因果関係、予測、検索

ピーター・スパーツ、クラーク・グリモア、リチャード・シャイネス

カーネギーメロン大学 哲

学科

私の両親、モリスとセシル・スパルテスへ - 追伸

ルシール・リンチ・シュワルツ・ワトキンス・スピード・ティンドール・プレストンを偲んで - C. G.

マーサへ、彼女のサポートと愛に対して - R.S.

数多くの*原因によって影響を受けたデータ扱う　のが*　　統計学の*主な仕事である。実験では、複合的な原因を切り離すことを目指します。することで を取り除く。　　すべて そのうちの一つを除いて を取り除く、*

というか　むしろ　を集中的に研究し、他の研究をできるだけ減らすことである。　　状況が許す限り　　が許す限り、　　にすることです。　比較的　小さな　残っている。統計データ　　否定された　この　資源である、受けなければならない　受け入れるのために　　分析しなければならない。多くの原因の影響を受けるデータを分析対象として受け入れなければならない。　を発見しなければならない。その　データ　　どの原因が重要なのか、そして観察された効果のうちどの程度がそれぞれの作用によるものなのかを、データそのものから発見しようとしなければならない。

--G.U.ユール、M.G.ケンドール 1950年

その　の理論　の　　エスティメーション　が論じている　その 原理を　を説明する。　　観測データを用いて、数値ではわからない理論量を推定したり、その値を明らかにしたりする原理を説明する。　　に入る　　に入る　私たちの　仕様　　　の　その　因果関係のある　を操作している。

--ロナルド・フィッシャー　　卿（1956年

ジョージ・ボックスは、「複雑な システム　　が乱されたときに何が

起こるかを知る唯一の方法は、乱すことである」と述べている（ほぼ）。

を乱すことだ。　を乱すことである、　ではなく　単にを観察することである。　観察するだけでなく 受動的に観察するのではなく　を受動的に観察するだけではない。"　この言葉は　という言葉は　注意の言葉　については　"自然な　実験"　は　不快なほど　強い。しかし、今日の世界では、どちらかといえば、弱すぎるものとして受け入れる以外に方法はないだろう。

--G.モステラ、J.テューキー、1977

因果推論は、統計学の問題の中でも最も重要で、最も繊細で、最も無視されている問題の一つである。

-- P. Dawid, 1979

目次

謝辞

本書のアイデアの源泉の1つは、私たちが10年前にピッツバーグ大学で始めた仕事にあります。世紀末のチャールズ・スピアマンのプロジェクトに始まり、ハーバート・サイモン、ヒューバート・ブラロック、ハーバート・コストナーの研究を含む、心理測定、経済、社会学の文献から、因果関係、統計、検索に関する多くのアイデアを引き出しました。

翌年に出版されたJudea Pearlの『*Probabilistic Reasoning in Intelligent Systems*』によって、我々はこの事業に対する新しい視点を得ることができた。パールの本は、発見を主目的とするものではなかったが、条件付き独立性と因果構造をいかに一般的に結びつけるかを示してくれた。この結びつきは、一般的で信頼できる発見手順を確立するために不可欠であることがわかった。それ以来、私たちはパールやダン・ガイガー、トーマス・ヴァーマとの文通や会話、そして彼らのいくつかの論文から利益を得ている。Pearlの研究は、Wermuth（1980）、Kiiveri and Speed（1982）、Wermuth and Lauritzen（1983）、Kiiveri, Speed and Carlin（1984）の論文を参考にしており、1980年代前半にはすでに因果推論の厳格な研究の基礎が出来上がっている。Paul Hollandは数年前にRubinのフレームワークを紹介してくれたが、有向グラフモデルとの論理的なつながりに気づいたのはつい最近である。さらに、J. Whittaker (1990)の無向グラフモデルの特性に関する優れた説明にも助けられた。

ピッツバーグ大学のGregory Cooperからは、データ、コメント、ベイズアルゴリズム

、ALARMネットワークの画像と説明を提供してもらい、多くのことを学びました。また、Kenneth Bollen 氏からは、長年にわたって有益なことを学びました。Chris Meekは、Rubin、Pratt、Schlaiferの様々な主張を有向グラフモデルの公理から導き出す重要な定理を得るために不可欠な手助けをしてくれた。

カーネギーメロン大学統計学部のSteve Fienbergと数名の学生がグラフィカルモデルに関するセミナーに参加してくれ、そこから多くのことを学ぶことができた。また、Elizabeth Slateには、Rubinフレームワークのいくつかの論文について案内してもらったことに感謝している。ナンシー・カートライトには、前著で採用され、今回も継続されているアプローチについて、丁寧かつ鋭い批判をしていただいた。彼女のコメントによって、私たちは『ルービン』のパラメータに関する研究に取り組むことになった。

第4章Brian Skyrmsには長年にわたる関心と励ましを、Marek Druzdzelには有益なコメントと励ましをいただいた。また、Linda Bouck、Ronald Christensen、Jan Callahan、David Papineau、John Earman、Dan Hausman、Joe Hill、Michael Meyer、Teddy Seidenfeld、Dana Scott、Jay Kadane、Steven Klepper、Herb Simon、Peter Slezak、Steve Sorensen、John Worrall および Andrea Woodyにも助けていただいた。また、リック・リントゥールスト博士と連絡を取ってくれたアーネスト・セネカに感謝します。特に、博士論文を公開してくれたリントゥールスト博士に感謝します。

私たちの仕事は、多くの機関によって支えられてきました。彼ら、そして彼らに代わって意思決定を行った人たちは、私たちの感謝に値する。カーネギーメロン大学、全米科学財団の科学史・科学哲学、経済学、知識・データベースシステムの各プログラム、海軍研究局、海軍人事研究開発センター、ジョン・サイモン・グッゲンハイム記念財団、Susan Chipman、Stanley Collyer、Helen Gigley、Peter Machamer、Steve Sorensen、Teddy Seidenfeld、Ron Overmanなどである。Navy Personnel Research and Development Centerのおかげで、多くの困難なデータ分析問題に触れることができ、そこから多くのことを学ぶことができた。

序文

本書は、分野を問わず、科学的な説明を得るため、あるいは行動、実験、政策の結果を予測するために統計的手法を用いることに関心のある方を対象としています。

G. Udny Yuleの研究の多くは、統計的サンプルからいつ、どのようにして因果関係を確実に推測し、その比較強度を推定するかを研究することを目的とした統計学のビジョンを示している。Yuleの研究は、Ronald Fisherの概念に取って代わられ、実験と非実験の間に根本的な断絶があり、統計学は無作為化実験がなければ因果推論に役立たないというものである。このような状況に対して、統計学関係者は時折、懸念を表明するが、それは当然であると我々は考えている。私たちの研究は、理論統計学の事業とその潜在的な実用的利益に関するYuleの概念のようなものへの回帰を意味します。

もし20世紀の知的歴史がそうでなかったとしたら、私たちの仕事が属する学問分野があったかもしれません。しかし、そうではありません。私たちは、統計学、コンピュータ科学、哲学に属する資料を開発しています。この組み合わせは、これらの科目の専門家にとって完全に満足のいくものではないかもしれません。しかし、私たちは、それがその目的にとって満足のいくものであることを望んでいます。私たちは、訓練された統計学者でもなければ、付き合いのある統計学者でもありません。社会科学や行動科学、疫学、経済学、市場調査、工学、そして応用物理学でさえも、無作為化実験から得られたものではないデータから因果関係を推測するために

統計手法が日常的に用いられ、政策や操作、実験の効果を予測するために標本統計が用いられているという事実に我々は衝撃を受けている。これらの用途がなければ、統計学という職業ははるかに小さなビジネスになっていただろう。統計学の専門家の多くは、このような利用によって繁栄している学問分野が、その聴衆にその利用が不当であると保証することを特に奇妙だとは思わないかもしれませんが、私たちには実に奇妙に感じられるのです。学問の外にいる私たちから見ると、このような目的への統計の応用に関する最も緊急な問題は、因果関係の推論や操作の効果の予測が信頼性をもって行える条件と行えない条件に関するものであり、最も緊急なニーズは、これらの問題に取り組むための原則的で厳密な理論にあります。

の問題です。彼らの著書の証言から判断すると、非常に多くの統計学者がそのような理論は不可能であると考えている。我々は、実験的試行以外の統計から原因を推論する可能性に反対する一般的な議論は不健全であり、実験と観察の研究デザインの原則を根本的に分離することは賢明でないと考えている。実験デザインと観察デザインは常に同じ推論を可能にするとは限りませんが、それらは統一された原則に従うものです。

私たちが開発した理論は、過去15年間に統計界で築かれた仮定から必然的に導かれたものです。理論の基礎となる構造は、基本的に公理的である。因果構造と確率分布の関係に関する2つの独立した公理を与え、そこから、様々な背景仮定の下で、統計的制約から確実に推論できる因果関係と予測の特徴とできない特徴を推論することになる。すべての公理のバージョンは、Lauritzen, Wermuth, Speed, Pearl, Rubin, Pratt, Schlaifer などの論文で見ることができます。ほとんどの場合、長期頻度を決定する予感として緩やかに考えることができる確率分布の観点から理論を展開するが、確率分布の多くは（規範的な）主観的信念度として理解することもでき、ベイズ的応用にも言及することがある。公理からは、推定、サンプリング、潜在変数の存在と構造、回帰、識別不能関係、実験計画、予測、シンプソンのパラドックスなどに関するさまざまな定理が導かれる。その中でも、因果推論によく使われる統計的手法が根本的に最適でないこと、そして、サンプルデータに基づいて行われた統計的判断の結果から因果関係を推測する、漸近的に信頼性が高く計算効率の良い探索手順が存在することを発見しました。(これから説明する手順は、確率変数の独立性に関する統計的判断を必要とする。このような手順が「漸近的に信頼できる」というのは、必要な統計的判断の結果がそれぞれ研究対象の集団で真である場合に、正しい情報を提供するということである)

序文

この本の大部分は数学である。公理が受け入れられるところでは、探索手順の存在を含む定理も受け入れられなければならないのである。私たちが説明する手順は、線形データと離散データの両方に適用でき、変数間の因果関係が十分に疎であり、かつサンプル

を十分に大きくする。これらの手順は、コンピュータプログラムであるTETRAD IIに実装されており、執筆時点では一般に公開されている[1]。

信頼性の高い発見方法の存在とその性質に関する定理は、それ自体、短期的な探索方法の信頼性については何も教えてくれない。我々が説明する方法は、予測不可能な一連の統計的判断を必要とし、それを仮説検定として実装したものである。このような場合によくあることだが、サンプルが少ない場合、個々の検定の*p値は*、探索手法の第1種の過誤確率を正しく推定できないことがある。我々は、モンテカルロ法を用いた模擬データに対する様々な手続きの広範なテストの結果を提供し、これらのテストは、シミュレーションの条件下での信頼性についてかなりの証拠を与える。シミュレーションは、私たちが説明するどの検索方法についても、その誤り確率を簡単に推定する方法を示している。また、本書では、医学研究者が救急医療の診断指標とその原因をモデル化するために作成した大規模な疑似経験的データセットと、非常に多くの経験的データセットの研究も行っており、そのほとんどは、仕様検索の文脈で他の著者によって議論されてきた。

さらに、因果関係と確率の関係を正しく理解することで、実験と観察の比較力、シンプソンのパラドックス、回帰モデルの誤差、レトロスペクティブとプロスペクティブのサンプリング、変数選択の危険性など、統計文献の多様なトピックを解明することができることを示すこともこの作品の目的である。私たちが考慮しない関連する話題も数多くあります。離散的な潜在変数による推定の問題、統計的決定の最適化、サンプリングデザインの多くの詳細、時系列、「非回帰的」因果構造--すなわち、フィードバックを伴うシステムの有限グラフ表現--の完全な理論などが含まれる。

序文
因果関係、予測、探索」は、教科書を意図したものではなく、関連する道具を備えていません。未解決の問題はあるが、演習問題はない。教科書では、たとえそうでなくても、すべてが完全で整然としたものであるかのように提示されるはずです。本書ではそのような見せかけは一切なく、各章には未解決の問題や未解決の疑問が豊富に含まれています。教科書は通常、あまり間を置かずに論点を整理するものだが、本書はかなり間を置いている。

本書の様々な定理は、グラフ理論的な性格を持つものが多く、その多くは統計学では全く馴染みのない種類の長くて難しい格論である。本書では

[1]UNIXオペレーティングシステムを搭載したワークステーションとPASCALコンパイラ、ネットワーク接続をお持ちの方ならどなたでもご利用いただけます。IBM互換の80-386および80-486パーソナルコンピュータ用に、より柔軟性の低いバージョンのプログラムも用意されています。Richard Scheines（RS2L@andrew.cmu.edu）までご連絡ください。

を除くすべての証明は、本書の巻末にある章に掲載した。詳細な証明が出版された文献にある場合は、読者にそれを紹介することにした。重要な結果の証明が出版されていない、あるいは容易に入手できない場合は、その実証を詳細に説明した。

本書の構成は次の通りである。第1章は、現在の統計実務の文脈で本書の動機に関わり、結果の一部を宣伝している。第2章では、この研究に必要な数学的な考え方を紹介し、第3章では、形式的な枠組みに因果関係の解釈を与え、公理を示し、それが失敗しやすい状況を指摘し、いくつかの基本定理を提供する。次の2章では、測定された変数に影響を与える測定不能な共通原因が存在しないことが知られている、あるいは仮定されている状況において、いくつかの基本的な問題に対する公理の2つの帰結を解明している。第4章では、因果関係の仮説が統計的に区別できないための必要条件と十分条件を、いくつかの意味ごとに図式化したものを示す。第5章では、統計学で一般的に推奨されるモデル指定手順の特徴を批判し、公理が適用され、測定されない共通の原因が働いていないと仮定して、母集団分布の特性から因果構造に関する正しい情報を抽出する実現可能なアルゴリズムについて述べる。このアルゴリズムは、様々な経験的サンプルとシミュレートされたサンプルに対して説明される。第6章では、第5章の分析を、測定されていない共通の原因が測定された変数に作用しないと仮定できない文脈に拡張する。理論的にも実践的にも、この章と次の章が本書の中心をなすが、特に難しい。第7章は、操作、政策、あるいは実験の効果を予測するという基本的な問題を扱っている。この章では、有向グラフモデルとドナルド・ルービンの予測分析のための「反実仮想」のフレームワークを簡単に結びつけることができる。第8章では、前章までの成果を回帰のテーマに適用する。標準的な統計的仮定が満たされている場合であっても、重回帰は大標本の限界においてさえ因果関係の影響を評価するための欠陥のある信頼できない方法であり、様々な自

動回帰モデル仕様検索は問題を悪化させるだけであることを論証する。第6章のアルゴリズムが原理的に信頼性が高いことを示し、様々なシミュレーションと経験的なデータセットで、様々な重回帰手続きに対してこれらのアルゴリズムの性能を比較する。第9章では、レトロスペクティブサンプリングとプロスペクティブサンプリングの問題、実験デザインと観察デザインの比較力、変数の選択、倫理的な臨床試験のデザインなど、前章の結果を踏まえて実証研究のデザインについて考察する。最後に、喫煙と肺がんをめぐる論争のいくつかの側面を振り返って、この章を締めくくります。第10章と第11章では、さらに線形ケースを検討し、測定されたデータ間の因果関係を発見したり、精緻化したりするためのアルゴリズムを分析する。

線形システムにおける測定不能な変数。第12章は、様々な未解決問題の簡単な考察である。証明は第13章で行われる。

私たちはこの作品を自己完結的なものにしようと努めたが、それは認めざるを得ないほど困難なものである。読者は、Pearl (1988)、Whittaker (1990)、Neopolitan (1990)を事前に読んでおくと、より効果的である。

表記 規約

テキスト

本文中では、各専門用語が定義されている箇所は太字で表記しています。

変数:	大文字、イタリック体、例: X
変数の値:	小文字、イタリック体（例: $X = x$
セット:	大文字で、太字で、例: **V**
変数の集合の値: 小文字、太字（例: $\mathbf{V} = \mathbf{v}$	
Xのメンバーで、**Y**のメンバーでない人:	**XY**
誤差変数:	ε, δ, e
XとYの独立性	$\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$
Zを条件としたXと**Y**の独立性:	$\mathbf{X} \perp\!\!\!\perp \mathbf{Y}\vert\mathbf{Z}$
X □**Y:**	**XY**
XとYの共分散:	COV(X,Y)またはγ_{XY}
XとYの相関関係：	ρ_{XY}
XとYのサンプル相関：	r_{XY}
XとYの部分的な相関関係	
セット**Z**のすべてのメンバーに対して制御する:	$\rho_{XY.\mathbf{Z}}$

私たちが考えるすべてのグラフにおいて、頂点はランダムな変数である。したがって、「グラフの変数」と「グラフの頂点」という言葉を同じように使うことにする

。

数値

図番号は、図のすぐ下の中央に記載し、各章の 1 から始まる。必要に応じて、測定された変数と測定されていない変数を区別するために、測定された変数をボックスで囲み、測定されていない変数を丸で囲む（エラー用語を除く）。e、 , で始まる変数は、「誤差」または「外乱」変数と理解する。例えば、下図では、Xと*Yは測定*され、*Tは測定されず*、誤差項である。

ε δ ε

表記規則

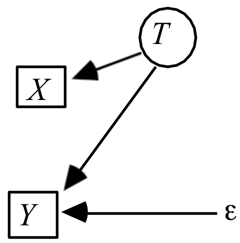


図1

図2のような測定変数と非測定変数の区別が不要なグラフでは、変数を枠で囲んだり、丸で囲んだりすることはしない。

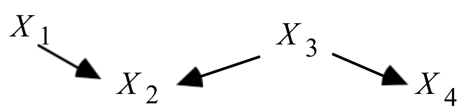


図2

ここでは、簡単のため、離散的な確率変数に対する確率分布について述べ、証明する。しかし、適切な可積分条件の下で、離散変数を連続変数に、確率分布を密度関数に、和を積分に置き換えることで、密度関数を持つ連続分布に簡単に結果を一般化することができる。

変数の集合の記述がグラフGと*G*内の変数の関数である場合、*G*を関数のオプション引数とする。例えば、**Parents**(*G*,*X*)はグラフ*G*の*X*の親である変数の集合を表し、文脈からどのグラフが参照されているかが明らかな場合は、単に**Parents**(*X*)と書くことにする。

分布が確率変数の集合**Oに対して**定義されている場合、分布を次のように呼びます。*P*(**O**)です。ランダム変数の分布間の方程式は、方程式の分布がすべて定義されているランダム変数のすべての値に対して真であると理解される。例えば、Xと*Y*がそれ

ぞれ0または1の値をとり、$P（X＝0）\neq 0$、$P（X＝1）\neq 0$のとき、$P（Y|X）＝P（Y）$

は、$P（Y＝0|\mathrm{X}＝0）＝P（Y＝0）$、$P（Y＝0|\mathrm{X}＝1）＝P（Y＝0）$、$P（Y＝1|\mathrm{X}＝0）＝$

$\mathrm{P}（\mathrm{Y}＝0）$、を意味します。

$= P(Y=1)$、$P(Y=1|\mathrm{X}=1) = P(Y = 1)$となります。

表記規則

次のような性質を持つ特殊な和算記号、$\vec{\sum}$を使うことがある：

(i) ランダム変数の集合が特別な和の記号の下に書かれている場合、和はランダム変数自体ではなく、ランダム変数の値の集合に対して取られることが理解される、

(ii) 条件付き確率分布がこのような総和記号の範囲に現れる場合、総和は条件付き確率分布が定義されている確率変数の値に対してのみ取られるものとする、

(iii) 特殊和記号の下に、その記号の範囲内の条件付き確率分布が定義されている確率変数の値が存在しない場合、和は0に等しくなります。

例えば、X、Y、Zがそれぞれ0か1の値をとるとします。そのとき、もし $P(Y=0,Z=0) \neq 0$

$$\vec{\sum_X} p(x|y=0,z=0) = p(x=0|y=0,z=0) + p(x=1|y=0,z=0)$$

ただし、$P(Y=0,Z=0)$=0とすると、$P(X=0|Y=0,Z=0)$と$P(X=1|Y=0,Z=0)$は定義されていないことになるので

$$\vec{\sum_X} p(x|y=0,z=0) = 0$$

ここでは、変数の空集合について、以下のような慣例を採用する。もし、$\mathbf{Y} = \emptyset$ ならば

(i) $P(\mathbf{X}|\mathbf{Y})$は$P(\mathbf{X})$を意味します。

(ii) $\rho_{\mathbf{XZ.Y}}$**はXZを**ρ意味します。

(iii) $\mathbf{A} \perp\!\!\!\perp \mathbf{B} | \mathbf{Y}$**はAを**意味する$\perp\!\!\!\perp$ **B**.

(iv) $\mathrm{A} \perp\!\!\!\perp \mathrm{Y}$はいつもそうです。

第1章

はじめに・広告掲載について

%0.1 課題

統計学の教科書には、因果関係を問う興味深い例が載っています：ハロタンはエーテルよりも手術死の原因になったのだろうか？カリフォルニア大学の大学院への女性の入学率が低いのは、女性に対する差別が原因なのだろうか？喫煙は癌の原因になるのか？実験計画、無作為化、重回帰における共線性、観察研究対実験研究など、統計教育における入門的、あるいは高度なトピックの多くに、原因の特定に関する問題がつきまといます。しかし、相関関係は因果関係ではない、という標準的な警告を除いては、教科書には因果関係と確率の関連についての体系的な議論はほとんどないに等しい。確率と統計的推論の数学は明示されているが、確率関係と因果関係の関連はほとんど黙示されている。少なくとも計量経済学以外の分野では、予測についても同じことが言える。教科書では、政策介入が問題となるケースを検討しているが、観測や実験の統計的分析と政策、行動、操作の効果の予測との間の関連について、体系的なことは何も教えてくれないのである。

さらに不思議なことに、多くの統計学の教科書は、自分たちが説明する方法は、無作為の非管理標本から因果関係を推論したり、操作の効果を予測したりするために

は信頼できるもの*ではないと*主張し、ありとあらゆる統計手法がこれらの目的には同様に無力であると述べたり示唆したりしています。もしこのような大胆な主張が広く信じられたら、このような主張が書かれた書籍の市場はほとんどなくなり、統計学という科目に対する関心も少なくともいくらかは失われるでしょう。例えば、線形回帰は、サンプルデータに直線を当てはめる手段として、また変数の新しい値を予測する手段として教えられています。しかし、回帰の実際の応用の多くは、回帰変数が*操作さ*れたとき、つまり、行動や政策によって回帰変数に新しい分布が強制されたときに、変数の値を予測することである。回帰の教科書の著者は、そのような用途がなければ

その本の読者の多くは失われてしまうだろう。控えめに言えば、回帰は実験以外のサンプルから原因を推測したり、操作されたシステムについて予測したりするのには使えないということである。

統計的標本から因果構造を推論することの可能性に反対する議論は、各書籍によって驚くほど一様である。例えば、Mosteller and Tukey (1977)は、*回帰が*誤った因果推論をもたらす事例を豊富に提供している。Rawlings (1988)は、応用回帰の教科書の主な例として、因果推論の問題を数章にわたって追求し、回帰はその目的には使えないと結論付けている。しかし、このような推論問題に対して回帰が最適に信頼できるという実証がなければ、これらの考察は、回帰を教えるのではなく、より良い方法を探すべきだということを主張しているに過ぎない。さらによくある教科書の議論は、回帰が因果関係の推論にできないことは、何もできないことを証明しようとするものである。この議論は、統計的依存関係による因果構造の決定が不十分であることを主張するものである。読者は、変数Xと*Yの*標本値の間に「関係」--つまり、何らかの統計的依存関係--が観察されるかもしれないと言われる（Younger, 1978; Rawlings, 1988）。

(i) *XがYを*引き起こすとき、

(ii) *Yが*Xを引き起こすとき

(iii) がそれぞれを引き起こす場合、

(iv) の両方を引き起こす第三の変数がある場合、

(v) サンプルが代表的でない場合、または

(vi) Xと*Yの*値が時系列を形成する場合。

このようなリストには、「原因を特定するためには、統計的関係の存在ではなく、実験が必要である」（Younger, p.176）という警告が必然的に付されているものもあ

る。この議論の論理的な強さを考えてみよう：

XとYの2つの変数だけを測定した場合、(i)(ii)(iii)(iv)を区別できないことは十分正しいが、追加の変数を測定した場合、これらの代替因果関係を区別できないことの証明は何であろうか。計量経済学の道具変数法（Bowden and Turkington, 1984）が有名な例ですが、他の文脈では、一対の変数に関連するパラメータの識別可能性や推定特性が、追加の変数を測定すると変化することが知られています。因果構造の特定も同じではないだろうか。否定的な答えを確信している教科書の数々は、その証明をどこに隠しているのだろうか。あるいは、(v)を考えてみよう。代表的でないサンプルへの訴えは、限りなく赤信号である。事実上、すべての統計的推定手続きは、サンプリングが不規則であれば信頼できない。

また、抽選の運が悪ければ、どのような推定法も迷走してしまう。統計的推定に要求されるような信頼性の要求を因果推論に適用するのは、何か不誠実な感じがします。(vi)のみが健全な指摘をしている。Yule(1926)は、増減する傾向のある変数の時系列サンプルは相関があることを指摘した。このような相関は、測定されない共通の原因によるものだと結論づけられるかもしれないが、Yuleは不特定の理由からその仮説を否定した。非定常時系列はいくつかの点で特殊であり、例えば、分布の混合を表している。このようなケースを除外する条件は何か、また、分布、あるいは分布と因果構造に関する特定可能な一般条件があり、それが得られると原理的に因果推論が可能になるのか、回帰や他の標準的な手法がこのような推論に最適なのか、といった疑問を持つ人もいるであろう。このような疑問は、教科書では決して提起されない。例えば、推定量が偏ったり矛盾したりする状況を知っていながら、そのどちらでもない状況を論じようとしない推定に関する教科書や、潜在変数モデルの中には識別できないものがあると指摘しながら、識別できないものはないと結論づける教科書をどう思うだろうか?

教育学は一般に受け入れられている統計理論を反映しており、いくつかの重要な例外を除いて、統計理論は因果推論に関する問題を回避するためにつま先立ちをしている。それでも、統計学の多くの研究課題は、その動機が細部に隠されていることもあるが、基本的には信頼できる因果推論に関する問題によって動機づけられている。

*倒れやすさ*に関する統計文献は、その一例である。数年前、National Halothane Study (Bunker, Forrest, Mosteller and Vandam, 1969)では、当時利用可能だった手法では分析できないほどの関連変数があることに気づいた。この問題は、代替麻酔薬の効果と

副作用に関する一連の因果関係の問題であった。ある変数*A*が別の変数*B*に及ぼす影響の存在と強さについて、**K**を適切に含むより小さな変数セット**K**（Aと*B*を含む）を分析することによって、同じ結論が得られるのはどんな場合か？ もう少し正確に言うと、Aと*B*を含む変数の集合**K**を分析することで、**K**に含まれないAと*B*の共通の原因があるとしても、変数*A*が*B*を引き起こすかどうかをいつ、どのようにして確実に決定できるのか、またそのような状況で、*A*を操作した場合の*B*への影響をいつ、どのようにして確実に予測できるのか、ということです。この問題はハロタン研究に限ったことではなく、因果関係を立証・評価しようとするほとんどすべての非実験的研究で生じる問題である。この問題は、非実験的な疫学につきまとうものである。喫煙が肺癌を引き起こすという疫学的な議論に対するFisher（1959）の批判は、悲観的な答えの一部であった。

Bishop、Fienberg、Holland（1975）の影響力のある本は、ハロセン研究の問題によって動機づけられ、潜在変数の存在下で因果関係を識別し推定するという問題を、まったく別の問題に置き換えた。ある変数をマージナルアウトすることによって得られたモデルの対数線形パラメータが、より大きな、マージナルアウトされていないモデルの対応するパラメータと同じであるとき、[1] 他の著者も、この崩壊性の問題の変形を生み出し、時には対数の線形形式から外れている（Asmussen and Edwards, 1983; Whittaker, 1990）．因果関係のある変数の集合に対する対数線形統計モデルが、変数の部分集合に対して折りたたまれるとき、折りたたまれた対数線形パラメータが、限界集合の変数間の因果関係について正しく特徴付けると考える理由はない。

「モデル選択」あるいは「仕様探索」は、因果推論に関する問題がひどく曖昧になっている研究分野である。対数線形モデル、構造方程式モデル、回帰モデルなど、統計的な「モデル」には、しばしば二つの異なる役割がある。一つは、変数の集合の中で可能な確率分布のクラスを制限し、その制限を満たす分布のファミリーをパラメトリック化する役割である。例えば、対数線形モデルは、任意のセルの確率の対数の線形展開において、特定のパラメータが消失することを指定することによって与えられる。この点で、仮説選択の重要性は、制限とパラメトリゼーションが、分布を理解し効率的に推定するのに役立つことである。計量経済学、疫学、市場調査などでは、確率分布を*変化させる*ような行動や出来事の影響について予測することが多いのですが、このようなモデルが持つもう一つの役割は、予測に役立つことです。これらは因果関係の主張であり、実際の確率分布の推定から導かれるものではなく、統計モデルの制約が表現される表現のさらなる解釈に依存するものである。因果関係の解釈とともに用いられる統計的仮説は、正しいか正しくないかのどち

らかに見えるが、その違いは重要である：喫煙は肺がんの原因になる。

したがって、我々素朴な部外者は、線形またはロジスティック回帰因子の選択、対数線形パラメータの選択、構造方程式のパラメータの修正などに関する研究は、確率分布の特徴の推定に類似した因果構造の推定問題の一種として、モデル選択を追求すると期待するかもしれない。理論的な研究としては、大標本数の極限で構造に関する正しい情報を与える手続きを見つけることができる条件を調査し、その誤差を特徴付けることが期待されるかもしれない。

---

1 崩壊性の疑問の発生について、Steve Fienberg氏の解説に感謝する。

の確率を計算し、そのような手続きの他の計算的、統計的特性を記述することである。しかし、統計学の文献には、この種の研究はほとんどないに等しい。その代わり、モデル選択と仕様の探索は、推定とは全く異なる扱いを受けてきた。テキストや論文には、特定の探索手順（例えばステップワイズ回帰）が失敗する可能性のある条件が書かれていることがありますが、探索方法は、その漸近的信頼性についての関連した保証がありません。モデルを特定するための一般的な方法の信頼性についての大規模なシミュレーション研究もあまりない。

統計学において、一方では統計的従属性、他方では因果的従属性の関係に関係する他の多くのトピックが、その関係の理論に導かれることなく扱われてきた。統計的に区別できない構造方程式モデル（Basmann, 1965; Stetzl, 1986; Joreskog, 1990）、シンプソンの「パラドックス」の議論、実験デザインの多くの問題（回顧的サンプリングと前向きサンプリング、実験におけるランダム化など）がそれである。

なぜ統計学の応用では因果関係の推論に関わることが多く、統計学の理論では因果関係の推論に関わることが少ないのでしょうか。因果関係と確率を結びつける数学的分析の試みを避ける理由として、因果関係という考え方は、数学が避けるべき多くの形而上学的な濁りを含んでいるからということがあげられる。ある人は因果関係を完全に確率関係で説明しようとし、ある人は因果関係を反実仮想の条件で特徴づけようとする。どちらが正しいのか、もし反実仮想的な性格付けが部分的にでも正しいのであれば、一体何を意味しているのか、誰が判断できるのだろうか。(しかし、未定義のものについては、厳密な理論が存在するはずはない、と考える。

統計理論における因果関係の軽視に関するこの説明は、2つの点で満足のいくものではありません。第一に、因果関係の「定義」がないからといって、統計学者が実験

的な文脈での因果推論について気楽に話すことはできないし、ある変数が他の変数を引き起こすという事実そのものが、実験によってその事実を*発見*するかどうかに常に本質的に依存するとは限らない。第二に、因果関係の概念は多くの点で曖昧であり、形而上学的な論争に巻き込まれているが、*確率の*概念も同様である。確率のどの解釈も、曖昧な反実仮想や不思議議な性質に訴えている。古典的な定義では、「等価な可能性」のケースに訴えた。限界周波数解釈は、現実の有限の経験的連続と想像上の無限連続を関連付けなければならない。主観的なベイズ解釈は、特に不明瞭な心理学的概念である*信念に*依存しており、複雑なケースでは、認知的に制限された人間にはインスタンス化できない信念の度合いの割り当てを必要とする（例えば、Fine、1973を参照）。文献にあるベイズモデルはほとんどない。

この事実は、確率は*理想的な合理的*行為者の信念の度合いであるという主張によって回避されることがあるが、これはまた別の非存在のクラスである。確率という基本的な概念が不明瞭であったからこそ、この考え方は大きな実を結ぶことができたのであり、そのプロセスの重要な一歩は、この理論に明確な数学的形式を与えることであった。その結果、まず特殊な確率分布とその性質を解析的に特徴付けることができ、その後、測度論、コルモゴロフ公理とその変種によって実現した。因果関係の概念は、17世紀以来、確率論の発展と並行し、動機づけられてきたものであり、科学的実践における確率の評価と密接に関係していることは否定できない。では、なぜ因果関係の概念に明確な数学的形式を与え、因果形式と確率形式の関係を科学的実践を反映する形で明示しないのだろうか。統計学の文献には、まさにそれを実現しようとする2つの試みがあり、それぞれ貴重なものであるが、それ自体では十分ではない。

変数の集合間の因果関係を数学的に表現することは、今世紀に入ってから統計学の文献に多く見られるようになった。Sewell Wright (1934)は、因果構造を表現するために有向グラフを使用した。グラフの頂点は変数を表し、ある変数から別の変数への有向辺は、最初の変数が2番目の変数に直接影響し、他のすべての変数を一定にしても妨げられない影響であるという主張を表している。それ以来、回帰モデル、因子モデル、連立方程式モデル、時系列モデルなどの表現に有向グラフが使われることがある。数学的な観点からは、有向グラフは、統計モデルが、片側に1つの変数が存在し、もう片側に変数の影響として扱われる代数方程式のセットで指定される場合に暗黙的に使用される。

それ自体は、有向グラフによる因果関係の数学的表現としては些細なことである。

興味深いのは、グラフ構造と確率の制約を結びつける何らかの条件が与えられたときである。方程式系と誤差変数の独立性の仮定は常に有向グラフを決定し、時には明示的に与えられ、相関と部分相関の制約を伴う（Simon, 1954; Blalock, 1961）、これは正規分布の独立性と条件付き独立性の制約と同じだからである。しかし、社会科学者は、グラフと確率分布を結びつける一般的な原理を明示せず、文献は、さまざまな特殊なケースの分析にとどまった（Blalock, 1971）。グラフ表現を用いずに、ライヘンバッハ（1956）やサプス（1970）を筆頭とする科学哲学者たちは、因果関係の概念そのものを統計的依存関係（サプスの場合は時間順序）の観点から分析することを試みている。明示的で一般的な数学的接続

しかし、有向グラフと確率分布の関係は、10年ほど前まで紹介されていなかった。

Kiiveri and Speed (1982)は、サイクルのない有向グラフで表される因果依存関係を、条件付き独立制約に関連付けた。彼らは、Yが*X*の原因ではなく、*X*が*Y*に影響を与えるとしても、*Y*の直接原因の中間集合**Zを通してのみで**あり、**Zの**変数が一定であれば、*X*の変動が*Y*の変動を生じない場合、XとYは**Zを**条件として独立であるという考えのいくつかの同等バージョンを策定した。この考え方を形式的に記述したものを*マルコフ条件と*呼び、本書ではそのような定式化の一つを用いる。KiiveriとSpeedは、様々な線形モデルで必要とされる相関の消失や部分相関に関する社会科学者の主張が、マルコフ条件から導かれることを示した。さらに彼らは、有向無サイクルグラフのマルコフ条件を満たす厳密に正の確率密度は、グラフの構造によって決まる「因数分解」を認めることを示した。有向グラフのマルコフ条件を満たす結合密度は、各変数に1項ずつ、グラフ内の親変数に対するその変数の条件付確率を与える項の積と等しくなければならない。したがって、離散変数の場合、グラフが

$$X \longrightarrow Y \longrightarrow Z$$

図1

を満たす必要があります。

$$p(x,y,z) = p(x)\, p(y\,|x)\, p(z|y).$$

グラフに適用されるマルコフ条件によって要求される条件付き独立性制約は、因数分解の結果である。このようにして、応用統計学で広く行われている実践が、エレガントな形式的基礎を与えられ、Wermuth（1980）らによって以前に行われていた分布の因数分解に関する統計的研究と結びついた。その後、マルコフ条件の結果や有向グ

ラフと分布の接続に関する追加の制約が多くの著者によって開発された（Wermuth and Lauritzen, 1983; Pearl, 1988; Wermuth and Lauritzen, 1990）。

統計学には、確率と因果関係を結ぶもう一つの糸がある。因果関係の概念と操作の効果の予測可能性との関係を強調するものである。これは、実験計画に関する研究から生まれたものである。1935年、Neymanは、ある単位で行われた実験から得られた結論が、反事実的であることを指摘した。

あるものはあるように扱われ、あるものは別のように扱われる。このような研究の結論は、しばしば、すべてのユニットが同じように扱われた*場合に*どうなるかについてであり、したがって、ある意味で、実験が調べなかった条件や、実際にサンプリングしなかった分布についてである。Donald Rubin (1974, 1977, 1978, 1986)と彼に続く数人の人々(Holland, 1986; Pratt and Schlaifer, 1988)は、因果関係の仮説を、値が観測されることのない確率変数のファミリーを仮定していると解釈しました。実験では、値が観測されないランダム変数は、ある治療を受けたユニットが、代わりに実験の別の治療条件を受けた*場合*、結果変数が*持つであろう*（しかし実際には持っていない）値を表します。例えば、ある治療があるユニットに適用された場合、そのユニットの治療は、他のユニットに同じ治療を適用した場合の結果に影響を与えないと仮定することができる。Rubinは、このフレームワークを用いて、実験計画における無作為化の重要性を主張し、治療が結果変数と共分散する変数の値によって決定される試験によって、ある変数の他の変数に対する効果を推定する方法を示している。プラットとシュライファーは、介入や操作の下での条件付き確率の不変性を予測するためのルールを与えた。変数Xが値x*を持つように強制*された集団におけるP（Y）*が*、操作されていないユニットの集団におけるP（$Y|X=$x）と等しくなるのはいつなのか。直感的には正しいが、そのルールは一般原理から明示的に導き出されたものではない。

私たちの視点から見ると、Rubin, Holland, Pratt and Schlaiferの理論は、本質的に、介入や操作の効果を予測できる特別なケースについての説明である。この作品では有向グラフの手法は使われていませんが、条件付き独立関係を大いに利用しています。そして、それがRubinのフレームワークをマルコフ条件と結びつけている。因果構造に対してマルコフ条件は、各変数X*の*親に対する条件付き確率という観点から、確

率分布の因数分解を提供します：$P(X \mid V1...Vk)$である。Xの「直接操作」とは、$P(X \mid V1...Vk)$ を他の分布$P^*(X \mid V1...Vk)$ に変更するが、元の因数分解における他の条件付き確率は変更しない介入と考えることができます。したがって、直接操作の結果は、P（$X \mid V1...Vk$）をP^*（$X \mid V1...Vk$）に置き換えることによって得られる、新しい因数分解を持つ新しい分布になる。そこで例えば、因果構造がX <- Z -> Yで、因数分解が$P(X \mid Z)P(Y \mid Z)P(Z)$ であるとする。ある介入によって、Xの分布が$P^*(X \mid Z)$に変化したとします。そして、その介入が集団のすべてのユニットに適用された場合、新しい共同分布は$P^*(X, Y, Z) = P^*(X \mid Z)P(Y \mid Z)P(Z)$ となります。この例では、Yの限界分布が変化しないので、このような介入のYへ*の効果は*、予測するのが些細なことです。

操作」の定義が適切であれば、マルコフ条件は、先ほど説明した原理の一般化を伴う。この定理が定式化されたことはないようだが、計量経済学におけるショックの扱いや、実験計画に関する多くの議論に、この定理の例が暗黙のうちに含まれている。Rubin（1977）の代替操作による微分効果の不偏推定に関する主張、PrattとSchlaiferのルールはすべて、問題となっている介入が「操作」の特別なケースであるという仮定から直ちに導かれる。第7章で示すように，彼らの分析は，介入がXを元の操作されていないシステムにおけるその原因から統計的に独立させるという特別な場合に適用される.

マルコフ条件は神から与えられたものではなく、本書の過程で説明する様々な理由で失敗する可能性があります。条件」に基づく推論の信頼性は、実質的な仮定が成立する場合にのみ保証される。しかし、条件は十分に弱いので、それが適用されると考える理由がしばしば存在する。本書のほとんどの研究では、マルコフ条件と、変数間の因果関係のグラフにマルコフ条件が適用されているために、変数間のすべての条件付き独立関係が発生すると仮定する、さらなる条件を組み合わせている。この仮定は忠実条件と呼ばれ、因果関係のグラフが確率分布に関連づけられるとき、そのグラフに適用されるマルコフ条件が、その分布で成り立つ条件付き独立関係の*すべてを*特徴づけるという主張として形式的に考えることができる。非公式には、忠実条件とは、条件付き独立関係が、パラメータ値の偶然ではなく、因果構造に起因するという仮定と考えることができます。変数が線形関係にあると仮定した特殊なケースの結果を調べることもありますが、そのような場合には、パラメータに対する「自然な」確率分布に対して、忠実性の条件のバージョンがほぼ常に成立することを示すことができます。

統計と因果関係に対するグラフと実験計画法の統合の直接的な価値は美学的なものである。しかし、美学はそれどころではない。統一理論の本当の価値は、多くの研究テーマを再構成して、基本的な問題をより直接的に扱うことができるようにすることと、実用的な推論問題に対する新しいアルゴリズム技法を提供することにある。

マルコフ条件と忠実条件、そしてその結果は、統計学の応用において基本的であると考えられる因果関係についての多くの疑問を分析するための枠組みを提供するものである：回帰法はいつ因果関係の推論に使えるのか? 回帰法はいつ因果関係の推論に信頼できるのか、どのような条件で最適なのか。より信頼性の高い方法はあるのか? 様々なモデル選択手続きは有効か?

信頼できるのか? より信頼性の高い方法はあるのか? 2つの代替的な因果構造が統計的に区別できないのはどのような場合か? 統計から潜在変数の存在を検出できるのはどのような場合か? 因果構造が不明な場合、信頼性の高い予測は可能か? 測定された統計的依存関係を作り出すために作用する潜在的変数が存在するかどうかが事前に分からない場合、信頼性の高い予測を行うことができるか? 因果推論の信頼性はサンプリング手順にどのように依存するか? レトロスペクティブ・サンプリングはプロスペクティブ・サンプリングと同様に因果推論に有効か? 因果構造について、実験ではできるが観察ではできない、あるいは、信じられないことだが、その逆の判別は可能か? どのような付加的な仮定のもとで、潜在的な構造のどのような特徴を確実に推論することができるのか? グラフと実験デザインの枠組みを統一することで、これらの疑問のすべてに関連する結果を得ることができます。その結果は、統計的な常識のバージョンであることもあるが、むしろ多くの場合、既成の意見を覆すものである。我々が得た結果は多くの点で不完全であるが、因果関係、予測、探索の問題が多くの統計的問題の基礎となる研究分野を形成していることを示すには十分であり、統一理論の中では信頼性に関する本質的な質問を無視する必要も、エラソーな技術的質問に頼る必要もないことを示すものであった。

この理論の実用的な結果には2種類ある。統一理論は、様々な因果関係の推論が信頼性をもって行えない正確な条件をもたらし、統一理論は、統計的決定から因果構造を推論するためのアルゴリズムをもたらし、そのアルゴリズムは、適切な仮定の下、漸近的に信頼できることを特別だが明確に定義できる意味で示すことができる。このような証明があれば、統計学は計算機科学と協力し、計算上実現可能で、実現可能な統計的判断の集合を必要とする信頼できるアルゴリズムを得ることができる。これらのアルゴリズムをコンピュータに実装することで、小・中標本に対する手続

きの信頼性と、仮定違反に対する頑健性をシミュレーション手法で調べることができる。理論から導き出したアルゴリズムのほとんどは、TETRAD IIプログラムの実験バージョンに実装されており、様々なシミュレーションテストの結果が本書の中で紹介されている。

%0.2　広告掲載について

このセクションの各例は、他の章でも扱われています。ここでの役割は、多くの数学を含む長い本を読みこなす忍耐力を読者に呼び起こすことである。しかし、これらの例は、それ自体、何かの信頼性を証明するものではありません、

その多くは、マルコフ条件と忠実性条件と正しい統計的判断を行う能力だけを仮定して、後に信頼性を証明する手順を示している。

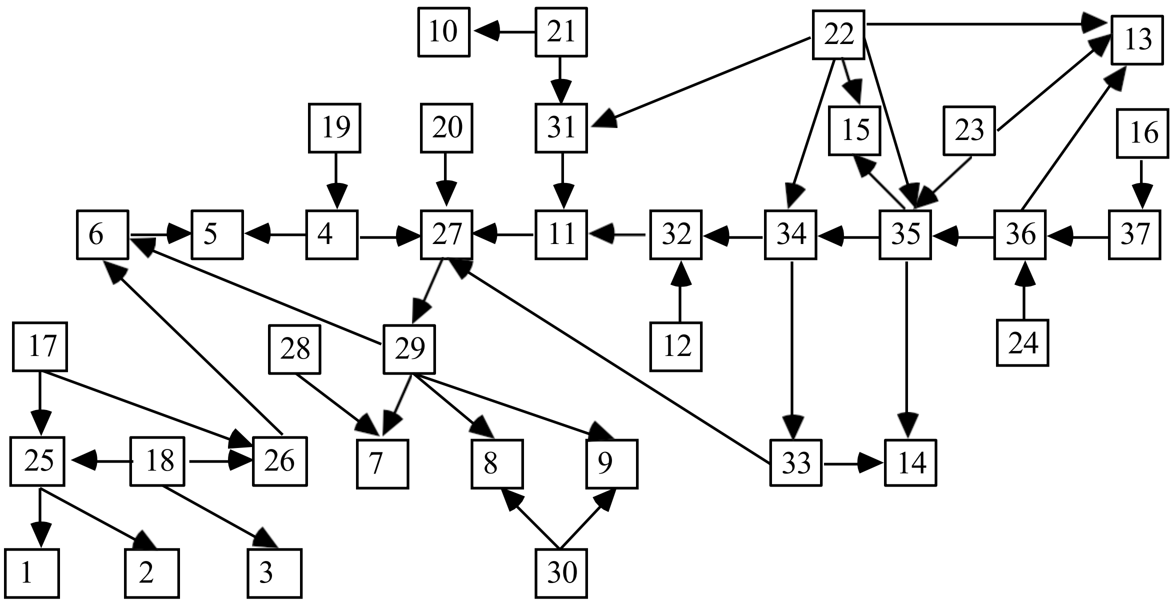
%0.2.1　データからのベイズネットワーク

下図はALARMネットワークと呼ばれ、救急医療システムをモデル化するために開発されたものである（Beinlich, et al. 1989）。変数はすべて離散的で、2、3、4個の異なる値をとる。ほとんどの場合、矢印は、ある変数が他の変数の原因とみなされることを示す。ネットワークを構築した医師はまた、ネットワークに確率分布を割り当てた。各変数*V*は、*V*に向けられたエッジを持つ変数の値の各ベクトルを条件とする確率分布を与えられる。このようなネットワークの1つの使用法は、ネットワーク内の他の変数の値（そのユニット）の測定から新しいユニットに対するいくつかの変数の値の確率を計算する。もちろん、このようなネットワークは離散データの統計モデルでもあり、対数線形モデル、ロジスティック回帰モデル、無向グラフ独立性モデルなどの代替となるものである。有向グラフは37個の変数と46個のエッジを持つ。Herskovits and Cooper (1990)は、この図を用いて、20,000人の救急医療統計のシミュレーションを行った。

多くの統計的検索手順では、37は無限大であるのと同じかもしれません。一般的な形式論で離散データのモデルを探索するために一般的に推奨されている手順は、10または12の変数で止まってしまう。これに対して、HerskovitsとCooperは、因果の方向と一致する変数の事前線形順序が与えられた場合、離散標本統計からこのグラフの隣接関係を回復する際にたった1つのエラーを起こす高速ベイズ手順を述べている。我々は、ALARMネットワークのほぼ全て（隣接関係とエッジの方向の両方を含む

）をサンプルデータから回復する手順について説明する。例えば、グラフの依存関係を線形とみなして生成されたデータから、15秒以内にネットワークの大部分を復元する、通常のワークステーション上での実装がある[2]。同じネットワークから生成されたHerskovitsとCooperの離散データでは、数分で構造の大部分を同等の信頼性でコンピュータによって復元することが可能である。確率の最尤推定値は直接得ることができる。これらの研究では、グラフに関する事前情報は一切与えられず、ほとんどの場合、コンピュータが矢印の*方向を決定している*ことを強調しておく。

2Decstation3100です。



ALARMの信念ネットワーク

KEYです：

1 - 中心静脈圧

2 - 肺毛細血管楔入圧
3 - 左室不全の病歴
4 - 全周囲抵抗
5 - 血圧

6 - 心拍出量

7 - 血圧から得られる心拍数

ニター
8 - 心電図から得られる心拍数
9 - オキシメーターから得られる心拍数

10 - 肺動脈圧
11 - 動脈血酸素
るものです。

12 - エンリッチド・エスオーピー
13 - 換気圧31
14 - 呼気中の二酸化炭素含有量

15 - 分量、測定値
16 - 分量、計算
17 - 低ボリューム血症

18 - 左室不全

19 - アナフィラキシー37

20 - 麻酔・鎮痛が不十分

な場合

21 - 肺塞栓症
22 - 挿管状態
23「換気チューブのよじれ
24 - 換気チューブが外れている

25 - 左心室終末-拡張期容積

26- 脳卒中量モ

27 - カテコラミン値
28 - 心拍数測定における誤差の原因
低心拍出
29 - 真の心拍数
飽和度30 - 心拍数の測定誤差は、以下の理由によ

電気メス

- シャント
32 - 肺動脈酸素量
飽和状態
33 - 動脈炭酸ガス含有量
34 - 肺胞換気量
35 - 肺換気

36 - 気管内チューブで測

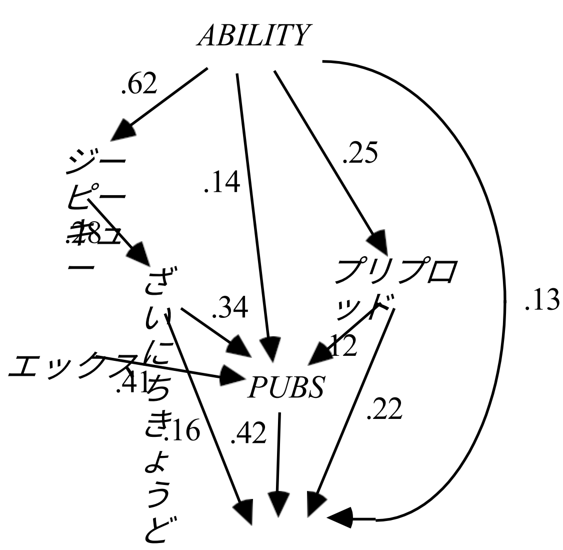
定した換気量

- 換気装置で測定した分換気量

図2

%0.2.2　データからの構造方程式モデル

ロジャースとマラント（1989）は、なぜある学術心理学者が他の人より多く出版するのかに興味を持った。大学院に入る前にすでに持っている能力なのだろうか。性別？大学院でのトレーニングの質？大学院で形成された出版習慣？最初の仕事の質？ロジャースとマラントは調査を行い、これらの変数と他の変数のそれぞれについてデータを収集した。常識的に考えて、彼らの変数間で起こりうる因果関係のほとんどの方向が決定される。例えば、性別と他の変数のいくつかが因果関係にある場合、他の変数が性別を引き起こすからということはありえない。最初の仕事の質や、卒業後の論文発表率が、その人の大学院の質を引き起こすことはありえない。ロジャーズとマラントは、経済学の2つの理論を適用して2組の連立方程式を作り、社会学の別の理論を適用し、さらに社会心理学の別の理論を適用して、それぞれの連立方程式を推定してテストしました。これらの理論は、それぞれ適合性テストに失敗した。ロジャースとマラントは、これらの否定されたモデルで仮定された依存関係を組み合わせ、さらにもっともらしいと思われる依存関係を追加して、新しいモデルを得、推定、テストし、11の直接的因果関係および関連する標準化係数のグラフとして提示した：



サイテス

図3

GPQ＝大学院のプログラムの質　*PUBS*＝大学院卒業後の出版物

QFJ = クオリティ・オブ・ファースト・ジョブ　*CITES*＝引用の必要性

PREPROD＝大学院での論文発表

RodgersとMarantoは論文の冒頭で、因果関係の依存関係は確率からは推論できず、"理論 "からしか推論できないので、手の込んだ演習が必要であると述べている。しかし、常識的な変数の時間的順序を考えれば、コンピュータが検索アルゴリズムを実行すれば、数秒後には、11の仮説的な接続のうち10をデータから直接生成し、最小の推定線形係数を持つ3つの依存関係のうち1つだけを除外することができます。

第5章と第8章では、理論的根拠に基づいて長々と正当化されてきた線形モデルの特徴が、我々が説明する手順によって経験的な共分散から直ちに生み出される様々な事例を紹介している。また、社会科学文献にある様々な結論がデータによって支持されないことを示す、もっともらしい代替モデルをアルゴリズムが生成するケースについても説明する。

%0.2.3　リグレッサーの選択

回帰は、政策の効果に関する予測を得るために、実証研究で最もよく使われる方法である。多くの場合、ほぼ同時に測定された変数から、測定されない要因が結果変数と1つまたは複数の回帰変数に影響しないという保証がない場合が多い。回帰は、このような役割を果たすと、何かと厄介であり、教育学や政策をめぐる議論において、回帰が重要視されていることを正当化するのは難しい。

誤差項と未測定変数Tを持つ一次方程式と、母集団の各単位に想定される因果依存関係を示す有向グラフで与えられる次の例を考えてみる。方程式は

$$Y = a1X_1 + a2X_5 + a3T + \varepsilon_y$$
$$X_1 = a4X_2 + \mathrm{a}5X_4 + \varepsilon_1$$
$$X_3 = a6X_2 + a7T + \varepsilon_3$$

誤差項を表さない有向グラフは、次のようになります:

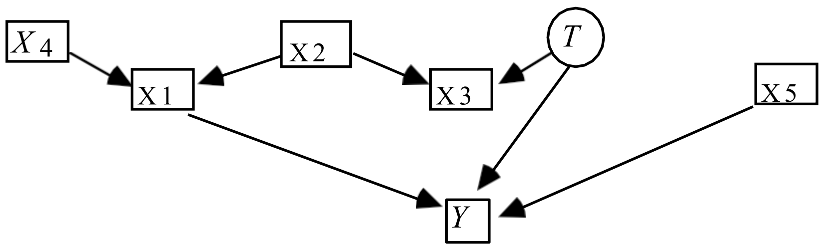


図4

定義づけのために、誤差項を含むすべての外生変数が独立で、平均がゼロ、分散が単位で正規分布し、aiはいずれも消失しないと仮定する。このような母集団で*Y*を*X 1*-*X 5*に重回帰させることを考える。*X 3*の回帰係数は、*X 3*が*Y*に及ぼす影響を信頼できない（偏った、矛盾した）推定値として提供する。この事実は、教科書でしばしば指摘されるが、通常は因果関係について明示的に言及しない。しかし、因果関係の明示的な議論を避けているためか、教科書[3]は、*ある回帰因子Xと結果変数Yが測定不能な共通原因を持つ場合、Xに直接影響を与えるか、Xと測定不能な共通原因を持つ他のすべての回帰因子のYへの影響度の推定値が同様に信頼できなくなる*という、より重要な事実を見落としています。図4の例では、*X 1*と*X 5*だけが、他の変数に媒介されない*Y*の直接原因であるにもかかわらず、*X 4*の回帰係数だけがゼロになる。同じような現象は、回帰因子の1つが結果変数の原因ではなく、むしろ効果である場合にも起こりうる。本書の後半でシミュレーション・スタディで示すように、回帰のための通常のコンピュータによるモデル選択手法は、このような場合、通常失敗します。実際、大きなサンプルを使ったシミュレーション・スタディでは、これらの手法は、単にすべての変数で回帰して有意な回帰因子を選択するより悪い結果になります。

統一理論は、このようなモデルが使われるときに一般的に行われる統計的仮定（線形性、均等分散性など）のもとで、回帰器構造に関する正しい情報を漸近的に生成

するアルゴリズムを導き出すものである。先ほどの例では、この手順により図5のようなものが得られ、これを「部分指向性誘導経路グラフ」と呼ぶ：

[3]回帰に関する約30冊の教科書を非ランダムに調査したところ、どの本にもこの事実が記載されていなかった。

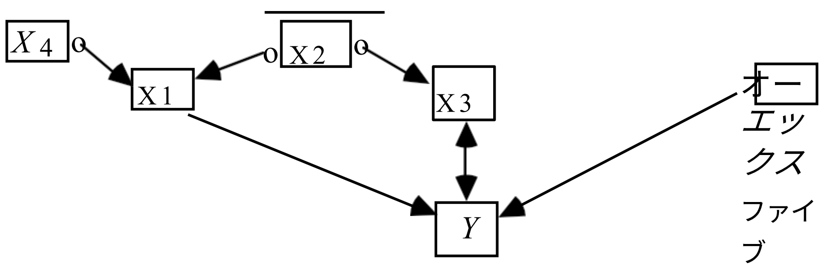


図5

両頭の矢印は、測定されていない共通の原因があることを示し、辺の端にある丸は、アルゴリズムがその端に矢じりがあるべきかどうか、つまり、o->辺が->なのか<->なのか判断できないことを示します。X_2の上のバーは、X_2に隣接するエッジの小さなoマークが両方とも矢じりである可能性がないことを示す。もし、マルコフ条件と忠実性条件が保証される文脈で得られたなら、このグラフは、$X1$がYの直接原因であること、X_5がYの直接原因かもしれないこと、$X3$がYの原因ではなく、YがX_3の原因ではないこと、$X3$とYが共通の原因の結果であること、他のX変数がYの直接原因ではないこと、を結論づけることができるであろう。

8章で詳しく説明する、2つの経験的な例を考えてみよう。1つ目は、6,000人以上のテスト結果のサンプルで、7つのテストスコアと、7つのテストスコアのうちの3つ、およびデータに含まれていない他のテストの平均である複合スコア、*AFQTの*値が、各被験者の値として含まれています。したがって、*AFQTの*他のテストに対する依存性は線形である。問題は、7つのテストのうち、どれが*AFQTの*構成要素であるかを特定することである。*AFQTを*他の7つのテストの得点に線形重回帰すると、すべてのテストに有意な回帰係数が得られる。これに対して、マルコフ条件と忠実条件から導くアルゴリズムは、*AFQTの構成要素である*3つのテストを正しく見つけることができます。この場合、変数が測定されていない共通原因の複雑な構造によって関連しているため、おそらく回帰は失敗する。

はじめに

Rawlings（1988）は、Cape Fear EstuaryのSpartina草の45のサンプルについての研究（Linthurst, 1979）を記述している。スパルティナのバイオマスの他に、成長に関連すると考えられる14の変数が測定された。バイオマスを14の他の変数に線形重回帰したところ、銅とカリウム濃度の2つの栄養変数だけが有意な係数を持ち、生物学的には妥当でない結果が出た。同じデータで、代替アルゴリズムでは、pHがサンプルの成長を制御していることがわかり、この結果は実験によって確認されました。この場合、重回帰は失敗する可能性があります。なぜなら、各回帰係数の有意性検定では、他の13の変数をコントロールする必要があり、その結果、重回帰係数は実質的に減少してしまいます。

サンプルサイズが45から32になり、部分相関があまり大きくない代替案に対してあまり力を発揮しないテストになります。一方、私たちのアルゴリズムでは、このような場合、1つ以上の他の変数を制御して、部分相関が消失することを検定する必要がありません。

%0.2.4　　実験を伴わない因果推論

Fisher（1959）は、喫煙と肺がんに関する疫学文献を批判する中で、2つの変数の相関関係では、直接的な効果と測定されていない共通の原因を区別することはできないと強調した。この不満は、Brownlee（1965）が、*喫煙と健康に関する*最初の*外科医総長の報告についてアメリカ統計協会誌*に発表したレビューでも同じことが述べられている。FisherもBrownleeももちろん正しいのですが、図6に示すようなケースを考えてみましょう：

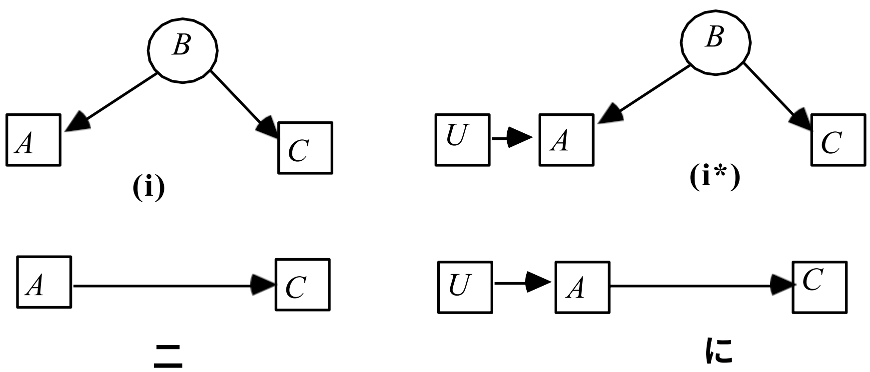


図6

*B*が未測定であるとする。フィッシャーの主張は、Aや*Cに関する*非実験的な研究では、たとえ母集団のすべての単位が全く同じ因果構造を持つとしても、仮説（i）と仮説（ii）を区別できないが、実験では区別できるということである。しかし、フィ

ッシャーが素通りした点もある。

変数Uが測定され、*A*を引き起こす（または*A*と共通の原因を持つ）場合、アスタリスクのついた図のように、*UがAを通してのみCに*影響を与える場合、構造(i*)、(ii*)は実験コントロールなしで区別することができます。*Uに関する*これらの制約は、データとは無関係に知られる必要はなく、分布を生成する因果過程について真であることが必要である。

ブラウンリーは、喫煙が癌の原因であり、喫煙と癌には測定不能な共通原因もある、というケースを強調した。第9章で見るように、そのケースには、以下のような特徴がある。

は興味深い複雑さを持つが、適度な予備知識があれば、適切な観察によって構造(i)および(ii)と区別することが可能である。

一般に、非実験的な証拠では、測定された統計的依存性が測定されていない要因によるものかどうかを判断することはできないという前提があります。このような判断は、実際には、本当の知識よりも、利便性や学問の慣習に左右されることが多いと思わざるを得ません。しかし、一般的な条件下では、連続的な線形関連変数であれ、離散的な変数であれ、この仮定は誤りである。測定不能な要因が働いているかどうかという事前の仮定なしに、因果構造に関する詳細な推論が可能な場合があり、同じような場合、因果構造を変えずに特定の変数を直接操作する政策の効果について、信頼できる予測が可能な場合もある。例えば、長方形の変数が測定され、楕円の変数（*環境汚染*と*遺伝子型*）*が測定されていない*、*測定された呼吸機能障害に*関連する以下の全く仮説的な因果構造を考えてみる。

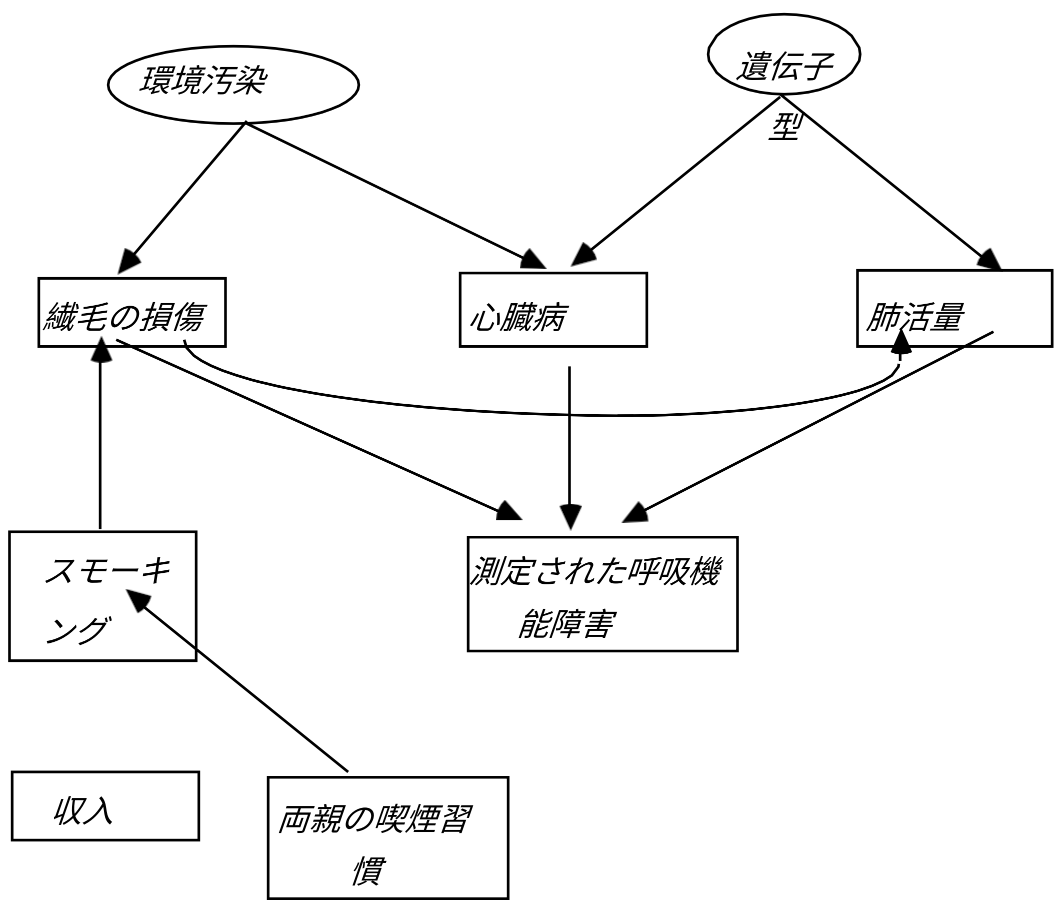


図7

図7の構造が、多項分布に従ってサンプルを抽出した集団における、これらの変数間の真の因果的依存関係を表しているとする。マルコフ条件と忠実条件は、測定された変数間の条件付き独立性と依存性の関係から、図8のグラフを生成するアルゴリズムをもたらします。

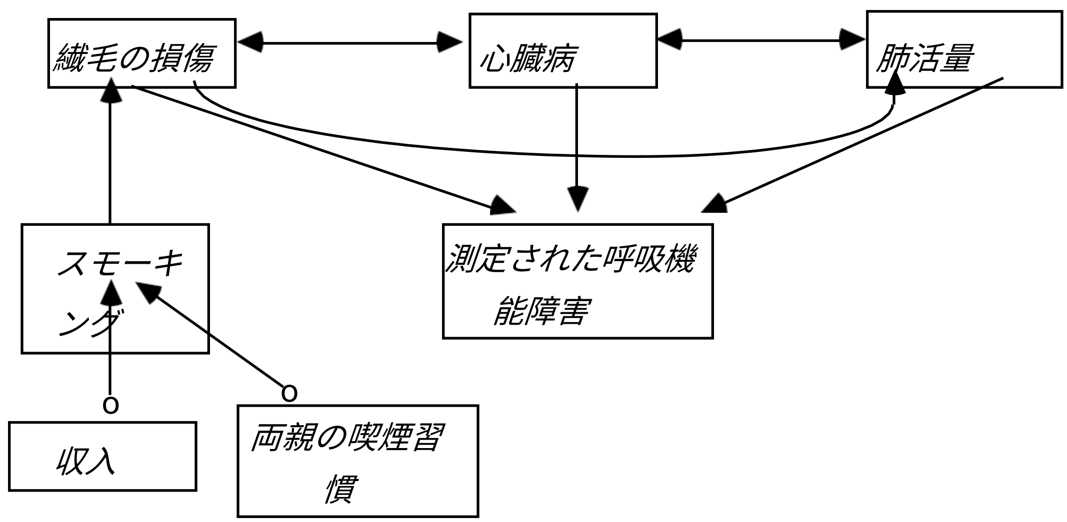


図8

両頭の矢印は、測定されていない共通原因の存在を示し、2つの辺の端にある円は、アルゴリズムがそれらの端に矢印があるべきかどうかを決定できないことを示す。この仮想的な例では、喫煙が間接的に肺活量に影響するという結論を導くのに統計学は十分であり、測定されていない因子の存在や非存在を事前に仮定することなく、変数間の因果依存関係についての他のほとんどの質問にも統計学は答えている。図8の図に含まれる集団に関する情報が与えられ、十分なデータがあれば、繊毛損傷、心臓病、肺活量、呼吸機能障害に対する禁煙の効果を確実に予測することができる。このアルゴリズムのバリエーションは、連続的で線形に関連する変数に適用される。

%0.2.5　観察されないものの構造

心理学的、社会学的テストや質問紙を使用する調査者は、それ自体には何の興味もない何百もの項目の回答を持つことが多い。興味あるのは、その項目が示す人物やシステムの特徴の性質と因果関係である。様々な項目がクラスターを形成しており、そのクラスターには共通点がある、と実質的な理由やその他の理由で考えられることがある。

測定されない原因同じクラスターに属する項目の中には、クラスターの全メンバーに共通する原因以外に、他の共通の原因がある場合があり、クラスターの一部のメンバーは、そのクラスターに属するかどうかにかかわらず、他の項目に対する反応によって影響を受ける場合があります。直線性などの単純化された仮定がなされたとしても、クラスターに関する仮定とデータから、それぞれのクラスターの測定されていない原因の因果関係について信頼できる結論を引き出すことは、全く絶望的だと思われがちである。しかし、変数が近似的に線形関係（または二項関係）であり、クラスターが*あまり*混同されていなければ、未測定変数間の因果関係に関する信頼できる情報を得ることができる。これらの因果関係が十分に疎であり、モデリングの仮定が適用される場合、因果構造をほぼ一意に特定することができる。そのために必要な手続きは完全に自動化されている。図9のグラフを考えてみよう。このグラフでは、16項目の回答のうち5項目（Y_1 , Y_2 , Y_4 , Y_5 , Y_{13} ）が混同されている。

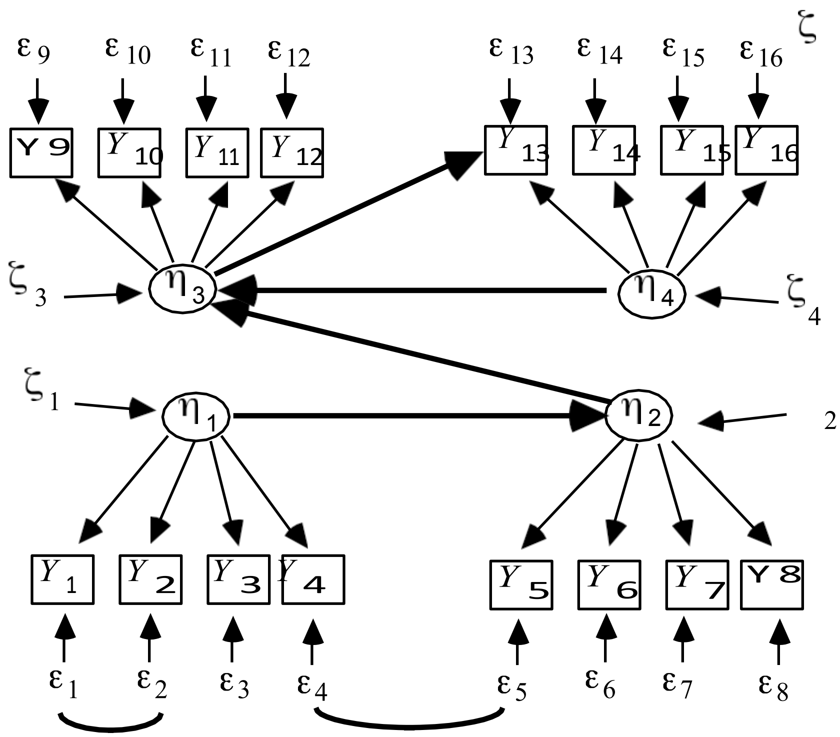


図9

はじめに
Y変数と4つのクラスター間の相関のみを入力とすると、これから導く手順は、潜在変数（「」変数）間のグラフィカルな接続を、統計的に区別できない$_{1\text{-}2}$接続の代替順序まで正しく再構築することができる。

%0.3 テーマ

本書の基本テーマは、因果依存性と確率依存性の間に系統的なつながりがあることである。このつながりを利用することで、信頼性の高い因果推論を行うための限界を理論的に確立することができる。因果推論において何がどのように間違ってしまうかは、因果性と確率の結びつきを理論的に理解することで初めて明らかになる。重要な制限の1つは、代替的な因果構造のグラフ理論的表現間の等価関係の特徴を証明することによって確立することができる。事実上、有向グラフの分類を、統計的区別不能の様々な関係まで調べることになる。一方、ある確率分布と一致するすべてのグラフのクラスの特徴付けを考えることができる。また、因果関係と統計的構造の間の典型的な（我々の場合は公理的な）接続が破られる状況を特徴づけることによって、因果推論の限界を調べることができる。この研究は、今世紀初頭にYuleとPearsonによって始められた。この限界の中で、サンプルの統計的性質から因果構造を探索する手続きの信頼性と計算機的性質を厳密に調べることができる。よく定義された手順の漸近的な信頼性は数学的に決定され、短期的な信頼性はシミュレーション研究によって実験的に推定される。その結果、モデル仕様探索の理論は、推定理論の特徴を持つようになった。

現在の統計的なモデリングや検索手順のほとんどは、かなり小さなモデルしか分析できません。例えば、対数線形モデルの一般的な探索・推定手順は、現実的なサンプルサイズでは、変数が想定する値の数に応じて、約12個かそれ以下の数で停止します。探索の効率を考慮すれば、変数間の因果関係が疎な場合に、100個以上の変数を持つ線形システムに対して、（明確な仮定の下で）信頼できる因果推論を行うア

ルゴリズムを見つけることが可能である。スパース構造上の離散変数では、数千単位のサンプルで50以上の変数を持つモデルを確実に同定することができる。

潜在変数に関して、統計学的実践はほとんど分裂状態である。統計モデルは、未測定変数を含むか含まないかのどちらかで提示されるが、統計的考察から未測定変数を含むか含まないかについて信頼できる論拠を示すことはほとんどない。様々なモデル構築技術は、測定されていない共通原因が唯一可能な種類であるかのように、あるいは、あたかもそれらが存在せず、全く無関係であるかのように進行する。このような仮定は必要ない。

測定不能な共通原因の有無、およびそれらの因果関係についての情報を得るための、漸近的に信頼できる方法が存在する。測定不能な共通原因の存在については、マルコフ条件と忠実性条件を仮定して、有益な十分条件が存在する。これらの条件は、潜在変数が存在するか否かにかかわらず、因果関係の結論と予測を確実に導き出すことができる定理をもたらします。線形性が仮定される場合、測定不能な共通原因の存在を特定するために使用できる、四元表現定理と呼ばれるものに要約される、より強力な定理が存在します。

本書のもう一つのテーマは、サンプルで調査された集団に適用される政策の効果を正しく予測するのに十分な（あるいは必要な）条件であり、調査前にその因果構造がわからないということである。ルービン、プラット、シュライファーのような統計学者たちは、次のような問いを重視してきた：観察集団や実験集団において、Xがある値を持つように*強制*されたときの**Zに対する**Yの条件付き分布は、いつX, **Zに対する**Yの条件付き確率と等しくなるのだろうか？これらの答えは、様々な反事実の主張が知られていることを前提としている。我々の用語では、これは研究対象のシステムの因果構造の側面を知っていることに相当する。その結果、次のような自然な疑問が生じる：上記の問いに答えるために必要な因果関係の知識は、どのような場合にサンプルデータから得ることができるのだろうか？我々は、この問いに答える結果を述べ、ZとX*に対する*Yの条件付き確率がXの操作によって不変になる場合について、サンプルデータから情報を与えることを目指す。しかし、私たちの考えでは、予測に関する根本的な問題はこれである：X_1、…、X_n*の分布が直接操作される場合*、X_1、…、X_n*が操作されなかった観察集団や実験集団の***Y**、**Z**、X_1、…、X_n、*その他の変数の分布から、***Zを***条件とする変数集合***Yの***分布は、サンプルの各ユニットでどのように算出できるのか。*確率と因果構造の形式的なつながりが答えを決めている

ので、その一部を解き明かしていきます。

検索アルゴリズムの信頼性を厳密に調査した結果、一般的に使用されている統計的検索手続きは、因果関係の推論には最適ではないという結論に直接つながる。この批判は回帰に最も重くのしかかる。より良い方法が利用でき、多くの回帰問題に簡単に適用できる。自動化されたモデル探索手順、特に線形回帰とロジスティック回帰における最も根拠のある反論は、その手順が、しばしば事前知識と一致する代替の因果仮説に対して漸近的に信頼できないことです。母集団の分布に関する情報が完全であるという理想的なケースにおいて、手順が正しい答えを与えるのであれば、より良い検定や計算効率の良いアルゴリズムを探し回ることができます。しかし、回帰や他の多くの自動化された技術の場合のように、確率的関係

と因果関係のマッチングが正しくない場合、どんなに統計的に繊細でも、良い推論にはなりません。

さらに、本書のもう一つのテーマは、実証研究のデザインにおける因果関係の推論の重要性である。相関関係は因果関係ではない、という格言の真偽は、統計的な依存関係から因果関係の依存性について何が判断*できるのか*、どのような条件のもとで判断できるのかを考え抜く上での障害になるかもしれません。対数線形法に関する最近の教科書（Christensen, 1990, p. 279）は、この問題を、大多数の統計学者によって支持されるだろうと思われる発言で開閉している：

> 因果関係は、データ分析で立証できるものではありません。因果関係を証明するためには、数値操作の域を超えた論理的な議論が必要です。例えば、うまくデザインされたランダム化実験は因果関係の結論の根拠となり得るが、観察研究の分析では相関関係の情報しか得られない。観察研究を因果関係の推論の根拠とする場合、相関関係から因果関係へのジャンプは非統計的な根拠に基づいてなされなければならない。

この文章は、実験的研究の因果構造に関する情報と観察的研究の因果構造に関する情報を対比しており、その表現する見解は至って標準的である。しかし、それは正しいのだろうか？ 因果構造と確率の関係を正式に理解すれば、実験と観察の比較に関する疑問は、実験デザインと観察デザインから抽出できる因果情報の数学的研究によって答えることができる。実験の威力と観測の無力さという標準的な主張は、誤解を招くというよりも、むしろ誤りであることが判明した。実験データと観察データから引き出せる因果関係の結論と予測には、それぞれ系統的な類似性があることが証明された。信頼性の高い因果関係の推論と予測を行うための実証研究の設計

の意味を理解するためには、複雑な（そしてまだ完全ではない）理論が必要である。

推論と予測について得られた結果は、被験者の選好が受ける治療に影響を与えることができる「倫理的」臨床試験を設計する問題を含む、実験設計に関する多くの論争的なテーマに対して興味深い示唆を与えるものである。発見可能な様々な経験的事実に応じて、実験結果から関連情報を得る力を損なうことなく、患者の選好がそのような役割を果たすことを可能にするデザインについて説明する予定である。

私たちが用いる方法は非ベイズ的なものであるが、この主題の広い問題はベイズ統計学と非ベイズ統計学の間の区分を越えており、ベイズ的な手続きによって関連する推論結果を得ることができるいくつかの方法を示すことになる。このテーマに適した数学的手法には、尺度理論よりもグラフ理論や計算理論が多く含まれる。経験的な手法では、分析的な結果が得られない場合に信頼性の根拠を示すためにコンピュータシミュレーションを多用する。方法論的な伝承のさまざまな部分が、このテーマですでに得られている結果によって否定され、また他のものは新しい視点と意義を与えられている。

本書の結果が実用的な科学的推論に対してどのような価値を持つかは別として、因果関係、予測、探索を調査する体系的で比較的無視されている研究分野を一緒に説明するものである。このテーマは、フィードバックを伴うシステム、ある種の推論の必要条件と十分条件、手続きの短距離信頼性、最適な探索手続きの存在、最適な統計的決定、モデルの識別不能特性、情報量と計算可能性のトレードオフなど、整った、そしてほとんど整わないオープンな質問でいっぱいです。

第2章

形式的な前置き

この章では、本書を通して使用されるいくつかの数学的概念を紹介する。この章は、我々が使用する形式的な装置の数学的に明示的な定義を提供することを目的としている。グラフ理論における標準的な概念について、正当な理由により非標準的な定義を用いることがあることを読者は警告しておく必要がある。読者が相関分析を含む有限数学と統計学の知識をある程度持っていることを前提としているが、そうでなければこの章には本書で必要となる数学的概念のすべてが含まれている。ここで定義された同じ数学的対象のいくつかは、次章で特別な解釈を与えられるが、ここではすべてを完全に形式的に扱う。

有向グラフ、無向グラフ、誘導経路グラフ、部分配向誘導経路グラフ、パターンなど、さまざまな種類のグラフを検討することができる。これらの異なる種類のオブジェクトはすべて、頂点の集合と辺の集合を含む。これらのオブジェクトは、いずれも頂点と辺の集合を含むが、含まれる辺の種類に違いがある。これらの違いにもかかわらず、無向パス、有向パス、親などの多くのグラフィカルな概念は、これらの異なる種類のオブジェクトのすべてに対して統一的に定義することができる。私たちが必要とするオブジェクトに対してこのような統一性を持たせるために、私たちはグラフの理論における慣習的な定義を修正する。

2.1 グラフ

図1に示す無向グラフには、無向の辺（例えばA - B）だけが含まれています。

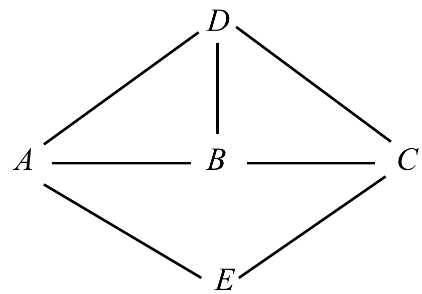


図1

図2に示す有向グラフは、有向の辺（例：*A*→*B*）だけを含んでいます。

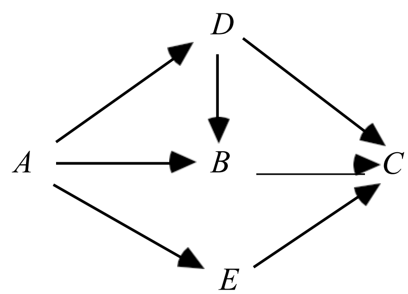


図2

図3に示す誘導パスグラフは、有向辺（例：*A*→*B*）と双有向辺（例：*B* <-> *C*）*の両方*を含む。(誘導パスグラフとその使い方については、第6章で詳しく説明します)。

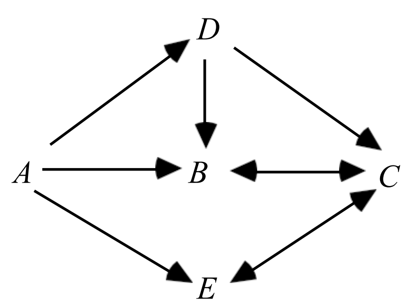


図3

図4に示す部分配向誘導パスグラフは、有向辺（例：*B*→）を含む。

F）、双方向のエッジ（例：*B* <-> *C*）、無指向のエッジ（例：*E* o-o *D*）、および部分

有向のエッジがあります。

の辺を持つ（例：*A* o-> *B*. ）。(部分配向誘導パスグラフとその用途については、第6章で詳しく説明する)。

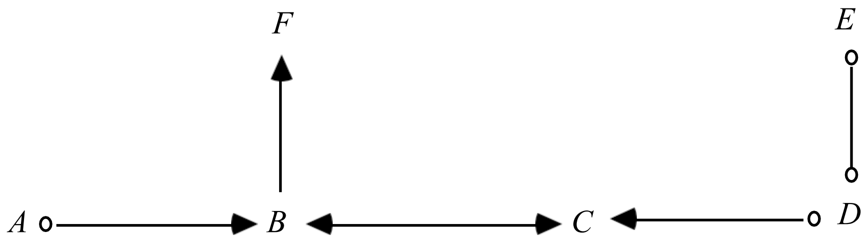


図4

図5に示すパターンは、無向きの辺（例：*A* - *B*）と有向きの辺（例：*A*）を含んでいます。-> *E*).(パターンとその使い方については、5章で詳しく説明します)。

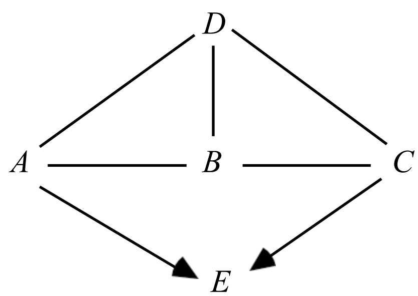


図5

通常のグラフ理論の定義では、グラフは順序付きペア**<V,E>であり**、Vは頂点の集合、Eは辺の集合である。**Eの**メンバーは、頂点のペア（有向グラフでは順序ペア、無向グラフでは無順序ペア）である。例えば、エッジ*A*→*B*は順序付きペア<A,*B*> で表されます。有向グラフでは、辺を表す頂点の組の順序が、実質的に辺の一端にある矢じりを示す。しかし、無向きの辺の端に付けられるマークは、もっと多様なものが必要である。一般に、辺の端は無印でも、矢じりでも、"o "でもよいことにする。

したがって、エッジの種類を完全に指定するためには、両端の変数と**マークを**指定

する必要がある。例えば、「A o-> B」の左端は、次のように表すことができる。

は順序ペア[A, o][1]、右端は順序ペア[B, >]として表すことができる。順序対の最初のメンバーはエッジの端点と呼ばれ、例えば[A, o]では端点はA です。エッジ全体は端点を表す順序対の集合で、例えば{[A, o], [B, >]}のようなものがあります。エッジのどちらの端が先に記載されているかは関係ないので、エッジ{[B, >],[A, o]}は{[A, o],[B, >]}と同じである。

$A \rightarrow B$のような有向辺は、Aの端点にマークがないことに注意。Aの端点のマークは空であると考えるが、順序ペアを書き出すときは、空のマークを表すために*EM*という表記を使う（例：[A，*EM*]）。

より正式には、**グラフは、**Vは頂点の空でない集合、Mはマークの空でない集合、Eは形式の順序付きペアの集合である順序付きトリプル**<V,M,E>であると**言いますが、これは
{[V1,M1],[V2,M2]}, ここでV1とV2は**Vに**、$V_1 \neq V_2$、M1とM2は**Mに**含まれる。
フィードバックがあるシステムの議論では、どのグラフでも、頂点V1とV2のペアが**E**内の多くとも1つの集合に出現すること、言い換えれば、任意の2つの頂点の間に多くとも1つの辺があることを常に仮定します。G = **<V,M,**E> の場合、*G*は**Vを超えるもの**であると言う。

例えば、図2の有向グラフは、＜{A，*B*，*C*，*D*，*E*}，{*EM*，＞} と表すことができる、{{[*a*,*em*],[*b*, >]}, {[*a*,*em*],[*e*, >]}, {[*a*,*em*],[*d*, >]}, {[*d*,*em*],[*b*, >]}, {[*d*,*em*],[*c*, >]}、{*b*,*em*],[*c*, >]}, {[*e*,*em*],[*c*, >]}}>となります。

Eの各メンバー{[V_1,M*1*],[V_2,M*2*]}を**辺と**呼びます（例：図2の{[*A*,*EM*],[*B*,>]}）。 辺の中の各注文ペア[V_1,M*1*}を**辺端と**いいます（例：[*A*,*EM*]は[E]の辺端となります。{[*a*,*em*],[*b*,>]} とする)。辺{[V_1,M*1*],[V_2,M*2*]}の各頂点V1を辺の**端点と呼ぶ**(例えば、A

は{[A,EM],[B,>]}の端点となる。) V_1、V2はEに端点V_1、V*2*がある場合にのみG内で**隣接する**(例えば、図2においてAとBは隣接するがAとCは非該当。)

無向グラフとは、マークMの集合＝｛EM｝であるグラフのことである。有向グラフとは、マークM＝｛EM，＞｝の集合で、**Eの**各辺について、一方の辺の端がマークEM、他方の辺の端がマーク"＞"であるグラフである。

無向グラフでは有向辺は存在しない。）辺{[A,M_1],[B,>]}はBに**入り**、辺{[A,EM],[B,M_2]}はAから**出る**。もしAからBへの有向辺があるならばAはBの**親**、Bは**子**

1順序ペアA , Bを<A , B>のように角括弧で表すのが通例だが、エッジの端点には角括弧を使い、角括弧が矢じりと誤解されないようにするためである。

(Vの頂点のすべての親の集合を**Parents(V)**、**Vの**頂点のすべての子の集合を**Children(V)**と表記する。）頂点*Vの***indegreeは**その親の数、outdegreeはその子の数、**degreeは***Vに*隣接する頂点の数に等しい（有向グラフでは、頂点のdegreeはindegreeとoutdegreeの和に等しくなる）。図2において、*Bの*親はAと*D*、*Bの*子は*C*である。したがって、*Bは*次数2、次数1、次数3である。

ここでは、グラフ内の無向きの経路を、グラフ内で隣接する頂点の列として扱うことにする。つまり、パス上で隣接するすべてのペア*X* , *Yに対し*て、エッジが存在する。

グラフの{[*X* ,M1],[*Y* ,*M*2]}となる。例えば、図2において、配列<*A*,*B*,*C*,*D*>*は*、配列内で隣接する変数の各組（Aと*B*、Bと*C*、Cと*D*）*が*グラフ内で対応する辺を持つので無向パスとなる。パスの辺の集合は、端点がシーケンス内で隣接している辺からなる。図2では、パス<A,*B*,*C*,*D*>*の*辺は{[*A*,*EM*],[*B*,>]}、{[*B*,*EM*],[*C*,>]}、{[*C*,>],[*D*,*EM*]}となっています。

より正式には、グラフ*Gの**A-B*間の**無向パスとは、***A*で始まり*B*で終わる頂点の並びで、その並びで隣接する頂点Xと*Yの*ペアごとに、*Gに*辺{[*X* ,M1],[*Y* ,M2]}が存在する。**辺**{[*X* ,M1],[*Y* ,*M*2}はXとYが*U*で（いずれかの順番で）互いに隣接するときのみ**パス***U* **内になる**。Xと*Yの*間の辺が経路*Uに*ある場合、XとYは*U*上で**隣接して**いるとも言う。Xと*Yの*間の無向パスで*Xを*含む辺が*Xから外れ*ている場合、その**パスは***Xから外れ* **ていると言い**、同様に、Xと*Yの*間のパスで*Xを*含む辺が*Xに入っている*場合、その**パスは***Xに入って***いると言う** 。証明を簡単にするために、1つの頂点からなる列を**空パスと**呼ぶことにする。どの頂点も2度以上含まない経路は**非周期的であり**，そうでない場合は**周期的である**．2つのパスが**交差するのは、**それらが共通の頂点を持つ場合であり、そのような共通の頂点を**交点と**する。経路Uが<U1, ...*Un*>、パスVを＜

Un， ....,Vm> とすると、UとVの**連結**は<U1,． ,Un,V1,・・・。,Vm> をU &Vと表記する。Uと空のパスの連結はU、空のパスとUの連結はUである。 通常、「パス」という言葉を使うときは非周期的なパスを意味し、周期的なパスを指すときは常に形容詞を使う。

グラフGにおけるAからBへの**有向パスとは、**Aで始まりBで終わる頂点の並びで、並びの中で隣接し、並びの中でこの順番で出現する頂点X , Yのペアごとに、G中に辺{[X ,EM],[Y ,>]}が存在するもので、Aはそのパスの**ソース**、Bは**シンクとする。例えば、**図2の＜A,B,C＞はソースA、シンクCの有向パスである。これに対し、図2の＜A,B,D＞は無向パスであるが、BとDがその順番で配列中に出現するので有向パスではなく、辺{[B ,EM],[D, >]}はGにない。

(ただし、$\{[D,EM],[B,>]\}$はGに含まれる。) したがって、有向パスは無向パスの特殊なケースである。UからVへの有向辺e（$U \rightarrow V$）に対して、**head**(e)=V、**tail**(e)=U。**有向無サイクルグラフは、**有向循環パスを含まない有向グラフである。

部分配向誘導経路グラフにおけるA-B間の**半有向経路とは、**AからBへの無向経路Uのうち、Aを指す矢尻を含む辺がないものである（すなわち、U上のAに矢尻はなく、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）。もちろん、すべての有向経路は半有向であるが、「o」の端点があるグラフでは、有向ではない半有向の経路が存在する可能性がある。

グラフは、その頂点のすべてのペアが隣接している場合、**完全で**ある。図6に完全無向グラフを示す。

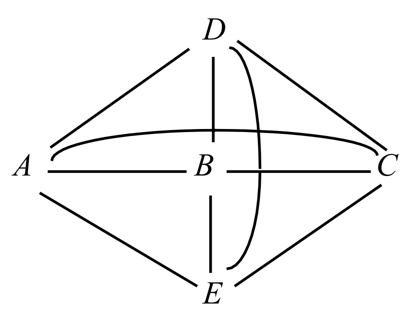
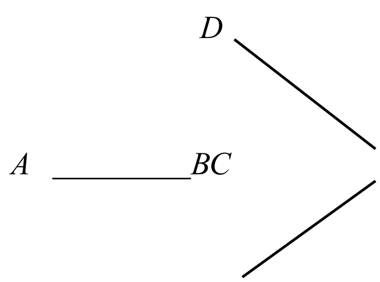


図6

グラフは、任意の2つの頂点間に無向きのパスが存在する場合、**連結されている。**
図1～図6は連結しているが、図7は連結していない。



E

図7

<**V**,**M,**E>**のサブ**グラフとは、**V'**が**Vに**含まれ、**M'が**M**に**含まれ、**E'**が**Eに**含まれるような任意のグラフ<V**'**,**M',E**'>である。の**部分グラフ**

の

V'上の<V,M,E>**は、V'**が**Vに**含まれる場合、ある辺が**Eに**あり、かつ両端点が**V'**にある場合にのみ**E'**にある部分グラフ <V**',M,E**'> です。

グラフ*G*の**閥とは、***G*の部分グラフのうち、完全なものを指す。例えば、図1では、部分グラフ*G*'＝(1)である。

<{*A*,*B*,D},{EM},{{[*A*,*EM*],[*B*,*EM*]},{[*B*,*EM*],[*D*,*EM*]},{[*A*,*EM*],[*D*,*EM*]}}>

は頂点*A*,*B*,*D*を持つ閥である。*G*中の頂点集合が*G*中の他の閥に適切に含まれない閥は**最大で**ある。図1では、*G*'と*G*"の両方が<{*A*,*B*},{*EM*}である、{{[*A*,*EM*],[*B*,*EM*]}}> は閥であるが、*G*"は*G*'と異なり、*G*"は*G*'に適切に含まれるため、最大値ではない。[2]

つまり、頂点*X*、*Y*、*Z* が三角形を形成するのは、*X* と*Y* が隣接し、*Y* と*Z* が隣接し、*X* と*Z* が隣接するときだけである。グラフ*G*において、頂点*V*が**無向パス***U***上のコライダー**であるのは、*U*上に*V*を端点として含む2つの異なる辺があり、両方とも*V*に入るときだけである。それ以外の場合、*V*は*U***上の非コライダーで**ある。グラフ*G*において、頂点*V*が*U*上のコライダーであり、Vが*U*上の異なる頂点V_1および$_V2$に隣接し、V_1および*V*2が*G*において隣接しないとき、頂点*V*は*U*上の**無遮蔽コライダー**である。

V.図2において、*A*、*B*、*C*、*D*、*E*はすべて*C*の祖先であるが、Aも*C*も*C*の親にはなっていない。同様に、Cは*A*、*B*、*C*、*D*、*E*の子孫であるが、Aや*C*の子ではない。すべての頂点*V*はVから*V*への有向（空）パスの元であるから、各頂点はそれ自身の子孫であり先祖であるが、もちろん、それ自身の親でも子でもない。

2.2 確率

グラフの頂点は、実線の写し、非負の実線の写し、整数の区間のいずれかに値を持つ確率変数になります。

2特に統計学の分野では、「閾」を我々が定義した*最大閾の*ように理解する作家もいる。

グラフの頂点上の共同分布とは、これらのオブジェクトのデカルト積上の可算加法的確率測度を意味する。2つの確率変数X, Yが**独立で**あると言う。　の結合密度が、XとYのすべての値に対するXの密度とYの密度の積であるとき、2 つの確率変数 X , Y は独立であると言う。これを　と書きます。ある変数の集合が別の変数の集合から独立であることを主張するとき、私たちは明白な方法で一般化する。ある確率変数の集合が**共同で**　**独立で**言うとき、その集合の任意の2つの不連続な部分集合が互いに独立していることを意味します。X、Yのすべての値、およびZのすべての値zに対して、**Zを**与えたXの密度と**Zを**与えたYの*密度の*積が等しく、**zの**密度が0に等しくないとき、確率変数**X、Y、Zの**集合について明白な方法で一般化します。**XがZを**与えられた**Yから**独立している場合、**X Y|Z と**書き、**条件付き独立性の**順序は**Zの**変数数に等しいと言います。

離散的な場合、**Vの**すべての値vに対して$P(\mathbf{v})\neq 0$である場合に限り、**V**上の分布が正であると言う。(一般に、**V上の分布は、**密度関数がすべての**vに対して**非ゼロである場合に正である）。Vが**V'**に含まれる場合、そして

$$P(\mathbf{V}) = \sum_{\mathbf{V}(\text{ブイ})} \overrightarrow{P(\mathbf{V'})} \text{ である。}$$

は、$P(\mathbf{V})$が**V**上の$P(\mathbf{V'})$の**余白であると**言うことにする。

2.3 グラフと確率 分布

分布に真に存在する条件付き独立関係を、いくつかの異なるグラフで表現することを検討する。

2.3.1 Directed Acyclic グラフ

有向非循環グラフは、確率分布における条件付き独立関係を表現するために使用することができます。

与えられたグラフGと頂点*W*に対して、*W*の親の集合を**Parents**(*W*)とし**Descendants**(*W*)を*W*の子孫の集合とする。

> **マルコフ条件：V**上の有向非循環グラフGと確率分布*P*(**V**)がマルコフ条件を満たすのは、**V**内のすべての*W*に対して、*W*が**Parents**(*W*)が与えられたときに、**Vの**独立性（**Descendants**(*W*) □**Parents**(*W*) ）がある場合に限られる。

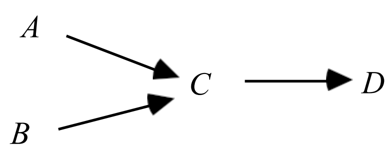


図8

(Pearl(1988)の用語では、GはPの**I-mapで**ある。図8において、マルコフ条件は以下の条件付き独立関係を伴う。

$$A \perp\!\!\!\perp B$$
$$D \perp\!\!\!\perp \{A, \mathrm{B}\} \mid C$$

f(**v**)≠0となる**Vの**全ての値に対して、マルコフ条件を満たす結合密度関数f(V)は次式で与えられる。

$$f(\mathbf{V}) = \prod_{V \,\square\, \mathbf{V}} f(\mathbf{V}|\mathbf{Parents}(V))$$

ここで、$f(V|$**Parents**$(V))$は、*V*の親である頂点の集合(空かもしれない)を条件とする*V*の密度を表す(Kiiveri and Speed, 1982を参照。).**Parents**(*V*) = □ ならば、$f(\mathbf{V}|Parents(V)) = f(V)$ という表記規則を思い出してください。）

離散変数に対する共同分布が図8のマルコフ条件を満たす場合、次のように因数分解することができる：

$$p(a,b,c,d) = p(a)\, p(b)\, p(c|a,b)\, p(d|c)$$

[3] 些細な独立関係、例えば、*C*□ /*whi*□は定義上真である。 ⍰ Φ｜φは定義上真である。

は、$P(A,B,C,D)\neq 0$となるようなA、B、C、Dのすべての値についてである。有向無サイクルグラフGにおいて、次数が0の頂点は外**生的**であるといいう。Gが分布Pに対してマルコフ条件を満たす場合、外因性変数V_1、v2のすべての組に対して、PのV_1 v*2*がある。

最小条件とは、直感的に言えば、グラフの各辺が、そうでなければ得られるであろう何らかの条件付き独立関係を阻止することである。

最小化条件：Gを**V**上の有向無サイクルグラフ、Pを**V**上の確率分布とすると、<*G*,P>が最小条件を満たすのは、頂点集合**Vを**持つGの固有部分グラフHについて、<*H*,*P*>がマルコフ条件を満たさない場合のみである。

図8の例に戻ると、マルコフ条件を満たすが、Aが$\{B,C,D\}$から独立している分布P'は、A-C間の辺を取り除いた部分グラフに対してもP'がマルコフ条件を満たすので、最小性条件を満たさない。Pearl (1988)の用語では、分布P(**V**)**が**有向無サイクルグラフGに対してマルコフ条件と最小性条件を満たす場合、GはPの**最小Iマップとなる**。

ある分布Pが有向無周期グラフGに対してマルコフ条件と最小化条件を満たす場合、GはPを**表すと**言うことにする。任意の有向無周期グラフGとマルコフ条件と最小化条件を満たす確率分布Pについて、変数AとBが統計的に依存する場合、以下のいずれかとなる：

(i) Gの中にAからBへの有向パスが存在する、または
(ii) Gの中にBからAへの有向パスが存在する：または
(iii) があり、変数Cと、CからBへ、CからAへの有向パスがG内に存在する。

異なる頂点AとBの間の**トレックは**、AとBの間の有向パスの無順序のペアです。

同じ発生源を持ち、発生源においてのみ交差する*B*。一対のパスの発生源はトレックの**発生源とも**呼ばれる。なお、トレックのパスのうち1つは空のパスであってもよい。

2.3.2 Directed Independence グラフ

有向独立グラフは、確率分布の条件付き独立関係を表現するもう一つの方法（ほぼ同等）である。有向無サイクルグラフGが$P(\mathbf{V})$の**有向独立性グラフ**であるとする（Whittaker 1990）。

Gに$A \rightarrow B$が発生するのは、～（$V \neq A$かつ$V > B \perp\!\!\!\perp B \mid \mathbf{K}(B)$）、ここで$\mathbf{K}(B)$はすべての であるような$A$頂点$V$）だけである場合。

定理2.1$P(\mathbf{V})$が正分布であるとき、**Vの**変数の任意の順序に対して、Pはその順序に対する$P(\mathbf{V})$の有向独立グラフのマルコフ条件と最小性条件を満たす。

分布Pが正でない場合、与えられた変数の順序に対するPの有向独立グラフが、Pが最小条件とマルコフ条件を満たす有向無サイクルグラフの部分グラフである可能性がある（Pearl, 1988）。

2.3.3 誠実さ

任意のグラフが与えられたとき、マルコフ条件は独立関係の集合を決定する。これらの独立関係は、マルコフ条件によって与えられた独立関係を持つすべての確率分布が、さらにこれらの独立関係を持つという意味で、他の独立関係を含むことがある。一般に、マルコフ条件を満たすグラフG上の確率分布Pは、グラフに適用されたマルコフ条件が内包する独立関係以外に、他の独立関係を含むことがある。例えば、図9のグラフのマルコフ条件を満たす分布では、グラフがその独立性を内包していなくても、AとDが独立である場合がある。

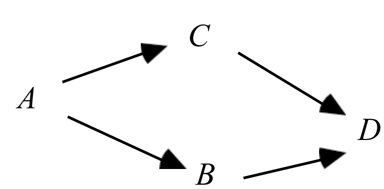


図9

線形モデルでは、Cに対するDとAに対するCの偏回帰係数の積が、Bに対するDとAに

*対するBの*対応する積をキャンセルする場合、このような独立性が生じることがある
。

*Pの*条件付き独立関係のすべてが、*Gに*適用されるマルコフ条件によって包含されるならば、PとGは**互いに忠実で**あると言う。さらに、ある分布*Pが***忠実で**ある有向無周期グラフが存在する場合、その分布Pは**忠実であると**言うことにする。Pearl (1988)の用語では、Pと*Gが*互いに忠実である場合、以下のようになります。

分布*P*が有向無サイクルグラフ*G*に忠実である場合、Xと*Y*の間にトレックが存在する場合に限り、XとYは依存する。

2.3.4 d- 分離

Pearl (1988) に従い、グラフ *G* において、*X* と *Y* が *G* の頂点で、$X \neq Y$ であり、**W** が *G* の頂点で *X* または *Y* を含まない集合であるとき、*X* と *Y* は、(i) *U* 上のすべてのコライダーが **W に**子孫を持ち、(ii) *U* 上の他の頂点が **W** にないような無向パス *U* が *X* と *Y* 間に存在しないときのみ *G* において **W を**与えられたとき **d 分離**するとする。$X \neq Y$、かつXとYが**Wに**ない場合、XとYは**Wを**与えられたときに**d-**分離しない場合に限り、与えられた集合**Wをd-連結して**いると言う。**U**, **V**, Wを*G*の頂点の不連続集合とし、Uと**V**が空でないとき、Uと**Vの**カルテシアン積におけるすべてのペア<U,V>が**Wを与えられたときにd-**分離している場合に限り、UとVは**Wを**与えられたときにd分離するという。**U**、**V**、Wが*G*の頂点の不連続集合で、UとVが空でない場合、Uと**VがWでd-**separatedでない場合に限り、UとVは**Wで**d-connectedと言う。d-connectednessの例として、次の有向無サイクルグラフがある（ただし、第6章で説明する誘導経路グラフなど他の種類のグラフにもこの定義は適用できる）：

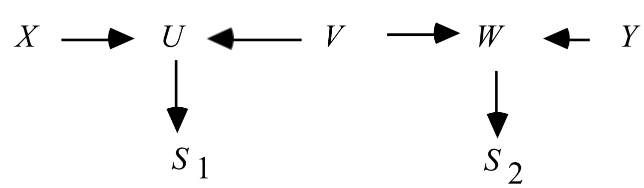


図10

XとYは空集合があればd個に分離される

XとYはd-connected与えられた集合$\{S_1, S2\}$ *である。*

形式的な前置き
XとYは集合{S_1,S_2,V}が与えられるとd分離される。

2.3.5 リニア 構造物

V上の有向非周期グラフ*G*が分布*P*（**V**）を**線形に表現するのは、V'**上の有向非周期グラフ*G'*と以下のような分布*P' '*（**V'**）が存在するときだけである。

(i) Vは**V'**に含まれる；

(ii) **Vの**各内生的（つまり正の次数を持つ）変数*Xに対して*、**V'ivVには、**ゼロ次
ε
数、正の分散、アウト度が1に等しく、Xから*Xへの*有向辺を持つ固有の変数*Xが存在します*；

(iii) Gは**V**上の*G'*のサブグラフである；

(iv) *Gの*各内生変数は、*G'*の親の一次関数である；

(v) *P' '*(**V'**)では、*G'*の任意の2つの外因性変数間の相関はゼロである；

(vi) *P*(**V**)は、*P' '*(**V'**)の**Vに対する**マージンである。

V'ivVのメンバーを**誤差変数と**呼び、*G'*を**展開グラフと呼ぶことにします**。有向無サ
ρ ρ
イクルグラフ*Gは*、*Gが*線形に表現するすべての分布において$_{AB.\mathbf{H}}$＝0**である**場合にのみ、$_{AB.\mathbf{H}}$＝0を**線形に含意します**（分布にはすべての偏相関が存在するとします）。*Gが**P*（**V**）を線形に表現するなら、＜G，*P*（**V**）＞対は有向無サイクルグラフ*G*と線形モデルであると言います。

2.4 無向きの独立性 グラフ

条件付き独立性に関する統計的仮説の表現として、*無向*グラフによる表現がよく知られている。有向グラフによる表現と無向グラフによる表現、この2つの表現は密接な関係にあるが、混同しないようにすることが重要である。

無向独立グラフ*Gが*分布*Pを***表現**する場合、Aと*Bの*間のすべての無向パスが**Cのメンバーを**含む場合に限り、Aと*Bが*ある集合**C**を条件として独立であることを示す。

ある有向無周期グラフGと忠実な確率分布を考えるとして

P.つまり、*U*は*G*と同じ頂点セットと*G*と同じ隣接関係を持つ無向グラフである。Iは、今与えられた定義に従って形成された分布*P*の無向独立グラフであるとする。Iと*U*が同じになるのは、Gに遮蔽されていないコライダがない場合のみである（Wermuth and Lauritzen 1983）。

2.5 決定論的および擬似的な非決定論的 システム

ここでは、専門的な意味で決定論的システムの概念を用いることにする：有向無サイクルグラフ*G*で表される確率変数の集合**V**上の結合確率分布*P*は、*G*の零でない次数の各頂点が、*G*における その直接の親である頂点の関数である場合に**決定論的で**ある。「関数」とは、親の各頂点に一意の値を与えるごとに、従属頂点の値が一意になることをいう。

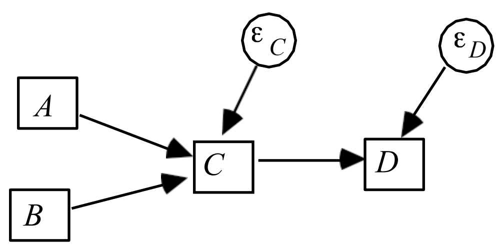


図11

図11のグラフ・分布の組が決定論的であるが、CとDは測定されていないとする。測
ε
定された変数、すなわち*A*、*B*、*C*、*D*だけを考えると、変数のいくつかは統計的に依存しているが、どの変数も他の変数の値によってその値が一意に決まることはないことがわかる。Cと*D*は「隠れた」変数であり、これを追加すると決定論的になるのだが、このシステムは決定論的でないように見える。さらに、システムを決定論的にするために、2つの測定変数が同じ隠れ変数に依存することを仮定する必要はなく、「隠れ」変数の間の依存性を仮定する必要もない。測定変数間の有向非循環グラフで表される分布が決定論的ではないが、このように有向非循環グラフで表される分布に埋め込み可能である場合、その分布は擬似決定論的であるという。

これに対して、図12を考えてみましょう。ここでもA,B,C,D,だけが測定されたとする。この場合、以下のどちらかでなければ、隠れ変数を追加してシステムを決定論的にすることはできない。

隠された変数が関連付けられているか、少なくとも1つの隠された変数が少なくとも2つの測定変数に隣接している。

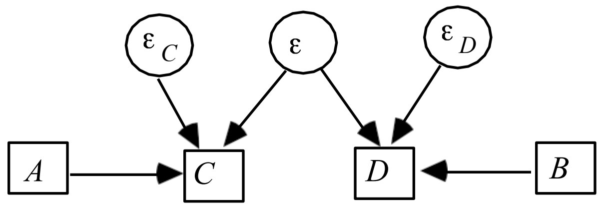


図12

より正式には、<*G*,*P*>は**擬似非決定論的であり**、PはV上の確率分布、GはV上の有向非周期グラフであるが、*G*が*P*の決定論的グラフでなく、**V**を適切に含む変数集合**V'**に対する分布*P'*と有向非周期グラフ*G'*が以下のように存在する場合に限る。

(i) *G'*は*P'*の決定論的なグラフである;

(ii) Gは**V**上の*G'*のサブグラフである;

(iii) **Vの**頂点が**V'の**頂点の祖先となることはない;

(iv) は、**V'**ivの2つの頂点を結ぶトレックの元となる頂点はない;

(v) Pは*P'*のマージナルです;

(vi) *GはPを*表します。

<*P*,*G*>が**線形擬似決定性であると**言う場合、<*P*,*G*>が擬似決定性であり、さらに*G'*において、**V'**の各頂点がその親の線形関数であることを意味します。有向無サイクルグラフで線形に表現される分布は擬似非決定論的である。(グラフと分布のブール型疑似非決定論的ペアなどにも同様の定義が適用される）。

2.6 背景 備考

Lauritzen, Speed and Vijayan (1978) の純粋なグラフ理論的研究と、対数線形モデルにおける統計的研究 (Bishop, Fienberg and Holland, 1975) から、1980年に Darroch, Lauritzen and Speed は、条件付独立の対数線形仮説の無向グラフ表現を導入した。Kiiveriの論文に基づき、Kiiveri and Speed (1982)は、Markov Conditionのバージョンを導入し、再帰的因果モデルという概念を定義し、以下の結果を得た。

多項分布の最尤推定を行い、離散変数と連続変数の両方への応用について体系的な調査を行った。その直後、Kiiveri, Speed and Carlin (1984)は、さらに形式的な基礎を発展させた。Wermuth and Lauritzen (1983)は、再帰的ダイアグラム、あるいは我々が有向独立グラフと呼ぶものの概念を導入した。最小性と忠実性の定義はPearl (1988)によるものである。

第3章

因果関係と予測：公理と説明

因果の性質に関する見解は、因果関係をある種の確率的関係として分析するもの、因果関係をある種の反実仮想的関係（操作や介入に関係する反実仮想的関係もある）として分析するもの、そして因果について全く語らないことを好むものに大別される。我々は因果の定義を提唱しないが、この章では、我々の使い方を体系化し、因果構造を確率、反実仮想、操作と結びつける仮定を明示しようとするものである。適切な形而上学的回旋によって、この仮定はこれらの観点のいずれからも支持されうるし、おそらく最後の観点さえも含まれるであろう。

3.1 条件付き

知的な計画を立てるには、通常、行動の結果を予測することが必要です。行動は状態を変化させるので、まだ行っていない行動の結果を評価するには、未来の条件文の真偽を判断する必要がある--*Xがそうであったとしたら、Yはそうであっただろう*。過去の慣行や政策の影響を判断するには、「もし*Xがそうであったなら、Yもそうであっただろう*」という反実仮想の文の真偽を判断する必要がある。

未来条件や反実仮想条件が成立する条件を詳細に記述することは、よく知られた難しい哲学的な問題である。Lewis（1973）は、カンガルーが松葉杖を使う状況を想像しても、「*カンガルーに尾がなければ転倒する*」は真であると指摘している。つまり、カンガルー用の松葉杖の希少性とカンガルーが松葉杖を使いたがらないことを考えると、もしカンガルーにしっぽがなかったら、転倒してしまうだろう、ということです。しかし、この直感を正確にするのは簡単ではない。

因果法則が反事実を伴うことは広く認識されている。実際、このことは、因果法則を、いわば偶然に成立する一般化とは区別する特徴であると考えられていることがよくある。*あなたのポケットにある硬貨はすべて銀でできている*」は、「*もしこの1円玉があなたのポケットにあれば、それは銀でできている*」という反事実を含意していない。しかし、*電子と陽電子の衝突でエネルギーが放出されるという*因果律は、*もしこの電子がこの陽電子と衝突したら、エネルギーが放出*されるという反事実を内包している。

因果関係の規則性と将来の条件文や反実仮想文の真偽との関連性から、因果関係の構造を発見することは、多くの文脈で知的計画にとって不可欠である。自動車事故の致死率と車の重量を関連付ける一次方程式は、ある集団では正しいかもしれないが、それが世界の強固な特徴を記述していない限り、法律で車の重量が操作された場合に致死率がどうなるかを予測することはできない。母集団における値の分布をパラメトリックに正確に表現しても、変数間の因果構造を反映しない限り、計画には使えないかもしれない。

3.2 因果関係

私たちは、因果関係を特定の事象間の関係であると理解しています：何かが起こり、他の何かを引き起こす。それぞれの原因は特定の出来事であり、それぞれの結果は特定の出来事である。ある出来事Aは、複数の原因を持つことができ、そのどれもが単独で*A*を生じさせるのに十分ではない。因果関係は、他律的、無反射的、かつ、

非対称的であると仮定する。すなわち、i) Aが*B*の原因であり、Bが*C*の原因である場合、Aは*C*の原因でもある、ii) イベント*A*はそれ自体を引き起こすことはできない、iii) *A*が*B*の原因である場合、*B*は*A*の原因ではない、です。

3.2.1 Direct vs. Indirect 因果関係

直接的な原因と間接的な原因の区別は、一連の出来事に対する相対的なものである。Cがマッチを打ったという出来事で、*A*がマッチに火がついたという出来事であり、それ以外の出来事を考慮しない場合、Cは*A*の直接的な原因である。しかし、もし*B*：マッチの先端の硫黄が酸素と結合するのに十分な熱を得たとしたら、もはや*C*

したがって、*Cが**B*を引き起こし、*Bが**A*を引き起こす場合、*Bは*Cと*Aの*間の**因果的媒介者**であると言うのである。

文脈と事象のセットを固定した上で、ある事象が他の事象の直接原因となるとはどういうことか。直感的には、*Aの*直接原因となる事象が発生すれば、*Aが*発生するかどうかは、*Aの*間接原因となる事象が発生するかどうかとは、もはや関係がないことになる、ということです。直接的な原因は、間接的な原因を効果から*遮断するのである。*ある子どもが保育園で水ぼうそうにかかり、ウイルスに感染して発疹が出た場合、感染は発疹の発生から暴露という事象を遮断する。いったん感染すれば、発疹が出るかどうかは、保育園でウイルスにさらされたか、土曜日の朝のプレイグループでウイルスにさらされたかということとは関係ない。

VをCと*Aを*含む事象の集合とする。*Cが***V{A}**に含まれる集合**Cの**メンバーで、(i) **Cの**事象が*Aの*原因であり、(ii) **Cの**事象が発生すると、**V**{A}（{A} □ **C**）の事象が発生するかしないかに関わらず、*Aの*原因となり、 (iii) (i) と (ii) を満たす**Sの**部分集合がない、という場合だけ、**Vに対して***Aの***直接**原因とすることができる。

3.2.2 イベントと 変数

因果関係を経験的に推定できる確率と結びつけるためには、事象を分類しなければならない。ある現実の、あるいは可能性のある事象を集め、ある種類であると宣言し、他の現実の、あるいは可能性のある事象から区別して、他の種類に集めなければならない。最も単純な分類では、事象をある種類、例えば、日食やダウ・ジョーンズ工業平均の下落として記述し、ある種類の各事象Aを、*Aの*非発生である事象¬*A*と対にします。このような分類により、原因としての*変数について*分かりやすく話

すことができるようになります。そのためには、ある種の事象やその不在を値として受け取るブール変数を導入します。**ブール変数*C*がブール変数*A*を引き起こすのは、**少なくとも1つの対（*C*, ¬C）が少なくとも1つの対（*A*, ¬A）を引き起こす場合だけである、と言う。通常、2つのタイプのイベント間の因果関係がある程度の一般性を持っていない限り、イベントをタイプに集め、そのような変数の間の因果関係を調べることに悩む人はいないでしょう--つまり、タイプ*A*のイベントの多くがタイプ*C*のイベントを原因として持ち、タイプ*C*のイベントの多くがタイプ*A*の影響を持つか、あるいは全く持たないかです。

事象は、ある種の*X*の事象がある種の*Y*の事象を引き起こし、ある種の*Y*の事象がある種の*X*の事象を引き起こすように、変数Xと*Y*に集約することができる。このような場合

は、変数間の因果関係に一義的な方向性がないことになる。このケースについては、第12章で検討する。

ある鍋の水を摂氏100度にするような、ある量がある値をとる事象もある。多くの種類の尺度は、特定のシステムが尺度の値をとる、あるいは尺度の値の集合の中の値をとるという、可能な出来事の配列と関連している。また、このようなスケールの変数は、少なくとも特定の時間間隔の特定のシステムについては、原因と結果であると言うことができます。ある特定のシステム*Sについて*、*Qの*値（または値の集合）qと*Rの*値（または値の集合）*rが*あり、*QがS*で値*qをとる事象がRがS*で値*rを*とる事象を引き起こす可能性がある場合、**尺度変数*Qは*** *S*で**尺度変数*Rを*引き起こすと**言う。実際には、異なる尺度の値の間の因果関係が、特定の値や特定のシステムに限定せずより一般的であると考えて初めて尺度というものを用いる場合が多い。Kがシステムの集合体である場合、**Kの**すべてのシステム*Sに対して*、*QがSにおいてRを*引き起こすという条件で、変数*Qが***Kにおいて**変数*Rを*引き起こすと言う。

なぜなら、Aと*Bという*2つの変数がどんなに離れていても、*Aの*値の変動が*Bの*値の変動をもたらすようなシステムの配置が、たとえ非常に可能性が低くても、物理的に可能な*場合が*ほとんどだからです（独裁者が、シカゴの出産件数が中国の茶価の関数であるように状況を整えることは可能でしょうが...）、変数間の因果関係を厳密に適用すれば、ほとんどすべての自然変数が他のほとんどの自然変数の原因となり得るでしょう。したがって、厳密には、変数の因果関係の定義は、他の変数の可能な値の集合に対する相対的なものであるべきだが、ここではこの形式を無視して、文脈を信頼することにする。直接原因の概念は、変数間の因果関係の定義と明らかに平行して、事象から変数へと一般化される：変数Cは、**Vに対する**変数*Aの*直接的

な原因である。

(i) *C* *は* **V** に含まれる変数の集合 **C の**メンバーであり、(ii) **C の**変数が値 **c を取るなら、V** の他の変数の値がどうであれ *A* *に*値 *a* *を*取らせるような **C の**変数の値 **c** と *A* の値 *a* *の*セットが存在し、 (iii) (i) と (ii) を満たす **C** の適正部分集合は存在しない。*X が*{X ,*Y* ,*Z*} *に対するY の*直接原因であり、{*X* ,*Y* ,*Z*} *に対するZ の*直接原因である場合にのみ、変数Xが変数YとZ*の***共通原因であるという**。**Vの**中に*A* で始まり*B* で終わる変数の並びがあり、その並びの中でその順番で隣接する変数Xと*Y の*各組について、*X が***Vに対して**Yの**直接原因となる**場合、**Vに対して***A からB への*因果の**連鎖が**あると言う。

(i)Aが*B*の原因である場合、*A*は**Vに対する***B*の直接原因または間接原因である。(ii)*A*、*B*、Cが**Vに**あり、**Vに対する**AからBへの因果連鎖が*C*を含まない場合、*A*および*B*を含む任意の集合**V'**に対して、**Vに対する**Aから*B*への因果連鎖は*C*を含まない。あるシステムにおいて、2つの変数の一方が他方の原因である場合、あるいは共通の原因を持つ場合に、**因果的につながって**いると言うことができる。母集団の**因果**構造とは、**V**を変数の集合、Eを**Vの**順序対の集合とした順序対**<V,E>であり**、<*X* ,*Y* >は、**Vに対して***X*が*Y*の直接原因である場合のみ**Eに含まれる**。どの母集団を意図しているかが明らかな場合は、明示的に言及することはない。*P*(**V**)が因果構造*C*=**<V,E>を**持つ集団における**V**上の分布である場合、*C*が*P*(**V**)を**生成**したと言う。二つの因果構造<V,E>と

<V**'**,**E**'>が**同型であるのは、**Vから**V'**への一対一の関数*f*があり、**Vの***A*および*B*の任意の2つのメンバーについて、<*f*(*A*),*f*(*B*)>がE**'**にある場合に限り、<*A*,*B*>が**Eにあるとき**である。変数の集合**Vは、**母集団において、**V**内の任意の2つ以上の変数のすべての共通原因が**V**内にあるか、母集団のすべてのユニットに対して同じ値を持つ場合にのみ、母集団に対して**因果的に十分である**[1] 我々はしばしば、明示的に集団に言及せずに因果的十分性の概念を使用します。

3.2.3 例

単純なデジタル論理回路は因果構造の具体例である。しかし、このような回路素子の記述があれば、その回路に関連するどの事象が他のどの事象を引き起こすかについて、ほとんどの人が合意できるという利点がある。下図の素子では、変数X_1、X_2、X_3は、対応する線に電流が流れるか流れないかで、1、0の2つの値を持ち、半円は「and」ゲートを表す。電流は上から下へ流れる。変数X_3の値は、X_1とX_2の値の単純なブール関数であり、「-」がブール乗算を表す場合、$X_3 = X_1 - X_2$となる。

[1] 厳密には、ある集団に対する**Vの**因果的充足性には、*X*が**Vに**含まれず、**Vの**2つ以上の変数の共通の原因である場合、**Vの**すべての変数の結合確率が、集団に生じる*X*の各値で同じであることが必要である、とする。

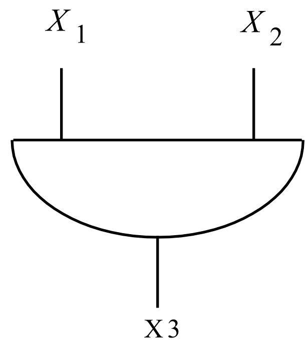


図1

X_1が値1をとる事象とX_2が値1をとる事象は、それぞれX_3が値1をとる事象の原因であると理解する。ブール変数X_1とX_2は、それぞれブール変数X_3の原因であると言う。

因果構造の形は、関係する変数の種類や、それらの間の関数の特定のクラスには依存しない。同型の因果構造は、連続変数の線形依存性のシステムによって実現されるかもしれない。ある線路の電圧を表す3つの異なる変数X_1、X_2、X_3は、正の実数範囲にある。その中に電圧の和を出力する機構があるとする（図2）。

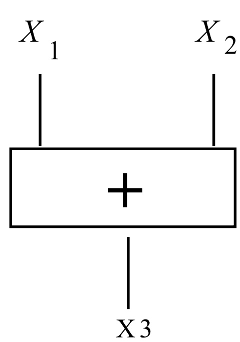


図2

この場合、$X_3 = X_1 + X_2$となるが、因果構造は図1の因果構造と同型であり、X_1とX_2それぞれがX_3の原因となっている。

$X3$は$X1$とX_2の効果、$X3$は$_X1$とX_2の関数というように、因果的な依存関係と機能的な

依存関係があることを示唆している。

各変数や事象をその直接の原因である変数の関数として表現することにより、機能的依存を因果的依存から推論することができる。逆に、ある方程式がシステムを正しく記述しているからといって、そのシステムの直接的な因果的依存関係が方程式の機能的依存関係に反映されていると推論することはできない。例えば、$X_3 = X_1 + X_2$ という方程式があるシステムで正しい場合、$X_2 = X_3 - X_1$ という方程式も同様にそのシステムで正しいが、X_1 と X_2 が X_3 を引き起こす場合、通常、X_3 と X_1 は X_2 を引き起こさない[2]。

3.2.4 因果関係を有向グラフで表現する

直接原因の概念を用いれば、有向グラフで因果構造を表現することは簡単である：

> **因果表現規約：** 有向グラフ*G*＝＜V，**E**＞**は、***G*の頂点が*C*の変数を表し、*A*が**V**に対する*B*の直接原因である場合にのみ、*G*にAから*B*への有向辺が存在するとき、ユニットの集団に対する因果的に十分な因果構造*C*を表す[3].

因果構造を表す有向非周期グラフを**因果グラフと**呼ぶことにする。以下の図3は、図1、図2に示した回路装置の因果グラフである。

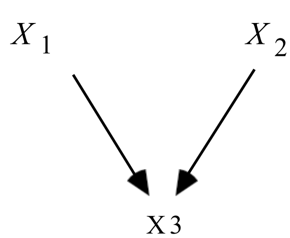


図3

前回の定義と同様に、*G*が因果グラフで、*G*中の頂点Xと、Xから*Y*への*Z*を含まない

有向パス、X から*Z*への*Y*を含まない有向パスが存在する場合、X は Y と*Z*の**共通の原因であると**言うことにする。

---

[2]Simon（1953）は、同定可能性の概念を用いて、システムを記述する方程式の集合から因果構造を導き出す一般的な方法を提案し、同じ論文の後半では、線形係数の摂動に対する不変性を用いた因果関係の説明も提案している。

[3]変数の因果関係は他動的かつ無反射であると仮定されるため、因果構造を表す有向グラフは非周期的でなければならない。環状有向グラフを導入するには、体系的な再解釈が必要である。

因果関係表現規約には重要な限界があります。例えば、薬物AとBがともに症状Cを軽減するが、Bを含まないAの効果は非常に些細なもので、B単独の効果はそうでないとする。本章で検討した有向グラフ表現では、この相互作用を表現する手段がなく、AとBがそれぞれCに影響を及ぼす他の状況と区別することができません。この相互作用は、グラフに関連する確率分布によってのみ表現されます。別の例として、単純なスイッチについて考えてみよう。図4のように、電池Aには充電と非充電の2つの状態があるとする。電池Aの充電は、スイッチBがオンであれば電球Cを点灯させるが、そうでない場合は点灯しない。AとBを独立な確率変数とすると、AとCはBと空集合に条件付きで依存し、BとCはAと空集合に条件付きで依存し、AとBはCに条件付きで依存する。したがって、A、B、Cの分布を表す有向無サイクルグラフは、上図のようになる。この結論は、情報が十分でないことを除けば、何の問題もない。AとCの依存関係は、すべて$B = 1$という条件によって生じる。$B = 0$のとき、AとCは独立である。

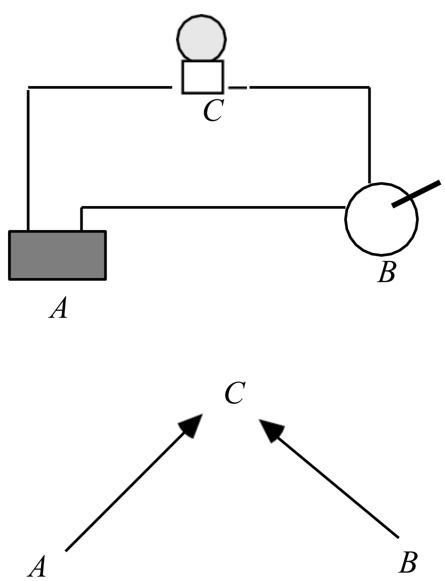


図4

離散データでは、条件付き独立性の事実が知られていれば、スイッチ変数を特定す

るので、より良い表現は、ある変数の特定の親をスイッチとして特定する。しかし、この種の一般的な表現は、しばしば把握が容易ではない。[4] 有向無サイクルグラフ表現を拡張してスイッチを表現する最近の研究は、Geiger and Heckerman (1991) に記述されている。

---

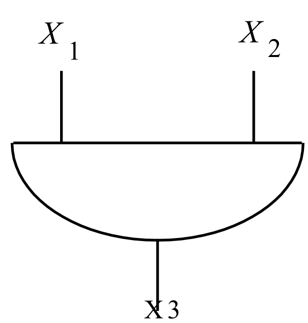
[4]より実用的な配置としては、因果関係グラフを推論する以外に、変数の操作の効果に関するユーザーの質問に答えるクエリーシステムが考えられます。

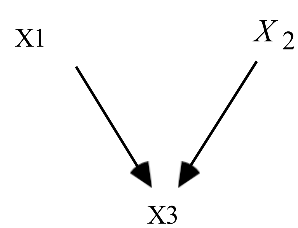
3.3 因果関係と 確率

3.3.1 Deterministic Causal Structures

近似的に図1と図2の装置は**決定論的であり、すなわち、**効果はその直接的な原因の決定論的な関数である。もし各効果が母集団におけるその直接原因の線形関数であれば、そのシステムは母集団における**線形決定論的因果構造であると言う。**

因果グラフの変数のうち、次数がゼロ、すなわち因果入力がないものを外**生**変数という。図3の因果グラフでは、$X\,1$と$X\,2$が外生変数である。外生的でない変数は**内生的である**。決定論的因果構造では、外因性変数の値が、残りの変数の値を一意に決定する。





回路図因果関係 **グラフ図5**

図1の回路素子とその因果関係グラフを考える。

5.この装置が説明通りに動くかどうかを検証する実験を想像してください。外生変数に値を割り当てる、つまり、$X1$とX_2に電流を流すかどうかを決め、X_3に電流が流れているかどうかを読み取るのです。この実験を次のような表で表すことができる。

X1	X2	X3
1	1	?
1	0	?
0	1	?
0	0	?

通常、設計通りに動作することは確認できたが、どの程度の頻度で、どのような形で故障するのかを知りたかったとする。いくつかの試行について、$X1$とX_2にランダムに値を割り当て、X_3に電流が流れたかどうかを読み取ることができる。つまり、外生変数の集合が占める可能性のある各状態に確率を割り当てることができるのです。例えば、以下のような感じです、

$$p(x_1 = 1, x_2 = 1) = 0.2$$
$$p(x_1 = 1, x_2 = 0) = 0.3$$
$$p(x_1 = 0, x_2 = 1) = 0.2$$
$$p(x_1 = 0, x_2 = 0) = 0.3$$

この因果構造は（外生変数がランダムであっても）決定論的であるため、外生変数に対する確率分布は、システム内の変数セット全体に対する結合分布を決定する。この例では、(X_1, X_2, X_3)に対する共同分布は次のようになる：

$$p(x_1 = 1, x_2 = 1, x_3 = 1) = 0.2$$
$$p(x_1 = 1, x_2 = 1, x_3 = 0) = 0.0$$
$$p(x_1 = 1, x_2 = 0, x_3 = 1) = 0.0$$
$$p(x_1 = 1, x_2 = 0, x_3 = 0) = 0.3$$
$$p(x_1 = 0, x_2 = 1, x_3 = 1) = 0.0$$
$$p(x_1 = 0, x_2 = 1, x_3 = 0) = 0.2$$
$$p(x_1 = 0, x_2 = 0, x_3 = 1) = 0.0$$
$$p(x_1 = 0, x_2 = 0, x_3 = 0) = 0.3$$

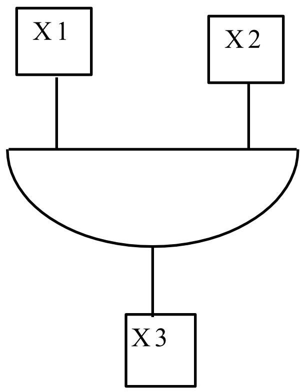
この分布は、図5の因果構造によって生成されると言います。

この例は、回路のサンプリング方式を調べるためではなく、決定論的因果関係装置によって確率分布がどのように生成されるかを説明するために使用します。決定論的因果構造と確率分布の関連について、我々が行う唯一の仮定は以下の通りである。

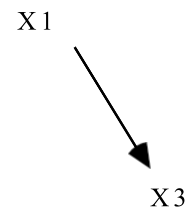
外生変数が生成しうる分布は、外生変数に対して許容される分布に含まれる。*我々は、外因性変数が、因果的に十分な構造を持つ変数上の確率分布において、共同で独立であると仮定する。これは*、統計的依存性が因果関係によって生み出されるという実質的な仮定であり、表現に関する慣習でもある。ある構造における外生変数が独立でない場合、因果グラフは不完全であり、グラフに表現されていない、統計的依存性をもたらすさらなる因果メカニズムが存在すると予想される。入力変数の一部が他の変数の原因になっているか（この場合、我々は等閑視しており、因果グラフは実際には構造の因果構造のグラフではない）、あるいは、観測された変数の非定常な共通の原因が構造の記述に含まれていないかのいずれかである。

3.3.2 疑似非決定論的・非決定論的因果関係 構造体

実際には、人々が測定する変数が互いの決定論的な関数であることはほとんどない。ある集団の変数の集合V**に対する**因果構造で、ある変数がV**における**その直接的な原因の決定論的関数でないものを、その集団の**非決定論的因果構造と**呼ぶことにする。非決定論的因果構造は、擬似的な非決定論的であるかもしれない。例えば、図1、図5に示した装置がラインX_3の電流を制御しているとする。また、X_2を隠して、X_1とX_3だけが因果構造に現れるとする。それでも私たちは、X_1がX_3の原因であるという仮説を立て、図6の右側の因果関係グラフを形成するかもしれない。



実際の回路図



仮説の因果関係グラフ

図6

実際の回路装置が生成する結合分布$P(X_1, X_2, X_3)$が3.3.1節で図5について与えられたものと同じだと仮定すると、観測分布$P(X_1, X_3)$はちょうど$P(X_1, X_2, X_3)$のマージン、すなわち

$$
\begin{aligned}
p(x_1 = 1, x_3 = 1) &= 0.2 \\
p(x_1 = 1, x_3 = 0) &= 0.3 \\
p(x_1 = 0, x_3 = 1) &= 0.0 \\
p(x_1 = 0, x_3 = 0) &= 0.5
\end{aligned}
$$

観測された分布では、X_3は明らかに直系の*親*X_1の関数ではなく、因果構造は非決定論的であるように見えます。この構造を擬似非決定論的と呼ぶ。より正式には、因果構造C＝＜V,**E**＞**が**集団に対して**擬似非決定論的で**あるのは、Cが集団に対する決定論的因果構造ではなく、**Vを**適切に含む変数集合**V'**に対する集団に対する因果構造C'が以下のように存在する場合に限る。

(i) C'は母集団に対する決定論的因果構造である;

(ii) AとBが**Vにある**場合、<A,*B*>が**Eに**あるのは、<*A*,*B*>が**E'**にある場合に限られる;

(iii) **Vの**変数が**Vの**変数の原因であることはない;

(iv) は、**V'**音の2変数に共通する原因である;

Cの機能依存関係がすべて線形である場合、その構造は**線形擬似不確定性**であるという。社会科学における構造方程式モデルは、通常、擬似不確定性因果構造であると仮定される。このようなモデルの誤差項は、しばしば原因が省略されたものと解釈される。

変数の集合**V**上の集団に対する**真に不確定な**因果構造とは、擬似不確定ではない不確定な因果構造である。真に非決定論的な構造、真に非決定論的な巨視的構造であっても、その変数が因果構造を持つものが存在することは、少なくとも考え得ることである。ここでは、擬似非決定論的構造で成り立つ条件付き独立性と因果構造の関係が、真に非決定論的な因果関係でも成り立つと仮定するが、後述するように、この仮定が慎重になされなければならない量子力学系が存在するようである。測定変数が他の測定変数の正確な関数である場合の議論については、3.8節を参照のこと。

3.4 アクシオム

確率と因果グラフをつなぐ3つの条件について考える：因果マルコフ条件、因果最小条件、そして忠実条件である。これらの公理は独立ではない。条件の様々な部分集合の結果が、本書の過程で調査される。次節では、これらの条件に対する正当性や反論を検討するが、これらの条件の重要性は、その真偽はともかく、社会科学の文献で目にした因果的な意味を持つ統計モデルのほぼすべてが、この3つを満たしているという事実によって証明されている：もしそのモデルが真であれば、3つの条件すべてが満たされるであろう。3つ目の条件である「忠実性」に反するモデルを構築することは容易であるが、そのようなモデルは現代の実務ではほとんど存在せず、存在しても「忠実性」の結果である性質を持つという事実が、そのモデルに対する反論として受け取られる。第5章と第8章では、この3つの条件を満たす対数線形モデル、回帰モデル、構造方程式モデルについて、公表されているものを検討する。

3.4.1 因果的マルコフ 条件

因果グラフとそれが生成する確率分布を結びつける直感は、ある基本的な条件において統一され、一般化されます：

因果マルコフ条件： *G*を頂点集合**V**を持つ因果グラフ、Pを*G*で表される因果構造によって生成される**V**の頂点上の確率分布とする。Gと*P*は、**V**中のすべての*W*について、**Parents**(*W*)が与えられたときに、*W*が**V**（**Descendants**（*W*）∪ **Parents**（*W*））から独立している場合にのみ因果マルコフ条件を満たす。

*G*のCausal Markov条件を満たす*P*として分布する変数を持つ集団の因果依存関係を*G*が記述する場合、Pは*G*によって**生成さ**れると言うことがある。**Vが**因果的に十分ではなく、分布*P*を生成する因果グラフ*G*の変数の適正部分集合である場合、*P*の**V**上の限界にCausal Markov条件が成り立つとは*考えないことにする*。

第2章で説明した因数分解の結果は、因果マルコフ条件を満たす因果構造を持つシステムの集団における変数の集合**V**の結合確率分布に適用される。$P(V \mid \textbf{親}(V))$が、*V*の直接の原因である頂点の(おそらく空の)集合を条件とする*V*の確率を表すとすれば、次のようになる。

$$P(\mathbf{V}) = \prod_{V \in \mathbf{V}} P(\mathbf{V}|\mathbf{Parents}(V))$$

は、各$P(V|\mathbf{Parents}(V))$が定義される**V**のすべての値について。

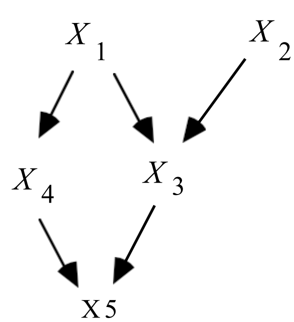


図7

図7のグラフに対して、マルコフ条件を直接適用すると、Gが生成する分布に関する

独立性の事実のリストが得られる。

$$X_1 \perp\!\!\!\perp X_2$$

$$X_2 \perp\!\!\!\perp \{X_1, X_4\}$$の場合

$$X_3 \perp\!\!\!\perp x_4 \mid \{x_1、x_2\}$$。

$$X_4 \perp\!\!\!\perp \{X_2, X_3\} \mid X_1$$

$$X_5 \perp\!\!\!\perp x_1、x_2\} \,/\, \{x_3、x_4\} \mid \{x_3、x_4$$

これらによって、例えば、他の独立関係が内包される。

$x_4, x_5\} \mid \perp\!\!\!\perp \{x_4, x_5\} \mid \{x_4, x_5\} \mid \{x_5, x_5 \;\; x_2 \mid \{x_1、x_3\}$とする。

条件付き独立の公理については、Pearl (1988)に記載がある。

3.4.2 因果の極小化 条件

我々は通常、確率と因果関係をつなぐさらなる条件を課す。この原則は、それぞれの直接的な因果関係が、そうでなければ得られるであろう何らかの独立性または条件付き独立性の関係を妨げると言うものである。例えば、次の因果グラフ*Gにおいて*、Cは*Aの*直接の原因である。

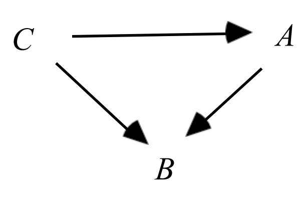


図8

$\perp\!\!\!\perp$ *CAが成立*する{A,*B,*C*}*上の分布*Pにおいて*、*Pが*マルコフ条件を満たすのは、次のような場合であってもである。

の場合、*C-A*間のエッジはグラフから削除されます。

因果最小条件：*G*を頂点集合**V**を持つ因果グラフとし、*P*を*G*が生成する**V**上の確率分布とする。<G, *P*>が因果最小条件を満たすのは、頂点集合**V**を持つ*G*のすべての適切な部分グラフHに対して

<*H*,*P*>はCausal Markov条件を満たさない。

考察するグラフには因果的な解釈を与えることがほとんどなので、以下、ほとんどの場合、この2つの条件を単にマルコフ条件と最小化条件と表現することにする。

3.4.3 忠実度 条件

因果グラフが与えられたとき、マルコフ条件によって独立関係のセットが決定される。これらの独立関係は、マルコフ条件によって与えられた独立関係を持つすべての確率分布が、さらにこれらの独立関係を持つという意味で、他の独立関係を含むことがある。一般に、マルコフ条件を満たす因果グラフ*G*上の確率分布*Pは*、そのグラフに適用されたマルコフ条件が内包する独立関係以外に、他の独立関係を含むことがある。しかし、そうではなく、*Pの独立*関係のすべてが、*Gに*適用されたマルコフ条件によって包含されるならば、Pと*Gは***互いに忠実**であると言う。さらに、ある分布*Pが*忠実である有向無サイクルグラフが存在する場合、その分布Pは忠実であると言うことにする。そこで、さらなる公理を考えてみます：

> **忠実性の条件***Gを*因果グラフ、*Pを**Gが*生成する確率分布とする。<G, *P*> *が*忠実条件を満たすのは、*P* で真となるすべての条件付き独立関係が、*Gに*適用される因果マルコフ条件によって包含される場合のみである。

分布*Pが**Gに*忠実であるのは、マルコフ条件と忠実条件の*両方を*満たす場合のみであることに注意してください。忠実性とマルコフ条件は最小性を伴いますが、最小性とマルコフ条件は忠実性を伴いません。私たちは、弱い方の公理を使うこともあれば、強い方の公理を使うこともある。忠実性は因果構造を発見するのに重要であることがわかり、また、確率分布と因果構造の間の「正常」な関係であることがわかる。

3.5 条件の検討

確率と因果が一緒になってこの条件を満たすと考えられるのはいつ、なぜか、また、この条件に反すると考えられるのはいつなのか。ある集団における変数の値は、どのような場合にこの条件に従って分布していると考えるべきか。

3.5.1 因果的マルコフと最小性 条件

外生変数が独立に分布する決定論的あるいは擬似非決定論的システムの因果グラフの頂点の確率分布を考えるなら、マルコフ条件を満たす必要がある。その証明は最終章で行われる。また、すべての擬似非決定論的システムにおいて、最小性条件が成立すると推測される。この条件の根拠は、この事実と、人間が大きく制御したり操作したりできるシステムに関する経験の歴史にある。電気的装置、機械的装置、化学的装置はすべてこの条件を満たしている。自動車力学から化学反応速度論、デジタル回路設計に至るまで、科学と工学の大きな領域は、故障の診断やメカニズムの推論にこの原理を用いなければ不可能であっただろう。

重要なケースでは、最小条件とマルコフ条件の適用が不明確な場合がある。1903年、*G*. Udny Yuleは統計学における属性の関連性の理論に関する基本的な論文を「異なる記録の混合によって引き起こされるかもしれない誤謬について」という項目で締めくくった。(ユールは ／AB｜C ／を「Cの宇宙におけるAと*B*の関連性」（p.131）を表すのに使っている）：

　この定理は、その逆の応用から、実用上かなり重要である;

つまり、｜AB／ガ感覚的に正または負の値を持つとしても、｜AB｜C｜と／AB｜γ｜がともにゼロでないことを確認することはできない。例えば、ある属性は男系にも女系にも遺伝しないかもしれないが、混合記録はかなりの見かけ上の遺伝を示すかもしれない。例えば、父親と息子の50%がその属性を持つが、母親と娘には10%しか持たないとする。この場合、どちらの血筋にも遺伝がないとすると、記録は（およそ）次のようなものにならざるを得ない
。

属性のある父親と属性のある息子:	25%
属性のある父親と属性のない息子:	25%
無属性の父親と、属性のある息子:	25%
無属性の父と無属性の息子:	25%

属性のある母親と属性のある娘:	1%
属性のある母親と属性のない娘:	9%
無属性の母親と、属性のある娘:	9%
阿母吽娘	81%

この2つのレコードを等しい割合で混合すると、次のようになります。

ぞくせいおやこ	13%
そうぞう	17%
そうぞう	17%
あおやまのこ	53%

つまり、2つの異なる記録が混ざり合うだけで、かなり大きな、しかし幻想的な継承が生まれるのである。同じような錯覚的な関連、つまり、最も明白な物理的意味を与えてはならない関連は、異なるレコードがプールされている場合や、1つのレコードだけが多くの異質な材料でできている場合にも、非常に多く発生することがある。

レコードの混合によって生じる架空の関連は、連続変数の場合に同じプロセスで生じる偽の相関に対応するもので、このケースはピアソン教授によって最近の回顧録で十分に議論された。相関がゼロである2つの別々の記録を一緒にプールすると、少なくとも変数の1つの平均が2つのケースで同じでない限り、偽の相関が必然的に発生する。

Yuleの例は、Causal Markov条件の問題を提示しているようです。**Vに関する混合**物を、**Vの**変数に関する異なる共同分布を持つ有限個の部分集団P_iの組み合わせからなる任意の集団とし、それぞれの分布はあるグラフについて因果的マルコフ条件を満たすとする。以下のような母集団を考える。

は、<G,p*1*>と<G,*P*$_2$>の構造の混合物であり、p1とp2は区別され、*G*のマルコフ条件を満たす。混合物の比率を*n*： *m*とする。

$P(X,Y,Z) = nP1(X,Y,Z) + mP2(X,Y,Z)$ とし、$n + m = 1$とする。少し代数学的に考えると$P(XY|Z) = P(X|Z)P(Y|Z)$であり、以下の場合のみである。

(1) $n2P1(X,Y,Z)P1(Z) + nmP2(X,Y,Z)P1(Z) + mnP1(X,Y,Z)P2(Z) + m2P2(X,Y,Z)P2(Z) = 1.$

$n2P1(X,Z)P1(Y,Z) + nmP1(X,Z)P2(Y,Z) + mnP2(X,Z)P1(Y,Z) + m2P2(X,Z)P2(Y,Z).$

n,m＞0であり、両分布において*X*,YがZの条件付きで独立である場合、すなわちp*1*(*X,Y*｜*Z*)=。
P_1（*X*｜*Z*）p*1*（*Y*｜*Z*）、P_2（*X*，*Y*／Z）＝P_2（*X*｜*Z*）P_2（*Y*／Z）とすると、式（1）は以下のようになる。

(2) p_2（*x*｜*z*）p_2（*y*｜*z*）＋p_1（*x*｜*z*）p_1（*y*｜*z*）＝p_1（*x*｜*z*）p_2（*y*｜*z*）＋p_2（*x*｜*z*）p_1（*y*｜*z*）。

確率分布を混合すると、可能な限りの条件付き*依存関係が*見つかるかもしれないというのが、古くからある結論だが、今でもかなり驚くべきことである。このように、多くの混合集団では、条件付き独立性と依存性は因果構造を知るための信頼できるガイドにはならないようです。

線形擬似不確定性システムの場合、それぞれが線形構造に関連する2つの異なる分布を持つ集団が混合されると、それぞれの別々の分布における消失相関は混合分布における消失相関を生じず、それぞれの別々の分布における消失部分相関は混合分布における消失部分相関を生じない。線形構造に基づくか否かにかかわらず、2つの分布の任意の混合物について、以下の場合に限り、2つの変数の共分散が混合物において消失することを確認するのは容易である。

$$k_1 COV_1(XY) + k_2 COV_2(XY) = k_1 k_2 [\mu_{1X}\mu_{2Y} + \mu_{1Y}\mu_{2X}] + k_1(k_1 - 1)\mu_{1X}\mu_{1Y} + k_2(k_2 - 1)\mu_{2X}\mu_{2Y}$$

ここで、集団1と集団2の割合をn：mとし、$k_1 = n/(n+m)$、$k_2 = m/(n+m)$、「μ_i」は集団iの平均を表す。

つまり、因果グラフG_1を持つ集団1と因果グラフG_2を持つ集団2が存在し、その合同集団はどちらのグラフに対してもマルコフ条件を満たさない分布になるという状況です。問題は、このような混合母集団が、以下に違反するかどうかです。

を満たしている。ある部分集団に属する原因が*Vの変数の共通の原因と正しくみなされる場合，混合集団では因果マルコフ条件は破られない*．その代わりに，因果マルコフ条件を満たすシステムの集団があるが，測定されていないかもしれない共通の原因（または原因）を持つ．場合によっては、部分集団に属する原因は、3.2.4節で検討したような潜在的なスイッチ変数のように作用するかもしれない。潜在変数の異なる値に対する条件付き分布は、異なる因果グラフに忠実な確率関係を決める。Yuleの例では、欠けている共通原因は性別である。別の例として、鉛と銅の混合標本を作る場合、各小集団内では密度と電気伝導度は独立だが、混合集団では統計的に依存することになる。これは、化学組成が密度と電気伝導度の共通の原因であるためと言うべきでしょう。また、関連する部分集団に属する原因が、*不自然な*種類と思われたり、少なくとも科学者が求めるような原因でない場合もある。このように、現代の認知神経心理学における重要な論争（Caramazza, 1986）は、例えばブロカ失語症の被験者のように、症候群によって選ばれた人々のサンプルに対する統計結果の使用に関するものである。このようなグループを研究する目的の1つは、2つ以上の正常な能力が、ブローカ失語症の患者において損傷した共通の原因を持っているかどうかを発見することであろう。例えば、ブローカ失語症患者のサンプルで、2つの認知能力のテストのスコアに相関が見られたとします。心理学者は、そのテストの成績には共通の潜在的な原因があると結論づけるべきでしょうか? しかし、その共通の原因とは、両方の技能の原因となる機能的能力（損傷しているか、そうでないか）である必要はない。むしろ、ブローカ失語症患者のサンプルは、さまざまな種類の脳障害を持つ人々の混合物であり、それぞれのサブグループの中で、問題の技能は独立して分布しているかもしれない。共通の原因は、ある集団に属することを表す変数に過ぎない。

混合物の統計が母集団の構成員に関する変数を反映しないような文脈もある。線形モデルでは、相関と偏相関は、線形係数と外生変数の分散によって決定される。これらのパラメータ自体は確率変数として扱うことができ、結果として得られる母集団分布は（一般に数え切れない）分布の混合となる。このような設定において、因果関係ではなく、統計的推論が広く研究されている（Swamy, 1971）。Xを確率変数とすると、Xの期待値を$E(X)$と表す。

定理3.1 M*を有向*非周期グラフGと線形係数a_{ij}*を持つ*線形モデルとする。M'を有向非周期グラフG*を持つ*線形モデルとし、M'の線形係数はM'の他のすべての確率変数から共同独立な確率変数a'_{ij}であり、E（a'_{ij}）＝a_{ij}*であるとする*。ここで、外生的な非

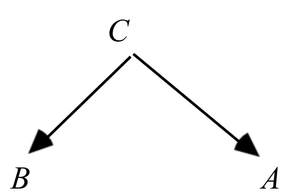
> 係数乱数変数はMとM'で同じである。そして、M'において$\rho_{AB.\mathbf{C}}=0$であるのは、Mにおいて$\rho_{AB.\mathbf{C}}=0$である場合のみである。

したがって、同じ因果グラフを持ち、パラメータが独立に分布する線形擬似不確定性因果十分系の混合物である集団は、測定不能な共通原因がなくても、そのグラフの因果マルコフ条件を満たすことになります。

プロの哲学者たちは、「因果マルコフ条件」の結果に対する批判を相次いで行っている。そのほとんどは、関連する潜在変数の省略に依存しているようだ。Wesley Salmon (1984)は、Causal Markov Conditionの観点からは適切に特徴づけられない「もう一つの、基本的に異なる種類の共通原因状況」が存在すると主張している。Salmonはこの別の因果関係を "interactive fork "と呼んでいる。

インタラクティブ・フォーク」の一例として、Davis(1988)のものがある：

> テレビのスイッチに不具合があると想像してください。スイッチがオンになると、音と映像の両方が出る。このとき、スイッチがオンで音が鳴る確率は、スイッチがオンで鳴る確率より大きい。(デイビス88号156頁)

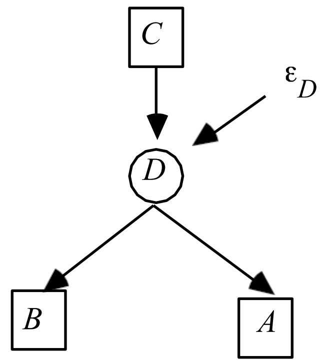


C＝スイッチオン
B = サウンドオン
A = スクリーンオン

図9

つまり、$P(B|C) < P(B|A \,\&\, C)$となる。

デイビスの例は、因果関係の状況を不正確に表現しており、次のように描くのがよいでしょう：



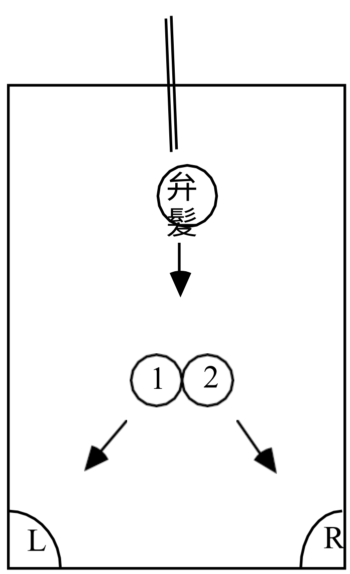
C＝スイッチオン B＝サウンドオン A＝スクリーンオン
D = 回路閉鎖

図10

回路の状態、あるいはスイッチイベントの下流にある何らかの変数が、AおよびBをを独立させました。

サーモン自身の図解では、プールゲームの次の例を少し変形して使っています（ここでは、彼のイベントをブール変数に置き換えています）。

CはAとBの両方に関係する因果関係の記述ですが、AとBはCを条件として独立したものではありません。



A＝Lに1球。
B＝Rの2球。
C = 手玉と1または2ボールの間のあらゆる種類の衝突。

図11

C（衝突があったこと）と*A*（ボール1がポケットに落ちたこと）を知っていると、*B* *が*起こったかどうか（ボール2がポケットに落ちたこと）については、*C*を知っているだけではわからない。Aと*B*は直接因果関係があるわけではないし、*C*を条件として独立しているわけでもない。

サーモンの例では、事象Cは、Aと共通の原因のすべてを完全に記述しているわけではありません。

*B.C*は、手玉と1個または2個の玉の間に何らかの衝突があったことを教えてくれるが、その衝突の性質については教えてはくれない。もし、先行事象がもっと情報量が多ければ、例えば、手玉が2つの標的の玉に当たったときの正確な運動量を特定することができれば、条件付き独立性が回復する。この例は、因果関係分析に何らかの代理変数が使われたり、ある変数の異なる値が折りたたまれたりするときに生じる、実際のデータ分析でおなじみの問題を反映しているだけである。これらの例は、因果マルコフ条件を疑う理由にはならないと我々は考えている。

Elliott Sober (1987)は、共通の原因が存在しない相関関係や、既知の共通の原因を条件付けた後に残存する相関関係が日常的に見られると論じている。イギリスのパン価格とヴェネツィアの海水面の相関は、いくつかの共通の原因（おそらく産業革命）を持っているかもしれないが、すべての依存関係を説明するには十分ではない。変数*A*が時間とともに増加する系列と変数*B*が時間とともに増加する系列を考えた場合、Aと*B*は因果関係がなくても、その時々のすべての単位から形成される集団の中で相関があることになる。このような組み合わせの集団は、明らかに時間値で与えられる集団の混合物である。

因果マルコフ条件にはもっと根本的な反論がある。すなわち、現在のところ、この条件が偽である非決定論的因果系が存在するということである。量子力学的な事象によって2つの粒子が生成され、それぞれ異なる方向に移動していく。保存則により、2つの粒子の動的変数は相関していなければなりません。例えば、一方がスピンアップの成分を持っていれば、もう一方はそのスピンダウン成分を持っているはずで

す。そこで、空間的に離れた2つのセンサーで2つの異なるスピンの成分を測定し、その相関関係を計算する実験を行います。粒子のペアが生成される瞬間に、*S*を条件として2つの粒子の動的変数が無相関であるようなシステムの状態*S*が存在するとします。J. S. Bell (1964) は、このような仮定に基づき、測定された力学的変数の相関を制約する不等式が成り立つことを論じた。この不等式は、ある種の量子力学的実験において破られる。このような実験では、相関する変数は空間的に離れたサブシステムに関連しているため、因果過程が「局所的に」、つまり距離を越えて瞬時に作用しないように制約する原則を捨てない限り、統計的依存性は、おそらくあるサブシステムがそのサブシステムに及ぼす影響による*ものではないだろう。*

を持つか、共通の原因によるものである。したがって、局所性原理を捨てない限り、「因果マルコフ条件」は偽であるように見える（Elby, 1992）。

私たちの考えでは、量子力学的実験において因果マルコフ条件が明らかに破綻していることは、他の文脈でマルコフ条件を放棄する理由としては不十分である。例えば、古典力学が文字通り偽であるからといって、軌道を計算する際に古典物理を使うことを放棄することはないのです。因果マルコフ条件は、実験室、医学、工学の分野で常に使われており、予期しない、あるいは予期しない統計的な落ち込みは、*一応*説明されるべきものである。もし、どこでもこの条件を放棄するならば、治療の割り当てと結果変数の値との間の統計的依存関係は、因果関係の説明を必要としなくなり、実験計画の中心的な考え方は消えてしまうことになります。これより弱い原則は一般に妥当とは思われない。例えば、*Yの*因果的な親が*Yを*より遠い*原因から*独立させる、とだけ言うのであれば、非常に奇妙な不連続性をもたらすことになる：*XがYに*最も影響を与えない限り、XとYは*Xの*親を条件として独立である。しかし、*XがYに*全く影響を与えないようになると、Xと*YはYの*親を条件として統計的に依存する可能性がある。

因果マルコフ条件の根拠は、第一に、外生変数が独立に分布する構造的に類似した擬似決定論的システムの集団に必ず当てはまること、第二に、反復プロセスを経て、その基本的性質をテストできるシステムに関する我々の経験のほとんどすべてによって支持されていることである。条件」に対する説得力のある反論は、「条件」が成立しない巨視的なシステムを示し、因果関係の説明を望む巨視的な自然・社会システムも「条件」を満たさないと考えるべき強力な理由を提示する必要があるであろう。しかし、そのようなケースはないように思われる。

3.5.2 誠実さとシンプソンのパラドックス

忠実性は、Simpsonが当初提示した「パラドックス」の変種を実現するケースで侵害されることがあります。ユールもピアソンも、2つの変数が部分集団では独立であっても、結合集団では従属であることを観察していたことはすでに見たとおりである。1948年、M. G. Kendallは『*統計学の上級理論*』*の中*で、2つの二項変数は独立であるが、第3の変数に依存するという逆の状況を説明する例を用いた。Kendallの例は、数年後のSimpson(1951)の論文で、因果従属性と分割表との関係に困難をもたらすと考え、ひねりを加えたものである。その後、この例が示す現象は、次のように呼ばれるようになった。

「シンプソンのパラドックス」。このような例は、因果関係と確率の関連性を議論する際の標準的なパズルとなりました。

Kendallの例5は次のようなものだった：

ある病気に対して多数の患者が治療を受け、回復した人数が注目される場合を考える。Aを回復、～Aを非回復、Bを治療、～Bを非治療[6]とし、その頻度を次のように考える。

	B	*~B*	合計
A	100	200	300
~A	50	100	150
合計	150	300	450

ここで、（*AB*）＝100＝（*A*）（*B*）／Nであり、属性は独立である。このように、治療が回復に影響を与えることはない。男性の性別をSM、女性の性別をSFとし、男女の頻度を次のようにする。

男性用

	ビーエスエム	*~ビーエスエム*	合計
空対地ミサイル	80	100	180
~アズム	40	80	120
合計	120	180	300

女性

	ビーエスエフ	*~B SF*	合計

ひょうじゅんへんせい	20	100	120
~ASF	10	20	30
合計	30	120	150

男性グループには、現在

[5]p.319.Qは、2×2表の1行目をa,b、2行目をc,dとしたときのYuleのQ＝（$ad-bc$）/（$ad+bc$）です。

[6](中略）ケンドールは、もちろん、記号がそれぞれの治療と回復の状態を示すものであり、その逆ではないことを意味しています。

$$QAB_{\cdot SM} = 0.231$$

と、女性グループにおいて

$$qab_{\cdot sf} = -0.429$$

したがって、男性では治療は回復と正の相関があり、女性では負の相関がある。両者の独立性が高いように見えるのは、下位集団でこれらの関連が相殺されたためである。

Kendallの例は、男性と女性の2つの分布の混合であり、一方の集団における2つの変数間の正の関連が、他方の集団における負の関連によって正確に相殺されるようなものである。このことは逆説的ではありませんし、同じ構造が主張される経験的な例を見つけることができます。混合分布は、すべてのユニットに共通する因果グラフに適用されるマルコフ条件から導かれない統計的独立関係を示すので、忠実な条件に違反することになる。

Kendallの分割表の説明は、ある集団では2つの変数の関連性が正であり、他の集団では負であるという事実に依存している。しかし、もし*両方の下位集団で関連が*正 *であるにも*かかわらず、混合集団ではそれが消失しているとしたら、何が起こっているのだろうか？ これこそ、1951年にSimpsonが投げかけた疑問である。シンプソンは次のような表と解説を与えている：

	男性		女性	
	未処理	処理済み	未処理	処理済み
アライブ	4/52	8/52	2/52	12/52
死者	3/52	5/52	3/52	15/52

今回は、男性でも女性でも治療と生存の間に正の関連があるが、表を合わせると、組み合わせた集団では治療と生存の間に関連がないことがわかる。とはどういうことでしょうか?

[7]Fienberg(1977)は、Darrochを引用して、「Yuleが1903年に発表した『属性の関連性の理論』の最後のセクションでこの問題を論じたので」(p.51)、このセクションの最初の文章を除いて、Yuleは実際に混合物の逆の問題、すなわち、変数が集団では統計的に依存するが、部分集団では独立である状況について論じたのです。

"賢明な "解釈はここですか? 男性に適用しても有益であり、女性に適用しても有益である以上、その治療が人種にとって無価値であるとして否定することはできない。

問題は、どのような因果関係でそのような表ができるのかということであり、この問題は正しくは「シンプソンのパラドックス」[8]と呼ばれる。

シンプソンの例では、変数G（男性・女性）、T（治療・未治療）、S（生存・非生存）が、因果構造に暗黙の制限を与える解釈を与えられている。私たちはこの例を読むとき、性別*Gは*治療Tや生存*Sによって*引き起こされることはないが、それらを引き起こす可能性はあると自然に考える。しかし、グラフ上では*TがSの*親であっても、分布上ではTと*Sは*独立しているので、シンプソンの表の分布は、ケンドールの例と同様に、GがTと*Sを*、Tが*Sを引き起こすグ*ラフの因果マルコフ条件を満足する。

仮に、Simpsonが例に挙げた変数に与えた解釈（結局のところ、完全に想像上のものである）を無視して、その解釈によって排除されるであろう因果構造を考えてみることにする。実質的な連想を避けるために、AをT、BをG、Cを*Sに置き換え*て、図12のグラフ(i)を得ることができる。シンプソンやケンドールのような分布は、図12のグラフ(ii)のように、AとCが隣接せず、それぞれが*Bを*引き起こすグラフで実現することもできる[9]。

先ほどの変数の置換で、シンプソンの分布はグラフ(ii)には忠実だが、グラフ(i)には忠実でなく、しかも分布に忠実なグラフは(ii)だけである。

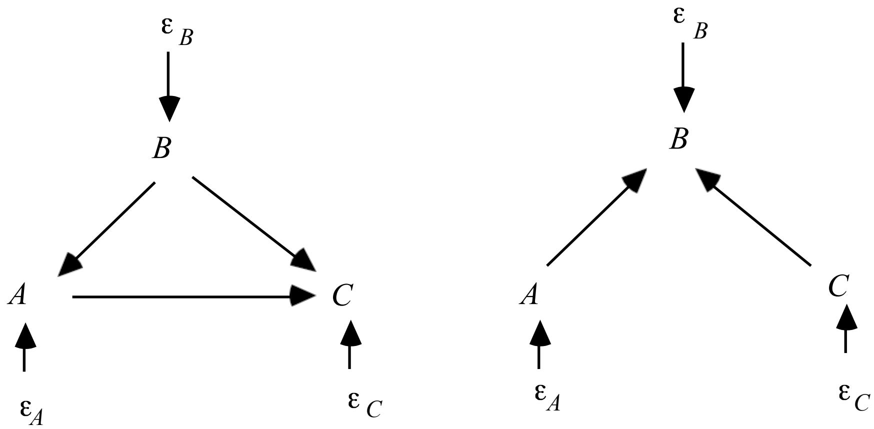
---

[8] その後の文献では、ある集団における独立性と依存性の関係が、下位集団における独立性と依存性の関係とどのように関連しているのか、またそのような事実の因果的意義について、他の多くの質問と混同されている。これらの疑問が混同されることの不幸な点は、これらの疑問には明確な答えがあることである。シンプソンに起因し、現在ではしばしば「シンプソンのパラドックス」と呼ばれているが、シンプソンが実際に提起した問題とは異なる状況を、コリン・ブライス（Colin Blyth）は次のように述べている（1972）：

同時に持つことが可能です。

(1) $p(a \mid b) < p(a \mid b')$ である。
と
(2) $p(a \mid bc) \square p(a \mid b'c)$ である。
(3) $p(a \mid bc') \square p(a \mid b'c')$

実際、Simpsonは(1)で等式、(2)と(3)で>を持つ。
[9] この点はBlalock(1961)に暗黙の了解があり、他の資料も同様であることは間違いない。



(i)　　　　**に**

図12

Judea Pearl (1988) は、グラフ (ii) のような因果構造が得られるとき、AとCが独立でありながら、Bを条件として依存すると考えるべき理由をベイズ的な例で説明している：あなたの車が発進するかどうかは、バッテリーが充電されているかどうか、タンクに燃料が入っているかどうかによって決まるが、これらの条件は互いに独立である。例えば、あなたの車のエンジンがかからないとします。その場合、あなたは、燃料タンクが空である確率とバッテリーが切れている確率があると考えます。次に、電池が切れていないことがわかったとする。その情報が加わると、燃料タンクが空である確率は変化しないのでしょうか?

仮に、AとCは独立だが、Bに依存するということがわかった場合、忠実性の条件から、もし因果構造があるとすれば、それは構造(ii)である。それでも、構造(i)は論理的に可能であり、もし変数がシンプソンが与えるような重要性を持っていれば、我々はもちろんそれを好むだろう。しかし、もし事前知識が構造(i)を必要としないのであれば、忠実性の条件を適用することで何を失うのだろうか。言い換えれば、分布に忠実でない因果構造を除外することで何を失うのだろうか。

線形の場合、パラメータ値（構造の線形係数と外生分散の値）は実空間を形成し、この空間の中でマルコフ条件が意味しない消失する部分相関を作る点の集合はルベーグ測度0となる。

定理3　.2： *M*を有向非周期グラフGと*n*個の線形係数a_1,..., anと外生変数v_1,..., v_kの*k*個の正変数を持つ線形モデルとする。とする。

> $M(\langle u_1,\ldots,u_n, u_{n+1},\ldots,u_{n+k}\rangle)$ は、値を指定することに矛盾しない分布である。$\langle u_1,\ldots,u_n, u_{n+1},\ldots,u_{n+k}\rangle$ for $a_1,\ldots,a_n$ and $v_1,\ldots v_k$.*Mのパラメータの値の空間*$\mathfrak{R}^{n+k}$上の確率測度*Pの*集合で、ルベーグ測度0を持つ$\mathfrak{R}^{n+k}$のあらゆる部分集合**V** **について**、$P(\mathbf{V})=0$とする。**Qの**すべての*qについて*、M（q）との確率分布inが*Gによって*線形暗示しない消光偏相関を持つように係数と分散値のベクトルの集合Qとし、すべてのP in ,$P(\mathbf{Q})=0$とする。

この定理は少し強化することができ、外因性変数と誤差変数のセットが共同独立である必要はなく、一対の独立で十分である。擬似非決定論的な場合、忠実性が破られるのは、変数間の関数従属性の非常に特殊な選択によってのみ可能である。外生変数が独立かつ正規分布している線形擬似不確定性システムの集団を考える。マルコフ条件が要求する条件付き独立関係は、線形係数のあらゆる可能な値に対して自動的に満たされる。それは、線形関数を構成する装置の働きによって保証されるからだ。しかし、マルコフ条件が要求*しない*条件付き独立関係は、装置の因果構造に忠実でない分布を特徴づけるもので、全く生成できないか、線形係数が非常に強い制約を満たす場合にのみ生成されるかのどちらかである。

同じモラルが他のクラスの関数にも適用される。離散変数については、3.2に類似した定理の正式な証明を試みていないが、直感的な根拠からそのような結果が期待されるはずである。グラフのマルコフ条件を満たす分布の因数分解式は、分布の自然なパラメトリゼーションを提供する。外生変数が*n個の値を持つ場合*、それは開区間（0,1）のコピーからなる*n-1個の*パラメトリック次元を決定する。内生変数*X*が*n個の値を*持つ場合、因数分解における条件付き確率P（X | **Parents** *（X））は*、*Xの*親の値の各ベクトルに対して（0,1）のコピーからなる別のn-1パラメータ*次元を*決定する。因数分解自体が内包しない条件付き独立関係を生成する確率値の集合は、このパラ

メータ空間では測度ゼロになると予想される。

3.6 ベイスターズ 解釈

我々はこの条件を、すべてのユニットが同じ因果構造を持つ集団における頻度について解釈してきた。この条件を、確率が主観的であるベイズ的解釈ではどうなるかを考えたい。現在の主観主義的な解釈では、確率は合理的で主観的な信念の度合いを理想化したものであるとする。厳密な主観主義では、有限の頻度はあり得るが、客観的な確率というものは存在しない。科学で研究されるシステムは決定論的であり、不確定性の外観は単にΘ無知によるものであると仮定するのである。ベイズ統計モデルの尤度構造は、ベイズ統計モデルでない普通の統計モデルと同じように見えることが多い。例えば、ベイズ線形モデルは、線形係数、分散、平均などを表すパラメータに対する分布を指定します。したがって、ベイズモデルは通常の線形モデルの混合物であり、測定された変数に対する共同分布は、このセクションで検討した条件を満たさない。

ベイズエージェントの信念度が、$f(X\,|\mathbf{Parents}(G,X)) = h(\mathbf{Parents}(G,X);\ )$という条件を満たす密度$f$で表Θされるとす Θる。ここで、パラメータは、その値がXの親に対する条件付き密度を決定する。ベイズ・エージェントは、.に対する分布もΘ持っているとする。 このような場合、因果マルΘコフ条件と因果最小性条件は、.に対するエージェンΘトの信念度を制約するものと理解される。を条件とする変数の主観的共同分布は条件を満たすが、通常、無条件の共同分布は条件を満たさない。

ここで、エージェントが代替可能な因果構造の集合**Gを**もてなし、**Gの**各構造*Gにお*Θ*いて*、以前のようにf（X｜Parents（G，X））＝h（**Parents**（G，X）；G）*を*保持するとする。そして、$_{G}$,G*を*条件としΘてエージェントの信念度を制約するために、因果

マルコフ条件と因果最小化条件を理解するのである。

このように理解すると、条件は信念の「合理的な」度合いに関する規範的な原則である。後の章では、臨床試験に関するベイズ的な提案を詳細に検討し、その提案が科学的専門家の信念の度合いについて行う仮定がマルコフ条件と一致することを論じることにする。

3.7 Axiomsの結果

因果マルコフ条件、最小性条件、忠実性条件の結果は本書を通して展開されるが、因果従属性と統計従属性の間のいくつかの重要なつながりをここで指摘しておく必要がある。

3.7.1 d- 分離

マルコフ条件は、因果グラフGが与えられたとき、Gに忠実な分布Pの独立性と条件付き独立性の関係を公理化するものである。例えば、頂点XとYの各対、およびXとYを含まない頂点**Q**の各集合について、XとYが**Q**を条件として独立かどうか、すなわち変数の集合間の原子的独立の事実をすべて知りたかったとする。マルコフ条件をGに直接適用する、つまり各頂点に対して定義を適用することは、一般に十分ではない。

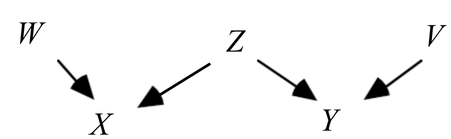


図13

例えば、図13のグラフに忠実な分布において、集合**Q**＝{Z}を条件としてXとYが独立かどうかを知りたかったとする。マルコフ条件を図13にそのまま適用すると、次のようになります：

$$W \perp\!\!\!\perp \{Z、Y、V\}。$$

$X \perp\!\!\!\perp \{Y, V\} \mid \{W, Z\}$。

$Z \perp\!\!\!\perp \{W, V\}$ です。

$Y \perp\!\!\!\perp W, X\} \mid \{V,Z\} \mid \{V,Z\} \mid \{V,Z\} \mid \{V,Z| \{V,Z\}$」です。

$V \perp\!\!\!\perp \{W, X, Z\}$

これらの事実が $\perp\!\!\!\perp XY \mid \{Z\}$ を内包していることは自明ではない。Pearl は、純粋にグラフィカルな

条件付き独立の特性化--これを**d-**separationと**呼ぶ--を**行い、Geiger、 Pearl, and Verma (Geiger and Pearl 1989a; Verma 1987)は、d-separationが、有向無サイクルグラフのマルコフ条件を満たすことによって生じる条件付き独立関係のすべてを特徴付けることを証明しました。

d-separationの定義は十分に直感的でないため、類推が役立つかもしれない。まず、無条件独立の状況を考えてみよう。グラフの無向きのパスを因果の流れを運ぶパイプと考え、各頂点はアクティブ（開いている）かインアクティブ（閉じている）かのバルブであるとする。もし頂点がコライダであれば、因果の流れはそこを通ることができないので、非活性である。例えば、図14上部の因果グラフにおいて、Xと*V* *は*空集合によってd-分離されているが、これは両者を結ぶ唯一の経路にあるYがコライダーであるためである。

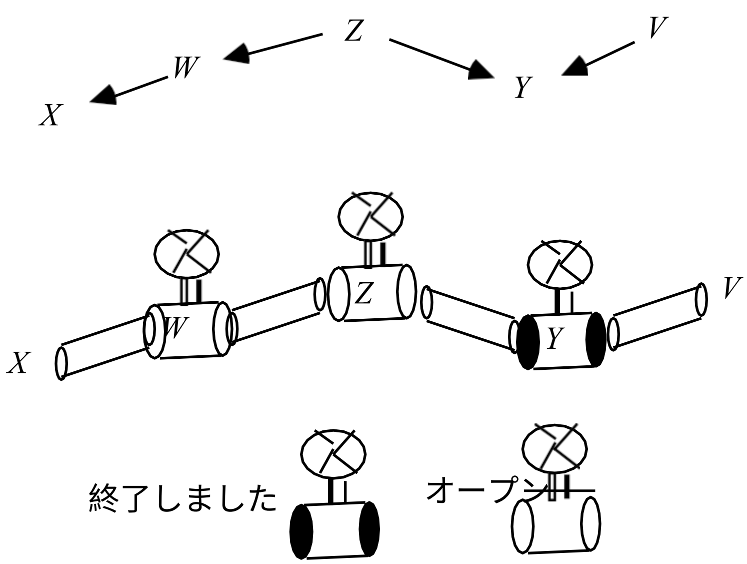


図14

*ノードへの条件付けはその状態をフリップフロップさせる。*図14のグラフにおいて、Xと*Yは*空集合が与えられるとd分離しないが、{*W*}、{Z}または{*W,Z}が*与えられるとd分離される。XとVは空集合が与えられるとd分離されるが、{*Y*}が与えられるとd

分離されることはない。ノンコライダーを条件付けると非アクティブになるというのは、マルコフ条件の直感と似ている。共通の原因を条件とすることで、効果は独立になり、より近接した原因を条件とすることで、効果は遠くの原因から独立になる。コライダーを条件付けるとアクティブになることは、前述の3.5.2節で述べたとおりである。

グラフ*G*が与えられたとき、任意の2つの頂点Xと*Y*が集合**Qによって**d個に区切られ、したがって*G*に忠実な分布において**Qを**条件として独立であるかどうかをチェックすることは、一見簡単である。パスは、その頂点がすべてアクティブである場合、すなわち、すべてのコライダーが**Qに**あり、その非コライダーがすべてQにある場合、アクティブである。

例えば図15では、XとYはUではd-separatedではないが、{V,Z} ではd-separatedである。

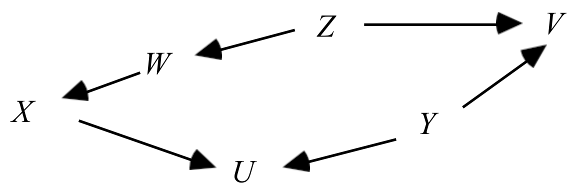


図15

X <- W <- Z -> V <- Yの経路はUが与えられたときに非活性であるが、XとYの間にアクティブな無向きの経路が1つある限り、Uが与えられたときにXとYはd-分離されることはない。

U.2番目の場合、{V,Z}の条件付けはX <- W <- Z -> V <- YのパスでVを活性化するが、条件付けはこのパス上でZを不活性化するので、不活性になる。X -> U <- Yのパスも{V,Z}にないパス上のUがコライダーなので{V,Z}が与えられて不活性である。XとYの間の無向きのパスはすべて{V,Z}で不活性化されるので、XとYは次のようにd分離される。

{V,Z} である。

しかし、残念ながら、その全貌はそれほど単純ではありません。コライダーの条件付けは、そのコライダーを活性化し、その子孫の条件付けも活性化する。例えば図16のグラフでは、WがUの子孫であるため、Wが与えられてもXとYはd分離されない。

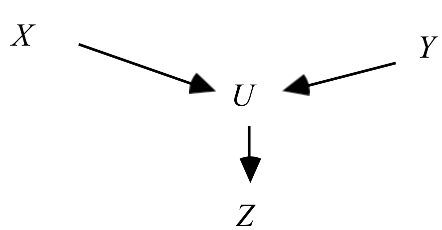


図16

有向無サイクルグラフGにおいて、XとYをGの頂点とし、$X \neq Y$、WをXまたはYを含まないGの頂点の集合とすると、XとYがGの**Wでd分離**するのは、（i）U上のすべてのコライダーが子孫を持つようなXとY間の無向パスUが存在しない場合に限る。

X≠Yで、XとYが**Wに**ない場合、Xと*Yが***Wに関して****d-**separatedでない場合に限り、XとYは集合**Wでd-connectedであるという**。**U**、**V**、**Wが***Gの*頂点の不連続集合であり、UとVが空でない場合、Uと**Vの**デカルト積におけるすべてのペア<U,*V*>が**Wを与えられた**ときにd-separatedである場合にのみ、Uと**V**は**W**を**与えられた**ときに**d-connected**であるという。

本質的な成果は以下の通りです：

> **定理3.3**：*P*(**V**)は、頂点集合**Vを**持つ有向非循環グラフ*Gに対して*、頂点**X**、**Y**、**Zの**すべての不連続集合に対して、Xと**Yが**Z**を**与えられたときにd分離する場合にのみ、**Zを**条件として独立である場合にのみ忠実となる。

定理3.4は、忠実性のより直感的な特徴付けを提供し、第5章で開発されたアルゴリズムの動機付けとなるものである。

> **定理3.4**：*P*(**V**)がある有向無サイクルグラフに忠実である場合、*P*(**V**)が頂点集合**Vを**持つ有向無サイクルグラフ*Gに*忠実なのは、以下の場合のみである。
>
> (i) *Gの*すべての頂点*X*，*Y*について，Xと*Yが*，*X*または*Yを*含まない*Gの*すべての頂点集合の条件下で従属する場合にのみ，Xと*Yが*隣接する；および
>
> (ii) Xが*Yに*隣接し、*Yが**Zに*隣接し、Xと*Z*が隣接しないようなすべての頂点*X*, *Y*, *Zに対して*、*X* -> *Y* <- Zは、*X*, *Zが**Yを*含み、*X*または*Zを*含まないすべての集合に対して従属条件である場合にのみ*Gの*サブグラフになります。

相関の研究は歴史的に正規分布と結びついており、その分布では消失性部分相関と条件付き独立は等価である。しかし、マルコフ条件と忠実条件は、消失性相関と偏相関を、正規性の仮定なしに、線形システムのグラフと因果構造に結びつけます。

このように、線形システムにとって、相関構造は因果構造へのガイドとなる。グラフ*G*の頂点*A*、Bと*G*の頂点のすべての部分集合**Cに対して**、$_{AB.\mathbf{C}} = 0$のときだけ、Aと*B*が**Cから**d個分離するとき、分布Pはグラフ*G*に**線形に忠実で**あると言うことにする。

定理3.5： Gが頂点集合**Vを**持つ有向無サイクルグラフで、AとBが**Vにあり**、Hが**Vに**含まれる場合、*G*は、Aと*B*が**Hを与えて**d-分離する場合にのみ ρ $_{AB.\mathbf{H}}=0$を線形に含めます。

分布PがグラフGに線形に忠実であるのは、頂点A、Bに対して以下の場合のみであることがわかる。

のGとGの頂点のすべての部分集合**Cは**、$_{AB.}$**Cの**場合のみ、**Cを**与えられたAおよびBはd-分離される。

= 定理3.5は、部分相関が消失する「再帰的」（すなわち非循環的）線形モデルにおいて因果構造を結びつけるパス分析の例（Wright, 1934; Simon, 1954; Blalock, 1961; Heise, 1975）のすべてを支える一般原則である。

この後の章では、分布が忠実であることを前提に、ある条件付き独立性や条件付き依存関係、あるいは消失性や非消失性の偏相関が因果構造から導かれることを頻繁に指摘することになる。逆に、ある条件付き独立関係や依存関係、あるいは偏相関の事実が与えられたとき、分布が忠実であれば、因果構造はある特性を持たなければならないことを、しばしば観察することになる。このような主張をするときは、必ず定理3.3、3.4、3.5の暗黙の帰結を利用していることになる。

3.7.2 マニピュレーション 定理

多くの実証研究の基本的な目的は、変化が自然に生じたものであれ、意図的な政策によって課されたものであれ、変化の影響を予測することである。観測された分布Pを利用して、ある変数の集合に新しい限界分布を課すような代替政策の効果について、どのように信頼できる予測を得ることができるだろうか。ある変数（例えば、薬物使用）の分布を直接変えるような政策を実施するという考え方は、結果として生じる分布PMANがPとは異なることを必要とします。Pだけでは$_{PMAN}$を予測することはできませんが、Pと因果構造は予測することができます。

仮に、外科医総監が喫煙の抑制を検討しているとして、"米国で誰も喫煙を許されな

かった場合、*Cancer*の分布はどうなるか？ "という質問をしたとする。ここで、**V** = *{飲酒、喫煙、がん}である*。説明のために、米国の実際の集団では、図17に示す因果構造が正しいと仮定する。

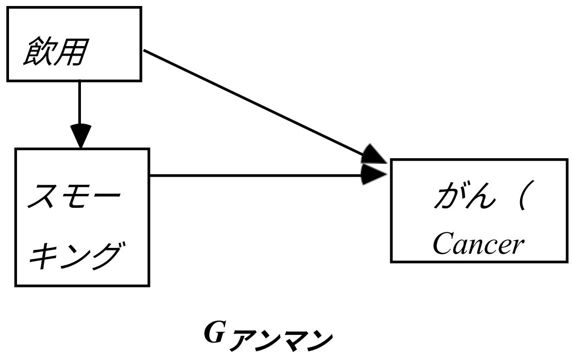
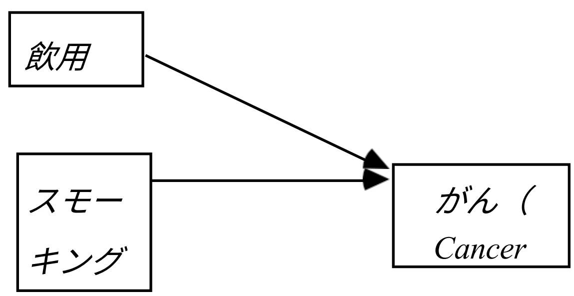


図17

実際にサンプリングされた（あるいはサンプリングと何らかの実験手順で作られた）集団を**操作されていない**集団、喫煙が禁止された仮想の集団を**操作された**集団と呼ぶことにする。仮に喫煙を禁止する政策が実行された場合、すべての人の喫煙を止めるという完全な効果が得られるが、母集団における*飲酒の*値には影響を与えないとする。すると、仮想の操作された集団の因果グラフは、操作されていない集団とは異なるものとなり、2つの集団で*Smokingの*分布が異なることになります。操作された因果グラフを図18に示す。



ジーメン

図18

操作されていないグラフと操作されたグラフの違いは、*GUnman*で操作された変数の親である頂点が、（操作の正確な形によっては）*GMan*で操作された変数の親でない場合があり、その逆もまた然りである。

喫煙を禁止することによって生じる*Smokingの*分布の変化をどのように表現すればよいのだろうか。一つの方法は、連邦政府の政策を表す変数の値が、二つの集団で異なることに注目することです。そこで、因果関係グラフにもう一つ、*Smokingの*原因である*Ban Smokingという*変数を導入することができます。すると、喫煙政策を表す新しい変数を含む完全な因果グラフは、図19のようになります。操作されていない

実際の集団では、*禁煙*変数が*オフ*で、仮説の集団では*禁煙*変数が*オンである*。実際の集団では$P(Smoking|Ban\ Smoking = off)$を測定し、喫煙が禁止された場合に生じるであろう仮想の集団では$P(Smoking = 0\ |Ban\ Smoking = on) = 1$である。**Vの**任意の部分集合**Xに対して**＝。

因果関係グラフの*{*喫煙、*飲酒*、*癌}を*、$P_{Unman(Ban\ Smoking)}(\mathbf{X})$とし、$P_{Man(Ban\ Smoking)}(\mathbf{V})$を$P(\mathrm{V}|\mathrm{Ban}\ Smoking = on)$とする。

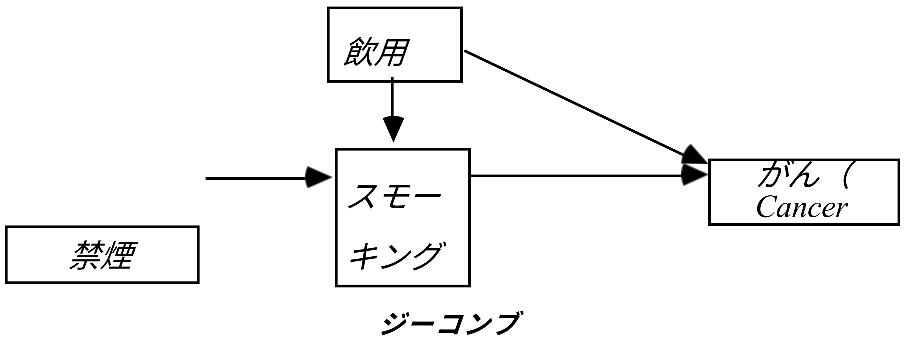


図19

ここで、$P_{Unman(Ban\ Smoking)}(Cancer|Smoking) = P_{Man(Ban\ Smoking)}(Cancer|Smoking)$ ($P_{Man(Ban\ Smoking)}(Cancer|Smoking)$ が定義される*Smoking*の値、つまり$Smoking = 0$の場合)かどうかを尋ねることができます。しかし、分布が忠実であれば、*Smoking*が与えられたときにCancerと*Ban* Smokingがd-separatedであるかどうかという問題に帰着します（この因果グラフではそうではありません）。さらに、$P_{Unman(Ban\ Smoking)}(Cancer)$
$\neq P_{Man(禁煙)}(Cancer)$は、Cancerが*Ban Smoking*からd分離されていないことを考えるとの空集合となる。しかし、これに対して$P_{Unman(Ban\ Smoking)}(\mathrm{Cancer}|\mathrm{Smoking}, Drinking) = P_{Man(Ban\ Smoking)}(\mathrm{Cancer}|\mathrm{Smoking}, Drinking)$ (for those of *Smoking* for which $P_{Man(Ban\ Smoking)}(\mathrm{Cancer}|\mathrm{Smoking}, Drinking)$ is defined, すなわち$Smoking = 0$) は、禁煙と癌が{喫煙、*飲酒*}*によって*d分離されているからです。この不変性の重要性は、観察された非喫煙者の部分集団における飲酒を与えた場合のがんの条件付き分布を考慮し、操作されていない集団における飲酒の分布を考慮することにより、喫煙が禁止された場合のがんの分布を予測できることです。

なお、$P_{Man(Ban\ Smoking)}(Cancer)$ の結論の*インプットの*1つは、禁煙が完全に成功し、*飲酒に*影響を与えないことである。この知識は、我々が行った*喫煙*、*飲酒*、*癌の*測定から得られたものではなく、他のソースから得られたものであると仮定されている。もちろん、その仮定が誤っていれば、$P_{MAN(Ban\ Smoking)}(Cancer)$ の計算が正しい

結果になる保証はない。仮に、喫煙を事実上禁止するのではなく、飲酒に影響を与えずに喫煙の可能性を低くするように介入する政策を考えた場合、操作変数*Ban* Smokingを含むシステム全体のグラフは図19と同じになり、グラフGUnmanは図17と同様になるが、操作したグラフ$_{GMAN}$は図17と同じようになり、むしろ18.介入しても、飲酒が喫煙に及ぼす影響を取り除くことはできない。

システムの予測分析には、3つの異なるグラフが含まれます：因果関係グラフ$_{GComb}$
操作内容を表す変数**Wを**含む因果関係グラフGUnmanと

*GComb*は、操作を表す変数を含まない変数集合**Vに対する***GComb*の部分グラフと、操作の結果生じる**V**内の変数間の因果関係を表す**Vに対する***GMan*のグラフである。*GMan*は、操作が*GUnman*の因果依存関係を「壊す」場合には*GUnman*の部分グラフとなり、そうでない場合にはGUnmanとGManは同じグラフとなる。

以下は正式な定義である：Gが変数**V**?**Wの**集合に対する有向非循環グラフであり、**V**∩**W**＝ϕである場合、**Wが***G*中の**Vに関して外生的であるのは、以下の**場合のみである。

GCombが変数**V**□**W の** 集合に対する有向非循環グラフであり、*P*（**V**□**W**）が*GComb*のマルコフ条件を満たす場合、**Wの**値を**w** 1から**w** **2に**変更することが**Vに関する***GComb*の**操作**であり、**WがVに関して**外生的である場合にのみ、*P*（**V|W = w 1**）≠ **P**（**V|W = w 2**）である。

PUnman(**W**)(**V**) = *P*(**V|W = w 1**), *PMan*(**W**)(**V**) = P(**V|W = w 2**) と定義し、同様に *P*(**V**) から形成される各種限界分布、条件付き分布についても定義する。

GCombを**結合グラフ、V**上の*GComb*の部分グラフを**操作されていないグラフ***GUnman*と呼ぶことにする。*(PUnman*(**W**)(**V**)は*GUnman*のマルコフ条件を満たすが、*GUnman*の部分グラフのマルコフ条件も満たす場合があることに注意されたい。これは、*G Comb*、ひいてはその部分グラフGUnmanには、操作された部分集団の分布を表すのに必要だが、操作されていない部分集団の分布を表すのには必要ない辺が含まれている場合があるからである)。

*V*が**Children(W)∩Vにある場合に限り、**Vは**Manipulated(W)**にある（つまり、Vは操作変数の1つによって直接影響を受ける変数である）；**Manipulated(W)**の変数は**直接操作されて**いると言うこともする。以下、**Wの**変数を**政策変数と呼ぶ**。

操作されたグラフG_{Man}は、$P_{Man(\mathbf{W})}(\mathbf{V})$がマルコフ条件を満たす$G_{Unman}$の部分グラフであり、**Manipulated**($\mathbf{W}$)のメンバーの親がG_{Unman}と異なるだけである。G_{Man}がどの部分グラフであるかは、操作の詳細と、$\mathbf{W} = \mathbf{w}_2$の部分集団の因果グラフがどうなっているかに依存する。例えば、喫煙が禁止されている場合、G_{Man}は所得と喫煙の間のエッジを含まない。一方、タバコに増税された場合、G_{Man}は所得と喫煙の間のエッジを含んでいる。このように定義された操作が与えられた場合、$P_{MAN(\mathbf{W})}(\mathbf{V})$がマルコフ条件を満たす$G_{Unman}$の部分グラフが必ず存在することを（第13章で）証明する予定です。マニピュレーションに関するすべての定理は、G_{Unman}の部分グラフであるG_{Man}のうち、以下のものに対して成り立つ。

$P_{Man(\mathbf{W})}(\mathbf{V})$はマルコフ条件を満たし、$G_{Unman}$とは**Manipulated**(**W**)のメンバーのせいぜい親が違うだけで、G_{Unman}とは異なる。

これらの定義は、**「操作の定理」を**内包しています：

定理3.6：（操作の定理）：頂点集合V⊡W上の有向非循環グラフGCombと、以下のマルコフ条件を満たす分布P(**V**⊡**W**)が与えられる。

G_{Comb}、**Wの**値をw1から**w2に**変更することが**Vに対する**G_{Comb}の操作である場合、GUnmanは操作されていないグラフ、GManは操作されたグラフ、である。

$$P_{Unman(\mathbf{W})}(\mathbf{V}) = \prod_{X \in \mathbf{V}} P_{Unman(\mathbf{W})}(\mathbf{X}|\mathbf{Parents}(G_{Unman}, X))$$

条件付き分布が定義されている**Vの**すべての値に対して、そのとき

$$P_{Man(\mathbf{W})}(\mathbf{V}) = \prod_{\mathbf{X} \in \mathbf{Manipulated}(\mathbf{W})} P_{Man(\mathbf{W})}(\mathrm{X}|\ \mathbf{Parents}(G_{Man}, X)) \times \prod_{\mathrm{X} \in \mathrm{V} \setminus \mathbf{Manipulated}(\mathbf{W})} P_{Unman(\mathbf{W})}(\mathrm{X}|\ \mathbf{Parents}(G_{Unman}, X))$$

。

は、各条件付き分布が定義されている**Vの**すべての値に対して、である。

この定理の重要性は、**Manipulated**(W)に含まれる各Xの因果構造と操作の直接効果（すなわち$P_{MAN(\mathbf{W})}(X\,|\mathbf{Parents}(X))$）がわかれば、操作されていない集団から共同分布が推定できることです。

操作の定理は、一対の変数間の因果機構が可逆的である場合には適用されない。この場合、一対の変数間の因果関係の方向が逆転した2つの部分集団が存在しうる。例えば、車のモーターが動くと車輪が回る（アクセルを踏んだとき）ことがあるが、車輪が回るとモーターが動く（車が坂道を転がるとき）こともある。私たちは、非

[10] この例を提案し、可逆的なメカニズムの問題を指摘してくれたMarek Druzdelに感謝する。

は、あるメカニズムが可逆的であるかどうかを判断するための方法です。喫煙と黄色い指のように、背景知識からメカニズムが可逆的でないことが明らかなケースもある、なぜなら黄色い指は喫煙の原因にならないからだ。また、関連する背景知識が得られない場合もあり、その場合は「操作の定理」が適用できるかどうかがわからない。

Rubin (1977; 1978) とそれに続く Pratt and Schlaifer (1988) は、観測されたシステムの集団が、すべての集団単位でいくつかの変数を直接操作することによって変化した場合、同じ変数の条件付き確率が等しくなるときの規則を提示している。彼らの様々なルールは、図17、18、19の考察で示された「操作の定理」の特殊な場合、つまり、ある変数が操作され、その介入によってその変数が操作されていないグラフの原因から独立する場合の直接的な結果であることを7章で示す。

操作定理はマルコフ条件の帰結であるため、別途正当化する必要はない。操作の定理は抽象的ですが、常に正しいとは言えないまでも、日常的に行われている推論を一般的に定式化したものに過ぎません。例えば、回帰モデルを用いて、回帰因子のいくつかに値を強制する政策の効果を予測する場合、操作の定理を応用していることになる。もちろん、回帰モデルの因果関係や統計的仮定が誤っていたり、実際に行われた変化が操作の条件を満たしていなかったりすれば、予測は正しく行われないかもしれない。このような失敗の例もあります。各ユニットの変数の値が他のユニットの値に依存し、その依存関係が因果関係グラフに表現されていない場合、操作の定理を適用すると、誤解を招く予測を与えることがある。公共政策の議論の中には、この要件に対する不合理な違反が例示されている。最近、自動車保険会社が出資する研究機関が、さまざまな種類の車の乗員の死亡率を、車の長さや重さなど

の変数に対して非線形回帰したところ、当然のことながら、車が小さいほど死亡率が高くなることがわかった。この統計的な分析結果をもとに、連邦政府が提案した自動車の小型化政策が高速道路での死亡率を高めると主張したのである。しかし、もちろん、あるサイズの車の死亡率は、その車に乗っている他の車のサイズの分布に依存します。

ある政策や介入によって、どの変数が直接影響を受けるかを間違えることがある。このような場合に「操作の定理」を暗黙のうちに適用すると、失望につながることがある。後章で述べるように、喫煙、肺がん、死亡率に関する文献には、予測が誤った例として、間違いなく、介入によってどの変数が直接操作されるかという誤った判断が原因であったことが、鮮明に示されている。

母集団（または標本）のユニット間の変数**Vの**セットの値の分布を意図的に変更するすべての介入は、**Vの**直接操作のための条件を満たす*必要が*あり、他のものはない理由はありません。しかし、実験計画における主な目的の1つは、実験操作が実際に意図された変数の直接的な操作であり、それ以外のものでないことを確認することである。例えば、ブラインドデザイン、ダブルブラインドデザインのポイントは、まさに治療変数のみの直接的な操作を実験で得ることである。動物を使った薬物試験における慢性的な傷の懸念は、本質的に、関心のある結果変数に関して、結果変数と薬理学的変数が直接操作されていることを懸念しているのです。一般に、介入によって直接操作される変数を間違えると、介入の結果の予測は失敗する。

このセクションでの我々の議論は、システムの因果構造が完全に知られていることを前提としている。第6章と第7章では、未知の因果グラフに忠実な分布の測定変数上のマージンであると仮定し、介入がここで定義した意味での直接操作にあたると仮定して、操作されていない分布から介入の効果をいつ、どのように予測できるかを検討します。

3.8 決定論

忠実性の条件が破られるもう一つの方法は、変数間に決定論的な関係がある場合である。ここでは、変数間の決定論的な関係によって、どのような余分な条件付き独立関係が内包されるかを判断するためのルールをいくつか挙げていくことにする。

グラフ内の変数間に決定論的関係があるとき、決定論的関係とマルコフ条件によっ

て、マルコフ条件だけでは包含されない条件付き独立性が存在する。例えば、Gが**V**上の有向無サイクルグラフで、**V**がZと*A*を含み、**Z**が*A*を決定する場合、*A*は**Z**が与えられた**V**（**Z** □{*A*}）から独立している。もし**Z**が*A*の親の適切な部分集合であれば、これは、*A*が**Z**与えられたその他の親から独立しており、また**Z**与えられたその子孫およびその非孫からも独立していることを内包している。

のメンバーはAの子であり、その場合、その子から考えると、Aは親を含む他のすべての変数から独立している。の親でない先祖が**Zに**含まれる可能性もある。

A.これらの場合、それぞれマルコフ条件だけでは内包されない条件付き独立関係を内包している。例えば、図20のグラフを考えてみよう。

$$\begin{tikzcd} A \arrow[r] \arrow[rr, bend left] & B \arrow[r] & C \end{tikzcd}$$

図20

マルコフ条件だけでは、A、B、Cの間に条件付き独立関係は成立しない。ただし、祖父母Aが孫Cを決定する場合は、$C \perp\!\!\!\perp B|A$である。親Bが子Cを決定するならば、$C \perp\!\!\!\perp A|B$である。子Cが親Bを決定するのであれば、$B \perp\!\!\!\perp A \mid C$.

したがって、d-分離可能性関係は、マルコフ条件と一組の決定論的関係が内包する条件付き独立性のすべてを捕らえることはできない。そこで、マルコフ条件と変数間の決定論的関係の集合が与えられたときに、変数の条件付き独立性を内包するグラフ条件を探すことにする。

Geigerは、Markov条件とMinimality条件、および変数間のある種の決定論的関係によってもたらされる条件付き独立性をグラフィカルに決定するための、簡単で証明可能な完全規則を提案した。Geiger (1990)に従って、Aと**Zを**含む**V**上の有向無サイクルグラフGにおいて、頂点AがG中の親の決定論的関数である場合、**決定論的変数**であると言う(変数AがG中に親を持たず、一定値を持つ場合、Aは決定論的変数であることに注意)。**X**、**Y**、Zが**Vの**変数の3つの不連続部分集合であるとき、**Xの**任意のメン

9バーと**Yの**任意のメンバーの間に、各コライダーが**Zに**子孫を持ち、*U*上の他の変数が**Zによって**関数的に決定されないような無向パス*Uが*存在しない場合に限り、**Zを**与えられたXとYは**D-分離する**。Geigerは、*Gの*マルコフ条件と最小化条件を満たすすべての分布と決定論的関係に対して、Xと**YがZを与えられたときに**独立である場合にのみ、XとYが**Zを**与えられたときにD分離することを示しましたが、我々はGeigerの規則が決定論的関係のもっと広いクラスで正しいことを証明します。

GをV上の有向非循環グラフとし、**Deterministic**（**V**）を**V**中の変数の順序付きタプルの集合とすると、**Deterministic**（**V**）中の各タプル*D*について、*D*が＜V1，...，Vn＞であればVnはV_1、...、Vn-1の決定性関数であり、V1の任意の部分集合の決定性関数でない。
V$_1$,...,*Vn-1*}は*Vn*を**決定するとも**言う。なお、Vnは*G*の祖先である可能性がある。V_1,...,Vn-1のメンバのうち。また、*A*が*B*を決定し、*B*が*A*を決定する場合、**Deterministic**(**V**)は<*A*,B>と<*B*,*A*>の両方を含む。**Deterministic**(**V**)は、変数間の決定論的な関係を伴う場合、それらの決定論的な関係は**Deterministic**(**V**)に含まれるという意味で完全であると仮定する。(例えば、*A*が*B*を決定し、*B*が*C*を決定する場合、*A*は*C*を決定する）**Det**(**Z**)は**Zの**ある部分集合によって決定される変数の集合であり、変数*A*が一定の値を持つ場合、それは空集合によって決定され、すべての**Zに対して** **Det**(**Z**)にあると言う。

Deterministic(**V**)は独立性だけでなく、変数間の依存性も内包することができることに注意。**Zが***A*を決定し、*Z*が**Zの**メンバである場合、*A*は*Z*が与えられた**Zpha**{*Z*}**に**依存する(他の依存関係も同様に**Deterministic**(**V**)が内包することができる)。もし**Deterministic**(**V**)と有向無サイクルグラフ*G*が両立しない場合、以下の定理3.7は空しく成立するが、**Deterministic**(**V**)と*G*が両立するかどうかを判定するテストがあることが望ましいことは明らかである。

ここでは、Geigerが考える決定論的関係の種類に限定されないように、D-separabilityの概念を拡張することにする。Gを頂点集合**Vを**持つ有向無サイクルグラフ、*Z*を*X*または*Y*を含まない頂点集合、*X*≠*Y*とすると、XとYはZと**Deterministic**（**V**）を与えられたとき、Xと*Y*の間に、*U*上の各コライダーが**Zに**子孫を持ち、*U*上の他の頂点が**Det**（**Z**）にない無向パス*U*が*G*に存在しない場合にのみ**D分離可能で**あり、それ以外

2の場合にはX

$\neq Y$であり、XとYが**Zに**ない場合、XとYはZと**Deterministic**(**V**)があればD-**connected** **である。**同様に、**X** , **Y** , Zが変数の不連続集合であり、XとYが空でない場合、Xと**Y** **の**デカルト積における各対$<X, Y>$がZと**Deterministic**(**V**)を与えられてD-separatedである場合に限り、またそれ以外の場合は**X** , **Y** , **Z**が不連続で、Xと**Y**が空でない場合はXとYがZと**Deterministic**(**V**)を与えて**D-connectedとなる。**

定理3.7：Gが**V**上の有向非循環グラフ、**X** , **Y**, Zが**Vの**不連続部分集合、P（**V**）がGのマルコフ条件と**Deterministic**（V）の決定論的関係を満たす場合、XとYがZと**Deterministic**（**V**）**が与えられた**D分離であれば、XとYはP中の**Zが**与えられた独立であるとする。

例えば、Gが図21のグラフで、**Deterministic**(**V**) = *{<A,B>, <B,C> だとする、<A,C>} となります。*

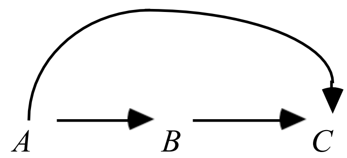


図21

Aと**Deterministic**(**V**)があればBとCはD分離され、*Bが*あればAとCはD分離されると**Deterministic**(**V**)です。

Gが図21のグラフのまま、**Deterministic**(**V**) = *{<A,B>, <B,A> となったとする、<b,c>, <c,b>, <a,c>, <c,a>} となる。*これまでのD分離可能関係に加え、今度は*A*とBは、*CがAを*決定するため、Cと**Deterministic**(V)が与えられ、D分離される。

親が子によって決定されるため、条件付き独立性が内包される場合もある。　図22のグラフを考えてみると、**Deterministic**(**V**) = *{<y ,w,z>,<z,*y>,*<z,w>} とする。*XとTはZと**Deterministic**(**V**)があればD分離される。なぜなら、*Zは*Yと*Wを*決定し、YとWはXと*Tの*間の唯一の無向パス上の非コライダーであるからである。

$$X \longrightarrow Y \longrightarrow Z \longleftarrow W \longleftarrow T$$

図22

最後に、*XがZのどの*親も決定しないにもかかわらず、*Zの*親でない祖先*XがZを*決定する可能性があることに注意する。 *Gを*図23のグラフ、**Deterministic**（**V**）＝*{<*X *,Z>} とする。X*、*R*、*Zは*それぞれ2つの値を持ち、Yは4つの値を持つとする。次のよ

うな分布を考える（ここで、各変数の親を条件とした確率を与える）：

$$P（X=0）=0.2$$

$$P(R = 0) = 0.3$$

$$p(y = 0|x = 0,r = 0) = 1$$

$$p(y = 1|x = 0,r = 1) = 1$$

$$p(y = 2|x = 1, r = 0) = 1$$
$$p(y = 3|x = 1, r = 1) = 1$$
$$p(z = 0|y = 0) = 1$$
$$p(z = 0|y = 1) = 1$$
$$p(z = 1|y = 2) = 1$$
$$p(z = 1|y = 3) = 1$$

事実上，YはRとXの両方の値を符号化し，ZはYを復号してXの値と一致させる．

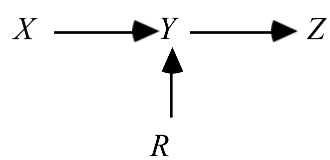


図23

Xと**Deterministic**(**V**)があればYとZはD分離し、Yと**Deterministic**(**V**)があればXとZはD分離することがわかる。

次の例は、与えられた有向無サイクルグラフGのマルコフ条件を満たす分布の集合と、マルコフ条件と決定論的な変数間の関係を満たす分布の集合の間の興味深い違いを指摘するものである。

Gを図24に示すグラフとする。任意の有向無サイクルグラフについて、そのグラフのマルコフ条件を満たす確率分布の集合には、そのグラフの最小化条件も満たす分布がある。しかし、**Deterministic**(**V**) = $\{<X, Y>\}$と仮定する。この場合、マルコフ*条件*と指定された決定論的関係を満たす分布のうち、最小性条件も満たす分布は存在しない。マルコフ条件と指定された決定論的関係を満たすすべての分布は、図24の$Z \rightarrow Y$エッジを含まない部分グラフに忠実である。このことは、有向無サイクルグラフGと決定論的関係の集合に対してマルコフ条件を満たすことによって付随する条件付き独立関係をすべて見つけるためには、頂点集合**V**を持つGの様々な部分グラフG'

において、**Vの** *各Yに対して* **Parents(** *G', Y* **)** のどの部分集合もYを決定しないD分離性を

検定する必要があることを示しています。

$$X \longrightarrow Y \longleftarrow Z$$

図24

このような決定論的な関係が得られる場合の因果関係グラフを構築するアルゴリズムについては検討しない。また、変数の集合**X**が変数*Y*を決定するかどうかを決定するためのテストについても検討しない。

3.9 背景 備考

因果関係と統計的制約の両方を表す仮説のあいまいな使用は、統計学とほぼ同じ歴史があります。現代では、世紀初頭にスピアマン(1904)がこの考え方を用いたことが、統計的心理測定の起源とされています。統計的仮説と因果関係の主張を同時に表す有向グラフは、Sewell Wright (1934)によって導入され、それ以来、特に線形モデルとの関連で使用されてきた。線形モデルと偏相関制約の間の接続は、いくつかの特殊なグラフについて、未測定の共通原因を持たない理論についてはSimon (1954) とBlalock (1961) によって、潜在変数を持つ理論についてはCostner (1971) とLazarsfeld and Henry (1968) によって説明されているが、一般的な特徴付けは現れていない。偏相関を持つ線形モデルのグラフ構造の分布のない接続がGlymour, Scheines, Spirtes and Kelly (1987)で開発されたが、一次偏相関のみで、環状グラフを含む。Geiger and Pearl (1989a) は、任意の有向無サイクルグラフに対して、忠実な分布が存在することを示しました。ここで定理3.5として与えられた一般的な特徴はSpirtes (1989)によるものだが、マルコフ条件、線形性、偏相関の関係はSimonとBlalockによってすでに理解されていたようで、Kiiveri and Speed (1982) に明示されている。操作可能性の定理は、実験計画や計量経済学のショック分析において、暗黙のうちに何度も使われてきたが、これまで明示的に定式化されたことはなかったようだ。この定理は、

Spirtes, Glymour, Scheines, Meek, Fienberg and Slate (1991) で初めて特別な事例として示された。最小性条件とd-分離性の考え方はPearl (1988)によるもので、d-分離性がマルコフ条件の結果を決定することの証明はVerma (1987), Pearl, and Geiger (1989a) によるものである。定理3.4を内包する結果がPearl, Geiger, and Verma (1990)によって述べられている。定理3.4はSpirtes, Glymour, and Scheines (1990c)で因果推論アルゴリズムの基礎として用いられた。D分離可能性はGeiger (1990)に記載されている。

第4章

統計的識別性

実験操作がなければ、統計的関係から因果構造を推論するためのあらゆる可能な方法の解像力は、統計的区別不能性によって制限される。もし2つの因果構造が同じ統計量を同じように説明できるのであれば、どんな統計量もそれらを区別することはできない。因果関係の仮説に対する統計的区別不能の概念は、因果構造を表す有向グラフと、関連する変数の共同分布を表す確率の間の接続に課す制限によって変化する。マルコフ条件と最小化条件を満たすことだけを要求する場合、2つの因果グラフは、同じクラスの分布が一方のグラフと他方のグラフでこれらの条件を満たす場合、区別がつかないことになる。分布がグラフ構造に忠実であることを要求すれば、別の統計的区別不能関係が得られ、さらに分布が線形構造に一致しなければならない場合など、さまざまなケースがある。それぞれのケースにおいて、問題はグラフ理論的に区別不能クラスを特徴づけることであり、そうすることによって初めて、因果グラフと分布を結びつける一般的な仮定では区別できない因果構造を一般的に理解することができる。

統計的性質から因果関係を推論するあらゆる可能な方法の解決力について、関連する多くの考察がある。グラフと分布のつながりに関する公理があるとき、公理を満たす*少なくとも一つの*確率分布を共有するために、二つのグラフはどのようなグラ

フ理論的構造を共有しなければならないか。例えば、2つの異なるグラフが、最小条件とマルコフ条件を満たす1つの同じ分布を認めるのはどのような場合か？ 二つの異なるグラフが、一方は最小性条件とマルコフ条件、他方は忠実性条件とマルコフ条件を満たす一つの同じ分布を認めるのはどんなときか？ 逆に、*ある*有向アシックスグラフに対してマルコフ条件と最小性条件（あるいはそれに加えて忠実性条件）を満たす確率分布が与えられた場合、その分布とこれらの条件に合致する*すべてのグラフの集合は*何であるか？ 最後に、計量理論的な問題がある。もし、「忠実性」のようなより限定的な仮定のもとで、因果構造を特定する手順が存在するならば、しかし、常に「忠実性」のような限定的な仮定のもとで、因果構造を特定するとは限らない。

マルコフ条件や最小条件などの弱い仮定のもとで、その手続きが失敗するケースはどの程度あり得るか? 分布の集合に関する様々な自然な尺度のもとで、例えば、グラフの最小性条件とマルコフ条件を満たすが、グラフに忠実でない分布の集合の尺度はどの程度か。

これらは、非実験データから構造に至るまで、人間であれコンピュータであれ、あらゆる可能な推論手順の限界に関する基本的な疑問である。測定された変数の系が因果的に十分であるとき、これらの質問の多くに対する答えを提供することになる。グラフが測定されていない共通の原因を表す変数を含むことができる場合、統計的区別不能性はあまり理解されない。

4.1 Strong Statistical Indistinguishability

2つの有向非循環グラフ*G*、*G'*は、同じ頂点集合**V**を持ち、*G*の最小条件とマルコフ条件を満たす**V**上のあらゆる分布*P*が*G'*の条件を満たし、その逆の場合のみ**強く統計的に区別できない（s.s.i）。**

もちろん、2つの構造がs.s.i.であるということは、因果構造が1つで同じであること、あるいは両者の違いがどのような手段によっても検出できないことを意味しない。XとYという2つの変数の相関から、XがYを引き起こすのか、YがXを引き起こすのか、それとも第3の共通原因Zが存在するのかを区別することはできないが、これらの選択肢は実験や他の手段によって区別することができる。

強い統計的区別不能性は、2つのグラフが同じ基礎となる無向グラフと同じ衝突を持

つという単純な関係で特徴づけられる：

定理4.1： 2つの有向無サイクルグラフG_1,G2は、(i)同じ頂点集合**Vを持ち**、(ii)頂点v1とv2がG1で隣接するのはG2で隣接する場合に限られ、(iii)**V**中のすべてのトリプルv1,v2,v3に対して、V_1 -> v2 <- v3はG2のサブグラフである場合に限り、かつ、そうである場合に、強く統計的に区別されない。

任意の有向無サイクルグラフGが与えられたとき、Gからのグラフs.s.i.は、Gのすべての衝突を保存するGの辺の方向の任意の反転セットによって得ることができる正確なものです。

G. 2つのグラフがs.s.i.であるかどうかの判断には、$O(n^3)$の計算が必要である。は頂点の数です。

図1のグラフG_1とG_2はs.s.i.ですが、G_1とG_3、G_2とG_3はs.s.i.ではないです。

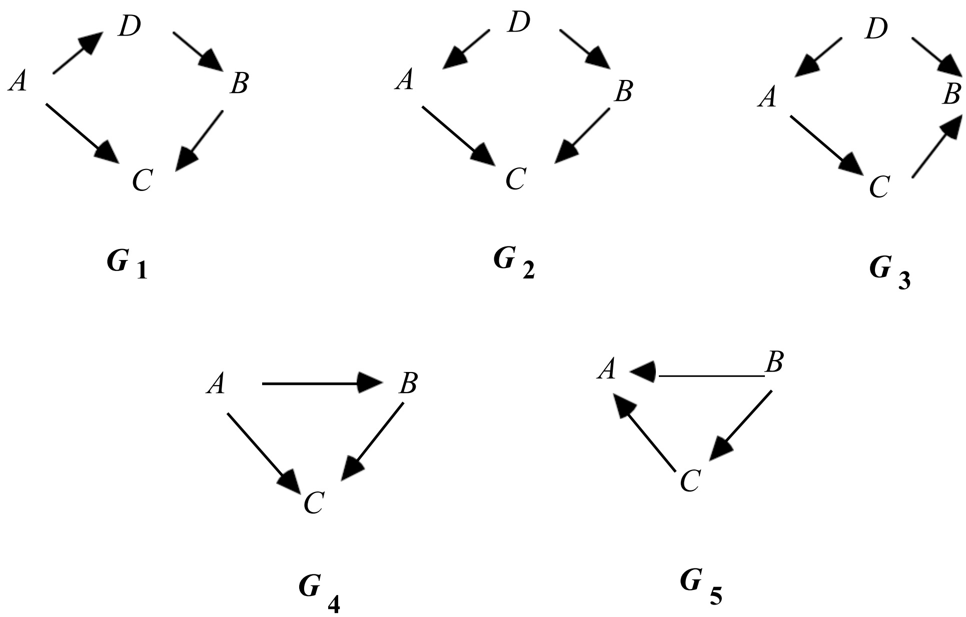


図1

ただし、変数の集合Vが、例えば既知の時間順序によって全順序化され、*P*(**V**)が正である場合、*P*(**V**)が最小性とマルコフ条件を満たす*ユニークな*グラフが存在することに注意してください。(Pearl1988のCorollary 3を参照)。

4.2 忠実な 識別可能性

Gに忠実な分布が*G'*にも忠実であり、その逆もまた然りである場合にのみ、2つの有向無サイクルグラフ*G*、*G'*は**忠実に区別できない**（f.i.）と言うことにする。問題は忠実な区別がつかないことをグラフィカルに特徴づけることである。

定理4.2：2つの有向無サイクルグラフGとHが忠実に区別できないのは、（i）同じ頂点集合を持ち、（ii）*Gにおいて*任意の2頂点が隣接するのは*Hにおいて*隣接する場合に限られ、（iii）*G*または*Hにおいて*X*は*Y*に*隣接し、Yは*Zに*隣接するがXは*Zに*隣接しないといった任意の3頂点*X*, *Y*, *Zが*、*Gにおいて*$X \to Y \leftarrow Z$*として*配向する場合、*Hにおいて*そう配向するときに限られる。

2つのグラフの忠実な区別がつかないかどうかは、*n*を頂点の数とすると$O(n^3)$で判断できる。

定理4.1および4.2から、2つのグラフが強く統計的に区別できない場合、忠実に区別できないことは明らかであるが、必ずしもその逆ではない。図1のグラフ*G*4とG5はs.s.i.ではないが、f.i.である。

f.i.グラフのクラスは、**パターンで**表現することができる。**パターンとは、**有向および無向のエッジを持つ混合グラフのことである。グラフ*Gは*、*以下の*場合にのみ、Π **によって表されるグラフの集合**に含まれる：

(i) Gは、以下のような隣接関係を持ちます。Π ;

(ii) Π において、*A*-*B*間のエッジが*A→Bを向いている*場合、*Gにおいて*A→B*を向い*ていることになる；

(iii) Π が*Gのパス*<X ,*Y* ,*Z*>上の非遮蔽コライダであるならば、YはGの <*X* ,*Y* ,*Z*>上の非遮蔽コライダである。

例えば、3頂点上のすべての完全無サイクル有向グラフの集合は、同じ頂点集合上の完全無向グラフからなるパターンで表現できる忠実な区別不能クラスを形成しています。有向無サイクルグラフの忠実な区別不能クラスのパターンが有向辺を持たず、純粋に無向であるとき、有向グラフが表す統計仮説は、そのパターンに対応する

無向独立グラフの統計仮説と同等となる。

4.3 Weak Statistical 識別可能性

前の2つのセクションで特徴付けられた区別不能関係は、与えられたグラフと同じクラスの確率分布を収容できるグラフを求めるものである。少なくとも部分的には、ある変数の集合に対する特定の確率分布から出発して、それらの頂点にあるすべての有向無尽グラフの集合を求め、そのグラフが矛盾しないようにすることで、状況を逆転させることができる。

を、与えられた分布で表現する。その答えは、確率と原因との関連に関する我々の仮定が、因果構造をどれだけ過小に決定しているかを特徴づける。マルコフと最小性だけを仮定すると、これらの2つの条件（正の下で）と有向独立グラフの定義条件の等価性は、与えられた分布*Pに対して*2つの条件を満たすすべてのグラフの集合を生成する（非現実的な）アルゴリズムを提供します。*Pの*あらゆる変数の順序に対して、その順序に適合する（すなわち、*GにおいてAがBの*子孫でない場合にのみ、順序において*AがBに*先行する）有向無サイクルグラフ*Gが*存在し、*Pの*最小性条件とマルコフ条件を満たす。これは、*Pの*順序と条件付き独立関係を想定し、有向独立グラフの定義を適用することで生成することができる。陽性を仮定しないアルゴリズムがPearl(1988)によって与えられている。そのアルゴリズムによると、*Ord*を変数の全順序とし、**Predecessors**(*Ord,X*) を順序*OrdにおけるXの*前任者とする。各変数Xについて、*GにおけるXの*親を、Pにおける**Rが**与えられたときに*Xが***Predecessors**(*Ord,X*) \Rから独立するような**Predecessors**(Ord,*X*)の最小の部分集合**Rとすると、***PはPの*最小性とマルコフ条件を満足することが分かる。

この場合、*Pに*忠実なすべてのグラフは、忠実な区別不能クラス、すなわち、*Pに*忠実な任意の1つのグラフからf.i.したすべてのグラフの集合を形成します。次の章では、分布の性質から忠実な区別不能クラスを生成するアルゴリズムをいくつか紹介します。

因果グラフと確率分布を結びつける公理が与えられたとき、どのようなグラフのペア*G*、*G'*に対して、Gと*G'*の両方について公理を満たす確率分布が*存在するかを*問うことは意味がある。ここで、2つのグラフが**弱忠実区別不能**（w.f.i.）であるのは、その両方に忠実な確率分布が存在する場合のみとする。また、2つのグラフが**弱く統計**

的に区別できない（w.s.i.）と言うのは、両者に対して最小値条件とマルコフ条件を満たす確率分布が存在する場合のみである。弱い忠実な区別が忠実な区別と等価であることが証明される：

定理4.3：2つの有向無サイクルグラフが忠実に区別できないのは、一方に忠実な分布が他方にも忠実である場合、およびその逆である場合に限られる。w.f.i.である場合に限り、f.i.である。

この定理は、忠実性が頂点集合上の確率分布の集合を、忠実な区別がつかないグラフの同値類に正確に対応する同値類に分割することを教えてくれる。ある分布があるグラフ*G*に忠実であるならば、その分布は*G*から忠実に区別できないすべてのグラフに忠実であり、またそのグラフにのみ忠実であることが導かれる。

一般にこれほど素晴らしいマッチングを期待する理由はない。最小条件とマルコフ条件だけを仮定するとする。2つの異なるグラフ、*G*、*G'*に対して、これらの公理を満たす分布*P*はどのような条件で存在するのだろうか。答えは*違います*。まさに、GとG'が強く統計的に区別できないときです。図2に示す2つのグラフはs.s.i.ではありませんが、両者に対して最小条件とマルコフ条件を満たす分布が存在します:

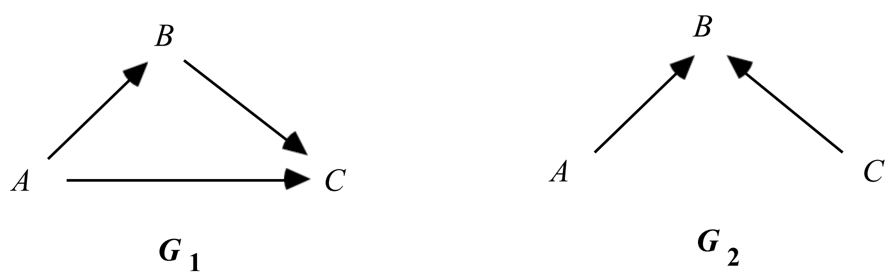


図2

シンプソンの「パラドックス」における分布は、すでに3章で見たように、その一例を示している。ある分布が2つのグラフGと*G'*に対して最小性条件とマルコフ条件を満たす場合、Gと*G'*は同じ辺と同じ衝突子を持ち、他の辺で適切な条件が満たされれば、一方のグラフのG*1などの*三角形が他方のグラフの*G*2などの衝突子に置き換えられることを除いて、推測されます。適切な」条件をどのように特徴づけるかはわからない。しかし、関連する興味ある性質があり、それを特徴付けることができる。

図2のグラフG1に忠実な分布がグラフG2に忠実であるはずがないが、G1の最小化条件とマルコフ条件を満たす分布は、グラフG2に忠実であることができる。このようなことは、いったいどんなときに起こりうるのでしょうか。言い換えれば、シンプソンの「パラドックス」の一般化は、どのようなときに起こりうるのだろうか。確率分布*P*が*G*の最小性・マルコフ条件を満たし、かつ*P*が*グラフHに忠実である*場合、Gと*H*の関係はどうなるのか?

定理4.4：確率分布Pが有向無サイクルグラフGとHのマルコフ条件を満たし、PがHに忠実である場合、すべての頂点X, Yについて、X, YがHで隣接していればGでも隣接している。

定理4.5：確率分布Pが有向無サイクルグラフGとHのマルコフ条件と最小値条件を満たし、PがグラフHに忠実な場合、（i）X -> Y <- ZがHにあり、XがHでZに隣接しないようなすべてのX, Y, Zに対して、X -> Yのどちらかが

を満たすような頂点の三重構造X, Y, Zに対して、(ii) Gにおいて <- Z が隣接している。

$X\rightarrow Y$□ZがGで、XがGでZに隣接していない場合、XがHでYに隣接し、YがHでZに隣接していれば$X\rightarrow Y$□Zとなる。

定理4.1：確率分布Pが有向無サイクルグラフGのマルコフ条件を満たし、Pが有向無サイクルグラフHに忠実で、GとHが（例えば時間によるように）$X\rightarrow Y$という変数の順序で$X<Y$のときだけ一致する場合、HはGの部分グラフとなる。

4.4 剛性 識別性

強い統計的区別不能、忠実な統計的区別不能、弱い統計的区別不能の概念に加えて、さらにもう1つの概念がある。2つの有向非周期グラフGとG'が、頂点の共通集合**O**上で、ある意味で統計的に区別できないとする。そうすると、実験なしに、**Oの**変数を測定しても、どちらのグラフがデータを生成した因果構造を正しく記述しているかを確実に決定することはできない。しかし、GやG'の変数以外の変数が測定され、**Oの**変数と適切な因果関係に立っていれば、GとG'を区別できるかもしれない。例えば、次の単純なグラフは、s.s.i.とf.i.の両方である（ここでA、Bは測定されて**O**内にあると仮定する）。

G_1　　G_2

図3

しかし、Aの原因である変数C、あるいはAと共通の原因を持ち、BとはAを介する可

能性を除いては何の関係もない変数Cも測定すれば、二つの構造を区別することができる。

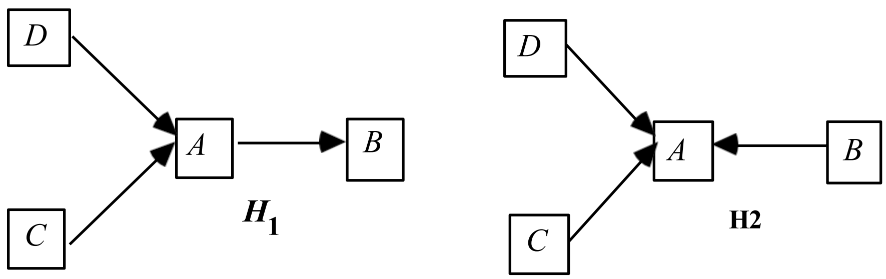


図4

図4のグラフはf.i.でもs.s.i.でもありません。w.s.i.でないグラフに埋め込むことができるw.s.i.構造の例を示すことは同様に簡単です。余分な変数を測定することによって区別できる因果的十分構造はどれですか。この問いに答えるには、さらにいくつかの定義が必要である。

$_G1$、$_G2$を共通の頂点集合**O**を持つ2つの有向非周期グラフとし、H_1、$_H2$を**O**を含む共通の頂点集合Uを持つ有向グラフとし、以下のようにする。

(i) は、**O**上の$_H1$のサブグラフがG_1、**O**上の$_H2$のサブグラフがG_2である;

(ii) $_{H1}$にあり$_G1$にない有向辺は$_H2$にあり、$_H2$にあり$_{G2}$にない有向辺は$_H1$にある

。

図3、4では、G_1、$_{G2}$は、**O**={A,B}、**U**={A,B,C,D}*に対して*$_{H1}$、$_H2$の**平行**埋め込みを持つ。そこで、2つのs.s.i.構造がさらなる変数の測定によって区別できるかどうかという問題は、次のようになります：その構造はs.s.i.ではない平行埋め込みを持つかどうか? もしそのような埋め込みがなければ、構造$_G1$と$_G2$は**厳密には統計的に区別できない（r.s.i.）と言うことになる。**

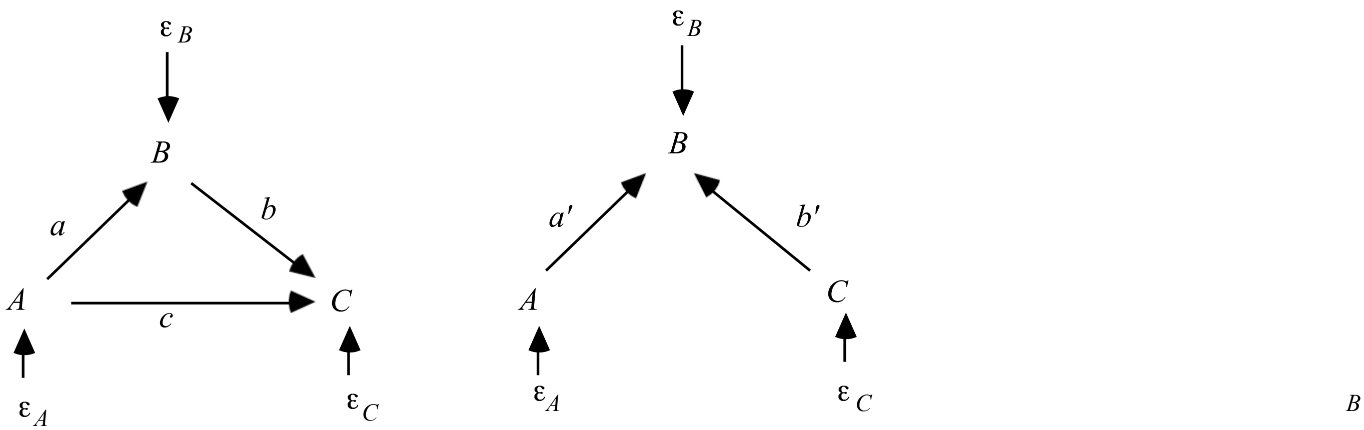
定理4.6：同じ頂点集合を持つ2つの異なるs.s.i.有向非循環グラフが剛体統計的に区別されないことはない。

つまり、正しい因果構造を持つ追加の変数が存在し、測定できるのであれば、測定された変数の因果的に十分な集まりの中の因果構造を原理的に特定することができる。定理4.6の証明は、忠実に区別できない構造に対する並行結果も示している。我々は、定理4.6の類似の結果が、正を仮定した弱い統計的識別可能性についても成り立つと推測する。

4.5 リニア の場合

パラメータ値によって、グラフでは線形にならない条件付き独立性やゼロ部分相関を強制することができる。図2のグラフ（図5では誤差変数が明示的に含まれている）はその可能性を示している。グラフの頂点をそれぞれ「誤差」変数に付随するものとして扱い、グラフと誤差変数で一次方程式の集合を決定させる。(私たち

は、誤差項を含むすべての外生変数の組がゼロ共分散であると仮定する)。結果は、外因性変数の共同分布の指定までは、構造方程式モデルである。各有向辺には線形係数が付きます。相関行列、したがってすべての偏相関は、線形係数と外生変数の分散によって決定される。



(i)　　　　　　　　**に**

図5

左の構造で$ab=-c$とすると、*A*,*C*は右のモデルのように無相関となる。このような現象は、グラフの構造ではなく、線形係数の値によって生じる部分相関の消失であり、相関関係から因果関係を推論しようとする試みを惑わせることになるであろう。この現象はいつ起こるのだろうか？ この疑問には、前章で線形忠実性が破綻する条件を検討したときに、すでに答えている。線形的な場合、パラメータ値--有向非周期グラフGを持つ構造の線形係数と外生分散の値--*は*実空間を*形成し*、この空間において、*G*が線形的に暗示しない消失する部分相関を作り出す点の集合はルベーグ測度0である。

定理3.2：有向非周期グラフGと*n個の*線形係数$a_1,\ldots,a_n$、外生変数$v_1,\ldots,v_k$の*k個*

の正分散を持つ線形モデルをMとする。$M(<u_1,\ldots,u_n,u_{n+1},\ldots,u_{n+k}>)$を、値を指定

することに矛盾しない分布とする。

$<u_1,\ldots,u_n, u_{n+1},\ldots,u_{n+k}>$ を $a_1,\ldots,a_n$ と $v_1,\ldots v_k$ について考える。を確率の集合とする。

Mのパラメータの値の空間$\mathfrak{R}^{n+k}$上の測度Pは、$\mathfrak{R}^{n+k}$のすべての部分集合**Vが**

ルベーグ測度0であるとき、$P(\mathbf{V})=0$である。Qを係数と分散値のベクトルの

集合として、**Qの**すべてのqについて、すべての確率分布

in with $M(q)$は、Gが線形に含意しない消失する偏相関を持つ。そして、すべてのP in Π $P(\mathbf{Q})=0$である。

この種の尺度理論的な議論は興味深いものであるが、完全な説得力を持つとは言えないかもしれない。というのも、一般的な線形モデルでは、因果関係のないものは値ゼロの線形係数によって示され、したがって尺度ゼロの集合を形成するので、推論のパリティによってすべてが他のすべてに因果的につながっていると主張することができるからである。Nancy Cartwright (1989)は、線形構造における独立関係は、因果構造だけでなく、線形係数や分散の特殊な値によって生じる可能性があるため、そのような関係から因果構造を推論するのは非合法であると反論している。つまり、真の因果構造をw.s.i.の代替案と区別できないような推論手順を否定しているのである。2つの構造がw.s.i.であってf.i.やs.s.i.でないことが不可能なのはどんな場合か、ある分布が2つのw.s.i.の因果構造に対してマルコフ条件や最小性条件を満たしているがs.s.i.やf.i.の因果構造ではない、という特別なマークや指標はあるか、といった問題です。これらの質問に対する答えは、基本的には、前節の定理の線形ケースへの応用に過ぎない。

ここでは、Cartwrightと同様に、変数の時間的順序が既知であることを仮定する。Pearl and Verma (Pearl 1988)は、正分布Pと与えられた変数の順序に対して、Pが最小条件とマルコフ条件を満たす有向無サイクルグラフが1つだけ存在することを証明しました。これは、与えられた相関行列を持つ正分布と、因果的に十分な変数の集合の与えられた順序に対して、分布を線形に表し、順序と一致する唯一の有向無サイクルグラフが存在することを意味する。

少なくともいくつかのケースでは、分布の陽性度をテストすることができます。(例

えば、2変量正規分布では、相関が1でなければ密度関数はどこでも非ゼロとなる）。これらの場合、与えられた変数の順序に対して、*P*がマルコフ条件と最小化条件を満たすユニークな有向無周期グラフが存在するか、そのような有向無周期グラフが複数存在することが検出可能であることがわかります。しかし、ある変数の順序に対して、*P*が*マルコフ*条件と最小化条件を満たすユニークな有向無サイクルグラフが存在するとしても、そのグラフを見つけるためのアルゴリズムは、テストする必要のある条件付き独立関係の数と順序のために、多数の変数に対して実行可能ではない。

ある分布が、それを生成した因果グラフに忠実であると誤って仮定したとする。これは、非公式に、忠実であると仮定しても真実でない場合、条件付き独立関係や特殊な条件による部分相関が消失することを意味します。

パラメータ値は、真の因果関係が*省略*された誤った因果推論をもたらすだけであり、それ以外の種類の誤りは生じない。このような事情が相関関係で明らかになるのはどのような場合か、検討することにする。

トレックは、両方のパスの元となる単一の共通頂点を持つ有向無サイクルパスの無順序ペアであることを思い出してください（ペアのパスの1つは、第2章で定義された空のパスであってもよい）。各変数の平均がゼロで、非エラー変数が単位分散を持つ標準化モデルでは、2つの変数の相関は、X, Yを結ぶすべてのトレックのうち、トレックのエッジに関連する線形係数の積の和で与えられる（この量を**トレック和と**呼ぶ）。例えば、図5の有向無サイクルグラフ(i)では、AとCの間のトレックサムは$ab + c$である。本節の例では、標準的なシステムを使用する。相関系は、以下の式により、あらゆる順序のすべての部分相関を決定する。

$$\rho_{XY.\mathbf{Z}\cup\{R\}} = \frac{\rho_{xy.\mathbf{z}} - \rho_{xr.\mathbf{z}} \times \rho_{yr.\mathbf{z}}}{\sqrt{1 - \rho_{xr.\mathbf{z}}^2} \times \sqrt{1 - \rho_{yr.\mathbf{z}}^2}}$$

再帰関係は、集合**U**上の2つの変数の間に、**Uの**メンバー上の部分相関をどのような順序でとっても同じ部分相関を与えるので、部分相関の消失は、標準化されたシステムの係数の方程式系に対応する。

ここで、正常で標準化されたシステムGの線形パラメータの特別な値が、ある偽の因果構造、例えばH*によってのみ*線形的に暗示される消失する部分相関を生み出すとします。そして、そのパラメータ値は、G*によって*線形的に暗示されない余分の消失する部分相関を生み出すはずです。どの部分相関も、変数のペアを結ぶトレックサムの関数であり、この場合のトレックサムには、以下の線形パラメータだけが関与します。

したがって、*G*が線形に含意しない追加の消失部分相関は、偶然の消失部分相関を生成するために満たされなければならない*G*のパラメータにおける（非線形）方程式系を決定することになる。(例えば、図5の有向無サイクルグラフ(i)では、*A*-*C*間の相関が0になるのは、*ab*＝-cという一つの方程式が満たされる場合のみである）。さて、あるGとある*H*（*G*の部分グラフ）に対して、これらの連立方程式は連立解を持たない場合がある。他の*G*と*H*の部分グラフでは、連立方程式は解を持つが、1つまたは複数のパラメータに対して有限の代替値しか許さず、ある誤差分散を消滅させる必要がある解しかない場合がある。このような解は、特別な相関制約によって「それ自身を与える」必要があります。

は、それ自体が消失する偏相関関係でない。を以下のように選択することを考える。GとH、それぞれのペアにおいてGは左手側、Hは右手側にある：

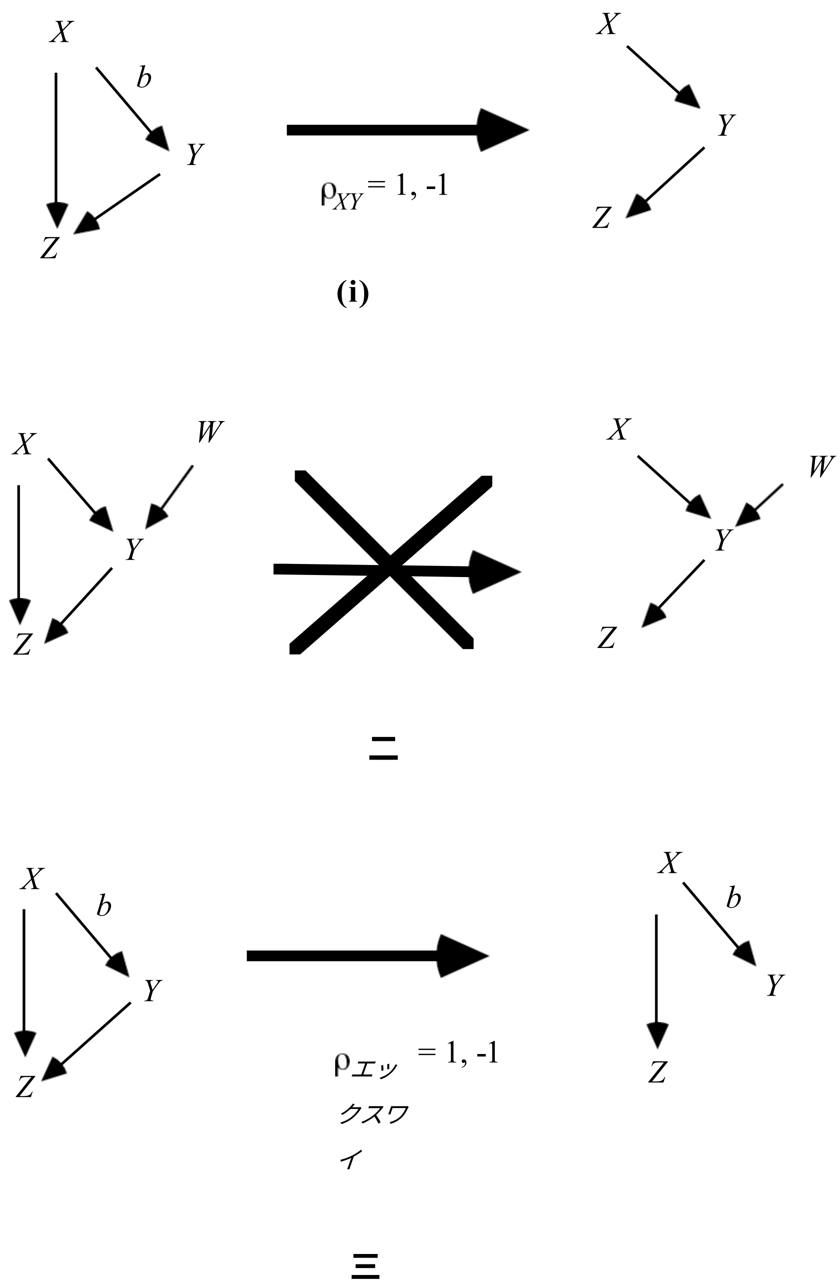


図6

(i)と(iii)では、左側のグラフの係数と分散を選んで、あたかもエッジが発生しないように見せることができるが、bと書かれた係数を1か-1のどちらかにすることによってのみ、エッジが発生しないように見せることができる。変数は標準化されている

ので、*Y*の誤差項が分散ゼロ、平均ゼロ、つまり消滅することが必要である。したがって、真のグラフが左側のグラフであり、パラメータ値が右側のグラフが線形に暗示する消失する部分相関を生み出すためには、変数Yは変数*X*の線形関数であり、変数*Xだけ*でなければならない。でない辺がある場合にも同じ結果が得られる。

最初の例と最後の例で排除されたものは、任意の長さの有向パスで置き換えられる。これらの場合、真のグラフが線形に暗示しない消失する部分相関を作り出す特別なパラメータ値が、相関によって明らかになることは明らかである。(ii)では、変数XとZの間のエッジは、真のグラフのパラメータ値のどのような選択によっても排除されているように見せることはできない。

我々は、事前の時間順序がなくても、*Gにおいて*3つの辺が三角形を形成しない限り、*Gの*パラメータ値がグラフHによって線形に暗示される消失部分相関のコレクションを正確に決定する場合--HがGのサブグラフであるかどうかにかかわらず--、消失部分相関が内包しない相関に対する特別な制約が存在すると推測する。

4.6 変数の再定義

これまでの区別不能の結果は、同じ頂点の集合を持つ代替グラフに関するものであった。頂点は乱数変数として解釈され、その値は何らかの測定システムの対象となる。新しい確率変数は、与えられた集合から、例えば線形結合やブール結合をとることによって常に定義することができる。任意の定義装置と、グラフと分布を結びつける公理があれば、固定された変数集合について考えたのと同様に、区別できないクラスについての疑問が生じる。変数集合**Vに対する**分布*Pはグ*ラフ*Gに*対応し、変数集合**V'**に対する分布*P'*は別のグラフ*G'*に対応する（Pと**Vから、**新しい変数を定義し、古い変数を無視し、マージナル化することで*P'*と**V'を得る**）。Gと*G'*の違いは重要でない場合もあり、それぞれのグラフがそれぞれの変数集合の間の因果関係を正しく記述していると言いたいだけかもしれない。しかし、元の変数が時間順に並

んでいて、変数の再定義によって、後の事象が前の事象を引き起こすような分布になる場合は、そうとは言えません。次のようなグラフの組を考えてみよう。

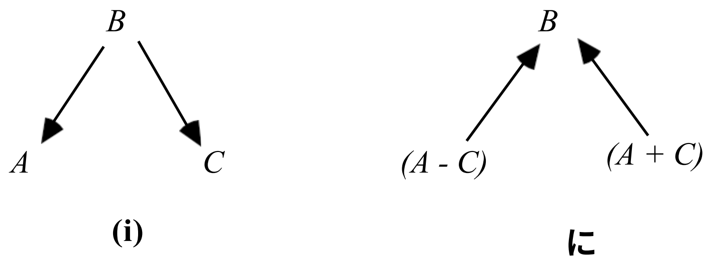


図7

有向無サイクルグラフ(i)において、AとCはBの影響であり、BがAとCより先に発生したとする。定義と周辺化の手続きにより、グラフ(i)に忠実な分布は、グラフ(ii)に忠実な分布に変換することができる。まず、AとCを標準化し、単位分散を持つ変数A'とC'を形成します。次に、変数（A'-C'）と（A'+C'）を考える。これらの共分散は、A'^2 - C'^2の期待値に等しく、ゼロである。単純な代数学は，Bを与えられた(A' - C')と (A'+ C') の偏相関が消えないことを示す．したがって、元の分布のマージナルは(ii)に線形に忠実であり、元の分布が正規分布であれば(ii)に忠実であることがわかる。

先ほど例示した変換は不安定であることに注意してください；A'とC'の分散が少しでも不等であれば、あるいは$xy + wz + \rho_{A'C'}(zy + xw) \neq 0$のような$x$、$y$、$z$、$w$の任意の値に対して変換が（$xA'$+ zC' ）と（yA'+ wC' ）を与えるならば変換後の分布上のマージンは（ii）ではなく、3変数上の完全グラフのすべての非周期方向、時間順序と矛盾しない仮説に対して誠実であろうと考えます。

別の見方をすれば、「偶然」偏相関が消失するような変数の変換は、「忠実な条件」の違反に過ぎない。図 8 の線形モデルを考えてみよう。

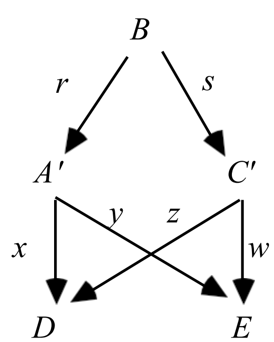


図8

$A' = rB + \varepsilon_{A'}$, $C' = sB + \varepsilon_{C'}$, $D = xA' + zC' + \varepsilon_D$, $E = yA' + wC' + \varepsilon_E$ とし、変数を標準化すれば、ρ_{DE}は$xy + zw + rysz + rxsw = xy + zw + rs(yz + xw)$ となり、$rs = \rho_{A'C'}$ なので前項式となる。ρ_{DE}＝0であれば、忠実条件違反となる。したがって、「偶然」ゼロ相関を生じるAおよびCの線形変換が得られる条件は、AおよびC間のトレが忠実条件違反で正確に相殺される条件と同一である。

$D＝A'＋C'$（すなわち$x＝z＝1$）、$E＝A'$-C'（すなわち$y＝$-w）のとき、図7の例が得られます。= 1）ここで誤差項の分散と平均はゼロに設定されている。この例では、忠実性に反するパラメータ値の集合がルベーグ測度0であるため、「偶然」相関が0になるAおよびCの線形変換の集合も0になる。

4.7 背景 備考

線形統計モデルの測定変数の値による過小決定については、「同定問題」として広く議論されており、特に計量経済学（Fisher, 1966）では、自由パラメータの推定を中心に議論が行われてきた。線形モデルで広く使われている「道具変数」という装置は、厳密な区別可能性に関する定理4.6の精神を受け継いでいるが、道具変数は、循環グラフや潜在変数を持つシステムでパラメータを識別するために使われている。結果変数が原因として扱われるように、純粋な線形回帰モデルを「書き換える」可能性は、以前から知られていたようで、Judea Pearlによって注目されたこの観測の原典は分かっていない。

本章で検討したような意味での統計的区別不能の説明は、Basmann (1965), Stetzl (1986), Lee (1987)によって提案されている。Basmannは、我々の用語で、環状グラフを持つ（すなわち「非再帰性」）すべての連立方程式モデルに対して、非環状グラフを持つ統計的に区別できないモデルが存在することを論じた。その結果が弱い区別不能の定理である（12章参照）。StetzlとLeeは、線形係数と分散の自由パラメータを持つ線形構造方程式モデルのみに焦点を当て、パラメータの最尤推定値、ひいては共分散行列の最尤推定値の観点から同等性を定義しています。Leeの論文では興味

深い試みがなされているが、一般的なグラフ理論的な特徴付けは行われていない。

パターンの概念と定理4.2は、Verma and Pearl (1990b)による。第6章で、因果的に不十分なグラフの区別不能関係に関するいくつかの結果を述べている。Suppes and Zanotti (1981)によるよく知られた結果は、離散変数の集合**X**上のすべての共同分布Pは、Tが**Xの**すべての変数の共通原因であり、他の有向辺がないグラフGのマルコフ条件を満たす**X**□{T}上のいくつかの共同分布*P*の*マージナルであると主張します。この結果は、弱い区別不能の定理と見なすことができる。

因果的に不十分な構造を認める場合特殊な場合を除き、P*は*G*に忠実であることはできない。

第5章

のための発見アルゴ
リズム
因果関係のある十分な構造

5.1 Discovery 問題点

発見問題とは、データ源となる構造、しかし調査者が調査前に知っている限り、そのどれもがデータ源となる構造である可能性がある、代替構造の集合からなるものである。どの構造であっても、実際の構造について解明しなければならないことがある。それは、ある仮説に決着をつけたいということかもしれません。その仮説は、ある可能性のある構造では真で、他の構造では偽であるかもしれませんし、ある種の現象についての完全な理論を知りたいということかもしれません。本書、そして社会科学や疫学の多くでは、発見問題における代替構造は、典型的には、頂点に共同確率分布を持つ有向非周期グラフのペアである。私たちは通常、因果関係を表すグラフの構造について何かを知りたいし、ある集団におけるグラフ内の変数の値の分布についても知りたいかもしれない。

例えば、ある変数についてはデータがあるが、他の変数についてはデータがない、データには実際の確率や条件付き独立関係が含まれている、あるいは、より現実的

には、単にランダムなサンプルの変数の値が含まれている、などである。理論的な議論では、通常、測定された変数間の真の条件付き独立関係がデータに含まれている発見問題を考えますが、例や応用では、常に統計的サンプルから推論することになります。

ある方法が極限で発見問題を解決するのは、サンプルサイズが無限に増えるにつれて、その方法が質問に対する真の答えや真の理論に収束していく場合である（*何であれ*）。

(事前知識と一致する)真実があるかもしれない。原因を推論するための手順は、代替可能性のいくつかについて、答えが出ないか間違った答えを与える場合、提起された問題を解決することはできない。しかし、代替構造のいくつかを除外したときに生じる、別の簡単な問題を解決することはできる。どの因果関係の発見問題が極限で解決可能か、またどのような方法で解決できるかは、数学的に確定した問題である。形而上学的な論争は、問題を解くことではなく、問題を動機づけることにあるのです。本書の残りの部分は、これらの形式的な問題の研究と、特定の答えの実用的な応用のための入門書である。

5.2 統計学における検索ストラテジー

統計学の文献には、データを使って代替分布の限定的なパラメトリゼーションを探索する手順がたくさんあります。統計的仮説の表現が、政策や実務の指針として、ある変数を操作した場合に何が起こるかを予測したり、ある変数を過去に操作した場合に何が起こったかを再現したりするために用いられる場合、統計的仮説は通常、因果関係の仮説でもある。その場合、最初の問題は、探索手順が因果構造を見つけるのに適しているかどうかということである。

統計学の文献で提案されている検索戦略の多くは、任意のモデル、あるいは制約を伴わない完全（あるいはほぼ完全）構造、あるいはすべての変数が独立である完全（あるいはほぼ完全）制約構造から始める、最良のみのビーム検索である。統計学者は、後者の手順を「前方探索」、前者の手順を「後方探索」と呼ぶことがあります。どの順序に従うかによって、これらの手順は、パラメトリゼーションのどの固

定パラメータを解放すると適合度が最も向上するか、あるいはどの自由パラメータを固定すべきかを決定するために、ある種の適合尺度を繰り返し適用します。そして、修正された構造を再推定し、停止基準を満たすかどうかを判断する。この種の「前方」手順はArthur Dempster (1972)によって共分散構造に対して提案され、「後方」手順は彼の弟子のNanny Wermuth (1976)によって、分布が「乗法的」である対数線形系と線形系の両方、つまり我々の言葉でいえばある有向無サイクルグラフのマルコフ条件を満たすものに対して提案されている。Byron(1972)とSorbom(1975)は、変異型線形システムに対して適合度統計を用いた前方探索アルゴリズムを提案し、そのバージョンがLISREL (Joreskog and Sorbom 1984) とEQS (Bentler 1985) の推定パッケージで自動化されてきた。その

後者のプログラムには後方探索手順も含まれている。対数線形パラメトリゼーションのための一般的な戦略のバージョンは、Bishop, Fienberg and Holland (1975), Fienberg (1977) by Aitkin (1979), Christensen (1990) などによって説明されている。同じ表現と検索戦略は、システム科学の文献でKlirとParviz (1976)などによって、「再構築可能性」分析というタイトルで使用されている。ロジスティック回帰におけるステップワイズ回帰の手続きも、同じ戦略のバージョンとみなすことができる。同じ戦略が無向グラフと有向グラフの表現に適用されている。それらはWhittaker (1990)によってさまざまな例で説明されている。

分布を推定するだけでなく、因果構造を特定したり、変数の一部を操作した結果を予測したりすることが目的である場合、これらの各ケースにおいて、一般的な統計的検索戦略は不満足である。(i)多くの因果関係のある仮説を除外し、因果関係のない多くの仮説を含む仮説空間を探索することが多い。(ii)分布の仕様により、統計的または計算上の理由から不必要に探索を制限する数値手続きを使用することが一般的。(iii)探索に単一の仮説を出力することを求める制限により、証拠があれば区別できないかもしれない代替仮説を出力しないことが伴うため、探索は非効率で信頼できない。これらの点については、それぞれ詳しく検討する。

5.2.1 間違った仮説 宇宙

正しい原因仮説を探索する場合、選択肢の空間は、可能な限り、背景知識によって除外されていないすべての原因仮説と、原因的解釈を持たない仮説を含む必要があります。1963年にBirchによって導入された対数線形フォーマリズムは、正しい因果仮説を見つけるという目標にうまく適応していない探索空間の重要な例を示してい

る。離散データの場合、より適切な探索空間は、対数線形仮説の結合のサブクラスであることが判明している。

対数線形フォーマリズムは、あらゆる次元の分割表を分析するための一般的な枠組みを提供します。離散的な場合、順序の有無にかかわらず、有限個の値をとる変数に関係する。例えば、4つの変数を持つシステムでは、i を1番目の変数の値、jを2番目の変数、kを3番目の変数、lを4番目の変数とすることにする。そして、特定のサンプルや集団において、x_{ijkl}は、第1の変数の値i、第2の変数の値j、第3の変数の値k、第4の変数の値lを持つユニットの数を表す。我々は、4つの（または他の数の）変数の値の特定のベクトルを "セル "と呼ぶことにする。形式論では、セル上の共同分布は、以下の対数の式で与えられる。

各セルの期待値で、いくつかのパラメータの和として表現される。例えば、m_{ijk}がセルi，j，kの期待値を表すBirchの表記では

$$\ln(m_{ijk}) = u + u_{1i} + u_{2j} + u_{3k} + u_{12ij} + u_{13ik} + u_{23jk} + u_{123ijk}$$

3つの二項変数の系では、7つのu項だけが独立であることができる。Birchのパラメトリクスの威力は、少なくとも2つの特徴にある。第一に、統計学で古くから研究されてきた多次元分割表の関連性を、あるパラメータがゼロであるという仮説として表現できることである。例えば、Bartlettは、3つの二値変数の間に「3因子相互作用」がないという仮説を、セルの確率の間に次のような関係で表現している：

$$P_{111}P_{122}P_{212}P_{221} = P_{222}P_{211}P_{121}P_{112}$$

Birchは、この条件を有限個のカテゴリーを持つ変数に一般化すると、様々なu項がゼロである場合にのみ得られることを示した。第二に、u項のいくつかをゼロにすることによって得られる各仮説について、様々なサンプリング手順に対して最尤推定値を得るための反復法が存在する。

Birchの結果は、いくつかの研究者によって拡張された。対数線形パラメトリゼーションにおける仮説は、特定のu項が消失する仕様として扱われるようになった。このような仕様のある形式については、期待される細胞数の直接の最尤推定値があり、他の仕様については最尤推定値に収束する反復アルゴリズムが開発されている。対数線形パラメトリゼーションの特定のクラスに焦点を当てるための様々な形式的動機付けが開発されている。例えば、Kullback (1959)は情報量に基づく距離尺度を用いて、最大エントロピー原理という少し異なる観点から同じように得られる対数線形関係のクラスを導きました。Fienberg(1977)などは、「階層モデル」-あるインデックスの集合を持つu項をゼロにすると、そのインデックスが最初の集合を含む他のすべてのu項もゼロになる-に注意を向けるよう促した。この制限の動機は、これらのパラメトリゼーションが分散分析と形式的に類似しており、例えばu_1項を第1変数の作用による大平均からの変動と考えることができるためである。

因果構造を対数線形形式論で表現することの難しさを知るために、前章の最も基本

的な因果関係、すなわち、$A \rightarrow B$□Cの任意のコライダーを考えてみましょう。

であり、Aと*C*は隣接していない。このような構造は、条件付き独立性に関する2つの事実に（忠実であると仮定して）対応する：第一に、AとCは*B*を含まないある変数集合を条件として独立であり、第二に、AとCは*B*を含むが*A*または*C*を含まないすべての集合を条件として従属である。この種の最も単純な場合、*A*、*B*、Cが唯一の変数である場合、Aと*B*は独立しているが*C*を条件として依存している。これらの関係を得る仮説は対数線形形式では*u*項が消えることによって表すことができない。バーチ自身も、3変数のシステムにおいて、2つの変数が独立であるという仮説は、3変数の一般的な対数線形展開におけるパラメータのどの部分集合の消失によっても表すことができないことを観察した。もちろん、限界独立仮説と*矛盾*しない対数線形仮説は存在するが、それを内包するものではない。

もう一つの不適切な検索空間は、LISRELプログラムによって提供されている。LISREL形式は、少なくともその著者であるJoreskogとSorbomが意図したように、測定されていない共通原因がない場合には、測定された変数間の因果関係に対応する構造を検索することができるが、測定されていない共通原因を含む構造を検索する場合には、測定した変数間の因果関係は禁止される。ユーザーはこの制限を回避する方法を見つけたが(Glymour, et al., 1987; Bollen, 1989)、むしろプログラムの作者は不満に思っている(Joreskog and Sorbom, 1990).LISRELがこのような特異性を持つのは、因子分析を祖先としているためであり、人為的に縮小された探索空間のもう一つの例である。Thurstone (1935)は、自分の「因子」を実際の原因としてとらえるのではなく、測定された相関関係を数学的に単純化したものに過ぎないことを注意深く、繰り返し強調した。もちろん、因子はすぐに仮説的な原因として扱われた。しかし、このように適用すると、サーストンの方法は、測定された変数自体の間の因果関係を*先験的に*排除し、測定された変数が測定されていない変数の原因であるとい

う可能性を排除し、潜在的変数の間の因果構造--相関関係のみ--を決定することができない。

5.2.2 Computational and Statistical Limitations（計算と統計の限界

検索によっては、可能な仮説空間のごく一部しか調べないものもあるが、これは各仮説を検証するために計算量の多い反復アルゴリズムが必要だからである。例えば、LISRELの自動モデル変更手順は、新しい仮説を検証するたびにモデル全体を再推定している。その結果、探索が遅くなるため、真実が隠されている可能性のある仮説空間の大部分を調べることができなくなる。

もう一つの共通した問題は、多くの検索が、信頼性のあるテストができない条件付き独立関係を決定することを要求することである。対数線形探索の多くは、真の構造がどうであれ、変数の総数から2を引いた大きさの変数集合を条件とする確率の推定を暗黙のうちに要求する。高次の条件付き確率の推定や高次の条件付き独立性の検定は、（特にいくつかの離散的な値をとる変数の場合）信頼できない傾向があり、妥当なサンプルサイズでは、変数の値の配列に対応するほとんどのセルが空かほとんど空になってしまうからである。この欠点は、対数線形形式にはない。Fung and Crawford (1990)が提案したグラフモデルの集合（無向独立グラフで表現できる階層的対数線形モデルのサブセット）を検索するための最近のアルゴリズムは、高次の条件付き独立性をテストする必要性を低減する。回帰係数が消えるという仮説の検定では、他の回帰係数の数だけサンプルサイズが減少するか、自由度が変化するため、妥当な代替案に対して検定がほとんど力を持たない可能性があるからである。

関連するが、同様に基本的な問題は、離散データのモデルを探索する際に、各段階（または任意の段階）でモデル推定を必要とする何らかの適合性の尺度を用いる場合、変数の数が増えるにつれて推定しなければならないセルの数が指数関数的に増加することである。最も単純なケースとして、変数が2値である場合、期待値を計算しなければならないセルの数は2^nである。$n = 50$の場合、セル数は天文学的な数字になる。

このような困難は、信頼性の高い検索手順にはつきものだと思うかもしれない。この章と次の章で見るように、そうではありません。

5.2.3 単一の 仮説を生成する

もし、ある種の証拠が、いくつかの代替仮説のうち、1つが正しいのか、それとも他の仮説が正しいのかを確実に区別することができないのであれば、適切な検索手順は、その事実を反映して、すべての仮説を出力する必要があります。このような状況で単一の仮説のみを出力することは、ユーザーを惑わせ、意思決定に不可欠な情報を否定することになる。

この種の欠陥の例として、LISRELとEQSのプログラムが挙げられる。これらのプログラムは、背景知識から構築された構造から始まり、ベストオンリービームサーチによって線形構造の中から因果モデルを探索する。各段階において、モデルのデータへの適合度を最も高めると判断される固定パラメータを解放する。異なる複数の固定パラメータを解放しても、適合度が全く同じになる可能性があるため、このプログラムでは任意の同点解消方法を採用している。探索の出力は単一の線形モデルであり、統計的に区別できない代替モデルは無視される。

後の章では、線形モデル用のLISRELとEQSプログラムに実装されている統計的探索手順の信頼性に関する大規模なシミュレーション研究を説明することにする。計算上の問題や、探索の様々な段階で区別できないモデルの中から任意に選択できることから、プログラムが最初に構造の大部分（正しい線形係数や分散も含む）を正しく与えた場合でも、データが生成された構造における依存関係を発見する上でこの手続きはほとんど意味をなさないことがわかった。この研究は測定不能な変数を持つシステムを対象としているが、因果的に十分なシステムを対象とした研究でも同様の結果が得られると予想される。

5.2.4 その他 アプローチ

統計的探索戦略が、生成とテスト・ベストのみの手順に限定されているという一般化には、いくつかの例外がある。Edwards and Havranek (1987)は、あるモデルがテストに合格すれば、より一般的なモデルもそうなり、あるモデルがテストに失敗すれば、より制限されたモデルもそうなるという仮定で、モデルを順番にテストする手順を説明しています。彼らの提案は、（あるパラメトリゼーションにおいて）可能

なすべての仮説が分類されるまで、棄却された仮説の境界集合と受け入れられた仮説の境界集合を追跡することである。EdwardsとHavranekは知らなかったようだが、同じ考え方は人工知能の文献で「バージョン空間」(Mitchell, 1977)という名で先に長く展開されていたようである。彼らが考えているアプリケーションでは、複雑さや信頼性の解析はできない。

5.2.5 ベイジアン メソッド

統計学における探索問題をベイズの観点から論じたものとしては、Leamer(1978)が最も有名である。Leamerの本には、新奇な仮説に出会ったときにベイズ派はどうすべきかという考察など、興味深い点がいくつもあるが、確実な探索のための方法は含まれていない。因果関係における回帰手法の利用を考える

例えば、Leamerは、意見によって支持された関連する回帰因子のセットを別々に分析し、これらの回帰因子のセットのそれぞれについて、パラメータの分布のベイズ的更新を別々に行うことを、次節で推奨している。どの変数が関心のある結果に実際に影響を与えるかを決定する問題は、事実上無視されている。

Cooper and Herskovits (1991, 1992)によって、より有望な探索のためのベイズ的アプローチが開発された。現在のところ、彼らの手順は離散変数に限定され、後の変数が前の変数を引き起こすことがないような全順序を必要とする。この順序に適合する各有向グラフには事前確率が割り当てられる。変数の共同分布は、グラフの各頂点にその親を条件とする分布を割り当て、これらの条件付き確率は各グラフの分布をパラメトリックにする。ディルシェレットプリオールを用いると、各グラフのパラメータに密度関数が課される。データはベイズの法則で密度関数を更新するために使われる。グラフの確率は、そのグラフに適合する分布に対する密度関数の積分となる。エッジの確率は、それを含むすべてのグラフの確率の和となる。CooperとHerskovitsは、貪欲なアルゴリズムを用いて、段階的に出力グラフを構築する。このアルゴリズムは、グラフの各頂点Xについて、Xの親でない個々の前任者をXの親セットに加える効果を考慮し、Xの親セットに加えることでXとその親からなる局所構造の事後確率を最も高める頂点を選択する。このようにして、局所構造の事後確率を増加させるXの親セットに追加できる頂点が1つもなくなるまで、Xに親を追加していく。このプログラムは、真のグラフが疎であれば、非常に大きな変数セットでも非常によく動作し、事前順序を持つ離散データでは、隣接関係を顕著な精度で決定するようである。密なグラフに対する精度は、現時点では不明である。

CooperとHerskovitsが開発したベイズ法は、適切な事前信念度を検索に使用できるこ

と、指定した事前分布とデータに一致する事後分布の比率でモデルが出力されること、適切な仮定[1]下で手法が正しいグラフに収束すること、などの利点を持つ。また、この手法は任意のグラフの組の事後確率の比を計算できるため、複数のグラフをグラフの確率で重み付けして推論することが可能である（ただし、一般的には、グラフの数が多いため、最も確率の高いグラフのみを考慮するヒューリスティックを用いなければならない）。

[1] 特に、その方法を理想化して貪欲なアルゴリズムをあきらめたとき。貪欲なアルゴリズムのため、隣接しない変数のペア、例えばXと*Y*の間に2つ以上のトレッキングがあり、それらの変数の間に密接な統計的関連が生じる場合、特定の検索手順は漸近的に信頼性が低くなると予想されます。これは、ALARMネットワークに誤って導入された1つのエッジの場合の状況である。実際には、このような構造は珍しく、誤差は許容範囲内である可能性があります。

の可能性がある)。この方法はディルシェレットプリオで機能する。なぜなら、関連する積分は解析的に利用可能であり、したがって、数値解析なしに事後密度を迅速に評価することができるからである。グラフの組合せ論に鑑みると、探索アーキテクチャの他の応用も同じ特徴を持たなければならない。一つの問題は、この方法を連続変数に拡張することであり、これは、グラフを記述するパラメータに対して迅速に更新できる共役プリオールのファミリーを見つけることに依存する。もう一つの、より基本的な問題は、変数の事前順序の要件を、計算可能性を維持したまま緩和できるかどうかに関わる。変数の順序を固定にすることで、組合せは非常に少なくなるが、多くの応用例では、そのような順序は不確実である可能性がある。第1章で紹介したALARMネットワークのデータを解析するのに必要な時間は約15分（Macintosh II）と高速であるため、Cooperたちは多くの順序を使用し、得られたグラフの事後確率を比較する方法を検討している。

5.3 Wermuth-Lauritzen アルゴリズム

1983年、WermuthとLauritzenは、再帰*的ダイアグラム*と呼ばれるものを定義しました。再帰ダイアグラムとは、有向無サイクルグラフGと、*その*グラフの頂点の全順序（$V_1 \to V_2$は、$V_1 < V_2$である場合にのみ起こる）である。さらに、$V_i < V_k$のときのみViがV_kの親となり、ViとV_kは順序においてV_kより前の他のすべての変数の集合に依存するような頂点上の確率分布P*が存在する。*Whittaker (1990)に倣って、このようなシステムを*有向独立グラフ*と呼ぶ。

この定義は、条件付き独立関係と変数の時間的順序から因果グラフを構築するアル

ゴリズムとみなすことができる。実際、一部の著者によってこのように使用されている（Whittaker, 1990）。変数の順序と条件付き独立関係のリストが与えられたら、時間順に変数を進め、各変数Vkから$V_i < V_k$となる各変数V_iに対して、定義の依存性テストを適用し、テストに合格したらV_i -> V_kを加える。この手順により、離散変数について、変数値のすべての組み合わせが正の確率を持つ忠実な分布の順序と独立関係から有向グラフが正しく復元される。いわば、因果的に十分な忠実系に対する発見問題が解決されたことになる。しかし、実際には、Wermuth-Lauritzenアルゴリズムは、非常に小さな変数集合の場合を除き、実現不可能である。したがって、残された課題は以下の通りである：

(i) 変数の順序があらかじめわかっているという条件を外す方法;

(ii) Wermuth-Lauritzen手続きの計算効率と統計的要件を改善する方法;

(iii) 因果的に十分な変数系という暗黙の制限をどう取り除くか。

本章では、これらの問題のうち、最初の2つを取り上げます。測定されていない共通原因、すなわち「潜在変数」が作用している可能性がある場合の因果推論の問題については、第6章で取り上げる予定である。

5.4 新しい アルゴリズム

因果構造を発見するためのいくつかのアルゴリズム（因果的充足性を仮定）について説明する。それらは変数の事前順序付けの必要性を排除し、2つを除くすべてがWermuth-Lauritzenアルゴリズムと比較して計算効率を改善し、統計的判断の困難さを軽減する。その中には劇的な改善もあれば、そうでないものもある。説明した各検索方法は、離散データでグラフィカルな対数線形モデルの検索にも使用できる。(変数の各トリプルについて、*X* -> *Y* <- *Z*が有向グラフに現れ、*X*が*Z*に隣接していない場合、Xと*Z*の間に無向のエッジを追加し、グラフからすべての矢じりを削除する。その結果、無向性の独立グラフとなる)。

以下の仮定のもと、本節で紹介するすべてのアルゴリズムは、母集団分布に忠実なグラフの特徴を証明的に回復することができます:

(i) 観測された変数の集合は、因果的に十分である。

(ii) 母集団のすべての単位が、変数間の因果関係を同じにする。

(iii) 観測された変数の分布は、因果構造の非周期有向グラフに忠実（離散の場合）、またはそのようなグラフに線形に忠実（線形の場合）です。

(iv) アルゴリズムが必要とする統計的な判断は、母集団に対して正しいものです。

第4の要件は不必要に強い。なぜなら、統計的な判断に誤りがあったとしても、多くの場合、アルゴリズムは成功するからである。しかし、これは医学、行動学、社会科学の文献に見られる因果関係の解釈を伴う特定の統計モデルのほとんどを保証するために必要な前提条件よりも強いものではありません。次の章では、これらの仮定を弱めた場合の結果について検討することにする。

実際には、アルゴリズムは共分散行列かセルカウントを入力とする。アルゴリズムがd-分離の事実を必要とする場合、離散の場合は条件付き独立性の検定を行い、線形連続の場合は部分相関の消失の検定を行う。(PがグラフGに忠実な離散分布である場合、AおよびBがCを与えられたときに条件付き独立である場合にのみ、変数の集合**Cを**与えられたときにd-分離され、PがグラフGに忠実な線形分布である場合、AおよびBは$\rho_{AB.\mathbf{C}} = 0$の場合にのみ**Cを与えられて**d-分離されるということを覚えておこう).このアルゴリズムは、与えられたd-分離可能性関係のセットを満たす有向無サイクルグラフの集合を構築する（そのようなグラフが存在する場合）。いずれの種類のテスト結果も、変数間のd-分離関係を決定するためにのみ使用されるので、アルゴリズムへの入力は単にd-分離関係そのものであるかのように話すことにする[2]。

グラフGが**d分割のリストLを忠実に表現する**のは、L中のd分割がすべて、かつGの真である場合に限られるとしよう。**d分割のリスト**Lが忠実に表現するのは、ある非周期有向グラフだけである。実際には、分布がそれを生成する因果構造に忠実であっても、サンプリングエラーや採用する統計検定の仮定に対する小さな違反によって集団の特性に関する判断に誤りが生じることがある。分布族の誤った指定やサンプリングのばらつきに対する手続きの頑健性は、モンテカルロ・シミュレーションの手法で調べることができます。

以下の各アルゴリズムは、有向無サイクルグラフのクラス、または有向と無向のエッジを持つ単一の混合オブジェクト（グラフのクラスを表すパターン）を出力として持つことができます。パターンが有向無サイクルグラフの集合を表すことを思い出してください。グラフGは、以下の場合に限り、以下の方法で表されるグラフの集合の中にある：

(i) Gは、以下のような隣接関係を持ちます。Π ;

(ii) Πにおいて、*A*-*B*間のエッジが*A*→*B*を向いている場合、*G*において*A*→*B*を向いていることになる；

(iii) Πが*G*のパス<X ,*Y* ,*Z*>上の非遮蔽コライダであるならば、YはGの <*X* ,*Y* ,*Z*> 上の非遮蔽コライダである。

もし、いずれかのアルゴリズムが、*G*に線形に忠実な分布からの共分散行列、または*G*に忠実な分布からのセルカウントを入力として使うなら、入力は*G*に**忠実なデータで**あると言うことにする：

---

[2]実際、任意の統計的制約は、グラフにおいて対応するd-分離関係が成立する場合にのみ、その制約が分布において満たされるような分布とグラフのペアリングのアルゴリズムの入力として使用することができます。

定理5.1： いずれかのアルゴリズムの入力が*G*に忠実なデータである場合、各アルゴリズムの出力は*G*を表現するパターンになる。

しかし、このアルゴリズムでは、d-separation factに含まれるすべての方位情報を明示的に特徴付けるパターンを常に提供するわけではありません。ある辺のある方位にのみ一致し、対応する矢印を明示的に含まないパターンを生成することができます。

5.4.1 SGS Algorithm

SGSアルゴリズム（Spirtes, Glymour and Scheines, 1990c）の正しさは、定理3.4から得られる：

定理3.4： *P*がある有向無サイクルグラフに忠実である場合、*P*が*G*に忠実であるのは次の場合のみである。

(i) *G*のすべての頂点*X*，*Y*について，Xと*Y*が，*X*または*Y*を含まない*G*のすべての頂点集合に対して条件付きで依存する場合にのみ，Xと*Y*が隣接する；そして．

(ii) Xが*Y*に隣接し、*Y*がZに隣接し、Xと*Z*が隣接しないようなすべての頂点*X*, *Y*, *Z*に対して、*X* -> *Y* <- Zは、*X*, *Z*が*Y*を含み、*X*または*Z*を含まないすべての集合に対して従属条件である場合にのみ*G*のサブグラフになります。

SGSのアルゴリズムです：

A.) 頂点集合**V**に完全無向グラフ*H*を形成する。

B.) 各頂点*A*、*B*のペアに対して、**V**{*A*,*B*}の部分集合**S**が存在する場合、*A*は以下のようになる

。

とBが**Sで**d個に分かれている場合、HからAとBの間のエッジを削除する。

C.) ステップB)の結果の無向グラフをKとする。頂点A、B、Cのうち、AとBの組とBとCの組がそれぞれKで隣接し（A - B - Cと書く）、AとCの組がKで隣接しないような3つの頂点について、AとCをd分割する{B}の部分集合**Sが**存在しない場合にのみ、A - B - CをA -> B <- Cとして方向づける。

D.) くり返す

$A \rightarrow B$、BとCが隣接し、AとCか隣接せず、Bに矢印がない場合、B-Cを$B \rightarrow C$として方向づける。

AからBへの有向パスと、A-B間のエッジがある場合、オリエント A-Bを$A \rightarrow B$とする。

エッジの向きを変えることができなくなるまで。

5.4.1.1　複雑さ

信頼性と効率性は別物です。SGSアルゴリズムのステップB)では、*G*内で隣接する変数の各ペアについて、残りの変数の可能な部分集合をすべて調べる必要があり、当然ながら指数関数的な探索となる。最悪の場合、信頼性を維持するためには、この複雑さは避けられない。最悪の場合、変数X, Yの条件付き独立関係を、そのペアを含まないすべての頂点の部分集合で調べないような手続きは、失敗することになります。

5.4.1.2　SGSの安定性

データが不完全な場合に、アルゴリズムが合理的な信頼性を維持できるかどうかを検討する必要があります。ここでは、安定性という概念を非公式に使用することにする：直感的に小さな入力の誤差が直感的に大きな出力の誤差を生む場合、そのアルゴリズムは安定ではない。SGSアルゴリズムでは、入力の直感的に小さな誤差は、入力に誤って含まれたり、誤って除外されたりする少数のd-分離関係で構成されます。ステップBの直感的に小さなエラーは、出力に誤って含まれる、または省略される数本の無向きの辺である。ステップCの直感的な小さなエラーは、いくつかのエッジの向きが間違っていることである。

SGSアルゴリズムのステップB)は安定である。例えば、入力から正しいd-separation関係が1つ省略されたとしても、**U**以外にX, Yがd-separationされる集合がない限り、アルゴリズムは正しい無向グラフを生成することができます。その場合でも、ステップBはX - Yの接続を仮定する誤りを犯すことになるが、それ以外の誤りはない。XとYが真のグラフで隣接していても、**Uが**与えられたときにXとYがd-separatedである

と誤って判断された場合、アルゴリズムはXとYを接続することに失敗するが、それ以外のエラーは生じない。

SGSアルゴリズムのステップC)は、安定性に欠ける。入力のいずれかの要素、つまり無向グラフかd-分離関係のリストのどちらかに小さな誤りがあると、出力に大きな誤差が生じることがある（そして、しばしば生じる）。それは、衝突で発生する辺がグラフの他の辺の向きを決めるからであり、入力エラーによってアルゴリズムが衝突を含むか含まないかを誤って決定した場合、そのエラーはグラフの他の多くの辺の向きに影響を与える可能性がある。

例えば、ステップC)の無向グラフの入力で、X, Zを結ぶ辺が誤って省略され、X - Y - Zが正しく入力されたとする。このとき、XとZがd分割されていない場合

Yを含みX，Zを含まない変数の部分集合によって、アルゴリズムが誤ってYでの衝突を要求し、この要求が他のエッジの配向に波及してしまう。あるいは、真の構造がYでの衝突を含むにもかかわらず、ステップC)の入力でX - Yが省略された場合、Y - Zに固有の向きが与えられず、この不確実性はZを含む経路上の他の辺の向きにまで波及する可能性がある。

また、基礎となる無向グラフが正しい場合でも、入力されたd-分離関係のリストに誤りがあるため、ステップC)で不安定さが生じることがある。C)の入力において、XはYに、YはZに隣接するが、XはZに隣接しない場合、Yを含む**Sが**与えられたXとZの間のd-分離関係が入力から省略されても、Yを含む他の集合がXとZをd-分離しない限り、方向エラーは発生しない。しかし、真の有向グラフにおいて、XとYの間、YとZの間の辺がYで衝突し、XとZを含むd-分離関係、Yを含みXやZを含まない集合Uが誤って入力に含まれると、アルゴリズムはYで衝突はないと結論づけ、この誤りが他の辺に波及することもある。

ステップC)を少し考えてみると、本節の冒頭で挙げた4つの仮定のいずれかに違反した場合、その出力は有向無サイクルグラフの集合体でない可能性があることがわかります。これは、必ずしもアルゴリズムの欠陥ではない。両頭のエッジは、因果構造が因果的に十分でない場合や、入力に誤差がある場合（サンプリングのばらつきによる）などに発生することがある。両頭エッジは、測定されていない共通原因の存在を特定する理論的な役割を担っており、この問題については次章で詳しく説明する。

5.4.2 PC Algorithm

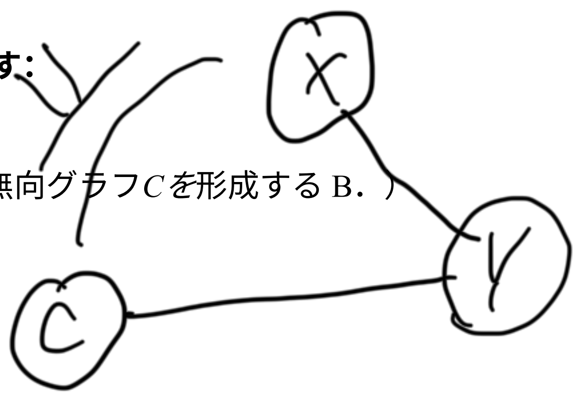
最悪の場合、SGSアルゴリズムは、条件付き独立関係や消失性部分相関に基づくあらゆるアルゴリズムと同様に、頂点の数に応じて指数関数的に増加するd-分離テストの数を必要とします。しかし、SGSアルゴリズムは非常に非効率的である。なぜなら、真のグラフのエッジでは、最悪のケースは期待されるケースでもあるからである。グラフ*G*にある任意の無向きの辺に対して、アルゴリズムの段階B)で実施しなければならないd分離テストの数は、真のグラフの連結性に影響されないため、疎なグラフであっても、頂点数が増加するにつれてアルゴリズムは急速に実行不可能となる。計算可能性の問題だけでなく、このアルゴリズムはサンプルデータに適用した場合の信頼性にも問題がある。標本分布から高次の条件付き独立関係を求めることは、一般に低次の独立関係を求めることよりも信頼性が低くなる。例えば、3つの値をとる37個の変数があるとして

の2つの変数の条件付き独立性を判断するためには、335の異なる状態の頻度間の関係を考慮する必要がある。

忠実な分布に対するSGS手順と同じ入出力関係を持ちながら、疎なグラフに対しては、離散の場合には高次の独立関係の検定を必要とせず、いずれにしてもできるだけ少ないd-分離関係の検定を必要とするアルゴリズムを希望する。以下の手順（Spirtes, Glymour, and Scheines, 1991）は、完全な無向グラフの形成から始まり、0次の条件付き独立関係を持つ辺を削除してグラフを「薄く」し、1次の条件付き独立関係で再び薄くし、といった具合に繰り返す。条件付き変数の集合は、条件付き変数の一方または他方に隣接する変数の集合の部分集合であればよい。

有向無サイクルグラフCにおいてAに隣接する頂点の集合を**Adjacencies**(C,A)とする(アルゴリズムでは、グラフCは継続的に更新されるので、**Adjacencies**(C,A)はアルゴリズムの進行とともに常に変化する)。

PCのアルゴリズムです：



A.)頂点集合**V**に完全無向グラフCを形成する B.）

$n = 0.$

くり返す

くり返す

Adjacencies$(C,X)\setminus\{Y\}$のカーディナリティがn以上であるようなCで隣接する変数XとYの順序ペアを選択し、カーディナリティnの**Adjacencies**$(C,X)\setminus\{Y\}$の部分集合Sと、もしXとYは、**S**がCからエッジ$X - Y$を削除し、**S**を**Sepset**(X,Y)と**Sepset**(Y,X)に記録したときにd-separatedとなる；

Adjacencies(C,X) \{Y}がn以上のカーディナリティを持つような隣接する変数XとYのすべての順序ペアと、カーディナリティnのAdjacencies (C,X) \{Y}のすべての部分集合**Sが**d分離のテストを受けるまで;

n=n+1である;

隣接する頂点X, Yの各順序ペアについて、**Adjacencies**(C,X) \{Y}がnより小さいカーディナリティを持つまで。

C.)ペアX, YとペアY, ZがそれぞれCで隣接するが、ペアX, ZがCで隣接しないような頂点X, Y, Zの各トリプルについて、Yが**Sepset**(X ,Z)にない場合に限り、X - Y - ZをX -> Y <- Zとして方向づける。

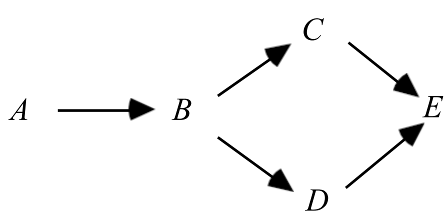
D. 繰り返す

$A \rightarrow B$、BとCが隣接し、AとCが隣接せず、Bに矢印がない場合、B-Cを$B \rightarrow C$として方向づける。

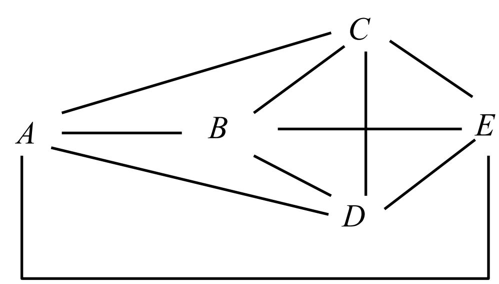
AからBへの有向パスと、A-B間のエッジがある場合、オリエントA-Bを$A \rightarrow B$とする。

エッジの向きを変えることができなくなるまで。

図1は、PCアルゴリズムの最初の2つの部分の動作をトレースしたものである：



真のグラフ



完全無指向性グラフ

---

n =　　0ゼロ次独立性がない

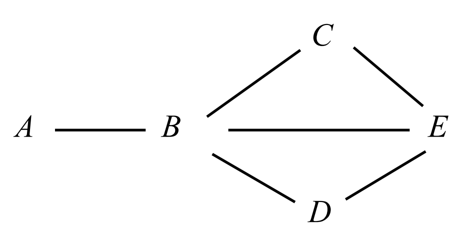
---

n =　　1一次独立性

結果としてのアジャカンシー

$A \perp\!\!\!\perp C \mid B$　　$AD \not\perp\!\!\!\perp B$

$A \perp\!\!\!\perp E \mid B$　　$CD \not\perp\!\!\!\perp B$



---

n = 2：　2次独立性

結果としてのアジャカンシー

$\perp\!\!\!\perp$

$B \quad E \mid \{C,D\}$

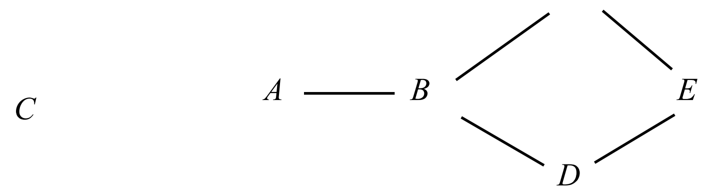


図1

この場合はそうではないが、アルゴリズムの段階B)は、真の有向グラフの隣接の集合が特定された後も、いくつかのステップでテストを続けることができる。図1の下部の無向グラフは、段階C)で部分的に方向付けされたことになる。その中で隣接が2つしかない変数のトリプルは、以下の通りである：

A - B - C;　A - B - D;
C - B - D　B - C - E;
B - D - E　C - E - D

*Eは***Sepset**(C,D)に含まれないので、C - EとE - DはEで衝突する。他の三角形はいずれも衝突相手を形成しない。このアルゴリズムが作り出す最終的なパターンを図2に示す。

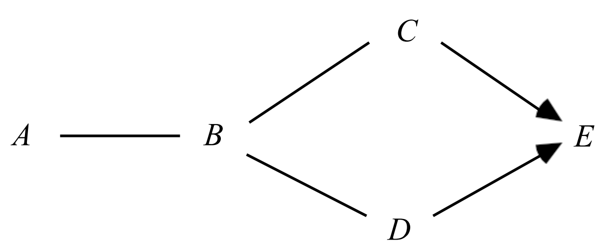


図2

図2のパターンは、忠実な区別不能クラスを特徴づける。図2の無向きの辺の向きは、Bでの衝突を含まないあらゆる向きが許容される。

5.4.2.1　複雑さ

グラフGに対するアルゴリズムの複雑さは、Gの最大次数によって境界を定められる。kを任意の頂点の最大次数とし、nを頂点の数とする。すると、最悪の場合、アル

ゴリズムが必要とする条件付き独立性テストの数は、以下のようになる。

(′・ω・ ˋ)□□□ク □□n −1□

^{2}□2□∑□ i□

i
=0

で結ばれている。

$$\frac{n^2 (n-1)^{k-1}}{(k-1)!}$$

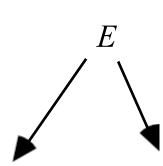
これは最悪の場合でも緩い上限である。nとkの最悪の場合、2つの変数がカーディナリティk未満の集合によってd分離されることはないと仮定しており、nとkの多くの値について、我々はその性質を持つグラフを見つけることができなかったのである。この問題の正式な予想複雑度解析はしていないが、最悪のケースは明らかに稀であり、最大次数kのグラフに必要な条件付き独立性検定の平均数はずっと少ない。実際には、100個もの変数を持つ疎なグラフを回復することが可能である。もちろん、計算量はkによって指数関数的に増加する。

このアルゴリズムの構造と、正しいグラフを見つけた後もテストを続けるという事実は、因果関係が疎であることが予想される非常に大きな変数セットに対する自然なヒューリスティック、すなわち、テストされる条件付き独立関係の順序の境界を固定するということを示唆している。

5.4.2.2　PCの安定性

理論的には、PCアルゴリズムはステップB)とC)の両方において不安定ですが、実際にはステップB)はステップC)よりもはるかに信頼できることが証明されています。

アルゴリズムのステップB)の初期段階で、誤って真のグラフからエッジを削除してしまった場合、真のグラフにない他のエッジが出力に含まれる可能性があります。次のような例を考えてみよう。



$$A \rightarrow B \rightarrow C \rightarrow D$$

図3

なぜなら、*E*は もはや*D*の隣接集合に含まれず、BとDはAと*C*のすべての部分集合に依存しているからである。エッジが誤ってグラフに残された場合、そのエッジの位置が

は、入力のd分割のリストに追加的なエラーはなく、理論的には配向できるはずのエッジが配向されないというエラーだけが発生します。

SGSアルゴリズムのステップC)と同じ理由で、アルゴリズムのステップC)が不安定になります。

PCアルゴリズムは、テストするd-分離関係の数が少ないため、SGSアルゴリズムよりも高速である。忠実なd分離関係のリストがあれば、2つのアルゴリズムは同じパターングラフのセットを出力します。しかし、サンプリングエラーなどにより、d-分離可能性関係のリストが忠実でない場合、2つのアルゴリズムは異なるパターングラフを出力することができる。次の例を考えてみよう。

$$A \longrightarrow B \longrightarrow C \longrightarrow D \longrightarrow E$$

図4

このグラフによると、Aと*E*は*B*、*C*、*D*の任意の空でない部分集合を与えられたときに互いにd分離する。もし、最初の無向グラフから*A*-Cと*E*-*C*のエッジを取り除いた後、Bと*D*の任意の空でない部分集合を与えられてAと*E*はd分離しないと誤って判断した場合、PCアルゴリズムは、Aと*E*の隣接の部分集合を与えられてAと*E*はd分離するかをテストするので、*A*-*E*間のエッジを誤って含めることになるだろう。一方、SGSアルゴリズムは、**V**{**A**,*B*}の任意の部分集合でAと*E*がd-separatedかどうかをテストするので、*C*でAと*E*がd-separatedなので、Aと*E*の間にエッジがないことを正しく認識するだろう。

一方、最初の無向グラフから*A* - Eと*B* - *E*のエッジを取り除いた後、*E*が与えられたときにAと*B*がd分離していると誤って判断した場合、SGSアルゴリズムは*A* -> *B*エッジを誤って取り除くことになります。もし*A* - Eと*B* - *Eのエッジが*先に削除された場合

、PCアルゴリズムは、*Eが与えられたときに*AとBが*d*-分離しているかどうかをテストしないので、*A -> B*エッジを正しく残すことができます。

PCアルゴリズムは「局所的」な情報を使ってエッジの有無を判断しようとするため、ある意味で不誠実な分布に「最も近い」グラフを生成することが保証されているわけではありません。次のような例を考えてみましょう。

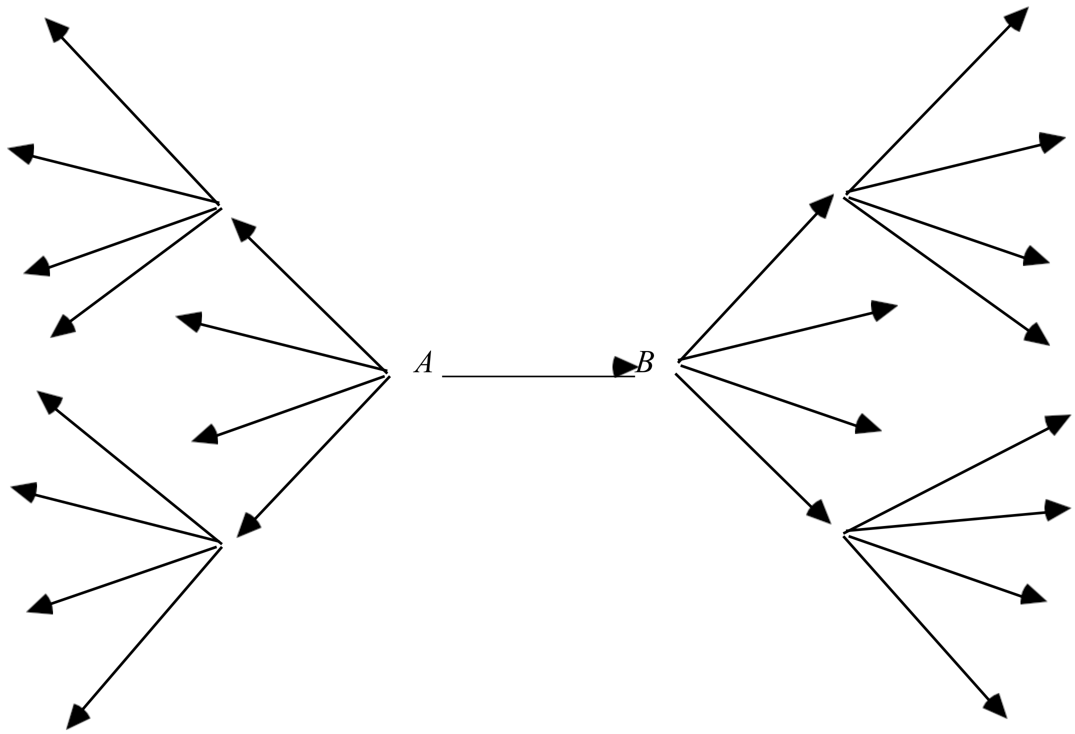


図5

このグラフに忠実な分布では、すべての変数が他のすべての変数に依存している。あるテストによって、Aと*B*が他の変数に対して独立であると判定されたとします。これは、パラメータの値が偶然に一致したためか、サンプリングエラーのためか、どちらかです。PCアルゴリズムは、この制約を満たすために、*A-B*のエッジを削除することになります。しかし、その際、グラフは切断される。その結果、*A*およびその左側のすべての子孫は、*B*およびそのすべての子孫から独立していることを意味するグラフとなる。つまり、1つの条件付き独立制約を満たすために、PCアルゴリズムは非常に多くの独立制約に違反するグラフを生成する可能性がある。多くのデータセットでは、2つの変数間の相関は消滅しないが、出力パターンがそれらを切断する。より高い信頼性を得るためには、この手順を修復アルゴリズムで補う必要があるが、離散的な場合はCooper and Herskovits Bayesian procedureで十分であろう。

5.4.2.3 PC* アルゴリズム

PCアルゴリズムは計算効率が良く、漸近的に信頼性が高いが、サンプルデータでは不必要なリスクを負うことになる。変数Aと*B*の間の無向きの辺を削除するかどうかを決定する際、この手続きは*A*の隣接集合のすべての部分集合と

有向無サイクルグラフに忠実な分布の場合、変数AとBが、**Parents**(*A*)または**Parents**(*B*)を与えられた条件付きで独立ならば、*A*-*B*間の無向パス上にある頂点のみからなる**Parents**(*A*)の部分集合または**Parents**(*B*)の部分集合で独立となる．それなら、*A*に隣接する変数の部分集合と、*B*に隣接する変数の部分集合で、Aと*B*の間の無向パス上にあるものが与えられたときに、Aと*B*の条件付き独立性をテストすれば十分である。

PCアルゴリズムとPC*アルゴリズムは、条件付き独立関係または相関の忠実なリストを入力とすると同じ出力が得られますが、サンプルデータのテストから決定した条件付き独立関係では異なる場合があります。PC*アルゴリズムは、PCアルゴリズムが犯すある種のエラーを回避することができます。しかし、PC*アルゴリズムが初期段階で*X*-*Y*間のパスを誤って切断した場合、無向グラフに*X*-*Y*エッジを誤って残す可能性があるが、同じデータがあればPCアルゴリズムがそのエラーを回避する可能性がある。さらに、PC*アルゴリズムが信頼性を向上させたとしても、ステップB)の各段階で、その段階で考慮するグラフ内のすべての無向パスを追跡しておかなければならないため、大きなコストで購入することになります。無向パスの数は一般に非常に多く、PC*アルゴリズムのメモリ要件は、変数の数が比較的少ない場合を除き、実現不可能であり、その場合は、PC*アルゴリズムが選択される可能性がある。しかし、真のグラフが疎である場合、無向グラフ*C*の平均次数が小さくなるまでPCアルゴリズムを使用することができ、それ以降はPC*アルゴリズムを使用することができます。この章の後半では、Christensen (1990)から引用した離散データに対する2つのアルゴリズムの性能について説明する。

5.4.2.4 テスト発注のためのスピードアップヒューリスティック

PCアルゴリズムのステップBでは、d-separationをテストするために、ある変数ペアと与えられたサイズのある部分集合**Sを**選択する。完全なグラフからエッジがより速く除去されるほど、アルゴリズムの後段で実施しなければならない探索が小さくなり、アルゴリズムがより速く実行される。したがって，最初にテストする変数ペア*A*，Bと部分集合**Sは**，*A*，*B*が**Sによって**d分離される可能性が最も高いものを選ぶのがベストである．我々は、テストの順序を選択する異なる方法を使用するPCアルゴリズムの3つのバリエーションを検討した。

ヒューリスティック1：変数の組と部分集合**Sを**辞書順にテストする。(これをPC-1と呼ぶことにする)。

ヒューリスティック2：まず、確率的に最も依存度が低い3つの変数ペアをテストする。条件付けの部分集合は辞書的な順序で選択される。(これをPC-2と呼ぶことにする）。

ヒューリスティック3：与えられた変数*Aについて*、まず、*Aに*最も確率的に依存する変数の部分集合を条件として、Aに最も確率的に依存しない変数*Bを*テストする（これをPC-3と呼ぶことにする）。

ヒューリスティック2の直感は、確率的依存度が最も高い変数は真のグラフで隣接している可能性が高く、したがって構築中のグラフから削除されることはない、一方確率的依存度が最も小さい変数は真のグラフで隣接している可能性が最も低いというものです。もちろん、このような関係が厳密に成立するわけではありません。

ヒューリスティック3の直感も同様である。変数*Aに*純粋に隣接していない変数*Bは*、*Aに*隣接する変数のある部分集合、または真のグラフで*Bに*隣接する変数のある部分集合が与えられると、*Aから*d分離される。*Aに対する*確率的依存度が最も高い変数は、真のグラフで*Aに*隣接する可能性が最も高いと仮定すると、Aに対する確率的依存度が高い*変数を*条件として、*Aに対する*確率的依存度が低い変数から*Aが*d分離されるかどうかをテストすることを示唆している。

5.4.3 IG（Independence Graph） アルゴリズム

Verma and Pearl (1990) は、SGSアルゴリズムのバリエーションを提案している。彼らの代替案では，有向無サイクル・グラフを探索する最初のステップは，無向独立グラフ *N* を構築することである．すなわち，各対の変数 *A*, *B* について，それらが他の

すべての変数の集合に条件付きで依存している場合，それらの間に無向エッジを導入する．有向非循環グラフに忠実な分布の無向独立グラフでは、どの変数の親も最大完全部分グラフ--徒党--を形成する。*N* で隣接する変数*A*,*B*の各ペアについて、*A*または*B*を含む*N*の閥の変数の部分集合が与えられたとき、AとBがd分離しているかどうかを判断します。

[3] 以下のヒューリスティクスにおいて、「確率的依存性が高い」とは、線形の場合は高い偏相関を、離散の場合は高いG^2統計量を意味する。

グラフの閥を決定することは不必要な計算を必要とし、最悪のケース以外では、2つの変数の条件付き独立性を、一方または他方の最大閥の全メンバーを条件として検定することは、不必要に高次の検定を伴うことになります。PCアルゴリズムのステップAを修正し、PCアルゴリズムの初期グラフを完全な無向グラフではなく、*無向独立グラフに設定し*、同じように進めるというものである。このアルゴリズムをIG(independence graph)と呼ぶことにする。

これらのアルゴリズムの効率は、明らかに独立性グラフをいかに簡単に構築できるかに依存する。相関行列の標準化逆行列の非対角要素は、対応する変数と残りの変数の間の偏相関係数の負である（Whittaker, 1990など参照）。したがって、線形ケースでは、標準化逆相関行列の項目が0でない場合に限り、Aと*Bの*間にエッジを置くことで独立性グラフを効率的に構築することができる。離散の場合、Fung and Crawford (1990)は最近、離散データから独立性グラフを構築する高速アルゴリズムを提案した。我々は、PCアルゴリズムの前処理として彼らの手順をテストしていない。

5.4.4 バリアブル セレクション

因果構造に関する予備知識は、実サンプルにおいて、これまで説明したアルゴリズムの結果をより有益なものにすることもあるが、変数の正しい選択は信頼できる推論に不可欠であり、そのためのアルゴリズム（少なくともこれらのアルゴリズム）は何の役にも立たない。

私たちは、変数を集約することもできるし、ある変数の異なる値を集約することもできる。第3章で取り上げたサーモンの架空の例のように、私たちは、より正確な自

然変数の不正確なバージョンである変数を測定することがあります。言い換えれば、他の変数に異なる影響を与える値を区別することができないのです。連続変数は、しばしば意図的にいくつかの離散的なカテゴリーに分類される。これは、分割表の手法が、関数従属性の形式（たとえば、線形かそれ以外か）について必要とされる実質的な仮定から解放された統計解析が期待できるためであり、分析すべき変数の一部が必ずしも離散ではなく、離散変数と連続変数の混合に関する問題に利用できる方法がほとんどないためでもある。時には、無知であろうと意図的であろうと、因果構造の異なる2つ以上の変数を1つの尺度に集約してしまうことがあります。集約と崩壊は、因果推論の信頼性にどのような影響を与えるのだろうか。

CがAとBの原因であり、Cの代用としてCほど正確ではなく、Cと完全には相関しないC'が使われた場合、C'を条件としてAとBが統計的に依存することがあることはすでに観察した。この種の例は、ある理論が、プロキシによって測定される原因を仮定するたびに現れる。例えば、Friedman（1957）は、消費は、代理人によってのみ測定できる「永久」所得によって引き起こされるという、多くの議論を呼んだ理論を提唱した。Friedmanの理論が真実であれば、消費と測定した所得を回帰すると、永久所得に対する消費の回帰係数の偏った推定値が得られ、消費と他の変数の間に説明できない相関を残すかもしれない。Klepper(1988)は、線形正規の場合、このような誤差がどのように有界化されるかを示した。

$C'=PROJ(C)$とし、$PROJ(C)$はCのn個の値の集合をm個の値の集合にマッピングする射影であるものとします。

< n個の値である。$P(A,B|C=c1)\neq P(A,B|C=c2)$、$PROJ(C=c1) = PROJ(C=c2)$ のようにCの値$c1$, $c2$が存在すれば、AおよびBはC'に基づいて独立ではない。独立関係は、変数の値を折りたたむことで消滅するのではなく、出現させることができる。変数B、Cが依存関係にあるとする。$C'=PROJ(C)$とし、$PROJ(C)$はCのn個の値の集合を$m<n$個の値の集合にマッピングする射影である。もし、$P(C = c1 \mid B) = P(C = c1)$となるような$C$の値$c1$が存在し、$PROJ(c1)$がユニークな逆数を持ち、$k$、$j$が1に等しくないすべてに対して$PROJ(c_k) = PROJ(c_j)$ なら、Bと C' は独立している。

Pearl (personal communication)は、非常に単純な集計が不誠実な分布を生み出すことを指摘している。Aが$C1$を、Bが$C2$を引き起こすとし、C1とC2はそれぞれ二値であり、変数間には他の因果関係はないとする。つまり$\{A, C1\}$は集合$\{B, C2\}$から独立しているが、AとC1は依存関係にあり、BとC2もそうである。C1とC2の異なる値のペアをコーディングする0, 1, 2, 3を取る変数Cを導入する。すると、A、B、Cの間の実際の

因果構造は、図6のようになる。

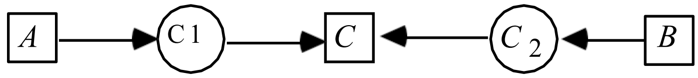


図6

しかし、合同分布では、Aと*B*は*C*を条件として独立しているので、合同分布はいかなる因果関係構造にも忠実ではない。この場合、分布の不忠実性は、変数間の決定論的関係によって不忠実となる分布のマージナルである、*C*を条件とするAと*B*の独立性

は、D分離可能性（第3章参照）を図6に適用することから直接導かれる。このようなことは、実際には時々起こるかもしれないが、常にテストして原理的に特定することが可能である：*A*を条件とすると、*C*の値は同じ条件付き確率で*C*の値を含む2つの同値類に分割され、*B*を条件とすると*C*の値は異なる同値類の組に分割される。*A*によって誘導される同値クラスを1つの変数の値、*B*によって誘導される同値クラスを別の変数の値とすると、$C1$と$C2$が回復する。

5.4.5 背景を取り入れる 知識を取り入れる

これらのアルゴリズムのユーザーは、検索を制約しうる多くの背景知識、あるいは少なくとも信念を持っている可能性がある。この知識は、グラフのある辺の存在や非存在についてかもしれないし、ある辺の向きについてかもしれないし、変数の時間順序についてかもしれない。この背景知識は、アルゴリズムにどのように利用されるのだろうか?

最も一般的な信頼できる事前信念は、発生時間によって変数を順序付けるか、部分的に順序付けるもので、*A*の測定が*B*の測定の前に行われたか、Aと*B*がそのように順序付けられた事象の正確な測定であると信じられている。このセクションのどのアルゴリズムも、このような知識を2つ利用するように簡単に修正することができる：

(i) *B*が*A*の現在の隣接関係のある部分集合を条件として*A*から独立しているかどうかをテストすることによって、Aと*B*が真のグラフで隣接しているかどうかを決定する際に、*A*より後の変数を含む変数の集合を条件として独立性をテストしないでください。

(ii) AとBが隣接し、Bが*A*より後である場合、エッジを$A \rightarrow B$のように方向付ける。

本書で紹介する例では、アルゴリズムはこのように変更されています。また、常識的な時間順序を使用することもありますが、そのような仮定がなされた場合は常に注意してください。

例えば、ある変数が他の変数に直接影響を与えるかどうかについての事前信念を、これらのアルゴリズムに組み込むことができる。事前信念が隣接を禁止する場合、アルゴリズムはその隣接をわざわざテストする必要はない。事前信念が、ある変数が他の変数に直接影響を与えることを要求する場合、他の辺の方向づけ手順において、対応する有向辺を課し仮定する。これらの手順は、制約のない探索の結果よりも事前の確信が優先されることを前提としており、この優先順位は必ずしも適切とはいえないが、それでもPCアルゴリズムによるTETRAD IIプログラムのバージョンに組み込まれている。

5.5 統計 決断

私たちが説明したアルゴリズムは完全にモジュール化されており、条件付き独立や消失する部分相関に関する必要な統計的決定を行うための任意の手順で適用することができます。その判断が優れていればいるほど、アルゴリズムに期待される性能は高くなります。条件付き独立関係の検定は、そのような決定の最も明白なクラスを形成するが、グラフ構造のd-分離可能性関係を与える任意の統計的制約があれば十分である。例えば、線形正規の場合、部分相関の消失は条件付き独立と等価であり、アルゴリズムが必要とする統計的判断は、部分相関が消失するという仮説の*t検定*で提供することができる。したがって、これらの仮定の下では、部分相関が消失するときに消失する統計量の検定で十分である。例えば、半部分相関係数の2乗に対する*F*検定が、対応する回帰係数に対する*t検定の*2乗に等しい（Edwards，1976）かもしれない。

本書の例では、Fisherの*zを用いて*$\rho_{XY.\mathbf{C}}=0$かどうかを検証しています：

$$z(\rho_{XY.\mathbf{C}},n)=\frac{1}{2}\sqrt{n-|\mathbf{C}|-3}\,\ln\left(\frac{1+\rho_{XY.\mathbf{C}}}{1-\rho_{XY.\mathbf{C}}}\right)$$

~~になります。~~ ←これ

$\rho_{XY.\mathbf{C}}$ = **C** を与えられた X と Y の母偏相関、｜C｜は **C** の変数数に等しい。X、Y、**C** が正規分布し、$r_{XY.\mathbf{C}}$ が **C** を与えられた X と Y のサンプル偏相関を表す場合、$z(\rho_{XY.\mathbf{C}},n) - z(r_{XY.\mathbf{C}},n)$ は標準正規分布（Anderson、1984）である。

離散的な場合、簡単のために2つの変数を考える。あるセルでのカウント数xijは、あ

るセルからN個のユニットをサンプリングして得られたランダム変数の値であると考えることにする。

[4] 因果的に十分な構造に対して、グラフGに適合する線形分布を課すことで得られる分布Pが、Gが線形に含意しない何らかの消失する偏相関を含意する場合、PはGに忠実ではないのでしょうか? PがGに*忠実でない*場合、Pは必ずしもGが線形に含意しない何らかの消失性偏相関を含意するのだろうか? どちらの質問に対しても、私たちはその答えを知らない。

多項分布である。x_{i+}は第1の変数が値i*を持つ*すべてのセルのカウントの合計を表し、同様にx_{+j}は第2の変数が値j*を持つ*すべてのセルのカウントの合計を表すとする。第1と第2の変数が独立であるという仮説のもと、確率変数x_{ij}*の*期待値は

$$E(x_{ij}) = \frac{x_{i+}x_{+j}}{N}$$

同様に、条件付き独立の任意の仮説におけるセルの期待値を、適切なマージンから計算することができる。例えば、第一の変数が第三の変数に対して条件付きで第二の変数から独立であるという仮説のもとでは、セルx_{ijk}*の*期待値は

$$E(x_{ijk}) = \frac{x_{i+k}x_{+jk}}{\text{エックスプラスケー}}$$

変数が3つ以上ある場合、この式は最初の3つの変数のi, j, k値の限界カウントの期待値を、他のすべての変数について合計することで得られるものに適用される。条件付き独立仮説が分布に与える独立制約の数は、条件付き独立関係の次数の指数関数であり、各変数が取りうる別個の値の数にも依存する。

このような独立性の仮説の検証には、とりわけ2つの統計が用いられてきた：

$$X^2 = \sum \frac{(Observed - Expected)^2}{\text{期待される}}$$

$$G^2 = 2\sum (Observed) \ln(\text{期待される})$$

はそれぞれ漸近的にχ^2として分布し、適切な自由度を持つ。本書の例では、**C を**条件として *A* と *B の*独立性を検定する場合の自由度を次のように計算する。変数*X* のカテゴリの数を返す関数を$Cat(X)$、**C の**変数の数を*n* とすると、検定の自由度(df)の数

は次のようになる：

$$df = \left(Cat(A)-1\right) \times \left(Cat(B)-1\right) \times \prod_{i=1}^{n} Cat(C_i)$$

構造的なゼロは存在しないと仮定する。ヒューリスティックな方法として、ゼロの項目がある分布の各セルについて、自由度の数を1つ減らす[5]。

セルの数は変数の数に対して指数関数的に増加するため、データ点数よりもはるかに多くのセルを持つケースを構築することは容易である。その場合、完全な結合分布のほとんどのセルは空となり、空でないセルでもわずかな数しかないことがある。実際，マージナル・トータルの一部がゼロになることは容易に起こり得ることであり，このような場合には，検定において自由度の数を減らさなければならない。信頼できる推定と検定のために、Fienbergはサンプルサイズを、検定中の仮説によって期待値が決定されるセルの数の少なくとも5倍とすることを推奨している。

離散データの場合、PCアルゴリズムにG^2を用いた独立性の検定を加えますが、シミュレーションの結果、X^2よりも正しいグラフになることが多いことがわかりました。

5.6 信頼性と確率 エラー

これまで説明してきたアルゴリズムのほとんどは、統計的な判断を必要とするもので、先ほど述べたように、仮説検定の形で実装することができる。しかし、検定のパラメータに通常の有意性を持たせることはできません。統計的検定の通常の快適さは、有意水準で、真の帰無仮説が検定で誤って失敗する限界の頻度を保証し、代替案に対する検出力で、指定した代替仮説が真である場合に偽の帰無仮説が棄却さ

れない限界の頻度を関数化したものである。非常に大きなサンプルを除いて、検索アルゴリズムで使用される検定の有意水準や検出力は、統計学的な判断をするために使用されるものではありません。

[5]ゼロエントリーのセルが与えられたときの自由度の減少を計算する正確な一般法則は知られていないようだ。Bishop, Fienberg, and Holland (1975)を参照されたい。

依存性は、検索に関する興味深いものの長期的な頻度を測定します。何がそうなのか?

検索手順で当然知っておきたいエラー確率には、以下のようなものがある:

1. モデルMが真であるとすると、サンプルサイズnにおいて、この手順がMと矛盾する結論を返す確率は何%か。

2. モデルM^*が真であるとすると、サンプルサイズnにおいて、手順がM^*と矛盾しないがMと矛盾しない結論を返す確率は何%か。

3. モデルMが真であるとき、サイズnのサンプルに対して、探索手順がMにない隣接を指定する確率は何パーセントか? ある探索手順がMにある隣接を省略する確率は何%か? ある探索手順が、Mにある辺にMにない矢尻を追加する確率は? 探索手順がM内の矢尻を省略する確率は? これらの確率は、ある特定の変数ペア、A,Bに対して何であるか?

大規模なモデルでは、ほとんどのサンプルで仕様の誤差が生じることが予想されるため、3のような質問が最も重要です。

これらの質問に対して分析的な答えを得ることはほとんど期待できない。仮説の数とサンプルサイズによっては、その確率は実際には有意水準よりもずっと高いかもしれない。しかし、いずれにせよ、誤った判断をする確率は、どの仮説をテストするかによって異なり、考慮されるすべてのアルゴリズムにおいて、それは実際の構造に複雑な形で依存する。さらに、各アルゴリズムは、必要な統計的判断が誤って行われても、正しい出力を出すことができる。例えば、グラフGにおいて、頂点Aと

*Bが*隣接していないとする。実際、Aと*BはC*、*D*、CとDなどの条件付きで独立であるとする。もし、探索手順でAとBが*Cの条件下で*独立であるという仮説が棄却された後、アルゴリズムがAとBが*Dの*条件下で独立かどうかを検証し、後者の独立を支持する判断を下した場合、先の誤りにもかかわらず、手順はAとBは隣接しないと正しく結論づけることになります。

特定のMと*M*について*、質問1、2、3の答えの推定値をモンテカルロ法で経験的に求めることができる。線形正規モデルのシミュレーション・パッケージは、現在

TETRAD II プログラムは、線形モデルおよび様々な分布を持つ離散変数モデルのためのシミュレーションパッケージを含んでいるため、市販の統計パッケージで一般的です。小規模なモデルであれば、100以上のサンプルを生成し、そのサンプルを検索処理にかけるのに数分しかかかりません。このプロセスは、私たちがシミュレーションのために*アドホックに*自動化したものであり、特定の検索結果の信頼性をテストするために一般的な方法で自動化することができますし、そうする必要があります。

5.7 見積もり

未測定の変数であっても、正規性の仮定のもとで因果仮説に従う最尤推定値を得る方法はよく知られている（Joreskog, 1981; Lohmoller, 1989））。また、正規性の仮定を放棄した場合には、通常の最小二乗法や一般化最小二乗法など、コンピュータによる様々な推定法が利用可能である。離散の場合、正の多項分布に対して、変数**Vの**集合に対するグラフの独立性制約に従うセルの最尤推定値（存在する場合）は、第3章の因数分解式の確率に限界度数を代入することで得られる（Kiiveri and Speed, 1982）。

$$P(\mathbf{V}) = \prod_{V \in \mathbf{V}} P\left(V \mid \mathbf{Parents}(V)\right)$$

測定された変数の共通の原因として働く測定されていない変数があるとき、我々が説明した手順から得られるパターンは、両端に矢印を持つエッジを持つことができます。このようなよくある状況では、離散的な測定変数の共同分布の最尤推定を得

る方法がわからないが、潜在変数を持つ対数線形モデルの推定に関する研究（Haberman, 1979）がガイドとして役に立つかもしれない。

5.8 事例と 応用例

模擬データと実データを用いたアルゴリズムを説明する。模擬データでは、現実的なサイズのサンプルに対するアルゴリズムの特性を例証する。経験的なケースでは、我々は

アルゴリズムが真実を生み出すかどうかは、しばしばわからない。しかし、少なくとも、研究者がデータの処理と説明に注意を払ったケースでは、アルゴリズムが公表された因果関係の説明を再現するか、ほぼ再現することは興味深いことである。また、このような美徳がないケースでは、アルゴリズムが公表された報告書で提唱された説明とは全く異なる説明を示唆することも興味深い。

PCアルゴリズムと別の手順であるFCI（Fast Causal Inference）アルゴリズムによって生成された回帰モデルと代替案の研究は、それぞれ第6章と第7章で潜在変数と予測について考察した後、第8章に延期される。

5.8.1 パブリッシングの原因 生産性

社会科学では、データセットの因果関係を説明するための「理論」の重要性が盛んに語られます。もちろん、データセットを説明する際には、常識に反する因果関係のグラフや変数の時間的順序に反するグラフは必ず排除されます。しかし、それに加えて、多くの実務家は、社会科学における観察データの因果的な説明を試みる際には、社会学、心理学、経済学、政治学などにおける原理の特殊性を経て、統計的依存関係だけから正しい説明を決定する可能性を否定することを要求する。このようなケースの多くでは、理論の必要性がひどく誇張されています。実際、科学文献にあるすべての「再帰的」構造方程式モデルについて、モデルの仮定が正しく、測定されていない共通原因が仮定されていない場合、分布が忠実であれば、母集団の統計的依存関係は、因果関係の有向グラフの基礎となる無向グラフを一意に決定します。そして、多くの場合、母集団の統計量だけでいくつかの、あるいはすべての辺の方向が決まる。変数が時間によって線形に並び、変数*A*が*B*より後に発生した場

合にのみ変数*B*の原因となりうる場合、統計的依存関係と時間順序は、分布が正でマルコフ条件と最小化条件が満たされていることだけを仮定して*一意の*有向グラフを決定する。因果関係の仕様を正当化するために文献を引用する努力は見当違いではないが、多くの場合、単位の近似的な均質性、サンプリングの仮定の正しさ、時には依存関係の線形性など、基本的な統計的仮定を確立することに努力を向ける方が良いだろう。

ここで、最近の、そしてかなり鮮明な例を紹介しよう。出版や引用率など、学問的成功の原因に関するかなりの文献がある。ロジャースと

Maranto (1989) は、社会学、経済学、心理学から引き出された学術的生産性の原因に関する仮説を検討し、「理論に基づく」複合的なモデルを作り上げました。

1966年から1976年の間に博士号を取得し、現在学術心理学者として働いているアメリカ心理学会の会員932名に勧誘と質問票を送付した。男女同数の心理学者が抽出され、心理学の学位を持っていない、心理学で最初の仕事をしなかったなどの回答者を削除すると、男性86人、女性76人のサンプルが得られた。

回答項目はグループに分類された。例えば、*ABILITY*グループは、被験者の学部におけるACT、*NMSQT*、選択性の平均スコア、Phi Beta Kappaの会員資格、卒業時の学部の優等生から構成されている。大学院プログラムの質（*GPQ*）*は*、学部教員の学問的質、全国ランキングを用いたプログラムの有効性、1978年から1980年の間に出版物を出した教員の割合、雑誌の編集者が学部教員にいるかどうかで構成されている。これらの回答項目は、未測定変数であるGPQと*ABILITYの*指標として、すなわち効果として扱われた。その他の指標は、最初の仕事の質（*QFJ*）、*SEX*、引用率（*CITES*）、出版率（*PUBS*）である。予備的な分析では、生産性の集約的な尺度（*PROD*）も使用された。そして、RodgersとMarantoが考えたさまざまな仮説は、線形「構造方程式モデル」として扱われた[6]。彼らは、クラスター変数間の相関を次のように報告している。

ABILITY	*ジーピーキュー*	*プリプロッド*	*ざいにちきょうどうたい*	*エックス*	*サイテス*	*PUBS*
1.0						
.62	1.0					
.25	.09	1.0				
.16	.28	.07	1.0			
-.10	.00	.03	.10	1.0		

.29	.25	.34	.37	.13	1.0	
.18	.15	.19	.41	.43	.55	1.0

[6]潜在変数であるGPQと*ABILITY*の出版生産性や*QFJ*など他の変数との相関をどのように求めたかは、論文からは明らかではない。例えば、因子構造を回帰モデルとして用い、被験者ごとに推定因子スコアを算出する方法や、構造方程式の自由パラメータに潜在変数の共分散を含め、LISRELなどのプログラムでその値を推定する方法などがある。一般に、これらの手順の結果は異なるだろう。

その後、社会学、経済学、心理学の各文献が示唆する因果理論について、非常に精緻な解説がなされる。RodgersとMarantoは、6つ以上の異なる構造方程式とそれに対応する原因理論を推定している。彼らが考える6つの構造とは、次のようなものである：

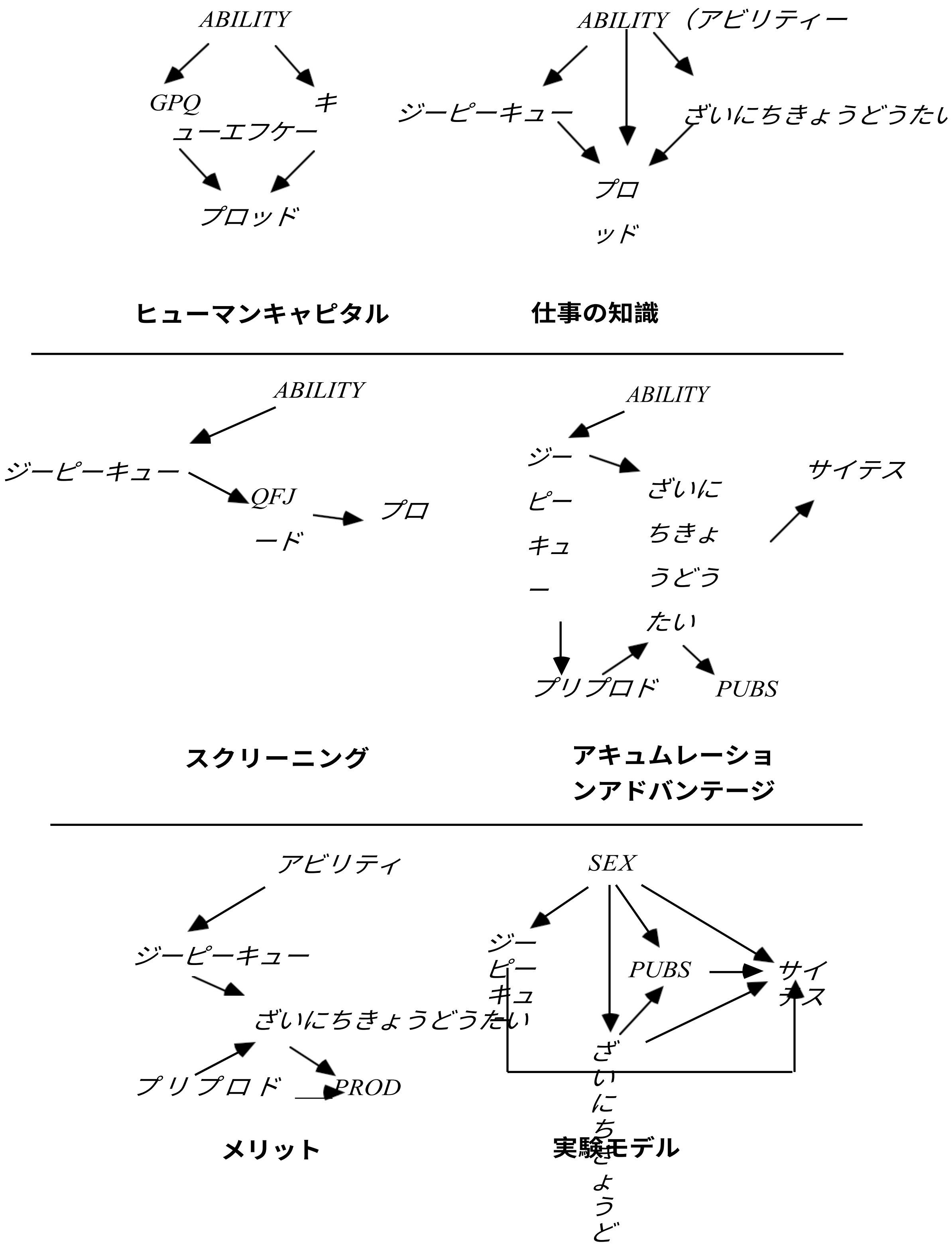


図7

グラフのラベルは、ロジャースとマラントがその因果関係グラフを導き出した社会科学的理論を端的に示しています。例えば、「人的資本」と「スクリーニング」のグラフは、次のようにして経済理論から得られたものである：

> 人的資本モデル（Becker, 1964）では、教育は関連する知識を伝達するため、生産性に直接的な影響を及ぼすとされている。人々は教育への投資

その限界費用（1年以上の教育による追加費用と逸失利益）が限界利益（1年以上の教育による生涯所得の増加）と等しくなるまで。能力の高い人は、能力の低い人に比べて、仕事と技能習得の両方でより生産的である。したがって、能力は生産性に直接的な影響を与えるが、能力の高い人ほど学校から多くのものを得られるため、教育を通じて間接的な影響を与える。また、職務経験もOJTを行うことで生産性を高める。人的資本の枠組みでは、教育の量だけでなく質も重要である。

スクリーニング仮説は、暗黙のうちに能力を生産性の主要な決定因子とみなしている。雇用主は最も生産性の高い応募者を雇いたいと考えるが、能力は直接観察できるものではない。個人は、雇用主に対して自分の能力を示す手段として教育に投資する。教育の限界費用は能力に反比例し、能力と教育水準の間に正の相関がある。したがって、学歴によって応募者を選別することで、雇用主は能力によって雇用することになる（Spence, 1973）。このモデルでは、教育は生産性に直接影響せず、能力との関連を通じてのみ影響する。教育の質のばらつきは、スクリーニング・モデルと一致する（Wise, 1975）。

実証モデル」は、社会理論に訴えない先行研究から得られたものである。

これらのモデルに基づく構造方程式系は、いずれも現象を救わない。しかし、「理論的」モデルのすべてのエッジを組み合わせ、さらにもっともらしいと思われる2つのエッジを加え、統計的に重要でない（0.05での）依存関係を捨てた結果、ロジャースとマラントが代わりに、データによく合う別の因果構造を提案しました:

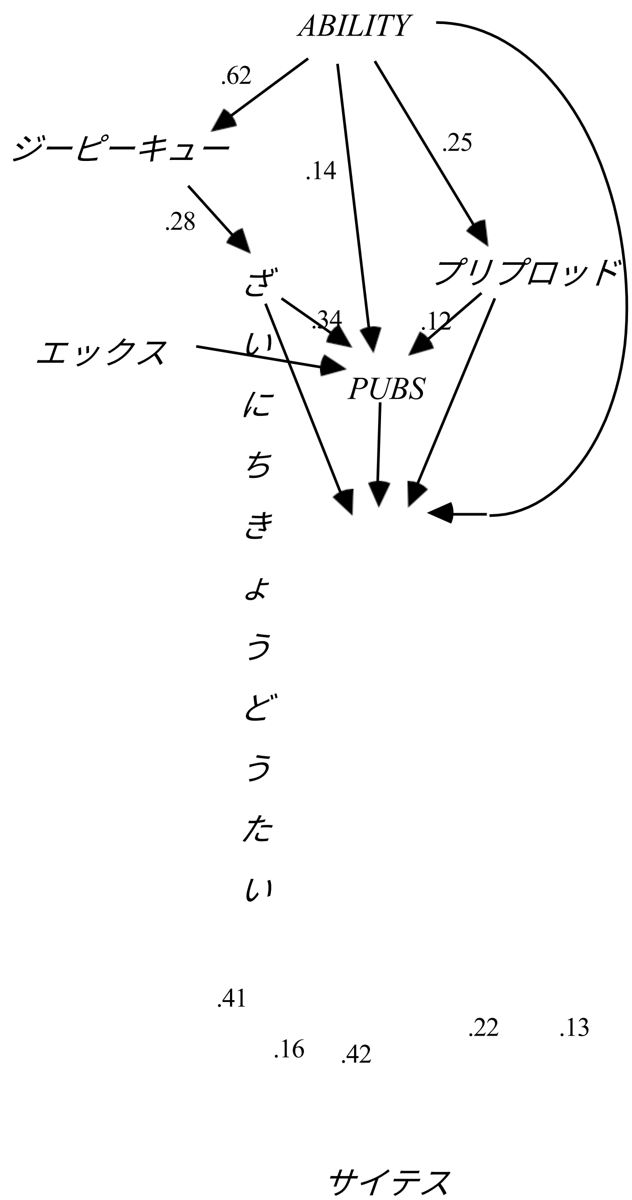


図8

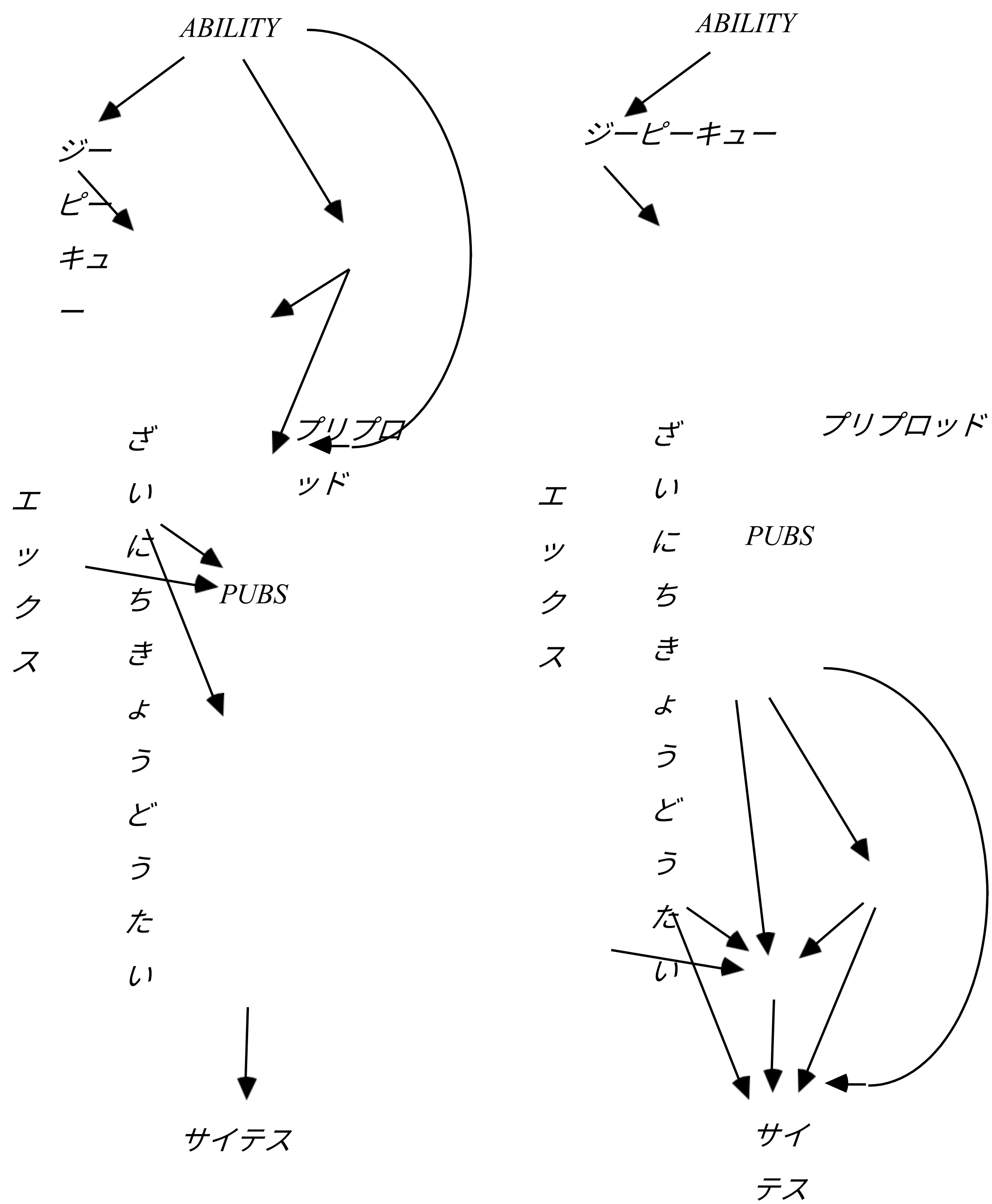
理論」を通したツアーはほとんど意味がなかったように見えるが、ロジャースとマラントはそうではないという:

> 観測された相関のパターンだけに基づく因果モデルは、非常に疑わしい。どんなデータでも、いくつかの代替モデルによって適合させることができる。そのため、理論に基づいた予想に基づき、最適なモデルを構築することにした。PUBSと*CITESという*2つの生産性指標と、5つ

> の因果関係の先行要因を用いて、6つの理論が特定するすべての経路を含む複合モデルを最初に推定した。このモデルでは、ABILITYと*PREPROD*の相関の観測値と予測値の間に大きな正の乖離が生じ、一つまたは複数の重要なパスが省略されていることが示唆された。6つの理論の最初の解釈を再検討した結果、2つの経路が見落とされているとの結論に達した。そのひとつは、ABILITYから*PREPROD*への道であり、もうひとつは、ABILITYから*PUBS*への道である。この2つの経路を追加し、有意でない経路はすべて複合モデルから削除して、最良の適合モデルにたどり着きました。

もしRodgersとMarantoの理論が完全に正しければ、その有向グラフの下にある無向グラフは条件付き独立関係によって一意に決まり、向きもほぼ一意に決まるはずです
。

PCアルゴリズムを適用して、常識的な時間順序との相関を、有意水準0.1の偏相関の検定で調べると、図9の左側のグラフのようになり、Rodgers and Marantoモデルと並べて表示された。



PC **出力ロジャースとマラントグラフ**

図9

RodgersとMarantoのモデルの1つのエッジを除くすべては、データとドメインの常識的な知識（変数の時間順序）から瞬時に生成されます。EQSはこのモデルの$\chi 2$を13.58、自由度を11、p=0.257としている。もし、検索
有意水準として.05を使用して手順を繰り返すと、プログラムは*PREPROD*→を削除します。
*PUBSの*エッジです。このモデルをEQSプログラムで推定しテストしたところ、$\chi 2$が19.2、自由度12、p値0.08と、誤差の確率ではなく、適合度の推定値としてとらえるべき数値です。

ロジャースとマラントのモデルの本質を見つけるために、常識以外の社会科学的理論が必要だという主張は、明らかに誤りである。また、ロジャースとマラントの探索の予備的な結果も、社会科学的な理論を信頼する根拠にはならない。これに対して、私たちはPCアルゴリズムの信頼性と限界について十分な知識を持っている。TETRADとEQSを使った研究全体にかかる時間は数分です。PCアルゴリズムではなく、SGSアルゴリズムを使って、モデルのわずかな変形を得ることができる。

5.8.2 教育と 豊かさ

Rindfuss, Bumpass and St. John (1980)は、既婚女性における結婚時学歴（*ED*）と第一子出産年齢（*AGE*）の相互影響に関心をもっていた。彼らは理論的な根拠から、図10の左のモデルについて長々と論じているが、上から順に回帰因子は以下の通りである：

DADSO＝　父親の職業

RACE＝レース　レース

NOSIB　＝兄弟姉妹がいないこと

FARM　＝ファームの背景

REGN　=アメリカの地域

ADOLF　＝対象者の幼少期の家族に2人の大人が存在すること

REL＝レル　宗教

YCIG =　タバコの煙

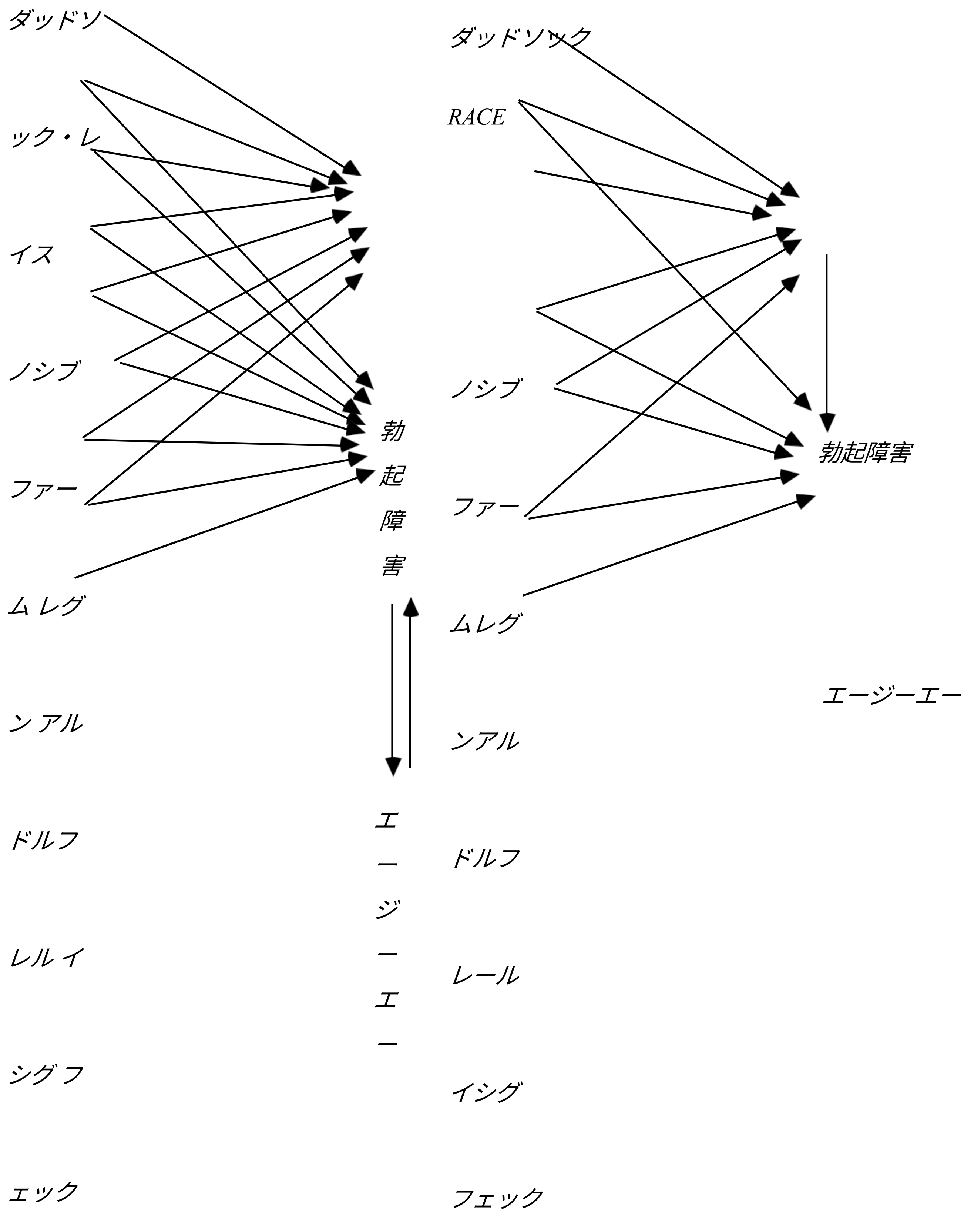
FEC　＝対象者が流産したかどうか。

回帰因子は相関がある。サンプルサイズは1766で、共分散は以下の通りである。

ダッドソ	*レース*	*ノシブ*	*ファーム*	*レグン*	*アドルフ*	*レル*	*イシグ*	*FEC*	*ED*	*AGE*
456.676										
-.9201	.089									
-15.825	.1416	9.212								
-3.2442	.0124	.3908	.2209							
-1.3205	.0451	.2181	.0491	.2294						
-.4631	.0174	-.0458	-.0055	.0132	.1498					
.4768	-.0191	.0179	-.0295	-.0489	-.0085	.1772				
-0.3143	.0031	.0291	.0096	.0018	.0089	-.0014	.1170			
.2356	.0031	.0018	-.0045	-.0039	.0021	-.0003	.0009	.0888		

18.66	-.1567	-2.349	-.2052	-.2385	-.1434	-.0119	-.1380	.0267	5.5696	
16.213	-.2305	-1.4237	-.2262	-.3458	.1752	.1683	.1702	.2626	3.6580	16.6832

驚いたことに、調査員は係数を推定したところ、$AGE \rightarrow ED$のパラメータがゼロであることを発見しました。EDと*AGE*が他の変数の原因ではないという事前情報があれば、PCアルゴリズム（検定に.05有意水準を使用）は、図10の右のようなモデルを直接見つけることができ、回帰変数間のつながりは描かれない。



Rindfuss, et al.理論モデル; **AGE→ED係数は統計的に有意ではない**

TETRAD IIモデル

図10

5.8.3 女性の オーガズム

Bentler and Peeler (1979) は、281人の女子大学生から性格と性的反応に関するデータを得た。その内容は、神経症（*N*）と外向性（*E*）を測定するEysenck Personality Inventory、異性愛行動目録（*HET*）、単性愛行動目録（*MONO*）、マスターベーションに対する否定的態度尺度（*ATM*）、性交およびマスターベーション体験の主観評価目録である。因子分析により、これらの回答から一次元的と思われる尺度が形成され、そのうちの2つの尺度（*SCOR*）と（*SMOR*）は、性交体験と自慰体験の主観的評価から得られた。

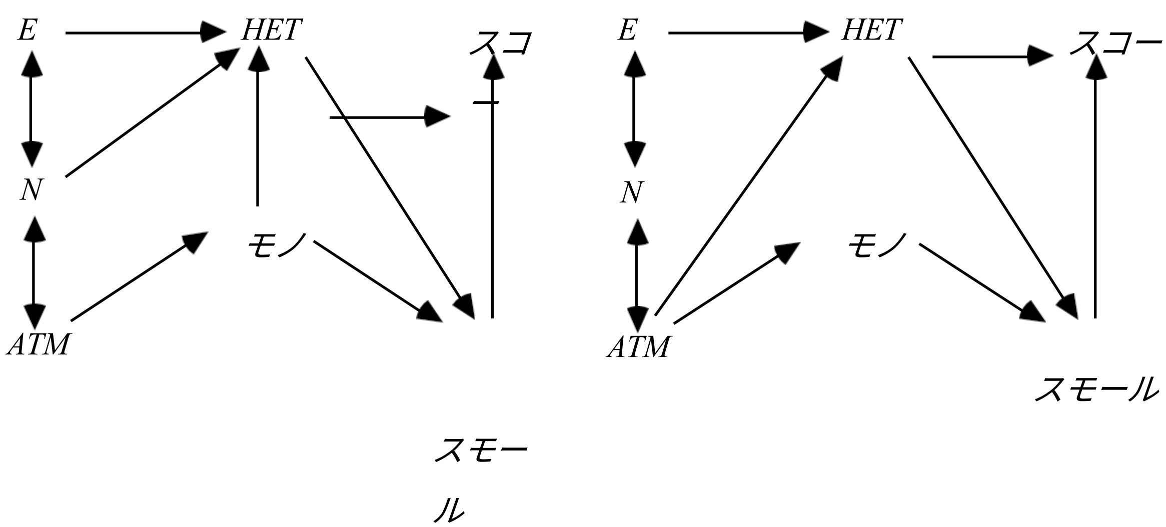
研究者らは、（1）マスターベーションとコイタスにおける主観的オーガズム反応は、それぞれ異なる内的プロセスによるものである、（2）外向性、神経質、そして、その内的プロセスによるものである、という二つの仮説に興味を持った。

自慰行為に対する態度は、SCORや*SMOR*で測定されるオーガズム反応性に直接的な影響を与えず、HETや*MONO*で測定される個人の性体験の履歴を通じてのみ、その現象に影響を与えるのです。

提示されたデータは、尺度と在庫スコアの相関関係のみであるため、今回は尺度の形成については触れないことにする：

E	*N*	*ATM*	*HET*	*モノ*	*スコー*	*スモール*
1.0						
-.132	1.0					
.009	-.136	1.0				
.22	-.166	.403	1.0			
-.008	.008	.598	.282	1.0		
.119	-.076	.264	.514	.176	1.0	
.118	-.137	.368	.414	.336	.338	1.0

BentlerとPeelerは、相関を説明するために2つの線形モデルを提供している。そのモデルと関連する漸近的なχ2の確率値を図11に示す。



モデル1

$p < .001$

モデル2

$p = .21$

図11

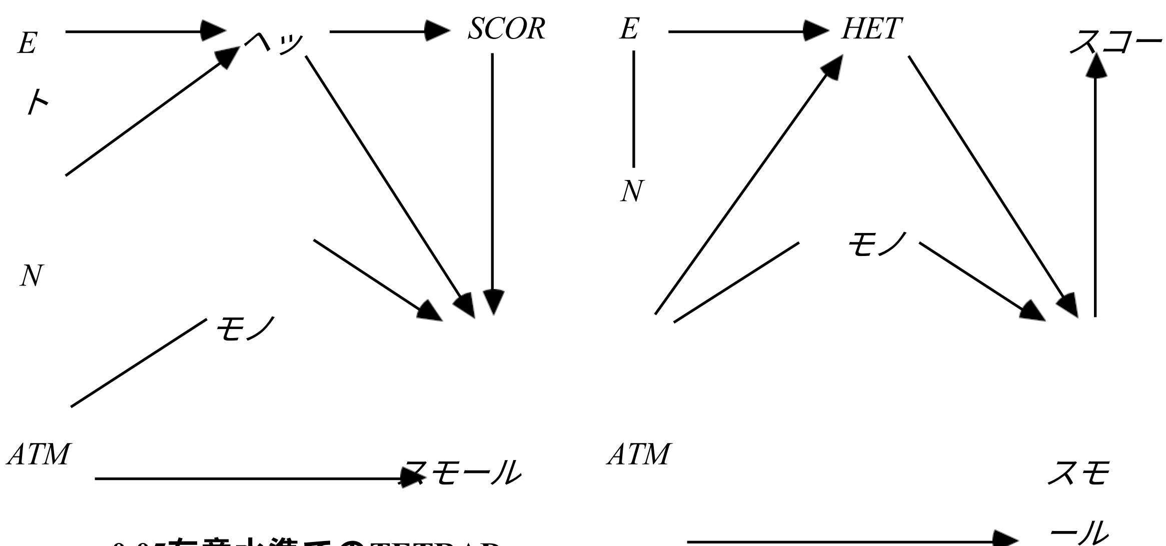
2番目のモデルだけが現象を救う。著者は次のように書いています。

> ...外向性（*e*）、神経症（*n*）、自慰行為に対する態度（*atm*）がオルガスム反応性に影響を与えるのは、その効果を通じてのみであるという仮説と一致するオルガスム反応性のモデルを開発することが可能であることがわかった。

これらの変数が異性愛（*het*）および自慰（*mono*）体験に与える影響。

その結果、仮説2は受け入れられるようである（p.419）。

この議論の論理は明らかではない。著者が述べているように、"データを同じようにうまく説明する他のモード（sic）も考えられるように開発されうることを忘れてはならない。"(p. 419).しかし、例えば、*ATM*がSCORや*SMOR*に直接影響を与えるモデルによって、データが同じようにうまく説明できるのであれば、仮説2を受け入れるべき理由はない。PCアルゴリズムを使えば、そのようなモデルを容易に見つけることができる。



0.05有意水準でのTETRADパターン

0.10有意水準でのTETRADパターン

$p = .148$

図12

図12の右側のモデルは、自由度12で漸近的なχ^2値が17、$p(\chi^2)$=0.148である。

PCアルゴリズムは、データに基づいて否定できないモデルを発見し、BentlerとPeelerに反して、マスターベーションに対する態度がマスターベーション中のオーガズム

体験に直接影響することを仮定するものである。

5.8.4 アメリカの職業 構造

Blau and Duncan (1967) のアメリカの職業構造に関する研究は、全米科学アカデミーから社会研究の模範的作品として賞賛され、ある統計学者 (Freedman 1983a) から科学の乱用として批判されている。ブラウとダンカンは、20,700人のサンプルを使って、1962年の職業（*OCC*）を決定する際の教育（*ED*）、最初の仕事（J*1*）、父親の教育（*FE*）、父親の職業（*FO*）の役割について予備理論を提示した。

彼らは、無向きのエッジが説明のつかない相関関係を表す次のようなグラフで理論を提示しています：

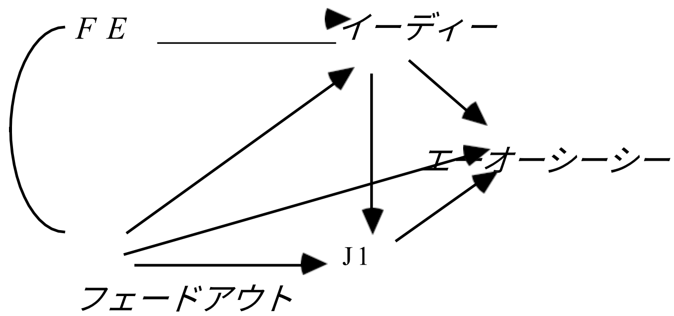
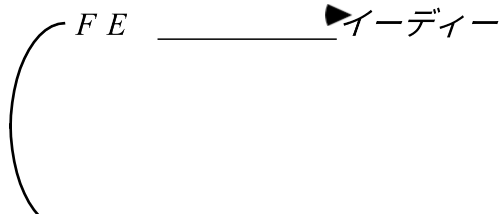


図13

BlauとDuncanは、依存関係は直線的であると主張している。彼らの顕著な結論は、父親の学歴が職業と最初の仕事に影響を与えるのは、父親の職業と対象者の学歴を通してのみであるということである。

BlauとDuncanの理論は、Freedmanによって恣意的で正当性がなく、統計的に不十分であると批判された（Freedman 1983a）。実際、この理論をEQS（Bentler 1985）やLISREL（Joreskog and Sorbom, 1984）のプログラムの漸近的χ2尤度比検定にかけたらのモデルは決定的に棄却され（$p < .001$）、Freedmanはブートストラップテストでも棄却されると報告している。

従来の有意水準0.05で偏相関の消失を検定した場合、常識的な時間による変数の順序が与えられ、ブラウとダンカンの共分散からPCアルゴリズムは以下のグラフを生成する：



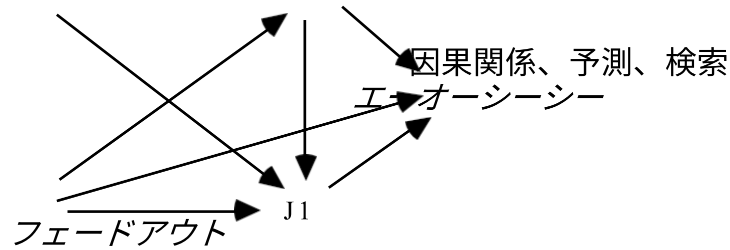


図14

この場合、すべてのコライダは三角形で発生し、遮蔽されていないコライダは存在しない。したがって、データは因果的なつながりの方向を決定しないが、時間順序はもちろん各辺の方向を決定する。によって隣接が生じることを強調する。

を、事前の制約なしに、データから完全に作成しました。このモデルは、同じ尤度比検定で$p > .3$となっています。

このアルゴリズムは、BlauとDuncanの理論に、FEと$_{J1}$の間の直接的なつながりを追加する。FEと$_{J1}$の間の接続は、消失性部分相関の検定に使用する有意水準が0.0002である場合にのみ消失する。1962年のFEから*OCC*への有向エッジと一致する消失部分相関のコレクションを決定するためには、消失部分相関の仮説を0.3以上の有意水準で棄却する必要がある。有意水準0.0001のデータで見つかった条件付き独立関係は、BlauとDuncanの有向グラフに忠実である。

Freedmanは、アメリカの集団では、これらの変数間の影響は家族ごとに異なると考えるべきであり、したがって、集団のすべての単位が同じ構造係数を持つという仮定は不当であると主張している。同様の結論は、別の方法でも得られる。第3章で述べたように、ある集団が、因果構造は同じだが分散と線形係数が異なる線形システムの部分集団の混合物からなる場合、係数が独立に分布しているか、混合物が特別な割合でない限り、集団の相関はどの部分集団の相関とも異なり、それぞれの部分集団で独立した変数が全体では相関している可能性がある。異なる線形構造を持つ部分集団が混合され、これらの特別な条件が得られない場合、相関関係から得られる有向グラフは通常完全なものとなる。BlauとDuncanのデータを当てはめるには、完全グラフになるには1辺だけ足りないグラフが必要であることがわかる。

ダンカン、フェザーマン、ダンカン(1972)が同じ実証研究から構築した別の線形モデルでも、同じモラルがより鮮明である、というより、より鮮明である。*FEは*父親の学歴、*FOは*父親の職業的地位、SIBは回答者の兄弟数、*ED*は回答者の学歴、OCCは回答者の職業的地位、*INC*は回答者の所得を意味し、彼らは社会経済的背景と職業的

達成に関する以下のモデルを開発した。

この場合、両頭の矢印は単に残差の相関を示すだけである。このモデルは自由度が4であり、EQSの尤度比検定に完全に失敗している（χ2が165）。この相関行列をTETRAD IIプログラムに渡すと、明らかな時間的順序である

の変数があれば、PCアルゴリズムは完全なグラフを作成する。

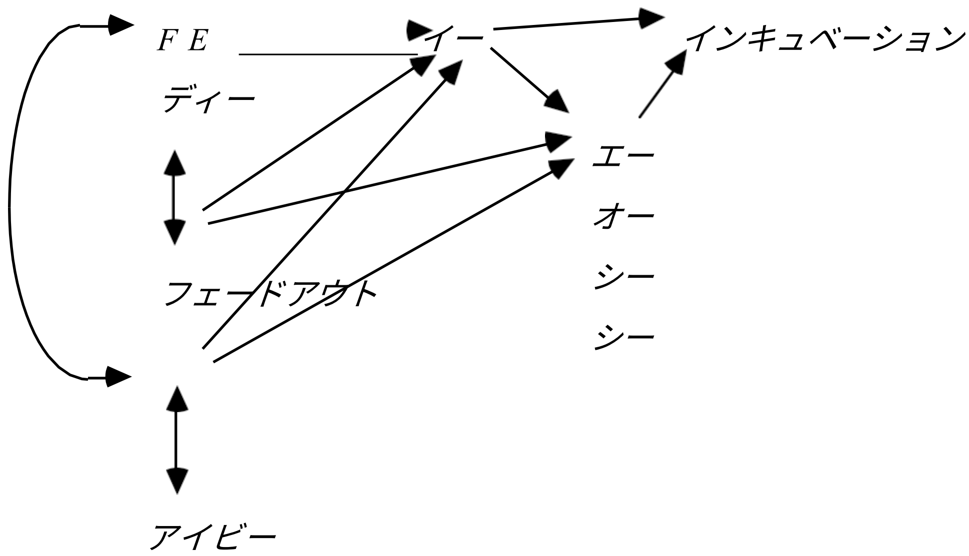


図15

5.8.5 ALARM ネットワーク

救急医療の因果関係をシミュレートするために開発されたALARMネットワークを思い出してください:

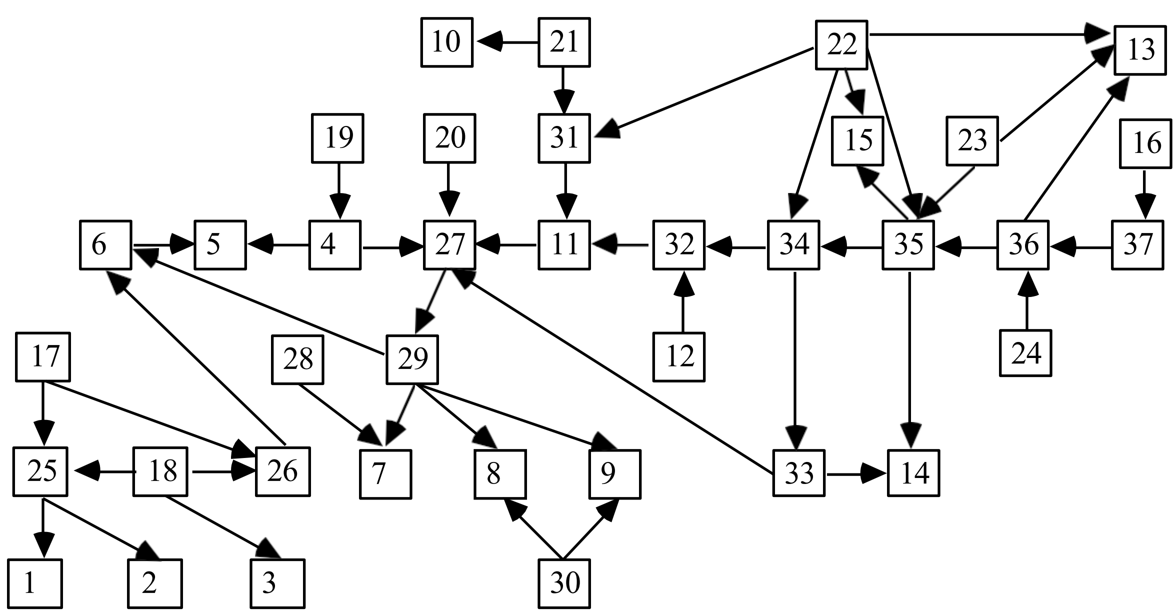


図16

SGSとPC*のアルゴリズムは、これほど大きな問題では実行できません。我々は、PCアルゴリズムをALARMネットワークの線形バージョンに適用した。同じ有向グラフを使い、グラフの各有向辺に0.1～0.9の値を持つ線形係数をランダムに割り当てた。次数ゼロの変数について結合正規分布を用い、3組のシミュレーションデータを作成し、それぞれのサンプルサイズは2,000とした。共分散行列と

PC-1 アルゴリズムを実装した TETRAD II プログラムのバージョンに、サンプルサイズが与えられました。この実装は共分散行列を入力として受け取り、パターンを出力する。変数の方向に関する情報はプログラムには与えられなかった。Decstation 3100で実行したところ、各データセットについて、プログラムがパターンを返すのに要した時間は15秒未満であった。各試験において、出力されたパターンはALARMネットワークの2つのエッジを省略し、1つのケースではALARMネットワークに存在しないエッジを1つ追加しました。

関連するテストでは、さらに10個のサンプルを生成し、それぞれ10,000個のユニットを用意しました。結果は以下のように採点された：母集団の相関関係からPCアルゴリズムが生成するであろうパターンを**真のパターンと呼びます**。また、アルゴリズムがサンプルデータから推測したパターンを**出力パターンと**呼びます。**エッジ存在エラーオブコミッション（Co）**は、変数のペアが、出力パターンでは隣接しているが、真のパターンでは隣接していない場合に発生します。*A-B*間のエッジ*e*が真のパターンと出力パターンの両方に存在する場合、*e*が出力パターンでは*A*を矢印としているが真のパターンでは矢印としていない場合（*B*についても同様）、**エッジ方向エラーオブコミッションが発生します（省略エラー（Om）は**それぞれのケースで同様に定義されます）。結果は、それぞれの種類の可能な誤りの数に対する実際の誤りの数の比率の試行分布の平均として表わされる。サンプルサイズ10,000での結果は以下のように要約される：

#trials	%Edge Existence Errors (エッジ存在エラー)		エッジディレクションエラー	
	委員会	省略	委員会	省略
10	.06	4.1	17	3.5

100個の変数を持つ同様の連結グラフのデータについて、10回の試行で、PC-1アルゴリズムは平均134秒、PC-3アルゴリズムは平均16秒を要しました。

Herskovits and Cooper (1990) は、ALARMネットワークのために、2、3、4個の値を持つ変数を用いて離散データを作成しました。彼らのデータをもとに、PCアルゴリズムによるTETRAD IIプログラムは、無向グラフのほぼすべてを再構成し（ある試験では2つの辺を省略し、別の試験では1つの辺を追加した）、ほとんどの辺を正しく配向させた。また、ほとんどのエッジが両方向に配置されるエラーが発生した。同じネットワークの線形データで使用したのと同じ指標で分解すると、（サンプルサイズ10,000でHerskovits and Cooperから得たシミュレーションデータで）結果は以下のようになります：

トライアル	%Edge Existence Errors (エッジ存在エラー)		エッジディレクションエラー	
	委員会	省略	委員会	省略
1	0	4.3	27.1	10.0
2	0.2	4.3	5.0	10.4

5.8.6 処女性

Reiss, Banwart and Foreman (1975) は、学部生の女性を対象に、婚前交渉に対する態度、大学の避妊クリニックの利用、処女性など、いくつかの態度の関係を検討したレトロスペクティブ研究である。クリニックを利用したことのある女性とそうでない女性の2つのサンプルが得られたが、年齢、教育、親の教育など、関連する背景変数に有意な差はなかった。Fienbergは、3つの変数についてクロスクラス化したデータを示している：婚外性交に対する態度（E）（常に間違っている、常に間違っていない）、処女性（V）、避妊クリニックの利用（C）（利用した、利用しなかった）。変数はすべて二値である。PCとSGSの手続きにより、すぐに以下のパターンが得られる：

$$E \text{ ——— } V \text{ ——— } C$$

図17

一つの賢明な解釈は、態度が性行動に影響を与え、それがクリニックの使用を引き起こすというものである。Fienberg (1977)は、対数線形法で同じ結果を得ている。

5.8.7 リーディング クラウド

Coleman (1964) は、3398人の小学生を対象に2回のインタビューを行った研究につい

て述べている。それぞれのインタビューでは、自分が「有力な群衆」の一員であるかどうか、また有力な群衆に対する態度が好ましいか好ましくないかを判断するよう求められた。このデータは、Goodman (1973a, b)とFienberg (1977)によって再分析されている。Fienbergの表記法を用いて、Aと*B*を一次面接での質問、Cと*D*を二次面接での対応する質問を表すとする。データは、Fienbergによって次のように与えられている：

			セカンドインタビュー			
会員意識			+	+	-	-
			+	-	+	-
	会員意識					
ファーストインタビュー	+	+	458	140	110	49
	+	-	171	182	56	87
	-	+	184	75	531	281
	-	-	85	97	338	554

Fienbergは、対数線形解析後の結論を、図18のパス図にまとめている。彼は、両頭の矢印をどのように解釈するかについては説明していない:

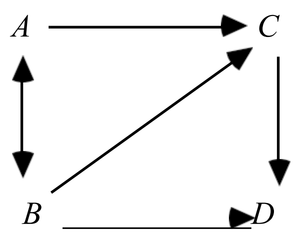


図18

PCアルゴリズムに、Aと*B*の後にCと*D*が発生することを伝え、検定の有意水準を通常の0.05とすると、プログラムは図19のようなパターンを作成する:

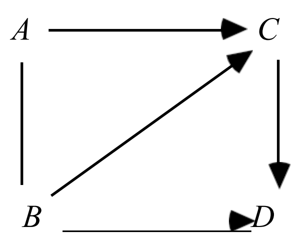


図19

PCモデルの無向きの辺をAから*B*への有向きの辺とすると、様々な細胞数の期待値が

得られ、Fienberg（p.127）の期待値とほぼ同じになることがわかった。

カウント[7] ただし、これはほぼ完全なグラフであり、サンプルが異なる因果構造の混合物であることを示す可能性がある。

5.8.8 大学 計画に与える影響

Sewell and Shah (1968) は、ウィスコンシン州の高校3年生10,318人をサンプルに、5つの変数を調査した。変数とその値は以下の通りである：

SEX	[男性＝0、女性＝1］です。
IQ＝知能指数、	[最低＝0、最高＝2]。
CP = カレッジプラン	[yes = 0, no = 1]です。
PE＝親の励まし	[低＝0、高＝1]。
*SES*socioeconomic status	[最低=0、最高=3] 彼

らは次のような因果関係の仮説を提示している：

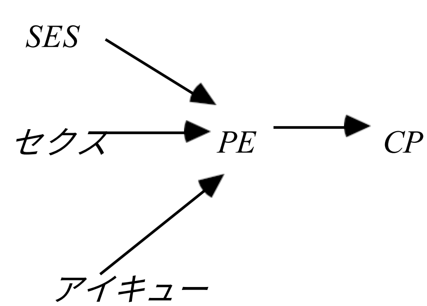


図20

このデータはFienberg（1977）によって再分析され、対数線形モデルを用いて因果関係の解釈を与えようとしたが、グラフ的な解釈を与えることができないモデルを発見した。

時間ごとに変数を以下のように順序付ける事前情報がある場合。

1	*エックス*		
2	*アイキュー*	*ピーイー*	*SES*
3	*商業手形*		

7わずかな差は、丸め誤差に起因するものと推定されます。

というように、後の変数が前の変数の原因であることを特定できないように、PCアルゴリズムによる出力は構造になっています：

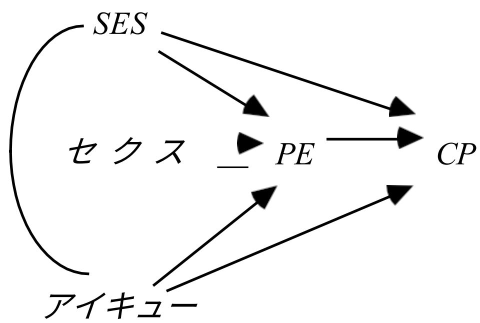


図21

このプログラムは、IQと*SES*の間のエッジを方向付けることができません。子供の知能が家族の社会経済状態を引き起こすということは非常に考えにくいので、唯一の賢明な解釈は、*SES*が*IQ*を引き起こすか、あるいは両者に共通の測定不能な原因がある、ということです。前者を選ぶと、有向グラフができ、その結合分布は標本から直接推定することができます。例えば、男性が大学進学の計画を持つ確率（*CP*）の最尤推定値は0.35であり、女性の確率は0.31であることがわかる。このサンプルで判断すると、*IQ*が低く、親の励ましがなく、社会経済的地位（*SES*）が低い子供が大学進学を計画する確率は0.011、さらに悩ましいことに、同じ条件で*IQ*が高い子供が大学進学を計画する確率は0.124しかない。

5.8.9 中絶 意見

Christensen (1990) は、Race (*R*) [white, non-white], Sex (*S*), Age (*A*) [six categories], Opinion (*O*) on legalized abortion (supports, opposes, undecided) を変数とするデータセットで対数線形モデルの選択と探索手順を説明しています。前方選択法では、43個の

対数線形モデルのフィッティングが必要である。後方消去法では22回、Aitkinによる方法では6回、Wermuthによる別の後方消去法では23回のフィッティングが必要である。これらの方法はいずれも、大規模な変数セットでは全く機能しない。Christensenは、「最良の」対数線形モデルは、最大クリークが[RSO]と[OA]である無向条件独立グラフモデルであることを示唆している。これを図22に示す。

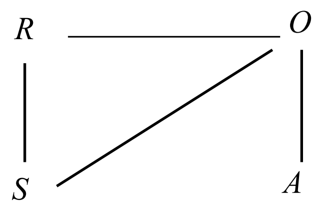


図22

その後、Christensenは、データの再帰的因果モデル（Kiiveri and Speed, 1982の用語で）を提案する。彼は実質的な根拠として、混合グラフを提案している

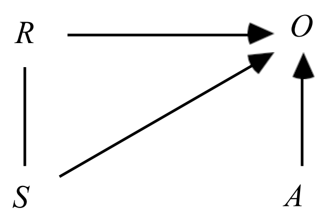
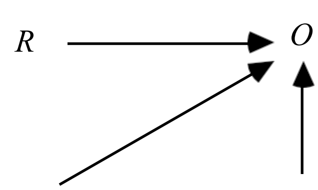


図23

で、"Rと*S*の間の無向きの辺は...Rと*S*の間の相互作用を表している "と言っている。図23は、これまで説明した意味での因果関係モデルではない。図23は、*R*-*S*エッジの2つの方向をメンバーとする因果グラフの等価クラスを表すパターンと解釈できるが、Christensenのデータでは Rと*S*はほとんど独立である.

この例はPC*アルゴリズムを使うには十分小さく、独立性検定に有意水準0.05を用いると、まさに図24が得られます。忠実であると仮定すると、図24の統計的仮説は、図22の対数線形モデルで要求される、*O*を条件とする{*R*,S}と*A*の独立性と矛盾している。



S A

図24

有意水準がやや低い（.01）R と *O* は独立と判断され、同じアルゴリズムで *R*→*O* の接続が省略される。有意水準.05のこのデータでは、PCアルゴリズムも図24のグラフを作成するが、*R*→*O* の接続は省略される。PC*アルゴリズムとPCアルゴリズムの出力の違いは、次のように生じる。両アルゴリズムとも、中間段階で図24の無向グラフを生成する。その無向グラフにおいて、*A* はRと *O* の間のどの無向パス上にも存在しない。そのため、PC*アルゴリズムは、*A* に対するRと *O* の条件付き独立性をテストせず、*R* - *O* エッジを残したままにしておく。一方、PCアルゴリズムは、*A* 上のRと *O* の条件付き独立性をテストし、肯定的な結果を得て、*R* - *O* エッジを削除する。

5.8.10 ランダム グラフを用いたシミュレーションテスト

本章で取り上げたアルゴリズムの速度と信頼性を検証するため、SGS、PC-1、PC-2、PC-3、IGの各アルゴリズムを、多数の模擬例でテストした。グラフ自体、線形パラメータ、サンプルは全て擬似的にランダムに生成されたものである。本節では、線形データと離散データの両方についてサンプル生成手順を説明し、線形データの場合のシミュレーション結果を示す。離散データでのシミュレーション結果は、回帰の章で検討する。

グラフの頂点の平均次数は2、3、4、5、変数の数は10または50、サンプルサイズは100、200、500、1000、2000、5000である。これらのパラメータの各組み合わせについて、10個のグラフを作成し、各グラフに忠実な1つの分布を求め、その分布から1つのサンプルを採取した。

計算量に限界があるため、SGSアルゴリズムは10変数のグラフでのみテストされた。

5.8.10.1 サンプル生成

すべての擬似乱数はUNIXの "random "ユーティリティによって生成されました。各サンプルは3段階に分けて生成される：

(i) グラフは擬似的にランダムで生成されます。

(ii) 線形係数（線形ケース）または条件付き確率（離散ケース）は、擬似的にランダムに生成されます。

(iii) モデル用のサンプルは擬似的にランダムで生成されます。

それぞれのステップについて、より詳しく解説していきます。

(i) ランダムグラフ生成器の入力は、平均次数と変数数である。変数は、エッジが下位の変数から上位の変数にしか行かないように順序付けられ、サイクルの可能性を排除している。いくつかの手順では辞書的順序を使用しているため、変数名はランダムにスクランブルされ、エッジで接続された変数ペアの間に系統的な辞書的関係が得られないようにした。各変数ペアには、以下の確率*pが割り当てられている。*

$$\frac{\text{平均度数 変数数}}{1} -$$

各変数のペアについて、0から1までの区間の一様分布から数値が引かれる。
描かれた数が*p*.8以下の場合のみ、エッジがグラフに配置される

(ii) 連続分布のシミュレーションでは、各内生変数に「誤差」変数を導入し、グラフの各辺に対して.1〜.9の線形係数の値をランダムに発生させた。離散の場合は、変数の値の範囲を手で選択し、*n個の*値をとる各変数について、カットオフポイントのランダムな選択によって単位区間を*n個の*サブ区間に分割する。そして、単位区間に分布（例えば、一様）が課される。

(iii) 離散の場合、生成された各分布に対して、各外生変数について、分布に従って0から1.0の間の乱数を生成し、その数が入るカテゴリに従って変数値を割り当てることによって、各サンプルユニットが得られる。内生変数の値は、変数の親の得られた値に対する条件付き確率で与えられる確率でランダムに値を選択することによって得られた。線形の場合、誤差項を含む外生変数は標準正規分布から独立に生成され、内生変数の値はその親の一次関数として計算された。

8この方法で「現実的な」グラフが生成されるかどうかはわからない。この方法で生成されたグラフのうち、好ましくない特徴として、孤立した変数が存在することがある。非公式に調べたところ、アラームネットワークに似ていないトポロジーがあった。

5.8.10.2　結果

前述したように、信頼性にはいくつかの側面がある。真のグラフにある無向きの辺を省略したり、真のグラフで隣接していない頂点間の辺（有向、無向を問わない）を含むことで、手順が誤ることがある。真のグラフにない辺については、その向きについての事実はないが、真のグラフにある辺については、真のグラフにある矢印を省略したり、真のグラフにない矢印を含むことによって、手順が誤ることがある。5.8.5節と同じ方法でエラーを数える。

各手順は、すべての試行で有意水準0.05を使用して実行されました。テストされた5つの手順は、同じように信頼できるわけでも、同じように速いわけでもありません。SGSアルゴリズムは最も遅いですが、いくつかの点では信頼できることがわかります。次のページのグラフは、その結果を示しています。グラフ上の各ポイントは数字で、これはデータを生成する有向グラフの頂点の平均次数を表している。10個の変数を持つランダムに生成されたグラフに基づく線形モデルのデータについて、サンプルサイズに対するPC-1 PC-2、PC-3、IG、SGSアルゴリズムの実行時間と信頼性をプロットし、同様に50個の変数を持つランダムに生成されたグラフに基づく線形モデルについてこれらのアルゴリズムのうち最初の4つの信頼性をプロットする。それぞれの場合、結果は次数2、3、4、5のグラフについて別々にプロットされている。

以下のような定性的な結論が得られます。

矢印とエッジの省略率は、サンプルサイズ1000程度まではサンプルサイズに応じて劇的に減少し、それ以降はずっと緩やかに減少しています。

矢と刃の委託率は、矢と刃の省略率に比べ、サンプルサイズによる劇的な変化はあまりない。

変数の平均次数が高くなるにつれて、平均エラー率は非常にほぼ直線的に増加するが、PC-2アルゴリズムは、グラフの平均次数が高い場合、エッジの省略に関して他のアルゴリズムよりも信頼性が低い傾向があることがわかる。

PC-1、PC-3、IG、SGSの各アルゴリズムは、長所と短所を補うものである。グラフがスパースでない場合、どの手順もすべての次元で信頼できない。信頼できる次元の1つは、辺の追加である：真のグラフで2つの頂点が隣接していない場合、グラフの平均次数やサンプルサイズに関係なく、これら4つの手順のいずれによっても誤って出力される可能性はほとんどない。

一方、平均次数が高くサンプルサイズが小さい場合、各手順の出力は真のグラフのエッジの50%以上が省略される傾向があることがわかった。サンプルサイズが大きく平均度が低い場合は、真のエッジの数%しか省略されないが、平均度が高い場合は、サンプルサイズが大きくても省略されるエッジの割合が大きい。例えば、サンプルサイズ5000、平均次数5の場合、PC-1は真のグラフのエッジの30%以上を省略している。

アローコミッションエラーはエッジコミッションエラーよりもはるかに一般的であ

る。グラフに矢印が発生しない場合、サンプルサイズが大きく、真のグラフが低次でない限り、どの手順でもかなりの確率で矢印が出力されることになる。平均次数が2程度の10変数でサンプルサイズが1,000以上の場合、SGSとIGのアルゴリズムはかなり信頼性が高く、矢印の誤差は6%程度である。同じ条件でPC-1、PC-2アルゴリズムの誤差は約20%である。SGSアルゴリズムの場合、これらの関係は矢印の省略の問題で逆転します。真のグラフに矢印がある場合、その手順が出力に矢印を含めない確率はどのくらいでしょうか。その答えは、PC-1とPC-2では約8%、SGSでは約20%である。IGアルゴリズムは、低いサンプルサイズでは矢印の省略に対する信頼性が非常に低いものの、高いサンプルサイズではわずかに信頼性が高いだけです。

PC-3アルゴリズムのリターンタイムは、他のアルゴリズムに比べて劇的に小さくなっています。また、実行時間も平均次数によって急激に増加することはありませんが、平均次数が増加するにつれて、この手順はより多くのエッジコミッションエラーを生成します。

この結果は、問題の大きさ、答えが欲しい質問、出力の特徴に応じて、プログラムを合理的に様々な方法で使用できることを示唆している。信頼性については、離散的な場合にもほぼ同じ結論が期待できるが、絶対信頼度は低くなる。変数の数が多い場合は、SGSアルゴリズムが全く実行できないことを除けば、同じパターンが成り立つはずである。

これらの手順で生成されたグラフの局所的な「修復」、特にエッジの省略エラーとアローのコミッションエラーについて、さらに研究が必要である。この方法が確率1で正しい判定に収束するためには、判定に用いる有意水準はサンプルサイズが大きくなるほど小さくなるはずで、より高い有意水準（例えばサンプルサイズが100未満の場合は.2、100～300の場合は.1）の使用により、サンプルサイズが小さい場合の性能が向上する可能性があります。

5.9 結論

本章では、非常に多くの変数がある場合でも、疎な因果構造を確実に復元できるいくつかのアルゴリズムについて説明し、その応用例を示す。これらのアルゴリズムは、離散的な場合は条件付き独立性の検定、線形的な場合は消失する部分相関の検

定を使って実装されている。我々は、関連する確率関係を決定するためのこれらの検定の使用が最適であるとは主張しないが、統計的決定法の改良は、アルゴリズムに前置きすることができる。PC*とSGSアルゴリズムを除いて、説明した手順は、真の因果グラフが疎である限り、多数の変数に対して実行可能である。

また、少なくとも因果関係グラフの隣接関係を見つけるためには、非常に単純な代替手段をいくつか用意すれば、それなりにうまくいくはずである。

5.10 背景 備考

発見問題という考え方は、すでに推定問題の概念に含まれており、推定量が一貫しているという要求は、本質的に、ある種の発見問題を解くという要求である。この考え方を一般的な非統計的設定に拡張したものが、Putnam (1965)によって提案され、独立してGold (1965)によって提案され、その後、コンピュータ科学、数学言語学、論理学の文献で広範囲に展開されている (Osherson, Stob and Weinstein, 1986).

因果的・統計的仮説のための多かれ少なかれ体系的な探索手順は、今世紀初頭のSpearman (1904) とその弟子たちの著作に見出すことができる。段階的探索のベイズ版は、Harold Jeffreys (1957)によって提案された。Thurstone（1935）の因子分析は、正確な発見問題から切り離されたアルゴリズムによる探索を開始した。Thurstoneは因子分析をデータの単純化を見つけるための装置以上のものとは考えておらず、同様の考え方は、その後の多くの統計的探索の提案で述べられている。探索に関する膨大な統計文献は、ほとんどフィッティング関数の最適化にのみ焦点を当てている。

SGSアルゴリズムは1989年にGlymourとSpirtesによって提案され、Spirtes, Glymour and Scheines (1990c) に掲載されました。Verma and Pearl (1990b)は、その後、クリークを調べるより効率的なバージョンを提案した。PCアルゴリズムのバージョンはSpirtes and Glymour (1990)によって開発された。ここで紹介するバージョンは、アルゴリズムのステップC)の効率において、PearlとVermaが提案した改良を含んでいる。ベイズ発見手続きはHerskovitsの論文(1992)で研究されている。

再帰的因果モデル」の最尤推定手順は、Kiiveri (1982)の博士論文で開発されたもので

ある。この構造の数学的特性は、Kiiveri, Speed and Carlin (1984)でさらに説明されている。

第6章

因果的充足性を持たない発見アルゴリズム

6.1 はじめに

前章では、一般的な統計的ファンタジー、すなわち、典型的なデータセットでは、測定された変数間の統計的依存性の一部が測定されていない共通の原因によるものでないことが知られていることに従った。我々はほとんど常に、測定した変数の原因のすべてを測定することができず、2つ以上の測定した変数の原因である変数の測定に失敗することが多い。社会科学のデータを調べると、ある研究の変数が他の研究の変数と関連しているように見えることが多いという印象がある。記録保存の慣習から、計量経済学者は、ある経済に関する研究において、他の経済に関する研究で因果的な役割を果たすと考えられる変数を無視せざるを得ないことがある（Klein, 1961）。心理学、社会心理学、計量経済学の多くの研究では、関心のある本当の変数が測定されていないか、プロキシや "指標 "によってのみ測定されていることがある。危険因子への曝露が病気を引き起こすことを示すと主張する疫学研究では、統計的な関連が危険因子と病気の共通の原因によるものではないことを示すことが議論の負担となる。想像しうるすべての関連性を測定できるわけではないので、測定されていない変数が関連を「混乱」させないことを証明できなければ、議論は根本的に不完全なものとなってしまう。もし私たちが考えるように、関連する変数が測

定されていないかどうかを考慮せずに信頼できる実証研究を進めることができないとすれば、発表された対照のない実証研究はほとんど信頼できないことになる。

実験的研究でも非実験的研究でも、測定されていない変数の存在が認識されないと、測定された変数間の因果関係についての誤った結論や、これらの変数の一部を操作する政策の効果についての誤った予測につながることがあります。測定されていない共通原因の有無を特定するために、信頼できるデータに基づく方法が用いられるまでは、観察データからの因果関係の推測は、よくても推測に過ぎず、悪くても疑似科学に過ぎない。*そのような方法はあるのだろうか?* この問題は、統計学の最も重要な理論的課題の一つであるはずである。

測定不能な共通原因を検出するための統計的手法、すなわち疫学者が好む用語では「交絡」は、主に心理測定学で開発されたもので、共通原因の存在と数に関する基準が、世紀末から特殊な統計モデルのために求められてきた。その結果、線形システムに関する文献には、潜伏変数に関する基準（例えば、Charles Spearman (1904) vanishing tetrad differences）が含まれているが、線形性を仮定しても必要でも十分でもないことが証明された。2つの潜在的共通原因の基準はKelley (1928)によって紹介され、関連する基準は因子分析で使用されているが、すべての統計的依存関係が測定されていない共通原因によるものであると仮定しない限り、それらは正しくない。測定変数が離散的で、その値が観察されない連続ベクトル・パラメータの確率関数である問題では、 その次元性に関して多くの基準が開発されている（Holland and Rosenbaum, 1986）。Suppes and Zanotti (1981) は，離散変数について，すべての測定変数が測定されていない共通の原因の効果であり，測定変数のすべてのペアが潜在変数の条件で独立である，形式的な潜在変数モデルが常に存在することを示した．彼らの議論は、モデルがマルコフ条件のみを満たすことを前提としており、分布が忠実であることが要求される場合には、結果は成立しない。

疫学者の間では、（Breslow and Day, 1980; Kleinbaum, Kupper and Morgenstern, 1982）危険因子Aへの曝露と疾患Bとの間の統計的依存関係が「因果関係」であるかどうかを決定する手段として、Surgeon General's report on Smoking and Health（1964）で紹介された基準が、時に今でも提唱されており、明らかにAがBを引き起こしAとBには共通の原因がないという意味となっている。その基準は、(i)投与量に伴う反応の増大、(ii)危険因子と疾患との統計的依存関係が特定の疾患サブグループとリスク曝露の特定の条件に固有であること、(iii)統計的関連性が強いこと、(iv)危険因子への曝露

がリスク増大期間に先行すること、(v)代替説明が欠如していること、である。

測定された変数の共通の原因がすべて測定されている因果的に十分なシステムにおいてさえ、このような基準は、原因と相関する変数を分離することができない。測定されていない "交絡因子 "の問題に対処することさえできないのです。基準(v)は回避である。非管理下の研究の問題は、まさにデータの代替説明が多すぎることである。基準(iv)は、測定された共通原因があるか測定されていない共通原因があるかを判断する上で全く役に立たない陳腐なものである。基準(iii)は、「もし観察された関連が因果関係ではなく、単に他の因子と疾患との因果関係の反映であるならば、この後者の因子は前者の因子よりも疾患と（相対リスクで）強く関連していなければならない」（Breslow and Day, 1980）という理由で擁護されている。

ブレスローとデイに代わって、複数の共通原因が存在するすべての仮説に対して単純性を訴えることができるかもしれないが、複数の原因メカニズムが存在する医学では、それはありえない主張である。最初の2つの基準は、Aと*B*が共通の原因を持っている状況とそうでない状況とを分けるものではない。

この章では、測定されない共通原因の存在が、測定された変数間の因果関係について調査者をどのように惑わすか、また、測定されない共通原因の存在をどのように検出できるかについて、多かれ少なかれ体系的に説明する。これらの問題は、一般的なケースと、すべての構造が線形であるケースとで、別々に扱う。しかし、この章の中心的な目的は、マルコフ条件と忠実性条件を仮定すると、測定された変数の系が因果的に十分であるかどうかについての予備知識がなくても、適切なサンプルデータから原理的に信頼できる因果推論を行うことができることを示すことである。

6.2 PCアルゴリズムと潜在的な 変数

PCアルゴリズムを少し修正することで、測定されていない変数が存在する場合でも、因果構造に関する正しい情報が得られるというのが自然な考え方である。*P'*が因果グラフに忠実な**V**上の分布で、Pは**Vに**適切に含まれる**O**上の*P'*の余白であるとする。すでに見たように、測定されていない共通の原因がある場合、PCアルゴリズムの出力は*A* <-> *B*の形の双方向エッジを含むことができる。*A*-*B*間の双方向エッジは、**Oに対して***A*および*B*を直接引き起こす測定されていない原因*C*があることを意味すると解釈できる。メタシンボルとして "*"を使い、矢印が持ちうる3種類のエンドマー

クのいずれかを表す：EM（空マーク）、">"、"o "です。

Modified PC Algorithm（修正PCアルゴリズム）：

A.)頂点集合**V**に完全無向グラフ*C*を形成する B．）

$n = 0$.

くり返す

くり返す

Adjacencies$(C,X)\setminus\{Y\}$のカーディナリティが*n*以上であるようなCで隣接する変数XとYの順序ペアを選択し、カーディナリティ*n*の**Adjacencies** $(C,X)\setminus\{Y\}$の部分集合Sと、もしXとYは、**S**が与えられたときに、*C*からエッジ*X* - *Y*を削除し、**S**を**Sepset**(X,Y)と**Sepset**(Y,X)に記録することでd-separatedとなる。

Adjacencies$(C,X)\setminus\{Y\}$のカーディナリティが*n*以上であるような隣接する変数XとYの順序付けられたペアの全てとカーディナリティ*n*の**Adjacencies** $(C,X)\setminus\{Y\}$の全ての部分集合がd-分離について検査されるまで。

$n = n + 1$である。

隣接する頂点*X*, *Y*の各順序ペアについて、**Adjacencies**$(C,X)\setminus\{Y\}$が*n*より小さいカーディナリティを持つまで。

C.) ステップB)の結果のグラフを*F*とする。*F*において XとYが隣接している場合、XとYの間の辺を*X* o-o *Y*として方向づける。

D.) ペア*X*, YとペアY, Zがそれぞれ*F*で隣接するが、ペア*X*, Zが*F*で隣接しないような頂点*X*, *Y*, *Z*の各トリプルについて、*Y*が**Sepset**(X,Z)にない場合に限り、*X* *-* *Y* *-* *Z*を*X* *-> *Y* <-* *Z*として方向づける。

E.) くり返す

A *-> *B*, *B* *-* *C*の場合、AとBは隣接しておらず、Bに矢印があるわけではありません。

B *-* *C*とすると、*B* *-* *C*を*B* -> *C*として方向付ける。

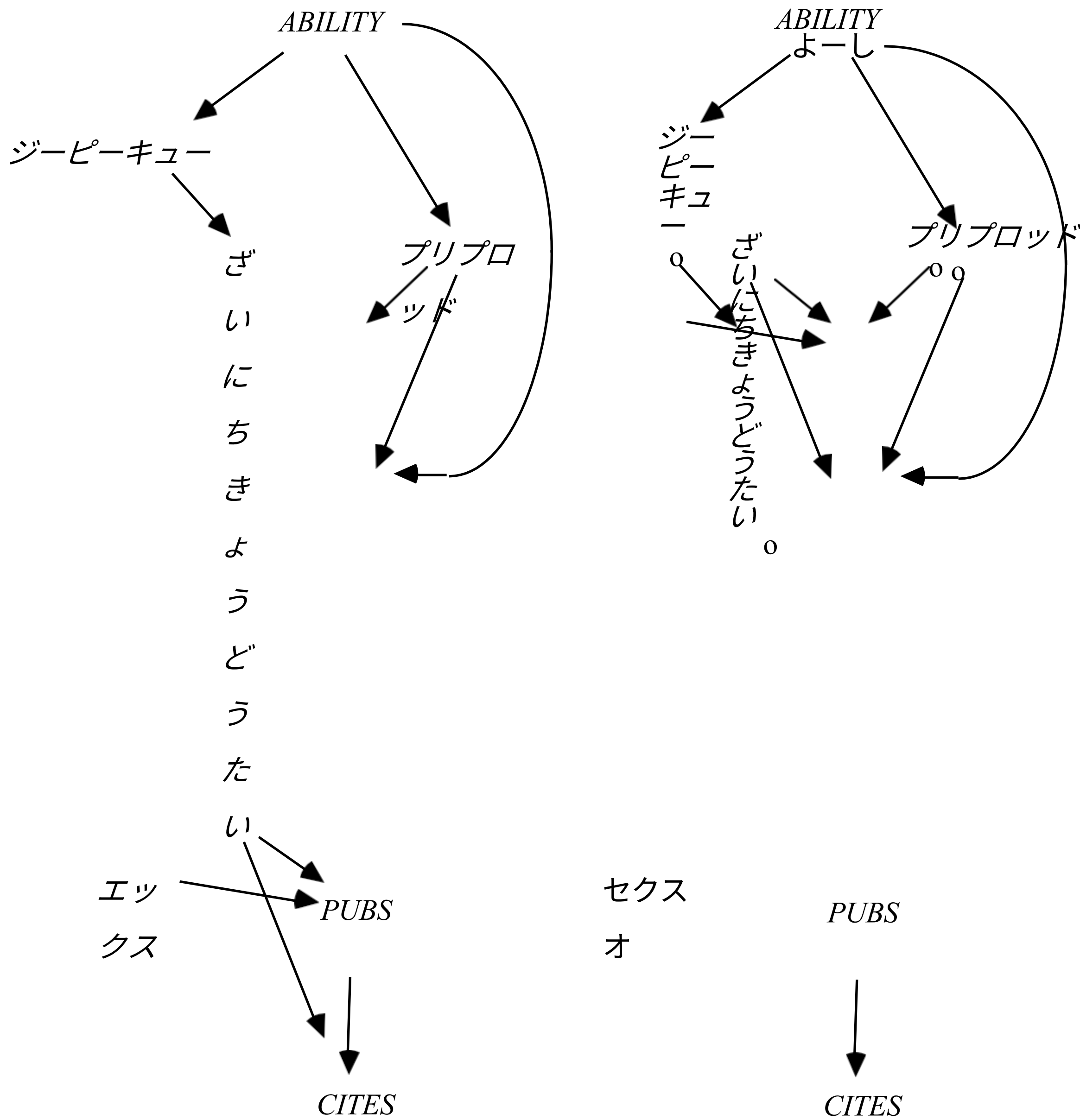
Aから*B*への有向パスがあり、Aと*B*の間にエッジがある場合、エッジを*A* *-> *B*のように方向付ける。

エッジの向きを変えることができなくなるまで。

(*X**-**Y*を*X**->*Y*のように方向付けるというのは、エッジの*X*端に同じマークを残し、エッジの*Y*端に矢じりをつけるという意味です)。

この修正を前章の例に適用した結果は、まったくもって理にかなっている。例えば、図1では、RodgersとMarantoのデータからPCアルゴリズムによって有意水準0.1で得られたモデルと、以下のようなモデルが示されています。

のグラフに忠実な分布から、修正PCアルゴリズムによってPC出力される。(いずれの場合も、変数の既知の時間順序が制約として課される)。



PC出力 **修正PC出力**

図1

修正PCアルゴリズムの出力は、例えばGPQと*ABILITYは*測定不能な共通原因によってつながっている可能性があるが、PUBSは共通原因によって妨げられることなく、

*CITESの*直接的原因であることを示している。一つの頂点が、隣接しない頂点と接続する2つ以上の辺に「o」記号を持つ場合、特別な制約が適用される。例えば、*ABILITYは*GPQと*PREPROD*へ*の*エッジを持ち、それぞれ*ABILITYの*端に "o "が付いている。そしてGPQと*PREPRODは互いに*隣接していない。その場合、2つの "o "の記号は*両方とも*矢じりにはなり得ない。なぜなら、もしそうであれば、GPQとPREPRODは*ABILITYを*条件として依存することになり、その代わりに修正パターンはそれらが独立であることを内包しているからである。

サンプリングされた分布が未測定の変数を持つグラフに忠実な分布のマージンである多くの場合（おそらくほとんどの実用的な場合）、PCアルゴリズムのこの単純な修正は、必要な統計的決定が正しく行われれば正しい答えを出します。

6.3 失敗談

残念ながら、このPCアルゴリズムの素直な修正は、一般的には正しくない。想像上の例でその理由を説明します。

*X*とYという2つの変数の測定されていない共通の原因を認識しなかったために起こる単純な間違いは、誰もが知っている。*X*が*Y*を引き起こすと考え、*X*を操作すれば*Y*の分布が変わると予測することがその誤りである。Xと*Y*の共通の原因を省くと、第3の変数*Z*が*Y*を直接引き起こすと誤って考えてしまうようなケースである。想像上のケースを考えてみよう：

> ある化学者が次のような問題を抱えている。化学者は、Aと*B*が結合して中間体*C*を経由して*D*を生成する別のメカニズムがあると考えている。試薬Aと*B*は容易に入手できるが、入手した試料には様々な量のDやその他の不純物が混じっている可能性がある。彼は不安定な中間体*C*の濃度を光度計で測定し、*D*の平衡濃度を標準的な方法で測定することができる。Aと*B*の濃度を操作することはできるが、*C*の濃度を操作する手段はない。
>
> 化学者は次のような実験計画を決定する。Aと*B*の濃度を10種類ずつ変えて、試薬を混合し、*C*の濃度をモニターし、*D*の平衡濃度を測定する試行を100回行う。もしAと*B*が*C*を介さずに*D*を生成する代替メカニズムがあるならば、*C*の濃度がすべて同じであるサンプルにおいて、Aと*D*、Bと*D*には正の相関があるはずだと化学者は考えている。

化学者は、*A*、*B*の平衡濃度と*D*の平衡濃度が、*C*を考慮すると相関していることを発見し、*A*、*B*が反応して直接*D*を生成する場合のように、別のメカニズムを確立したと発表する。

しかし、1年も経たないうちに彼の説は否定された。同じ試薬を使って、別の化学者が同様の実験を行ったが、マスキング剤が中間体*C*と反応し、*D*が生成されないようにした。最初の化学者の手順では何が問題だったのか?

*C*の操作の代わりに統計的制御を行うことで、化学者は、測定されていない変数による限界確率分布が、2つの変数の間に偽の直接的な関係があるように見せることができるという事実に反している。化学者が描いたメカニズム図はグラフG_1であり、これが観察された結果を生み出す一つの方法である。残念ながら、グラフG2のようなメカニズムで結果が出ることもあり、化学者のケースでは、試薬中の不純物(*F*)がCと*D*の両方の原因になっている:

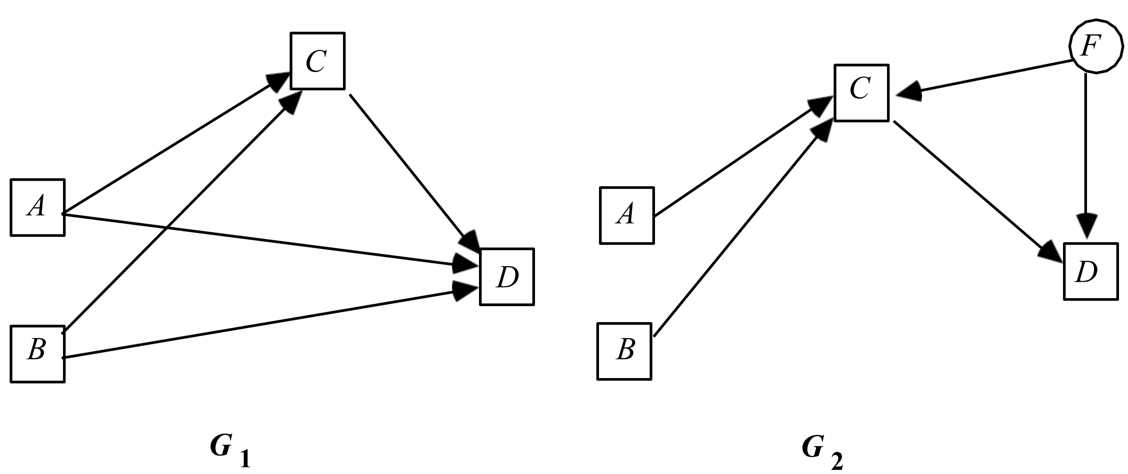


図2

一般的なポイントは、2つの測定変数Cと*D*に作用する理論変数*F*が、存在しない*A*-*D*間、*B*-*D*間の因果関係を示唆する統計的依存関係を作り出すことができるということ

である。忠実な分布の場合、SGSやPCアルゴリズムを用いると、G_2のような構造は、出力にAからDへの有向エッジを生成することになる。

同じ点をより解析的に見ると、次のようになる：変数**Vの**集合に対する有向非循環グラフ*Gにおい*て、AとDが*G*内で隣接する場合、**V**{*A*,*D*}*の*任意の部分集合が与えられると、Aと*Dは*d分離されない。したがって、因果的充足性の仮定の下で、Aと*Dが* *V*{**A**,*D*}*の*任意の部分集合を条件として独立である場合に限り、**Vに対して***Aが**D*の直接原因であるか、Dが*A*の直接原因であるかのいずれかが成り立つ。しかし、**Oが因果**的十分性でない場合、*O*{**A**,*D*}*の*あらゆる部分集合を条件としてAと*Dが*独立であれば、*Aが***Oに対して***D*の直接原因であるか、Dが**Oに対して***A*の直接原因であるか、Aと*D*の両方に共通の原因となる潜伏変数*Fが*存在するか、とは言えない。

これは図2の$_{G}2$で示され、**V** = {*A*,*B*,*C*,*D*,*F*}、**O** = {*A*,*C*,*D*} *である*。AとDは**O**{**A**,*D*}のどの部分集合でもd-separatedではないので、*Gに*忠実な分布のどの限界でも、Aと*Dは* **O**{*A*,*D*}のどの部分集合を条件としても独立ではなく、修正PCアルゴリズムではAとDの間に辺を残すことになる。しかし、*Aは***Oに対する***D*の直接的な原因ではなく、Dは**Oに対する***A*の直接的な原因ではなく、Aと*D*の潜在的な共通原因は存在しない。図2に示す有向無サイクルグラフ$_{G1}$は、*Aが**D*の直接的な原因であり、Aから*D*へ*Cを*通らない経路が存在するため、グラフ$_{G2}$と同様に{A，*C*，*D*}上のd-分離関係の集合を持つ。したがって、忠実な分布が与えられた場合、両者を条件付き独立関係だけで区別することはできない。

上述のPCアルゴリズムの単純な修正に関するさらなる根本的な問題は、PCアルゴリズムによって構築されたグラフに双方向エッジを認めると、Aと*Bが***Oの**ある部分集合でd分離される場合、**Adjacencies**(*A*)や**Adjacencies**(*B*)のある部分集合でd分離することがもはや当てはまらないという点である。図3 の グラフで、T_1、$_{T}2$が未測定であると仮定した場合を考える。

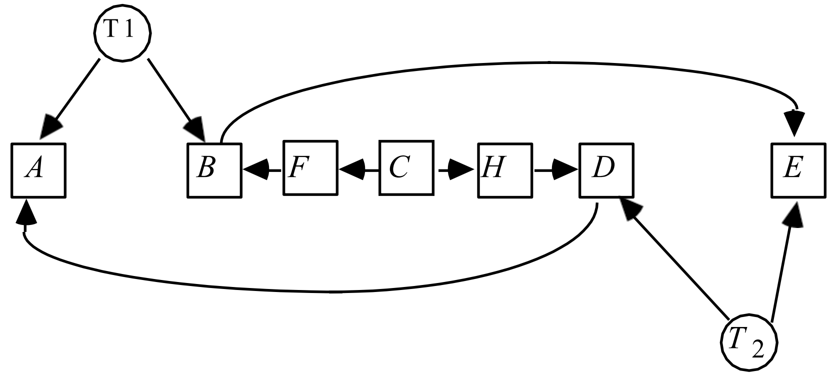


図3

測定された変数のうち、**Parents**(A) = {D} と **Parents**(E) = {B}ですが、AとEは{B}のどの部分集合でも{D}のどの部分集合でもd-separatedにはならず、それらをd-separateする唯一の集合はF、C、Hを含む集合です。修正PCアルゴリズムは、C、F、HがAまたはEに隣接していないことを正しく検出します。その後、Cを含む任意の部分集合でAとEがd-separatedであるかどうかをテストすることに失敗します。

{B,C,D}となり、AとEが誤って隣接したままになってしまう。つまり、アルゴリズムのある段階で構築されたグラフの局所的な特徴（すなわち隣接関係）のみを調べることによって、グラフからどのエッジを削除するかを決定することは不可能であるということである。同様に、PCアルゴリズムの出力に双指向性のエッジが認められると、アルゴリズムの所定段階で構築されたグラフの局所的特徴（すなわち、共通の端点を共有するエッジのペア）を調べることによって、エッジの向きに関するすべての情報を抽出することは不可能である。

このような問題があるため、完全な一般性のためには、PCアルゴリズムと出力の解釈に大きな変更を加えなければならない。我々は、真のグラフが疎であり、多くの双方向エッジが連鎖していない場合、大規模な変数セットで実行可能な、楽観的に高速因果推論（FCI）アルゴリズムと呼ぶ手順が存在することを示す。このアルゴリズムは、潜在変数が作用する可能性がある場合、測定された分布が真のグラフのマルコフ条件と忠実条件を満たす分布のマージンであると仮定して、因果構造に関する漸近的に正しい情報を与える。FCIアルゴリズムは、修正PCアルゴリズムの誤りを回避し、場合によってはより多くの情報を提供する。

例えば、図4の想像上の構造からボックス化された変数に対するマージナル分布がある場合、修正PCアルゴリズムは図5の第1図に示す正しい出力を与えるが、FCIアル

ゴリズムは図5の第2図に示す正しい、より情報量の多い結果を与える：

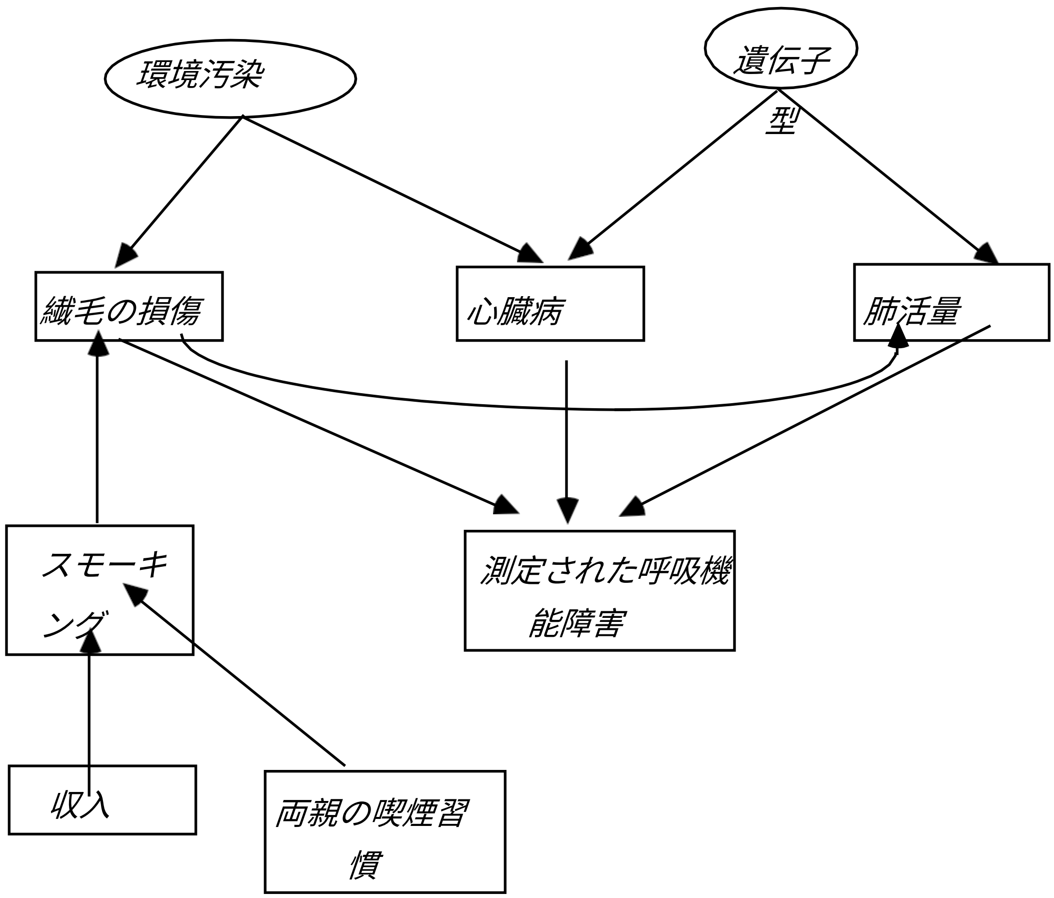


図4

図5では、両頭の矢印は、測定されていない共通原因の存在を示し、修正PCアルゴリズムと同様に、o->の形のエッジは、アルゴリズムが、エッジの一端の円が矢じりになるべきかどうかを決定できないことを示します。変数の集合{*Ciliadamage*, *Heartdisease*, *Lung capacity*, *Measured breathing dysfunction}*間の隣接関係は完全なグラフを形成するが、それでもFCIアルゴリズムによってエッジを完全に方向づけることができることに注目。

FCIアルゴリズムの導出には、様々な新しいグラフ概念とかなり複雑な理論が必要である。我々は、VermaとPearlの誘導パスと誘導パスグラフの概念を導入し、これらのオブジェクトが因果構造に関する情報を提供することを示す。そして、データから誘導パスグラフのクラスを推論するアルゴリズムについて考える。

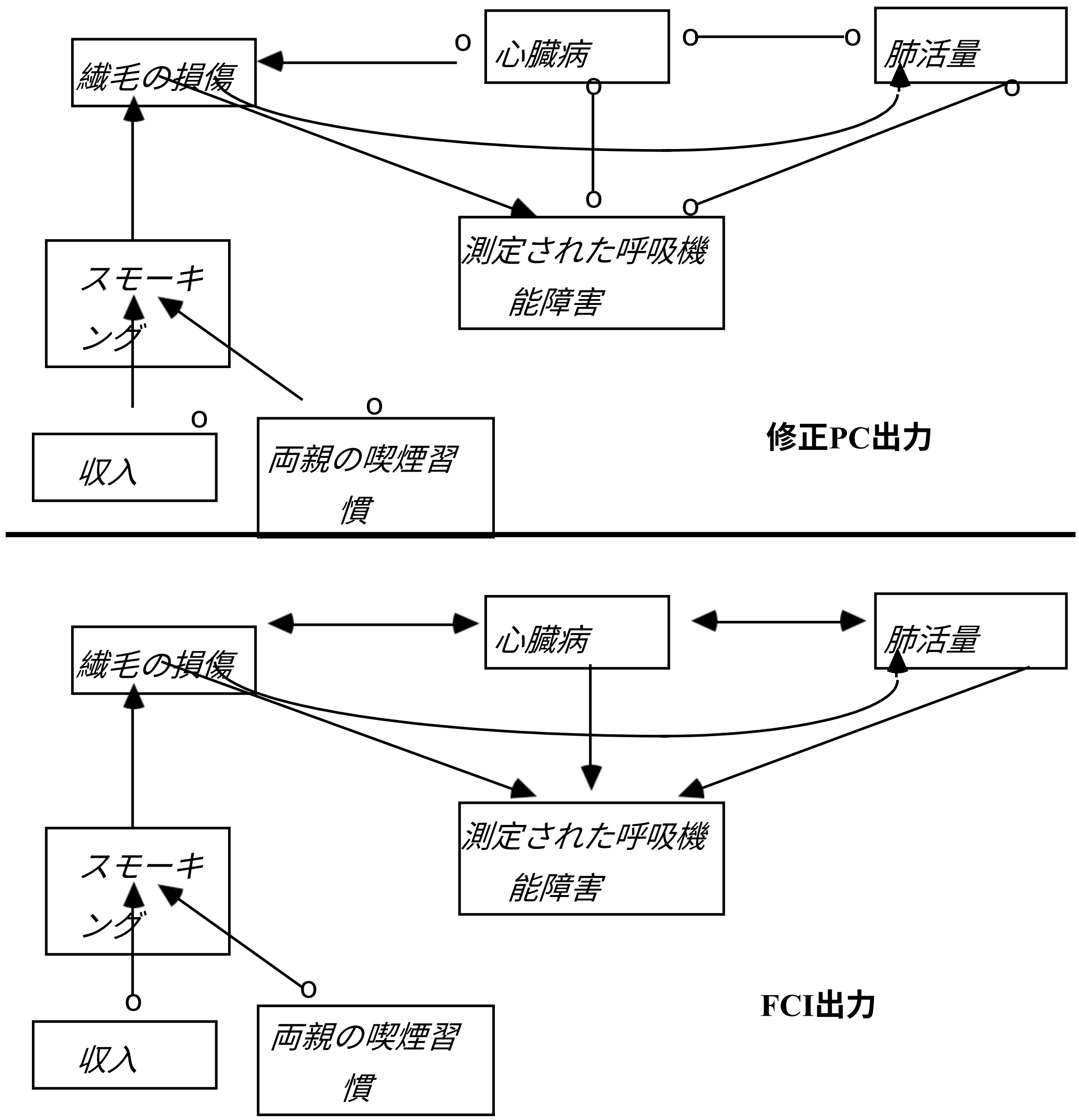


図5

6.4 パスの誘導

Verma and Pearl (1991)は、変数**V**の集合に対する有向無サイクルグラフGと、**V**の部分集合である**O**が与えられたとき、**O**の2変数がd-分離しない条件を特徴付けた。Gが変数の集合**V**上の有向無サイクルグラフ、Oが A と *B* を含む**V**の部分集合、*A*≠*B* とすると、A と *B* 間の無向パスUは、終点を除く *U* 上の**O**のすべてのメンバーが *U*

上のコライダーであり、U上のすべてのコライダーがAまたはBのいずれかの祖先

である場合にのみ**Oに対して誘導パスとなる**。

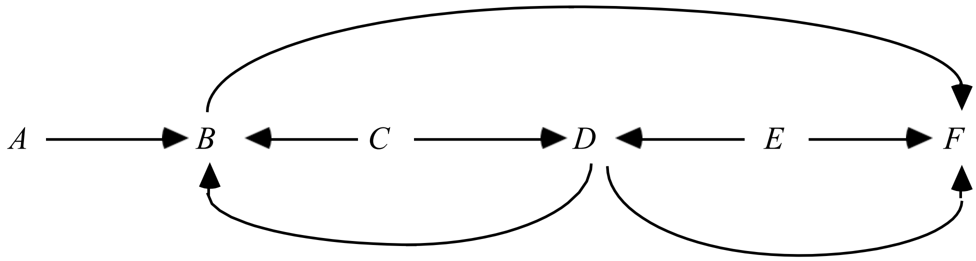


図6：グラフ$_{G}$3

例えば、グラフ$_{G}$*3においてパス*、パス*U* = <*A*,*B*,*C*,*D*,*E*,F> は、**O** = 上の誘導パスである。{同様に、*Uは***O** = *{*A,*B*,*F}*上の誘導パスである。しかし、Uは**O** = {A,B,F}上の誘導パスではない。

Cが**Oに**あるから*{A*,*B*,*C*,*D*,F*}*だが、Cは*Uの*コライダではない。

> **定理6.1**：Gが頂点集合**Vを**持つ有向非循環グラフで、**Oが**Aと*Bを*含む**Vの**部分集合であるとき、Aと*Bの*間に部分集合**O**上の誘導経路が存在する場合に限り、*O*{**A，**B*}の*任意の部分集合**Zによって**AとBはd分割されない。

定理6.1と、Uが**O** = *{A*,*B*,*D*,*F}*上の誘導経路であることから、AとFは*{B*,*D}のす*べての部分集合が与えられるとd-connectedであることがわかる。グラフ$_{G}$*3 では***O** = *{A*,*B*,*C*,*D*,*F}*上のAと*Fの*間に誘導経路が存在しないので、Aと*Fは*{*B*,*C*,*D}の*ある部分集合（この場合、*{*B,*C*,*D}*そのもの）があればd分離しているとなる。

6.5 パス グラフの誘導

V上のグラフ*Gにおける***Oに対する**誘導経路は、Verma and Pearl (1990b)に記述されて

いる（ただし名前はない）以下の構造で表すことができる。*G'*は、Oが*G*の頂点の部分集合であり、Aと*B*が**Oに**ある場合に限り、変数Aと*B*の間に*A*に矢尻を持つ辺が存在し、**Oに対して**Aと*B*の間に*A*に入る誘導経路が*G*に存在する場合に限り、**有向無サイクルグラフ*G*のO上の誘導経路グラフで**ある（第2章の記法を用いると誘導経路グラフのマークの集合は{>、EM}である）。誘導経路グラフでは、2種類のエッジがある：*A*→*B*は、*A*→*B*にかかるすべての誘導経路が、A→Bにかかるすべての誘導経路を含む。

AとBの間の**O**が*Aから出てBに入ること*、A <-> B*は***Oの**上にAに入ってBに入る誘導経路があることを意味する。この後者のエッジはAとBの共通原因が潜在している場合にのみ発生する。

図7〜図9は、**O** = {A,B,D,E,F}、**O** = {A,B,D,E,F}に対するG3の誘導パスグラフを表している。それぞれ{A,B,D,F}と**O** = {A,B,F}*とする。ただし、G3において*<B,D>は**O** = {A,B,D,E,F} 上のBとD*の間の*誘導経路でD*から外れているが*、**O** = {A,B,D,E,F} 上にD*に*入る別の誘導経路 <B,C,D> *があるた*め誘導経路グラフではBとD*の間の辺は*Dで矢尻を持つ。**O** = {A,B,D,E,F}*上の*誘導経路グラフにはAとFの間の辺はないが、**O** = {A,B,D,F}と**O** = {A,B,F} 上の誘導経路グラフには、AとFの間の辺がある。

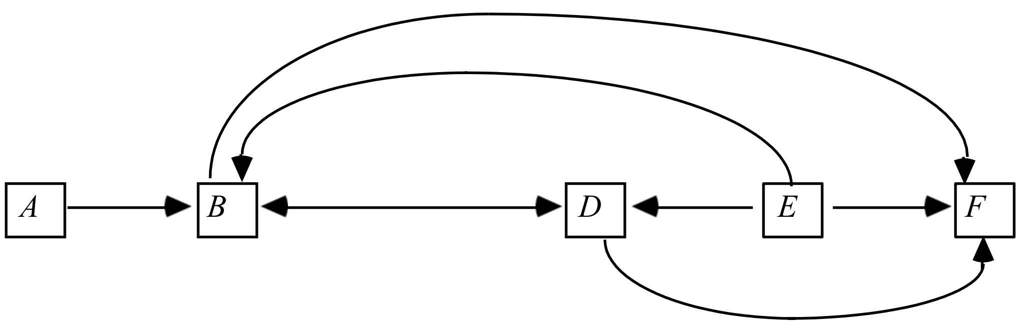


図7：*G* 3の{*A*, *B*, *D*, *E*, F}*にわたる*誘導パスグラフ

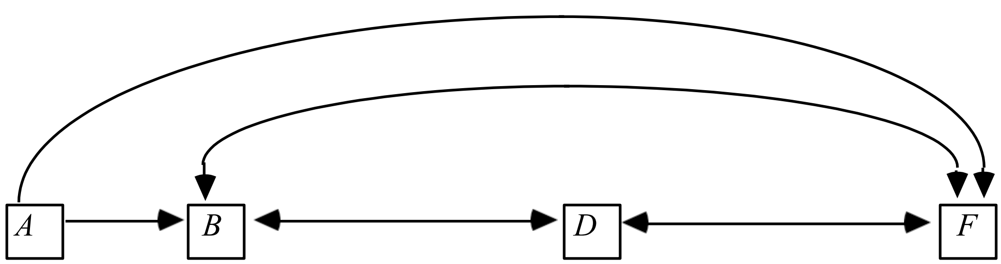


図8：*G* 3の{*A*, *B*, *D*, F}*にわたる*誘導パスグラフ

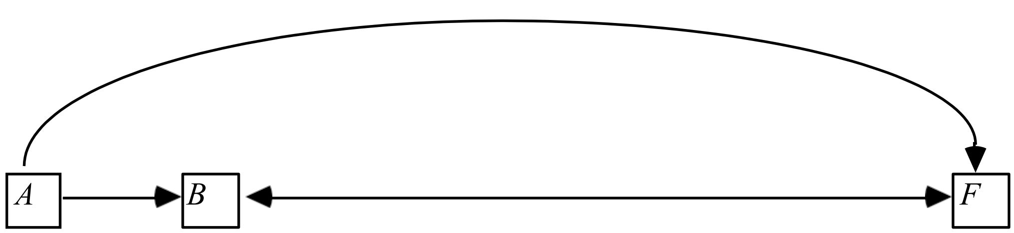


図9：G_3 Over {A, B, F}の誘導パスグラフ

有向パスで発生しうるエッジの種類が1つの矢尻を持つエッジのみであり、無向パスは1つの矢尻または2つの矢尻を持つエッジを含むことができれば、d-分離性の概念を誘導パスグラフに修正なしに拡張することができる。Gを有向無サイクルグラフ、G'を**O**上のGの誘導経路グラフ、**X**、**Y**、Sを**Oに**含まれる変数の不連続集合とすると、XとYはGの**S**が与えられたときのみ**d-分離可能である**場合、G'の**S**は**d-分離される**。

誘導経路グラフと有向無サイクルグラフのd-separability関係は、両頭の矢印が非常に重要な違いになっています。**O**上の有向非循環グラフにおいて、**O**{**A**,*B*}の任意の部分集合でAとBがd-separatedであれば、**Parents**(*A*)か**Parents**(B)でAとBがd-separatedとなる。これは誘導パスグラフでは成り立たない。例えば、**O** = {*A*,*B*,*C*,*D*,*E*,*F*,*H*} 上の図3の誘導パスグラフである誘導パスグラフG_4では、**Parents**(*A*) = {*D*}, **Parents**(*E*) = {*B*}ですが、Aと*E*は{*B*}の任意の部分集合、{*D*}の任意の部分集合ではd分離せず、Aと*E*をd分離する集合はすべて*C*, *H*, *F*を含みます。

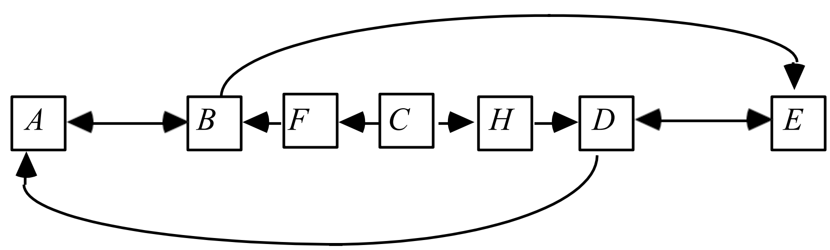


図10：誘導パスグラフG_4

しかし、誘導経路グラフの頂点集合には、d-separabilityに関する限り、有向無サイクルグラフの親集合とよく似た振る舞いをするものがある。

G'を**O**上の誘導パスグラフとし、$A \neq B$とすると、*V*?**D-SEP**(*A*,*B*)は、$A \neq V$であり、*U*上のすべての頂点が*A*または*B*の祖先であり、（終点を除いて）*U*上のコライダ

であるような*A-V*間の無向パス*U*がある場合に限るものとする。

定理6.2：**O**上の誘導経路グラフ*G'*において、AとBが**Oにある**とき、*A*が*B*の祖先でなく、AとBが隣接していなければ、**D-SEP**（*A*,*B*）が与えられてAとBはd分離していることになる。

したがって、AとBが**Oの** *すべての*部分集合を条件として依存するかどうかを決定しなくても、Aと*B*が誘導経路グラフで隣接するかどうかを決定することができる。

Oが因果的に十分な変数集合でない場合、**O**{**A**,*B*}*の*すべての部分集合を条件としてAと*Bが*依存する場合、Aと*Bの*間の誘導経路の存在を推論することはできるが、*Aが***Oに対して***Bの*直接原因であるか、*Bが***Oに対して***Aの*直接原因であるか、Aと*Bの*共通原因が潜在しているかを推論できない。それでも、次のレンマに示すように**O**上のAと*Bの*間の誘導経路の存在は、Aと*B*間の因果関係に関する情報を含む。

Lemma 6.1.4：GがV上の有向無尽グラフで、Oが**Vの**部分集合でAと*Bを*含み、GがAと*Bの*間の**O**上の誘導経路で*Aから外れて*おり、AとBが**Oにある**とき、*Gに*Aから*Bへの*有向経路が存在する。

OがVの部分集合であり、**O**上のAと*Bの*間に*Aから*外れる誘導経路があると判断できる場合、*Aが*Bの（おそらく間接的な）原因であると推論できることは、Lemma 6.1.4から成り立つ。したがって、**O**上の分布から**O**上の誘導経路グラフの特性を推論できる場合、測定できなかった変数があっても変数間の因果関係について推論できることになる。次節では、**O**上の分布から**O**上の誘導経路グラフの性質を推論するアルゴリズムについて説明する。

6.6 部分指向性誘導パス グラフ

部分配向誘導パスグラフは、いくつかの種類のエッジを含むことができる：*A* -> *B*, *A* o-> EM（空マーク）、">"、"o "の3種類の端のいずれかを表すメタシンボルとして "*" を使用し、"*"シンボル自体は部分配向誘導パスグラフには現れない。(また、誘導経路グラフで発生しうる2種類の端(EMまたは">")を表すメタシンボルとして "*"を使用

する)。

有向無サイクルグラフGに対して、π**O上の**誘導パスグラフG'を持つ部分配向誘導パスグラフは、G'の隣接関係、およびG'と同じd-連結関係を持つすべての誘導パスグラフに共通するG'の辺の一部の配向を表すことを目的としています。G'を**O**上の誘導パスグラフとすると、**Equiv**(G')はGと同じd-連結関係を持つ同じ頂点上の誘導パスグラフの集合であり、**Equiv**(G')のすべての誘導パスグラフは同じ隣接関係の集合を共有する。以下の定義を用いる：

π **は、以下の**場合に限り、**O上の誘導経路グラフG'を持つ有向無サイクルグラフGの部分配向誘導経路グラフである。**

(i). の*A*-*B*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→*B*、*B*→*A*、*A* o-> *B*、*B* o-> *A*、*A* o-o *B*、または*A* <-> *B*;

(ii). π と*G'*は同じ頂点を持つ;

(iii). π と*G'*は同じ隣接関係を持つ;

(iv). *A* o-> *B* が in πであるならば、**Equiv**(*G'*) のすべての誘導経路グラフ *X* において、*A* -> *B* のいずれかである。

または*A* <-> Bは*X*にある;

(v). にある場合、π*A*→Bは**Equiv**(*G'*)のすべての誘導パスグラフにある;

(vi). において、*A* <u>*-* *B*</u> *-π* Cが存在する場合、**Equiv**(*G'*)の任意の誘導パスグラフ

において、Aと*B*、Bと*C*間のエッジが*B*で衝突することはない;

(vii). にある場合、*A*π<-> *B* は **Equiv**(*G'*) のすべての誘導パスグラフにある;

(viii). にある場合、π**Equiv**(*G'*)のすべての誘導パスグラフ*X*において、*A*→*B*のいずれかが成立する。

B、*A* <-> *B*、または*A* <- Bは*X*にある。

(厳密には、部分配向誘導パスグラフは、下線によって追加された余分な構造のため、我々が定義したようなグラフではない。)エッジ*A* *-o *B*は、**Equiv**(*G'*)のどの部分集合においても、Aと*B*の間のエッジが*B*に入ることもBから出ることも拘束しないことに注意する。πGの部分配向誘導経路グラフの隣接関係は、*O*{**A**,*B*}のあらゆる部分集合が与えられたときに、Aと*B*がd-connectedである場合にのみ、Aと*B*を隣接させることによって構築することができる。π

各辺*A* *-* *B*を*A* o-o *B*として配向させるだけである。もちろん、*G*に対するこの特定の部分配向誘導パスグラフは、π*G'*の配向が**Equiv**（*G'*）のすべての誘導パスグラフに

共通する特徴について非常に示唆に富んでいない。例えば、図11は、測定された呼吸機能障害の原因に関する想像上のグラフを再び示している。図12は、**O**＝｛繊毛*損傷*、*喫煙*、*心臓病*、*肺活量*、測定さ*れた呼吸機能障害*、*収入*、*両親の喫煙習慣*｝上のグラフ$_{G}$5の不案内な部分配向誘導パスグラフを示す。

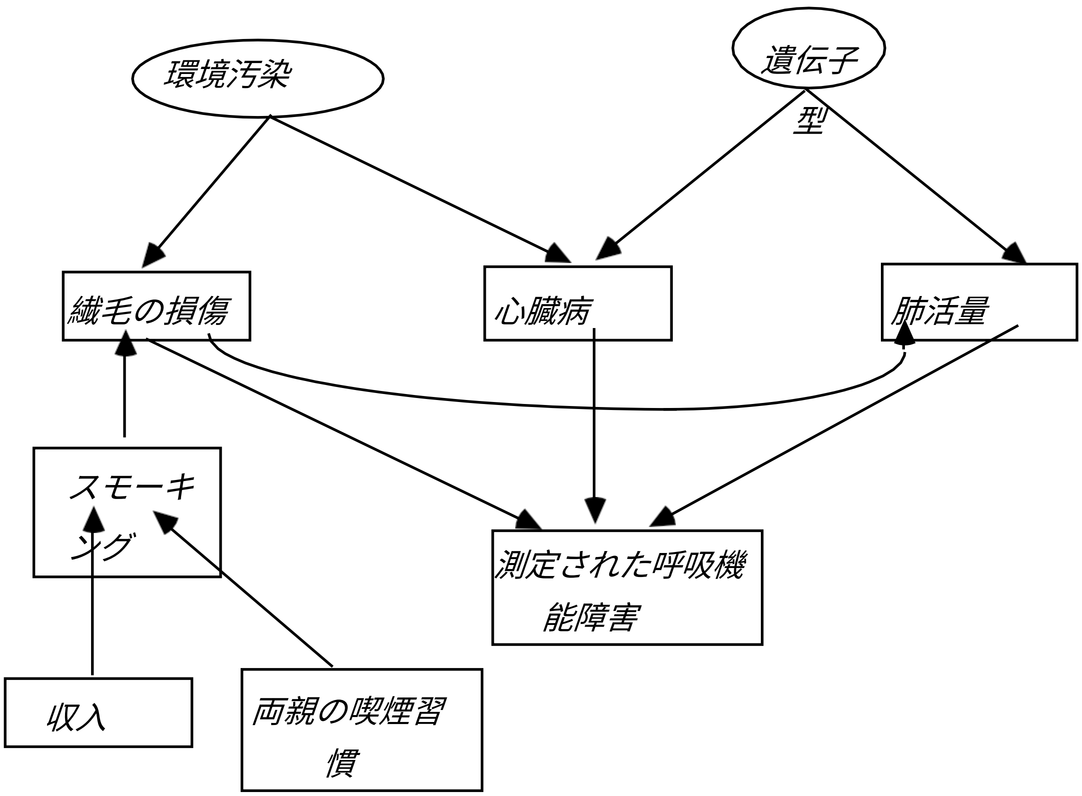


図11：グラフG_5

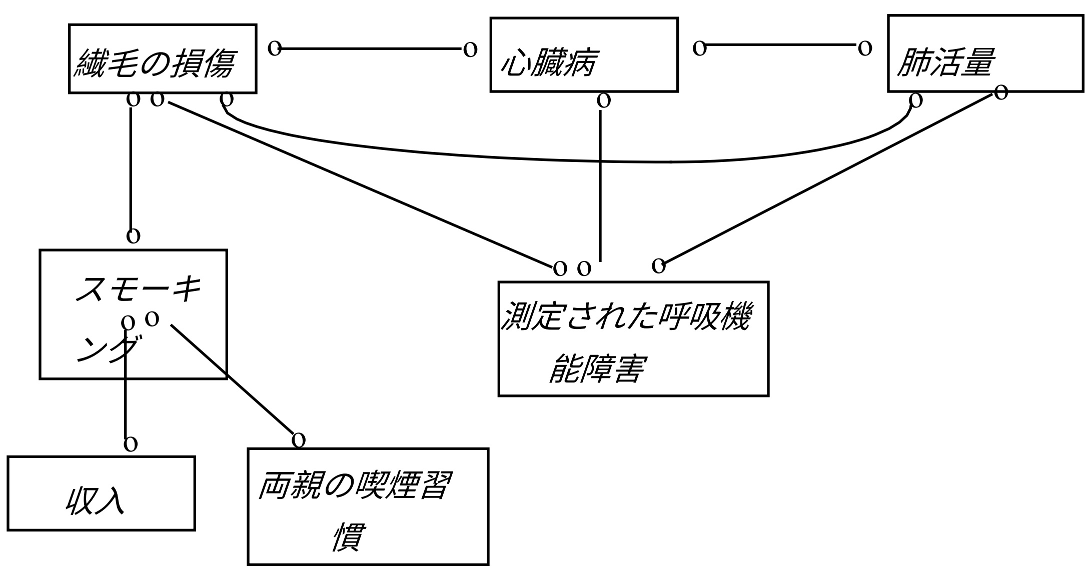


図12：G_5のO以上の非情報的な部分指向性誘導パスグラフ

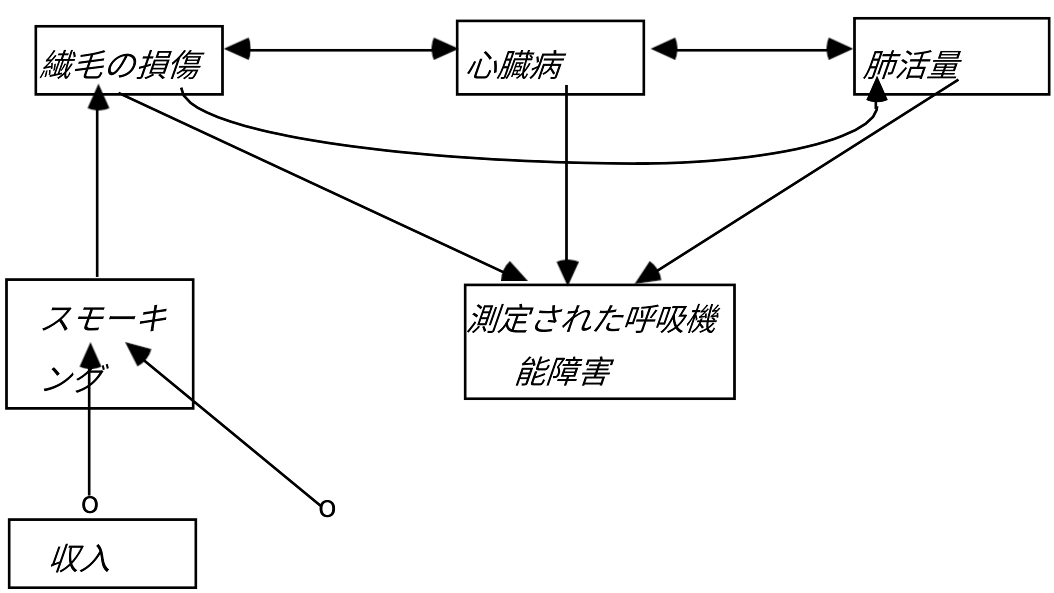
*B*が*U*の終点であるか、*U*が*A* <- *B*のいずれかの部分経路を含むような頂点*A*および*C*が存在する場合に限り、*B*が無向経路*U*上の**確定的な非結合員**であると言おう。

- *C*、*A* *-* *B* -> *C*、または *A* *-* *B* *-* *C*. *G* の**最大情報量の部分配向誘導パスグラフ**
π
で、誘導パスグラフ *G*'、

(i) エッジ*A* *-o *B*は、AとBの間のエッジが、**Equiv**(*G*')のあるメンバーでは*B*に入り、**Equiv**(*G*'.)の他のメンバーでは*Bから出る*場合にのみ出現する、および

(ii) *A*-*B*間、*B*-*C*間のすべての辺の組について、その辺が*B*で衝突するか、*B*で明確に非衝突となるかのいずれかである。

このような*G*の最大情報量の部分配向誘導経路グラフは、*G*'と同じ隣接関係を持つあらゆる可能な誘導経路グラフを構築し、*G*'と同じd-connection関係を持たないものを捨て、どの配向特徴が**Equiv**（*G*'）のすべてのメンバーに共通しているかを追跡するという単純だが非効率なアルゴリズムで配向させることができる。もちろん、これは完全に計算上実現不可能である。図13は、**O**＝｛*Ciliadamage*｝上のグラフG5の最大配向の部分配向誘導パスグラフを示す、
喫煙、*心臓病*、*肺活量*、*測定された呼吸機能障害*、*収入*、*両親の有無*
喫煙習慣｝です。



両親の喫煙習慣

図13：最大情報量の部分指向性誘導経路のグラフ

G_5 Over O

我々の目標は、有向無サイクルグラフG*に対して*、計算可能な限り多くの方位情報を含む部分方位誘導パスグラフを構築するアルゴリズムを述べることである。我々が提案するアルゴリズムは、2つの主要な部分に分かれている。まず

部分的に配向された誘導経路グラフの隣接関係が決定される。そして、可能な限り辺の向きを揃える。

6.7 Latent Common Causesを用いた因果推論のためのアルゴリズム

アルゴリズムを説明するために、もう少し定義が必要である。部分配向誘導経路グラフにおいて π .

(i). Aは*B*の**親であり**、*A* -> *B* in π .である場合にのみ、*B*の親となる。

(ii). Bは，経路<A,*B*,C>に沿った**コライダーであり**，*A* *-> *B* <-* *C* in π .のときだけである.

(iii). Bと*A*の間のエッジが*A*に**入る**のは、*A* <-* *B* in π .のときだけである。

(iv). Bと*A*の間のエッジは、*A* -> *B* in π の場合にのみ、*A*から**外れる。**

(v). 部分配向誘導パスグラフにおいて π'において、*U*は*B*の**明確な判別パス**である。 *U* が*B* を含む *X* と *Y* の間の無向パスであり，$B \neq X$，$B \neq Y$で，*U* 上の *B* と終点を除くすべての頂点が *U* 上の衝突者または明確な非衝突者である場合にのみ *B* の確定**識別パスとなる**.

(i) *U*上でVと*V'*が隣接し、*V'*が*U*上でVと*B*の間にある場合、*U*上で*V* *-> *V'*となります、

(ii) *V*が*U*上のXと*B*の間にあり、Vが*U*上のコライダーである場合、*V* -> *Y* in π, else *V* <-* *Y* において

(iii) *V*が*U*上のYと*B*の間にあり、Vが*U*上のコライダーである場合、*V* -> *X* in π, else *V* <-*. *X* in π

(iv) π において、XとYは隣接しない。

図14は、確定識別パスの概念を説明する図である。

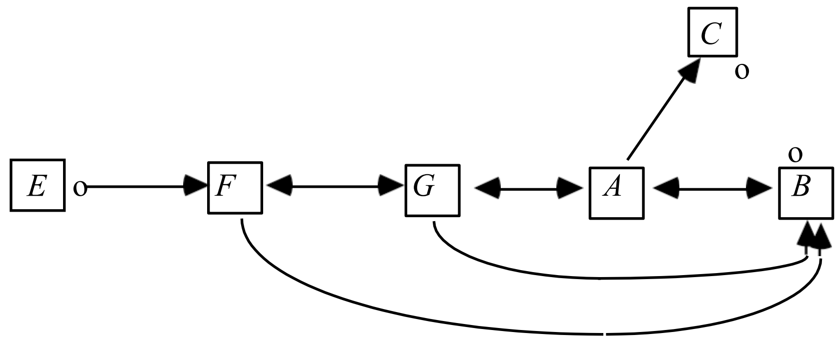


図14：＜E，*F*，*G*，*A*，*C*，*B*＞は*C*の確定識別パスである。

実際には、因果推論アルゴリズムと高速因果推論アルゴリズム（このセクションで後述）は、共分散行列またはセルカウントを入力として受け取る。アルゴリズムが必要とするd-separationの事実は、条件付き独立性の検定（離散の場合）または消失する部分相関の検定（線形、連続の場合）を実行する手順である。(Pがグラフ*G*に忠実な離散分布である場合、*A*および*B*が**Cを**与えられたときに条件付き独立である場合にのみ、*A*および*B*は変数**Cの**集合を与えられてd-分離され、Pがグラフ*G*に忠実な線形分布である場合、*A*およびBは$_{AB.\mathbf{C}} = 0$の場合にのみ**Cを与えられて**d-分離されるということを覚えておこう.）どちらのアルゴリズムも、ある有向無周期グラフ*G*の部分配向誘導パスグラフを構成し、Gには測定変数と未測定変数の両方が含まれる。

因果推論アルゴリズム[1]

A). 頂点集合**V**に完全無向グラフ*Q*を形成する。

B). **V**の任意の部分集合**S**が与えられたとき、AとBがd個に分離されている場合、*A*間のエッジを削除する。

と*B*に、**S**を**Sepset**(*A*,*B*)と**Sepset**(*B*,A)に記録します。

C). ステップB)の結果のグラフをFとする。各辺をo-oとする。頂点*A*,*B*,*C*の各

トリプルについて、ペア*A*,Bとペア*B*,*C*がそれぞれ*F*で隣接するがペア*A*,*C*は*F*

で隣接しない場合、*B*が**Sepset**(*A*,*C*)にない場合にのみ*A* *-> *B* <-* *C*として、*A*

- *B* *-* Cとして配向し、かつ*B*が**Sepset**(*A*,*C*)にある場合にのみ*A* <u>*-* *B*</u> *-* *C*

として、*A**-* *B**-* *-* *C*に配向する。

D). くり返す

Aから*B*への有向パスがあり、エッジ*A* *-* *B*がある場合、*A* *-* *B*を次のように方向付ける。

A *-> *B* です、

else もしBが<A,*B*,C>に沿ったコライダーで、Bが*D*に隣接し、*A*および

Cが*D*からd-connectedでない場合、*B* <-* *D*として*B* *-* *D*を方向づける

、

elseは，Uが*M* inのためにAと*B*の間の明確な差別的な経路であり，Pと

Rは*U*上で*M*に隣接し、*P* - *M* - Rは三角形である。

Mが**Sepset**(*A*,*B*)にある場合、Mはサブパス*P*上のノンコライダーとしてマークされる <u>*-*。</u>
<u>*M*</u> *-* *R*

else *P* *-* *M* *-* *R* は *P* *-> *M* <- * *R* として方

向づけられる。 else *P* <u>*-> *M*</u> *-* *R* ならば *P* *->

M -> R として方向づける。

エッジの向きを変えることができなくなるまで。

CIやFCIのアルゴリズムが、*G*に線形に忠実な分布の**O**上のマージンからの共分散行列や、*G*に忠実な分布の**O**上のマージンからのセルカウントを入力として使う場合、入力は*G*に忠実な**O**上のデータであると言うことにする。

定理6.3： CIアルゴリズムの入力が*G*に忠実な**O**上のデータである場合、出力は**O**上の*G*の部分配向誘導パスグラフとなる。

図11のグラフに忠実な**O**={繊毛*損傷*、*喫煙*、*心臓病*、*肺活量*、*呼吸機能障害*、*収入*、*両親の喫煙習慣*}のデータを入力した場合

1CIアルゴリズムの原型の誤りを指摘してくれたThomas Verma (personal communication)に感謝する。

をCIアルゴリズムに変換すると、図13に示す**O**上の最大情報量の部分配向誘導パスグラフが出力される。

残念ながら、Causal Inference (CI) アルゴリズムは、隣接関係の構築方法から、多数の変数に対して実用的ではありません。Aと*B*が**O**{*A*,*B*}のある部分集合でd-分離している場合に限り、完全グラフからAと*B*間の辺を削除することは理論的に正しいが、これは2つの理由から非現実的である。第一に、Aと*B*の条件付き独立性を検定するO{*A*,*B*}の部分集合が多すぎる。第二に、離散分布の場合、サンプルサイズが膨大でない限り、他の変数の大きな集合を条件とする2変数の独立性を信頼できる検定がない。

Xと*Y*のような与えられた頂点のペアが誘導経路グラフで隣接していないと判断するためには、**O**{**X** , *Y* }のある部分集合が与えられたときに、XとYがd-separatedであることを見つけなければならない。もちろん、Xと*Y*が*誘導経路グラフで隣接し*ていれば、O{**X** ,*Y* }のすべての部分集合が与えられれば、d-connectedである。実際にO{*X* ,*Y* }の全ての部分集合を調べることなく、Xと*Y*が**O**{**X** ,*Y* }の全ての部分集合でd-connectedであることを決定できるようにしたい。

原因十分集合**V**上の有向非循環グラフにおいて、PCアルゴリズムを使用することで、実行するd-separationテストの順序と数を減らすことができる。これは、XとYが**V**{**X** ,*Y* }の任意の部分集合によってd-separationされる場合、それらは**Parents**(*X*) または **Parents**(*Y*) によって d-separation されるという事実である。PC アルゴリズムがグラフを構築している間は、どの変数が **Parents**(*X*) や **Parents**(*Y*) に含まれるかはわからないが、アルゴリズムが進むにつれて、いくつかの変数は *X* や *Y* に確実に隣接しないので、**Parents**(*X*) や **Parents**(*Y*) に含まれないと判断できるようになる．これ

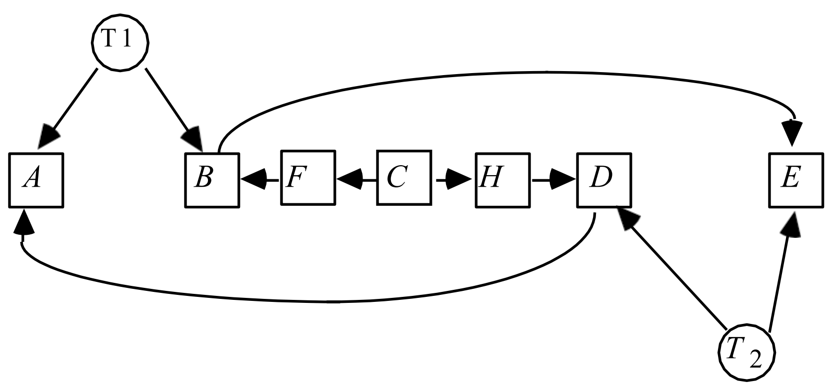
により、PCアルゴリズムが実行するd-分離テストの回数と順序が（SGSアルゴリズムに比べて）少なくなります。

一方、O上の誘導パスグラフは、**O**{**X** ,*Y* }のある部分集合でXと*Yが*d-separatedであれば、**Parents**(*X*)か**Parents**(*Y*)でXと*Yが*d-separatedであるということはない。しかし、XとYが*O* ,*Y* }の*ある*部分集合でd-separatedであるならば、Xと*Yは***D-Sep**(*X*)か与えられた**D-Sep**(*Y*)でd-separatedである。もし、ある変数*Vが*D**-Sep**(*X*)にも**D-Sep**(*Y)にも*ないことがわかっていれば、XとYが*Vを*含む集合によってd-separatedかどうかを調べる必要はありません。もう一度言いますが、グラフを構築するまで、どの変数が**D-Sep**(*X*)や**D- Sep**(*Y)* にあるかはわかりません。しかし、アルゴリズムが進むにつれて、ある変数がD-**Sep**(*X)*または**D-Sep**(*Y)*に*ない*ことを判断できるアルゴリズムがあるのです。

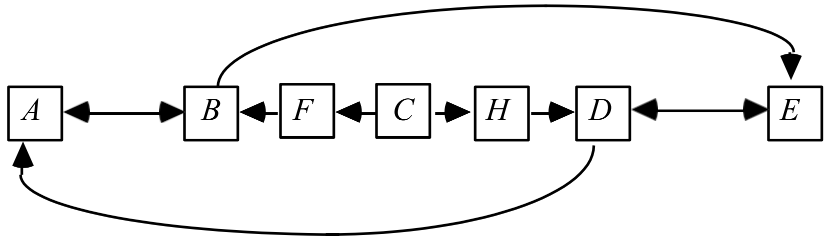
図3の有向無サイクルグラフをGとする（以下、図15で再現）。**O**＝{*A*,*B*,*C*,*D*,*E*,*F*,*H*}上の*Gの*誘導パスグラフを*G'*とする。*A*およびEは、任意の

は、*A*に隣接する変数または*D*に隣接する変数の部分集合（いずれの場合も{*B*,*D*}）である。Aと*E*の誘導経路グラフにおいてAとEは隣接していないので、*O*{**A**,*E*}のある部分集合が与えられると、それらはd-separatedである。したがって、**D-Sep**(*A*,*E*) ({*B*,*D*,*F*}に等しい)か**D-Sep**(*E*,*A*)) ({*B*,*D*,*H*}に等しい)によってそれらはd-分離される。(この場合、Aと*E*は*D*-**Sep**(*A*,*E*)と**D-Sep**(*E*,*A*)の両方によってd分離されます)。問題は、*O*{**A**,*E*}のすべての部分集合でAと*E*が*d*-separatedかどうかをテストせずに、{B,*D*,*H*}または{B,*D*,*F*} でAと*E*がd-separatedかどうかをテストすることをどうやって知ることができるか?

変数*V*が*G*'の**D-Sep**(*A*,*E*)にあるのは、*V*≠Aで、*A*-*V*間に終点を除くすべての頂点がコライダーであり、各頂点が*A*または*E*の祖先である無向パスがある場合に限る。変数*V*がこうしたパスにないことを判定する方法を見つけることができれば、*A*およびEが、*V*を含む任意のセットでd分割されたかどうかをテストしなくてもよい（もちろん*V*が**D-Sep**（*E*,*A*）になければ）。アルゴリズムの任意の段階において、これまでに構築されたグラフを .



グラフ*G*



パスグラフ*G'*を誘導する

図15

FCIアルゴリズムは、完全なグラフからどのエッジを削除するかを3つのステージでπ決定する。最初の段階は、PCアルゴリズムの最初の段階と同じである。完全な無向グラフを初期化し、XとYの間の辺がπ、XまたはYに隣接する頂点の部分集合が.FCIでd分割されている場合、その辺を削除する。これにより、誘導経路グラフにない辺の多くは除去されるが、おそらく全てではないだろう。この操作を図15のグラフに忠実なデータに対して行うと、図16のようなグラフになる。

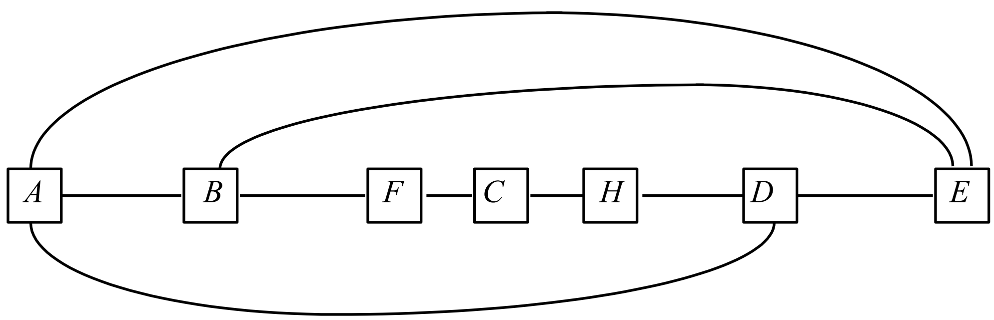


図16

AはFまたはHまたはCに隣接せず、EはFまたはHまたはCに隣接しないと正しく判断したアルゴリズムは、AとEがF、HまたはCを含む変数の部分集合によってd個分離されているかどうかをテストしなかったので、手順のこの段階でAとEはまだ隣接していることに注意してください。

次に、PCアルゴリズムと同様に、衝突の有無を判断して辺を方向付ける。この段階のアルゴリズムのグラフを図17に示す。

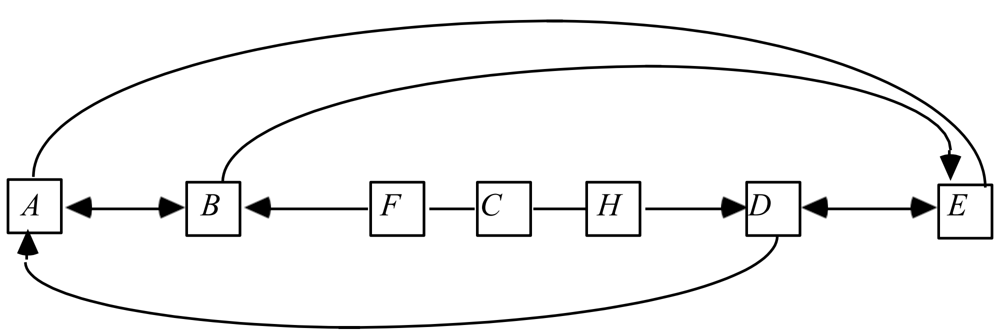
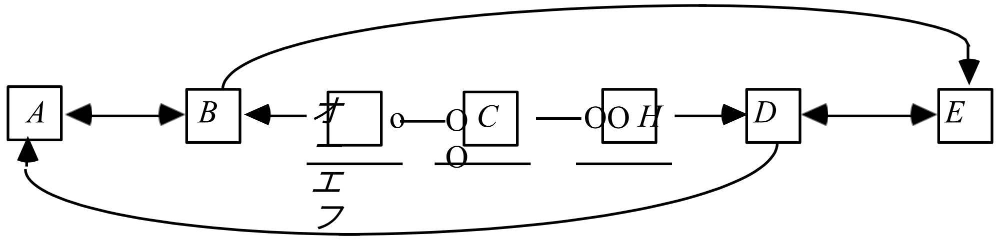


図17

図17は、基本的に図15のグラフに忠実なデータを与えられたPCアルゴリズムが、ステップA)、B)、C)を行った後に構築するグラフである。

これで、いくつかの頂点が**D-Sep**(*A*,*E*)にも**D-Sep**(*E*,*A*)にもないことを決定できる。正しい隣接関係を見つけるために、これらの頂点を含む*O*{**A**,*E*}の任意の部分集合で*A*および*E*がd分離しているかどうかをいちいちテストする必要はない。アルゴリズムのこの段階では、頂点*V*が*G'*の**D-Sep**(*A*,*E*)にあるための必要条件は、*A*-*V*間に無向パス*U*が存在し、終点を除く各頂点がコライダーであるか、三角形の中にあるためその向きが隠されているかのいずれかである。したがって、CとHは**D-Sep**(*A*,*E*)に、CとFは**D-Sep**(*E*,*A*)に間違いなく存在しない。**Possible-D-Sep**(*A*,*E*)に置く*G'*では、間違いなく決定していない頂点はすべて**D-Sep**(*A*,*E*)になく、同様に**Possible-D-Sep**(*E*,*A*)にもない。この場合、**Possible-D-Sep**(*A*,*E*)は{*B*,*F*,*D*}、**Possible-D-Sep**(*E*,*A*)は{*B*,*D*,*H*}です。ここで、もしAと*E*が*O*{**A**,E}の任意の部分集合でd-separatedならば、**Possible-D-Sep**(*A*,*E*)のある部分集合か**Possible-D-Sep**(*E*,*A*)のある部分集合でd-separatedであることがわかる。この場合、Aと*E*は*Possible-D*-**Sep**(*A*,*E*)の部分集合（この場合は集合全体）でd-separatedであることがわかるので、Aと*E*の間の辺を削除する。

正しい隣接関係のセットが得られたら、すべての辺の向きを変え、因果推論アルゴリズムで行ったのと全く同じように、辺の向きを変えながら進めていきます。その結果、図18に示すような出力が得られました。



FCI出力

図18

与えられた部分的に構成された部分配向誘導パスグラフに対して π について、

π

Possible-D-SEP(*A*,*B*)は次のように定義される：*A*≠*B* のとき、*V*が**Possible-D-**

Sep(*A*,*B*)に入るのは、*V*が

≠*A*であり、*A*-*V*間に無向パス*U*が存在する。 π のA-V間の無向パスUが存在し、すべてのサブパスに対して

*U*の<*X* ,*Y* ,*Z*>は、*Y*がサブパス上のコライダーであるか、Yが確定的なノンコライダーでないかのどちらかであり、かつ、オン

Uで、*X*、*Y*、Zが三角形を形成している。π

この**Possible-D-Sep**(*A*,*E*)の定義を用いると、**Possible-D-Sep**(*A*,*E*)に含まれないすべての頂点は π *G*'の**D-Sep**(*A*,*E*)にも含まれないことを証明することができます。ただし、Possible**-D-Sep**(*A*,*E*)に含まれるメンバーの中には、G'のD-Sep(A,E)に含まれないものがあることから、そのことを判断できる可能性がある。

をG'の**D-Sep**(A,E)の中に入れる。**Possible-D-Sep**(A,E)のサイズを小さくすること（アルゴリズムが実行するd分離性のテストの数と順序が減るように）と、セットのサイズを小さくするために必要な余分な作業を行い、それでもG'の**D-Sep**（A,E）のスーパーセットであることを保証するというトレードオフが明らかに存在する。最適なバランスが何であるかはわからない。Gが疎である場合（すなわち、各頂点がG内の多数の他の頂点に隣接していない場合）、アルゴリズムは、多数の変数を含む任意の**Cについて、**AとBが**Cを与えられて**d分離されるかどうかを決定する必要はない。

高速因果推論アルゴリズム

A). 頂点集合**Vに**完全無向グラフQを形成する。

B). $n = 0$.

くり返す

くり返す

Adjacencies(Q,X) \\{Y}のカーディナリティがn以上であるようなQで隣接する変数XとYの順序ペアを選択し、カーディナリティnの**Adjacencies** (Q,X) \\{Y}の部分集合**Sを与え、**XとYが**Sから**d分離する場合QからXとY間の端を削除し**Sepset** (X ,Y), **Sepset** (Y ,X) に記録せよ。

Adjacencies(Q,X) \\{Y}のカーディナリティがn以上であるような隣接する変数XとYの順序付き変数ペアのすべてと、カーディナリティnの**Adjacencies**(Q,X) \\{Y}のすべての部分集合**Sが**d分離をテストされるまで;

n＝n＋1である;

隣接する頂点X, Yの各順序ペアについて、**Adjacencies**(Q,X) \\{Y}がnより小さいカーディナリティを持つまで。

C). ステップB)の結果の無向グラフをFとする。各辺をo-oとする。頂点A,B,Cのうち、A,BのペアとB,CのペアがそれぞれFで隣接するが、A,CのペアはFで隣接しないような3つの頂点について、A *-* B *-* CをA *-> B <-* Cとして方向づけるのは、Bが**Sepset**(A,C)にない場合のみとします。

D). Fにおいて隣接する各変数A、Bの組について、Fにおける**Possible-D-SEP**(A,B)□{A,B} の任意の部分集合Sまたは**Possible-D-SEP**(B,A)□{A,B} の任意の部分集合**S**でA、Bがd分割されていればA、B間の辺を取り除いて**S** **Sepset**(A,B) 、 **Sepset**(B,A) に記録せよ.

そして、このアルゴリズムは、任意の変数XとYのペア間のエッジをX o-o Yとして再配列し、因果推論アルゴリズムのステップC）とD）と同じ方法でエッジの再配列を進める。

定理6.4：FCIアルゴリズムの入力が*G*に忠実な**O**上のデータである場合、出力は**O**上の*G*の部分配向誘導経路グラフとなる。

高速因果推論アルゴリズム(FCI)は、*G*に忠実な分布の測定変数に対するマージンから正しい統計的決定を与えると、常にグラフ*G*の部分指向誘導パスグラフを生成する。このアルゴリズムが完全かどうか、すなわち、すべてのケースで最大情報量の部分指向誘導パスグラフが生成されるかどうかはわからない。

CIアルゴリズムと同様に、FCIアルゴリズムの入力が図11のグラフに忠実なデータである場合、出力は図13の最大情報量の部分配向誘導経路グラフとなる。

O上の同じFCI部分配向誘導パスグラフを持つ2つの有向非周期グラフGと*G'*は、**Oの**メンバーだけを含む同じd-connection関係を持つ。

公証6.4.1Gが**V**上の有向アシリックグラフ、*G'*が**V'**上の有向アシリックグラフ、OがVと**V'**の部分集合であるとき、Gと*G'*は**O**上のFCI部分配向誘導路グラフが同じである場合にのみ**Oの**変数のみの間で同じd分離関係を持つ。

有向無サイクルグラフ*G*が与えられたとき、**O**上の*G*のFCI部分配向誘導経路グラフから、**Oの**メンバーだけを含むどのようなd-分離関係が*G*の真であるかを決定することが可能である。部分配向誘導経路グラフにおいて、$X \neq Y$で、XとYが**Zに**ない場合、Xと*Y*の間の無向パス*U*は、*U*上のすべての衝突者が**Zに**子孫を持ち、*U*上のすべて π の確定非衝突者が**Zに**なく、*U*上の他のすべての頂点が**Zに**ないが**Zに**子孫を持つ場合にのみ、**Zを**与えられたXと*Y*を**確実にd-接続する**。部分配向誘導経路グラフにおいて、**X**、**Y**、Zが変数の不連続集合であるとき、**Xの**あるメンバーが**Zを**与えられた

Yのあるメンバーとd-連結している場合に限り、**XはZを**与えられた**Yと**確実にd-連結する。

定理6.4.2： GがV上の有向アシリックグラフ、Oが**Vの**部分集合、**O**上$^{\pi}$の*G*のFCI部分配向誘導パスグラフ、**X**、**Y**、Zが**Oの**不連続部分集合のとき、**Xが***G*中の**Zを**与えられた**Yに**$^{\pi}$d-接続されるのは、**Xが**.の**Zを**与えられた**Yに**確実にd-接続する場合のみ。

これらのコロラリーはSpirtes and Verma (1992)で証明されています。

6.8 検出可能な因果関係 影響力に関する定理

本節では、部分配向誘導パスグラフから、多くの異なる種類の因果推論を導き出すことができることを示す。

定理6.5：**O**上の有向無サイクルグラフ*G*の部分配向誘導パスグラフであり、π、Aから*B*への有向パス*U*が*G*に存在する場合、Aから*B*への有向パスがGに存在する。

GがV上の有向無サイクルグラフで、Oが**V**に含まれる場合、CIアルゴリズムへの入力が**O**上の*G*に忠実なデータであれば、CIアルゴリズムの出力を**O**上の*G*の**CI部分指向誘導経路グラフと**呼ぶことにする。部分配向誘導経路グラフにおけるAから*B*への**半指向**経路とは、Aから*B*への無向経路Uで、どの辺も*A*を指す矢尻を含まない、つまり*U*上の*A*に矢尻がなく、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*端に矢尻は存在しない。

定理6.6：**O**上の有向無サイクルグラフ*G*のCI部分配向誘導パスグラフであり、π、Aから*B*への半有向パスがGに存在しない場合、Aから*B*への有向パスは*G*に存在しない。

異なる変数Aと*B*の間のトレックは、Aから*B*への有向パス、*B*から*A*への有向パス、または頂点Cから*A*、*B*それぞれへの有向パスのペアで、*C*でのみ交差するものである。以下の定理は、部分配向誘導パスグラフの辺が、終点以外の測定頂点を含まない

グラフのトレックを示すときの十分条件を示す。

定理6.7：**O**上のπ有向無サイクルグラフ*G*の部分配向誘導パスグラフであり、AとBがπ、で隣接し、Aと*B*の間にエッジを除く無向パスが存在しない場合、*G*には*A*と*B*の間に*A*または*B*以外の**O**内の変数を含まないトレクが存在する。

定理6.8：**O**上の有向無サイクルグラフ*G*のCI部分配向誘導パスグラフであり、AからBへのあらゆる半方向パスが、*G*のAから*B*へのあらゆる有向パスが**C**の何らかのメンバーを含む場合、。

定理6.9：有向無周期グラフの部分配向誘導パスグラフである場合
O上の*G*で、*A* <-> *B* in , あとき、*G*にAと*B*の共通原因が潜在的に存在する

。FCIアルゴリズムでも同様の結果が得られる。

これらの定理の適用を説明するために、最大情報量の部分的な
G5の因果構造の図13の指向性誘導パスグラフ。定理6.5を適用すると、*喫煙は繊毛の損傷*、*肺活量*、*測定された呼吸機能障害を*引き起こすと推論される。定理6.6を適用すると、*喫煙は心臓*病や*収入*、*親の喫煙習慣を*引き起こさないことが推論される。*所得が喫煙を*引き起こすのか、喫煙と*所得に*共通の原因があるのか、測定された変数間の条件付き独立関係から判断することは不可能である。測定変数間の統計から、*繊毛*損傷と*心臓*病には潜在的な共通原因がある、*繊毛損傷は心臓病を*引き起こさない、*心臓*病は*繊毛損傷を*引き起こさない、と判断される。

ここで、次章でより深く掘り下げることになるトピックを記しておく。一般に、変数Aから変数*B*への因果経路があるかどうかを判定することを目的とした研究のデザインにおいて、Aと*B*の間の関連を媒介する可能性のある変数だけでなく、*A*の原因となり得るものを測定することが有効であることを示唆している。

6.9 非依存性 拘束性

因果的に不十分なグラフに適用されるマルコフ条件と忠実条件は、条件付き独立関係ではない測定変数の限界分布の制約を伴うことがあり、したがって、FCIアルゴリズムでは使用されない。Thomas Verma (Verma and Pearl, 1991)による図19の例を考えてみよう。

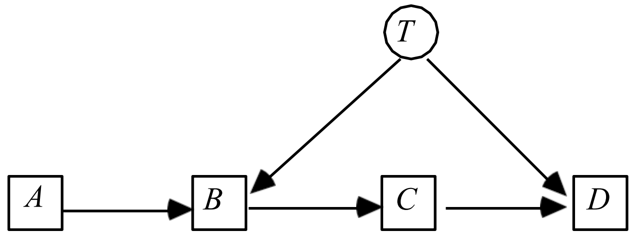


図19

Tが未測定であると仮定する。すると、グラフ全体に忠実な共同分布は、以下の制約を満たす必要がある。

$$\sum_{\vec{B}} p(b|a)p(d|b,c,a)$$

は、Cと*D*の値のみの関数である。

$$\sum_{\vec{B}} p(b|a)p(d|b,c,a) = \sum_{\vec{B}\vec{T}} p(b|a)p(d|b,c,a,t)p(t|b,c,a) = \sum_{\vec{T}} p(d|c,t)\sum_{\vec{B}} p(b|a)p(t|b,a)$$

(*A*、*Bは*{*C*、T}*を与えられたDから*独立し、Cは{*A*、*B*}*を与えられたTから*独立するから)。したがって

$$\sum_{\vec{T}} p(d|c,t)\sum_{\vec{B}} p(b|a)p(t|b,a) = \sum_{\vec{T}} p(d|c,t)p(t|a) = 。$$

$$\sum_{\vec{T}} P(D|C,T)P(T) = g(C,D)$$

(TとAは独立しているため）。

この制約は、AからDへの有向辺がグラフに追加された場合は伴わない。つまり、条件付き独立関係という形ではなく、潜在的な構造を特定するために原理的に利用できる、さらなる周辺構造が存在するということである。次節で線形モデルに目を向けると、同じような点が見えてきます。

6.10 一般化された統計的識別可能性と 直線性

何らかの理由で、調査が線形構造と、各乱数変数がその親と測定不能な要因の線形関数であるという仮定に一致する確率分布に限定されたとする。直線性のような制限の効果は、そうでなければ区別がつかないような因果構造を区別できるようにすることである。それは、その制約が、マルコフ条件、最小条件、忠実条件によって要求される条件付き依存関係や独立関係と一緒に、測定された変数に追加の制約を伴うからです。これらの追加制約は、条件付き独立関係の形式でない場合があります。線形の場合、通常はそうではありません。例えば、*X*変数が測定され、*T*変数が未測定である、以下の2つの構造を考えてみましょう。

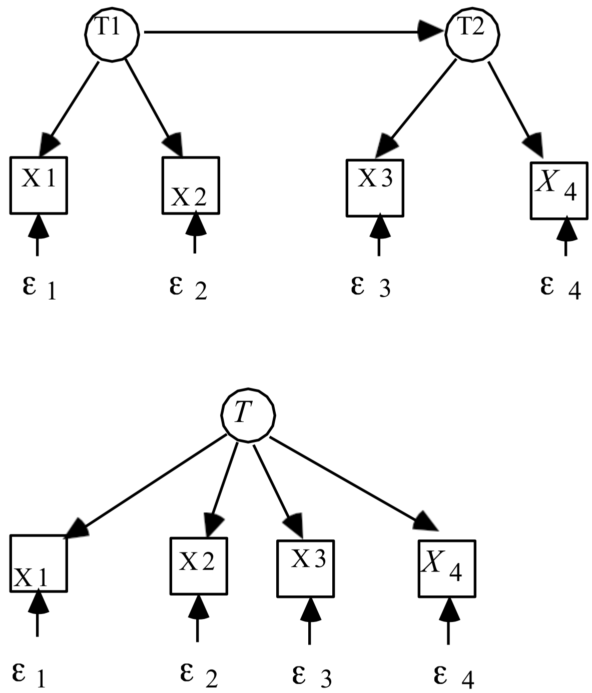


図20

これらの構造はそれぞれ、測定された変数に対する限界分布において、変数のすべてのペアが測定された変数の他のすべてのセットに対して条件付きで依存することを意味する。それぞれの場合、*X変数の*最大情報量の部分配向誘導パスグラフは完全な無向グラフである。これらの変数の間の条件付き独立関係を調べても、どのような構造が得られるかはわからない。しかし、もし線形性が要求されるなら、どのような構造が得られるかは簡単にわかる。なぜなら、線形性が仮定された場合、2番目の構造は、測定された変数の相関に関する以下の3つの制約をすべて含み、1番目の構造は、これらの制約のうち最初のものだけを含むからである（ここで、下付き文字を避けるために、$X1$と$X2$の相関を$\rho 12$とする）：

$$\rho 13 \rho 24 - \rho 14 \rho 23 = 0$$
$$\rho 12 \rho 34 - \rho 14 \rho 23 = 0$$
$$\rho 13 \rho 24 - \rho 12 \rho 34 = 0$$

今世紀初頭、Charles Spearman（1928）は、このような制約を**バニシング・テトラッ**

ド差と呼んだ。

このように、線形性制約の下での統計的区別不能の特徴は、全く新しい問題であり

、一般的な解決策を提示することはできないものである。とは言えない。

例えば、条件付き独立関係と四分子差の消失が、測定不能な変数を持つ線形構造の忠実な区別不能クラスを共同で決定することである。例えば、以下の各線型構造は、*A*、*B*、*C*、*Dに関する*限界分布において単一のテトラッド差が消失することを伴い、これらの変数に対して完全無向グラフからなる部分配向誘導路グラフを持つ：

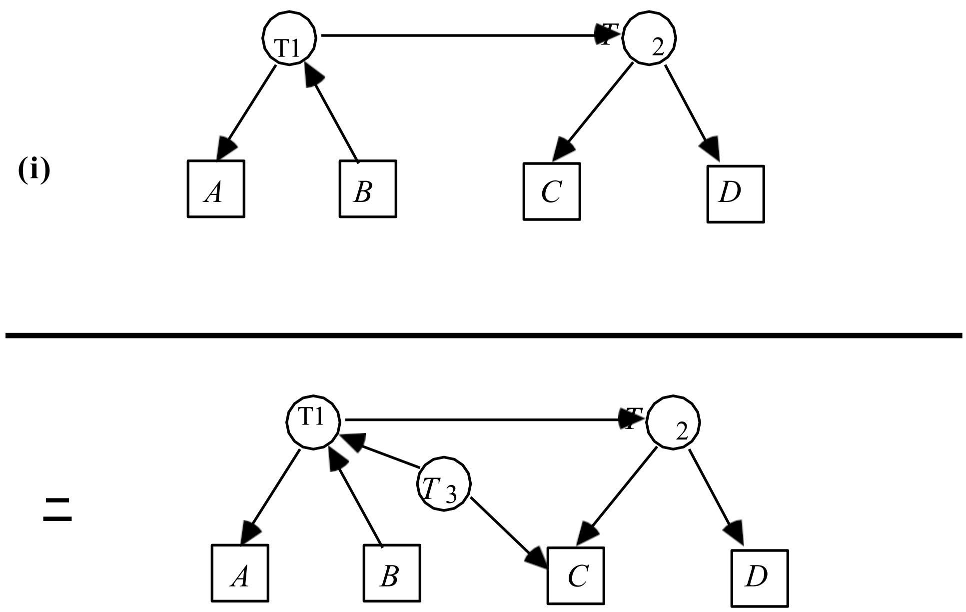


図21

しかし、この2つのグラフは、線形構造のクラスにわたって忠実に区別できない。構造(ii)は、AとBの相関が正、BとCの相関が正、AとCの相関が負である線形性に合致した分布を許容する。構造(i)は、マージンがこの条件を満たす、線形性と一致する分布を認めない。

構造(i)と(ii)は、社会科学の文献に見られる測定不能な変数を含む線形因果構造の典型的なものではありません。実用的には、消失性四分位数制約の検証は、部分的にしか線形でない構造であっても、代替的な因果構造を区別する強力な手段となる。消失性四分子差の仮説に対する検定は、1920年代にWishartによって正規変数を仮定

して導入され、Bollen (1989)によって漸近的に分布のない検定が記述されている。

消失性テトラッド差を利用したアルゴリズムについては、本書の後半で説明し、図解していきます。その利点を生かすためには、以下のことを判断できるようにする必要があります。

測定されていない共通の原因を持つ構造、あるいは持たない構造が、測定された変数の間で特定の消滅する四分木の差を伴うとき、アルゴリズム的にどうなるのか。この疑問は重要な定理につながる。

6.11 四元代表制 定理

ある有向無サイクルグラフ*G*で線形に表現される分布を**線形**モデルと呼ぶことにする。(線形モデルは、それを表現する有向無サイクルグラフGと、線形係数と、次数0の変数（誤差項を含む）の独立限界分布によって一意に決定されます（少し正式な定義は13章にあります）。

まずいくつかの用語がある：頂点Iと*J*の間のトレック*T*(*I*,*J*)*があるとき*、***I***(*T*(*I*,*J*))は*T*(*I*,*J*)のソースから*I*への有向パスを、***J***(*T*(*I*,*J*))は *T*(*I*,*J*) のソースから*J*への有向パスの集合を示している。（トレック中の有向パスが空のパスでもいいことに注意）**T**(*I*,*J*)はIと*J*間のすべてのトレックからなるセットを表す。

有向無サイクルグラフ*Gにおいて*、**T**(*K*,*L*)中の全ての*T*(*K*,*L*)と**T**(*I*,*J*)中の全ての*T*(*I*,J)について、*L*(*T*(*K*,*L*))と*J*(*T*(*I*,*J*))が頂点*Q*で交差していればQは*LJ*（*T*（*I*,*J*），*T*（*K*，*L*））という**詰所と言える**。同様に、**T**（*K*，*L*）内のすべての*T*（*K*，*L*）と**T**（*I*，*J*）内のすべての*T*（*I*，J）について、*L*（*T*（*K*，*L*））とすべての*J*（*T*（*I*，*J*））*が頂点Q*で交わり、かつすべての*T*（*I*．*L*)が**T**(*I*,*L*)内で、すべての*T*(*J*,*K*)内で、*L*(*T*(*I*,*L*))と*J*(*T*(*J*,*K*))も*Q*で交わるならQは*LJ*（*T*（*I*，*J*），*T*（*K*，*L*），*T*（*I*，*L*））**チョークポイントである．**トレックの定義も参照。

線形モデルにおけるテトラッド差の消失に関する基本定理はこうである：

四分木表現定理 .10：有向無サイクルグラフ*G*において、*LJ*

（*T*（*I*,*J*）,*T*（*K*,*L*）,*T*（*I*,*L*）,*J*（*K*））または*IK*（*T*（*I*,*J*）,*T*（*K*,*L*）,*T*（*J*,*K*）

）チョークが存在します。

は、*G*が$_{IJ\,KL} - _{IL\,JK} = 0$を線形に含意する場合にのみ、点である。

定理6.10の帰結として

定理6.11：非周期グラフGが$\rho_{IJ}\rho_{KL} - \rho_{IL}\rho_{JK} = 0$を線形に含意するのは、$\rho_{IJ}$または$\rho_{KL} = 0$、$\rho_{IL}$または$\rho_{JK} = 0$のいずれか、あるいは$\rho_{IJ.\mathbf{Q}} = \rho_{KL.\mathbf{Q}} = \rho_{IL.\mathbf{Q}} = \rho_{JK.\mathbf{Q}} = 0$のように$G$に確率変数の（おそらく空）集合**Qが存在する**ときのみである。

定理6.10は、任意の有向無サイクルグラフが線形に含意する消失性四分子差を計算するための高速アルゴリズムを提供する。定理6.11は、測定されていない共通原因が線形構造で作用している場合を判定する手段を提供する。後の章では、測定不能な変数間の因果関係の構造を調べるための、これらの事実の意味するところを説明する。

6.12 一例です：数学のマークと因果関係 解釈

Whittaker(1990)は、統計学におけるグラフィカルモデルに関する最近のテキストのいくつかの箇所で、Mardia, Kent and Bibby (1979)のデータセットについて、力学、ベクトル、代数、解析、統計という5つの数学科目の試験における88人の学生の成績について述べている。この例は、四分木表現定理の使用法の一つを説明するもので、Whittakerが説明する方法と我々の方法との解釈の重要な違いについてコメントする機会を与えてくれる。このデータの分散・共分散行列は次の通りである：

メカニクス	*ベクター*	*アルジェブラ*	*分析*	*統計情報*
302.29				
125.78	170.88			
100.43	84.19	111.60		
105.07	93.60	110.84	217.88	
116.07	97.89	120.49	153.77	294.37

これらのデータが与えられると、PCアルゴリズムはすぐに次のようなパターンを決定する：

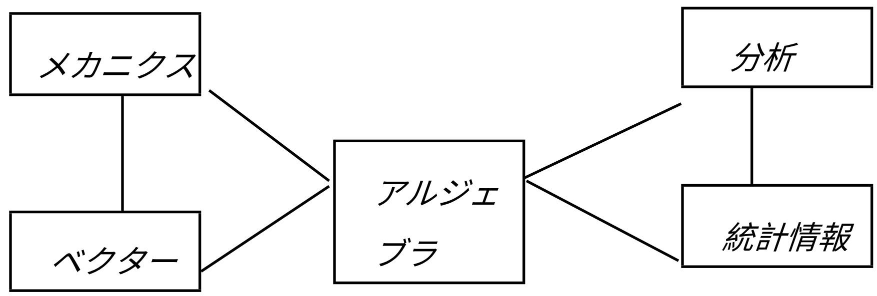


図22

Whittakerは別の解釈で同じグラフを得た。無向独立グラフとは、*G*を無向グラフ、Pを分布としたとき、*G*の頂点*X*、*Y*が*G*の他のすべての頂点の集合を条件として独立であれば隣接しない、あるいは逆説として、*X*、Yが*G*の他のすべての頂点の集合を条件として従属であれば、*X*、Yは*G*で隣接する、などの組をいう。 無向独立グラフは因果構造の多くを隠し、ときには独立関係の多くを隠す。このように、変数Xと*Z*が変数*Y*の原因であるが、Xと*Z*は統計的に独立しており、因果関係が全くない場合、無向性独立グラフはXと*Z*の間に辺を持つ。事実上、独立グラフは変数集合の適切な部分集合の間に保持される条件付き独立関係を表すことができない。

忠実な分布（またはサンプル）から得られるすべての*無向きの*パターングラフは、その分布から得られる無向きの独立グラフのサブグラフである。この場合、2つのグラフは同じになるが、一般的にはそうである必要はない。

Whittakerは、無向性独立グラフを特定することが4つの理由で重要であると主張している：(i)複雑な5次元の物体を2つの単純な3次元の物体--グラフの2つの最大閥--に還元する、(ii)変数を2組にグループ化する、(iii)試験の成績における異なる科目間の相互関係を分析する際に、*代数学を*一つの重要な試験として強調する。 (iv) *統計学を*予測するには代数と解析だけで十分、*力学は*代数と*ベクトル*で十分だと主張してい

る、しかし*代数学を*予測するにはすべての4点が必要だ（p．6)

第二の理由は第一の理由の単なる結果であり、第一の理由はあまり意味がないように思われる：変数が5つあることを認識することの負担はそれほど大きくない。統計学には、データを単純化するという理由で表現を導入し、実際にはそのような縮小の対象を原因として扱うという長い伝統がある。例えば、サーストン以降の因子分析の歴史がそれである。しかし、因子分析のように、因果関係のある結論が引き出されるのは

独立性グラフからの信頼性は低い。3つ目の理由は漠然としすぎていて、あまり意味がないように思われる。第4の理由による主張は正しいが、「予測する」ということが、コーチングによって意図的に変化させた変数の値を予測することとは無関係であると理解される場合のみである。このような教育データの統計解析は、因果関係を重視する傾向があり、そのような目的のためには、指示されたグラフモデルが仮説をよりよく表現していると思われます。

潜在変数の消失性四分木検定である定理6.11を適用すると、測定変数間の消失性部分相関では説明できない消失性四分木差が4つ存在することがわかる。これは、潜在変数が関与する消失性偏相関によって内包されることを示唆しており、潜在変数の導入を示唆する。試験対象の数学的構造から見て自然な考え方は、*代数学の知識が*指標となり、それがベクトルや*力学*で測定される*ベクトル代数の知識の*要因となり、また解析や*統計に*影響する実解析の*知識の*要因になるというものである。すると、データの説明は次のようになります:

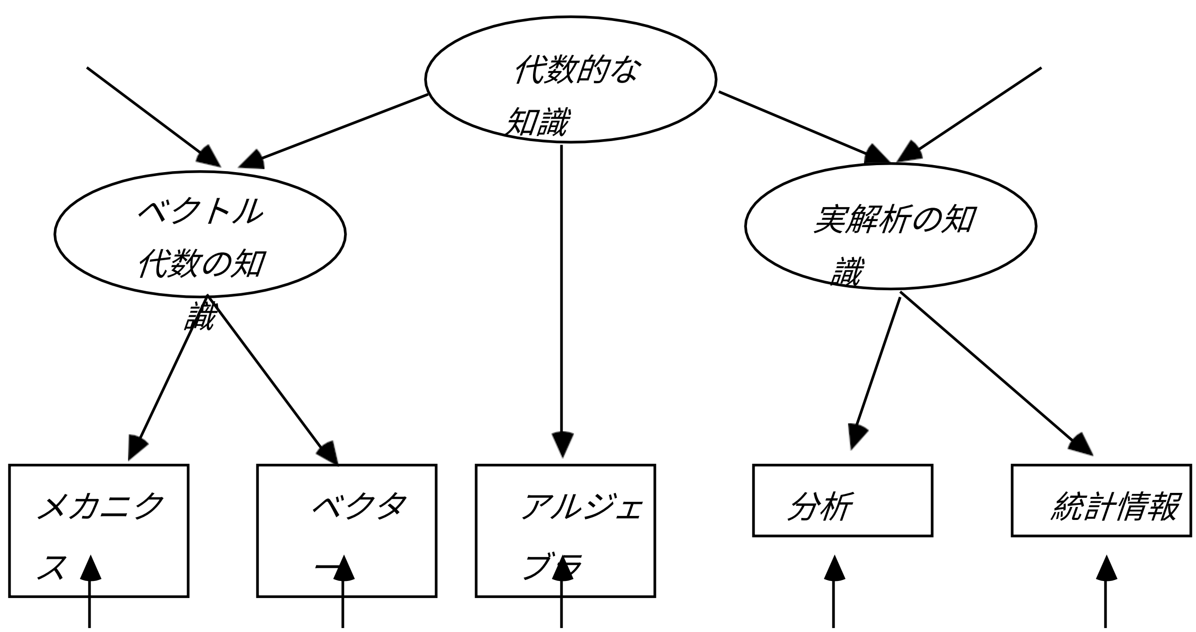


図23

表記のない矢印は、その他の変動要因を示している。忠実な分布と線形性を仮定すると、このグラフは、データが示唆する測定変数間の消失する一次偏相関を伴わない。しかし、*代数*的知識以外の要因による*代数学の*ばらつきが十分に小さければ、このグラフに忠実な線形分布は、まさにそれらの消失する部分相関を近似的に与えるだろう。

この構造（線形と仮定）には8つの消滅するテトラッドの違いが含まれ、TETRAD IIプログラムはそのすべてを識別して検定し、棄却できない（$p > .7$）。このモデル自体を尤度比検定で帰無仮説として扱うと、p値は約0.9となり、これはWhittakerが無向グラフ独立モデルについて報告している値とほぼ同じである。

6.13 背景 備考

一連の論文（Pearl and Verma 1990, 1991, Verma and Pearl 1990a, 1990b, 1991）において、VermaとPearlは、変数**Oの**集合に対する有向無サイクルグラフ*G*のパターン（あるいは「完成ハイブリッドグラフ」）と呼ぶ構造を出力する「Inductive Causation」アルゴリズムについて述べている。誘導経路、誘導経路グラフ、定理6.1の証明の重要な考え方は、すべてこの論文に示されている。残念ながら、この論文で主張されている帰納的因果律アルゴリズムの出力に関する2つの主要な主張（彼らのレンマA2と定理2で与えられる）は、偽である（Spirtes, 1992を参照）。

帰納的因果律アルゴリズムの初期のバージョンでは、*A* -> Bと*A* o-> *B*を区別していなかったため、定理6.5のように*A*が*B*を引き起こすことを推論するために使用することができなかった。この区別は、Spirtes and Glymour (1990) の定理6.5と定理6.6を証明するために（別の表記で）導入されたもので、Verma and Pearlは、その後のバージョンの帰納法アルゴリズムにこれを組み込んだ。Inductive Causation Algorithmは、エッジの方向付けに明確な識別パスを使用しないため、場合によってはFCI手順よりも少ない方向情報を与える。Inductive Causation Algorithmの出力には、三角形内の確実

に衝突しない辺と、単に向きを変えていない辺を区別する表記がない。CIアルゴリズムと同様に、*O*{**A**,*B*}のすべての部分集合を条件として、いくつかの変数のペアの独立性をテストする必要があるため、Inductive Causation Algorithmは大量の変数に適用することができない。

消失性四分子差は、Spearmanとその追随者たちによって、モデル仕様の原則的な手法として用いられた。彼らの手法については、Glymour, Scheines, Spirtes and Kelly (1987)で簡単に説明されている。Spearmanは、1916年から1935年にかけて、Godfrey Thomsonの一連の論文の中で、消失する四分子差から共通の原因を推定することに異議を唱えた。我々の用語では、Thomsonのモデルはすべて線形忠実性に違反している。

第7章

予測

7.1 はじめに

多くの実証研究の基本的な目的は、変化が自然に生じたものであれ、意図的な政策によって課されたものであれ、変化の影響を予測することである：環境中の鉛の発生源を減らすと、暴露された地域の子供たちの知能は向上するのか？環境中の鉛の発生源を減らせば、曝露された地域の子供たちの知能は向上するのか。タバコへの課税を強化すれば、肺がんは減少するのか。これらの効果はどの程度なのだろうか？ある畑にある種の小麦を植えた場合と別の種の小麦を植えた場合の収穫量の差はどうなるのか、すべての子どもにポリオワクチンを接種した場合としなかった場合の人口あたりのポリオ患者数の差はどうなるのか、仮釈放者に月600ドルを6ヶ月間与えた場合と何も与えなかった場合の再犯率の差はどうか、中年喫煙者に禁煙支援を行った場合の肺がん死亡の減少はどうか、ガロンあたり1ドルの追加税を課した場合のガソリン消費量はどうなのか。

無作為化試験に見られるような実験デザインの一つのポイントは、統計的な観点から、対応する治療法が一般的な政策となり、あらゆる場所で適用された場合に生じる分布そのものであるサンプルを*作成*しようとすることである。このような仮定の

下での実験では、統計的推論の問題は従来通りであり、簡単であるとは言い難く、政策結果の予測は原理的に問題ない。しかし、社会科学、疫学、経済学、その他多くの分野の実証研究では、観測されたサンプルが、ある政策を採用した場合に生じる分布そのものであることを知らない、あるいは合理的に仮定することができない。政策を実施することで、観測された標本に表れない形で関連する変数が変化する可能性がある。推論の課題は、受動的な観察または準実験的な操作に対応する分布から得られたサンプルから、政策が課された場合に生じるであろう分布に関する結論に移行することである。私たちの考えでは、統計的推論の最も基本的な問題の一つは、そのような推論が可能であるとすれば、それはいつなのか、ということである、

どのような手段で。モステラとテューキーによれば、その答えは「決して」である。この答えが分析に耐えられるかどうかは、これから見ていくことにしよう。

7.2 予測 問題点

予言の可能性は、少なくとも以下のような様々な種類の状況において分析することができる：

ケース1：因果関係グラフ、どの変数が直接操作されるのか、直接操作がその変数に何をするのかがわかっている。我々は、直接操作されない変数の分布を予測したい。より正式には、直接操作される変数の集合**X**、操作された分布の*P*(**X |Parents(X)**)、操作された集団の**Parents(X)** が操作されていない集団の**Parents(X)** の部分集合であることが分かっているのである。これは、Rubin、Holland、Pratt、Schlaiferが扱う状況であり、この場合、因果グラフと操作の定理は、操作されていない分布からの限界条件付き確率の観点から操作された分布を計算する関連式を規定する。後者はサンプルから推定することができ、計算された操作された分布の適切なマージンを取ることによって、**Xの**直接操作の下での**Y**（または**Zの**条件付き**Y）の分布を**求めることができる。

ケース2：直接操作される変数の集合**X**、操作された分布における*P*（**X｜Parents**（**X**））、操作された集団における**Parents**（**X** ）が操作されていない集団における**Parents**（**X** **）の**部分集合であること、測定変数が因果的に十分であることが分かっている、ケース1とは違って因果グラフが分かっていない。因果グラフはサンプルデ

ータから推測する必要がある。この場合、サンプルとPC（または他の）アルゴリズムは有向グラフのクラスを表すパターンを決定し、そのクラスの特性は、**Xを**直接操作した後の**Yの**分布が予測できるかどうかを決定します。

ケース3。難しい、興味深い、現実的なケースは、直接操作される変数の集合Xを知っていて、操作された集団の*P*（**X**｜**Parents**（**X**））を知っていて、操作された集団の**Parents**（**X**）が操作されていない集団の**Parents**（**X**）**の**部分集合であるが、事前知識とサンプルは、測定した変数に測定されていない共通の*原因があるかも*しれないという可能性を開いている場合に発生する。もし観察研究が根拠のない先入観なしに扱われるなら、きっとそれが典型的な状況なのだろう。それは

モステラーとテューキーが、コントロールされていない観察から予測することは不可能であると結論づけたのは、主にこのケースに起因している。Xを直接操作したときのYの分布、あるいは**Zに対する**Yの条件付分布を予測するという基本的な問題の一つの捉え方は、次のように定式化できる：*部分的に方向づけられた誘導経路グラフと、操作されていない分布の（観測変数上の）マージナルで真となる条件付き独立性事実だけが与えられたときに、予測に必要な条件と、予測に必要な条件を見つける。観測された分布から予測された分布の特徴を計算する方法を示す。*本章の最終的な目的は、この問題の部分的な解決策を提供することである。

これらのケースを順番に取り上げていくことにする。ケース1は簡単ですが、Rubinの理論との関連で時間をかけています。ケース2は非常に簡単に扱います。ケース3は、より典型的で理論的に最も興味深い推論問題であると我々は考えている。読者は、証明が先送りされたとしても、問題は複雑で困難であることを警告される。

7.3 ルビン・ホランド・プラット・シュライファー理論[1]

Rubinのフレームワークには、シンプルで魅力的な直観があります。実験や観察研究では、ある集団からサンプルを採取します。母集団の各単位は、子供であれ、国民経済であれ、化学物質のサンプルであれ、特性の集まりを持っています。母集団に含まれるユニットの特性のうち、いくつかは性質*的なもの*です。例えば、ガラスの花瓶は壊れやすく、強く叩くと割れてしまう性質を持っています。壊れやすい花瓶は、叩かれないと壊れないのです。同様に、子供の集団では、それぞれの読み聞かせプログラムに対して、その読み聞かせプログラムを受けたらあるテスト後の点数（またはテスト点数の範囲）を出すという性質を持っています。実験研究において、異なる単位に異なる処理を施すとき、私たちは、一部の単位だけがその性質が現れる状況にさらされたデータから、単位の性質（またはその平均値、またはその平均値の差）を推定しようとしているのです。Rubinは、このような各気質量Qと関連する治療変数Xの各値xに、母集団の各ユニットに対する値が、そのユニットに治療x

予測
*が与えられた*場合に*Q*が持つであろう値、言い換えれば、システムが*X*値を*x*に等しくすることを強いられた場合にQとなる確率変数、$Q_{X_f=x}$を関連づける。

1この項は、Spirtes, Glymour, Scheines, Meek, Fienberg and Slate, 1992を参考にしています。

x*1*、そのユニットについてQの値が測定された場合、測定されたQの値は、そのユニットについて

$Q_{Xf}=x1$です。

実験では，一組の値＜x, $y_{Xf}=x$＞を与えることができ，ここで$y_{Xf}=$xは確率変数$Y_{Xf}=x$の値である．しかし，治療*x1を与えられた*ユニット*iについて，*我々は，$Y_{Xf}=x2$，$Y_{Xf}=x3$*などの*値も知りたい．*これは，*ユニット*iが*治療*x2*または*x3に*さらされた場合，つまり，これらのユニットのX値が*x1ではなくx*2または*x*3に強制された場合，ユニットiが示すことになるY値をそれぞれ表すXの各可能値について知りたい．これらの未観測の値は、システムの因果構造に依存する。例えば，ユニットiが処理x*2におい*て示すべきYの値は，他のユニットに与えられた処理に依存するかもしれない．ここでは、このような依存性はないと仮定するが、因果構造とRubinの反実仮想変数の間の他の種類の接続を詳細に調べることにする。

Rubinの枠組みにおける典型的な推論問題は、一部のメンバーだけが治療*x*を受けたサンプルから、母集団のすべてのユニットについて、X*の*ある値*x*に対する$Y_{Xf}=x$*の*分布を推定するというものである。X*に*一意的な値を強制するのではなく*，*X*に*ある特定の値の分布を強制することを考えたり，他の変数Z*の*（強制されない）値に応じてX*に*異なる特定の分布を強制することを考えたりする．我々の「実験」は純粋に観察的で，X*が*値*x*であると観察されるときにユニット*iについて*変数Q*の*値*q*が必ずしも$Q_{Xf}=x$と同じではないことができるかもしれない．このような様々な問題に対する回答は、引用した論文に記載されています。例えば、我々の言い換えでは、PrattとSchlaiferは次のように主張している：

すべての単位が、YがXと場合によっては他の変数の効果であり、X以外のYの原因が測定されないシステムであるとき、Xのすべての値xについて、X＝xに対するYの条件

付き分布が$f_{Y|X=x}$と等しくなるためには、Xと各乱数変数$f_{Y|X=x}$（xはXのすべての可能な値の範囲）が統計的に独立していることで十分であり「ほぼ必要」です。

我々の用語では、$X=x$上のYの条件付き分布が、Xのすべての値xに対して$f_{Y|X=x}$と等しいとき、X上のYの条件付き分布が "不変 "であると言い、彼らの用語では、"観測可能 "であると言う。プラットとシュライファーの主張は、いくつかの例で明確にすることができ、また、このフレームワークの適用におけるいくつかの暗黙の前提を説明するのに役立つ。Xと観測されないUがYの唯一の原因であり、それらは互いにいかなる因果関係も持っていないとする。

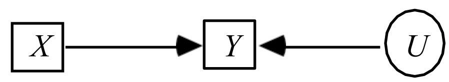


図1

簡単のために、依存関係はすべて線形で、*X*、*Y*、*U*のすべての可能な値とすべてのユニットについて、*Y*＝*X*＋*U*と仮定します。Xfは、集団内のすべてのユニットに*強制される*可能性のある*X*の値を表すとします。Xは観測変数であり、Xfは観測変数ではない。Xは確率変数であり、Xfは確率変数ではない。表1の値を考える。

表1

X	*Y*	*U*	*エックスエフ*	UXf=1	YXf=1
1	1	0	1	0	1
1	2	1	1	1	2
1	3	2	1	2	3
2	2	0	1	0	1
2	3	1	1	1	2
2	4	2	1	2	3

簡単のために、各行（確率変数でない*Xf*は無視する）が等確率であるとする。ここで、*X* と *Y* の列は、測定された変数の可能な値を与える。Xfは，あるユニットに強制される可能性のある*X*の値を示す変数で，その列は*Xf*以降に表を続けていない.

= 1.*UXf=1*の列は、Xが値1を持つことを強制されたときの*U*の値の範囲を表し、*YXf*=1は、Xが値1を持つことを強制されたときの*Y*の値の範囲を示す。表中の*YXf*=1は、*Xf*の値と*UXf*=1の値によって一意に決まり、*X*の値とは無関係であることに注意。

この表は、プラットとシュライファーの主張である「*YXf* =1は*X*に依存せず、*X* = 1を条件とする*Y*の分布は*YXf* =1の分布に等しい」を示しています。

$U=_{UXf}=1$、$_{YXf}=1=1+_{\mathrm{UXf}}=1$とすることで、表を作成した。つまり、$X$の分布を除けば、すべてのユニットに$X$の値を強制しても、因果構造、確率構造は全く変化しないと仮定して表を作成したのである。によって

を$Y_{Xf=2}$ = 2 + $U_{Xf=2}$で適用すると、Xf=2のときにPrattとSchlaiferの主張を満たす値を得るために表を拡張することができます。

Pratt と Schlaifer のルールに従って、*X* に対する *Y* の条件付確率が直接操作の下で不変 *でない* 別の例を考えてみる。この場合、XはYを、*U*は*Y*を引き起こすが、Xと*U*の間には、従来通り、いかなる因果関係も存在せず、さらに、測定不能な変数VがXと*Y*の共通の原因となっている、図2のような状況である。

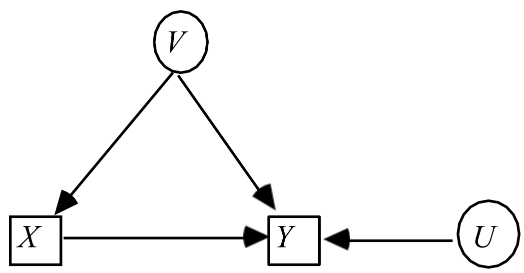


図2

表1と同じ規約で、次のような分布を考えてみましょう：

表2

X	*V*	*U*	*Y*	*エックスエフ*	VXf=1	UXf=1	YXf=1
0	0	0	0	1	0	0	1
0	0	1	1	1	0	1	2
0	0	2	2	1	0	2	3
0	0	3	3	1	0	3	4
1	1	0	2	1	1	0	2
1	1	1	3	1	1	1	3
1	1	2	4	1	1	2	4
1	1	3	5	1	1	3	5

ここでも、すべての行が同じ確率であると仮定し、確率変数でないX_f*の値を無視する。*$Y_{Xf=1}$がXの値に*依存するように*なったことに注意。そして、PrattとSchlaiferが要求するように、$X = 1$に対するYの条件付き分布は、$Y_{Xf=1}$の分布と等しくないのです。

の分布が、X＝1が強制され、したがってXf＝1となるようにテーブルを構成した。$U_{Xf=1}, V_{Xf=1}$ は Xf から独立している。言い換えれば、方程式系が

$$y = x + v + u\,x$$
$$= v$$

は、X、Y、U の値を求め、仮定 $U_{Xf=1} = U$、$V_{Xf=1} = V$、方程式は

$$Y_{Xf=1} = Xf + V_{Xf=1} + U_{Xf=1}$$

は、$U_{Xf=1}$、$V_{Xf=1}$、$Y_{Xf=1}$の値を決定するために使用されました。強制システムは、図3に描かれた図によって記述されているかのように扱われた。

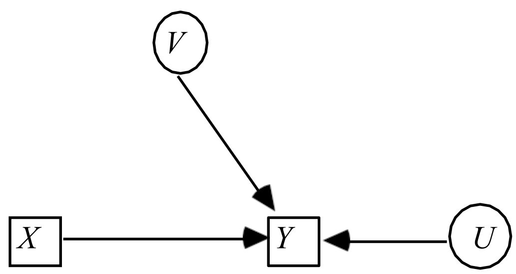


図3

例えば、$Y = X + U$ であるが、Y と X *の両方に*依存する変数 V *が存在し*、図 4 のような系になるとする。

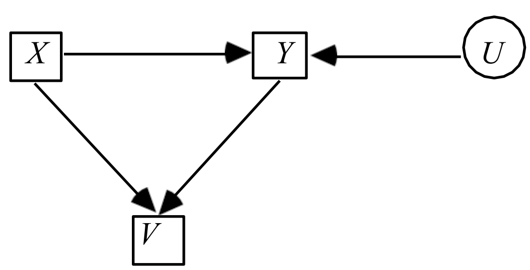


図4

ここで、$Y = X + U$、$V = Y + X$ と仮定して得られた値の表を示します。これらの関係

予測
は、X を直接操作しても変更されません:

表3

X	*Y*	*V*	*U*	*エックスエフ*	VXf=1	UXf=1	YXf=1
0	0	0	0	1	2	0	1
0	1	1	1	1	3	1	2
0	2	2	2	1	4	2	3
1	1	2	0	1	2	0	1
1	2	3	1	1	3	1	2
1	3	4	2	1	4	2	3

ここでもすべての行が等確率であると仮定する。YXf=*1はXに*依存せず、YXf=1は*X*=1の条件下で*Y*と同じ分布を持つことに注意。したがって、PrattとSchlaiferの原理は再び満たされ、さらに*Xに対するYの*条件付き確率は不変である。この表は、操作された系が操作されていない系と全く同じ方程式系を満たすと仮定し、実質的に図4の依存関係のグラフが*Xに*値を強制することによって変化しないと仮定して作成された。

我々が再構築したPrattとSchlaiferのルールは、Markov Conditionの結果である。Rubinが説明した他の例も同様である。この関係を明確にするために、いくつかの結果が必要である。ここでは、第3章で紹介した技術的な定義を前提とし、さらにいくつかの定義が必要である。

Gが変数**V**□**Wの**集合に対する有向非循環グラフであり、Wが*Gにおける***Vに関して**外生的であり、YとZが**Vの**不連続部分集合であり、*P*（**V**□**W**）が*Gに対する*マルコフ条件を満足する分布であるとする、とし、**Manipulated**(**W**) = **X** とすると、*P*(**Y**|**Z**)は、**Wを**w1から**w2に**変更することによって*G*中の**Xを**直接操作しても、両者が定義されているところでは$P(\mathbf{Y}|\mathbf{Z}, \mathbf{W}=\mathbf{w}_1) = P(\mathbf{Y}|\mathbf{Z}, \mathbf{W}=\mathbf{w}_2)$のときだ

け**不変であるとする。**なお、*P*(**Y**|**Z**)が*G*中の**X**を**W**を変えて直接操作しても不変であるための十分条件は、*G*中の**Z**が与えられたときに**W**が**Y**から d分離されることである。YとZを含む有向非循環グラフ*G*において、**ND**(**Y**)は**Y**に子孫を持たないすべての頂点の集合である。 **Y**∩**Z** = ⍰のとき、*V*が**IV**(**Y**,**Z**)（**Z**与えられた**Y**に対する情報変数）内にあるのは、**Z**与えられた**Y**と d-連結し、*V*は**ND**（**YZ**）内にない場合のみである。**(Y∩Z＝φ の 場合～** *W*が**IP**(**Y**,**Z**)（*W*は**Z**を与えられた**Y**の情報変数である親を持つ）にあるのは、*W*が**Z**のメンバーであり、かつ*W*が**IV**(**Y**,**Z**) □**Y**に親を持つ場合だけである。ここでは、以下の結果を用いる。

定理7.1：GCombが**V**?**W**上の有向無サイクルグラフで、Wが*GComb*の**Vに関して**外生的、YとZが**Vの**不連続部分集合、*P*（**V**?**W**）が*GComb*のマルコフ条件を満足する分布であるとき、**X**∩**Z** のメンバが *GUnman* の **IP**(**Y**,**Z**) のメンバでなく、**X** **Z** のメンバが *GUnman* の **IV**(**Y**,**Z**) のメンバでないとき、 *P*(**Y**|**Z**) は *GComb* の **X** を $\mathbf{w}_1$ から $\mathbf{w}_2$ に変えて直接操作しても変化しない。

定理 7.1 の重要な点は、*GComb* において **W を** $\mathbf{w}_1$ から $\mathbf{w}_2$ **に**変えるという **X の**直接操作の下で *P*(**Y**|**Z**) が不変であるかどうかは *GUnman* の特性だけで決まるということです。したがって、**W** や *GComb* を指定せずに、*P*(**Y**|**Z**) が *GUnman* の **X** の直接操作のもとで不変であることを述べることがある。

前述の各例とPrattとSchlaiferの一般法則は、定理7.1の補題の結果である：

公則7.1：GCombが**V**?**W**上の有向無サイクルグラフで、Wが*GComb*の**Vに関して**外生的で、XとYが**Vに**あり、*P*（**V**?**W**）**が***GComb*のマルコフ条件を満たす分布であるとき、*P*（*Y* / X）*は*、*GUnmanにおいてX*への無向パスが頂点の空集合を与えてXと*Y*をd接続しない場合、**Wをw**1から**w2に**変更し*GComb*の*X*を直接操作しても変化しない。同様に、(1) *Y* が *X* の（直接・間接）原因でなく、(2) *GUnman* において *X* と *Y* の共通の原因がない場合。

グラフの用語では、PrattとSchlaiferの主張は、「観測可能性」（不変性）のために、Gと*G'*（*G*から*X*へのすべての辺を除去して得られる操作システムのグラフ）、およびそれらの関連確率が、*X*上の*Y*の同じ条件分布を与えることを要求することになる。この主張の充足側を特徴づけているのが公理7.1である。PrattとSchlaiferは、この条件を "almost necessary "と言っている。これは、彼らの条件の先行詞が成立せず、結果的に成立する場合があるということであり、さらに、先行詞が成立しない場合、

条件付き確率が特別な制約を満たさない限り、結果的に成立しないということであると考えることができる。同様のことが、我々が与えたグラフ条件にも当てはまる。XとY*の*間に、*Xに*入る空集合が与えられ、*Xを*直接操作したときのY*の*確率が、*Xに対するYの*元の条件付き確率に等しいd-connecting pathが存在するケースが存在する。この場合も、条件付き確率が制約を満たす場合にのみ先行詞が破綻し、結果詞が成立するので、この条件は "ほぼ必要 "と言える。

*X*に値を強制したときの*Y*の分布は、*Xに対するYの強制*されていない条件付き分布からは予測できないが、それでも、*X*に値を強制したときの**Zに対する***Y*の条件付き分布は、X と**Zに対する***Y*の強制されていない条件付き分布から予測できるということがありうる。 PrattとSchlaiferは、X と*Y*に加えて、さらにいくつかの変数**Z**が測定されている場合の一般論について検討している。プラットとシュライファーは、Xと**Zに対する***Yの強制*されない条件付き分布が、**X**が特定の値を持つことを強制された集団における**Zに対する***Y*の条件付き分布と等しいとき、YとXに関係する法則は「**Zを伴って**観測可能」であると言う。

すなわち、*Xの任意の値xに対して、Xの分布が$_{ZXf=x}$のzの値に対する$_{YXf=x}$の条件付き分布から独立であること、Xが値を持つことを強制されたとき*、PrattとSchlaiferは付随物を伴う観測可能性の十分条件と「ほぼ必要」条件を主張*する。*
x.この規則も、定理7.1の特殊なケースである。

Rubinによる例を考えてみる。(ルービンのXはプラットとシュライファーのZ、ルービンのTはプラットとシュライファーのX）。ある教育実験において、読書プログラムの課題*T*が、1つ以上の測定不能な共通原因*V*を共有するあるテスト前の変数*X*の無作為抽出値に基づいて、ポストテストの得点である*Y*に割り当てられた場合、集団内のすべての学生に治療*T*＝1を与えた場合と、すべての学生に治療*T*＝2を与えた場合の*Y*値の平均差を予測したいと思います。実験の状況は、図5のようになる。

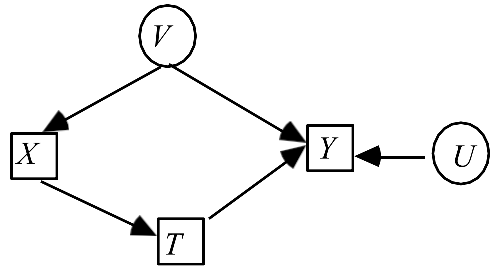


図5

実験サンプルが十分に代表的であれば、RubinはYの不偏推定値を求める
τ
ことができるとしている：*X*の値を1から*K*まで*k*とし、*T*＝1、*X*＝*k*の条件下での*Y*の平均値を*Y1*kとし、*Y2*kについても同様とする。*T*＝1、*X*＝*k*とした場合のサンプル中のユニット数を*n1*kとし、*n2k*についても同様とする。数 *n1* と n2 はそれぞれ *T* = 1 と *T* = 2 のサンプルに含まれるユニット数の合計を表す。

$\overline{Y}_{Tf=1}$ = 治療1が全ユニットに強制された場合のYの期待値 とする。Rubin によれば、$\overline{Y}_{Tf=1}$ を次のように推定する：

$$\sum_{k=1}^{K}\frac{n1k+n2k}{n1+n2}\overline{Y1k}$$

および見積もり　によるものです：

$$\sum_{k=1}^{K}\frac{n1k+n2k}{n1+n2}\left(\overline{Y1k}-\overline{Y2k}\right)$$

この選択の根拠は明らかではないかもしれない。もし、すべてのユニットがT＝1を強制される仮想的な集団に注目するならば、ルービンの暗黙の独立性の仮定から、次の導出が示すように、操作された集団を図6のような因果構造を持っているかのように扱うことは明らかである。

$$\overline{Y}_{Tf=1}=\sum_{\vec{Y}} Y\times P(Y_{Tf=1}) ＝。$$

$$\sum_{\vec{Y}} Y\times\sum_{k=1}^{K} P(Y_{Tf=1}\mid X_{Tf=1}=k, T_{Tf=1}=1)\,P(X_{Tf=1}=k\mid T_{Tf=1}=1)\,P(T_{Tf=1}=1) ＝。$$

$$\sum_{\vec{Y}} Y\times\sum_{k=1}^{K} P(Y_{Tf=1}\mid X_{Tf=1}=k, T_{Tf=1}=1)\,P(X_{Tf=1}=k)。$$

$P(T_{Tf=1}=1)=1$であり、$X_{Tf=1}$と$T_{Tf=1}$は図6に示す因果関係グラフによれば独立なので、上式の第二等式が成り立つ。定理7.1により、$P(Y_{Tf=1}|X_{Tf=1},T_{Tf=1})$ と $P(X_{Tf=1})$はともに図5のグラフのTを直接操作しても不変である。これは以下の式を内包している。

$$\overline{Y}_{Tf=1}=\sum_{\vec{Y}} Y\times\sum_{k=1}^{K} P(Y_{Tf=1}\mid X_{Tf=1}=k、T_{Tf=1}=1)P(X_{Tf=1}=k)＝。$$

$$\sum_{k=1}^{K} P(X=k) \times \sum_{Y} Y \times P(\mathrm{Y} \mid X=k, T=\mathbf{1}) = \frac{n1k + n2k}{\mathrm{n1} + \mathrm{n2}} \times Y\overline{1k}$$

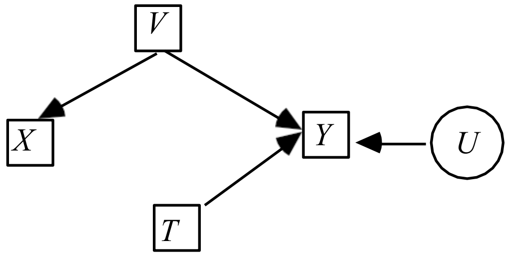


図6

なお、XとTは、$X_{Tf=1}$と$T_{Tf=1}$とは異なり、独立ではない。ルービンが$X_{Tf=1}$と$T_{Tf=1}$が独立であると黙認しているのは、操作された集団の因果グラフが図6のグラフであり、操作されていない集団の因果構造である図5のグラフではないことを暗に想定していることになる。$Y_{Tf=2}$も同様な方法で導出できる。

のファッションを提案します。

ルービンの理論の再構成は、母集団のすべてのユニットが関連する変数について同じ因果構造を持つことを前提としているが、もちろん、ユニットが他の点で均質であることを前提としているわけではない。母集団（および推論を行うサンプル）の因果構造がユニットごとに異なっていても、プラットとシュライファーの法則に従った予測に必要な反事実を知っている人がいるかもしれないことは考えられることである。例えば、AとBには測定不能な共通の原因がなく、BはAを引き起こさないことが知られていて、母集団は実際にはAがBを引き起こす系とAとBが独立した系が混在している可能性があるとする。この場合、Aが$A=a$の値を持つように強制された場合のBの分布は、$A=a$を与えられたBの条件付き確率から予測することができ、実際にその確率は同じである。この場合、また他の因果構造が混在する集団の予測の場合も同様で、PrattとSchlaiferのルールを適用して得られる予測は、因果的に均質な部分集団のそれぞれにおいて、関連する条件付き確率が不変であるかどうかを考慮すれば、マルコフ条件から導くことができる。したがって、AとBに因果関係がない場

合、$P(B|A=a)$はAが値aを持つことを強制されたときのBの確率に等しく、AがBを引

き起こす場合、$P(B|A=a)$はAが値aを持つことを強制されたときのBの確率に等しく

、この2つの原因構造を持つシステムのいかなる混合においても確率は同一である。

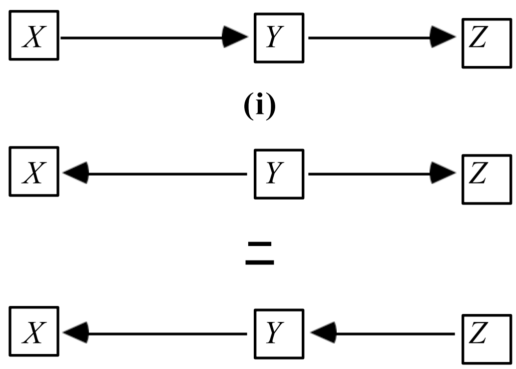
7.4 Causal Sufficiencyによる予測

Rubinのフレームワークは、2つの次元で特殊である。それは、既知の様々な反事実的（あるいは因果的）特性を仮定し、条件付き確率の*不変性を*扱うものである。しかし、我々はデータを考える前に因果構造や反事実がわからないことが非常に多く、不変性*そのものに*興味があるのではなく、予測の道具としてのみ興味があるのです。目標をより明確にする必要がある。調査者は、**Oを**含む未知の頂点集合**Vを**持つ未知の因果グラフ$_{GUnman}$に忠実な分布の**O**上のマージナルである分布$_{PUnman}$(**O**)を知っている（または推定している）とする。彼女は、直接操作される**Oの**メンバーである変数*X*と、$_{GMAN}$*における**X**の*直接原因となる変数**Parents**（$_{GMAN}$、*X*）*を知ってもいる*。彼女は、Xが直接操作される唯一の変数であることを知っています。最後に彼女は、操作が*X*に何をするか知っている、つまり、$_{PMan}$(*X* |**Parents**($_{GMan}$,*X*))を知っている。このような状況で、未知の因果グラフがどうであれ、観測されていない変数に対する操作された分布と操作されていない分布がどうであれ、また、操作が今述べた仮定と一致してどのようにもたらされたとしても、$_{PMAN}$（**Y**｜**Z**）が一意に決まるならば、**Zに**基づく**Yの**分布は**予想可能**である。目標は、**Zを**条件とする**Yの**分布がいつ予測可能か、そしてどのようにして予測を得るかを発見することである。

$_{PUnman}$(**O**)が、操作されていないグラフ$_{GUnman}$に忠実な分布の**O**上のマージナルであるという仮定は、いくつかの理由で失敗することがある。まず、分布の特定のパラメータ値のために失敗する可能性がある。Wが政策変数の集合である場合、**w 2**（操作された）部分集団が**w 1**（操作されていない）部分集団にない依存関係を含むために失敗することもある。例えば、電池と電球が、スイッチを含む回路で接続されているとする。スイッチの状態を*W*、スイッチがオフである操作されていない部分集団がw_1、スイッチがオンである操作された部分集団が$_{w2}$とする。$_{w}$*1*の部分集団では

、電球は常に切れているので、電球の状態（オンかオフか）は電池の状態（充電されているかいないか）に依存しません。一方、$w2$部分集団では、電球の状態は電池の状態に依存します。したがって$GComb$には電池の状態から電球の状態への辺があり、$GUnman$（$GComb$のWを除いた部分グラフ）にも電池の状態から電球の状態への辺があることになる。このことは、$w1$部分集団の電池の状態と電球の状態に関する共同分布が$GUnman$に忠実でないことを意味します。予測アルゴリズムの結果は、操作によって追加の依存関係（自分の背景知識の一部である場合もない場合もあります）を導入しない状況でのみ信頼できます。）

一般的だが未知の因果構造を持つシステムに対して因果的に十分であると正しく信じられている変数の観測から、介入や政策の効果の予測を行いたいとする場合である。この場合、サンプルとPC（または他の）アルゴリズムは有向グラフのクラスを表すパターンを決定し、そのクラスの特性は、*X*を直接操作した後の*Y*の分布が予測できるかどうかを決定します。例えば、パターンが*X* - *Y* - *Z*であり、図7のグラフの集合を表すとする。



(iii) 図7

これらの因果グラフのそれぞれについて、*X*を直接操作した後の*Yの分布を*計算することができるが、その結果は最初のグラフと他の2つのグラフとで異なる。それぞれのグラフの$P_{Man}(Y)$は、操作の定理と適切なマージンをとることから計算することができ、それぞれのグラフの結果を以下に示す:

(i) $$P_{Man}(Y) = \sum_{X} P_{Unman}(Y \mid X) P_{Man}(X) \neq P_{Unman}(Y)$$ です。

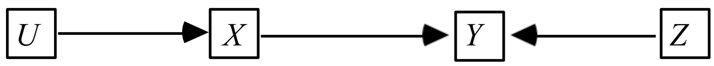
(ii) $$P_{Man}(Y) = P_{Unman}(Y)$$

(iii) $$P_{Man}(Y) = \sum_{Z} P_{Unman}(Y \mid Z) P_{Unman}(Z) = P_{Unman}(Y)$$

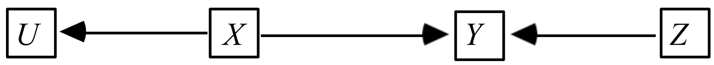
集団のすべてのユニットが同じ値のXを持つことを強制される場合、(i)操作されたYの分布は、操作されていないYの分布と等しくない。(ii)と(iii)の場合、操作されたYの分布は、操作されていない分布と等しくなる。パターンがないため

は、これらの構造のうちどれが正しいかを教えてくれ、*X*の操作上の*Y*の*分布を*予測することはできない。

例えば、*U - X -> Y <- Z*のパターンは、図8のグラフのいずれかを表すことができます。



(i)



(ii) 図8

各グラフの$PMan(Y)$は、操作の定理と適切なマージンを取ることで計算できます。各グラフの結果は以下の通りです：

(i) $$PMan(Y) = \sum_{\vec{X}} PUnman(Y \mid X) PMan(X)$$

(ii) $$PMan(Y) = \sum_{\vec{X}} PUnman(Y \mid X) PMan(X)$$

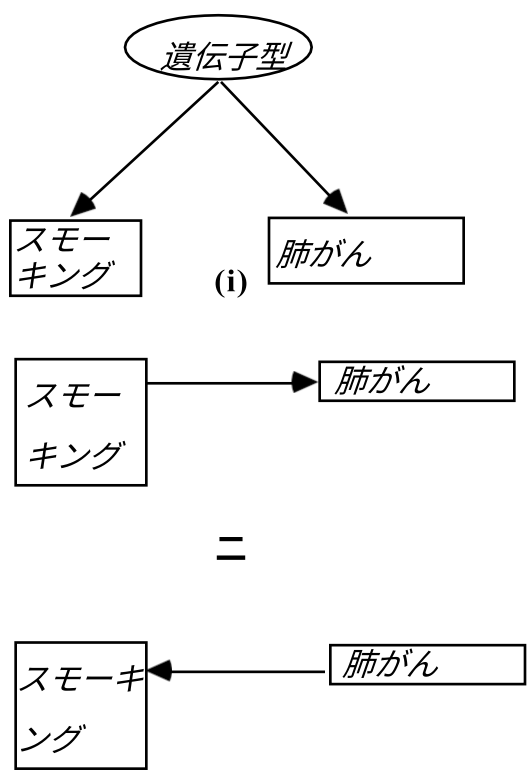
(ただし、$PMAN(Y)$は(i)と(ii)で同じですが、$PMAN(U, X, Y, Z)$は(i)と(ii)で同じではないので$PMAN(U, X, Y, Z)$は予測不能です)。

構造が因果的に十分であることが分かっている場合、パターンを見つけて操作定理を適用し、パターンが表すすべてのグラフについて適切なマージンをとることによって、変数の分布（または変数のある集合の別の集合に対する条件分布）の予測可能性を決めることができます。すべてのグラフが同じ結果を与える場合、それが予測である。様々な計算上のショートカットが可能であり、そのいくつかは次節で述べる予測アルゴリズムで説明する。

7.5 因果関係のない予測 充足感

操作されたシステムの因果構造が観察されたシステムの因果構造と異なることが分かっていても、観察されたシステムの因果構造が不明で、観察された統計的依存関係が観察されない共通の原因によるものであることが分かっていても、その原因が分からないという最も深刻なケースにようやく行き着く。これは、モステラやテューキーが非実験的研究の典型と考える状況であり、私たちも同意見である。問題は、それにもかかわらず、予測が可能な場合があるのかどうか、可能だとすればいつ、どのように可能なのかということである。

次のような些細な例を考えてみましょう。喫煙と肺がんだけを測定した場合、両者には相関があることがわかる。この相関は、図9に描かれた3つの因果関係グラフのいずれかによって生み出される可能性がある。



(iii) 図9

3つのグラフはすべて、図10に示すように、同じ最大情報量の部分配向誘導パスグラフをもたらす。

図10

グラフ(i)や(iii)で*Smokingを直接操作した*場合、*P*(*Lung cancer*)は変化しないが、グラフ(ii)でSmokingを直接操作した場合、*P*(*Lung cancer*)が変化する。つまり、測定された変数の限界分布から、*Smokingの直接操作の効果を*予測することはできないのです。

操作定理によれば、各有向非循環グラフ*GUnman について*、分布を*PUnman*(**W**)(**V**|*Parents*(*GUnman*,*V*))の形の項の積に分解するとき、*PUnman*(**W**)(*X* |**Parents**(*GUnman*,*X*)を*PMAN*(**W**)(*X* |**Parents**(*GMan*,*X*)に置換するだけで変数*Xに対する*操作効果が算出できる。）(と置き換える（ここで、GManは操作されたグラフである）。この単純な置換は、*PUnman*(**W**)(*X* |**Parents**(*GUnman*,*X*))以外の因数分解における各項は、*GUnmanにおけるXの*いかなる直接操作に対しても不変であることが保証されており、したがって操作されていない集団における頻度から推定できる。

ここで、*P*(**O**)が有向無サイクルグラフ*GUnmanに*忠実な分布*P*(**V**)のマージナルであり、*GUnmanの*部分配向誘導パスグラフである、因果的に非充実なケースにこの戦略を一般化してみることにしましょう。我々は、*PUnman*(**V** |*M*(*V*)) (ここで、集合**M**(*V*)のメンバーシップは*Vの*関数である)という形式の項の積であり、*PUnman*(*X* |**M**(*X*))を除く各項が、*Xの*すべての直接操作の下で不変であり、そのためのすべての有向無周期のグラフがあるPPの分布の因子化を探索できる。

このような因数分解が見つかれば、*PMan*(*X* |**Parents**(*GMan*,*X*))を*PUnman*(**V** |*M*(*X*))に置き換えることで操作の効果を予測することができる。(ここでGManは操作されたグラフ）、因果的に十分な場合と同じように。が存在する多くの有向無サイクルグラ

フのうち、どのグラフが有向無サイクルグラフであるかはわからないだろう。

O上の部分配向誘導パスグラフが実際に分布を生成した。しかし、$P_{Man}(\mathbf{Y}|\mathbf{Z})$はそれぞれ同じになるので、それは問題ではないだろう。これは本質的に我々が採用する戦略である。$P(X \mid \mathbf{Parents}(G_{Unman}, X))$以外の各項が$G_{Unman}$*における***Xの**直接操作の下で不変である因数分解を単に構築できる因果的に十分な場合とは異なり、因果的に十分でない場合には、各項が以下のようになる因数分解を見つけるために、異なる因子分解から*探さなければならない*。

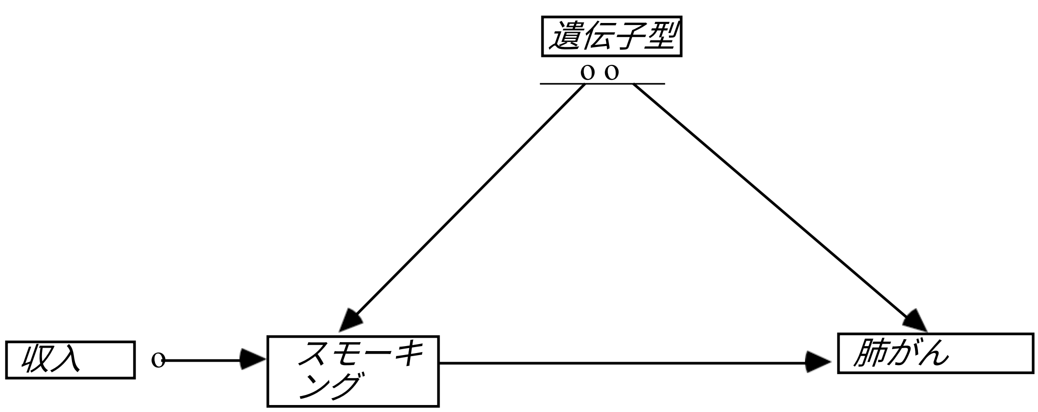
ただし、$P_{Unman}(X \mid \mathbf{M}(X))$は、**O**上の部分配向誘導経路グラフが.Oに等しいすべての有向無サイクルグラフ*G*に対して、*X*のすべての直接操作の下でπ不変である。幸いなことに、私たちが

がわかるように、$P(\mathbf{O})$の可能なすべての因数分解を検索する必要はない。

この戦略の詳細を具体化し、例を示す。FCIアルゴリズムを用いて、G_{Unman}の**O**上のπ部分配向誘導パスグラフを構築する。第6章で説明したVermaとPearlの例を考慮すると、**O**π上の部分配向誘導経路グラフを構成するグラフの中には、どのような分布も表現できないものがあるかもしれないことに注意する。

をマージナル$P_{Unman}(\mathbf{O})$とするのは、非依存性の制約があるためである。本書で展開された理論から、予測可能性を決定し、原理的に可能な限り予測を得る計算手順を提供することは望めない。なぜなら、グラフが限界分布に伴いうるすべての制約について理解していないからである。しかし、条件付き独立性制約のみを考慮することで、予測可能性の十分条件を提供することができる。

ここで、この戦略をより詳しく説明する例を挙げます：*遺伝子型*（*G*）、*喫煙*（*S*）、*所得*（*I*）、*両親の喫煙習慣*（*PSH*）、*肺がん*（*L*）*を*測定したとする。操作されていない分布は、図11に示すような部分配向誘導パスグラフを持つ操作されていないグラフに忠実であるとする。



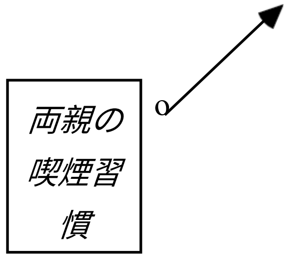
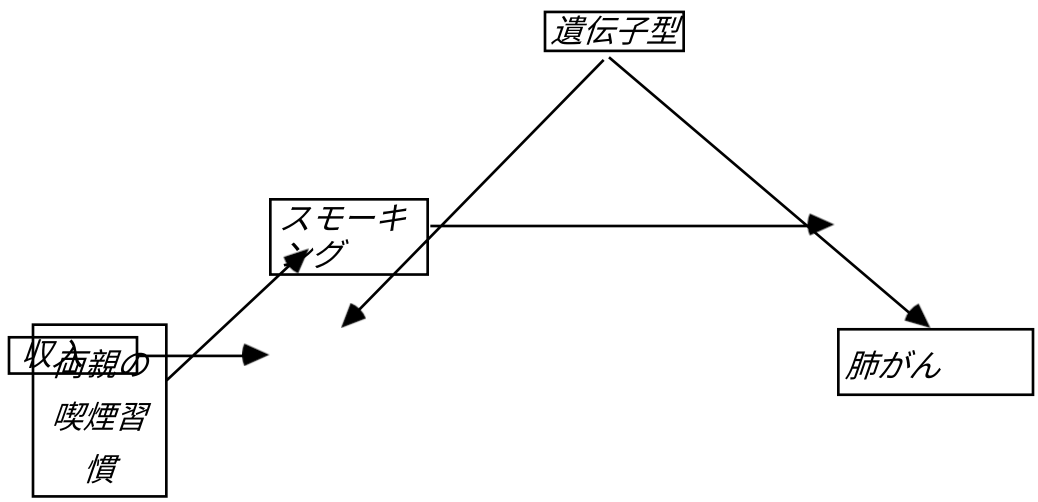


図11

部分配向誘導パスグラフでは、所得と*喫煙に*共通の測定不能原因があるのか、*両親の喫煙*習慣と喫煙に共通の測定不能原因があるのかがわからない。

未測定の原因、などなど。測定された分布は、例えば図12のようなT_1、T_2が未測定であるようないくつかの構造のいずれかによって生成されるかもしれない。

操作されたグラフにおいて、所得と*親の喫煙習慣が**Smoking**の*親にならないように*Smoking*を直接操作すると、どのグラフから限界分布が得られても、部分的に方向づけられた誘導経路グラフと操作定理から、Smokingが直接操作されると、操作された集団では、結果として図13のような因果グラフになることがわかります。



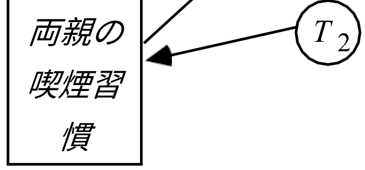


図12

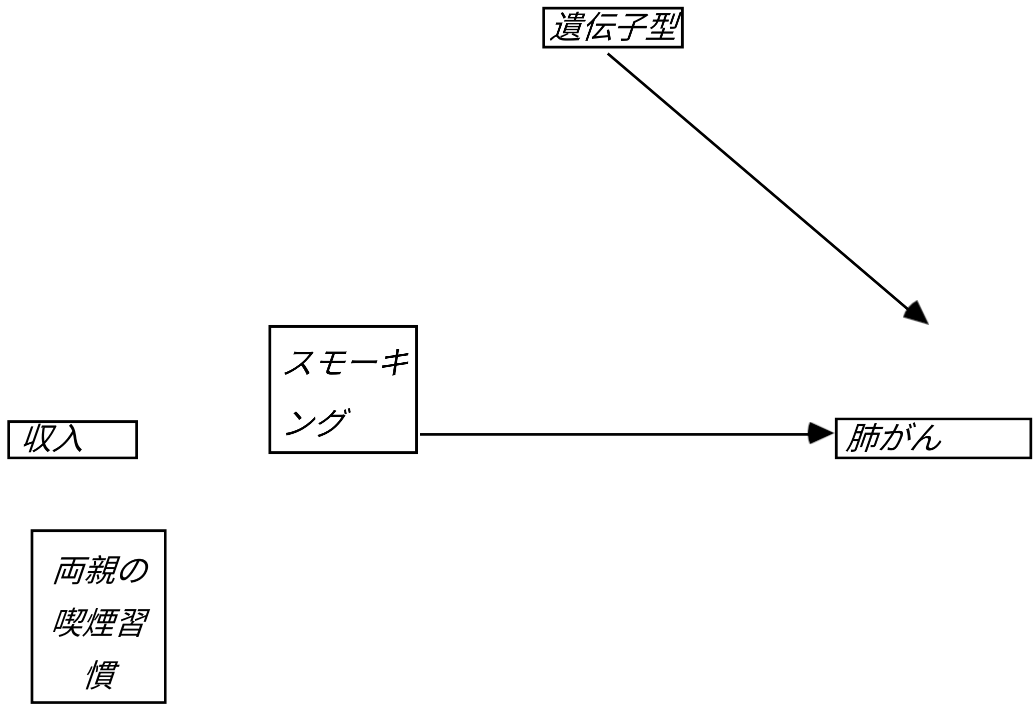


図13

この場合、*Smoking*を直接操作した場合の*Lung cancer*の分布を決定することができます。3つのステップを踏んでいる。ここでは、各ステップを実行した結果を示すだけである。各ステップがどのように行われるかは、次の章で詳しく説明する。

まず、部分配向誘導経路グラフから、操作されたグラフの共同分布を因数分解する方法を見出す。測定された変数上の分布を*PUnman*、*Smoking*を直接操作した結果の分布を*PMAN*とする。部分配向誘導経路グラフから次のように決定される。

$$PMan(I, PSH, S, G, L) = PMan(I) \quad PMan(PSH) \quad PMan(S) \quad PMan(G) \quad PMan(L \mid G, S)$$

ここで、*I*＝所得、*PSH*＝親の喫煙習慣、*S*＝喫煙、*G*＝遺伝子型、*L*＝肺がん。これは直前のグラフに対応する*PMan*の因数分解で、*Smoking*を直接操作した結果を表しています。

次に、部分配向誘導経路グラフから、先ほど与えられた結合分布の式のうち、どの因子が$P_{MAN}(L)$を計算するのに必要かを判断することができます。この場合、$P_{MAN}(I)$と$P_{MAN}(PSH)$は無関係であることが証明され、次のようになります:

$$P_{Man}(L) = \sum_{G、S} \overrightarrow{P_{Man}(S)} \times P_{Man}(G) \times P_{Man}(L|G,S)$$

第3に、部分配向誘導経路グラフから、$P_{MAN}(G)$と$P_{MAN}(L|G,S)$が対応する操作されていない確率である$P_{Unman}(G)$と$P_{Unman}(L|G,S)$とそれぞれ等しいことが判断できる。さらに、$P_{Man}(S)$は操作される量であるため、既知であると仮定される。したがって、$P_{Man}(L)$の式の3つの要素はすべて既知であり、$P_{Man}(L)$を計算することができる。

この例は、MostellerとTukeyが書いた観察からの予測に対する悲観論は、彼らが書いた時点では正当化されていたかもしれないが、根拠が十分でなかったことを示すのに十分であろう。

例でスケッチしたアルゴリズムは、以下により正式に説明される。ここでは、簡単に参照できるように、各ステップに文字でラベルを付けている。$P_{Unman}(\mathbf{V})$は操作前の分布、$P_{Man}(\mathbf{V})$は操作後の分布であり、**X**中の単一変数Xは分布$P_{Man}(X|\mathbf{Parents}(G_{Man},X))$を持つように操作されると仮定する、ここでGMANは操作後のグラフを示す。$P_{Unman}(\mathbf{V})$は操作されていないグラフG_{Unman}に忠実であり、$\mathbf{Parents}(G_{Man},X)$は既知、$P_{MAN}(X|\mathbf{Parents}(G_{Man},X))$は既知、そして$P_{MAN}$（**Y**｜**Z**）の予測に関心があると仮定します。予測アルゴリズムは、$P_{Unman}(\mathbf{O})$がグラフG_{Unman}のマルコフ条件を満たすなら、$P_{MAN}(\mathbf{O})$も同様であり、したがって$P_{Unman}(\mathbf{Y}|\mathbf{Z})$の任意の因数分解式は$P_{MAN}(\mathbf{Y}|\mathbf{Z})$の式でもある、という事実によって単純化される。グラフG'の変数の全順序*Ord*がG'に対して**許容される**のは、以下の場合だけであることを思い出してほしい。

≠Bであり、G'にAからBへの有向パスが存在する場合、AはOrdにBより先行する。が

O上のGのFCI部分配向誘導経路グラフであるときπ Xが**Definite-Non-Descendants**(**Y**)にあるのは、**Y**中の任意のYからXへの半有向経路が.Ord中に存在しないときのみである。有向無サイクルグラフGが分布Pの最小I-mapであるのは、PがGのマルコフ条件と最小性条件を満たす場合のみであることを思い出してください。

。

A). $P_{Man}(\mathbf{Y}|\mathbf{Z})$ = 不明です

予測アルゴリズム

B). P_{Unman}(**O)**から部分配向誘導パス　　　　π グラフを生成する。

C). *Ordにおける**X**の*前任者が許容される　　π forinの各変数の順序について。

は Parents($_{GMan}$,*X*) □**Definite-Non-Descendants**(*X*) C1)

に等しい。その順序付けのためのP_{Unman}(**O**)の最小I-map

*F*を形成する;

C2).Fから$_{PUnman}$(**Y**|**Z**)の式を抽出し、それをEとする;

C3.各$V \neq X$について、Eにおける項$_{PUnman}$(V|**Parents**(F,V))が$_{GMan}$において不変である場合、Xのときに

が直接操作される場合は

C3a）である。戻る　$_{PMan}$(**Y**|**Z**)　=　E',　ここで E'　は　に等しい。　に等しい。
E　を除く　その

$_{PUnman}$(X |**Parents**(F,X)) は $_{PMan}$(X |**Parents**($_{GMan}$,X)) に置き換えられます。

C3b)を終了する。

(このアルゴリズムは、変数の集合Xを操作する場合にも適用できる。**Xの**各Xについて、Xの前任者がすべて**Definite-Non-Descendants**（X）または**Parents**（$_{GMAN}$，X）にあり、**Xの**変数間に因果関係がなく、**Xの**あるXが**Xに**ない変数Vの親**ならば**、**Xの**すべてのメンバーがVの前任者となるという変数の並びを見つけることが可能なら、である。）この説明では、重要な詳細が省かれています。部分配向誘導経路グラフ（ステップB）、$_{PUnman}$(**V**)が与えられた変数の順序に対して最小条件とマルコフ条件を満たすグラフ（ステップC1）、$_{PMAN}$(**Y**|**Z**)の式E（ステップC2）、GUnmanが何か分からない時に$_{PUnman}$(**Y**|**Z**)の式に現れる与えられた条件付き確率項が$_{GUnman}$で直接操作した時に変化しないかどうかはどう判断すればよいか（ステップC3）？ 詳細は後述する。

ステップB： FCIアルゴリズムでステップB)を実施する。

ステップC$O\{X\}$内の全てのVに対して、$_{PUnman}$(V|**Parents**(F,V))が$_{GUnman}$のXの直接操作の下で不変である$_{PUnman}$(**Y**|**Z**)の式を生成すればステップC1）とC2）が*成功であるとする*。ある有向非循環グラフがC1)とC2)を成功させるような変数の順序があれば

π

、.PUnman(Y|Z)のアルゴリズムが正しいかどうかは、そのような順序に依存しないことに注意されたい。

しかし、この推測が正しくない場合、アルゴリズムは、より大きな変数順序の集合を探索する他のアルゴリズムよりも情報量が少なくなる)。

ステップC1：与えられた順序*Ord*に対して、*Ord*における*V*の前任者を**Predecessors**(*Ord*,*V*)とする。**O**上の*F*の各*V*について、**Parents**(*V*)を**Predecessors**(*Ord*,*V*)の最小の部分集合とし、*V*が**Parents**(*V*)を与えられた**Predecessors**(*Ord*,*V*)◆Parents(*V*)から独立しているようにします。そして、*F*は*P*(**O**)の最小I-mapである。Pearl (1988)を参照。*P*(**O**)が忠実な分布*P*(**V**)のマージナルであるという仮定の下で、**Parents**(*V*)が与えられたときに*V*が**Predecessors**(*Ord*,*V*)の各メンバーが*V*から独立しているかどうかをテストすることによって、**Predecessors**(*Ord*,*V*)の**Parents**(*V*)Γの独立をテストできます。これは明らかに、小さな変数の集合が**Parents**(*V*)と等しいかどうかを最初にテストすることを示唆しています。

誘導経路グラフG'と許容可能な全順序*Ord*に対して、*W*が*Ord*において*V*に先行し、かつ*U*上の終点を除く各頂点が*Ord*において*V*に先行し*U*上のコライダーであるような無向パス*U*がWと*V*の間に存在する場合にのみ*W*は**SP**（*Ord*， G', *V*）（G'における*V*の先行者を分離し、順序*Ord*とする）内にある。Gが**V**上の有向無サイクルグラフ、GIPが**O**上の*G*の誘導パスグラフ、Ordが*GIP*に受け入れられる順序、*P*（**V**）が*G*に忠実であるとすると、**O**内の各*X*について**Parents**（*X*）＝*SP*（*Ord*,*X*）が*P*（**O**）の最小Iマップとなる有向無サイクルグラフ*GMin*となる。もちろん、一般に*GIP*を与えられることはない。しかし、部分的に指向性のある誘導経路グラフを構築し、**SP**(*Ord*,*X*)の探索を絞り込む変数のセットを特定することができます。部分配向誘導経路グラフと*Ord*が許容できる場合、*V*≠Xで、Vと*X*の間に無向パス*U*が存在し、*U*上の*X*以外のすべての頂点が次のような場合にのみ*V*を**Possible-SP**（*Ord*,*X*）に入れるとする。

*Ord*における*X*の前任者であり、端点を除く*U*上のどの頂点も*U*上の**Definite**-non-colliderではない。**O**上の部分配向誘導パスグラフ

π π π π

、*Ord*を許容するための　順序付けに対して、Vは、*V*≠*X*である場合にのみ、Vと*X*との間に無向パス*U*があり、*U*上の*X*以外のすべての頂点が*Ord*における*X*の前任者で、端点を除くすべての*U*上の頂点が*U*上のColliderであるようである。

定理7.2： *P*(**O**)を**V**上の*G*に忠実な分布のマージン、**O**上の*G*の部分指向誘導経路グラフ、Ordを**O**上の部分指向誘導経路グラフで許容できる**O**の変数の順序とすると、 π *P*(**O**)の最小Iマップ$_{Gmin}$で、**Definite-SP**（*Ord*，*X*）に於いて、次のようなものが存在します。 π

は、**Parents**($_{GMin}$,*X*)に含まれ、それは、.の**Possible-SP**(*Ord* ,*X*)に含まれる。

定理7.2を*P*(**O**)の最小Iマップを探索するためのヒューリスティックとして利用するこ

π

とができます。この手順がヒューリスティックに過ぎないのは、次の理由による。

部分配向誘導経路グラフを持つ**O**上の任意の誘導経路グラフに対して許容できない順序付けを特定することはできるが、**O**上の部分配向誘導経路グラフを持つ任意の誘導経路グラフに対して許容できる順序付けが常に確実に許容できると言い切ることはできない．その場合、**Definite-SP**(*Ord*,*X*)**を**含み、**Possible-SP**(*Ord*,*X*)に含まれる集合**Mは、Mを**与えられた*X*から独立した**Predecessors** (*Ord*,*V*) \Mにならない可能性があります。

を広く検索します。

ステップC2：*P*が有向無サイクルグラフ*G*のマルコフ条件を満たす場合、次のレンマは、*P*(**Y**|**Z**)の式*E*を決定する方法を示す。(関連する結果としては、Geiger, Verma, and Pearl 1990を参照)

レンマ3.3.5：*P*が有向無サイクルグラフ*G*のマルコフ条件を満たす場合、その上に**V**、その後

$$P(\mathbf{Y}|\mathbf{Z}) = \frac{\sum\limits^{\rightarrow}_{\mathbf{iv(y,z)}\ w\ \square\mathbf{iv(y,z)}\square\mathbf{ip(y,z)}\square\mathbf{y}} \prod P(\mathbf{W}|\mathbf{Parents}(W))}{\sum\limits^{\rightarrow}_{\mathbf{iv(y,z)}\square\mathbf{y}\ \mathbf{w}\square\mathbf{iv(y,z)}\square\mathbf{ip(y,z)}\square\mathbf{y}} \prod P(\mathbf{W}|\mathbf{Parents}(W))}$$

を、因数分解の条件分布が定義され、*P*(**z**)≠0となる**Vの**すべての値について定義する。

ステップC3：以下の定理7.3、7.4を用いて、与えられた条件分布がπ*GUnmanにおけるXの*

π

直接操作の下で不変であるかどうかから判断する。**O**上の部分配向誘導経路グラフ

π

であるとき、**O**上の部分配向誘導経路グラフの無向パス*U*上の頂点Bは、Bが*Uの*端点であるか、*U*上に*A* <u>*-* *B*</u> *-* C、*A* *-* *B* -> C、または*A* <- *B* *-* *Cという*辺がある場合にのみ*U*上の**確定無接触**者となります。

O上の指向性誘導経路グラフπは、*U*上のすべてのコライダーが**Zの**メンバーへの半指向性経路の元であり、すべての定常非コライダーが**Zに**含まれない場合にのみ、**Zを**与えられたAと*B*間の可能**性のあるd-連結経路である**。 Yと**Zが不連続であれば**、*Xが***Zに**存在し、Xと**Z***{***X** }与えられた**Yの**ある*Yとの*間に*Xから*外れないd-連結経路が存在する場合のみに、*Xは***Possibly IP**(**Y**,**Z**) となる 。Yと**Zが**不連続の場合、*Xが***Zに**なく、**Zを**与えられたXと**Yの**ある*Yとの間にd*-connecting pathがあり、Xから**Y**⍰**Zの**メンバーへの半直接パスがあるときのみ、*Xは***Possibly-IV**（**Y**,**Z**）に存在することに注意

。

7.3および7.4はまた、ステップC1およびC2を成功させる変数の順序が受け入れられる有向無サイクルグラフ*G*が存在する場合、その順序が受け入れられる最小のIマップも同様であることを含意する。

定理7.3： Gが**V**?**W**上の有向無サイクルグラフ、Wが*G*の**Vに関して**外因性、Oが**Vに**含まれる、GUnmanが**V**上の*G*のサブグラフ、*GUnman*の**O**上のFCI部分指向誘導パスグラフ、YとZが**Oに**含まれる、**XがZに**含まれる、YとZが不連続、π**X**中のXが**Possibly-IP**（**Y**，**Z**）にない、なら。

P(**Y**|**Z**)は、*G*の**X**を直接操作して、**W**の値を次のように変えても不変である。

w $_1$ から **w** $_2$ へ。

定理7.4：GがV□W上の有向無サイクルグラフ、Wが*G*の**V**に関して外因性、Oが**V**に含まれる、GUnmanがV上の*G*のサブグラフ、π*GUnman*の**O**上のFCI部分指向誘導パスグラフ、**X**、**Y**、**Z**が**O**に含まれる、**X**、**Y**、Zがペア的に不π一致、**X**中のXも**Possible-IV**（**Y**，**Z**）inに含まれていなければ、*P*（**Y**｜**Z**）はか

は，W の値を **w** $_1$ から **w** $_2$ に変更することによって，*G* の **X** を直接操作しても不変である．

予測アルゴリズムは、*PUnman*(**W**)(**V**)から部分配向誘導パスグラフを構築することに基づいている。図14のモデルを考え、*X*、*Z*、*T* の関係がグラフG_1 で線形であり、*W* が政策変数である。

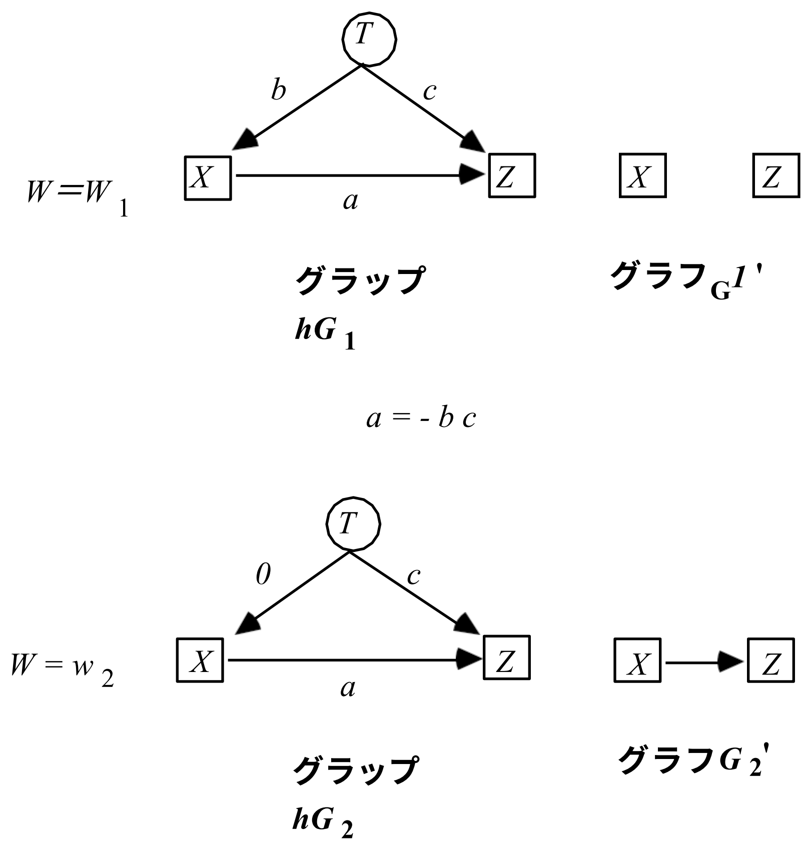


図14

$a = -bc$のとき$W = $ w1のときX, Z, Tにかかる分布はG_1に忠実ではないが、XとZにかかる分布はG_1'に忠実である。つまり、W=w1のときのXとZの分布は有向無サイクルグラフに忠実であるが、その分布を生成した因果過程のグラフには忠実でない。グラフG_2は、Wの値をw1からw2に変えることでXを直接操作した場合のモデルを描いたもので、これによりTの係数が

の式を0にし、Xに新たな分布を与える。操作されたXとZの分布はG_1'のマルコフ条件を満たさず、むしろグラフG_2'のマルコフ条件を満たし、G_1'が含まないXとZの間の辺を含んでいる。もし、操作されていないXとZの分布から部分配向誘導パスグラフを構成するとしたら、それはエッジを含まず、Zの分布は操作された分布と操作されていない分布で同じであるという予測を立てるが、それは間違いである。したがって、予測アルゴリズムは、操作されていない分布が操作されていないグラフ（結合グラフには$X \to Z$の辺が含まれるため、$X \to Z$の辺を含む）に忠実である場合にのみ、正しいことが保証される。

この仮定は、一見するとそれほど制限的なものではありません。例えば、*喫煙が癌に*与える影響について実験を行ったとしよう。各被験者に1日に吸うタバコの本数を次のように割り当てることにした。実験に参加する各被験者に対してダイスを振り、1の目が出たらタバコを吸わない、2の目が出たら1日に10本吸う、などの割り当てをする。**W**＝ {実験}、**V**＝ {*死*、*喫煙*、*飲酒*、*がん*} とする。図15は、実験対象者と非実験対象者、*GUnman* を合わせた集団の因果関係グラフである。政策変数は *Experiment* で、非実験者集団の全員に対して同じ値（0）、実験者集団の全員に対して同じ値（1）である。*Die*は、実験集団のメンバーに対して異なる値をとるので、政策変数ではない。

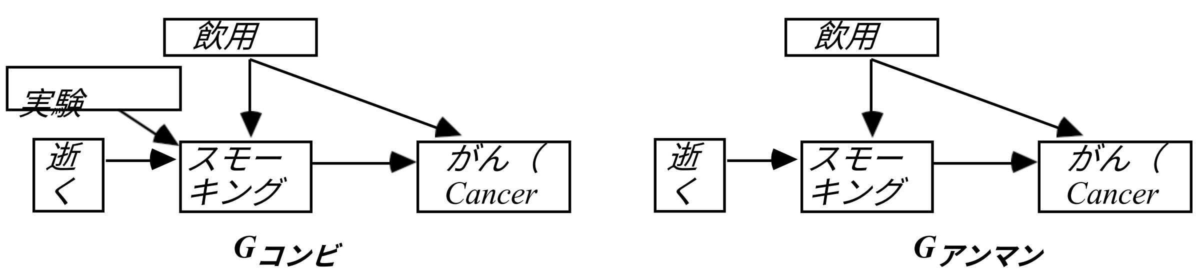
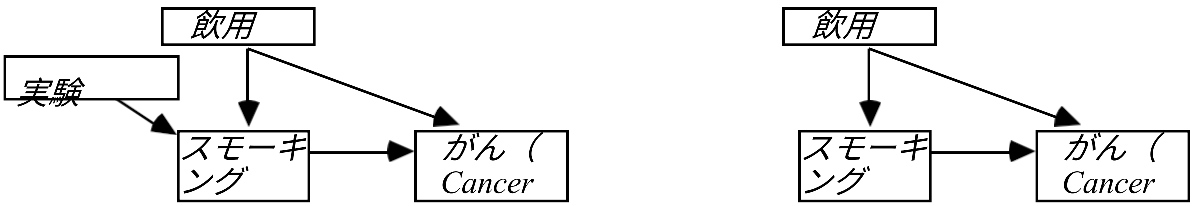


図15

この場合、ダイスの出目と被験者の喫煙本数は非実験集団では独立であるが、*GUnman*では両者の間にエッジが存在するため、*PUnman*（**V**）**は***GUnman*に忠実であるという仮定は明らかに誤りである。しかし、変数の部分集合**V'**={*喫煙*、*飲酒*、*癌*} を考えたとする。**V'**をマージナル化した結果の因果グラフは図16のようになる。この場

合、$P_{Unman}(\mathbf{V'})$はG_{Unman}に忠実である。操作された集団では*喫煙の*原因であるが、操作されていない集団ではそうでない変数は分析を複雑にするので、一般には単に考慮しないことにする。因果関係から除外しても問題はない。

グラフは、測定された変数の集合に対して、それらが操作された変数のみの直接的な原因である限り、削除することができます。これにより、それらを取り除いた後に残る変数の集合は、因果的に十分であることが保証される。



GComb　　　　**G*アンマン***

図16

定理7.5： Gが**V**□**W**上の有向非循環グラフ、Wが*G*中の**Vに関して**外因性、GUnmanが**V**上の*G*の部分グラフ、*PUnman*(**W**)（**V**）＝*P*（**V**｜**W**＝**w** **1**）が*GUnman*に忠実、**Wの**値を**w** 1から**w** **2に**変えることが*G*中の**Xを**直接操作する場合、予測アルゴリズムが正しいことになります。

予測アルゴリズムは完全ではなく、原理的に計算可能なのに*PMan*(**Y**|**Z**)が不明と言う場合があります。

7.6 例

まず、前章の仮想的な例として、図17の有向無サイクルグラフと、図18の**O** = {*所得*
π
、親の*喫煙習慣*、*喫煙*、*繊毛損傷*、*心臓病*、*肺活量*、*呼吸*機能*障害の測定値*} *に*かかる部分配向誘導パスグラフを考えてみる。PUnmanは*GUnman*に忠実であると仮定する
、

であり、操作されたグラフにおいて、所得と*親の喫煙*習慣は、親が*スモーキングをする*。Prediction Algorithmを使用して結論を出します。

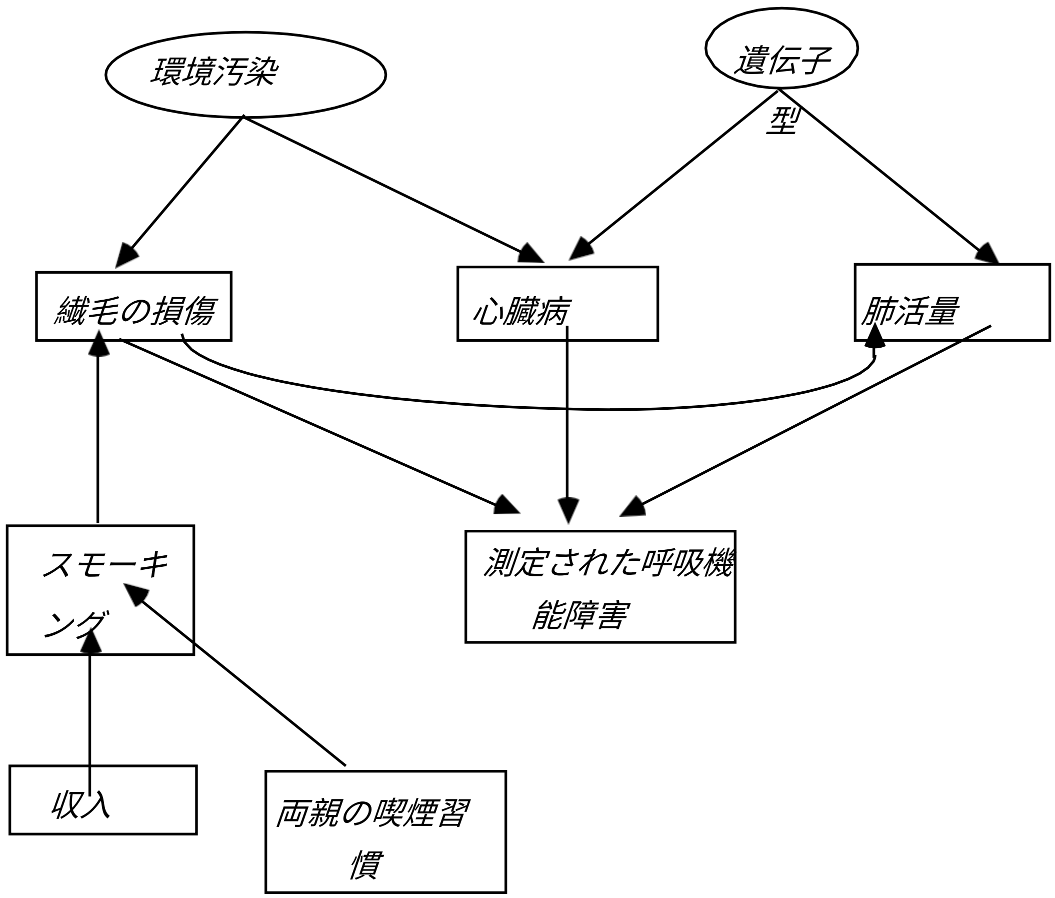

環境汚染
遺伝子型
繊毛の損傷
心臓病
肺活量
スモーキング
測定された呼吸機能障害
収入
両親の喫煙習慣

図17

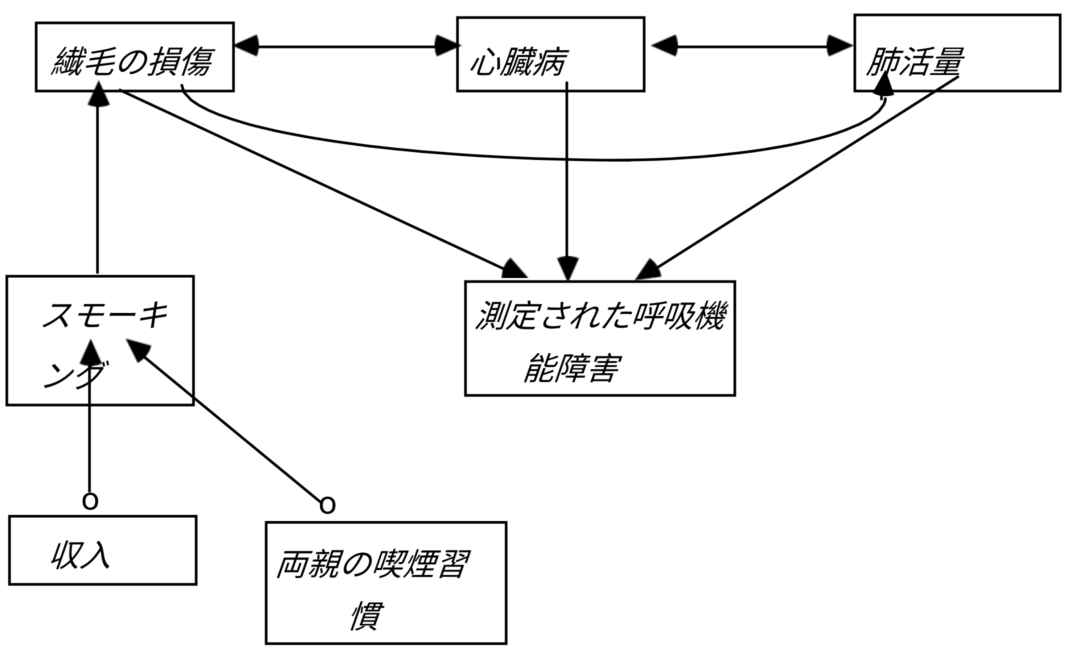

繊毛の損傷
心臓病
肺活量
スモーキング
測定された呼吸機能障害
o
o
収入
両親の喫煙習慣

図18

の共同分布の全体が一致することを決定するプロセスを、少し詳しく示すことにする。*{所得、両親の喫煙習慣*、*心臓病*、*肺活量*、*呼吸*機能*障害の測定値}*は、*喫煙を*直接操作することで予測可能である。変数名を以下のように略称することにする：

収入	I
保護者の喫煙習慣について	PSH
スモーキング	S
繊毛の損傷	C
心臓病	H
測定された呼吸機能障害	M
肺活量	L

まず、変数の順序を決めることから始めます。順序付けには2つの制約がある。まず、Sに先行する変数は、**Definite- Non-Descendant**(S)にある変数だけであり、次に、その順序は、部分指向の誘導経路グラフに受け入れられるものである。つまり、I、PSH、HがSに先行する。第二に、部分配向誘導パスグラフで受け入れられるためには、S、C、L、Mがその順序で出現しなければならない。これらの制約に適合する順序*Ord*を任意に1つ選択する：$I, PSH, H, S, C, L\ M.$ (直接操作される変数の前任者である変数間の順序は、直接操作される変数の前任者である変数だけを含む各項が常に不変であるため、問題にならないことに注意)

$P_{Unman}(I,PSH,S,C,H,M,LC)$ が最小条件とマルコフ条件を満たす有向グラフを生成する。この場合、図18の部分配向誘導経路グラフに対して許容されるあらゆる順序は、誘導経路グラフに対しても許容される順序であると判断することができる。したがって、定理7.2を適用することができる。結果として得られる因数分解は、$P_{Unman}(I)$ x $P_{Unman}(PSH)$ x $P_{Unman}(H)$ x $P_{Unman}(S|I,PSH)$ x $P_{Unman}(C|S,H)$ x $P_{Unman}(L|C,H,S)$ x

予測
$P_{Unman}(M|C,H,L)$ です。

ここで、考えている条件分布を予測するために、因数分解された分布のどの項が必要かを判断します。共同分布全体を予測するのだから、因数分解された分布のすべての項が必要であることは些細なことである。

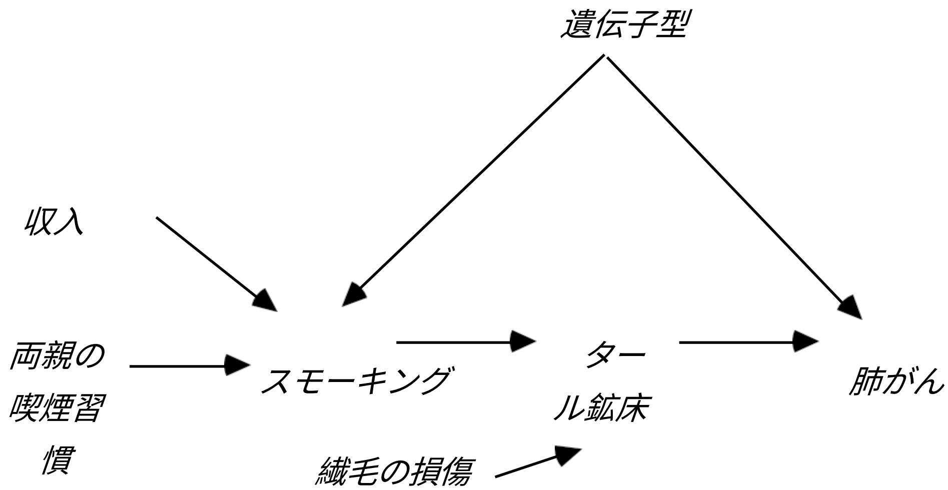
最後に、因数分解された分布の$P_{Unman}(S|I,PSH)$以外の各項がG_{Unman}のSの直接操作で不変かどうかを、部分配向誘導パスグラフを用いて検証する。$P_{Unman}(I)$、$P_{Unman}(PSH)$、$P_{Unman}(H)$は定理7.4により不変である。

はSからI,H,PSHへの半指向性のパスは存在しない。$_{PUnman}(C|S,H)$は定理7.3により不変である。なぜなら、Hが与えられたときにSからCへd-connectする可能性のあるすべてのパスは*Sから外れているからだ*。$_{PUnman}(L|C,S,H)$は定理7.3により不変である。
7.4は、C、H、Lを考えると、SとMの間にd個の接続可能な経路が存在しないため。

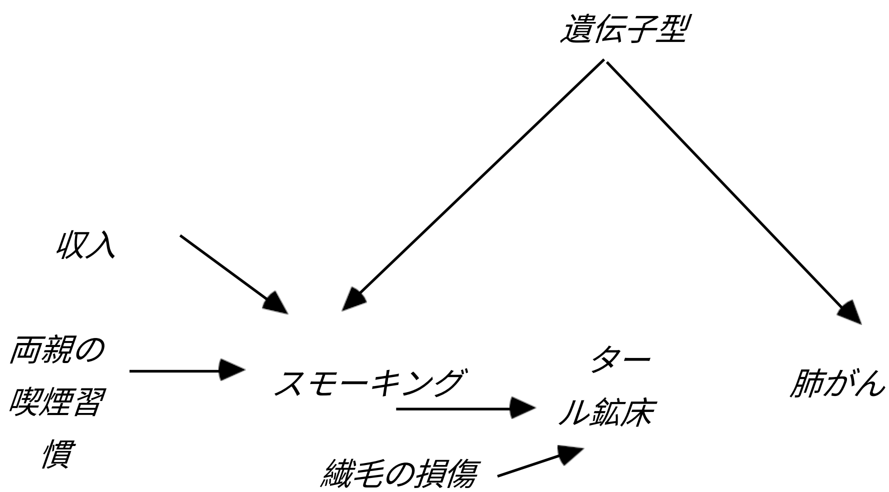
したがって、$_{PMan}(I,PSH,H,S,C,L,M) = {}_{PUnman}(I)\ {}_{PUnman}(PSH)\ {}_{PUnman}(H)\ {}_{PMan}(S)$ です。$_{PUnman}(C|S,H)\ {}_{PUnman}(L|C,H,S)\ {}_{PUnman}(M|C,H,L)$.

この場合、$_{PUnman}(I,PSH,H,S,C,L,M)$の式の中で、与えられた変数の順序に対して、$_{PMAN}(S)$を除くすべての項が$_{GUnman}$の*Smoking*を直接操作しても不変であるため、探索は単純であった。もしこのテストに失敗したら、$P(I,PSH,H,S,C,L,M)$の因数分解表現で$_{PMAN}(S)$以外の項が不変であるか、順序が尽きるまで、異なる順序の変数で処理を繰り返せばよいのである。

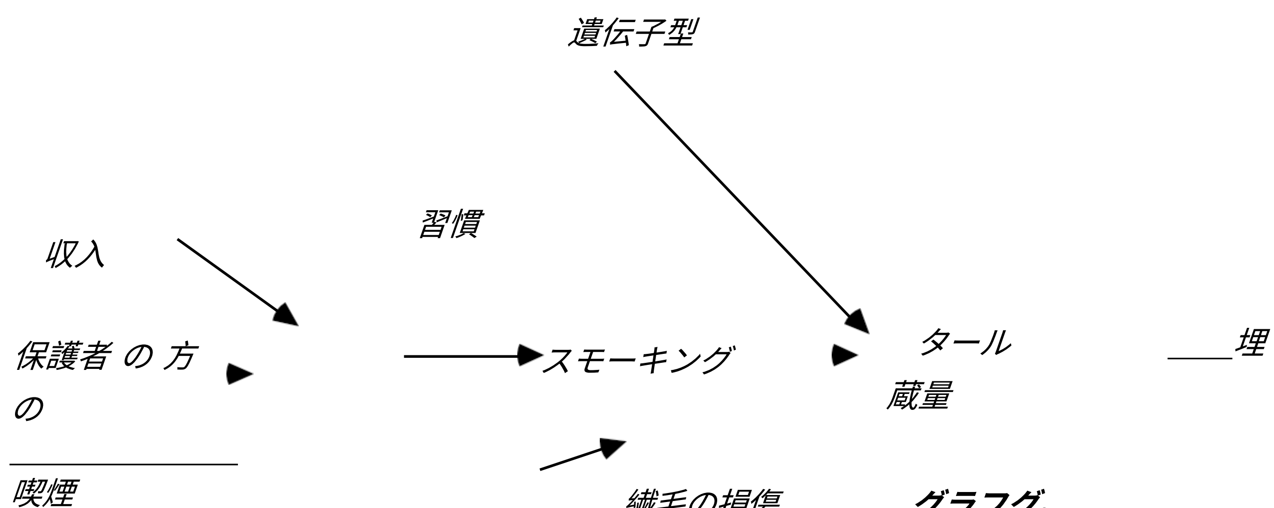
次の例として、図19の喫煙と*肺がんの関係について*、3つの代替モデルを考えてみる。G_1では、*喫煙は肺がんを*引き起こし、喫煙と*肺がんの*共通の原因がある、$_G2$では、*喫煙は肺がんを*引き起こさないが、肺がんと*喫煙の*共通の原因がある、$_G3$では、*喫煙は肺がんを*引き起こすが、喫煙と*肺がんの*共通の原因はない。



グラフグ$_{1}$



グラフG_{2}



グラフグ$_{3}$

肺がん

図19

O=上のG_1、G_2、G_3の最大情報量の部分配向誘導パスグラフを示す。

{図20に｛喫煙、*肺がん*｝を示します。

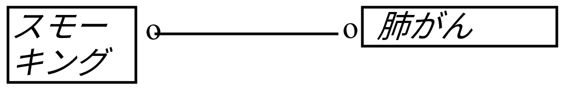


図20

この部分配向誘導パスグラフからは、*SmokingがLung cancerを*引き起こすのか（G_3のように）、*SmokingはLung cancerを*引き起こさないがSmokingと*Lung cancerの*共通の原因があるのか（G_2のように）、*SmokingがLung cancerを*引き起こすが共通の原因もあるのか（G_1のように）を判断することはできない。また、*Smoking*を直接操作した場合の*Lung cancerの*分布は予測できない。もし、変数<*Smoking,Lung cancer*>*の*順序を試すなら、Prediction Algorithmを適用するためには、*P*(*Lung cancer*|*Smoking*) がG_{Unman}*のSmokingの*直接操作の下で不変であることを示す必要があります。しかし、定理7.3を用いて*P*(*Lung cancer*|*Smoking*) が不変であることを示すことはできない。なぜなら、*Smoking* o-o *Lung cancer* edgeは、*Smokingから*外れない空集合が与えられたときにSmokingと*Lung cancerの*間にd-connecting pathが存在し得ることを保証しているからである。これはこの手法の極めて一般的な特徴で、直接操作される変数Xと*Y*の間に*X*端に "o "を持つ辺があるときは、*Yの*条件分布を予測することはできない。もちろん、この特徴はそれ自体、*P*(*Lung cancer*)が他の方法で予測できないことを示すものではありません（この例では明らかに予測できないですが）。

ただし、**O**＝｛喫煙、*肺がん*、*収入*｝*であったとする*。真のグラフをG_2とすると、部分配向誘導パスグラフは図21のようになる。

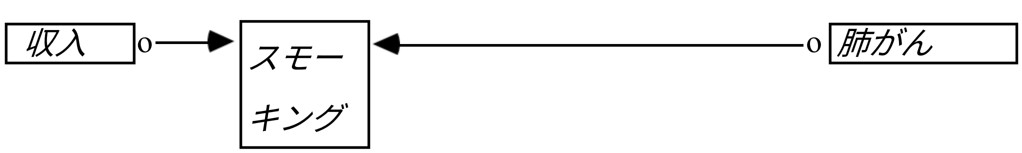
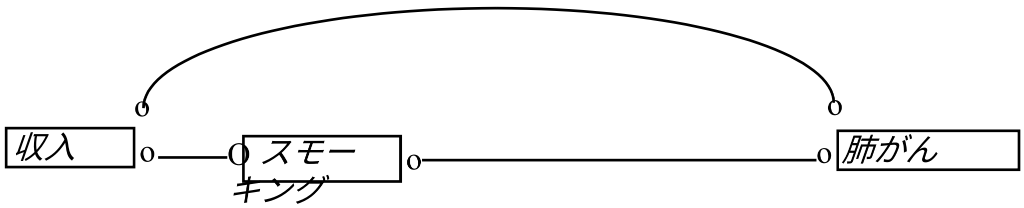


図21

前章の結果より、喫煙から*肺がんへの*半直接パスが存在しないため、*喫煙は肺がんの*原因にならないと結論づけられる。この場合、*P*(*Lung cancer*)は*GUnmanにおけるSmokingの*直接操作の下では不変であるため、*PMAN*(*Lung cancer*)は予測可能である。

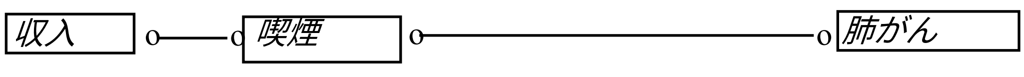
O＝ *{肺がん*、*喫煙*、*収入}* に対するG_1、G3の部分配向誘導パスグラフ（図22に示す）には、以下のように判断するための十分な情報が含まれていない。

*喫煙は肺がんを*引き起こす。それぞれのケースで*Smoking* o-o *Lung cancer*のエッジが存在するので、P_{Man}(*Lung cancer*)を予測するために予測アルゴリズムを使用することはできないことになる。



G **1の部分指向性誘導パスグラフ**

Over O = {*肺がん、喫煙、収入*}。



G **3の部分指向性誘導パスグラフ**

Over O = {*肺がん、喫煙、収入*}。

図22

真のグラフがG_3*であれば*、図23のように、部分配向誘導経路グラフで直接つながっていない2つの*喫煙の*原因も測定することで、*喫煙が肺がんを*引き起こすと判断することが可能である。部分配向誘導経路グラフにはSmokingから*Lung cancer*への有向パスが存在するので、前章の結果により、データを生成したプロセスの因果グラフにはSmokingから*Lung cancer*への有向パスが存在し、*Smoking*は*Lung cancer*の原因となる。Prediction Algorithmの出力は以下の通りである：

$$P_{Man}(\text{肺がん}) = \sum_{\text{スモーキング}} \overrightarrow{P}_{Man}(\text{喫煙})\, P_{Unman}(\text{肺がん}|\text{喫煙})$$

なお、*両親の喫煙*習慣と*所得が*無相関であることや、*Smoking*の直接の親であることは必要ない。Smokingから*Lung cancer*へのエッジは、第3の変数Vで衝突するエッジを持ち、部分配向誘導パスグラフで隣接しておらず、Vから*Smoking*への有向パスUが存在し、Uのすべてのサブパス$<X, Y, Z>$ について、X, Y, Zが三角形を形成しないような変数のペアによって配向される。

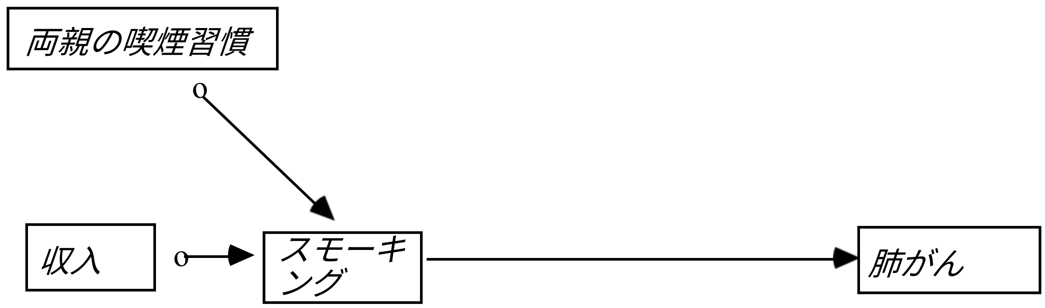


図23

残念ながら、G1が真の因果グラフである場合、*喫煙が肺がんの*原因であるかどうかを判断することはより困難である。**O**={喫煙、*肺がん*、*収入*、両親の*喫煙習慣*}で、G*1*が真の因果グラフである場合、さらなる背景知識がなければ、*喫煙が肺がんを*引き起こすかどうかを判断することはできない。図24は、部分配向誘導経路グラフにおいて、Smokingから*Lung cancer*へのエッジは、Incomeと*Parents' smoking habits*と三角形になっているため、両端に「o」がついて配向していることを示す。*Smoking* o-o *Lung cancer*のエッジの存在から、Prediction Algorithmでは*Smoking*が直接操作された場合の*P*(*Lung cancer*)を予測することはできないことがわかります。

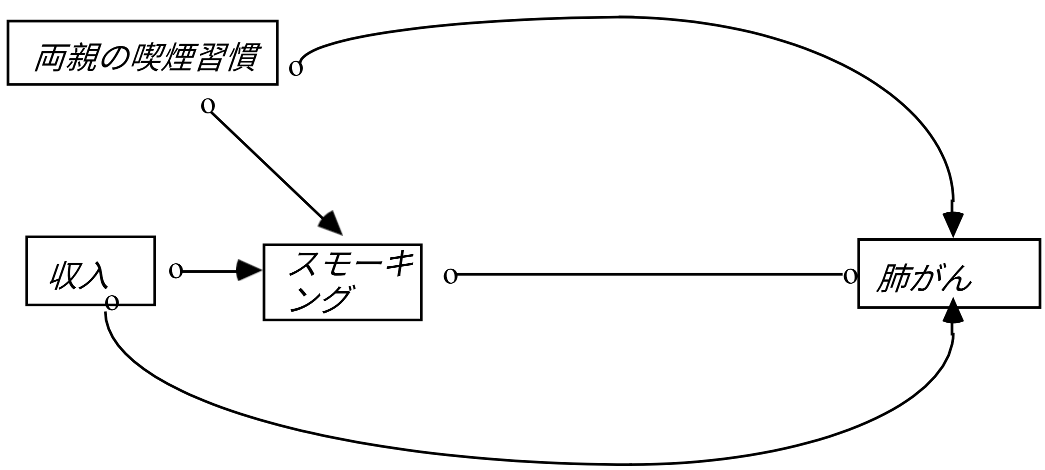


図24

*所得が肺がんを*直接引き起こすわけではないというのはもっともな話である。もし

、背景知識として、Incomeと*Lung cancer*の間に因果関係がある場合、Smokingから*Lung cancer*への因果経路が含まれていることが分かっていれば、部分配向誘導経路グラフから*Smoking*が*Lung cancer*を引き起こすと結論づけることができる。

あるいは、G1が正しいモデルであれば、喫煙と*肺がんの*間に因果関係がある*タール預金*などの変数を測定することで、*喫煙*ががんの原因であることを判定しようとするかもしれない。部分的に方向づけられた誘導パスグラフでは、所得と*肺がんの*間に誘導エッジが残っていますが、*所得*、*喫煙*、*タール*預金は三角形にはなっておらず、喫煙から*タール預金*へのエッジは方向づけられる可能性があります。残念ながら、図25が示すように、この場合、*Tar*預金と*肺癌の*間のエッジの一端が「o」の状態で配向されたままになるので、部分的に配向された誘導経路グラフはまだ*喫煙が肺癌を*引き起こすことを含意しない。また、*Smoking* o-o *Lung cancerの*エッジがあるため、*PMan*(*Lung cancer*)はPrediction Algorithmで予測することができない。

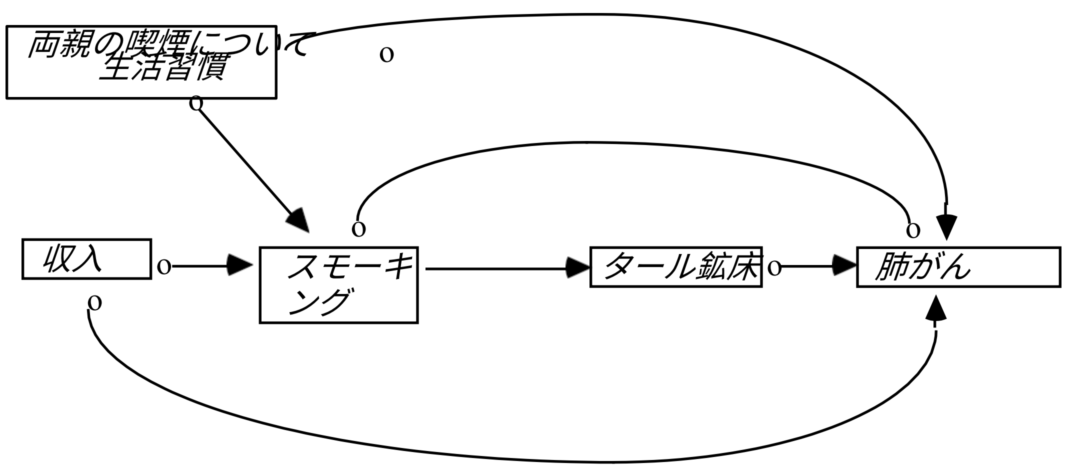


図25

しかし、G1が正しいモデルで、Smokingと*Lung cancerの*間の変数、例えば*Tar* depositsと、*Tar depositの*別の原因、例えば*Ciliadamageを*測定すると、*SmokingがLung cancerを*引き起こすと判断できる。図26参照。しかし、*Smoking* o-> *Lung cancerの*エッジがあるため、Prediction Algorithmで*PMAN*(*Lung cancer*)を予測することはできない。

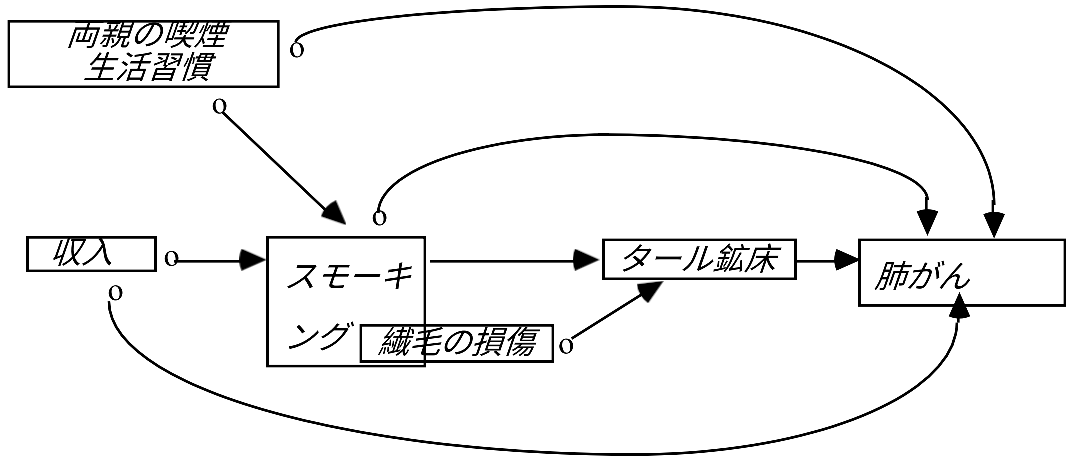


図26

また、喫煙と*肺がんの共通の原因*（この場合は*遺伝子型）をすべて測定することで*、*所得-喫煙-肺がんの*三角形を壊し、喫煙が*肺がんの*原因であると判断することができます。喫煙と*肺がんの*共通の原因をすべて測定することで、部分配向誘導パスグラフから所得と*肺がんの*間のエッジが削除される。これにより、*Income*、*Smoking*、*Lung cancer*の三角形が崩れ、図27のようにSmokingから*Lung cancer*のエッジはIncomeと*Smoking*のエッジによって方向づけられる。また、P_{MAN}(*Lung cancer*)は予測可能である。Prediction Algorithmの出力は次の通りである：

$$P_{Man}(LungCancer) =$$

$$\sum_{\text{喫煙},\text{遺伝子型}} \vec{P}_{Man}(\text{喫煙})\ P_{Unman}(\text{遺伝子型})\ P_{Unman}(\text{肺がん}|\text{喫煙},\text{遺伝子型})$$

もちろん、喫煙と*肺がんの*共通原因を*すべて*測定することは、その共通原因の多さからも、測定困難性からも（*遺伝子型の場合と*同様）、困難であろう。そのため、共通原因が1つでも未測定である限り、誘導経路グラフは*所得-喫煙-肺がんの*三角形となり、*喫煙-肺がん*間のエッジは向きを変えることができない。

図27の部分配向誘導パスグラフからは、GenotypeがSmokingと*Lung cancer*の共通の原

予測
因であるかどうかは判断できないが、Smokingと*Lung cancer*の共通の原因が*あること*

*は*判断できる。

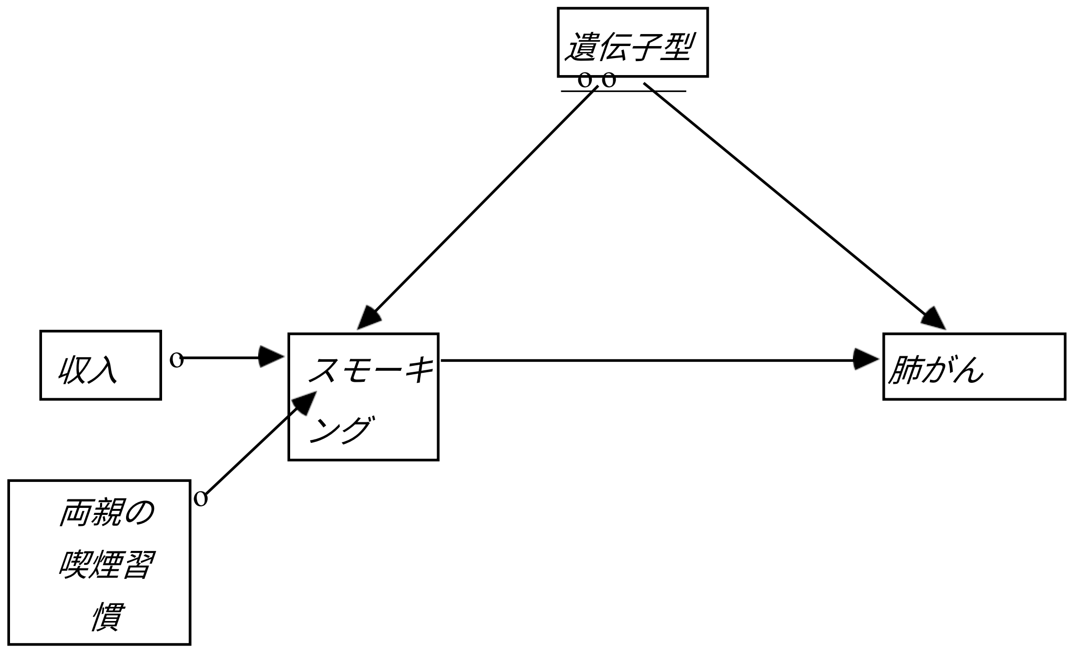


図27

7.7 結論

ここで開発された結果は、操作されていないシステムの観察から操作の効果の予測を得ることができ、制御されていない観察から実験結果の予測を行うことができるケースが存在する可能性があることを示している。次章では、実際のデータ解析の問題からいくつかの例を検討する予定である。我々の予測に関する十分条件が最大情報量に近いかどうかは分からず、この問題についてはまだかなりの理論的研究が必要である。

7.8 背景 備考

もちろん、これに適合する応用例は数多く存在するが、本章で展開される理論の先

取りは、Rubinが創始した研究の伝統の中にしか見出すことはできなかった。介入によって直接操作された一つの変数*X*がその親から独立する場合に適用される操作の定理の特別なケースは、1991年のセミナーでFienbergが独自に予想したものである。

第8章

回帰・因果・予測

回帰は特別なケースであって、特別なテーマではない。回帰研究における因果推論の問題は、前の章で検討した問題の例であり、解決策も同様にそこに見出される。回帰について特異なのは、因果推論にあまり適さない技法が、その目的のためにこれほど広く使われるようになったということだけである。

8.1 Regression Fails to Measure Influence（回帰が影響力を測定できない場合

回帰モデルは，回帰変数 **X が**結果変数 Y に及ぼす「影響」を推定するために一般的に使用される[1] 。変数間の関係が線形であれば，各 X_i について，他のすべての **X** 変数が強制的に一定である場合に X_i の単位変化によって生じるであろう Y の期待変化は，α_i というパラメータで表すことができる．X_i と Y が1つ以上の測定不能な共通原因を持つ場合，α_i の回帰推定値が不正確になることは明らかであり，広く指摘されている（たとえば，Fox, 1984）．

このようなエラーを避けるために、調査者はしばしば、潜在的な回帰変数のセットを拡大し、元の回帰係数が X_i と相関しているかどうかを判断することが推奨される（Pratt and Schlaifer, 1988）。

1線形回帰では、*Xi*の*Yに対する*"直接的な影響 "とは、(i)*Xi*の単位変化により母集団の各メンバーに生じるであろう変数*Y*の値の変化で、他のすべての**X変数は**変化しないように強制されている。例えば、(ii) **Xiの**単位変化に対する*Y*の母集団*平均*変化、(iii) *Xi*の単位変化に対する各人口の*Y*の変化、(iv) *Xi*の単位変化に対する*Y*の人口平均変化、などの意味が与えられるかも知れない。(iii)と(iv)の解釈では、*Xi*が*Y*に影響を与える他の回帰変数にも影響を与える場合、回帰係数は信頼できない推定値となる。解釈(ii)は，単位が均質で，確率的特性がサンプリングによるものであれば，(i)と等価である。そうでなければ，特別な場合，たとえば，確率変数としての線形係数が独立分布する場合（この場合も，ここで与えられた分析が適用できる（Glymour，Spartes and Scheines, 1991a））以外は，解釈 (i) で回帰は信頼できないことになる．

これは，もし交絡する共通原因があれば，それを測定して明らかにすることを期待して，回帰因子の数が安定したままであることを意味する。回帰推定値は，回帰因子の数が増えると不安定になることがよく知られているが，それは，たとえば，追加の回帰因子が，以前の回帰因子と結果変数の共通の原因であるかもしれないからである (Mosteller and Tukey, 1977)。他の回帰因子が追加されたときのXの回帰係数の安定性は，Xと結果変数に共通の原因がないことの証拠であると考えられる．

しかし、回帰変数が統計的に従属する場合、回帰変数X iと結果変数Yの測定不能な共通原因の存在が、*他の*回帰変数X_kの影響度の推定を偏らせることがあり、Yに全く影響を与えず、Yと共通原因もない変数が、有意な回帰係数を与えることがあることはあまり知られていないようです。その誤差はかなり大きくなる可能性がある。より大きな変数の集合で回帰し、安定性をチェックするという戦略は、この問題を改善するどころか、むしろ悪化させるかもしれない。また、測定された回帰候補変数の1つがYの原因ではなく、*効果である*場合にも、同様の困難が生じる可能性がある。このような状況は、コントロールされていない研究で時々発生すると思われる。

具体的には、外生変数と誤差変数がすべて無相関で共同正規分布、誤差変数の平均値がゼロ、線形係数がゼロでないことを指定します。X変数とY*だけが*測定されていると仮定する。一次方程式の各セットには、非エラー変数間の仮定された因果関係や機能依存関係を示す有向グラフが付属している。大きなサンプルでは，これらの構造のそれぞれからのデータに対して線形重回帰は，集合$\{X_1, X_2, X_3, X_5\}$のすべての変数に非ゼロの回帰係数を与えるが，X_2はこれらの構造のいずれにおいてもYに直接影響せず，X_3は構造（i），（ii）においてYに直接または間接的に影響せず，構造（iii），（iv）におけるX_3の効果は測定されていない共通の原因によって混同されている．X_2とX_3の影響に関する回帰推定値は、4つのケースすべてで不正確となる。

もし、回帰因子の仕様探索で、(i)または(ii)のX_1単独、または$X1$とX_5、あるいは(i), (ii), (iii), (iv)のX_5単独が選ばれたなら、これらの変数の回帰はY*に対する*それらの影響について一貫した、不偏の推定を与えるだろう。しかし、市販の統計パッケージの教科書的な手続きでは、これらのすべてのケースで$\{X1\}$や{X 5}を特定することができない。

$\{X5\}$ または $\{X_1, X5\}$ を適切な回帰因子のサブセットとする。

$Y = a_1 X_1 + a_2 X_5 + \varepsilon_Y$

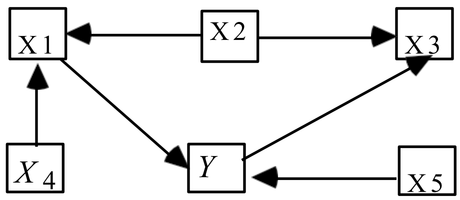
$X_1 = a_3 X_2 + a_4 X_4 + \varepsilon_1$

$X_3 = a_5 X_2 + a_6 Y + \varepsilon_3$

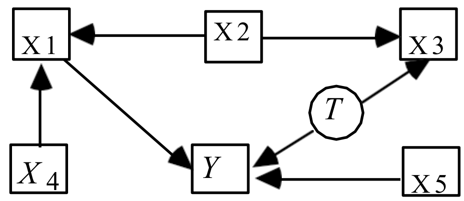
$Y = a_1 X_1 + a_2 X_5 + a_3 T + \varepsilon_Y$

$X_1 = a_4 X_2 + a_5 X_4 + \varepsilon_1$

$X_3 = a_6 X_2 + a_7 T + \varepsilon_3$



(i)



二

$Y = a_1 X_1 + a_2 X_5 + a_3 T2 + a_4 X_3 + \varepsilon_Y$

です。

$X_1 = a_5 X_2 + a_6 X_4 + \varepsilon_1$

$X_2 = a_7 T1 + \varepsilon_2$

$X_3 = a_8 T1 + a_9 T2 + \varepsilon_3$

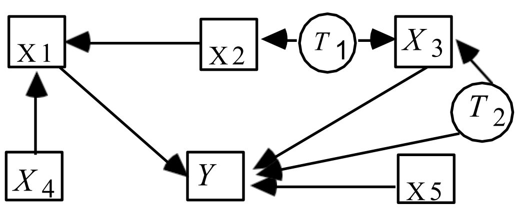
$Y = a_1 X_1 + a_2 X_5 + a_3 T2 + a_4 X_3 + \varepsilon_Y$

$X_1 = a_5 X_2 + a_6 X_4 + \varepsilon_1$

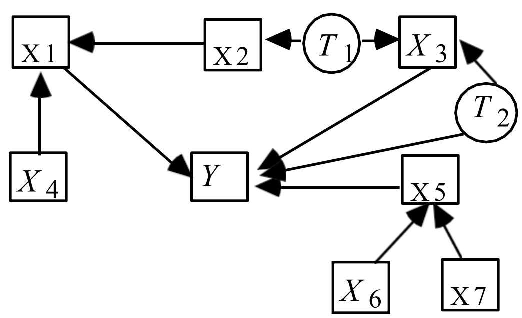
$X_2 = a_7 T1 + \varepsilon_2$

$X_3 = a_8 T1 + a_9 T2 + \varepsilon_3$

$X_5 = a_{10} X_6 + a_{11} X_7 + \varepsilon_5$



(三)



四

図1

シミュレーションで難易度の高い例を作ることは容易である。構造(i)を用いて、線形係数の値を20セット作成した。半分は正、半分は負で、それぞれ絶対値が0.5より大きい値である。係数値の各系統について、それらの値を構造(i)の係数に代入し、外生変数に無相関の標準正規分布を使用して、5,000個の無作為標本を作成した。各サンプルはMINITABに与えられ、すべてのケースでMINITABは$\{X_1, X_2, X_3, X_5\}$が有意な回帰係数を持つ回帰因子の集合であることを発見した。MINITABの

STEPWISE手続きで選択されたのは

MallowのC_pまたは調整済みR^2の最低値で選択すると、STEPWISE手順と同様の結果が得られました。

この困難は，結果変数と回帰候補変数のすべての共通原因を測定すれば改善されるが，残念ながら，回帰手法では，その条件がいつ到達したかを教えてくれるものはない．そして，余分な回帰候補変数の追加は，問題を解決するよりもむしろ問題を引き起こすかもしれない．構造 (i) と (ii) では，X_3が測定されなければ，X_2の回帰推定値は一貫していて不偏である．

XiとYが測定不能な共通原因を持つ（またはXiがYの効果である）ような、任意の回帰因子Xiと直接的に原因となる、または共通の直接測定不能な共通原因を持つ回帰因子Xkの影響度の推定に誤差が生じることになります。真の構造と係数の値によっては、誤差が非常に大きくなる可能性がある。結果変数に何の影響もない変数が、他のどんな単一の回帰因子よりも大きな標準化回帰係数を持つケースを構築するのは簡単である。カテゴリー・データでは全く平行な問題が発生する。定理3.4を思い出してください:

定理3.4： *P*があるグラフに忠実である場合、*P*が*G*に忠実であるのは次の場合だけである。

(i) *G* のすべての頂点 *X*, *Y* に対して、*X* と *Y* が *X* または *Y* を含まない *G* の頂点集合のすべてにおいて条件付きで依存する場合にのみ、*X* と *Y* が隣接すること；および

(ii) *X*が*Y*に隣接し、*Y*が*Z*に隣接し、XとZが隣接しないようなすべての頂点*X*, *Y*, *Z*に対して、*X* -> *Y* <- Zは、*X*, *Z*が*Y*を含み、*X*または*Z*を含まないすべての集合に対して従属条件である場合にのみ*G*のサブグラフになります。

この定理の最初の部分を考えると、図1の構造(i)において、回帰手順がX_2を*Y*に直接影響する変数として誤って選択する理由がわかります。構造と分布はマルコフ条件

と忠実条件を満たしていますが、線形回帰は、X iとYの偏相関が他の*すべての***X**変数を制御しても消滅しない場合に、変数X_iをYに影響すると考えます。定理3.4の(i)の部分は、回帰の基準が不十分であることを示している。定理3.4から直ちに、マルコフ条件と忠実条件を仮定すると、変数の集合**Xに対する**Yの回帰は、真の構造において、どの**X**変数もYの効果でない、またはYと共通の未測定の原因を持っていない場合、**X**変数の影響力の不偏または一貫した推定値をもたらすだけであることがわかる。

重回帰法が適用される典型的な経験的データセットには相関のある回帰因子があり、非管理的な研究では、測定されていない共通の原因が結果変数と回帰因子の両方に作用していないことを知ることはまれであるため、この問題は常在化している。このように、統計的手法の最も一般的な使い方の1つは、精巧な推測に過ぎないように思われる。

8.2 解決策とその応用例

PC、FCI、またはその他の信頼できるアルゴリズムと、前章の適切な定理を適用して、どのX変数が結果Yに影響を与えるか、与えないか、測定値から答えられないかを判断します。次に、適切と思われる方法で依存関係を推定し、前章の結果を適用してX変数を操作した場合の効果の予測を得ます。余分な理論は必要ない。経験則とシミュレーションの両方で、いくつかの例を挙げることにする。

まず、構造(i)の20個のサンプルについて、PCアルゴリズムの実装は、もちろん潜在変数がないと仮定しているが、すべてのケースで、$\{X_1, X_5\}$をYに直接影響する変数として選択していることに注目する。一方，FCIアルゴリズムの実装では，そのような仮定はなく，すべてのケースで，X_1はYに直接影響し，X_5はその可能性があり，他の変数は影響しないとしています.

図1の他の3つの構造のそれぞれにおいて、十分に大きなサンプルでは、重回帰法は、常にX_2とX_3を「有意」または「最善」または「重要」な変数に含めて、同等のエラーを起こす。これに対して、FCIアルゴリズムと定理6.5〜6.8は、次のような結果をもたらします:

構造	直接的な影響力	直接的な影響なし	未確定
二	X1	X_2、X_3、X_4	X5
三	X1	X_4	X_2、X_3、X_5
四	X_1、X_5	X_4、X_6、X_7	X_2、X_3

これらのケースでは、FCI手続きは、$X2$と$X3$がYに直接的な影響を与えないことを確実に決定するか、あるいは、それらが根拠のない直接的な影響を与えるかどうかを決定することができないと決定する。

8.2.1 武装勢力の資格の構成要素 テスト

AFQTは、アメリカ軍で使用されているテストバッテリーです。以下に挙げるような多くの構成テストがある：

算術的推理（*AR*） 数値演
算（*NO*） 単語知識（*WK*
）

さらに、以下のテストを含む多くのテストは、*AFQTの*一部ではありませんが、AFQTやその構成要素と相関があります：

数学的知識（*MK*） 電子情報
（*EI*） 一般科学（*GS*） 機械
的理解力（*MC）*

6224人の軍人を対象にしたこれら8項目のスコアをもとに、AFQTを他の7つの変数に線形重回帰すると、7つすべてに有意な回帰係数が得られ、*AFQTの*線形成分であるテストを区別することができない。共分散行列は以下の通りである。

$n = 6224$

エーエフキューティー	*ノー*	*ダブリューケー*	*拡張現実*	*エムケー*	*イーアイ*	*MC*	*GS*
253.9850							
29.6490	51.7649						
60.3604	6.2931	41.967					
57.6566	14.5143	16.0226	40.9329				
29.3763	18.2701	13.2055	20.6052	40.7386			
36.2318	2.10733	22.6958	16.3664	12.1773	63.1039		
35.8244	4.45539	17.4155	20.3952	16.459	35.1981	62.9647	

38.2510	5.61516	27.1492	14.7402	14.8442	29.9095	26.6842	48.9300

AFQTが他のどの変数の原因でもないという事前情報があれば、TETRAD IIのPCアルゴリズムは、{*AR*, *NO*, WK}を隣接する唯一の変数として正しく選び出すことができます。

AFQT の構成要素となりうる唯一の変数である。(Spirtes, Glymour, Scheines and Sorensen, 1990）.[2]。

8.2.2 スパルティィナ バイオマスの原因について

回帰に関する最近の教科書（Rawlings 1988）は、変数の関連に因果関係があると考えるのが妥当な生物学的研究に対して、回帰の原理と技法を巧みに説明している。14個の変数のうち、結果変数であるスパルティィナ草の重量に最も影響を与えるのはどれか。この例は、回帰の教科書全体に与えられた主要な応用例であるため、第13章まで読んだ読者は、その方法がこの問題に関してほとんど有益な情報をもたらさないことに驚くかもしれない。

Rawlingsによると、Linthurst（1979）は、ノースカロライナのCape Fear Estuaryの9つの地点から、それぞれ5つのSpartina草と土壌のサンプルを得た。スパルティィナの質量（*BIO*）の他に、各サンプルについて14の変数が測定された：

遊離硫化物 (*H2S*) 塩分
濃度 (*SAL*)
pH7での酸化還元電位（$_{EH}$*7*）
水中での土壌pH（*PH*）
pH6.6における緩衝酸度（
BUF） リン濃度（*P*） カリ
ウム濃度（*K*） カルシウム
濃度（*CA*） マグネシウム濃
度（*MG*） *ナトリウム*濃度

（*NA*） マンガン濃度（*MN*

） 亜鉛濃度（*ZN*） 銅濃度

（*CU*）

アンモニウム濃度($NH4$)

2実は、7つのテストがすべて*AFQTの*構成要素であるという誤った情報があり、SGSアルゴリズムで初めてそうでないことが判明しました。

相関行列は以下の通りである[3]:

ビオ	*H2S*	*SAL*	EH7	*PH*	*BUF*	*P*	*K*	*CA*	*MG*	*NA*	*MN*	*ZN*	*CU*	NH4
1.0														
.33	1 .0													
-.10	.10	1.0												
.05	.40	.31	1.0											
.77	.27	-.05	.09	1.0										
-.73	-.37	-.01	-.15 -.95	1.0										
-.35	-.12	-.19	-.	31-.40	.38	1.0								
-.20	.07	-.02	.42 .07	.	02-	-.23	1.0							
.64	.09	.09	-.04	.88 -.79		-.31	-.26	1.0						
-.38	-.11	-.01	.	30-.18	.13	-.06	.86	-.42	1.0					
-.27	.00	.16	. .06	34-.	04-	-.16	.79	-.25	.90	1.0				
-.35	.14	-.25	-.11	-.48	.42	.50	-.35	-.31	-.22	-.31	1.00			
-.62	-.27	-.42	-.23	-.72	.71	.56	.07	-.70	.35	.12	.60	1.0		
.09	.01	-.27	.09	.18	-.14	-.05	.69	-.11	.71	.56	-.23	.21	1.0	
-.63	-.43	-.16	-.24	-.75	.85	.49	-.12	-.58	.11	-.11	.53	.72	.01	1.0

データ解析の目的は、これらの変数のうち、どれが野生のスパルティナのバイオマスに最も影響を与えるかを、後の実験的研究のために明らかにすることでした。温室での実験では、野生での因果関係を推定することになる。最良のケースでは、観察研究の統計分析によって、温室でのスパルティナの生育に影響を与える変数が正しく選択されることを望むかもしれない。最悪の場合、観察研究では間違った因果構造を見つけたり、野生では成長に影響を与えるが（例えば、競合する種の成長を抑制または促進する）温室では影響を与えない変数を見つけたりすることになると思われる。

Rawlingsは、SAS統計パッケージを使用して、変数セットを重回帰で分析し、次に2つのステップワイズ回帰手順で分析しました。回帰因子の候補が大きすぎるためか

、回帰因子のすべての可能な部分集合を検索することは行われなかった。その結果は次の通りである：

3 Rawlings (1988)に記載されている相関行列は、CUとNH_4の相関を0.93と誤って記載しています。

(i) *を他のすべての変数に重回帰すると、有意な回帰係数は*K*と*CU*のみであった*

;

(ii) は、いずれも*PH*、*MG*、*CA*、*CU*を回帰変数とするモデルであり、これらの変数のみを用いた重回帰では、いずれも有意な係数が得られた;

(iii) 1変数ずつ単純回帰すると、*PH*、*BUF*、*CA*、*ZN*、$_{NH}$*4が*有意な係数を示した

。

人はどう考えればいいのだろうか。ローリングは、"どの結果も生物学者を満足させるものではなく、結果の矛盾が混乱を招き、生物学的に重要だと予想される変数が有意な効果を示さなかった "と報告しています。(p. 361).この分析はリッジ回帰によって補完され、係数の推定値の安定性を高めているが、問題となる点--重要な変数の特定--についての結果は、最小二乗法とほとんど同じである。また、Rawlingsは、主成分の因子分析と、成分のさまざまな幾何学的プロットも提供している。これらの計算は、測定された変数のどれがスパルティナの成長に影響を与えるかについての情報を提供しない。

*例えばPHはBUF*と高い相関があり、*PHの*代わりに*BUFをMG*、*CA*、*CUとともに*用いると有意な係数が得られることに注目し、ローリングスは彼の著書が扱う手順のこの利用を事実上あきらめた:

> 通常の最小二乗回帰は、すべての変数がモデル内にある場合、相関する複合体のどの変数も重要ではないと示すか、自動変数選択技術を使用する場合、複合体を表す変数の1つを恣意的に選択する傾向がある。本当に重要な変数が、その寄与が相関する変数に奪われているために、重要でないように見える

ことがある。逆に、重要でない変数が、本当の原因因子と関連しているために、重要であるように見えることがある。因果関係の有無にかかわらず、回帰結果を利用して独立変数に「相対的重要性」を付与することは、共線性の存在する場合には特に危険である。(p. 362)

Rawlingの結論は、重回帰や従来の回帰因子の選択方法については正しいが、より信頼性の高い推論手続きについては正しくない。もし私たちがPC

---

[4]SASプログラムのPROC REGの「最大R二乗」「ステップワイズ」オプションです。

PHは、他のすべての変数と区別され、*PHを*コントロールすると、他のすべての変数（*MGを*除く）と*BIOとの*相関が消失するか消滅する[6]。このアルゴリズムは、部分相関の消失の検定に.05の有意水準を用いても、0.1の水準を用いても、0.2の水準を用いても、*BIOに*隣接する唯一の変数が*PH*であることを見出した。これらの場合、PCアルゴリズムやFCIアルゴリズムでは、PHと*BIOを*直接接続することができるという結果が得られます。もしシステムが線形正規で因果マルコフ条件が成立するならば、この集団では、*PHが*一定であれば、*BIOに対する*他の回帰因子の影響は遮断されることになる。もちろん、変数の値の範囲が広ければ、*BIOが*回帰因子に線形に依存すると考える理由はほとんどないし、このサンプル内のばらつきを生み出すのに影響力のない因子は、今後も影響力を持たないだろう。また、この分析では、PHと*BIOの関係が*1つ以上の測定されていない共通の原因によって交絡されているかどうかも判断できないが、この場合の理論の原則はそうではないことを示唆している。もし、PHと*BIO*に共通の未測定の原因*Tが*あり、他の13の変数のうち*Zが*PH*を*引き起こすか、*PHと*共通の未測定の原因を持っているなら、Zと*BIOは*PH*を*条件に相関があるはずだが、そうではないようである。

このプログラムと理論から、もし*PHが*サンプルのような値（ほとんど*PH*5以下か*PH*7以上）になるように強制されれば、サンプルで証明された範囲内で他の変数を操作しても、スパルティナの成長に影響はないだろうと予想することができる。この推論は少し危険です。なぜなら、温室で制御された条件下で植物を育てることは、野生での成長に関連する変数を直接操作することにはならないかもしれないからです。例えば、自然界で*PHの*変動がスパルティナの生育に影響を与えるのは、主に温室に存在しない競合種の生育に影響を与える場合、温室での実験は、このシステムにとって*PHを*直接操作することにはならないだろう。

Linthurstの論文の第4章は、PCアルゴリズムの分析を部分的に裏付けている。Linthurstが説明する実験では、塩性湿地の小川岸からスパルティナのサンプルを採取した（観察研究で使用した場所とは異なる場所と思われる）。3×4×2（*PH*×*SAL*×*AERATION*）の4ブロックからなるランダム化完全ブロックデザインで、温室に移植した後、植物に共通の養液を与え、その養液を変化させた。

5この場合の母集団の定義は不明確であり、いずれにせよかなり狭く描かれる必要がある。6より正確には、0.05において、*MG*を除くすべての回帰因子の*BIO*との偏相関は、*PH*を含むある集合がコントロールされると消滅する。

PHとSALと*AERATION*の値です。*AERATION*変数は、この実験では重要でないことが判明しました。酸性の値は*PH*4、6、8であった。養液の*SAL*は、15、25、35、45 ‰に調整した。

Linthurst は、*PH* 6 では生育が *SAL* と変化したが、他の *PH* 値である 4 と 8 では変化しなかったが、*SAL* のすべての値で生育が *PH* と変化することを発見した (p.104) 。各変数は、植物のミネラルレベルと相関があった。リンツァーは、極端な *PH* 値が植物の成長を制御する可能性のある様々なメカニズムを考えた：

> pH4と8では、塩分濃度は本種の性能にほとんど影響を与えなかった。成長反応の決定にはpHがより支配的であったようである。しかし、pHと塩分の影響も考慮しない限り、組織濃度の高低が植物のパフォーマンスに与える因果関係を示す証拠はないようだ。(p.108)

> 両極端なpHの全体的な効果は、根に直接ダメージを与え、それによってその膜透過性とそれに続く選択的な取り込み能力を変化させることを示唆している。(p. 109).

観察データと実験データを比較すると、PC アルゴリズムの結果は基本的に正しく、2 つの手順でサンプリングした個体群の変動を通じて外挿することができるが、中立に近い *PH* 値を通じて外挿することはできない。PC 検索の結果、非実験サンプルにおいて、観察された空中バイオマスの変動は、おそらく *PH* の変動に起因しているが、他の変数の変動には起因していないことがわかった。Rawlings が報告した観察データ（p.358）では、ほとんどすべての *SAL* 測定値が 30 前後であり、極端な値は 24 と 38 である。実験的研究に比べて、野生のサンプルではむしろ制限された変動が観察された。しかし、野生で観測された *PH* の値は、両極端に集中している。6 の半

分の *PH* 単位の範囲内にある観測は 4 つだけで、5.6 から 7.1 の間の *PH* 値では全く観測されていない。PHと*SAL* の観測値については、実験結果は観察研究の結果と非常によく一致しているようだ：*PH*値が極端であれば、*SAL* の小さな変動はスパルティナの成長に影響を与えない。

8.2.3 外国投資が政治に与える影響 弾圧

Timberlake and Williams (1984)は、第三世界諸国への外国投資が独裁を促進すると主張するため、回帰分析を用いた。彼らは政治的排除（*PO*）（すなわち独裁）を測定した、

1973年の外資導入率（*FI*）、1975年のエネルギー開発（*EN*）、そして市民的自由（*CV*）。市民的自由は1～7の順序で測定され、値が低いほど市民的自由が*大きいことを示す。*72の「非中核」国に対するそれらの相関は以下の通りである：

プライベートオファーリング	*フィー*	*EN*	*かくのうようき*
1.0			
-.175	1.0		
-.480	.330	1.0	
.868	-.391	-.430	1.0

彼らの推論は不当である。彼らのモデルとSGSアルゴリズムから得られたモデルを図2.7に示す（有意水準.12で部分相関の消失を検定する）。

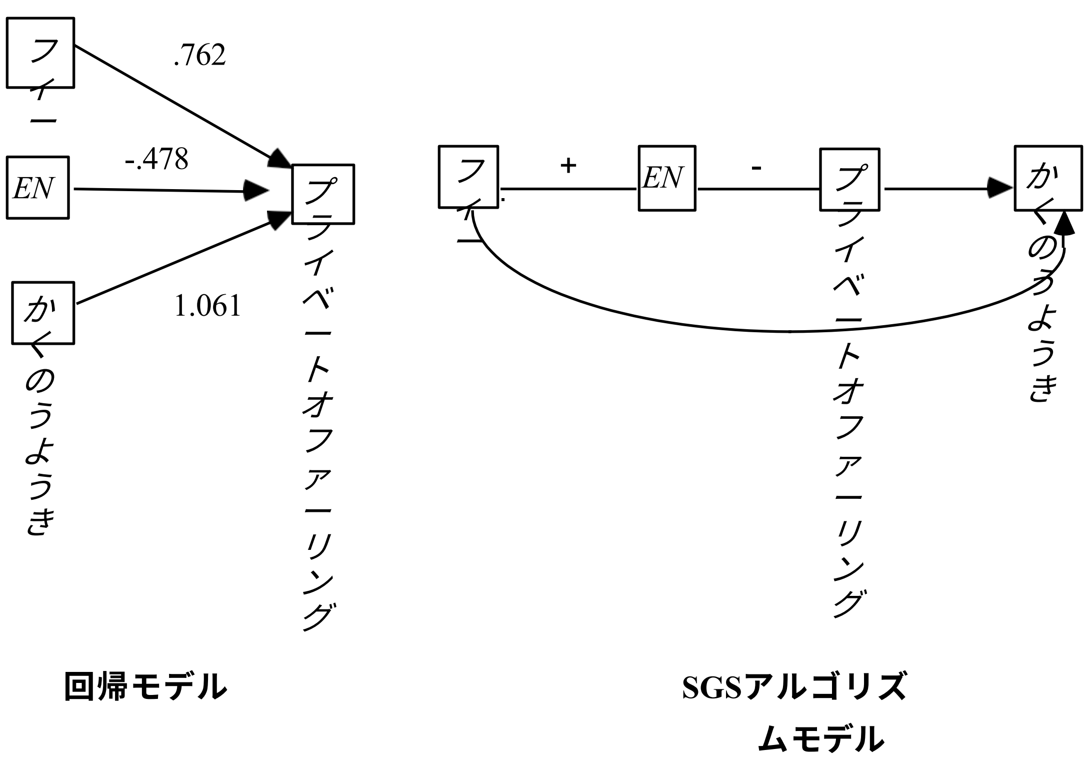


図2

SGSアルゴリズムでは、*FI-EN*および*EN*-POのエッジを方向付けることはできませんし、それらが少なくとも1つの測定されていない共通原因によるものかどうかも判断できません。SGSアルゴリズムモデルの最尤推定は、*POに対するFIの*影響（もしあれば）が負であることを必要とし、モデルはEQSプログラムによる尤度比テストに容易に合格します。SGSアルゴリズムモデルのいずれかが正しいとすると、TimberlakeとWilliamの回帰モデルは、図1の構造（i）のように、結果変数の影響を回帰変数としてとらえたケースと思われる。

このデータ分析では、測定されていない共通原因がないことを前提としています。相関関係を同じ有意水準でFCIアルゴリズムにかけると、次のような部分配向誘導パスグラフが得られます：

7有意水準が低い検索では、FIと*EN*間の隣接関係が取り除かれている。

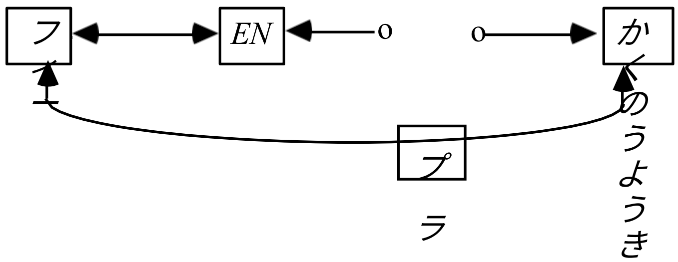


図3

このグラフは、依存関係の必要な符号とともに、外国投資とエネルギー消費は、外国投資と市民的自由のように共通の原因を持ち、エネルギー開発は政治的排除に影響しないが、政治的排除はエネルギー開発に負の影響を与える可能性があり、外国投資は政治的排除に直接または間接的に影響しない、と述べています。

8.2.4 その他のシミュレーション 研究内容

以下のシミュレーション研究では、TimberlakeとWilliamsが作成した回帰に存在すると思われるいくつかの混乱を示す図4のグラフから作成したデータを使用しました。

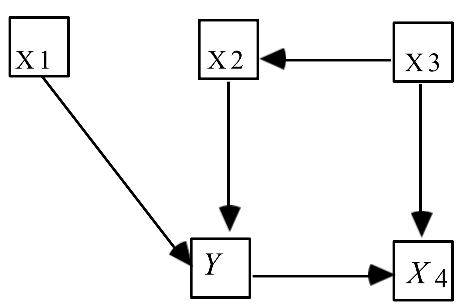


図4

線形と2値変数の離散の両方について、SGSアルゴリズムを用いてサンプルサイズ2,000と10,000のそれぞれで100回の試行が実行された。同様のセットを、線形変数と3元変数についてPCアルゴリズムで実行した。(結果は、辺の存在と方向に関するエ

ラーと、回帰因子の正しい選択について別々に採点された。図4のグラフのパターンを真のパターンと呼ぶことにする。エッジの存在に関するエラー（Co）は、変数のペアが出力パターンでは隣接しているが、真のパターンでは隣接していない場合に発生することを思い出してください。エッジの方向性エラーは、真のパターンと出力パターンの両方に存在するエッジにおいて、出力パターンには矢印があるが、真のパターンにはない場合に発生します。のエラー

omission (Om)は、それぞれのケースでアナログ的に定義されている。結果は、それぞれの種類の可能な誤りの数に対する実際の誤りの数の比率の試行分布の平均値として集計される。*Y*の実際の原因が両方（Both）正しく特定された（誤った原因はない）試行と、*Y*の原因が両方ではなく一方（One）正しく特定された（やはり誤った原因はない）試行の割合が、サンプルサイズごとに記録される:

バリアブルタイプ	#trials	n	エッジの存在		エッジ方向		両方正解	(1)正解
			コ	アム	コ	アム		
SGS								
リニア	100	2000	1.4	3.6	3.0	5.4	85.7	3.6
リニア	100	10000	1.6	1.0	2.7	2.2	90.0	7.0
バイナリ	100	2000	0.6	16.6	29.5	21.8	38.0	34.0
バイナリ	100	10000	1.2	7.4	30.0	9.1	60.0	25.0
PC								
リニア	100	2000	6.0	2.0	1.0	6.2	80.0	15.0
リニア	100	10000	0.0	1.0	2.5	2.9	95.0	0.0
ターナリー	100	2000	3.0	1.0	29.1	8.3	65.0	35.0
ターナリー	100	10000	3.0	2.0	10.8	1.2	85.0	15.0

離散データに対するSGSアルゴリズムとPCアルゴリズムの結果の違いは、前者の場合は2値変数、後者の場合は3値変数を選択したことに起因している。離散変数による統計的独立性の検定は、変数が2つ以上の値を持つ場合に、より強力になるようです。

予測や政策のために、最後の2列の数字は、シミュレーションの統計的仮定が満たさ

れ、サンプルが大きく、図4のような因果構造が得られた場合、この手順がかなり確実に結果変数の本当の原因を見つけることを示唆している。このような場合、回帰は、すべての回帰因子が結果変数に影響を与えることを発見します。

8.3 仕様の誤り確率 検索の場合

我々は、データから因果構造を特定するための様々なアルゴリズムが、必要な統計的判断が正しく行われれば正しいことを示したが、中小規模のサンプルにおける様々な種類の誤差の確率については結果を示していない。検定に関するNeyman-Pearsonの説明では、帰無仮説が真である場合に棄却される確率（タイプI）と、対立仮説が真である場合に帰無仮説が棄却されない確率（タイプII）という2つの誤差の尺度が普及している。これに対応して、ある探索手順がある標本からモデル*Mを*得たとき、そのモデル*Mが*真であれば、その探索手順がそのサイズの標本でそれを発見しない確率を求めることができ、代替案*M'*があれば、*M'*が真であれば、探索手順がそのサイズの標本で*Mを発見する*確率を求めることができる。また、探索手順の結果に対する誤り確率を、それぞれタイプI、タイプIIの誤り確率と呼ぶことにする。特に小さな標本では、条件付き独立を決定するのに用いた検定の有意水準や検出力は、探索手順における対応する誤りの確率の信頼できる指標とはならないかもしれない。

探索手順の誤差確率は解析的に求めることはほぼ不可能であり、代わりにモンテカルロ法を使用することを推奨している。ある手順でサイズ*nの*サンプルから*Mが*得られる場合、*Mを推定*し、推定したモデルを用いてサイズ*nの*サンプルをいくつか生成し、それぞれについて探索手順を実行し、*M*以外のものが見つかる頻度を数える。もっともらしい、あるいは興味深い代替モデル*M'*に対しては、*M'*を推定し、その推定モデルを用いてサイズ*nのサンプルをいくつか生成し、それぞれについて検索手順を実行*し、*Mが見つかる頻度をカウントする*。ここでは、タイプIIエラーの確率が非

常に高いケースを例に、仕様検索のエラー確率の決定を説明する。

Weisberg,(1985)は、回帰が異常な結果をもたらす実験的研究を用いて、外れ値を検出する手順を説明する：

> ラットの肝臓に存在する特定の薬物の量を調べるための実験が行われた。19匹のラットを無作為に選び、体重を測定し、軽いエーテル麻酔をかけ、薬剤を経口投与した。大きな肝臓は小さな肝臓よりも与えられた用量をより多く吸収すると考えられたため、動物が実際に受けた用量は、体重1キログラムあたり40mgの薬物とほぼ決定されました。

実験仮説は、投与量の決定方法について、肝臓に占める投与量の割合（Y）と体重（X_1）、肝臓重量（X_2）、相対投与量（X_3）との間に関係がないことである。(p. 121-124)

Yを(X_1, X_2, X_3)に回帰すると、仮説と一致しない結果が得られる。Yの係数は体重(X_1)と投与量(X_3)に対してともに有意であり、一方は他方によって決定されているにもかかわらずである。Weisbergのデータに対して次のような回帰値が得られる（標準誤差は括弧付きで係数の下に、t統計は標準誤差のすぐ下に示されている）：

$$\begin{array}{lllll} Y & = -3.902*x_1 + & .197*x_2 + & 3.995*x_3 + & \varepsilon \\ & (1.345) & (.218) & (1.336) & \\ & -2.901 & .903 & 2.989 & \end{array}$$

X_2を含まない重回帰では、X_1とX_3についても有意な回帰因子が得られる。

.05レベルである。しかし、ワイズバーグは、YをXのどれか1つに回帰させることはできないとしている。

はその水準で有意である。したがって、いくつかの統計的判断の結果は矛盾しており、例えば、$\rho_{X1Y}=0$, $\rho_{X2Y}=0$, $\rho_{X3Y}=0$ だが、$\rho_{X1Y.X3} \neq 0$.つ

このような矛盾については、いくつかの見解がある。一つは、使用された特定の有意水準が大きく影響しているというものである。相関の消失と偏相関の仮説を棄却するために0.01水準が使われた場合、Yとの相関は消失し、他の1つの変数をコントロールする偏相関も消失する。しかし、X_1とYの偏相関を他の回帰因子の両方で制御しても棄却できず、矛盾が残ることになる。また、連続した統計的判断の結果に矛盾が生じることは、特に小さなサンプルでは予想されることであり、可能であれば、最も信頼できる統計的判断に基づいて推論を行うべきだという考え方もあります。今回のケースでは、サンプルサイズの関係でどの統計的検定も検出力が低いですが、偏相関の順位が低いほど検出力が高くなります。経験則では、余分な変数をコントロールすることは、データポイントを捨てることである。したがって、この場合、PCアルゴリズムは偏相関を考慮せず、消失相関のみから、*X変数のどれもが*Y変数を引き起こさないと結論づけます。Weisbergは代わりに19個のデータポイントのうち1個を除外することを推奨し、その後残りのデータを使って重回帰を行うと、有意な回帰係数は得られない（.05）。

実験設定から、体重と肝臓の重量は投与量に因果的に先行し、それ自体が結果、すなわちラットの体内で見つかった薬物の量に先行すると仮定できる。