

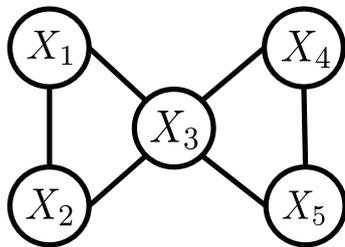
## グラフィカルモデルの選択における KOO 法の適用

公立諏訪東京理科大・共通・マネジメント教育センター 櫻井 哲朗  
 広島大・理・名誉教授 藤越 康祝

本報告では、変数間の関係性を表すグラフィカルモデルについて関心がある。ここでは、変数間の独立または条件付き独立によってモデル化することを考える。このとき、変数を量的変数とし、さらにその変数に多変量正規分布を仮定することで、変数間の条件付き独立は偏相関係数がゼロかゼロでないかの問題として扱うことができる。つまり、 $p$  次元多変量正規分布に従う確率変数ベクトル  $\mathbf{X}$  において、 $X_i$  と  $X_j$  以外の変数をまとめた  $\mathbf{X}_{\overline{ij}}$  を与えたもとでの  $X_i$  と  $X_j$  の母偏相関係数  $\rho_{ij|\overline{ij}}$  は次のような関係をもつ。

$$\mathbf{X}_{\overline{ij}} \text{ を与えたもとで } X_i \text{ と } X_j \text{ が独立である: } X_i \perp X_j | \mathbf{X}_{\overline{ij}} \Leftrightarrow \rho_{ij|\overline{ij}} = 0$$

グラフィカルモデルでは、このような変数間の関係を変数を頂点としたグラフとして表す。偏相関係数がゼロでない変数間同士を線で結び、偏相関係数がゼロのところは結ばない。例えば、下図のような場合は  $X_1$  と  $X_4$  は線で結ばれていないので偏相関係数がゼロであることを表しており、他にも線で結ばれていないところは偏相関係数がゼロに対応する。



対応する偏相関行列:

$$\begin{pmatrix} - & & & & \\ \rho_{12|12} & - & & & \\ \rho_{13|13} & \rho_{23|23} & - & & \\ 0 & 0 & \rho_{34|34} & - & \\ 0 & 0 & \rho_{35|35} & \rho_{45|45} & - \end{pmatrix}$$

このようにして、変数間の関係は多変量正規分布  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  のもとでは共分散構造の問題としてとらえることができ、それはさらに線をどのように結ぶかという図形的な問題として表すことができる。このとき、いろいろなモデルを考えることができ、 $p$  変数あればそのモデル数は  $2^{p(p-1)/2}$  にのぼる。ここでは、AIC のようなモデル選択規準量を与えることで最適なモデルを見つける。しかし、 $p$  が増えれば候補のモデルは膨大な数となり、その中から最適なモデルを見つけるのは計算が困難となる。そこで本報告では、Zhao et al.(1986) 等で提案された計算量が少ない新たな変数選択法を用いて、この問題を解決する。ここでは、この選択法を KOO 法と呼ぶ。

KOO 法は、次のような方法のもと最適なモデルを決める。標本数  $n$  として  $p$  個の変数があり、この変数の中から最適なモデルを構築することを考える。このとき、 $i$  番目、 $j$  番目の変数に対して次の規則を用いて最適なモデルを決める。

$$\begin{cases} MV_{\overline{ij}} - MV_{\text{Full}} > d & \Rightarrow i \text{ 番目, } j \text{ 番目の変数を線で結ぶ} & i, j = 1, \dots, p, \\ MV_{\overline{ij}} - MV_{\text{Full}} \leq d & \Rightarrow i \text{ 番目, } j \text{ 番目の変数を線で結ばない} & i \neq j \end{cases}$$

ここで、 $MV_{\text{Full}}$  は全ての変数を線で結んだモデルから計算される値であり、 $MV_{\overline{ij}}$  は全ての変数を線で結んだモデルから  $i$  番目、 $j$  番目の変数間の線を取り除いたモデルから計算される値である。上記の KOO 法が高次元漸近的枠組み  $p/n \rightarrow c < 1$  のもとで、一致性を持つための十分条件について考察し、その結果に対し数値実験によって妥当性を検証する。

### 参考文献

1. ZHAO, L. C. , KRISHNAIAH, P. R. and BAI, Z. D. (1986). On determination of the number of signals in presence of white noise. *Journal of Multivariate Analysis*, 20, 1-25.