、ヒトが因果関係を学習する際に以下の4つの信念を抱いていると仮定する。(1) 背景原因 (B)⁷⁾と原因候補 (C)は独立に結果 (E)に影響する。(2) 背景原因によって結果が発生することはあるが、抑制されることはない。(3) 背景原因と原因候補の因果パワーはそれらの生起頻度とは無関係である。(4) 原因が存在しない時に結果が生じることはない。これらの仮定が満たされる場合、発生的因果パワー q_G は以下の式から算出される。

$$q_G = \frac{\Delta P}{1 - P(e|\neg c)} \quad (9)$$

$q_G = 1$ は原因が結果を必ず発生させることを意味し、 $q_G = 0$ は影響を及ぼさないことを意味する。背景原因によって結果が常に生起する、つまり天井効果が生じている場合 (i.e., $P(e|c) = P(e|\neg c) = 1$), 因果パワー q_G は定義されない。また、背景原因が存在しない場合 (i.e., $P(e|\neg c) = 0$), $q_G = \Delta P$ となり、共変動と因果は一致する。Buehner et al. (2003) は ΔP を一定に保ちつつ、 $P(e|\neg c)$ を体系的に操作し、 $\Delta P = 0$ の事態を除いて実験参加者の評定値がパワー PC 理論の予測値と一致することを報告している。

5.2 因果サポートモデル

Griffiths & Tenenbaum (2005, 2009) は、これまでパラメータ推定の問題として扱われていた因果強度の判断をモデル比較の問題として捉え直し、原因候補と結果の間に因果関係が存在するかという構造の判断から説明を試みた。ここで言うモデルとは、図3に示すような背景原因 (B) のみが結果 (E) に影響を及ぼす因果モデル (Graph 0) と背景原因 (B) と原因候補 (C) が結果 (E) に影響する因果モデル (Graph 1) のことである。背景原因と原因候補が独立に結果を生起させる場合、結果の生起確率は noisy-OR 関数を用いて以下のように算出される。

$$P(e|b, c; w_0, w_1) = w_0b + w_1c - w_0w_1bc \quad (10)$$

ここで w_0 は背景原因の因果強度を、 w_1 は原因候補の因果強度を表す。また、 b と c はそれぞれ背景原因と原因候補の有無を表し、存在するときは1、

7) 厳密には背景原因 (B) は原因候補 (C) 以外の全ての原因を表し、その中には観測不能な原因も含む。

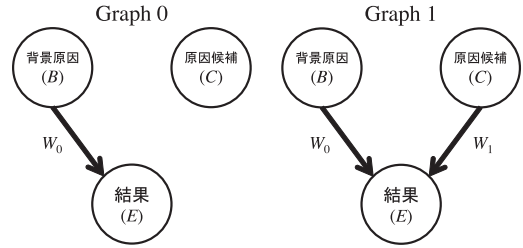


図3 背景原因 (B) と原因候補 (C) と結果 (E) の因果モデル。 w_0 と w_1 は背景原因と原因候補の因果強度をそれぞれ表す。

存在しないときは0になる。なお、発生的因果パワー q_G は noisy-OR 関数の下での w_1 の最尤推定量となることが示されている⁸⁾。

ベイズの定理を適用することにより、得られたデータが Graph 0 によって生じたか、Graph 1 によって生じたかは対数オッズの形で以下のように表現できる。

$$\begin{aligned} & \log \frac{P(\text{Graph 1}|D)}{P(\text{Graph 0}|D)} \\ &= \log \frac{P(D|\text{Graph 1})}{P(D|\text{Graph 0})} + \log \frac{P(\text{Graph 1})}{P(\text{Graph 0})} \end{aligned} \quad (11)$$

この式の右辺の第一項である対数尤度比はどちらの仮説が尤もらしいかを表す指標であり、ベイズファクター (Kass & Raftery, 1995) と呼ばれている。Griffiths & Tenenbaum (2005) はこの対数尤度比を因果サポートと名づけ、データが Graph 0 よりも Graph 1 を支持する指標とした。

$$\text{causal support} = \log \frac{P(D|\text{Graph 1})}{P(D|\text{Graph 0})} \quad (12)$$

なお、それぞれの尤度は以下の式から算出される。

$$P(D|\text{Graph 0}) = \int_0^1 P(D|w_0, \text{Graph 0}) P(w_0|\text{Graph 0}) dw_0 \quad (13)$$

$$\begin{aligned} P(D|\text{Graph 1}) &= \int_0^1 \int_0^1 P(D|w_0, w_1, \text{Graph 1}) P(w_0, w_1|\text{Graph 1}) dw_0 dw_1 \end{aligned} \quad (14)$$

ここで $P(w_0|\text{Graph 0})$ と $P(w_0, w_1|\text{Graph 1})$ は

8) ΔP は線形加算関数 (i.e., $P(e|b, c; w_0, w_1) = w_0b + w_1c$) の下での w_1 の最尤推定量となることが示されている。

因果強度の事前分布となっており、事前分布には 0 から 1 までの一様分布が想定されている。因果サポートモデルは、 ΔP ルールやパワー PC 理論では説明困難であった $\Delta P = 0$ の事態における実験参加者の評定値に適合するだけでなく、サンプルサイズの効果や不完全な分割表における判断や比率データからの学習を説明することに成功していた (Griffiths & Tenenbaum, 2005)。

Carroll, Cheng, & Lu (2013) は、複数の原因事象が単一の結果事象に影響する実験事態において連合学習の代表的なモデルと因果ベイズネットの比較を行っている。このような事態では 2 つの因果モデルの比較ではなく、複数の因果モデルの比較が行われる。連合学習モデルと因果ベイズネットのシミュレーションの結果と行動実験の結果を照らし合わせたところ、因果ベイズネットが支持された。また、Griffiths, Sobel, Tenenbaum, & Gopnik (2011) は、因果関係が存在する確率はどのくらいか、因果関係は決定的か確率的かといった先行知識の影響を検討し、それらが因果ベイズネットでうまく説明できることを報告している。

5.3 SS パワーモデル

Lu, Liljeholm, Cheng, & Holyoak (2008) は、因果強度の事前分布に偏りが無い一様分布を想定するよりも、因果の単純性 (Chater & Vitányi, 2003) を反映した分布を想定する方が適切であると主張している。因果関係における単純性とは、原因事象の数を最少にしつつ因果強度を最大にする因果モデルへの選好である。Lu et al. (2008) は (14) 式における因果強度の事前分布に必要な十分な因果関係への選好を組み込み、この事前分布に基づくベイズモデルを SS パワーモデルと名づけた。背景原因と原因候補が結果を発生させる場合、事前分布は以下のように表現される。

$$P(w_0, w_1 | \text{gen}, \text{Graph } 1) \propto (e^{-\alpha w_0 - \alpha(1-w_1)} + e^{-\alpha(1-w_0) - \alpha w_1}) \quad (15)$$

ここで α は必要十分な原因への選好を示すパラメータであり、 $\alpha = 0$ となる時、事前分布は一様分布となる。この事前分布は原因候補の因果強度と背景原因の因果強度が両立しないこと、つまり、一方の値が高く想定される場合は他方の値が低く想定され

ることを示している。また、SS パワーモデルでは「原因は結果をどの程度発生させるか？」という強度の判断と「因果関係はどのくらいありそうか？」という構造の判断は区別されており、前者には w_1 の事後分布の平均値が、後者にはベイズファクターが用いられる。

SS パワーモデルは、Perales & Shanks (2007) がメタ分析を行うために収集した複数の実験のデータセットに対して因果サポートモデルよりも良い当てはまりを示すだけでなく、強度の判断と構造の判断の違いや発生的因果関係と抑制的因果関係の差異も記述できていた。一様分布よりも単純性に基づく事前分布を支持する研究は他にも存在するが (Powell, Merrick, Lu, & Holyoak, 2016), SS パワーモデルとは異なる事前分布も提唱されている。Yeung & Griffiths (2015) は繰り返し学習 (iterated learning) の手法を用いて実験参加者の事前分布を推定した結果、原因候補と背景原因ともに因果強度は高く想定されていることを報告している。両者の実験事態は大きく異なるため、どのような要因によって事前分布が変化したかは不明瞭であるが、実験課題に使用したカバーストーリーが領域固有の先行知識として事前分布に影響した可能性は十分にあり得る。

ここまで因果ベイズネットを用いたモデルが因果強度の判断をうまく説明できることを示してきた。ここで興味深いことは、これまで対立してきた異なるモデル (e.g., ΔP ルール, パワー PC 理論) が因果ベイズネットという大きな枠組みの中で特殊な因果モデルの 1 つとして捉え直されたことである。ただ一方で、モデルを複雑化してデータへの適合度が上昇するのは当然と言えば当然のことである。 ΔP や因果パワーは分割表の頻度情報から簡単に算出できるが、因果サポートモデルや SS パワーモデルの予測値を算出する際にはマルコフ連鎖モンテカルロ法を用いて事後分布を推定しなければならない。モデルの複雑性という点では Hattori & Oaksford (2007) の二要因ヒューリスティックモデル⁹⁾は特筆に値する。このモデルは 2×2 の分割表において

9) 二要因ヒューリスティックモデルは以下の式で表現される。

$$H_G = \lim_{d \rightarrow \infty} \varphi = \sqrt{P(e|c)P(c|e)} = \frac{a}{\sqrt{(a+b)(a+c)}}$$

セル d の無視は、この情報が何も起きていないことを表しており、無数に存在していて数えられないことに起因する。

セル d を除く 3 つの頻度情報しか使用しないにも関わらず, Perales & Shanks (2007) のデータセットに十分に高い適合度を示している. このような状況の中でベイズ的アプローチに求められることは, 他のモデルが適用できない実験事態において因果ベイズネットの予測と一致した結果が得られるか否か検討することであろう. 例えば, これまで提唱されてきたモデルの多くは 2 値変数を対象としており, 連続変数を扱うことができない. 一方, 因果ベイズネットでは数理的基盤が構築されているため, 因果モデルには様々な変数や関数を組み込むことができる. 因果ベイズネットの更なる検証に加え, 連続変数を対象とした因果強度の判断に関する実証的知見はほとんど存在していないため, 今後の研究が期待される.

6. 因果ベイズネットによる因果推論

因果構造の学習や因果強度の学習が因果の知識の獲得を扱うのに対して, 因果推論では獲得した因果の知識をどのように使用するかということに焦点が当てられる. 一般的な実験事態では, 実験参加者に因果構造を教示して事象間の確率関係を学ばせた後で, ある事象の状態から別の事象に対して推測を行うことが求められる (see Rottman & Hastie, 2014 for a review). 例えば, Waldmann & Hagmayer (2005) の実験では, 3 種類の架空の物質の関係として 1 つの原因が 2 つの結果をもたらす共通原因の因果モデル (i.e., $E_1 \leftarrow C \rightarrow E_2$) が提示された. 実験参加者はこれら 3 つの物質の共変情報を学習した後, 一方の結果事象 (E_1) の状態から他方の結果事象 (E_2) の状態を推測することが求められた. ここでは, 因果ベイズネットと特に関連の高い 3 つのトピックについて述べる.

6.1 結果から原因への診断的推論

結果事象から原因事象を推測することは診断的推論やベイズ推論と言われる. 結果事象 (E) が生起している時に原因事象 (C) が生起している確率はベイズの定理によって以下のように算出される.

$$\begin{aligned} P(c|e) &= \frac{P(e|c)P(c)}{P(e)} \\ &= \frac{P(e|c)P(c)}{P(e|c)P(c) + P(e|\neg c)P(\neg c)} \quad (16) \end{aligned}$$

この式において原因事象が生起する確率 $P(c)$ は基準率 (base rate) と呼ばれ, 診断的推論に基準率が正しく反映されないことは基準率錯誤 (base rate fallacy) と言われている (e.g., Bar-Hillel, 1980). 基準率は常に無視されるというわけではなく, 自然頻度の形で表記された場合には考慮されやすくなることが知られている (see Barbey & Sloman, 2007 for a review).

Krynski & Tenenbaum (2007) は診断的推論における因果の知識の重要性を指摘している. 基準率錯誤の代表例である乳がん問題では, 乳がんである確率 $P(c)$, 乳がんである時に検査で陽性になる確率 $P(e|c)$, 乳がんでない時に検査で陽性になる確率 $P(e|\neg c)$ が与えられた上で, 検査が陽性の時に乳がんである確率 $P(c|e)$ を推測することが求められる. この問題の因果モデルは乳がんと検査結果で構成される単一の因果関係 (i.e., $C \rightarrow E$) であり, そのパラメータは基準率 $P(c)$ と原因がある時に結果が生起する条件付き確率 $P(e|c)$ によって記述される. 彼らは因果モデルに明示されないパラメータ (e.g., $P(\neg c)$, $P(e|\neg c)$) は考慮されにくくすると仮説を立て, 乳がんでない時に検査結果が陽性になることの原因として他の疾患を加え, 実験課題の因果構造を共通結果の因果モデル (i.e., $C_1 \rightarrow E \leftarrow C_2$) に変更した. 乳がんでない時に検査で陽性になる確率 $P(e|\neg c)$ は, 共通結果の因果モデルでは $P(e|c_2)$ となり, 因果モデルに明示されるパラメータとなる. 実験の結果, 偽陽性の検査結果が他の疾患によってもたらされると教示された場合には基準率錯誤が減少することが明らかとなった.

Meder, Mayrhofer, & Waldmann (2014) は診断的推論をパラメータ推定の問題ではなくモデル選択の問題として捉え直し, 因果ベイズネットを用いた新たな計算モデルとして構造帰納モデル (structure induction model) を提唱している. Griffiths & Tenenbaum (2005) が因果強度の学習を 2 つの因果モデル (図 3 参照) の比較によって説明したように, Meder et al. (2014) は観察されたデータを生成する因果モデルの不確実性が診断的推論に反映されたと考えた. すなわち, 背景原因 (B) のみが影響する因果モデル (i.e., Graph 0: $B \rightarrow E$) と原因事象 (C) と背景原因 (B) の両方が影響する因果モデル (i.e., Graph 1: $B \rightarrow E \leftarrow C$) のどちらの可能性が高いかによって, 判断が柔軟に変化するわ

けである。ベイズの定理に基づく予測では $P(c|e)$ が一定である限り診断的推論が変化することはないが、構造帰納モデルによると $P(e|c)$ が低い場合は背景原因のみが影響する因果モデル (i.e., Graph 0: $B \rightarrow E$) の可能性が高くなり、結果事象の生起は原因事象が生起していたことの証拠とはなりにくくなる。彼らは $P(c|e)$ を一定に保ちつつ $P(e|c)$ を変化させることで2つの因果モデルの可能性を操作し、実験参加者の診断的推論が構造帰納モデルの予測と一致することを報告している。また、このモデルは原因から結果への予測的推論では因果モデルの不確実性の効果は僅かであることを予測しており、実際そのような結果が得られている (see also Fernbach, Darlow, & Sloman, 2011)。

6.2 因果推論における観察と介入

ある事象の状態が判明した時に他の事象の状態について推測することは観察に基づく推論であり、因果モデルとそのパラメータによってその内容が決まる。一方、事象に対して操作を行った場合の推論は介入に基づく推論であり、両者は明確に区別される。例えば、外に出たときに地面が濡れていれば雨が降ったかもしれないとその原因を推測することができるが、庭に水を撒いた後に地面の濡れから降雨に対して推測を行うことはない。介入によって制御された事象 (e.g., 地面の濡れ) は本来の原因 (e.g., 降雨) とは独立になるからであり、このことはグラフ切除や無効化と言われる (Pearl, 2000; Sloman, 2005)。グラフ切除は介入された事象に対するエッジを取り除くことで表現され、単一の因果関係では原因事象の生起確率は基準率に従う。

$$P(c|do(e)) = P(c|do(\neg e)) = P(c) \quad (17)$$

因果推論における観察と介入の違いは多くの研究によって示されてきた。Sloman & Lagnado (2005) は複数の事象が関連している事態で特定の事象に対して介入を行った場合に他の事象についてどのような推測がなされるか検討した結果、介入された事象が本来の原因と独立に扱われていることを示した。Waldmann & Hagmayer (2005) は、事象の生起情報を表に記載した要約提示の因果学習課題を用いて、ヒトが観察による学習から介入の結果を正しく予測できることを報告している。同様の結果は、事象の生起情報を1試行毎に提示する

事例提示の因果学習課題においても得られている (Meder, Hagmayer, & Waldmann, 2009; 斎藤・嶋崎, 2012)。また、ラットでさえも観察と介入を区別して因果推論を行っていることを示唆する研究も報告されている (Blaisdell, Sawa, Leising, & Waldmann, 2006; Leising, Wong, Waldmann, & Blaisdell, 2008; but see also Dwyer & Honey, 2009)。

6.3 因果マルコフ条件に関する推論

因果マルコフ条件によると、共通結果の因果モデル (i.e., $C_1 \rightarrow E \leftarrow C_2$) では結果事象の状態が不明な場合、2つの原因事象は独立関係にあるが、結果の状態が明らかになると両者には従属関係が生じる。そのため、結果 (E) が生じた時に一方の原因 (C_1) が生起している確率 $P(c_1|e)$ と結果 (E) と他方の原因 (C_2) が生じた時に一方の原因 (C_1) が生起している確率 $P(c_1|e, c_2)$ は異なるものになる。結果の生起は他方の原因によって説明されるため、前者の確率 $P(c_1|e)$ が後者の確率 $P(c_1|e, c_2)$ よりも高くなることが知られている (see Khemlani & Oppenheimer, 2010)。片方の原因の生起によってもう片方の原因の生起確率が減少することは、心理学の領域では割引効果 (discounting effect) として (e.g., Kelley, 1973)、人工知能の領域では説明による取り除きの効果 (explaining away effect) として知られている (Pearl, 2000)。割引効果を反映した判断が行われることは多くの社会心理学的研究によって示されてきた (e.g., Morris & Larrick, 1995)。

一方、共通原因の因果モデル (i.e., $E_1 \leftarrow C \rightarrow E_2$) では因果マルコフ条件と一致しない判断が報告されている。共通原因モデルにおいて原因事象 (C) の状態が明らかになると、それぞれの結果事象 ($E_1 \cdot E_2$) は独立関係になることが因果マルコフ条件では想定されている。つまり、結果 (E_1) が生起するかは原因 (C) によって決まるため、もう一方の結果 (E_2) の状態は影響しないわけである (i.e., $P(e_1|c) = P(e_1|c, e_2) = P(e_1|c, \neg e_2)$)。しかしながら、他方の結果が生起している場合には推測対象となっている結果事象の生起確率を高く見積もることが知られており (e.g., Rehder & Burnett, 2005)、このことはマルコフ条件の逸脱 (Markov violation) と言われている。

因果ベイズネットはヒトがどのように判断すべきかを扱う規範的モデルであるが、比較的単純な実験事態ではその予測と一致した合理的な判断が見られることから、ヒトがどのように判断するかを扱う記述的モデルとしても採用されてきた。しかしながら、マルコフ条件の逸脱は記述的モデルとしての因果ベイズネットに対する明確な反証事例であり、この逸脱に対しては様々な説明が試みられている。例えば、Rehder & Burnett (2005) は因果関係のメカニズムを新たなノードとして因果モデルに加え、この潜在変数がそれぞれの事象に影響を及ぼすことで逸脱が生じると説明している。Park & Sloman (2013) はメカニズムに関する知識を操作した実験を行い、この仮説を支持する結果を報告している。また、Rehder (2014) は因果の方向性を考慮しない推論によってマルコフ条件の逸脱が生じていると考え、無向グラフを用いたモデルを構築して説明を試みている。ただし、これらはともにアドホックな説明であり、マルコフ条件の逸脱に一定の示唆を与えはするものの、一部の推論にのみ適用して他の推論に適用しないことについては恣意的と言わざるを得ない。Rottman & Hastie (2016) は様々な因果モデルを用いてマルコフ条件に関する推論を網羅的に検討し、ノードの追加や無向グラフによる説明が一部のデータとしか一致せず、全てのデータ一致するわけではないことを報告している。マルコフ条件の逸脱に関する研究はまだ少なく、因果ベイズネットが規範的モデルから記述的モデルとなるためには実証的知見の集積が必要不可欠である。

7. ベイズ的アプローチに対する批判および今後の展望

ベイズ的アプローチは様々な領域における実験結果の説明に成功してきたが、その一方で一部の研究者からは異議も唱えられている。ここではこれらの批判が因果性の学習と推論について当てはまるかどうか考察し、ベイズ的アプローチの今後の展望について述べる。Jones & Love (2011) はベイズ的アプローチの問題点として以下の2点を指摘している。1点目はベイズモデルが心理学的過程や概念を考慮していないことである。因果ベイズネットは何を計算するかという計算レベルのモデルであり、いかに計算するかというアルゴリズムレベルのモデルではない (cf. Marr, 1982)。そのため、どのよう

なメカニズムで行動が生じたかについては何も述べていない。彼らの批判は至極当然なものであるが、ベイズ的アプローチは計算レベルの説明からスタートしただけであり、アルゴリズムレベルの説明も試みられており (e.g., Lu, Rojas, Beckers, & Yuille, 2016)、この点について評価を下すのは早計であろう。また、因果パワーや因果モデルといった概念は因果ベイズネットが適用される前から提唱されており、因果ベイズネットによって明確に定義されるようになったことは興味深い。連合学習・ヒューリスティック・マルコフ連鎖モンテカルロ法など計算方法の候補は数多く存在するが、どのようなメカニズムとなっているかは認知神経科学との協力によって今後明らかにされるだろう。

ベイズ的アプローチに対する2点目の批判はベイズモデルの設定に対する根拠が不十分なことである。ベイズモデルを構築する際には仮説空間・事前分布・尤度などを設定する必要があるが、どのように設定するかはモデルの構築者に委ねられている。十分な根拠のない仮定に基づいて様々なモデルを自由に構築できる結果、ベイズモデルとデータが一致したとしても従来の結果を異なる枠組みで説明したに過ぎないと批判している。確かに因果ベイズネットの適用を試みた初期の研究 (e.g., Gopnik et al., 2001) は従来のモデルでも十分説明可能であったが、その後の研究では因果ベイズネットでしか説明できない実験結果が数多く報告されている (e.g., Waldmann & Hagmayer, 2005; Carroll et al., 2013)。また、これらの研究の多くでは複数のベイズモデルを構築してその比較検討が行われており、恣意的な設定によってデータと適合しているわけではないことが示されている。Jones & Love (2011) の批判は一部の研究に当てはまるかもしれないが、因果性の学習と推論に関しては的外れと言っても過言ではないだろう。

本論文で紹介した実験結果の多くはヒトが因果ベイズネットと一致した合理的な判断を行うことを示しているが、一方でヒトは不合理な判断を行うこともある。具体的には、制御幻想や迷信行動のように実際には因果関係がないにも関わらずであると判断することやマルコフ条件の逸脱のように無関係な事象によって判断が変化することが挙げられる。このような行動は合理的に信念を更新していくことを想定しているベイズモデルから導かれることはな

い。この問題に対する1つの解決策は二重過程理論 (Evans, 2008; Evans & Stanovich, 2013) を採用し、因果性の学習と推論を熟慮のプロセスと直観的プロセスの相互作用として説明することである。熟慮のプロセスに因果的思考を、直観的プロセスに連合的思考を想定すれば、因果ベイズネットの予測と一致しない結果も説明可能になる。

因果性の学習を連合学習の観点から説明する試みは古くから行われており (e.g., Dickinson et al., 1984), 因果関係と相関関係の区別を必要としない実験事態では多くの現象の説明に成功している (Shanks, 2007 for a review). Rescorla-Wagner モデル (Rescorla & Wagner, 1972) に代表される連合学習理論では、事象の共生起によって表象間に連合が形成され、一方の表象が活性化された場合には他方の表象も活性化されることが想定されている。マルコフ条件の逸脱を連合学習の観点から考えた場合、共通原因の因果モデル (i.e., $E_1 \leftarrow C \rightarrow E_2$) において他方の結果の有無によって判断が変化することは (i.e., $P(e_1|c, e_2) \neq P(e_1|c, \neg e_2)$), 連合を通じた表象の活性化の有無として解釈することができる。また、Rescorla-Wagner モデルでは関係が有ることの学習は関係が無いことの学習に比べて連合強度の変化量が大きくなるようにパラメータが設定されているため、実際には無関係な事象に対して学習初期には関係があると判断することを予測している。合理的なベイズモデルでは説明困難とされるこれらの不合理的な判断は連合学習モデルの予測と一致しているが、このことは連合学習のプロセスが不合理であることを意味しているわけではない。連合的思考によって因果関係を捉えようとする時に不具合が生じ得るというわけである。

因果性の学習と推論において直観的プロセスがどのように機能するかはよく分かっておらず、またこれら2つのプロセスがどのように相互作用するかはほとんど明らかになっていない。アルゴリズムレベルでのベイズモデルの構築とそれぞれのプロセスが優位になる事態での実験的検討により、その一端を明らかにすることができるだろう。心理学的妥当性の高い記述的モデルとなるためには、合理的側面だけでなく不合理的側面も説明することが求められる。

8. おわりに

本論文では、因果構造の学習・因果強度の学習・因果推論という3つの側面における因果ベイズネットの予測を示し、関連する実験結果を概観し、今後の方向性について考察を行った。実験結果の多くは因果ベイズネットの予測と一致しており、その心理学的妥当性が示唆された。一見すると因果ベイズネットによって多くの問題が解決したように感じられるかもしれないが、因果を記述する手段を得てようやくスタート地点に立てたと考えた方が適切かもしれない。マルコフ条件の逸脱が生じるのはなぜか、合理的なベイズモデルにおいて不合理な判断が生じるのはなぜか、そしてこれらはどのようなメカニズムによって発生しているのか、明らかにすべき課題はたくさんある。

文 献

- Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, **30**, 241–254.
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, **44**, 211–233.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53**, 370–418.
- Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, **3**, 385–402.
- Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science*, **311**, 1020–1022.
- Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**, 708–731.
- Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **29**, 1119–1140.
- Burns, P., & McCormack, T. (2009). Temporal information and children's and adults' causal inferences. *Thinking & Reasoning*, **15**,

- 167–196.
- Carroll, C. D., Cheng, P. W., & Lu, H. (2013). Inferential dependencies in causal inference: A comparison of belief-distribution and associative approaches. *Journal of Experimental Psychology: General*, **142**, 845–863.
- Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, **10**, 335–344.
- Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, **7**, 19–22.
- Chater, N., Tenenbaum, J. B., & Yuille, A. (Eds.) (2006). Probabilistic models of cognition [Special issue]. *Trends in Cognitive Sciences*, **10**.
- Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, **104**, 367–405.
- Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, **79**, 102–133.
- Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, **9**, 309–347.
- Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, **10**, 294–300.
- Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, **14**, 1–13.
- de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives [Foresights: Its logical laws, its subjective source]. *Annales de l'institut Henri Poincaré*, **7**, 1–68.
- Deverett, B., & Kemp, C. (2012). Learning deterministic causal networks from observational data. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 288–293. Austin, TX: Cognitive Science Society.
- Dey, D. K., & Rao, C. R. (2005). *Handbook of statistics 25: Bayesian thinking, modeling, and computation*. Amsterdam, Netherlands: Elsevier. (繁榘 算男・岸野 洋久・大森 裕浩 (監訳) (2011). 『ベイズ統計分析ハンドブック』. 東京: 朝倉書店.)
- Dickinson, A., Shanks, D. R., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology*, **36A**, 29–50.
- Dwyer, D. M., Starns, J., & Honey, R. C. (2009). “Causal reasoning” in rats: A reappraisal. *Journal of Experimental Psychology: Animal Behavior Processes*, **35**, 578–586.
- Eberhardt, F., & Scheines, R. (2007). Interventions and causal inference. *Philosophy of Science*, **74**, 981–995.
- Eells, E. (1991). *Probabilistic causality*. Cambridge, United Kingdom: Cambridge University Press.
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, **59**, 255–278.
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, **8**, 223–241.
- Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, **140**, 168–185.
- Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **35**, 678–693.
- Fishburn, P. (1986). The axioms of subjective probability. *Statistical Science*, **1**, 335–345.
- Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.
- Glymour, C. (2003). Learning, prediction, and causal Bayes nets. *Trends in Cognitive Sciences*, **7**, 43–48.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, **111**, 3–32.
- Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from pat-

- terns of variation and covariation. *Developmental Psychology*, **37**, 620–629.
- Gopnik, A., & Tenenbaum, J. B. (Eds.) (2007). Bayesian networks, Bayesian learning and cognitive development [Special section]. *Developmental Science*, **10**, 281–364.
- Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, **35**, 1407–1455.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, **114**, 211–244.
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, **51**, 334–384.
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, **116**, 661–716.
- Hájek, A. (2003). Interpretations of probability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012 ed.). Retrieved from <http://plato.stanford.edu/archives/win2012/entries/probability-interpret/>
- Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science*, **31**, 765–814.
- Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning*, **20**, 197–243.
- Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, **62**, 135–163.
- Jenkins, H., & Ward, W. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, **7**, 1–17.
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, **34**, 169–231.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, **28**, 107–128.
- Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, **116**, 20–58.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, **55**, 271–304.
- Khemlani, S. S., & Oppenheimer, D. M. (2010). When one model casts doubt on another: A levels-of-analysis approach to causal discounting. *Psychological Bulletin*, **137**, 195–210.
- Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge, United Kingdom: Cambridge University Press.
- Körding, K., & Wolpert, D. (2004). Bayesian integration in sensorimotor learning. *Nature*, **427**, 244–247.
- Kousta, S. (Ed.) (2010). Approaches to cognitive modeling [Special section]. *Trends in Cognitive Sciences*, **14**, 339–364.
- Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgement under uncertainty. *Journal of Experimental Psychology: General*, **136**, 430–450.
- Kushnir, T., Wellman, H. M., & Gelman, S. A. (2009). A self-agency bias in preschoolers' causal inferences. *Developmental Psychology*, **45**, 597–603.
- Lagnado, D. A., & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **30**, 856–876.
- Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **32**, 451–460.
- Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond covariation: Cues to causal structure. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*, 154–172. Oxford: Oxford University Press.
- Leising, K. J., Wong, J., Waldmann, M. R., & Blaisdell, A. P. (2008). The special status of actions in causal reasoning in rats. *Journal of Experimental Psychology: General*, **137**, 514–527.

- Lewis, D. (1973). Causation. *The Journal of Philosophy*, **70**, 556–567.
- Lu, H., Rojas, R. R., Beckers, T., & Yuille, A. L. (2016). A Bayesian theory of sequential causal learning and abstract transfer. *Cognitive Science*, **40**, 404–439.
- Lu, H., Yuille, A., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, **115**, 955–984.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W.H. Freeman.
- Mayrhofer, R., & Waldmann, M. R. (in press). Sufficiency and necessity assumptions in causal structure induction. *Cognitive Science*.
- McCormack, T., Frosch, C., Patrick, F., & Lagnado, D. (2015). Temporal and statistical information in causal structure learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**, 395–416.
- Meder, B., Gerstenberg, T., Hagmayer, Y., & Waldmann, M. R. (2010). Observing and intervening: Rational and heuristic models of causal decision making. *The Open Psychology Journal*, **3**, 119–135.
- Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory & Cognition*, **37**, 249–264.
- Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review*, **121**, 277–301.
- Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review*, **102**, 331–355.
- Neapolitan, R. E. (2004). *Learning Bayesian networks*. Upper Saddle River, NJ: Prentice Hall.
- Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review*, **111**, 455–485.
- Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the Markov property in causal reasoning. *Cognitive Psychology*, **67**, 186–216.
- 大塚 淳 (2010). ベイズネットから見た因果と確率. 『科学基礎論研究』, **38**, 39–47.
- Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. *Proceedings of the 7th Annual Conference of the Cognitive Science Society*, 329–334.
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, United Kingdom: Cambridge University Press.
- Perales, J. C., & Shanks, D. R. (2007). Models of covariation-based causal judgment: A review and synthesis. *Psychonomic Bulletin & Review*, **14**, 577–596.
- Powell, D., Merrick, M. A., Lu, H., & Holyoak, K. J. (2016). Causal competition based on generic priors. *Cognitive Psychology*, **86**, 62–86.
- Psillos, S. (2005). Causality. In M. Horowitz (Ed.), *New dictionary of the history of ideas*. New York, NY: Charles Scribner's Sons.
- Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive Psychology*, **72**, 54–107.
- Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, **50**, 264–314.
- Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Appleton-Century-Crofts.
- Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological Bulletin*, **140**, 109–139.
- Rottman, B. M., & Hastie, R. (2016). Do people reason rationally about causally related events? Markov violations, weak inferences, and failures of explaining away. *Cognitive Psychology*, **87**, 88–134.
- 斎藤 元幸・嶋崎 恒雄 (2012). 因果推論に対する因果ベイズネットの適用の妥当性について—観察と介入の違いに関する実験的検討—. 『認知科学』, **19**, 380–390.
- Saito, M., & Shimazaki, T. (2012). Strategy

- changes in causal structure learning: The role of task complexity. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 2264–2269. Austin, TX: Cognitive Science Society.
- Shanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology*, **60**, 291–309.
- Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. New York, NY: Oxford University Press.
- Sloman, S. A., & Lagnado, D. A. (2005). Do we 'do'? *Cognitive Science*, **29**, 5–39.
- Sloman, S. A., & Lagnado, D. (2015). Causality in thought. *Annual Review of Psychology*, **66**, 223–247.
- Sobel, D. M., & Kushnir, T. (2006). The importance of decision making in causal learning from interventions. *Memory & Cognition*, **34**, 411–419.
- Sobel, D. M., & Sommerville, J. A. (2009). Rationales in children's causal learning from others' actions. *Cognitive Development*, **24**, 70–79.
- 園信太郎 (2014). 『確率概念の近傍—ベイズ統計学の基礎をなす確率概念—』. 東京: 内田鶴園.
- Spirtes, P., Glymour, C., & Scheines, P. (1993). *Causation, prediction, and search*. New York, NY: Springer-Verlag.
- Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, **27**, 453–489.
- Suppes, P. (1970). *A probabilistic theory of causation*. Amsterdam, Netherlands: North-Holland.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, **331**, 1279–1285.
- 戸田 正直 (1982). 主観確率の測定法について. 『行動計量学』, **9**, 1–11.
- 豊田 秀樹 (編) (2015). 『基礎からのベイズ統計学—ハミルトニアモンテカルロ法による実践的入門—』. 東京: 朝倉書店.
- 植野 真臣 (2013). 『ベイジアンネットワーク』. 東京: コロナ社.
- Verma, T. S., & Pearl, J. (1990). Equivalence and synthesis of causal models [Technical report R-150]. *Department of Computer Science, University of California, Los Angeles*.
- Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**, 216–227.
- Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgments of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 174–188.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford, United Kingdom: Oxford University Press.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, **114**, 245–272.
- Xu, F., & Griffiths, T. L. (2011). Probabilistic models of cognitive development [Special issue]. *Cognition*, **120**, 299–412.
- Yeung, S., & Griffiths, T. L. (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology*, **76**, 1–29.

(Received 1 July 2016)

(Accepted 16 Dec. 2016)



齋藤 元幸 (正会員)

2008 年関西学院大学文学部総合心理科学科卒業。2010 年関西学院大学大学院文学研究科博士課程前期課程修了。2013 年関西学院大学大学院文学研究科博士課程後期課程単位取得退学。博士 (心理学)。

2013 年から関西学院大学大学院文学研究科研究員、2015 年よりカリフォルニア大学ロサンゼルス校客員研究員。推論や意思決定に関心を持ち、因果性の学習と推論に関する心理学的研究に従事。EuroCogSci2011 Best Student Paper Award 受賞。日本心理学会、日本認知科学会、Cognitive Science Society 各会員。