

グラフィカルモデルによる統計的因果推論

宮川 雅巳*

Statistical Causal Inference Using Graphical Models

Masami Miyakawa*

確率変数を頂点とする無向グラフ, 有向グラフあるいは連鎖グラフで表される統計モデルをグラフィカルモデルという。このモデルは変数間に成り立ついくつかの独立性・条件付き独立性を規定する。グラフィカルモデルをデータから推測する統計的作業がグラフィカルモデリングである。対象とするデータ形式は, 複数の個体の各々について複数の変数が測定された多変量データであるので, 多変量解析法の1つといえる。本論文では, 変数間の先行関係に関する知識と多変量データがもつ関連情報から, 有向グラフあるいは連鎖グラフで表現される統計モデルを推測する方法について述べる。この過程で基本的な役割を果たすのが無向グラフで表現される統計モデルであることを強調する。得られたグラフの矢線に因果的関係の解釈が可能となるとき, そのグラフを因果ダイアグラムと呼ぶ。因果ダイアグラムの利用方法として, 介入効果とそれを識別するための共変量選択基準, 多元分割表での交絡因子の同定と併合可能の判定基準を紹介する。

1. はじめに

統計的因果推論は, 研究者のもつ定性的因果仮説に統計的データを組み合わせることで, 興味ある変数間の因果的関係について定量的評価を行う方法論である。この方法の起源は, Wright[58], [59]の「パス係数の方法」に遡ることができる。これは現在, パス解析あるいは構造方程式モデルと呼ばれており, その内容は

- ・パスダイアグラムによる定性的因果仮説の表現
- ・パス係数による変数間の相関係数の分解
- ・直接効果と総合効果の峻別

からなる (Bollen[4]p. 32)。構造方程式モデルは統計的因果モデルであり, パス係数とは, 単なる関連の記述的尺度でなく「因果的関係」を表したものであることに注意を要する。すなわち, パス係数による方法とは, データを生み出す因果過程をパスダイアグラムが正しく表現しているときに限り, 因果的関係を相関の測定から推測する方法である。この大前提が理解されていないために, パス解析あるいは構造方程式モデルはたびたび不適切に用いられきたことが指摘されている (Freedman[15], Pearl[38])。

また, 統計的因果モデルには, パス解析とは異なるストリームがある。それは, Neyman-Rubinの潜在反応モデルである (Neyman[34], Rubin[45])。このルーツは, Fisher[14]が考案した統制された無作為化実験での要因効果の考え方である。この要因効果の実質的意味を“counterfactual” (仮定法過去) の概念を用いて定式化したのが Neyman-Rubin の因果モデルである (たとえば Holland[22], 田口[52]4.5節)。Rosenbaum and Rubin[43]では, 無作為化実験での要因効果を観察研究において推測するための試みがなされている。なお, このストリームにおい

* 東京工業大学 大学院社会理工学研究科 経営工学専攻, 〒152-8552 東京都目黒区大岡山 2-12-1

では、観察研究で推測対象とする因果的關係は、処理と反応との關係に限定されている（たとえば Rosenbaum[42]1.1節）。処理とは人為的介入であり、まさに“counterfactual”な状況を想定できるものである。

一方、グラフィカルモデルは、純粋な確率論的立場から独立性・条件付き独立性を規定した統計的関連モデルとして出発した。それは、分割表における Birch[1]の対数線形モデルと多変量正規分布における Dempster[12]の共分散選択モデルの統合と、無向グラフ表現を利用した両者での独立性・条件付き独立性の精緻化である。多変量データに基づき、グラフィカルモデルを統計的に推測する統計的方法としてグラフィカルモデリングが開花した。この側面からのまとまったテキストとして、Whittaker[57]、Edwards[13]がある。

さて、統計的因果モデルと統計的関連モデルを結びつける上で、重要な役割を果たしたのは有向グラフあるいは連鎖グラフで表現される統計モデルである。ただし、有向グラフのモデルも独立性・条件付き独立性という確率論的性質を規定するだけで、因果關係を規定しない。この立場からのテキストとして、Cox and Wermuth[8]、Lauritzen[25]が既に刊行されている。和書では、宮川[27]、日本品質管理学会テクノメトリックス研究会編[35]がある。

一方で、グラフでの矢線に因果的解釈ができることを前提として、有向グラフを利用した統計的因果推論を展開する立場がある。この場合の有向グラフは因果ダイアグラムと呼ばれる。この立場では、Pearl[36]、Spirtes, Glymour and Scheines[50]のテキストが代表的である。統計的因果分析に有向グラフを取り込むことで、構造方程式モデルと潜在反応モデルを結びつける試みもなされている（Galles and Pearl[16]）。

本論文では、あくまで確率論的性質を規定した有向グラフの統計的推測に重点を置き、そこでの矢線に因果的解釈ができるかどうかは研究者の判断に任せるという立場をとる。この立場こそが「因果關係とは何か」「因果律とは何か」に深く立ち入らずに、統計的因果推論を客観的な統計的方法にする公理論的アプローチと考えるからである。これは、確率の現実的な意味を問わない公理論的確率と同じ立場である。この立場の上で、有向グラフの矢線に因果的解釈が可能なときの因果ダイアグラムの積極的な利用方法を論じる。

本論文の以下の構成は次の通りである。2章では、条件付き独立性を規定する統計モデルについて概説する。初めに、条件付き独立性に関する基本的な性質を述べ、変数間の条件付き独立性を無向グラフで記述したときのマルコフ性について説明する。次に、多変量正規分布での共分散選択モデルと多元分割表での対数線形モデル、およびそれらの統計的推測法であるグラフィカルモデリングについて述べる。3章では、有向グラフあるいは連鎖グラフで規定される統計モデルとその推測方法について述べる。ここでは、変数間の先行関係、順序関係に関する知識の程度に応じて、異なる手順がとられることを説明する。4章では、因果ダイアグラムが与えられたときに、これに基づく介入効果とその識別可能条件、交絡因子の同定とオッズ比の併合可能条件などについて述べる。5章では、残された課題、今後の展望について考察する。

2. 条件付き独立性を規定する統計モデル

2.1 条件付き独立性の基本的性質

3つの確率変数 X, Y, Z を考える。これらの同時密度関数を $f_{XYZ}(x, y, z)$ と記す。離散変数のときは同時確率関数とする。この同時密度関数から導かれる周辺密度関数を $f_X(x), f_{XY}(x, y)$ などと記す。また、これらから導かれる条件付き密度関数を $f_{Y \cdot X}(y|x), f_{XY \cdot Z}(x, y|z)$ などと記す。

x, y, z の取り得る任意の値に対して

$$f_{xyz}(x, y, z) = f_x(x)f_{yz}(y, z) \quad (2.1)$$

が成り立つとき、 X と (Y, Z) は独立であるといい、この関係を $X \perp\!\!\!\perp (Y, Z)$ と表記する。 $X \perp\!\!\!\perp (Y, Z)$ ならば、 $X \perp\!\!\!\perp Y$ かつ $X \perp\!\!\!\perp Z$ であるけれども、逆は一般には成り立たない。次に、 x, y, z の取り得る任意の値に対して

$$f_{xy \cdot z}(x, y|z) = f_{x \cdot z}(x|z)f_{y \cdot z}(y|z) \quad (2.2)$$

が成り立つとき、 X と Y は Z を与えたときに条件付き独立であるといい、この関係を $X \perp\!\!\!\perp Y|Z$ と表記する (Dawid[10])。 (2.2) 式は

$$f_{xyz}(x, y, z) = \frac{f_{xz}(x, z)f_{yz}(y, z)}{f_z(z)} \quad (2.3)$$

と書き改めることができる。

【定理 1】 因数分解基準

$X \perp\!\!\!\perp Y|Z$ の必要十分条件は、同時密度関数 $f_{xyz}(x, y, z)$ に対して

$$f_{xyz}(x, y, z) = g(x, z)h(y, z) \quad (2.4)$$

を満たす関数 g と h が存在することである。 □

以下では、 x, y, z の取り得る任意の値に対して $f_{xyz}(x, y, z) > 0$ とする。

【定理 2】

次の 4 つの内容は同値である。

- 1) $X \perp\!\!\!\perp Y|Z$ かつ $X \perp\!\!\!\perp Z|Y$
- 2) $X \perp\!\!\!\perp Y|Z$ かつ $X \perp\!\!\!\perp Z$
- 3) $X \perp\!\!\!\perp Y$ かつ $X \perp\!\!\!\perp Z|Y$
- 4) $X \perp\!\!\!\perp (Y, Z)$ □

この定理 2 は、グラフィカルモデルのマルコフ性を示す上で重要な役割を果たす。証明は例えば Lauritzen[25]3.1 節にある。これから次の系が導かれる。確率変数 W を新たに追加し、 $f_{xyzw}(x, y, z, w) > 0$ を仮定する。

【系 1】

次の 4 つの内容は同値である。

- 1) $X \perp\!\!\!\perp Y|(Z, W)$ かつ $X \perp\!\!\!\perp Z|(Y, W)$
- 2) $X \perp\!\!\!\perp Y|(Z, W)$ かつ $X \perp\!\!\!\perp Z|W$
- 3) $X \perp\!\!\!\perp Y|W$ かつ $X \perp\!\!\!\perp Z|(Y, W)$
- 4) $X \perp\!\!\!\perp (Y, Z)|W$ □

以上の性質は、確率変数を確率ベクトルに拡張したときにも成り立つ。

2.2 無向グラフで規定されるモデル

p 個の確率変数の集合を $V = \{X_1, X_2, \dots, X_p\}$ とする。この V を頂点集合とし、その直積 $V \times V$ のある部分集合 E を無向の辺の集合とするグラフ $G = (V, E)$ を考える。

V の部分集合 c において、 c の任意の頂点間に辺があるとき、 c は完全であるという。 c は完全で、かつ、 c に任意の頂点を加えた集合が完全でないとき、 c はクリークであるという。グラフ G のクリークの集合を $C = \{c_1, c_2, \dots, c_q\}$ と記す。クリーク c_j に含まれる変数のみを要素とする確率ベクトルを X_{c_j} と表記する。

【定義1】 無向独立グラフ

無向グラフ $G=(V, E)$ が $V=\{X_1, X_2, \dots, X_p\}$ の同時密度関数をクリークに従って因数分解されている形, すなわち

$$f_V(x_1, x_2, \dots, x_p) = \prod_{c_j \in C} g_j(\mathbf{x}_{c_j}) \quad (2.5)$$

の形に規定するとき, その無向グラフを無向独立グラフという。□

【例1】 図1に示す無向独立グラフを考える。クリークは $c_1=\{X_1, X_2, X_3\}$, $c_2=\{X_3, X_4, X_5\}$ である。この無向独立グラフが規定する同時密度関数のクリークに従う因数分解は

$$f_V(x_1, x_2, \dots, x_5) = g_1(x_1, x_2, x_3) g_2(x_3, x_4, x_5) \quad (2.6)$$

である。□

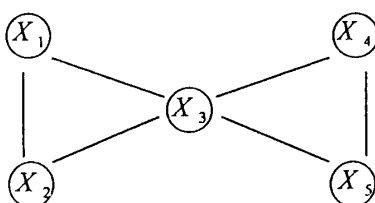


図1 無向独立グラフの例

さて, 無向独立グラフで成立している条件付き独立性を完全に記述するために, いくつかのグラフ用語を確認する。辺で直接結ばれた頂点は隣接しているという。ある頂点 X と隣接している頂点集合を境界といい, $bd(X)$ と記す。辺のつながりを道という。道で結ばれた頂点は連結しているという。2つの頂点 X_i と X_j を結ぶ任意の道が頂点集合 s のある要素を通るとき, s は X_i と X_j を分離しているという。2つの頂点集合 a と b において, 任意の $X_i \in a$ と任意の $X_j \in b$ を s が分離しているとき, s は a と b を分離しているという。

【定理3】 無向独立グラフでのマルコフ性

無向独立グラフでは次の3種類のマルコフ性が成り立ち, これらは同値である。

- 1) 対ごとのマルコフ性: 2つの頂点 X_i と X_j が隣接していなければ

$$X_i \perp\!\!\!\perp X_j \mid V \setminus \{X_i, X_j\} \quad (2.7)$$

が成り立つ。すなわち, X_i と X_j は残りの $p-2$ 個の変数を与えたときに条件付き独立である。

- 2) 局所的マルコフ性: V から X とその境界 $bd(X)$ を除いた頂点集合を X の非閉包と呼び, $nc(X)$ と表記したとき

$$X \perp\!\!\!\perp nc(X) \mid bd(X) \quad (2.8)$$

が成り立つ。すなわち, X の境界を与えたとき X は X の非閉包と条件付き独立である。

- 3) 大域的マルコフ性: たがいに排反な頂点集合 a, b, s において, s が a と b を分離していれば

$$a \perp\!\!\!\perp b \mid s \quad (2.9)$$

が成り立つ。すなわち, 排反な頂点集合はそれらを分離する頂点集合を与えたとき条件付き独立である。□

定義1からこれら3つのマルコフ性が導けること、およびこれら3つのマルコフ性が同値であることは、定理1、定理2および系1を用いることで示せる。

【例2】 図1の無向独立グラフにおいて、対ごとのマルコフ性より

$$\begin{aligned} X_1 \perp\!\!\!\perp X_4 | (X_2, X_3, X_5), & \quad X_1 \perp\!\!\!\perp X_5 | (X_2, X_3, X_4) \\ X_2 \perp\!\!\!\perp X_4 | (X_1, X_3, X_5), & \quad X_2 \perp\!\!\!\perp X_5 | (X_1, X_3, X_4) \end{aligned}$$

が成り立っている。次に局所的マルコフ性からは

$$\begin{aligned} X_1 \perp\!\!\!\perp (X_4, X_5) | (X_2, X_3), & \quad X_2 \perp\!\!\!\perp (X_4, X_5) | (X_1, X_3) \\ X_4 \perp\!\!\!\perp (X_1, X_2) | (X_3, X_5), & \quad X_5 \perp\!\!\!\perp (X_1, X_2) | (X_3, X_4) \end{aligned}$$

が読みとれる。最後に大域的マルコフ性からは

$$(X_1, X_2) \perp\!\!\!\perp (X_4, X_5) | X_3$$

という関係を得る。□

【系2】 無向独立グラフで2つの頂点 X_i と X_j が連結していなければ、すなわち両者を結ぶ道がないならば、 $X_i \perp\!\!\!\perp X_j$ であり、かつ、 $V \setminus \{X_i, X_j\}$ の任意の部分集合 a に対して $X_i \perp\!\!\!\perp X_j | a$ である。□

この系2の前半は、2つの頂点 X_i と X_j が連結していなければ $X_i \perp\!\!\!\perp X_j$ あることを主張しているけれども、この逆は一般に成り立たない。 $X_i \perp\!\!\!\perp X_j$ であっても、無向独立グラフで X_i と X_j が隣接していることがある。

2.3 共分散選択モデル

多変量正規分布は、平均ベクトル μ と分散共分散行列 Σ によって完全に記述される分布である。いま、 $\Sigma = (\sigma_{ij})$ に対して、その逆行列を $\Sigma^{-1} = (\sigma^{ij})$ と表記する。 X_i と X_j 以外のすべての変数を与えたときの X_i と X_j の母偏相関係数を $\rho_{ij \cdot rest}$ と記す。ここで“rest”は $V \setminus \{X_i, X_j\}$ の意味である。 $\rho_{ij \cdot rest}$ は $\Sigma^{-1} = (\sigma^{ij})$ の要素によって

$$\rho_{ij \cdot rest} = -\frac{\sigma^{ij}}{\sqrt{\sigma^{ii}} \sqrt{\sigma^{jj}}} \quad (2.10)$$

と表現される。よって「 $\sigma^{ij} = 0$ 」と「 X_i と X_j 以外のすべての変数を与えたとき X_i と X_j は条件付き独立である： $X_i \perp\!\!\!\perp X_j | rest$ 」は同値になる。そこで、多変量正規分布において、いくつかの σ^{ij} を0とおいた統計モデルを考えると、そのモデルはいくつかの条件付き独立性を与えるものになる。そのようなモデルを共分散選択モデルといい、Dempster[12]によって提案された。なお、分散共分散行列 Σ から導かれる母相関行列を $\Pi = (\rho_{ij})$ 、その逆行列を $\Pi^{-1} = (\rho^{ij})$ すれば、(2.10) 式は

$$\rho_{ij \cdot rest} = -\frac{\rho^{ij}}{\sqrt{\rho^{ii}} \sqrt{\rho^{jj}}} \quad (2.11)$$

と書き改めることができるので、相関行列から議論を始めてよい。

共分散選択モデルと無向独立グラフの対応関係は明らかである。定理3の「対ごとのマルコフ性」より、モデルで $\rho^{ij} = 0$ とした X_i と X_j を辺で結ばず、そうでない変数対を辺で結んだグラフが無向独立グラフに他ならない。

一般に、統計的モデリングのプロセスは

- ① モデルを設定する

- ② モデルの母数を推定する
- ③ モデルのデータへの適合性を評価する
- ④ モデルの変更を検討する

というサイクルを何度か回して、然るべきところでモデルの利用へと向かうものである。ここでは、①において多変量正規分布を仮定したもとの、ある共分散選択モデルを設定する。続く②の母数の推定では最尤推定法を用いることにする。Dempster[12]が与えた最尤推定の基本定理は次のものである。

【定理 4】

共分散選択モデルでの最尤推定値は、次の 2 条件を満たすものとして一意に定まる。

- 1) モデルで $\rho^{ij}=0$ とした (i, j) については、 ρ^{ij} の推定値は 0 になる。
- 2) モデルで $\rho^{ij}=0$ としていない (i, j) については、 ρ_{ij} の推定値は標本相関係数 r_{ij} になる。

□

Wermuth and Scheidt[55]は、この 2 条件を満たす推定値を求めるアルゴリズムをプログラムリスト付きで示している。

次の③では、最尤推定の延長として尤度比検定統計量を算出する。帰無仮説にあるのが①で設定している共分散選択モデルで、対立仮説にあるのが Π に何等制約を置かない多変量正規分布モデルである。尤度比検定統計量の対数を 2 倍した量を逸脱度 (deviance) と呼ぶ。これを求めると、上の 1), 2) の条件を満たす相関行列の推定量を $\hat{\Pi}$ 、標本相関行列を R とすれば、両者の行列式を用いて

$$dev = n \log \frac{|\hat{\Pi}|}{|R|} \quad (2.12)$$

という量が導ける。 n は標本数である。逸脱度は近似的に χ^2 分布に従い、自由度は 0 とおいた偏相関係数の数、すなわち無向独立グラフで削除した辺の数である。逸脱度によって、共分散選択モデルのデータへの適合性を評価するとき、その論理は基本的に仮説検定である。真の相関構造があるモデルに完全に一致することは現実にはない。そのため、標本数が多くなるにつれて、いつかはモデルが棄却されるという宿命をもつ。

④のモデルの変更には話を進めよう。共分散選択モデルの変更は、無向独立グラフでの辺の削除・追加に対応する。よって、変更前と変更後のモデルの意味がグラフによって視覚的にとらえやすい。これこそが、共分散選択モデルを無向グラフ表現する最大の利点である。ところで、①でのモデル設定がなされれば、②と③のステップは完全に機械的に行われ、解析者の判断の入る余地はない。よって、④こそが解析者の腕の見せ所となる。④での主業務は利用価値のあるモデルを選択することである。宮川・芳賀[30]が開発した解析システムでは、フルモデル(Π に何等制約を置かない多変量正規分布モデル)を初期モデルとした上で、モデルの変更作業を解析者に任せ対話型システムを実現している。

【例 3】 共分散選択モデリングの実行例

$n=70$ の個体の各々について 4 つの量的変数を測定し、その結果を (2.13) に示す標本相関行列にまとめた。この数値例は日本品質管理学会テクノメトリックス研究会編[35]5.1.2 節からの引用である。相関行列は対称行列なので、左下のみを表示する。

$$\begin{array}{cccc} X_1 & 1.000 & & \\ X_2 & 0.720 & 1.000 & \\ X_3 & -0.389 & -0.257 & 1.000 \\ X_4 & 0.631 & 0.487 & -0.559 & 1.000 \end{array} \quad (2.13)$$

この標本相関行列から標本偏相関行列を求めると (2.14) を得る。なお、偏相関行列と相関

行列を区別するために、偏相関行列では対角要素を－とする。

$$\begin{array}{cccccc} X_1 & & - & & & \\ X_2 & & 0.61 & & - & \\ X_3 & & -0.09 & & 0.07 & & - \\ X_4 & & 0.37 & & 0.09 & & -0.44 & & - \end{array} \quad (2.14)$$

$r_{13 \cdot 24} = -0.09$, $r_{23 \cdot 14} = 0.07$, $r_{24 \cdot 13} = 0.09$ と 3 つの標本偏相関係数の値が無視できるほど小さいことが観察される。そこで、この 3 つの母偏相関係数を 0 と置いた共分散選択モデルを採用する。このモデルの逸脱度の値は $dev = 0.823$ となる。 χ^2 分布に近似するときの自由度は $df = 3$ である。これより p 値を求めると $p = 0.8439$ となる。この共分散選択モデルを無向グラフ表示したものが図 2 である。定理 3 に示したマルコフ性から、このモデルで成り立っている様々な条件付き独立性が即座に読みとれる。□

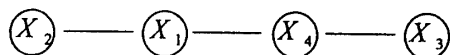


図 2 相関行列 (2.13) に当てはめた無向独立グラフ

2.4 グラフィカル対数線形モデル

分割表にまとめられる度数データに対する統計モデルは、サンプリング方式によって異なる。たとえば、最も簡単な 2 元分割表では

- ・総和のみが与えられる場合
- ・行和と列和のいずれかが与えられる場合
- ・行和と列和のいずれもが与えられる場合
- ・総和も与えられない場合

というサンプリングの仕方によって異なる。3 元以上になっても同様である。

3 元分割表で「総和のみが与えられる場合」の統計モデルを考える。項目を表す離散確率変数を X, Y, Z とし、各項目の水準数を I, J, K とする。個体がセル $x_i y_j z_k$ に属する確率を $p_{ijk} = P\{X = x_i, Y = y_j, Z = z_k\}$ と表記する ($i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K$)。また、それぞれの周辺確率を添字に + を用いた記号で表現する。 Z を与えたとき X と Y が条件付き独立とは、(2.2) 式の定義より、すべての i, j, k において

$$\frac{p_{ijk}}{p_{++k}} = \frac{p_{i+k}}{p_{++k}} \times \frac{p_{+jk}}{p_{++k}} \quad (2.15)$$

が成り立つことである。

セル確率 p_{ijk} の対数について

$$\log p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} \quad (2.16)$$

という線形加法構造を想定するモデルを対数線形モデルという。母数の無駄を省くために、右辺のすべての母数は、それがもつ任意の添字について和をとったときに 0 という制約（これを Birch[1] の制約という）を置く。

さらに、たとえば $(\alpha\beta)_{ij} = 0$ ならば $(\alpha\beta\gamma)_{ijk} = 0$ というように「低次の項が 0 ならば、その項の添字を含むすべての高次の項は 0」という制約を満たすモデルを階層的対数線形モデルという。

さて、(2.16) 式において、すべての i, j, k で

$$(\alpha\beta\gamma)_{ijk}=0 \quad \text{かつ} \quad (\alpha\beta)_{ij}=0$$

ならば、定理1の因数分解基準により、 $X \perp\!\!\!\perp Y|Z$ である。同様にして、4元以上の多元分割表に対する階層的对数線形モデルにおいても、ある項目対に対応する2因子交互作用項を0と置いたモデルでは、その項目対が残りの項目を与えたときに条件付き独立になる。

ところで、階層的对数線形モデルでは、母数の添字の包含関係から、極大項の集合を列挙できる。たとえば

$$\log p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \quad (2.17)$$

では、 $(\alpha\gamma)_{ik}$ と $(\beta\gamma)_{jk}$ が2つの極大項である。この極大項の集合を生成集合あるいは定義集合といい、対応する項目名を使って XZ/YZ と表記する。

階層的对数線形モデルを無向独立グラフに対応させよう。素直に考えると、2因子交互作用項が0である項目対を辺で結ばず、そうでない項目対を辺で結ぶ無向グラフを作成すればよいように思える。しかし、対数線形モデルでは、多変量正規分布の共分散選択モデルと異なり3次以上の高次の交互作用項があるため、上の操作ではモデルとグラフが1対1に対応しない。たとえば、項目数が3の完全グラフには、(2.16)式のフルモデルとともに、3因子交互作用項のみを0とした

$$\log p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \quad (2.18)$$

が対応してしまう。

そこで、グラフと1対1対応するように、階層的对数線形モデルの部分クラスを考える。Darroch et al.[9]は、与えられた無向グラフのクリーク集合が生成集合に一致する階層的对数線形モデルをグラフィカルモデルと呼んだ。3項目の完全グラフに対応するグラフィカルモデルは(2.16)式のフルモデルであって、(2.18)式のモデルではない。なお、ここでの「グラフィカルモデル」とは極めて狭義の意味であり、共分散選択モデルを含めてグラフィカルモデルと呼ぶのが一般的である。Whittaker[57]7章では、対数線形モデルでのグラフィカルモデルをグラフィカル対数線形モデルと呼んでおり、本論文もこの立場をとる。完全グラフ以外の無向グラフに対応するグラフィカル対数線形モデルは、必ず何らかの独立性・条件付き独立性を規定する。

ところで、グラフィカル対数線形モデルでない階層的对数線形モデルは利用価値が乏しいかといえ、決してそうではない。 X が $I=2$ の処理変数、 Y が $J=2$ の反応変数で、 Z が年齢層などの共変数のとき、(2.18)式のモデルは「各層でのオッズ比が層間で一様である」という意味のあるモデルになっている。グラフィカル対数線形モデルに限定したモデリングでは、このようなモデルが推測対象外になってしまうことに注意する。

階層的对数線形モデルが普及していく過程で重要な役割を果たしたのが、分解可能モデルあるいは乗法的モデルと呼ばれるモデルクラスの出現である。この概念は、ほぼ同時期に Goodman[19], Bishop[2], Haberman[2], Sundberg[51]らによって提起された。たとえば、4元分割表の場合に、生成集合が XYZ/YZW となるモデルでの同時確率関数 $p_{xyzw}(x, y, z, w)$ は、これから導かれる周辺確率関数によって

$$p_{xyzw}(x, y, z, w) = \frac{p_{xyz}(x, y, z)p_{yzw}(y, z, w)}{p_{yz}(y, z)} \quad (2.19)$$

という乗法的分解が可能である。この性質をもつモデルクラスを分解可能モデルあるいは乗法的モデルという。このモデルクラスを一般的に定義するには、ややわかりにくい表現が必要と

なる (例えば宮川[27]p.101 を参照)。これに対して Darroch et al.[9]は、グラフ表現による明解な判定基準を与えた。まず、分解可能モデルはグラフィカル対数線形モデルのサブクラスであること、さらに、あるグラフィカル対数線形モデルが分解可能モデルである必要十分条件は、無向グラフ表現において弦のない長さ 4 以上の閉路がないことである。分解可能モデルの意義は、3章に述べる有向独立グラフとの関係においても認識される。

【例 4】 図 3 に示す 2 つの無向独立グラフを考えよう。図 3 (a) は (2.19) 式のモデルに対応するものである。\$X-Y-W-Z-X\$ という長さ 4 の閉路において、\$Y-Z\$ が弦であるから、この閉路には弦がある。よって、このモデルは分解可能モデルである。一方、図 3 (b) では \$X-Y-W-Z-X\$ という長さ 4 の閉路に弦がないので、分解可能モデルでない。□

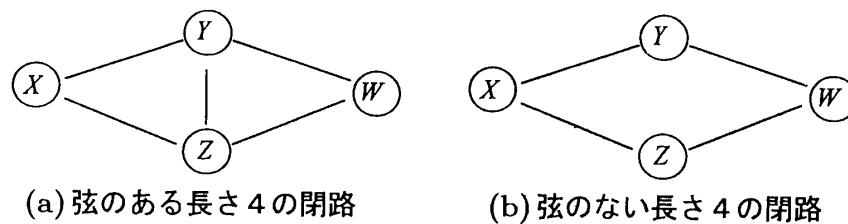


図 3 分解可能なモデルとそうでないモデル

モデルの統計的推測に話を進める。この節の冒頭に述べたように、分割表の統計モデルはサンプリング方式に依存する。しかし、推測対象の母数を各セルでの期待度数にすると、サンプリング方式によらない統一的な表現ができることが知られている (Bishop, Fienberg and Holland[3]3.2 節)。このとき、Birch[1]が与えた次の結果が本質的である。

【定理 5】

階層的対数線形モデルにおいて、生成集合の各要素に対応する周辺期待度数の最尤推定値はそこでの観測周辺度数に一致し、これらから任意のセルでの期待度数の最尤推定値は一意に定まる。□

3元分割表の場合で、この定理の意味を確認しよう。セル $x_i y_j z_k$ での観測度数を n_{ijk} とし、観測周辺度数についても添字に + をつけた記号を用いる。総和は n と記す。いま、(2.17) 式のモデルを想定すると $X \perp\!\!\!\perp Y | Z$ が成り立っており、(2.15) 式を変形すれば

$$np_{ijk} = \frac{np_{i+k} \times np_{+jk}}{np_{++k}} \quad (2.20)$$

を得る。左辺はセル $x_i y_j z_k$ の期待度数 μ_{ijk} である。右辺には生成集合 XZ/YZ の各要素での周辺期待度数とそれから導かれる z_k での周辺期待度数が登場している。よって、セル $x_i y_j z_k$ の期待度数の最尤推定値 m_{ijk} は、(2.20) 式右辺の周辺期待度数をそれぞれ観測周辺度数で推定したもののより

$$m_{ijk} = \frac{n_{i+k} \times n_{+jk}}{n_{++k}} \quad (2.21)$$

と求められる。これからわかるように、対数線形モデルが分解可能モデルであれば、同時確率関数は周辺確率関数によって表現されているので、セル期待度数の最尤推定値は観測周辺度数により閉じた形で求められる。

分解可能モデルはもともと分割表の統計モデルで生まれた概念であるが、上述のグラフ表現

によって多変量正規分布における共分散選択モデルにも適用されるものとなった。実際、共分散選択モデルにおいても、定理4を満たす最尤推定値は分解可能のときには明示表現が可能で、そうでないときには反復計算を要する (Speed and Kiiveri[49])。なお、分解可能でない階層的対数線形モデルでの母数推定のアルゴリズムとしてはIPF (Iterative Proportional Fitting) が知られている。IPFのアイデアはDeming and Stephan[11]に遡り、Haberman[20]にはプログラムリストが載っている。

ある階層的対数線形モデルを帰無仮説に、フルモデルを対立仮説にとったときの尤度比検定統計量から逸脱度が導かれる。3元の場合は

$$dev = 2 \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K n_{ijk} \log \frac{n_{ijk}}{m_{ijk}} \quad (2.22)$$

の形になる。 m_{ijk} は帰無仮説のもとでの期待度数の推定値である。4元以上になってもそのまま拡張される。逸脱度の自由度は、帰無仮説で0と置いた母数の実質的個数である。

【例5】グラフィカル対数線形モデリングの実行例

質問紙調査では各質問項目に対して「はい・いいえ」の質的反応を測定することが多い。いま、それぞれ2水準の4つの項目 X, Y, Z, W に対して、全部で $n=200$ 人の回答者が各項目に回答した結果を表1の4元分割表にまとめたとする (日本品質管理学会テクノメトリックス研究会編[35]8.2節より引用)。

表1 4元分割表の数値例

		z_1		z_2	
		w_1	w_2	w_1	w_2
x_1	y_1	36	8	32	5
	y_2	18	4	16	3
x_2	y_1	6	22	8	24
	y_2	4	6	4	4

一般に多元分割表は、項目数が増えると疎になる。そこで、項目間の関連状況を把握するには、いくつかの項目だけを取り出し他の項目について和をとった周辺分割表を作成するのが有効である。この作業を併合という。2つの項目を取り出せば2元分割表にまとめることができる。すべての項目対についてこれを行った結果をまとめたものをバート表あるいは多重クロス表という。表1に対するバート表が表2である。

表2 バート表の数値例 (表1に対応)

	x_1	x_2	y_1	y_2	z_1	z_2	w_1	w_2
x_1	122	0						
x_2	0	78						
y_1	81	60	141	0				
y_2	41	18	0	59				
z_1	66	38	72	32	104	0		
z_2	56	40	69	27	0	96		
w_1	102	22	82	42	64	60	124	0
w_2	20	56	59	17	40	36	0	76

バート表は相関行列と同様に対称行列なので、左下のみを表示する。対角に位置する項目自身の2元表は、各カテゴリーの度数を示すいわゆる単純集計である。異なる項目からなる 2×2 の分割表がいわゆるクロス集計に相当する。項目 X と W の間に強い関連があることが観察される。このことを意識して元の4元分割表を見直せば、当然ながらその傾向を確認できる。このようにバート表は2つの項目間の関連を把握するのに有効である。

ところが、ここで多元分割表固有の問題が発生する。多元分割表をむやみに併合して周辺分割表にまとめ、そこでの関連を結論にすると、とんでもない誤りを犯す恐れがある。これを指摘したのがシンプソンのパラドックスである(Simpson[48])。すなわち、多元分割表をいくつかの周辺分割表に併合する作業はとても大事なことであるけれども、併合してよい場合とない場合がある。それを元の多元分割表から識別しなければならない。そのような識別を行って、かつ、項目間の関連状況をコンパクトにまとめた数枚の周辺分割表を提示することが、多元分割表におけるグラフィカル対数線形モデリングの仕事である。

表1の4元分割表に対する対数線形モデルのフルモデルは

$$\begin{aligned} \log p_{ijkl} = & \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\delta)_{il} \\ & + (\beta\gamma)_{jk} + (\beta\delta)_{jl} + (\gamma\delta)_{kl} \\ & + (\alpha\beta\gamma)_{ijk} + (\alpha\beta\delta)_{ijl} + (\alpha\gamma\delta)_{ikl} + (\beta\gamma\delta)_{jkl} \\ & + (\alpha\beta\gamma\delta)_{ijkl} \end{aligned} \quad (2.23)$$

である。4つの項目では項目対に対応する2因子交互作用項が6つあるので、そのうちのどれが0と置けるかを検討していく。また、これにともなって高次の交互作用項も0と置く。考えられる6つのケースのそれぞれについて、逸脱度と p 値を計算し、それをまとめると表3を得る。なお、解析には宮川・芳賀[31]が開発したソフトを用いた。

表3 フルモデル(2.23)からのモデル変更の検討

	削除する交互作用項	自由度	逸脱度	p 値
①	$(\alpha\beta)_{ij}, (\alpha\beta\gamma)_{ijk}, (\alpha\beta\delta)_{ijl}, (\alpha\beta\gamma\delta)_{ijkl}$	4	2.69	0.610
②	$(\alpha\gamma)_{ik}, (\alpha\beta\gamma)_{ijk}, (\alpha\gamma\delta)_{ikl}, (\alpha\beta\gamma\delta)_{ijkl}$	4	1.28	0.865
③	$(\alpha\delta)_{il}, (\alpha\beta\delta)_{ijl}, (\alpha\gamma\delta)_{ikl}, (\alpha\beta\gamma\delta)_{ijkl}$	4	64.25	0.000
④	$(\beta\gamma)_{jk}, (\alpha\beta\gamma)_{ijk}, (\beta\gamma\delta)_{jkl}, (\alpha\beta\gamma\delta)_{ijkl}$	4	0.63	0.960
⑤	$(\beta\delta)_{jl}, (\alpha\beta\delta)_{ijl}, (\beta\gamma\delta)_{jkl}, (\alpha\beta\gamma\delta)_{ijkl}$	4	3.09	0.543
⑥	$(\gamma\delta)_{kl}, (\alpha\gamma\delta)_{ikl}, (\beta\gamma\delta)_{jkl}, (\alpha\beta\gamma\delta)_{ijkl}$	4	0.66	0.957

まず、 p 値の一番大きいケース④を採用する。ケース④で成立している条件付き独立性は $Y \perp\!\!\!\perp Z | (X, W)$ である。この要領で、次にどの2因子交互作用項を0と置けるかを検討していく。このような母数の減少法で次のグラフィカル対数線形モデル

$$\log p_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\delta)_{il} + (\beta\delta)_{jl} \quad (2.24)$$

に到達した。フルモデル(2.23)式と比べれば、1つの4因子交互作用項、4つの3因子交互作用項、4つの2因子交互作用が消えている。消えた母数の実質的個数が9であり、これが逸脱度の自由度となる。(2.24)式に対応する無向独立グラフは図4である。

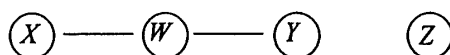


図4 表1の分割表に当てはめた無向独立グラフ

図4のモデルの主張点は、 $(X, Y, W) \perp\!\!\!\perp Z$ と $X \perp\!\!\!\perp Y | W$ である。前者より

$$p_{ijkl} = p_{ij+l} \times p_{++k+} \quad (2.25)$$

という関係が規定される。一方、後者より

$$p_{ij+l} = \frac{p_{i++l} \times p_{+j+l}}{p_{++++}} \quad (2.26)$$

が規定される。(2.26)式を(2.25)式の右辺に代入した形で、右辺の各周辺確率を観測周辺度数で推定すれば、結果として

$$m_{ijkl} = \frac{n_{i++l} \times n_{+j+l}}{n_{++++}} \times \frac{n_{++k+}}{n} \quad (2.27)$$

という式を得る。これに基づき、各セルでの期待度数を推定した結果が表4である。

表4 図4のモデルのもとでの期待度数

		z_1		z_2	
		w_1	w_2	w_2	w_2
x_1	y_1	35.07	8.07	32.38	7.45
	y_2	17.97	2.33	16.58	2.15
x_2	y_1	7.57	22.61	6.98	20.87
	y_2	3.87	6.51	3.58	6.01

各セルで、期待度数が観測度数に近い値になっていることが観察される。また、定理5より、この期待度数から作成した X と W の2元表、 Y と W の2元表、および Z の1元表はパート表でのそれに一致する。期待度数と観測度数の乖離具合を総合的かつ定量的に評価するために逸脱度を算出すると、その値は $dev=4.07$ となり、自由度は $df=9$ で p 値は $p=0.9069$ である。□

3. 有向独立グラフの統計的推測

3.1 有向独立グラフと条件付き独立性

この章では、有向グラフで表現されるグラフィカルモデルについて論じる。 p 個の確率変数の集合を $V=\{X_1, X_2, \dots, X_p\}$ とし、無向グラフと同様に V が頂点集合である。この直積 $V \times V$ の部分集合 E を有向の矢線の集合とする。 $(X_i, X_j) \in E$ のとき、 X_i から X_j へ矢線がある。初めに、一般的な有向グラフの用語を確認する。

- ① 親と子：頂点 X_i から X_j に矢線があるとき、頂点 X_i は X_j の親といい、 X_j は X_i の子という。 X_j の親の集合を $pa(X_j)$ と表記する。
- ② 道と有向道：頂点の系列 X_1, X_2, \dots, X_{n+1} は、すべての $i=1, 2, \dots, n$ で、 X_i から X_{i+1} に、あるいは X_{i+1} から X_i に矢線があるとき、長さ n の道という。特に、すべての $i=1, 2, \dots, n$ で、 X_i から X_{i+1} に矢線があるとき、長さ n の有向道という。
- ③ 非巡回的：長さ n の有向道 X_1, X_2, \dots, X_{n+1} で、 $X_1=X_{n+1}$ となるものを長さ n の巡回閉路という。巡回閉路がない有向グラフを非巡回的という。
- ④ 先祖と子孫：頂点 X_i から X_j に有向道があるとき、頂点 X_i は X_j の先祖といい、 X_j は X_i の子孫という。
- ⑤ 非子孫：すべての頂点から X_i と X_i の子孫を除いたものを X_i の非子孫といい、その集合を $nd(X_i)$ と表記する。

以上の準備をもとに、有向独立グラフの定義を与える。

【定義 2】 有向独立グラフ

非巡回的有向グラフ $G=(V, E)$ が $V=\{X_1, X_2, \dots, X_p\}$ の同時密度関数をグラフに従う逐次的因数分解されている形、すなわち

$$f_V(x_1, x_2, \dots, x_p) = \prod_{i=1}^p f_{i \cdot pa}(x_i | pa(x_i)) \quad (3.1)$$

の形に規定するとき、その有向グラフを有向独立グラフという。ここに、 $f_{i \cdot pa}(x_i | pa(x_i))$ は $pa(X_i)$ を与えたときの X_i の条件付き密度関数であり、 $pa(X_i)$ が空のときには X_i の周辺密度関数を意味する。□

グラフに従う逐次的因数分解に対して定理 1 の因数分解基準を適用すれば、次の性質を導ける。

【定理 6】 有向独立グラフでの局所的マルコフ性

有向独立グラフでは、任意の X_i に対して、その親を与えたときに、 X_i は親以外のすべての非子孫と条件付き独立である、すなわち

$$X_i \perp\!\!\!\perp nd(X_i) \setminus pa(X_i) \mid pa(X_i) \quad (3.2)$$

が成り立つ。親が存在しない場合は条件部が空となり独立性を意味する。□

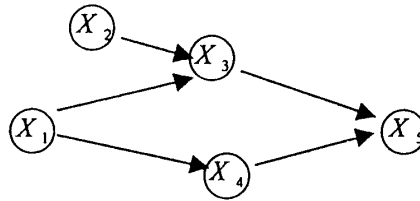


図 5 有向独立グラフの例

【例 6】 図 5 の有向独立グラフが規定する逐次的因数分解は

$$f_V(x_1, x_2, \dots, x_5) = f_1(x_1) f_2(x_2) f_{3 \cdot pa}(x_3 | x_1, x_2) f_{4 \cdot pa}(x_4 | x_1) f_{5 \cdot pa}(x_5 | x_3, x_4) \quad (3.3)$$

である。局所的マルコフ性が主張する独立性・条件付き独立性は

$$\begin{aligned} X_1 &\perp\!\!\!\perp X_2 \\ X_2 &\perp\!\!\!\perp (X_1, X_4) \\ X_3 &\perp\!\!\!\perp X_4 \mid (X_1, X_2) \\ X_4 &\perp\!\!\!\perp (X_2, X_3) \mid X_1 \\ X_5 &\perp\!\!\!\perp (X_1, X_2) \mid (X_3, X_4) \end{aligned} \quad (3.4)$$

である。□

有向独立グラフでの大域的マルコフ性の表現は 2 通りあり、いずれもやや準備がいる。初めに、Lauritzen et al.[26]による表現を述べる。

いま、頂点 X_k の 2 つの親 X_i, X_j において、 X_i と X_j の間でいずれの向きにも矢線がないとき、 X_k は $\{X_i, X_j\}$ に対する V 字合流点であるという。有向独立グラフの矢線をすべて無向の辺に変え、すべての V 字合流点に対して 2 つの親を辺で結ぶ。このようにして得られる無向グラフをモラルグラフという。図 5 の有向独立グラフに対するモラルグラフを図 6 に示す。

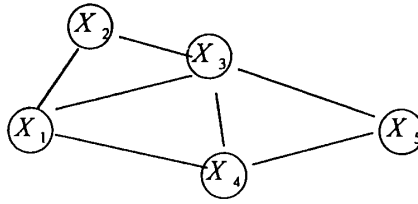


図6 モラルグラフの例 (図5に対応)

逐次的因数分解に定理1の因数分解基準を適用することで、モラルグラフで成立している独立性・条件付き独立性は、すべて元の有向独立グラフで成立していることがわかる。しかし、この逆は成り立たず、元の有向独立グラフで成立している独立性・条件付き独立性のすべてがモラルグラフに表現されているわけではない。そこで、次のような精緻化を行う。

頂点集合 V の部分集合 u において、任意の $X \in u$ の先祖がすべて u に含まれているとき、 u は先祖集合であるという。任意の部分集合 u に対して、 u を含む最小の先祖集合が存在し、これを u の最小先祖集合という。

【定理7】 有向独立グラフでの大域的マルコフ性 (1)

たがいに排反な頂点集合 a, b, s において、 $a \cup b \cup s$ の最小先祖集合が生成する部分有向グラフのモラルグラフにおいて、 s が a と b を分離していれば

$$a \perp\!\!\!\perp b | s$$

が成り立つ。 □

【例7】 図5の有向独立グラフにおいて、 $a = \{X_3\}$, $b = \{X_4\}$, $s = \{X_1\}$ とする。 $a \cup b \cup s$ の最小先祖集合は $\{X_1, X_2, X_3, X_4\}$ となる。この変数集合が生成する部分有向グラフのモラルグラフで、 s は a と b を分離しているので

$$X_3 \perp\!\!\!\perp X_4 | X_1$$

が成り立つ。 □

次に、Geiger et al.[12]による表現を示す。

【定義3】 有向分離

2つの頂点 X_i と X_j を結ぶすべての道のそれぞれについて、 $\{X_i, X_j\}$ と排反な頂点集合 s が次の条件のいずれかを満たすとき、 s は X_i と X_j を有向分離するという。

- 1) X_i と X_j を結ぶ道にV字合流点があるとき、その合流点とその子孫は s に含まれない。
- 2) X_i と X_j を結ぶ道にV字合流でない点があるとき、その点の中で s に含まれるものがある。

たがいに排反な頂点集合 a, b, s において、任意の $X_i \in a$ と任意の $X_j \in b$ を s が有向分離するとき、 s は a と b を有向分離するという。 □

【定理8】 有向独立グラフでの大域的マルコフ性 (2)

たがいに排反な頂点集合 a, b, s において、 s が a と b を有向分離していれば

$$a \perp\!\!\!\perp b | s$$

が成り立つ。 □

定理7と定理8のいずれにおいても、 s が空の場合には独立性を意味する。また、2つの定理の内容は同値であり、いずれも有向独立グラフで成立しているすべての独立性・条件付き独立

性を過不足なく表現している。ただし、これらの大域的マルコフ性は直接的な表現をしているだけであって、局所的マルコフ性も全体としてはこれと同値である。実際、例7で導いた $X_3 \perp\!\!\!\perp X_4 | X_1$ は、局所的マルコフ性が主張する $X_2 \perp\!\!\!\perp (X_1, X_4)$ と $X_3 \perp\!\!\!\perp X_4 | (X_1, X_2)$ に定理2と系1を適用することで導出される。

【例8】 図5の有向独立グラフにおいて、 $a = \{X_3\}$ 、 $b = \{X_4\}$ とする。 X_3 と X_4 を結ぶ道は、 $X_3 \leftarrow X_1 \rightarrow X_4$ と $X_3 \rightarrow X_5 \leftarrow X_4$ の2つである。後者において X_5 はV字合流点である。ここで $s = \{X_1\}$ とすれば、 s は前者の道に対して条件2)を満たし、後者の道に対して条件1)を満たす。よって、 s は a と b を有向分離しているのだ

$$X_3 \perp\!\!\!\perp X_4 | X_1$$

が成り立つ。 □

3.2 変数間の先行関係が既知の場合

いま、 p 個の変数があるとき、その間の先行関係が完全に与えられているとする。一般性を失うことなく、その順序を X_1, X_2, \dots, X_p とする。この先行関係を所与として、データから有向独立グラフを推測したい。有向独立グラフが規定する独立性・条件付き独立性を局所的マルコフ性で表せば

$$X_i \perp\!\!\!\perp nd(X_i) \setminus pa(X_i) | pa(X_i)$$

であるから、データからこれらが当てはまっているかどうかを統計的に判定すればよい。しかし「条件となる親が何か」は、有向独立グラフが完成されて初めて認識されるものだから、この基準はそのままは使えない。

幸いにして、次の手順を踏んでいけば、親を認識できなくてもデータから有向独立グラフを推測できる (Wermuth and Lauritzen[54])。 X_1, X_2, \dots, X_p という順序が与えられたとき、 $i < j$ なる X_i と X_j について

$$X_j \perp\!\!\!\perp X_i | \{X_1, X_2, \dots, X_j\} \setminus \{X_i, X_j\} \quad (3.5)$$

が成立しているかどうかを統計的に判定し、成立していないときに X_i から X_j への矢線を引き、成立していれば矢線を引かない。これをすべての $i < j$ で行う。すなわち

$$\begin{aligned} X_2 &\perp\!\!\!\perp X_1 \\ X_3 &\perp\!\!\!\perp X_1 | X_2, \quad X_3 \perp\!\!\!\perp X_2 | X_1 \\ X_4 &\perp\!\!\!\perp X_1 | (X_2, X_3), \quad X_4 \perp\!\!\!\perp X_2 | (X_1, X_3), \quad X_4 \perp\!\!\!\perp X_3 | (X_1, X_2) \\ &\vdots \end{aligned}$$

のそれぞれが成立しているかどうかを、逐次、判定していく。上の各行が部分変数集合に対する1回の無向グラフによるモデリングに対応する。このとき、たとえば3変数のモデリングで、 $X_1 \perp\!\!\!\perp X_2 | X_3$ という条件付き独立性は判定の対象外であることに注意する。仮にこの関係が認められても、それによって X_1 から X_2 への矢線を削除してはならない。この矢線の有無の判定は、2変数での $X_2 \perp\!\!\!\perp X_1$ に基づくものである。

【例9】 図5の有向独立グラフで表される統計モデルが真のとき、十分多量のデータのもとで第1種の誤りが起こらなければ、上の判定過程において

$$\begin{aligned}
& X_2 \perp\!\!\!\perp X_1 \\
& X_4 \perp\!\!\!\perp X_2 | (X_1, X_3), \quad X_4 \perp\!\!\!\perp X_3 | (X_1, X_2) \\
& X_5 \perp\!\!\!\perp X_1 | (X_2, X_3, X_4), \quad X_5 \perp\!\!\!\perp X_2 | (X_1, X_3, X_4)
\end{aligned} \tag{3.6}$$

が認識される。そして、これらから図5が作成できる。実際、定理2と系1を駆使することで、(3.6)式と(3.4)式とが同値であることを示せる。□

3.3 変数間の先行関係が部分的に既知の場合

いくつかの変数対には順序がついても、ある変数対には順序がつけられないという場面は多い。そのような場合は、有向の矢線と無向の辺を併用した連鎖グラフによる統計モデルでデータの関連構造をモデリングする。

いま、全部で7つの変数があるとする。それらを $V_1 = \{X_1, X_2, X_3\}$, $V_2 = \{X_4, X_5\}$, $V_3 = \{X_6, X_7\}$ の3つの部分集合に群分けしたとき、 V_1 , V_2 , V_3 の間には先行関係を想定できるが、各部分集合内では順序がつけられないとする。このとき、次の手順で連鎖グラフによるグラフィカルモデリングを行う。

手順1: V_1 の変数を対象にして、無向独立グラフを推測する。

手順2: V_2 と V_2 の変数を対象にして

2-1) $X_i \in V_1$, $X_j \in V_2$ なるすべての (X_i, X_j) に対して

$$X_j \perp\!\!\!\perp X_i | V_1 \cup V_2 \setminus \{X_i, X_j\}$$

が成立しているかどうかを判定し、成立していなければ X_i から X_j へ矢線を引き、成立していれば矢線を引かない。

2-2) $X_i, X_j \in V_2$ なるすべての (X_i, X_j) に対して

$$X_j \perp\!\!\!\perp X_i | V_1 \cup V_2 \setminus \{X_i, X_j\}$$

が成立しているかどうかを判定し、成立していなければ X_i と X_j を辺で結び、成立していれば辺で結ばない。

手順3: V_1 と V_2 および V_3 の変数を対象にして

3-1) $X_i \in V_1 \cup V_2$, $X_j \in V_3$ なるすべての (X_i, X_j) に対して

$$X_j \perp\!\!\!\perp X_i | V_1 \cup V_2 \cup V_3 \setminus \{X_i, X_j\}$$

が成立しているかどうかを判定し、成立していなければ X_i から X_j へ矢線を引き、成立していれば矢線を引かない。

3-2) $X_i, X_j \in V_3$ なるすべての (X_i, X_j) に対して

$$X_j \perp\!\!\!\perp X_i | V_1 \cup V_2 \cup V_3 \setminus \{X_i, X_j\}$$

が成立しているかどうかを判定し、成立していなければ X_i と X_j を辺で結び、成立していれば辺で結ばない。

このようにして推測されたグラフを連鎖独立グラフという。各手順のそれぞれは無向グラフによるモデリングに帰着する。そこでの要点は、手順2で V_1 内のモデリングを行わないこと、同様に手順3で V_1 と V_2 内のモデリングを行わないことである。4群以上になっても、手順2, 3と同様な処理を行い順次グラフを拡大していけばよい。連鎖独立グラフにおいても、有向独立グラフとほぼ同様なマルコフ性が成り立つ。これについては Lauritzen[25]3.2節, 宮川[27]3.3節を参照されたい。

3.4 変数間の先行関係が未知の場合

ここで扱う状況は、変数間の先行関係を考えることに意味があるが、それに関する知識が不足している場合である。言い換えれば、統計データが示す関連構造の背後に未知なる有向独立グラフが存在すると仮定できる場合である。

いま、解析対象にしている確率変数の集合 $V = \{X_1, X_2, \dots, X_p\}$ の同時密度関数がある有向独立グラフ G に従って逐次的に因数分解され、それから規定されるマルコフ性以外のいかなる独立性・条件付き独立性を付加的に有していないとき、その同時分布は G に対して忠実 (faithful) であるという (Spirtes, Glymour and Scheines[50]2.3節)。以下では、 G に従う同時分布は G に忠実であるとする。このとき、 V の変数を解析対象にして無向独立グラフを推測すると、十分多量なデータのもとで第1種の誤りが起こらなければ、それは未知なる有向独立グラフのモラルグラフになる。それは次の理由による。まず、モラルグラフでは、元の有向独立グラフで存在していた矢線はすべて辺として残っている。同時分布が G に忠実であるので、矢線で結ばれた変数間ではいかなる独立性・条件付き独立性も成立しない。次に、同時分布が G に忠実なとき、定理8の内容は必要十分条件になる。すなわち、たがいに排反な頂点集合 a, b, s において、 s が a と b を有向分離することは、 $a \perp\!\!\!\perp b | s$ が成り立つ必要十分条件になる。ところで、モラルグラフでは V 字合流点の親が辺で結ばれている。 V 字合流 $X_i \rightarrow X_k \leftarrow X_j$ の道には、 X_i と X_j の間に X_k しかないのので、定義3の条件より、 X_k を含む任意の頂点集合は X_i と X_j を有向分離しない。よって、 V に対する無向独立グラフにおいては V 字合流点の親の間に辺が存在する。一方で、既に3.1節に述べたように、逐次的因数分解に定理1の因数分解基準を適用することで、モラルグラフで成立している独立性・条件付き独立性は、すべて元の有向独立グラフで成立している。以上より、有向独立グラフに対応する無向独立グラフはそのモラルグラフになることがわかる。

ところで、4つの変数について無向独立グラフを推測した結果、図3(b)に示す弦のない長さ4の閉路が得られたとしよう。定義2に記したように、ここでは有向独立グラフを非巡回的有向グラフに限定している。図3(b)において、非巡回的であるという制約で矢印をつけると、必ず V 字合流点が発生する。そのような有向独立グラフのモラルグラフでは、 V 字合流点の親は辺で結ばれているはずである。よって、図3(b)がモラルグラフになるような有向独立グラフは存在しない。このように「分解可能でない無向グラフィカルモデルには、背後に有向独立グラフを想定できない」という点で、解釈の難しいモデルである (Cox and Wermuth[7])。

上述のようにモラルグラフは、有向独立グラフで成立している独立性・条件付き独立性をすべて表現しているわけではない。しかし、あらゆる部分変数集合に対して無向独立グラフを作成すれば、有向独立グラフで成立している独立性・条件付き独立性をすべて識別できる。多変量正規分布の場合、たとえば「 $X_4 \perp\!\!\!\perp X_3 | X_1$ 」という関係は、この3変数のみの相関行列から推測可能である。

このような部分変数集合の総数は、全変数の数を p としたとき

$${}_p C_2 + {}_p C_3 + \dots + {}_p C_p = 2^p - p - 1 \quad (3.7)$$

と膨大になるが、理屈としてこの作業が可能である。この膨大な数の無向グラフィカルモデリングの結果を次のような無向グラフにまとめる。

初めに全変数の完全グラフを用意する。そして、ある頂点对 (X_i, X_j) において、 (X_i, X_j) 以外の何らかの頂点集合 (空の場合も含む) を与えたときに条件付き独立性が成り立てば、 X_i と X_j の線を削除する。これをすべての頂点对について行って得られる無向グラフは、有向独立グラフのすべての矢線を辺に置き換えたものになる (Spirtes, Glymour and Scheines[50]5.4

節)。この無向グラフを SGS 無向グラフと呼ぶ。ただし、これが条件付き独立性を意味する無向独立グラフでないことに注意する。図 5 の有向独立グラフに対する SGS 無向グラフは図 7 になる。

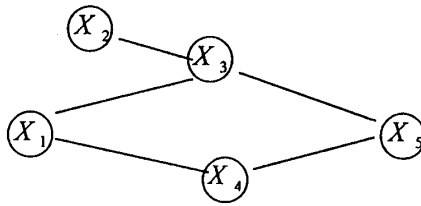


図 7 SGS 無向グラフの例 (図 5 に対応)

次の作業は、この SGS 無向グラフの辺に可能な限り矢印をつけることである。データのもつ関連情報のみから矢印をつけられるのであろうか。これを可能にするのが V 字合流の認識である。図 5 の有向独立グラフでは、 $X_2 \rightarrow X_3 \leftarrow X_1$ と $X_3 \rightarrow X_5 \leftarrow X_4$ という 2 つの V 字合流がある。上述のように、V 字合流点に任意の頂点を加えた頂点集合は 2 つの親を有向分離しない。すなわち、V 字合流点に任意の頂点を加えた頂点集合を与えたときに 2 つの親は条件付き独立にならない。これが V 字合流の著しい特徴となっている。

これに対して、図 5 での $X_1 \rightarrow X_3 \rightarrow X_5$ という部分に注目しよう。このとき、「 $X_1 \perp\!\!\!\perp X_5 | X_3$ 」は成り立たないが、「 $X_1 \perp\!\!\!\perp X_5 | (X_3, X_4)$ 」は成り立つ。すなわち、SGS 無向グラフで「 $X_i - X_j - X_k$ 」という X_i と X_k の間に辺のない道があるとき、この道が有向独立グラフで V 字合流でなければ、必ず、 X_j を含むある頂点集合 s において「 $X_i \perp\!\!\!\perp X_k | s$ 」が成り立つ。逆に、この道が V 字合流のときには、 X_j を含むいかなる頂点集合 s においても「 $X_i \perp\!\!\!\perp X_k | s$ 」は成り立たない。

以上の性質を利用して V 字合流を認識できる。それにはまず、SGS 無向グラフからすべての「 $X_i - X_j - X_k$ 」を列挙する。図 7 では

$$\begin{array}{ccc} X_2 - X_3 - X_1 & X_3 - X_1 - X_4 & X_1 - X_4 - X_5 \\ X_1 - X_3 - X_5 & X_2 - X_3 - X_5 & X_3 - X_5 - X_4 \end{array}$$

の 6 つがある。ここで、たとえば「 $X_3 - X_1 - X_4$ 」においては「 $X_3 \perp\!\!\!\perp X_4 | X_1$ 」が成り立つので、この道は V 字合流でないと判定できる。一方、たとえば「 $X_2 - X_3 - X_1$ 」では「 $X_2 \perp\!\!\!\perp X_1 | X_3$ 」「 $X_2 \perp\!\!\!\perp X_1 | (X_3, X_4)$ 」「 $X_2 \perp\!\!\!\perp X_1 | (X_3, X_5)$ 」「 $X_2 \perp\!\!\!\perp X_1 | (X_3, X_4, X_5)$ 」のいずれも成り立たないことより、この道が V 字合流だと判定できる。このように認識した V 字合流に矢線を入れ、それ以外は辺のままにしたグラフを SGS 連鎖グラフという。図 5 に対応する SGS 連鎖グラフは図 8 である。 X_1 と X_4 を結ぶ辺には矢印がついていないことに注意する。

この節では、変数間の関連情報のみから、すなわち、量的変数の場合は相関行列、質的変数の場合には多元分割表から、有向独立グラフを推測する方法について論じた。この方法の拠り所は V 字合流にあるので、背後にある有向独立グラフに V 字合流がなければ、全くお手上げで

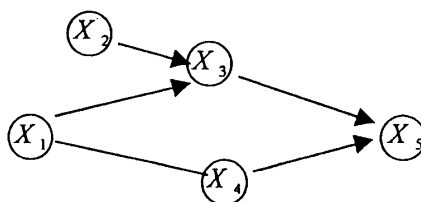


図 8 SGS 連鎖グラフの例 (図 5 に対応)

ある。変数の関連情報のみからは区別できない有向独立グラフは観察的に同値であると呼ばれる。それは

- 1) 同一の SGS 無向グラフをもつ
 - 2) V 字合流をなす道「 $X_i - X_j - X_k$ 」の集合が等しい
- という条件で記述される (Pearl[38])。

観察的に同値な有向独立グラフから1つを選択するには、統計データのもつ関連情報に加えて変数間の先行関係に関する情報が必要となる。それを可能にする強力な方法が実験研究である。図8では、 X_1 の値を人為的に変更したときに X_4 の分布が変われば、 X_1 から X_4 への矢印をつけることができる。

4. 因果ダイアグラムとその利用法

4.1 統計的因果関係と因果ダイアグラム

統計的因果推論の主要な目的として次の2つを挙げることができる。

- 1) 統計データにおいて実現している関連を統計的因果モデルで説明する。
 - 2) ある処理変量に人為的介入を施したときの反応変量の分布を推測する。
- 1) は科学的目的、2) は技術的目的といえる。1) のための方法論として、パス解析に始まる構造方程式モデルがある。パス係数による相関の分解は、まさしく観測された相関関係を統計的因果モデルで説明する強力な手段である。一方、2) のための方法論として、潜在反応モデルや観察研究における様々な個体分類・マッチング手法 (Rubin[46], [47], Rosenbaum and Rubin [43], [44], Rosenbaum[41]) などを挙げることができる。また、技術的目的という立場からは、操作可能性による次のような変量の分類は重要である。

統計的分析の対象は個体のある集団である。個体の測定によって得られる量において、その値がなぜ個体間で変動するかを知りたいという意味で科学的興味のあるものを反応という。個体に対して人為によって与える条件を処理という。処理を施す以前に個体が有している属性を共変量あるいは背景因子、付随変数などという。Cox[6]は、個体の固有属性に加えて、外的要因も共変量に含める立場をとっている。工業における工程解析を研究対象にするとき、温度や湿度といった環境要因を共変量として扱うことは有用である (黒木・宮川[23])。処理から影響を受け、反応に影響を与える可能性がある変量を中間特性という。

ところで「因果関係とは何か」については、古くから様々な哲学的論争がなされている (たとえば Holland[22]を参照)。Cox[6]は、統計的因果関係にも

- a) 2つの変量に対して、それらよりも先行する別な変量によって説明されない2変量間の統計的関連を統計的因果関係とする立場
- b) 統制された無作為化実験における複数の処理間での反応分布の違いを統計的因果関係とする立場

という2つの見地があるとしている。大雑把に言えば、a) は上の1) に対応し、b) は2) に対応する。また、上に述べた操作可能性からの変量の分類と対応させれば、b) では、基本的に処理から反応への因果関係のみを推測対象とする。一方、a) ではそれに加えて、共変量から中間特性や反応への因果関係、さらには共変量から処理への人為による因果関係をも解析対象とする。

一般に、因果分析、因果推論と呼ばれるものの目的には

- ・観察された結果に対する原因の究明
- ・観察された因果関係に対する因果メカニズムの解明
- ・観察された因果関係における因果的効果の定量的評価

があるとされている (Holland[22])。これらは相反するものでなく、たがいに補完しあうものであろうが、統計的因果推論においては主なる目的がどれであるかを明確にしておくことが重要である。たとえば、一般に中間特性というのは因果メカニズムの解明に役立つものだが、処理から反応への因果的効果の定量的評価には直接は役立たない。

さて、非巡回的有向グラフの矢線に因果的解釈が可能なグラフを因果ダイアグラムという。そのときの因果関係は広義の概念で a) に近い。一方で、その因果ダイアグラムをもとに技術的に利用可能な因果的効果の評価しようとするとき、対象となるのは b) の処理から反応への因果関係に限定するのが自然である。このあたりの整理が構造方程式モデルと潜在反応モデルを結びつける鍵となる。

Pearl[37][38]では、因果ダイアグラムを次のように定義している。

【定義4】 因果ダイアグラム

非巡回的有向グラフ $G=(V, E)$ と、その頂点に対応する確率変数の集合 $V=\{X_1, X_2, \dots, X_p\}$ が与えられている。グラフ G が変数間の因果的関係を

$$X_i = g_i(pa(X_i), \varepsilon_i) \quad i=1, 2, \dots, p \quad (4.1)$$

なる形に規定し、各変数がこの因果的関係に従って生成されるとき、 G を因果ダイアグラムという。ここに誤差変数 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ はたがいに独立とする。□

(4.1) 式は記述的関連モデルではなく、“data-generating process”を表す統計的因果モデル、すなわち構造方程式モデルである。

確率変数間の因果的関係が (4.1) 式によって規定されたとき、その同時分布において G に従う逐次的因数分解、すなわち (3.1) 式に与えた

$$f_V(x_1, x_2, \dots, x_p) = \prod_{i=1}^p f_{i \cdot pa}(x_i | pa(x_i))$$

が成立する (Pearl[37])。つまり、因果ダイアグラムは、それを形式的に有向独立グラフと見たときに規定されるすべての独立性・条件付き独立を与えることになる。

4.2 介入効果とその識別可能性

Pearl[37]は、因果ダイアグラムが与えられたとき、ある変量に対して、それへの有向道をもつ変量には操作をせず、その変量自身に対する外的操作によってある値に固定する行為を介入と定義した。その上で、変量 X に介入したことによる別な変量 Y への因果的効果を介入効果と呼び、その数学的定義を次のように与えた。このとき、 X への有向道をもつ変量のいくつかを観測したとしても、その値には依存せずに X の値を x に固定していること (これを無条件介入と呼ぶ) に注意する。

【定義5】 介入効果

頂点集合を $V=\{X, Y, Z_1, Z_2, \dots, Z_p\}$ とする因果ダイアグラム G において

$$f(y | set(X=x)) = \int \dots \int \frac{f_V(x, y, z_1, z_2, \dots, z_p)}{f_{X \cdot pa}(x | pa(x))} dz_1 \dots dz_p \quad (4.2)$$

を X の Y への介入効果という。 $set(X=x)$ は介入によって X の値を x にしたことを意味する。□

この $set(X=x)$ という記号は確率論にはない。 $f(y | set(X=x))$ をグラフで解釈すれば、 G より X へ向かう矢線をすべて除いた因果ダイアグラムで、 X を x に固定したときの Y の分布といえる。一般に、介入効果 $f(y | set(X=x))$ は通常条件付き密度関数 $f_{Y \cdot X}(y|x)$ とは異なる。

【例10】 簡単な場合として、図9を考えよう。図9(a)では

$$f_V(x, y, z) = f_X(x) f_{Z \cdot X}(z|x) f_{Y \cdot X}(y|z) \quad (4.3)$$

と G に従って逐次的因数分解されているので、(4.2) 式に代入すれば

$$f(y|\text{set}(X=x)) = \int f_{Z \cdot X}(z|x) f_{Y \cdot Z}(y|z) dz$$

となる。ここで $Y \perp\!\!\!\perp X|Z$ が成り立つので、 $f_{Y \cdot Z}(y|z) = f_{Y \cdot XZ}(y|x, z)$ である。また、 $f_{Z \cdot X}(z|x) = f_{XZ}(x, z)/f_X(x)$ であるから

$$f(y|\text{set}(X=x)) = \int \frac{f_{XYZ}(x, y, z)}{f_X(x)} dz = f_{Y \cdot X}(y|x) \quad (4.4)$$

となり、この場合の介入効果は条件付き密度関数に一致する。

また、図 9 (b) では

$$f_V(x, y, z) = f_Z(z) f_{X \cdot Z}(z|x) f_{Y \cdot Z}(y|z) \quad (4.5)$$

と逐次的因数分解されているので、(4.2) 式に代入すれば

$$f(y|\text{set}(X=x)) = \int f_Z(z) f_{Y \cdot Z}(y|z) dz = f_Y(y)$$

となり、この場合の介入効果は Y の周辺密度関数になる。

一方、図 9 (c) では

$$f_V(x, y, z) = f_Z(z) f_{X \cdot Z}(z|x) f_{Y \cdot XZ}(y|x, z) \quad (4.7)$$

と逐次的に因数分解されているので、(4.2) 式に当てはめて

$$f(y|\text{set}(X=x)) = \int f_Z(z) f_{Y \cdot XZ}(y|x, z) dz \quad (4.8)$$

となる。これらの介入効果が我々の常識に十分合うという点で、定義 5 に意義がある。□

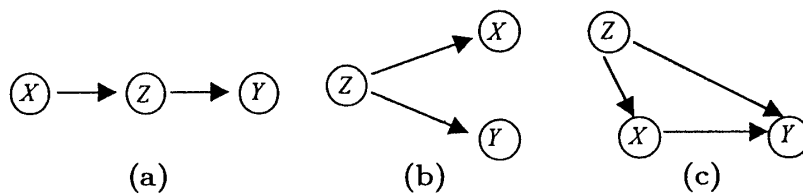


図 9 簡単な因果ダイアグラムの例

例 10 で行ったように、(4.2) 式の定義をもとに、因果ダイアグラムで成立している独立性・条件付き独立性による式変形を行うことで、介入効果について次の推測ルールが導かれる (Pearl[37]).

【定理 9】 推測ルール

- a) 因果ダイアグラム G より X へ向かう矢線をすべて除いたグラフにおいて、 X が Y と Z を有向分離するならば

$$f(y|\text{set}(X=x), z) = f(y|\text{set}(X=x)) \quad (4.9)$$

である。

- b) 因果ダイアグラム G より X から出る矢線をすべて除いたグラフにおいて、 Z が X と Y を

有向分離するならば

$$f(y|set(X=x), z) = f_{Y \cdot XZ}(y|x, z) \quad (4.10)$$

である。特別な場合として Z が空のときには

$$f(y|set(X=x)) = f_{Y \cdot X}(y|x) \quad (4.11)$$

である。

- c) 因果ダイアグラム G より X へ向かう矢線をすべて除いたグラフにおいて、 Z が X と Y を有向分離するならば

$$f(y|set(X=x), z) = f_{Y \cdot Z}(y|z) \quad (4.12)$$

である。特別な場合として Z が空のときには

$$f(y|set(X=x)) = f_Y(y) \quad (4.13)$$

である。 □

さて、図 9 (c) のケースに見たように、一般には $f(y|set(X=x))$ の表現には、 X と Y 以外の確率変数が登場する。いま、 $V = \{X, Y, Z_1, Z_2, \dots, Z_p\}$ の部分集合 $u = \{X, Y\} \cup s$ において、 u の同時密度関数によって $f(y|set(X=x))$ が一意に表現されるとき、 X の Y への介入効果は u において識別可能であるという。Pearl[37] は、識別可能となる十分条件としてバックドア基準を与えた。

【定義 6】 バックドア基準

因果ダイアグラム G において X から Y へ有向道があるとする。このとき、次の 2 つの条件を満たす頂点集合 s は (X, Y) についてバックドア基準を満たすという。

a) X から s の任意の要素に有向道がない。

b) G より X から出る矢線をすべて除いたグラフで、 s が X と Y を有向分離する。 □

バックドア基準を満たす頂点集合は必ず存在する。 X の親 $pa(X)$ や非子孫 $nd(X)$ はバックドア基準を満たす自明な集合である。また、ある頂点集合 s がバックドア基準を満たすとき、 s に X の子孫でない任意の頂点を加えた集合もまたバックドア基準を満たす。

【定理 10】 バックドア基準による識別可能性

因果ダイアグラム G において、頂点集合 s が (X, Y) についてバックドア基準を満たすならば、 X の Y への介入効果は $u = \{X, Y\} \cup s$ において識別可能であり、 $s = \{Z_1, \dots, Z_r\}$ と記したときに

$$f(y|set(X=x)) = \int \dots \int f_Z(z_1, \dots, z_r) f_{Y \cdot XZ}(y|x, z_1, \dots, z_r) dz_1 \dots dz_r \quad (4.14)$$

で与えられる。 □

【例 11】 図 5 の有向独立グラフが因果ダイアグラムとみなせるとしよう、 $X = X_4$ 、 $Y = X_5$ とすれば、 $\{X_1\}$ 、 $\{X_3\}$ はいずれも (X_4, X_5) についてバックドア基準を満たす。一方、 $\{X_2\}$ は (X_4, X_5) についてバックドア基準を満たさない。 □

さて、バックドア基準の効用は、回帰分析が使われる場面でより顕著である。このことを説明するために、(4.1) 式の特別な場合である次の線形構造方程式モデルを考えよう。

$$X_i = \sum_{X_j \in pa(X_i)} \alpha_{ij} X_j + \varepsilon_i \quad i = 1, 2, \dots, p \quad (4.15)$$

ここに、 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ はたがいに独立とする。 α_{ij} は X_j から X_i への矢線に対応する因果的効果を

表す母数でパス係数という。また、 X_1, X_2, \dots, X_p はすべて平均 0, 分散 1 に標準化されているとする。

【例 12】 図 5 の有向独立グラフが因果ダイアグラムとみなせるとき、これが規定する線形構造方程式モデルは

$$\begin{aligned} X_1 &= \varepsilon_1 \\ X_2 &= \varepsilon_2 \\ X_3 &= \alpha_{31}X_1 + \alpha_{32}X_2 + \varepsilon_3 \\ X_4 &= \alpha_{41}X_1 + \varepsilon_4 \\ X_5 &= \alpha_{53}X_3 + \alpha_{54}X_4 + \varepsilon_5 \end{aligned} \quad (4.16)$$

である。 □

因果ダイアグラムが変量間の定性的因果関係を正しく記述しているとき、統計データからパス係数の値を推定することができる。そのための方法が回帰分析である。回帰モデルは、 $V = \{X_1, X_2, \dots, X_p\}$ の同時分布から導かれる任意の条件付き分布について考えることができる。すなわち、回帰モデルは因果関係とは無関係な統計的関連モデルである。

このとき、線形構造方程式モデルでのパス係数 α_{ij} と、ある回帰モデルでの偏回帰係数 $\beta_{ij \cdot s}$ の関係が重要になる。ここに $\beta_{ij \cdot s}$ は、目的変数に X_i をとり、説明変数に X_j と $k \in s$ なる X_k をとった回帰モデルでの X_j の偏回帰係数である。すると、次の系が成り立つ (Pearl[38])。

【系 3】 パス係数と偏回帰係数との関係

因果ダイアグラム G において X_j から X_i へ矢線があるとする。 G よりこの矢線を除いたグラフで頂点集合 s が (X_j, X_i) を有向分離するならば、目的変数に X_i をとり、説明変数に X_j と $k \in s$ なる X_k をとった回帰モデルでの X_j の偏回帰係数 $\beta_{ij \cdot s}$ は、パス係数 α_{ij} に等しい。 □

さらに、系 3 の内容は次のように拡張される。因果ダイアグラムにおいて、頂点 X_j から X_i へ直接の矢線を含めて 1 つ以上の有向道があるとき、パス解析では、各有向道に沿ったパス係数の積をすべての有向道について足し合わせた値を X_j から X_i への総合効果と呼んでいる (Bollen[4]p.36)。このとき、次の系が成り立つ (Pearl[38])。

【系 4】 総合効果と偏回帰係数との関係

因果ダイアグラム G において、 X_j から X_i へ 1 つ以上の有向道があるとする。このグラフで頂点集合 s が (X_j, X_i) についてバックドア基準を満たすならば、目的変数に X_i をとり、説明変数に X_j と $k \in s$ なる X_k をとった回帰モデルでの X_j の偏回帰係数 $\beta_{ij \cdot s}$ は、 X_j から X_i への総合効果に等しい。 □

【例 13】 図 10 に示す因果ダイアグラム G を考えよう。これが規定する線形構造方程式の 1 つである

$$X_6 = \alpha_{62}X_2 + \alpha_{63}X_3 + \alpha_{64}X_4 + \alpha_{65}X_5 + \varepsilon_6 \quad (4.17)$$

に現れるパス係数 α_{63} を推定したいとする。 G より X_3 から X_6 への矢線を除いたグラフで、頂点集合 $\{X_2, X_5\}$ は (X_3, X_6) を有向分離するので、系 3 より

$$X_6 = \alpha_{62 \cdot 35}x_2 + \beta_{63 \cdot 25}x_3 + \beta_{65 \cdot 23}x_5 + \varepsilon_{6 \cdot 235} \quad (4.18)$$

という回帰モデルでの x_3 の偏回帰係数 $\beta_{63 \cdot 25}$ は、パス係数 α_{63} に等しい。

また、 X_3 から X_6 への総合効果は $\alpha_{63} + \alpha_{53}\alpha_{65}$ であり、これを推定したいとする。このとき、頂点集合 $\{X_2\}$ は (X_3, X_6) についてバックドア基準を満たすので、系 4 より

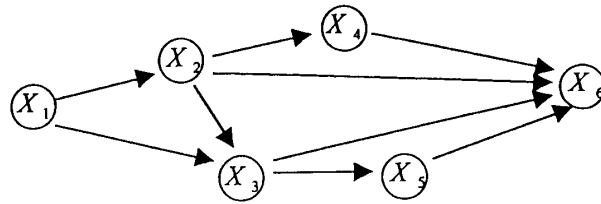


図10 因果ダイアグラムの例(2)

$$X_6 = \beta_{62 \cdot 3} x_2 + \beta_{63 \cdot 2} x_3 + \varepsilon_{6 \cdot 23} \quad (4.19)$$

という回帰モデルでの x_3 の偏回帰係数 $\beta_{63 \cdot 2}$ は、総合効果 $\alpha_{63} + \alpha_{53} \alpha_{65}$ に等しい。□

このように、バックドア基準は、回帰分析の目的を与えられた定性的因果仮説のもとでの因果的効果の推定においたとき、予測誤差による変数選択基準とは全く異なる変数選択基準を与える。

4.3 交絡因子の同定と併合可能性

医学・薬学で、処理と反応および共変量から分割表が構成されるとき、主たる関心事は処理から反応への統計的因果関係である。2.4節に述べたシンプソン・パラドックスの指摘により、共変量による層別が分割表解析の定石になっている。

Simpson[48]が与えた数値例を表5に示す。ただし、議論の都合上、 X と Y の位置を転置している。

表5 シンプソン・パラドックスの例 (Simpson[48])

	z_1		z_2		Zについて併合	
	y_1	y_2	y_1	y_2	y_1	y_2
x_1	4	3	2	2	6	6
x_2	8	5	12	15	20	20

すなわち、Simpson[48]が与えた数値例とは、 Z で層別したときの X と Y のオッズ比が層間で等しく1以下であるときに、 Z で併合したときの X と Y のオッズ比が1になる現象である。これを2.4節で用いた記号で表せば

$$\frac{n_{111} n_{221}}{n_{121} n_{211}} = \frac{n_{112} n_{222}}{n_{122} n_{212}} < 1$$

であるにもかかわらず

$$\frac{n_{11+} n_{22+}}{n_{12+} n_{21+}} = 1$$

となる現象である。これに先立ち、Yule[61]は、逆に

$$\frac{n_{111} n_{221}}{n_{121} n_{211}} = \frac{n_{112} n_{222}}{n_{122} n_{212}} = 1$$

であるにもかかわらず

$$\frac{n_{11+}n_{22+}}{n_{12+}n_{21+}} < 1$$

となる数値例を与えている。現在では、ユール・シンプソンのパラドックスといえ

$$\frac{n_{111}n_{221}}{n_{121}n_{211}} = \frac{n_{112}n_{222}}{n_{122}n_{212}} \leq 1 \quad (\text{あるいは} \geq 1) \quad (4.20)$$

であるにもかかわらず

$$\frac{n_{11+}n_{22+}}{n_{12+}n_{21+}} \geq 1 \quad (\text{あるいは} \leq 1) \quad (4.21)$$

となる現象（ただし，(4.20) 式と (4.21) 式の不等号の少なくとも 1 つは厳密に成り立つ）を指す (Mittal[32])。

また，Simpson[48]は，併合前の共通のオッズ比と併合後のオッズ比が等しくなる十分条件が

$$Y \perp\!\!\!\perp Z|X \quad \text{あるいは} \quad X \perp\!\!\!\perp Z|Y \quad (4.22)$$

であることを指摘した。Bishop, Fienberg and Holland[3]p.39 では，このオッズ比における併合可能性に対する必要十分条件が (4.22) 式であると記述されている。この記述が誤りであることが Whittemore[57]によって指摘される。(4.22) 式は， Z の水準数 K が $K=2$ のときには必要十分条件であるが， $K \geq 3$ では十分条件であって必要条件ではない。ただし，Whittemore[57]が与えた厳密な必要十分条件は，実務において使いやすいものでは必ずしもない。

さらに，Simpson[48]は，併合前の共通のオッズ比と併合後のオッズ比が等しくないときに，どちらのオッズ比に意味があるかについても議論している。表 5 に与えた数値例において， X ， Y ， Z の内容として 2 つの場合を与えている。1 つ目の場合は，トランプのカードの分類であり

X : 柄 (x_1 : 絵札, x_2 : 数札)

Y : 色 (y_1 : 赤, y_2 : 黒)

Z : 外観 (z_1 : 汚れている, z_2 : きれいな)

という項目である。乳児が遊んだトランプカード 52 枚をこれらの項目で分類したというものである。この場合は，項目 Z は結果系の反応であり，これによる層別は適当でない。よって， z_1 と z_2 のそれぞれでの X と Y のオッズ比には意味が乏しく， Z で併合した X と Y でのオッズ比に意味がある。

2 つ目の場合は，典型的な医学データの場合で

X : 処理 (x_1 : 無, x_2 : 有)

Y : 生死 (y_1 : 生, y_2 : 死)

Z : 性別 (z_1 : 男, z_2 : 女)

である。この場合は， Z はまさしく処理が施される以前に個体が有している属性，すなわち共変量であり，男女のいずれにおいても処理と生死に関連があることを示した各層でのオッズ比に意味がある。

この 2 つの場合で，併合前のオッズ比と併合後のオッズ比のどちらに意味があるかを判断する決め手は，項目間の因果関係において他にない。より一般に，分割表による処理と反応間の要因解析において，層別すべき共変量は「交絡因子」と呼ばれる。このとき，たびたび注意されるように，交絡因子とは，個体に処理が割りつけられる前に定まるもので，処理に影響を受けるものであってはならない(柳川[60]p.59)。補助変数のある実験データの解析法である共分

散分析でも、補助変数について同様な注意がなされている(田口[53]p. 807)。このような変量間の因果関係の記述には、因果ダイアグラムが最も簡潔で明瞭である。

Simpson[48]が与えた2つの場合について因果ダイアグラムを作成すれば、図11に示すものになると考えられる。図11(a)での「柄」と「色」から「外観」への矢線は、乳児が「赤札を黒札よりも好み、絵札を数札よりも好んだ」ことによる因果的効果を表す。図11(b)での「性別」から「生死」への矢線は「処理の有無のいずれにおいても、男性の生存割合が女性よりも大きい」という因果的効果を表し、「性別」から「処理」への矢線は「男女によって処理の割合が違う」という人為あるいはサンプリングによる影響を表す。このとき「処理を施すことで個体の性が変わる」という因果関係は考えられない。「処理」から「生死」への矢線は「男女のいずれにおいても、処理有での生存割合が処理無よりも大きい」ことを意味し、処理変量から反応変量への因果的効果という意味で、介入が可能な効果である。

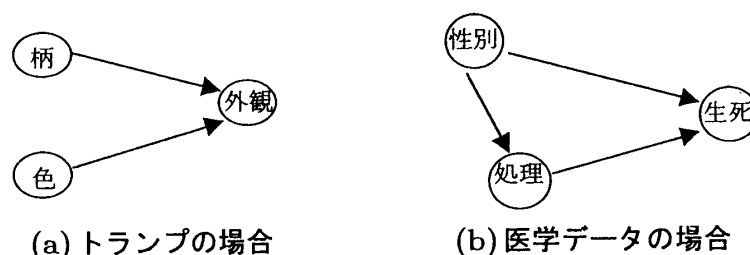


図11 Simpson (1951) の例に対する因果ダイアグラム

このように、変量間の因果関係を因果ダイアグラムで表現すれば、交絡因子の同定は容易である。すなわち、因果ダイアグラムより処理 X から出る矢線を除いたグラフにおいて、処理 X と反応 Y のいずれに対しても先祖となる変量が交絡因子である。逆に、因果ダイアグラムにおいて、処理 X の子孫で、反応 Y の先祖となる中間特性は交絡因子でなく、これによる層別は誤りである。

ところで、オッズ比における併合可能性を論じるとき、観測度数において、併合前の各層でのオッズ比および併合後でのオッズ比が厳密に等しいことは稀であり、実質的な対象は想定したモデルでの期待度数に対するオッズ比である。実際、オッズ比の併合可能性の十分条件である(4.22)式も、モデルがもつ性質と考えるのが現実的である。(4.22)式に対応する無向独立グラフは、図12の(a)と(b)である。一方、 X が処理、 Y が反応で、 Z が X に先行する共変量の場合を考えると、これらの無向独立グラフを与える有向グラフは図12(c)しかない。つまり、この場合の併合可能性の実質的な十分条件は、 $Y \perp\!\!\!\perp Z | X$ である。実際、図12(c)において、 Z から X を介さない Y への有向道がないので、 Z は交絡因子でない。

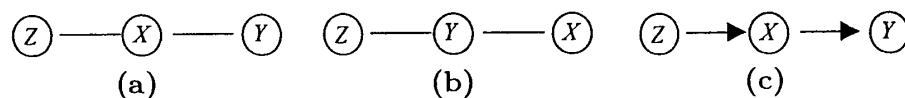


図12 併合可能となる統計モデル

4元以上の分割表においても、オッズ比の併合可能性の問題は重要である。いま、性別で層別した処理と反応の2元表でオッズ比がいずれも有意に1以下(処理有の生存割合が処理無よりも有意に大きい)であったとしよう。しかし、2つのオッズ比の一様性の検定(たとえばBleslow and Day[5]の検定)を行ったところ、一様性は棄却されたとする。その場合にも、その処理を施すことは有効であると判断するのが合理的であろう。ところが、男性での2元表を

別な共変量 W (例えば年齢)によって層別したときに、各年齢層でのオッズ比が1以上になる可能性がある。このとき、(4.22)式のアナロジーで

$$Y \perp\!\!\!\perp W | (X, Z) \quad (4.23)$$

が成り立っていれば、 Z と W の組合せで層別したときの X と Y のオッズ比は、 Z のみで層別したときのオッズ比に等しい。(4.23)式を与える因果ダイアグラムとして、図13(a)を考えることができる。

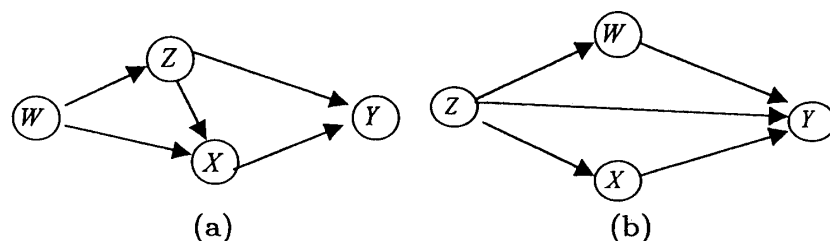


図13 4変量での因果ダイアグラムの例

図13(a)で、 Z が (X, Y) についてバックドア基準を満たしていることに注意する。しかし、分割表のオッズ比という母数においては、バックドア基準は併合可能性の十分条件にはなっていない。図13(b)はその反例である。図13(b)においても、 Z は (X, Y) についてバックドア基準を満たしているが、 Z と W の組合せで層別したときの X と Y のオッズ比が、 Z のみで層別したときのオッズ比と異なる数値例を作ることができる(森山[33])。しかし、次のようなことがいえる。

一般の分割表解析では、併合前と併合後でオッズ比の値が変わらないというのは、厳しい要求である。一方で、(4.20) (4.21)式で表されるユール・シンプソンのパラドックス(すなわちオッズ比における1以上と1以下の反転)は極めてミスリーディングな結論を導くので避けたい。このパラドックスを避ける十分条件も既に調べられており

$$X \perp\!\!\!\perp Z \quad \text{あるいは} \quad Y \perp\!\!\!\perp Z \quad (4.24)$$

であればパラドックスは生じない(Good and Mittal[18])。 $X \perp\!\!\!\perp Z$ は、処理有の割合を層間で一定にすることを意味し、予見研究での重要な指針である。一方、 $Y \perp\!\!\!\perp Z$ は反応有の割合を層間で一定にすることであり、回顧研究の設計に役立つ。

さて、 Z が (X, Y) についてバックドア基準を満たすとき、 X に対する Z 以外の任意の非子孫 W に対して

$$Y \perp\!\!\!\perp W | (X, Z) \quad \text{あるいは} \quad X \perp\!\!\!\perp W | Z \quad (4.25)$$

が成り立つ。 $Y \perp\!\!\!\perp W | (X, Z)$ が成立するときは、前述のように、オッズ比が併合可能である。一方、 $X \perp\!\!\!\perp W | Z$ のときには、(4.24)式と同じ理屈で、 Z と W の組合せで層別したときの X と Y のオッズ比と、 Z のみで層別したときのオッズ比において、ユール・シンプソンのパラドックスは生じない。このように、多元分割表におけるオッズ比の解析においても、バックドア基準は役立つ。

5. おわりに

本論文で述べた統計的因果推論とは、因果ダイアグラムを推敲するための一連の方法論とい

える。研究者のもつ定性的因果仮説が因果ダイアグラムに表現されたとき、その非巡回的有向グラフがすべての頂点对に矢線のある完全有向グラフでなければ、その定性的因果仮説は統計的仮説になる。すなわち、ある頂点对に矢線がないことが規定する独立性・条件付き独立性は、統計データのみから検証可能であり、Popper (1959) のいう科学的仮説の基準に適合する。

一方、変量に対する先験的知識と統計データから因果ダイアグラムを推論するためには、変量の先行関係とデータにおいて近似的に実現している独立性・条件付き独立性より、有向独立グラフあるいは連鎖独立グラフを統計的に推測することが有効な手段となる。経時的に測定された変量間や、共変量と処理変量の間には、明らかな先行関係がある。先行関係にある変量間の統計的従属関係に因果的解釈が可能なとき、有向独立グラフは因果ダイアグラムになる。有向独立グラフの統計的推測は、適当な変量集合に対する無向独立グラフの推測を反復することに帰着する。それは、量的データの場合は共分散選択モデリング、質的データの場合にはグラフィカル対数線形モデリングである。

無向独立グラフと有向独立グラフでのマルコフ性の理論は既に完成している。今後は新たな応用分野を開発することが課題である。例えば、素数べき型直交表実験で、主効果とともにいくつかの2因子交互作用を割り付けるための道具として線点図が知られている(田口[52])。線点図は因子を頂点とする無向グラフである。実験データを解析し、有意と判定された主効果および2因子交互作用から線点図を作成すると、無向独立グラフでのマルコフ性を利用することで、要因効果全体の構造を理解しやすくなる(宮川[28])。

3.4節に述べた有向独立グラフの推測では、SGS無向グラフを決定するための効率的なアルゴリズムに関心が持たれている。また、SGS連鎖グラフを有向独立グラフにするために、どのような実験を追加するのが最も効率的かというのも興味深い問題である。

4.2節に述べた因果ダイアグラムによる介入効果の推測論においては、分布あるいは母数としての識別可能条件が与えられたものの、有限個の統計データからの推定精度についてはほとんど調べられていない。例えば、一般にバックドア基準を満たす共変量集合は一意でないが、分布あるいは母数としての介入効果はそのような共変量集合に依らず一意に表現される。しかし、推定精度は共変量集合に依存する。宮川・黒木[29]は、2つのバックドア基準を満たす共変量集合がグラフ上で特別な関係にある場合に限って、推定精度からの共変量選択基準を与えたが、より一般的な議論は今後の課題である。

4.2節では単一の処理変量に介入する場合を定式化したが、複数の処理変量への同時介入も応用上価値の高い問題である。Pearl and Robins[39]は同時介入に対するバックドア基準を与え、Kuroki and Miyakawa[24]は同時介入に対するフロントドア基準を与えた。また、同時介入では、ある時点である処理を施し、その結果を見て次の処理条件を決めるという条件付き介入がより現実的である。この条件付き介入の推測論も現在発展中である。

参 考 文 献

- [1] Birch, M. W. (1963). Maximum likelihood in three-way contingency tables, *Jour. Roy. Statist. Soc. B*, **25**, 220-233.
- [2] Bishop, M. M. (1971). Effects of collapsing multi-dimensional contingency tables, *Biometrics*, **27**, 220-233.
- [3] Bishop, M. M., Fienberg, S. and Holland, P. (1975). *Discrete Multivariate Analysis, Theory and Practice*, MIT Press.
- [4] Bollen, K. A. (1989). *Structural Equations with Latent Variables*, John Wiley and Sons.
- [5] Breslow, N. and Day, N. E. (1980). *Statistical Methods in Cancer Research, Vol.I; The Analysis of Case-Control Studies*, IARC.

- [6] Cox, D. R. (1992). Causality: some statistical aspects, *Jour. Roy. Statist. Soc. A*, **155**, 291-301.
- [7] Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs, *Statistical Science*, **8**, 204-283 (with discussions).
- [8] Cox, D. R. and Wermuth, N. (1996). *Multivariate Dependencies; Models, Analysis and Interpretation* Chapman and Hall.
- [9] Darroch, J. N., Lauritzen, S. L. and Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables, *Ann. Statist.* **8** 522-539.
- [10] Dawid, A. P. (1979). Conditional independence in statistical theory, *Jour. Roy. Statist. Soc. B*, **41**, 1-31.
- [11] Deming, W. E. and Stephan, F. F. (1940). On a least square adjustment of a sampled frequency table when the expected marginal totals are known, *Ann. Math. Statist.*, **11**, 427-444.
- [12] Dempster, A. P. (1972). Covariance selection, *Biometrics*, **28**, 157-175.
- [13] Edwards, D. (1995). *Introduction to Graphical Modelling*, Springer-Verlag.
- [14] Fisher, R. A. (1935). *The Design of Experiments*, Oliver and Boyd.
- [15] Freedman, D. A. (1987). As others see us: a case study in path analysis (with discussion), *Jour. Educational Statist.*, **12**, 101-223.
- [16] Galles, D. J. and Pearl, J. (1998). An axiomatic characterization of Causal counterfactuals, *Foundations of Science*, **3**, 151-182.
- [17] Geiger, D., Verma, T. S. and Pearl, J. (1990). Identifying independence in Bayesian networks, *Networks*, **20**, 507-534.
- [18] Good, I. J. and Mital, Y. (1987). The amalgamation and geometry of two-by-two contingency tables, *Ann. Statist.*, **15**, 694-711.
- [19] Goodman, L. A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications, *Jour. Amer. Statist. Assoc.*, **65**, 226-256.
- [20] Haberman, S. J. (1972). Log-linear fit for contingency tables, Algorithm AS 51, *Appl. Statist.*, **21**, 218-225.
- [21] Haberman, S. J. (1974). *The Analysis of Frequency Data*, University of Chicago Press.
- [22] Holland, P. W. (1986). Statistics and causal inference, *Jour. Amer. Statist. Assoc.*, **81**, 945-970 (with discussions).
- [23] 黒木 学, 宮川雅巳(1999). 因果ダイグラムにおける介入効果の推定と工程解析への応用, *品質*, **29**, 237-247.
- [24] Kuroki, M. and Miyakawa, M. (1999). Identifiability criteria for causal effects of joint interventions, to appear in *Jour. Japan Statist. Soc.*, **29**.
- [25] Lauritzen, S. L. (1996). *Graphical Models*, Clarendon Press.
- [26] Lauritzen, S. L., Dawid, A. P., Larsen, B. N. and Leimer, H. G. (1990). Independence properties of directed Markov fields, *Networks*, **20**, 491-505.
- [27] 宮川雅巳 (1977). グラフィカルモデリング, 朝倉書店.
- [28] 宮川雅巳 (1997). 推定された線点図とその使い方, *品質*, **27**, 499-505.
- [29] 宮川雅巳, 黒木 学(1998). 因果ダイアグラムにおける介入効果推定のための共変量選択, 応用統計学会第20回シンポジウム講演予稿集, 47-52,
- [30] 宮川雅巳, 芳賀敏郎(1997). グラフィカル正規モデリングのための対話的データ解析システム, *品質*, **27**, 326-336.
- [31] 宮川雅巳, 芳賀敏郎(1996). グラフィカル対数線形モデリングによる多元分割表解析, 応用統計学会第18回シンポジウム講演予稿集, 70-75.
- [32] Mital, Y. (1991). Homogeneity of subpopulations and Simpson's paradox, *Jour. Amer. Statist. Assoc.*, **86**, 167-172.
- [33] 森山勝行(1999). 因果ダイアグラムに基づく多元分割表の併合可能性に関する研究, 東京工業大学大学院平成10年度修士論文.
- [34] Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles, *Ann. Agricultural Sciences*, 1-51, English translation by Dabrowska, D. M. and Speed, T. P. (1990). *Statistical Science*, **5**, 465-472.
- [35] 日本品質管理学会テクノメトリックス研究会編(1999). *グラフィカルモデリングの実際*, 日科技連出版.
- [36] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann.
- [37] Pearl, J. (1995). Causal diagrams for empirical research, *Biometrika* **82**, 669-709.

- [38] Pearl, J. (1998). Graphs, causality, and structural equation models, *Sociological Methods and Research*, **27**, 226-284.
- [39] Pearl, J. and Robins, J. (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables, In Besnard, P. and Hanks, S. editors, *Uncertainty in Artificial Intelligence*, **11**, 669-709, Morgan Kaufmann.
- [40] Popper, K. (1959). *The Logic of Scientific Discovery*, Harper and Row.
- [41] Rosenbaum, P. R. (1984). From association to causation in observational studies: the role of tests of strongly ignorable treatment assignment, *Jour. Amer. Statist. Assoc.*, **79**, 41-48.
- [42] Rosenbaum, P. R. (1995). *Observational Studies*, Springer-Verlag.
- [43] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity scores in observational studies for causal effects, *Biometrika*, **70**, 41-55.
- [44] Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score, *Jour. Amer. Statist. Assoc.*, **79**, 516-524.
- [45] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies, *Jour. Educational Psychology*, **66** 688-701.
- [46] Rubin, D. B. (1976). Multivariate matching methods that are equal percent bias reducing, I: some examples, *Biometrics*, **32**, 109-120.
- [47] Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies, *Jour. Amer. Statist. Assoc.*, **74**, 318-328.
- [48] Simpson, E. H. (1951). The interpretation of interaction in contingency tables, *Jour. Roy. Statist. Soc. B*, **13**, 238-241.
- [49] Speed, T. P. and Kiiveri, H. (1986). Gaussian Markov distributions over finite graphs, *Ann. Statist.*, **14**, 138-150.
- [50] Spirtes, P., Glymour, C. and Scheines, R. (1993). *Causation, Prediction, and Search*, Springer-Verlag.
- [51] Sundberg, R. (1975). Some results about decomposable (or Markov-type) models for multidimensional contingency tables: distribution of marginals and partitioning of tests, *Scand. Jour. Statist.*, **2**, 71-79.
- [52] 田口玄一 (1976). 第3版実験計画法 (上), 丸善.
- [53] 田口玄一 (1977). 第3版実験計画法 (下), 丸善.
- [54] Wermuth, N. and Lauritzen, S. L. (1983). Graphical and recursive models for contingency tables, *Biometrika*, **70**, 537-552.
- [55] Wermuth, N. and Scheidt, E. (1977). Fitting a covariance selection to a matrix, Algorithm AS 105, *Appl. Statist.*, **26**, 88-92.
- [56] Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, John Wiley.
- [57] Whittemore, A. S. (1978). Collapsibility of multidimensional contingency tables, *Jour. Roy. Statist. Soc. B*, **40**, 328-394.
- [58] Wright, S. (1923). The theory of path coefficients: a reply to Niles' criticism, *Genetics*, **8**, 239-255.
- [59] Wright, S. (1934). The method of path coefficients, *Ann. Math. Statist.*, **5**, 161-215.
- [60] 柳川 堯 (1986). 離散多変量データの解析, 共立出版.
- [61] Yule, G. U. (1903). Notes on the theory of association of attributes in statistics, *Biometrika*, **2**, 121-134.