構造方程式モデル

黒木　学

構造方程式モデルは，複数の変数間の従属関係を誤差の付随した線形・非線形方程式によって表現した統計モデルの1つであり，その起源は遺伝学におけるWrightのパス解析にまでさかのぼることができる．その後，構造方程式モデルは社会科学の分野で大きく発展し，現在では工学・医学・情報科学の分野においても利用されている．本稿で解説する線形構造方程式モデルは，構造方程式モデル理論の基礎をなすものである．

1．線形構造方程式モデルとは

構造方程式モデルは，複数の変数間の従属関係を誤差の付随した線形・非線形方程式によって表現した統計モデルであり，その起源は遺伝学におけるWright[6][7]のパス解析にまでさかのぼることができる．その後，構造方程式モデルは社会科学の分野で大きく発展し，現在では工学・医学・情報科学の分野においても利用されている（例えば，Bollen[1]，黒木[3]，宮川[4]，Pearl[5]）．特に，本稿で解説する線形構造方程式モデルは，構造方程式モデル理論の基礎をなすものであり，近年では，ノンパラメトリック構造方程式モデルの観点から再検討され，新たな概念が生み出されている．

さて，線形構造方程式モデルの理論ではグラフを用いてその性質が議論されることが多い．このことから，グラフ用語を用いて線形構造方程式モデルの基本的フレームワークを解説する．

2．グラフ用語

グラフ G は頂点の集合 $\boldsymbol{V}$ と，その直積 $\boldsymbol{V}\times\boldsymbol{V}$ の部分集合である矢線の集合 $\boldsymbol{E}$ によって，$G=(\boldsymbol{V},\boldsymbol{E})$ として表現される．2つの頂点 $\alpha,\beta\in\boldsymbol{V}$ に対して $(\alpha,\beta)\in\boldsymbol{E}$ かつ $(\beta,\alpha)\notin\boldsymbol{E}$ のとき，α から β に向きのある有向の辺（矢線）を引く．矢線のみから構成されるグラフを有向グラフという．α から β への矢線が存在するとき，α は β の親であるといい，β は α の子であるという．β の親全体からなる集合を $pa(\beta)$ と記す．異なる頂点の列 $\alpha_0,\alpha_1,\cdots,\alpha_n$ は，すべての $i=1,\cdots,n$ で $(\alpha_{i-1},\alpha_i)\in\boldsymbol{E}$ または $(\alpha_i,\alpha_{i-1})\in\boldsymbol{E}$ であるとき，長さ n の道という．特に，長さ n の道で，すべての $i=1,\cdots,n$ に対して $(\alpha_{i-1},\alpha_i)\in\boldsymbol{E}$ かつ $(\alpha_i,\alpha_{i-1})\notin\boldsymbol{E}$ であるとき，有向道という．長さ n の道 $\alpha_0,\alpha_1,\cdots,\alpha_n$ で，$(\alpha_{i-1},\alpha_i)\in\boldsymbol{E}$ かつ $(\alpha_{i+1},\alpha_i)\in\boldsymbol{E}$ であるとき α_i を合流点といい，そうでないとき，α_i を非合流点という．

3．モデルの概要

有向グラフ G とその頂点に対応する連続型確率変数の集合 $\boldsymbol{V}=\{V_1,V_2,\cdots,V_m\}$ が与えられている．グラフ G が確率変数間の関数関係を

$$V_i=\sum_{v_j\in pa(v_i)}\alpha_{v_iv_j}V_j+\epsilon_i\ (i=1,2,\cdots,m) \tag{1}$$

なる形に規定するとき，(1)式を線形構造方程式モデルとよび，グラフ G をパスダイアグラムという．本稿では，誤差変数 $\epsilon_1,\epsilon_2,\cdots,\epsilon_m$ は多変量正規分布にしたがうものとし，その平均ベクトルは $\boldsymbol{0}$ で，$\boldsymbol{V}$ に含まれる変数の分散は1であると仮定する．また，$pa(v_i)$ は V_i に対する説明変数からなる変数集合で V_i の親集合である．さらに，$\alpha_{v_iv_j}$ は V_j から V_i への矢線に対応する係数でパス係数，あるいは直接効果と呼ばれるもので，$V_j\in pa(V_i)$ に対して $\alpha_{v_iv_j}\neq 0$ とする．パスダイアグラムと線形構造方程式モデルとは一対一に対応する．構造方程式モデルは，$\boldsymbol{V}$ の要素間に一定方向の（半）順序関係がある逐次モデルと，フィードバック・ループを含む非逐次モデルに大別される．以下では，逐次モデルを対象に解説を行う．

例として，図1のパスダイアグラムを考えよう．図1に対応する構造方程式モデルは逐次モデルとなり

くろき　まなぶ
大阪大学 基礎工学研究科
〒560-8531 豊中市待兼山町 1-3

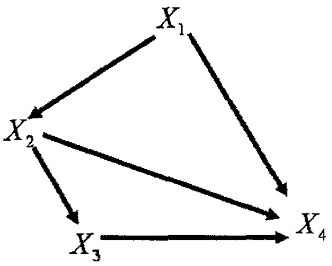

図1　パスダイアグラム

$$\left.\begin{aligned} V_4 &= \alpha_{v_4v_1}V_1+\alpha_{v_4v_2}V_2+\alpha_{v_4v_3}V_3+\epsilon_{v_4} \\ V_3 &= \alpha_{v_3v_2}V_2+\epsilon_{v_3} \\ V_2 &= \alpha_{v_2v_1}V_1+\epsilon_{v_2} \\ V_1 &= \epsilon_{v_1} \end{aligned}\right\} \quad (2)$$

で与えられる．

逐次モデルの場合，すべての変量が観測されていればパス係数は通常の回帰分析を逐次的に利用，あるいは最尤法を用いることにより推定することができる．一方，非逐次モデルの場合には，識別不足などの問題により，逐次モデルと比べてパス係数の推定が複雑となることが少なくない．

4.　モデルの性質

線形構造方程式モデルが持つ重要な特徴のひとつとして，「直接効果と総合効果の区別」がある．V_j から V_i への有向道のそれぞれについてパス係数すべての積を考えるとき，その総和を V_j から V_i への総合効果とよぶ．例えば，図1のパスダイアグラムにおいて，V_3 から V_4 への総合効果は，V_3 から V_4 への直接効果 $\alpha_{v_4v_3}$ と一致しているが，V_2 から V_4 への総合効果は $\alpha_{v_4v_2}+\alpha_{v_4v_3}\alpha_{v_3v_2}$ であり，V_2 から V_4 への直接効果 $\alpha_{v_4v_2}$ とは異なっている．統計的因果モデルとして線形構造方程式モデルが仮定できるとき，総合効果はある変数 V_i を外的操作により1単位変化させたときのもうひとつの変数 V_j の期待値の変化量を意味する(例えば，黒木[3]，宮川[4]，Pearl[5])．総合効果の推定法として，操作変数法（Bowden and Turkington[2]），バックドア基準（黒木[3]，Pearl[5]）などがある．

線形構造方程式モデルが持つもうひとつの重要な特徴として，「パス係数による変量間の相関係数の分解」がある．例として，図1のパスダイアグラムが与えられたとき，V_2 と V_4 の相関係数 $\rho_{v_2v_4}$ は，(1)式より

$$\rho_{v_2v_4}=\alpha_{v_4v_2}+\alpha_{v_4v_3}\alpha_{v_3v_2}+\alpha_{v_4v_1}\alpha_{v_2v_1} \quad (3)$$

と表現することができる．(3)式の第一項は V_2 から V_4 への直接効果である．第二項は V_2 から V_4 への有向道のうち V_3 を経由する道によって生成される統計的関連性の強さを意味しており，間接効果と呼ばれる．$\alpha_{v_4v_1}\alpha_{v_2v_1}$ は有向道によって生じることのない統計的関連性の強さを意味しており，疑似相関と呼ばれる．このことから，相関係数は直接効果，間接効果，疑似相関に分解することができることがわかる．一方，V_1 と V_3 の相関係数 $\rho_{v_1v_3}$ は，$\rho_{v_1v_3}=\alpha_{v_3v_2}\alpha_{v_2v_1}$ と記述できる．V_1 と V_3 の間には矢線がないので直接効果は0である．また，V_1 から V_3 への有向道に対応する間接効果のみによって相関が生成されており，V_4 を経由する道に対応する統計的関連性は含まれていないことがわかる．一般に，相関係数は直接効果，間接効果，擬相関によって記述することができ，合流点を含む道からは生成されないことが知られている．

参考文献

[1] Bollen, K. A.: *Structural Equations with Latent Variables*, Wiley-Interscience (1989).

[2] Bowden, R. J. and Turkington, D. A.: *Instrumental Variables*, Cambridge University Press (1984).

[3] 黒木学:「統計的因果推論―モデル・推論・推測―」, 共立出版 (2009).

[4] 宮川雅巳:「統計的因果推論―回帰分析の新しい枠組み―」, 朝倉書店 (2004).

[5] Pearl, J.: *Causality: Models, Reasoning, and Inference, The 2nd Edition*, Cambridge University Press, Wiley-Interscience (2009).

[6] Wright, S.: The theory of path coefficients: a reply to Niles' criticism, *Genetics*, 8 (1923) 239-255.

[7] Wright, S.: The method of path coefficients, *Annals of Mathematical Statististics*, 5 (1934) 161-215.