潜在変数と選択バイアスが存在する場合の因果推論

ピーター・スパートス、クリストファー・ミーク、トーマス・リチャードソン
カーネギーメロン大学哲
学科 15213, PA Pittsburgh

アブストラクト

潜在変数と選択バイアスの両方が働いている可能性がある場合、因果関係を発見するための一般的で情報量が多く、信頼性の高い手順があることを明らかにする。測定された変数間の条件付き独立性と依存性に関する情報があれば、潜在変数や選択バイアスが存在する場合でも、ある変数から別の変数への因果経路が存在すると再信頼的に結論付けるための十分条件と、そのような因果経路が存在しない場合に信頼性の高い結論付けるための十分条件が存在する。

1 はじめに

ランダムに選ばれていないサンプルから因果関係を推論することには、よく知られた問題がある。**Sprites**ら(1993)は、調査対象の変数と特定の因果関係を持つ特徴に基づいてサンプルが選択された場合、そうでなければ正しい発見アルゴリズムが大規模サンプルの限界においてさえ失敗することを示した。例えば、ある集団でXとYが独立であるにもかかわらず、たまたまXとYに影響される変数Zの値を用いて集団からサンプルを選択した場合、XとYは（Zに対する条件付けによって）集団では持たない統計的依存性をサンプルで持つことになる。Cooper (1995)は、この種のラ・テント変数を含む、より興味深い例をいくつか挙げている。選択バイアスを表現する方法と、選択バイアスがデータから検出できる特別なケースについては、Wermuth, Cox, and Pearl (1994)で議論されています。重要な問題は、潜在変数と選択バイアスの両方が働いている可能性がある場合、因果関係を発見するための一般的で情報量が多く、信頼できる手順があるかどうかということです。選択バイアスが適用されない場合、**FCIという**アルゴリズムがあり、後述する仮定の下で、（大標本限界において）ほぼ確実に、因果経路の有無に関する情報を含むそのような情報を与える。

は、1つの測定変数から別の測定変数へと変化する（Spines et al.1993）。我々は、出力の再解釈のもと、**FCI**アルゴリズムは、選択バイアスが存在する可能性がある場合にも適用されるが、出力は、選択バイアスがないことが知られている場合よりも一般的に情報量が少ないことを示した。また、潜在変数や選択バイアスが存在する場合でも、測定された変数間の条件付き独立性と依存関係に関する情報があれば、ある変数から別の変数への因果経路が存在すると再信頼的に結論付けるための十分条件と、そのような因果経路が存在しない場合に信頼性高く結論付けるための十分条件が存在することを明らかにした。

本論文では、変数のセットは太字で、定義された用語はイタリックで表記している。 グラフ理論的な用語は付録で定義している。**FCI**アルゴリズムとその正しさの証明は、Spines et at.(1993).スペースがないため、ここではそのアルゴリズムについて説明しない。この論文の本編では、選択バイアスが出力の解釈において必要とする適応と再解釈を説明することに焦点を当てる。本論文の付録で示した定理は、Spirtesら(1993)で示した、選択バイアスのない**FCIアルゴリズムの**正しさの証明を、面倒ではあるが選択バイアスのケースに単純に適応することで証明したものである。

2 有向無尽グラフ

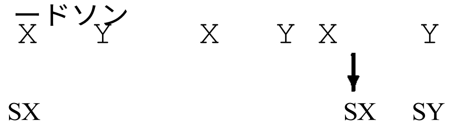
因子**分析**モデル、独立誤差を持つパスモデル、独立誤差を持つ再帰線形構造方程式モデル、各種潜在能力モデルなどは、すべて有向非循環グラフモデルの一例である。頂点Vの集合を持つ*有向*acyclic *graph* (DAG) Gは、因果的解釈と統計的解釈の両方を与えることができる。(統計的解釈ではDAGモデルはJ3ayesian Networkと呼ばれるPearl 1988、Spirtes et al.）テケ「Aは、ある変数セットVに関して、ある集団のメンバーにおけるBの直接的な原因である」を原始的とする。Vが変数の集合である場合、変数Vを持つ集団の因果DAG Gにおいて、AからBへのエッジが存在するのは、その集団のあるメンバーにおいて、AがVに対してBの*直接の*原因である場合に限られる。

を表現することができる。頂点Vの集合を持つDAG Gは、V上で許容される尺度とGのグラフ構造とを関連付ける制限の下、V上の確率尺度の集合を表すこともできる。1990年、我々は、変数の集合V上の確率測度が、頂点Vを持つDAG Gに対して、*局所有向Morkoo特性を*満たすのは、V内のすべてのWに対して、Wがその親の集合を条件として、すべての非親の非孫の集合から独立している場合のみであると言う。

3　セレクションバイアスの表現

我々は、標本の分布が、その標本が抽出された母集団分布と異なる2つの異なる理由を区別する。1つ目は、標本のばらつき、つまり*標本*ビンという単純なミリアール現象で、ある母集団分布に対して、有限の無作為標本から得られるパラメータ推定値は、一般に母集団のパラメータと正確に一致するわけではありません。第二の理由は、Vの変数間の因果関係や、標本の個体が集団から選択されるメカニズムによって、標本で期待されるパラメータ値と集団のパラメータ値に差が生じることがあるからである。この場合、その差は*選択*トリオによるものと言うことになる。サンプリングバイアスは、より大きなサンプルを採取することで改善される傾向があるが、セレクションバイアスはそうではない。この論文では、サンプリング・バイアスの問題は考えません。私たちは常に、無限大の理想化された選択された部分集団を扱っていると仮定しますが、選択バイアスがある可能性があります。

選択バイアスを表現する目的のために，Cooper (1995)に従って，各測定された確率変数Aについて，集団のそのメンバーについてAの値が記録された場合に1に等しく，そうでなければ0に等しいバイナリ確率変数 SAが存在すると仮定する．Vが変数の集合である場合，我々は常にVが3つの集合に分割できると仮定する：測定された変数の集合O（観測の略），Oに対する選択変数の集合S（選択の略），および残りの変数L（潜伏の略）．選択された部分集団におけるOの部分集合X上の限界分布では，選択変数の集合Sは，その値が選択された部分集団において常に1に等しいので，条件付きである．したがって、Oの不連続な部分集合X、Y、Zについて、XP Z|Yかどうかは判別できないが、XMZ|Y U（S＝1）かどうかは判別できると仮定することになる。(XMZ|Yは、XがYを与えられたZから独立していることを**意味し**、Yが空であれば単にXYZと書く。XのメンバがXだけである場合は
XMZ|Y の代わりに (X} MZl) と書く。 S の変数がすべて常に同じ値をとる場合もあるので、そのような場合は、単一の変数 S で選択を表すことにする。



(i) (ii) (iii) 図1：3つの因果関係DAG

図1に示したある集団の3つの因果DAGは、選択変数と非選択変数がどのように関係しうるかを示すもので、多くの異なる方法がある。(i)の因果DAGは、例えば、X値が記録された集団のメンバーとY値が記録された集団のメンバーが、一対の独立したコインの8ipsによってランダムに選択された場合に発生するものである。(ii)のDAGは、X値とY値の両方が記録されている（すなわち、Sx - Syで、サンプルに欠損値がない）ユニットを選択するために1枚のコインの8ipが使用された場合に発生することになる。(iii)のDAGは、例えばXが教育年数で、Xの値が高い人が低い人よりも教育に関するアンケートに回答し、サンプルに登場する頻度が高い場合に発生することになる。変数Yがある変数X / YのSxの原因である可能性、またSxが2つの異なる変数の原因である可能性を事前に排除しない。

見ていない部分集団、あるいは見ていない部分集団全体について、正しい因果関係の結論を導くためには、いくつかの仮定が必要である。まず、選択された部分集団から集団全体についての因果推論に興味がある場合を考えてみる。因果グラフの概念は、変数と母集団に関連するものである。したがって、母集団全体の因果グラフと選択された部分集団の因果グラフは異なる場合がある。例えば、ある薬が男性の生存率には影響しないが、女性には影響する場合、集団の因果グラフには薬から生存率へのエッジが存在するが、男性の下位集団の因果グラフには薬から生存率へのエッジが存在しない。このため、選択された部分集団の因果グラフから、集団または選択されていない部分集団の生存に関する因果的な結論を導くためには、次のような仮定をすることになる：

母集団推論の前提Vが変数の集合である場合、母集団のV上の因果グラフは、選択された部分集団と選択されていない部分集団のV上の因果グラフと同一である。

確率能力分布と因果関係に関するいくつかの付加的な仮定があり、以下の例で紹介する。最初の仮定は、因果関係DAGにおいて、各変数はその親（すなわち、親）が与えられたとき、その非孫（すなわち、間接的にも影響を与えない変数）からは独立である、というものである。

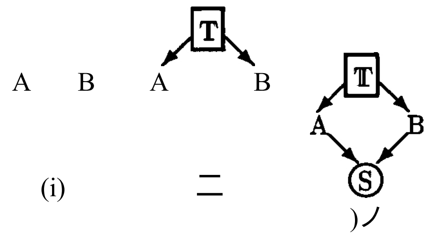


図2：選択バイアスを表現する

その直接の原因）。しかし、この原則は、次の例で示すように、任意の変数の集合や任意の部分集団には適用されないことを理解することが重要である。Pop' を母集団 Pop の部分集団とし、Pop' のすべてのメンバーに値 S = 1 が割り当てられるとする。 ここで、(i)と
(ii) はPop'の正しい因果DAGであり(ただし(ii)の方がより詳細)、図2の(iM)はPopの正しい因果DAGとなる。(図では、潜在変数を箱、選択変数を丸で囲んでいる）。また、AとBはPop'とPopの両方で従属するが、AとBはPop'ではTを与えられても従属するがPopでは従属しないとする。

(i) 図2のAはBの直接の原因ではなく、BはBの直接の原因ではないので、正しい因果関係DAGである。
A. ただし、Aと共通の原因が潜在的に存在することを示すものではないので、incomplèteであることに留意する。
B. しかも、Pop'ではAとBが依存関係にあるため、(i)では局所有向マルコフ特性が成立しないほどincomplèteである。図2のDAG(ü)は、AとBの因果関係をより詳細に描いたものであるが、Pop'においてAとBがTを与えられて依存しているため、変数集合｛A，B，T｝の因果関係DAGがPop'の(ü)について局所有向マルコフ特性を満たすとは言い切れない。図2の(lii)は、(i)、(ii)のいずれよりも詳細な因果関係DAGである。さらに、(A,B,T,S}は、母集団Pop（すなわち、Sに条件をつけない）において、AとBはTを与えられたら独立であるから、図2の（ni）に対する局所有向マルコフ特性を満たす。したがって、潜在変数を含めて、Pop' から Pop へと母集団を広げることによって、結局、その変

は、集団の異なるメンバー間の因果的相互作用（集団の個々のメンバーの特性間の因果的相互作用とは異なる）がある場合に成立し、またフィードバックを持つ集団では一般に成立しない。

因果的忠実性の仮定は、基本的に以下の通りです。

数の集合に対する原因DAGの局所有向マルコフ特性を満たす母集団とその中の変数に対する分布を見つける地点に到達する。

因果マルコフ仮定： 各集団Pop'と変数V'の集合に対して、Pop'がSによって選択される集団Pop、集合V D V' U S、V上のPopの因果グラフG、Gの局所有向マルコフ特性を満たすPopの分布P（V）が存在する（このような集合VをPop'、Pop、Sに*因果的に十分*と呼ぶ）。

因果的マルコフ仮定は、多くの場合

サブポピュレーションで真となる条件付き独立関係は、DAG の特定のパラメータ化のためではなく、構造的な理由（すなわち、DAG 構造のため）で真であること。

因果的忠実性の仮定Pop' をS が選択したPop の部分集団、V を因果的に十分な変数の集合、G を変数V を持つ集団Popの因果グラフ、P（V）を分布とすると Popにおける Vは、Vにおける任意の不連続なX、Y、Z に対して、因果マルコフAs-sumption とS = 1 がXMZ|Y U (S = 1) を内包する場合にのみ、XMZ|Y U (S = 1) となります。(この場合、P(V) はS によって選択されたPop'においてGに忠実であると言う)。

Causal Fbithfulness Condition は、様々な理由で失敗する可能性がある。それは、グラフ構造のせいではなく、ある条件付き依存関係がDAG のある特定のパラメタ化に対して成立し、他のパラメタ化に対しては成立しないことである可能性がある。また、変数間に決定論的な関係がある場合にも、Vi-olated になる可能性がある。しかし、Spirtes ら(1993) やMeek(1995) は、線形正規構造方程式モデルや離散ベイジアンネットワークの自然なパラメタリゼーションにおいて

統計学文献によく見られる因果関係モデル。
を含む、独立誤差を持つすべての再帰的構造方程式モデル。(線形再帰構造方程式モデルの紹介はBollen 1989を参照）。因果マルコフ仮定は、一般的には

、忠実度の違反につながるパラメタリゼーション集合はルベーグ尺度0であることを示している。

これらに似た公理（S に対する条件付けの制限なし）の正当性については、Spirtes ら（1993 ）に述べられている。以降の節で示すように、これらの仮定により、選択された部分集団の条件付き独立関係から、集団の因果グラフについて信頼性の高い推論を行うことができるようになる。もし、DAG G の局所的なdi-rected Markov 特性を満たすすべての分布で条件付き独立関係が真であれば、G は条件*付き独立関係を*継承すると言い、同様に、DAG G に忠実なすべての分布で条件付き依存関係 が 真 で あ れ ば 、 G entoiis *the conditioned dependence relationship* と言う。

4 実施例

ここで、様々な異なる仮定の下で、条件付きイン依存関係や依存関係のいくつかの異なるセットと、それらを生成した因果DAGについて何がわかるかを考察する。[1]

変数の集合V にかかる因果グラフG が与えられたとき、我々は

'例を詳細に追うことに興味のある読者のために、我々の仮定では、DAGが内包する条件付きイン依存関係はd-separation関係で与えられる。詳しくはPearl 1988を参照。

は、VSに含まれる任意の3つの不連続な変数セットX、Y、Zに対して、GがXMZ|Yをen-tailする場合に限り、XMZ|Y U Sを伴う場合、選択バイアスはないと言います。(これは一般に、選択された部分集団と集団の*分布が*同じであることを意味するのではなく、同じ条件付き独立関係が両者で成立することを意味することに注意)。例えば、Sの変数がVの他のどの変数とも因果関係がない場合に起こる。この場合、DAGを図に描くとき、Sの変数を省略し、Vの変数と同じ条件独立関係を持つエッジを省略する。

は、S の端点である。

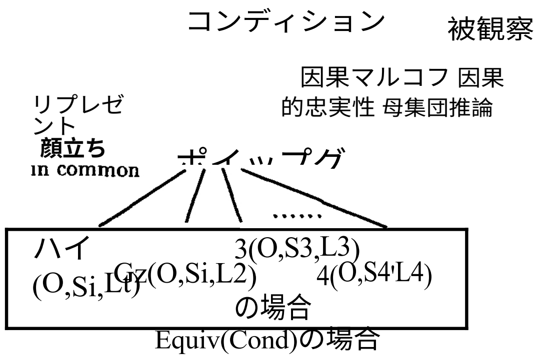
与えられたDAG Gに対して、Gの変数セットVを観察（O）、選択（S）、潜在（L）に分割する。(L)変数とし、**G(O,S,L)と**表記することにする。X、Z、YをO'の部分集合とし、検定可能な条件付き独立関係はXMZ|Y U (S = 1)だけと仮定し、これを観測*可能な*条件付き独立関係の集合と呼ぶことにします。X、Y、ZがOに含まれ、XMZ|Y U (S = 1)であれば、観測された条件付き独立関係であると言うことになる。Cond(O)は、Oの変数間の条件付き独立関係の集合である。DAG **G(O,S,L)** は、Oに含まれる各X、Y、Zに対して、以下の場合に限りXMZ|Y U (S = 1)を内包していればCond(O)を満たしている。

XMZIがCond(O)にある場合。 (OがCond(O)を内包する全てのDAGで同じであるが、Sはそうで

はないことに注意。これは、いくつかのDAGが崩壊することを許容しているためである。

を1つの選択変数に変換する)。文脈からOが何であるかが明らかな場合は、単にCondと書くことにする。しかし、全く同じCondを内包する多くの異なるDAGが存在する可能性がある。あるCondを内包する全てのDAGの集合をEquiv(Cond)と呼ぶことにする。

今、研究者が正しい因果関係DAGが何であるかを知らないが、おそらく選択した部分集団に対して条件付き独立関係の仮説検定を行うことによって



、Condが何であるかを決定できると想像してください。(後で見るように、Condのメンバーの多くはCondの他のメンバーを含むので、実際にテストする必要があるのはCondのメンバーの一部だけである)。この情報だけで、因果マルコフ仮定、因果忠実性仮定、および母集団推論仮定から、彼または彼女が結論付けられるのは、真の因果DAGはHi (O,Si.li) かG **(O,S2.L2)** か、または、以下のいずれかであるということだ。

G3(O,S3,L3) や 4(O,S4,$4) など、つまり真の因果DAG はEquiv(Cond) のあるメンバーです。もしEquiv(Cond)
が大きい場合、Equiv(Cond)のメンバー全員が何らかの重要な特徴を共有していない限り、この情報自体はあまり興味をそそられるものではありません。しかし、以下の例で示すように、Equiv(Cond)のメンバーには重要な特徴が共通して*いる*場合がある。

[2]Oの変数がすべて記録されている部分集団に限って見ている。X、Z、Yの条件付き独立関係を考えるとき、X、Z、Yに対応する選択変数が1に等しい値を持つ部分集団だけを見れば、より複雑なストーリーが語れる。

図3：一般的な戦略

我々の戦略は、Spirtes et a1. (1993)に記述されている選択バイアスのないストラットエギーを一般化したものである。(1993)で説明された選択バイアスのないstrat-egyを一般化したもので、Causal Markov Assumption、Causal Faithfulness Assumption、Population Inference Assumptionを用い、*部分指向誘導経路グラフ*（POIPG）と呼ばれるグラフオブジェクトを、Condから構築することになる。POIPGは（例と付録でより詳細に説明される）Equiv(Cond)のすべてのDAGに共通するある特徴を表現しています。[3]構築されたPOIPGから、Equiv(Cond)のすべてのDAGが他の興味深い特徴を共有していることを推論できる場合がある、例えば、それらはすべてAからBへの有向パスを含むかもしれない。この場合、CondからEquiv(Cond)のどのDAGが真の因果DAGであるかを正確に知ることはできないが、Equiv(Cond)のすべてのDAGがAからBへの有向パスを含むことを知っているので、真の因果DAGにおいてAがBの（おそらく間接）原因であると確実に結論づけることができる。この戦略は図3に概略的に示されている。以下の例では、観測された条件付き独立関係の特定のセットにこの戦略を適用し、DAGのどのような特徴を再信頼的に推論することができるかを示している。

4.1　例1

まず、観測された条件付き独立関係の集合があまり情報量が多くない、非常に単純な例から始めることにする。(簡単のため、以下のすべての例で、Sの変数はすべて同じ値をとり、したがって単一の変数Sで表現できると仮定する)。Oを{A,B}とし、観測された条件付き独立関係の集合Condlは空、すなわちCondl = 8とする。ここで、Equiv(Condl)にどんなDAGがあるのか調べたい。Vを因果的に十分な変数の集合とする。このようなDAGの最も単純な例は、V = O = {A,B}であり、*選択バイアスがない*場合である。(V=Oで選択バイアスがないことは、一般的に仮定か背景知識から来るもので、これらの条件を決定的に確認することは一般的に不可能であるため。

POIPGは、Verma and **Pearl** 1991で説明された（名前は付けられていない）構造の一般化であり、Wermuth, Cox, and Pearl 1994で用いられた表現スキームと共通する特徴がある。

をデータから得るだけである)。このような前提のもと、Condlを内包するDAGが2つだけ存在し、ラベルは
図 4 の (i) と (ii) である。一般に、潜在変数がなく選択バイアスもないとき、O(A,B}のすべての部分集合Xについて、AとBがX U（S＝1）を与えられて依存する場合にのみ、AとB間にエッジが存在する。

ここで、潜伏変数があるが選択バイアスはないとする。図4の(ni)と(vi)にそのようなDAGが2つ示されている。図4の(ni)と(vi)の例は、潜在変数が存在する場合、Opha(A,B}のすべての部分集合Xに対して、AとBがXUSで依存する場合にのみ、AとBの間にエッジが存在するとは限らないことを示している（選択バイアスなしなら、AとBがX U Sで依存するなら、XでA Bが依存するとだけ思い起こせばよいだろう。）

最後に、選択バイアスがある場合について考えてみよう。Equiv(Condl)におけるこのようなDAGの例を、図4の(iv)と(v)に示す。

Equiv(Cond1)のDAGは、特に潜在変数と選択バイアスの両方の可能性がある場合、ほとんど共通点がないように見える。Equiv(Cond1)のDAGは非常に多様であるが、すべてのDAGがEquiv(Condl)の中にあるわけではない。例えば、辺が全くないDAG **G(O,S,L)**はEquiv(Condl)には含まれない。

Equiv(Cond1)では、Oの各分割集合Xについて、X U S（この例ではXは単なる空集合）が与えられると、AおよびBは依存する。一般に、任意のOについて、この条件の簡単なグラフの特徴付けが存在する。DAG **G(O,S,L)**は、Oの各分割集合Xに対して、AとBの間にceHain種類の無向パスが**G(O,S,L)**に存在する場合に限り、X U Sが与えられるとAとBが従属することを含む。この無向きの経路を誘導経路という。UがDAG **G(O,L,S)の**OにおけるA-B間の誘導経路であるのは、U上のすべての衝突者が(A,B}に子孫を持つようなA-B間の非周期無向性経路である場合に限られる。また、終点を除くU上の非コライダーはO U Sに存在しない。(頂点Vが無向パスU上のコライダーであるのは、U上の隣接する2つの辺がVに入る、すなわち、U上にX -+ V +- YがUのサブパスとなる変数XおよびYが存在する場合に限る)

終点がXとYの誘導パスの向きは、XとYを含む誘導パス上の辺が、それぞれXとYに矢印の頭を持つかどうかで決まる。A-B間の誘導パスの向きは、図4に示すように4通りの可能性があります。Aから出てBに入る誘導経路（例：(iv)の誘導経路A -+ S +- B）、Aから出てBに入る誘導経路（例：(i)の誘導経路A -+ B、(v)の誘導経路A -r S +- T -+ B）、Bから出てAに入る誘導経路（例：(ii)の誘導経路B -+ A）、Aに入る誘導経路のうち、4つの経路が存在する。

AからBへの誘導経路（例：A +- T -+ Bの誘導経路で(iii) となり、(vi)の誘導経路A +- T -+ BとA +- U -r Bが存在する。）したがって、Condlが観測された場合、AとBの間に誘導経路があることはわかるが、その誘導経路の向きがどうなっているかはわからない。POIPGでAとBが隣接していることは、AとBの間に誘導経路があることを意味し、辺の両端にある「o」は誘導経路の向きを知ることができないことを意味する。誘導パスの存在と向きは、Equiv(Cond1)のDAGについて特に興味深い情報ではないのが普通である。POIPGから、AとBの因果関係について、もっと興味深いことがわかるだろうか。この場合、答えはノーである。しかし、次の例では、より興味深い結論を導き出すことができるケースを示す。

4.2 例2

O＝（A，B，C，D｝、Cond2＝｛DC（A，B｝|{C}｝とする、
ABB}と、これらによって包含される他のすべての条件付き独立性関係。Equiv(Cond2)において、潜伏変数がなく、選択バイアスもないDAGは、図5の(i)のみである。

ここで、V / O のように潜在変数を持つ DAG を考えるが、選択バイアスはないものとする。その場合、潜在変数の数に上限がなければ、Equiv(Cond2)のDAGは無限に存在し、そのうちのいくつかは図5の(ii)、(iii)、(iv)に示すようなものである。

ここで、選択バイアスのあるDAGを考えたとする。図4.2の(v)と(vi)は、Equiv(Cond2)であり、選択バイアスを持つDAGの例である。

図4.2のDAGに共通するものはあるのだろうか? Equiv(Cond2)のどのDAGでも、(A,D), (B,D), (A,B)のペアの間には誘導経路が存在しない。なぜなら、これらのペアはそれぞれ、X U Sの条件で独立であるというOの部分集合Xが存在する。これはPOIPGでは、AとD間、BとD間、AとB間の辺がないことで表現されている。さらに、図5のDAGでは、Aから出る誘導経路とAから入る誘導経路があるが、すべてCから入る誘導経路であることに注意されたい。Equiv(Cond2)を表す**POIPGでは**、AとCの間に誘導経路が存在するため、AとCは隣接している。エッジのA側の "o "は、A-C間の誘導経路のA側の向きがわからないことを**意味し**、エッジのC側の">"は、Equiv(Cond2)のすべてのDAGのA-C間の誘導経路がすべてCに入ることを意味している。

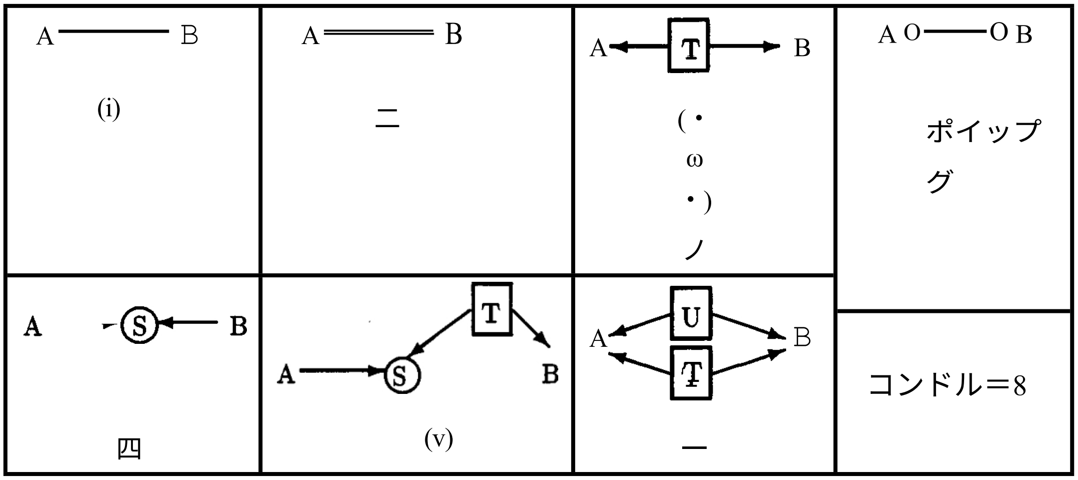


図4：Equiv(Condl)の一部のメンバー

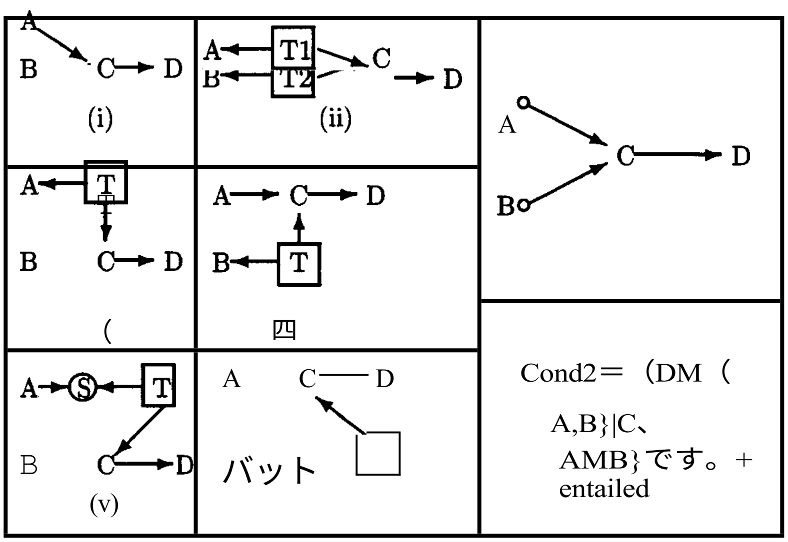


図5：Equiv(Cond2)の一部メンバー

Equiv(Cond2)のすべてのDAGにおいて、CとDの間の誘導経路はすべてCからDに抜けることが示されます。これらの事実は、POIPGにおいてCとDの間の辺がD端で「>」、C端で「-」を持つことで表現されています。

POIPGがエッジC -r Dを含むことから、Equiv(Cond2)のすべてのDAGにおいてSのメンバーを含まないCからDへの有向パスが存在すること、すなわちCがDの（おそらく間接的）原因であることが示されうる。また、Equiv(Cond2)のPOIPGは、Equiv(Cond2)のどのDAGにもDからA、DからBへの有向パスが存在しない、すなわちDはAまたはBの（直接または間接の）原因ではないことを伴うことが示せる。

4.3 例3

最後に、O = (A,B,C,D}, Cond3 = {DC{A,B}, As(C,D}}と、これらに付随する条件付き独立関係すべてを含む例を考える。 Equiv(Cond3)には、以下のようなDAGは存在しない。

V＝Oであることから、Equiv(Cond3)の各DAGは、どちらかの潜在変数を含むと結論付けられる。 (i)

と図6の(ü)は、Equiv(Cond3)の潜在変数を持つ DAG の例である。選択バイアスがない限り、これらの性質はEquiv(Cond3)のどのDAGでも成立することが示される。

ここで、選択バイアスを持つDAGも考えてみるとする。図6の(iii)は、Equiv(Cond3)に入る選択バイアスを持つDAGの例である。BとCの間の誘導経路は、図6の**すべての**DAGでBに入り、Cに入ることに注意。これは、Equiv(Cond3)の**すべての**DAGでBとCの間のすべての誘導経路がそうであることを示すことができる。ただし、Equiv(Cond3)の各DAGにおいて、CはBの子孫ではなく、BはCの子孫ではない。この性質は、潜伏変数や選択バイアスがある場合でも、Equiv(Cond3)のどのDAGでも成り立つことが示される。したがって、Cond3 の条件付き独立性が観測された場合、潜 在変数や選択バイアスがあっても、また、潜 在変数や選択バイアスの因果関係に関係なく、確実に結論づけるこ とができる。

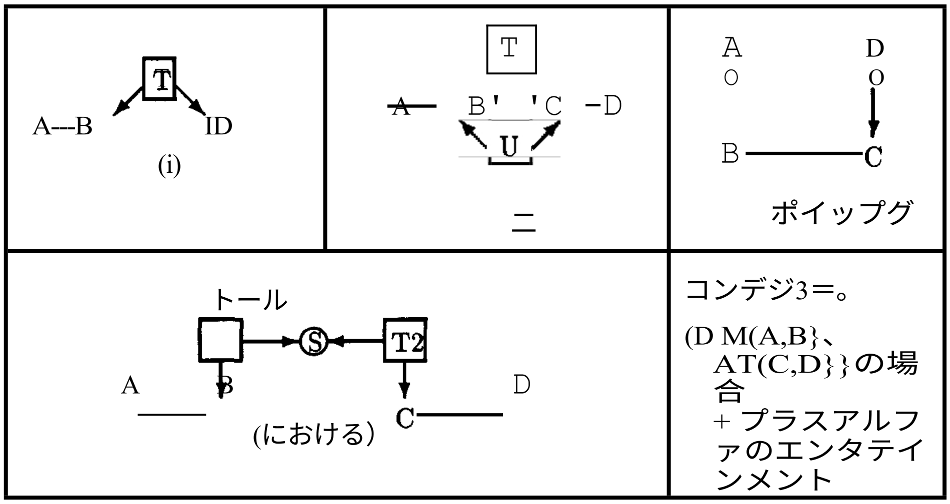


図6：Equiv(Cond3)の一部のメンバー

変数を他の変数に変換する場合、Cond3を生成した因果DAGでは、BはCの直接・間接の原因ではなく、CはBの直接・間接の原因でもない。

5 ぶんぽう POIPGの構築

POIPG は変数間の因果関係に関する貴重な情報を含んでいることがわかった。しかし、観測可能な条件付き独立関係の数はOのメンバー数に応じて指数関数的に増加する。さらに、依存関係の中には大きな変数集合を条件とするものがあり、妥当なサンプル**サイズでは**信頼性の高いテストができないことが多い。POIPGを構築することは可能か?

FCIアルゴリズム（Spirtes et al. 1993）は、Causal Markov Assump-tion 、 Causal Faithfulness Assumption、選択バイアスがないこと、独立性の関係が確実にテストできることの仮定で、正しいPOIPGを構築する。選択バイアスの可能性を許容する場合、Spirtes et at.(1993)に記述されたアルゴリズムは、依然として正しい出力を与える。(1993)のアルゴリズムはまだ正しい出力を与えるが、POIPGから引き出せる結論は、本論文の例と付録で説明した少し弱いものである。最悪の場合、FCI アルゴリズムは、任意の頂点が隣接する頂点の最大数を固定した場合でも、変数の数だけ指数関数的な時間を要する。しかし、シミュレーションデータ上では、真の**グラフが疎であれば**、最大100個の変数でアルゴリズムを実行できることが多い。これは、観測可能な条件付き独立関係の全セットを調べる必要がないためで、多くの条件付き独立関係は、他の条件付き独立関係によって包含されている。FCIアルゴリズムはこの事実を利用して、比較的小さな条件付き独立関係のセットをテストし、できるだけ少ない変数を条件とする依存関係をテストする。

6 APPENDIX

有向グラフは、頂点Vと辺Eの順序ペア（V,E）であり、各辺は、異なる頂点の順序ペアである。AからBへの辺がある場合、AはBの *一部*であり、AからBまたはBからAへの辺がある場合、AとBは隣接している。有向グラフGにおけるXq間の ***無向***ポトとは、l＜i＜nのとき、AiとA:+zがGにおいて隣接するような頂点（X , ... , Xq）の並びのことで、**Aiが**無向パスU**上の *コライダーとなる***のは、J'-iとAi+tからXiへの辺が
U.HiとAqの間の無向パスUは、以下のようになる
。
(とJ2 into（out of）A "の間のU上のエッジがある場合にのみ、Hiを*out* o/ Hiし、Aqについても同様である。有向グラフGにおけるHiからAqへの ***有向***パスとは、1＜*＜nのとき、GにおいてAiからAi+への有向辺が存在するような頂点（At、...、Aq）の並びである（パスは1つの頂点からなる場合もある）
。
頂点が複数回ある。有向グラフGのすべての有向パスが非周期的である場合、有向グラフはocpcJicである。AはBの *祖先* であり、BはA の ***子孫であるのは***
、A＝Bであるか、AからBへの有向パスが存在する場合のみである。

定理1は、Verma and Pearl (1991)の結果を一般化したものである。

定理1 *A DAiS* G O,S,LJ *etitoils that for all subsets X of O,* ***A is dependent on*** *B* ***9ioen (X*** U *iS) [A, B j* ***if and only iJ there is a*** **inducin9** *path be- tween A and B.*

DAG G'(O,S',L') は、Oのすべての不連続な部分集合X、Y、Zに対して、G'(O,S,L) が XMZ|Y U (S = 1) を伴う場合にのみ、Equiv (G(O,S,L)) に含まれます (S' = 1) 。

以下の定義では、記号 "*"をワイルドカードとして使い、部分配向誘導パスグラフwのあらゆる種類の辺の端点を表す。"*"自体は、この定義では決して出現しない。 **vは *部分配向誘導*____^9である。**

*有向非循環グラフ***G(O,S,L)***のパスグラフは、以下の*場合にのみ有効です。

(i) は、wの頂点の集合がOに等しい；

(ii) rにおいてAとBの間にエッジが存在する場合、それは以下の種類のいずれかである：A -r B、A o-+ B、A o-o B、またはA '-+ B。

(in) Equiv(G(O,S,L)) の各 (O,S',L') に対して A と B の間に誘導経路がある場合にのみ、inr に A と B の間のエッジが存在する；

(iv) の頂点AとBのペアの間には、最大で1つのエッジが存在する；

(v) A -* B が r にある場合、Equiv(G(O,S,L)) の各 Gt (O,S',L') において、A と B 間の誘導経路はすべて A から外れる；

(vi) A *-+ B が w にある場合、Equiv(G(O,S,L))の各 **Gi (O,S',L')** について、A と B 間の誘導経路はすべて B にある；

(vii) A *-* B *-* C が x にある場合、Equiv(G(O,S,L)) の各 **Gi(O,S',L')** に対して、U が A と B の間で B に入り、V が B と C の間で B に入るような誘導経路のペア U と V は存在しない。

(B-"AをA +- B、B o-+ AをA +-o Bと書くこともある)。

非公式には、POIPGの有向パスとは、同じ方向を指す"-r "個のエッジのみを含むパスのことである。

定理2 *rが部分配向のインティキリップ である* ***path*** *グラフであり、AからBへの有向パスUが存在す* ***in π,*** *とすると、POIPiS rを持つ*ecerp *DA61 G(O,iS,L)において、AからBへの有向ポスがあり、Aは*S*に降臨しない。*

定理3 *rが部分配向誘導パスグラフで、rにA +-+ Bがある場合、潜在変数があり、AからBへの有向パスがなく、POIPG zを持つDAG* G **O,S,LJ** *にBからAへの有向パスがないエンド。*

部分配向誘導経路グラフwにおけるAからBへの*半指向経路とは*、A-B間の無向経路非循環Uにおいて、どの辺もAを指す矢尻を含まない（すなわち、U上のAには矢尻がなく、XとYが経路上で隣接し、Xが経路上のAとYの間にある場合、XとY間の辺のX端に矢尻はない）。定理4、5、6は、一対の変数A、B間の因果経路上にどのような変数が現れるかについての情報、すなわち、それらの経路がどのように遮断されうるかについての情報を与える。

定理4 rをn個の*部分配向誘導経路グラフとし、Cのメンバーを含むu内のAからBへの半指向経路が*存在しない*場合、POIPG rを持つすべてのDAO* G{O,S,L,J *のAからBへのディレクト経路でCのメンバーを含むものは*S*のメンバーも含むことになる。*

定理S *rが部分配向誘導パスグラフで、*wに*AからBへの半指向パスが存在しない場合、POIPiS rを持つe "erp DAG* **G{O,S,LJ** *のAからBへのすべての有向ポスは*Sのn個の*メンバーを*含むことになる。

定理6 *rが部分性指向*誘導*経路グラフであり、e "er]j* 半方向*経路fi am A to Bが*'r*におけるCの**メンバーを含む場合**、POIPG rを持つすべてのDAG* **G{O,S,L}** *trのe "*erj/*方向経路jam A to Bはメンバ*ーa/ S U Cを含む。

謝辞

本論文の初期の草稿に有益なコメントをいただいた2名の匿名レフェリー、および有益な意見をいただいたClark GlymourとGreg Cooperに感謝したい。この論文のための研究は、Office of Navel Research grant ONR @N000l4-93- 1-0568の支援を受けました。

参考文献

Bollen, K. (1989). *潜在変数を用いた構造方程式* .New York： Wiley.

クーパー、G. (1995).選択バイアスが存在する場合のデータからの因果関係発見。In *Preliminary papers of the fifth international workshop on Artificial Intelligence and StatisticS,* Fort Lauderdale, FL、pp.140-150.

Lauritzen, S., A. Dawid, B. Larsen, and H. Leimer (1990).有向マルコフ場の独立性。*Networks 20.*

Meek, C. (1995).ベイジアンネットワークにおける強い完全性と忠実性。*人工知能における非誠実性(Uncer- tainty in Artificial Intelligence)議事録(Proceedings)。* To appear.

Pearl, 1. (1988). *インテリジェントシステムにおける確率的* Jteasoning.**San** Mateo: kforgao-Kaufznann.

Spirtes, P., C. Glymour, and R. Scheines (1993).アンティ*ーション、予測、および検索。*Springer-Verlagです。

Verma, T. and J. Pearl (1991).因果関係のある*モデルの*等価性と同義性。In *Proceedings of the Sixth C!onference* in *Art.Int.,* Mountain View, CA, pp.220-227.AIにおける不確実性のための協会（Association for Uncertainty in AI.

Wermuth, N., D. Cox, and J. Pearl (1994).一変量再帰回帰から導かれる多変量構造の説明国家。Technical Report Report 94-1, University of Mainz.

ライト，S. (1934).パス係数の方法。*数理統計学年鑑 5.*