

## VI-2-4. 因子分析 (Factor analysis : FA)

### VI-2-4-1. 因子分析とは何か

たとえば、重回帰は、

$$y = a_1x_1 + a_2x_2 + \cdots + a_px_p + e$$

あるいは、行列で表現すると

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \quad \mathbf{A} = (a_1 \quad \cdots \quad a_p)$$

$$\mathbf{Y} = \mathbf{XA}^T + \mathbf{E}$$

のようなモデルで、 $\mathbf{Y}$ と $\mathbf{X}$ が与えられて $\mathbf{A}$ を推定します。そうではなくて、因子分析では、 $\mathbf{Y}$ が与えられて、 $\mathbf{X}$ と $\mathbf{A}$ を推定します。もちろん、一意的にそんな解を求めることはできませんから、観測変数を複数にして、

$$\mathbf{Y} = \begin{pmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nm} \end{pmatrix}$$

これを手掛かりに、確率的にありそうな

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mp} \end{pmatrix}$$

を求めます。それが因子分析がやっていることです。重回帰分析は、被説明変数(目的変数)と複数の説明変数の関係性を分析しています。主成分分析は複数の変数間の関係性(データ構造)の把握です。因子分析では、複数の観測変数が与えられて、その関係を使って、観察されない潜在的な変数(因子)を推定するのです。海の観測をイメージしてください。普通観測データは、海の表面でしか獲れませんが、その表面的な観察変数から、海の底の直接観察されない現象(海流や生物の存在など)を推定しようとしています。コンピュータの計算力を使って、構造なども含めて、直接観察されない現象間のメカニズムを抽出する分析です。だから、因子分析は現代的な統計学発達の契機となったのです。因子分析は、主成分分析と対比して説明されることもあります。しかし、主成分分析は、データを並び替えてデータの構造を説明しているので、ランダム変動を成分の中に含んだままです。因子分析は、ランダムな変動を外に取り出して、潜在的な因子を見つけ出す分析です。

### VI-2-4-2. 因子分析のモデル

統計分析では、データは次のような構造になっていると考えています。 $\mathbf{Z}$ は結果として現れる現象、 $\mathbf{X}$ はその現象を説明する変数(因子得点)で、 $\mathbf{A}$ は現象を構成する説明変数間の関係、 $\mathbf{E}$ は説明変数の関係で説明できない部分です。 $\mathbf{E}$ を誤差、残差、独立因子などと呼びます。因子分析では、与えられた $\mathbf{Z}$ から $\mathbf{E}$ 最適化して $\mathbf{A}$ および $\mathbf{X}$ を推測します。 $\mathbf{Z}$ は複数の測定項目からなるサイズ $n$ のデータです。

$$\mathbf{Z} = \mathbf{X}\mathbf{A}^T + \mathbf{E}$$

$$\mathbf{Z} = \begin{pmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{pmatrix}_{n \times m}$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}_{n \times p}$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mp} \end{pmatrix}$$

$$\mathbf{E} = \begin{pmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nm} \end{pmatrix}_{n \times m}$$

これらを使って

$$\mathbf{A}^T = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1p} & \cdots & a_{mp} \end{pmatrix}$$

$$\mathbf{E} = \begin{pmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nm} \end{pmatrix} = \mathbf{Z} - \mathbf{X}\mathbf{A}^T = \begin{pmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{pmatrix} - \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1p} & \cdots & a_{mp} \end{pmatrix}$$

$$= \begin{pmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{pmatrix} - \begin{pmatrix} \sum_{k=1}^p a_{1k}x_{1k} & \sum_{k=1}^p a_{2k}x_{1k} & \cdots & \sum_{k=1}^p a_{mk}x_{1k} \\ \sum_{k=1}^p a_{1k}x_{2k} & \sum_{k=1}^p a_{2k}x_{2k} & \cdots & \sum_{k=1}^p a_{mk}x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^p a_{1k}x_{nk} & \sum_{k=1}^p a_{2k}x_{nk} & \cdots & \sum_{k=1}^p a_{mk}x_{nk} \end{pmatrix}$$

$$= \begin{pmatrix} z_{11} - \sum_{k=1}^p a_{1k}x_{1k} & z_{12} - \sum_{k=1}^p a_{2k}x_{1k} & \cdots & z_{1m} - \sum_{k=1}^p a_{mk}x_{1k} \\ z_{21} - \sum_{k=1}^p a_{1k}x_{2k} & z_{22} - \sum_{k=1}^p a_{2k}x_{2k} & \cdots & z_{2m} - \sum_{k=1}^p a_{mk}x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} - \sum_{k=1}^p a_{1k}x_{nk} & z_{n2} - \sum_{k=1}^p a_{2k}x_{nk} & \cdots & z_{nm} - \sum_{k=1}^p a_{mk}x_{nk} \end{pmatrix}$$

$$e_{ij} = z_{ij} - \sum_{k=1}^p a_{jk}x_{ik} = z_{ij} - (x_{i1} \quad \cdots \quad x_{ip}) \begin{pmatrix} a_{j1} \\ \vdots \\ a_{jp} \end{pmatrix} = z_{ij} - \mathbf{x}_i \mathbf{a}_j^T$$

ここで、 $\mathbf{x}_i = (x_{i1} \quad \cdots \quad x_{ip})$  と  $\mathbf{a}_j = (a_{j1} \quad \cdots \quad a_{jp})$  がベクトルであると考えます。このこと

によって、 $e_{ij}$ は $\mathbf{x}_i$ と $\mathbf{a}_j$ の二つのベクトルの関数になります  $z_{ij}$ は定数で、あらかじめ与え

られたもので、それ自身は $\mathbf{x}_i$ や $\mathbf{a}_j$ を含みません。

$$\begin{aligned}
e_{ij} &= f_{ij}(\mathbf{a}_j, \mathbf{x}_i) \\
\mathbf{E}\mathbf{E}^T &= \begin{pmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nm} \end{pmatrix} \begin{pmatrix} e_{11} & \cdots & e_{n1} \\ \vdots & \ddots & \vdots \\ e_{1m} & \cdots & e_{nm} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^m (e_{1j})^2 & \sum_{j=1}^m e_{1j}e_{2j} & \cdots & \sum_{j=1}^m e_{1j}e_{nj} \\ \sum_{j=1}^m e_{2j}e_{1j} & \sum_{j=1}^m (e_{2j})^2 & \cdots & \sum_{j=1}^m e_{2j}e_{nj} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^m e_{in}e_{i1} & \sum_{j=1}^m e_{in}e_{i2} & \cdots & \sum_{i=1}^m (e_{in})^2 \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^m (f_{1j}(\mathbf{a}_j, \mathbf{x}_1))^2 & \sum_{j=1}^m (f_{1j}(\mathbf{a}_j, \mathbf{x}_1))(f_{2j}(\mathbf{a}_j, \mathbf{x}_2)) & \cdots & \sum_{j=1}^m (f_{1j}(\mathbf{a}_j, \mathbf{x}_1))(f_{nj}(\mathbf{a}_j, \mathbf{x}_n)) \\ \sum_{j=1}^m (f_{2j}(\mathbf{a}_j, \mathbf{x}_2))(f_{1j}(\mathbf{a}_j, \mathbf{x}_1)) & \sum_{j=1}^m (f_{2j}(\mathbf{a}_j, \mathbf{x}_2))^2 & \cdots & \sum_{j=1}^m (f_{2j}(\mathbf{a}_j, \mathbf{x}_2))(f_{nj}(\mathbf{a}_j, \mathbf{x}_n)) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^m (f_{nj}(\mathbf{a}_j, \mathbf{x}_n))(f_{1j}(\mathbf{a}_j, \mathbf{x}_1)) & \sum_{j=1}^m (f_{nj}(\mathbf{a}_j, \mathbf{x}_n))(f_{2j}(\mathbf{a}_j, \mathbf{x}_2)) & \cdots & \sum_{j=1}^m (f_{nj}(\mathbf{a}_j, \mathbf{x}_n))^2 \end{pmatrix} \\
\mathbf{E}^T\mathbf{E} &= \begin{pmatrix} e_{11} & \cdots & e_{n1} \\ \vdots & \ddots & \vdots \\ e_{1m} & \cdots & e_{nm} \end{pmatrix} \begin{pmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nm} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n (e_{i1})^2 & \sum_{i=1}^n e_{i1}e_{i2} & \cdots & \sum_{i=1}^n e_{i1}e_{im} \\ \sum_{i=1}^n e_{i2}e_{i1} & \sum_{i=1}^n (e_{i2})^2 & \cdots & \sum_{i=1}^n e_{i2}e_{im} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n e_{im}e_{i1} & \sum_{i=1}^n e_{im}e_{i2} & \cdots & \sum_{i=1}^n (e_{im})^2 \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n (f_{i1}(\mathbf{a}_1, \mathbf{x}_i))^2 & \sum_{i=1}^n (f_{i1}(\mathbf{a}_1, \mathbf{x}_i))(f_{i2}(\mathbf{a}_2, \mathbf{x}_i)) & \cdots & \sum_{i=1}^n (f_{i1}(\mathbf{a}_1, \mathbf{x}_i))(f_{im}(\mathbf{a}_m, \mathbf{x}_i)) \\ \sum_{i=1}^n (f_{i2}(\mathbf{a}_2, \mathbf{x}_i))(f_{i1}(\mathbf{a}_1, \mathbf{x}_i)) & \sum_{i=1}^n (f_{i2}(\mathbf{a}_2, \mathbf{x}_i))^2 & \cdots & \sum_{i=1}^n (f_{i2}(\mathbf{a}_2, \mathbf{x}_i))(f_{im}(\mathbf{a}_m, \mathbf{x}_i)) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n (f_{im}(\mathbf{a}_m, \mathbf{x}_i))(f_{i1}(\mathbf{a}_1, \mathbf{x}_i)) & \sum_{i=1}^n (f_{im}(\mathbf{a}_m, \mathbf{x}_i))(f_{i2}(\mathbf{a}_2, \mathbf{x}_i)) & \cdots & \sum_{i=1}^n (f_{im}(\mathbf{a}_m, \mathbf{x}_i))^2 \end{pmatrix}
\end{aligned}$$

ここで、 $\mathbf{E}$ の各行と列を次のような横ベクトル、縦ベクトルだと考えます。

$$\mathbf{e}_i = (e_{i1} \quad \cdots \quad e_{im}) \text{ and } \boldsymbol{\varepsilon}_j = \begin{pmatrix} e_{1j} \\ \vdots \\ e_{nj} \end{pmatrix}$$

このように考えると行列 $\mathbf{E}\mathbf{E}^T$ と行列 $\mathbf{E}^T\mathbf{E}$ の各因子はベクトルの内積です。また、両ベクトルの対角因子は同一ベクトルの内積で、ベクトルの長さの2乗です。因子分析では、 $\mathbf{E}$ の最適化を、これらのベクトルの最適化することによって、誤差を最小にする、あるいは、推測値の妥当性の確率を最大化することだと考えます。最小二乗法で計算するのであれば、行列 $\mathbf{E}\mathbf{E}^T$ と行列 $\mathbf{E}^T\mathbf{E}$ の対角成分の和（トレース  $tr$ ）を最小化します。

$$tr(\mathbf{E}\mathbf{E}^T) = \mathbf{e}_1\mathbf{e}_1^T + \mathbf{e}_2\mathbf{e}_2^T + \cdots + \mathbf{e}_n\mathbf{e}_n^T = \sum_{i=1}^n \mathbf{e}_i\mathbf{e}_i^T = \sum_{i=1}^n \sum_{j=1}^m \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2$$

$$tr(\mathbf{E}^T\mathbf{E}) = \boldsymbol{\varepsilon}_1^T\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2^T\boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_m^T\boldsymbol{\varepsilon}_m = \sum_{j=1}^m \boldsymbol{\varepsilon}_j^T\boldsymbol{\varepsilon}_j = \sum_{j=1}^m \sum_{i=1}^n \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2$$

最尤法であれば、対角成分の誤差の確率の総乗を最大化します。

$$\begin{aligned} Prob(\mathbf{E}\mathbf{E}^T) &= \prod_{i=1}^n Prob(\mathbf{e}_i\mathbf{e}_i^T) = \prod_{i=1}^n Prob \sum_{j=1}^m \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2 \\ Prob(\mathbf{E}^T\mathbf{E}) &= \prod_{j=1}^m Prob(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^T) = \prod_{j=1}^m Prob \sum_{i=1}^n \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2 \end{aligned}$$

実際には、そのままでは微分できないので、対数をとることによって変数を分離します。

$$\begin{aligned} \log_e Prob(\mathbf{E}\mathbf{E}^T) &= \log_e \prod_{i=1}^n Prob \sum_{j=1}^m \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2 = \sum_{i=1}^n \log_e Prob \sum_{j=1}^m \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2 \\ \log_e Prob(\mathbf{E}^T\mathbf{E}) &= \log_e \prod_{j=1}^m Prob \sum_{i=1}^n \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2 = \sum_{j=1}^m \log_e Prob \sum_{i=1}^n \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2 \end{aligned}$$

ここで、 $\mathbf{a}_j$ を与えられた係数  $tr(\mathbf{E}\mathbf{E}^T)$ と $\log_e Prob(\mathbf{E}\mathbf{E}^T)$ を  $n$  個の項からなる多項式だと考えると、 $i$ 番目の項は $\mathbf{x}_i$ の関数で、 $i$ 番目以外のベクトルを含みません。ですから、個々の項ごとに微分したものを0にすることで、全体の $tr(\mathbf{E}\mathbf{E}^T)$ と $\log_e Prob(\mathbf{E}\mathbf{E}^T)$ の微分値を0とすることが出来ます。同様に、 $\mathbf{x}_i$ を与えられた係数とすれば、 $tr(\mathbf{E}^T\mathbf{E})$ と $\log_e Prob(\mathbf{E}^T\mathbf{E})$ はともに  $m$  個の項からなる多項式で、 $j$ 番目の項は $\mathbf{a}_j$ の関数で、 $j$ 番目以外のベクトルを含みません。個々の項ごとに微分したものを0にすることで、全体の $tr(\mathbf{E}^T\mathbf{E})$ と $\log_e Prob(\mathbf{E}^T\mathbf{E})$ の微分値を0とすることが出来ます。

$$\frac{d(tr(\mathbf{E}\mathbf{E}^T))}{d\mathbf{X}} = \sum_{i=1}^n \frac{\partial \left( \sum_{j=1}^m \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2 \right)}{\partial \mathbf{x}_i}$$

$$\frac{d(\log_e Prob(\mathbf{E}\mathbf{E}^T))}{d\mathbf{X}} = \sum_{i=1}^n \frac{\partial \left( \log_e \sum_{j=1}^m \left(f_{ij}(\mathbf{a}_j, \mathbf{x}_i)\right)^2 \right)}{\partial \mathbf{x}_i}$$

$$\frac{d(\text{tr}(\mathbf{E}^T \mathbf{E}))}{d\mathbf{A}} = \sum_{j=1}^m \frac{\partial \left( \sum_{i=1}^n (f_{ij}(\mathbf{a}_j, \mathbf{x}_i))^2 \right)}{\partial \mathbf{a}_j}$$

$$\frac{d(\log_e \text{Prob}(\mathbf{E}^T \mathbf{E}))}{d\mathbf{A}} = \sum_{j=1}^m \frac{\partial \left( \log_e \sum_{i=1}^n (f_{ij}(\mathbf{a}_j, \mathbf{x}_i))^2 \right)}{\partial \mathbf{a}_j}$$

これらのことから、 $\mathbf{E}$ の最適化を行列 $\mathbf{E}\mathbf{E}^T$ の対角因子の最適化とすれば、 $\mathbf{A}$ を所与の行列として与えて、最適化された $\mathbf{X}$ が得られます。反対に $\mathbf{X}$ を所与の行列として与えれば、最適化された $\mathbf{A}$ が得られます。これを繰り返せば $\mathbf{X}$ 、 $\mathbf{A}$ が最適化されます。もう一つの最適化として、 $\mathbf{X}$ を縦ベクトルを並べた行列としてみて、これらの直交化するために、行列 $\mathbf{E}\mathbf{E}^T$ の非対角成分を0にするということも考えられますが、これは直交性（因子間の独立性）にのみ着目した最適化です。この二つの最適化は、いつでも同時に成り立つわけではありません。因子分析では、対角成分の最適化だけを考えます。それが、因子分析のモデルです。そもそも、因子は独立したものですから、それらは直交関係であることを期待することはできますが、いつでも相関を持たない直交関係にあるとは言えません。それが、因子分析です。数学的なモデルとして整理すると次のようになります。

$$\mathbf{E} = \begin{pmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nm} \end{pmatrix} = \mathbf{Z} - \mathbf{X}\mathbf{A}^T$$

$$e_{ij} = f_{ij}(\mathbf{a}_j, \mathbf{x}_i)$$

$$\boldsymbol{\psi} = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m (f_{1j}(\mathbf{a}_j, \mathbf{x}_1))^2 & 0 & \cdots & 0 \\ 0 & \sum_{j=1}^m (f_{2j}(\mathbf{a}_j, \mathbf{x}_2))^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{j=1}^m (f_{nj}(\mathbf{a}_j, \mathbf{x}_n))^2 \end{pmatrix}$$

$$\boldsymbol{\varphi} = \begin{pmatrix} \varphi_1 & 0 & \cdots & 0 \\ 0 & \varphi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varphi_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n (f_{i1}(\mathbf{a}_1, \mathbf{x}_i))^2 & 0 & \cdots & 0 \\ 0 & \sum_{i=1}^n (f_{i2}(\mathbf{a}_2, \mathbf{x}_i))^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{i=1}^n (f_{im}(\mathbf{a}_m, \mathbf{x}_i))^2 \end{pmatrix}$$

最適化の方法として、最小二乗法では $\psi_1, \dots, \psi_m$  and  $\varphi_1, \dots, \varphi_n$ を最小化し、最尤法では $\log_e \text{Prob}(\psi_1), \dots, \log_e \text{Prob}(\psi_m)$ と $\log_e \text{Prob}(\varphi_1), \dots, \log_e \text{Prob}(\varphi_n)$ を最大化します。いずれにしても、計算するときは、適当な $\mathbf{A}$ あるいは $\mathbf{X}$ を所与のものとして与える必要があります。 $\mathbf{A}$ あるいは $\mathbf{X}$ を、最初にどのように与えるか、それについて様々な考え方や方法があります。因子抽出に様々な方法があるのはそのためです。その方法に違いはあっても、求めようとしている数理的なモデルはおなじです。一般的ではありませんが、一つの簡便なやり方と

して、初めに主成分分析の主成分を用いるという方法があります。分散の大きいいくつかの主成分が、全体を説明する大きな因子に近いだろうと考えることは、全く根拠がないわけではありません。しかし、主成分分析では誤差を外側に取り出しているのではなくて、マイナーな要因を無視しているだけですから、主成分は因子分析が求めようとしている因子ではありません。また、主成分は直交しています。したがって、主成分を因子として因子抽出を行うことはあまり勧められたことではありませんが、初期の因子分析のソフトでは、この方法がデフォルトとして使われています。

### VI-2-4-3. 因子抽出

上記の説明で、因子分析が想定しているモデルと解の形がわかりましたが、因子分析では説明すべき目的変数のベクトルが複数ありますから、線形代数的に、 $n$ 次元空間上に一次結合としてできたベクトルの先端と、観測変数のベクトルの先端を結ぶベクトルを残渣変数のベクトルとして、その長さを最小化するために残差ベクトルと一次結合して出来たベクトルを直交化するという方法で、決定論的に1段階で唯一解を求めることはできません。そうではなくて、説明変数（因子）を一次結合するための各ベクトルの重さを暫定的に与えて、個々の因子の座標空間中の方向（説明変数のベクトルの単位ベクトル）を最適化し、最適化した個々のベクトルを使って、ベクトルの重みづけを最適化するというのを、ある値の範囲に収束するまで繰り返して、漸近的に最適な重み（係数）と説明変数のベクトル（因子）を求めます。

分析の基本的な流れはこの通りなのですが、いろいろな計算の仕方があって、その説明が体系化されていないので、何を選択したらよいのか、私たちを混乱させます。しかし、大きく分けると、二つの方法しかありません。最尤法と最小二乗法です。最尤法でやるか最小二乗法でやるかは、データの分布の特性（分布の形）をどのように仮定しているかという問題で、本質的な問題です。

最小二乗法が想定しているデータの分布は、偏りなく、分布の中心をピークとして、左右平等（シンメトリック）にデータの頻度が低下している分布です。ですから、中心からの距離を、分布確率の代替変数として考えて、その最小化が確率の最大化だと捉えることが出来るのです。

最尤法は、直接に確率を計算します。したがって、データが標準化してあれば、データに偏りがあっても問題になりません。しかし、その計算には厄介な問題があります。確率を極大化する説明変数の値とは、導関数が0になる説明変数の値です。導関数を求めるために、確率の総乗を微分しますが、それは直接計算できないので、対数をとって、対数確率を微分します。対数をとってみると、和の対数になる項が出てきてしまいます。対数の総和は、個々の対数の微分の和ですから微分できますが、総和の対数は微分できないという問題です。この問題を、イエンゼンの不等式を使って、対数の総和に書き換えます。これは、ある仮定を前提とした近似計算なので、仮定した前提条件をデータが満たしていない

場合に、不適解が出ます。不適解とは、ある観測変数にかかわる変数の分散の合計が観測された実際の分散を超えてしまうことです。これだと誤差分散がマイナスになります。そんなことはあり得ないので、そのような解は不適解です。この問題が、因子抽出における抽出方法の問題の本質です。

複数の被説明変数が与えられて、変数と係数を同時に推定するのですから、一段階で唯一解を決定的な解として求めることは、どちらの方法でもできません。暫定的な係数のセットを与えて、その条件の下で最適な変数を推定し、その変数を与えて、係数を推定するということを繰り返して、漸近的に解に接近する方法を考えます。これを EM アルゴリズム (Expectation maximization algorithm) と言います。この方法はコンピュータの計算力に依存した計算です。EM アルゴリズムは最尤法との関係で説明されることが多いのですが、最尤法の時にのみに使える考え方ではありません。そうした過程を省略するためのアイデアはいくつかありますが、いずれも、与えられたデータは目的変数であるという、因子分析の基本的な考え方から外れていて、不適解が出てくる場合の代替的な方法です。あまり本質的ではないと思うので、ここでは、漸近的接近による方法を説明します。

## 最小二乗法による接近

### E-ステップ

因子分析のモデルは次のようになっています。

$$\begin{aligned} \mathbf{E} &= \mathbf{Z} - \mathbf{X}\mathbf{A}^T = \begin{pmatrix} e_{11} & \cdots & e_{1m} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nm} \end{pmatrix} \\ e_{ij} &= f_{ij}(\mathbf{a}_j, \mathbf{x}_i) \\ \boldsymbol{\psi} &= \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m (f_{1j}(\mathbf{a}_j, \mathbf{x}_1))^2 & 0 & \cdots & 0 \\ 0 & \sum_{j=1}^m (f_{2j}(\mathbf{a}_j, \mathbf{x}_2))^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{j=1}^m (f_{nj}(\mathbf{a}_j, \mathbf{x}_n))^2 \end{pmatrix} \end{aligned}$$

最小二乗法の最初の E-ステップでは、暫定的に **A** を与えて、独立因子の対角行列 **ψ** を最小二乗法で最適化することによって  $\mathbf{x}_n$  の期待値を求めます。  $f_{ij}(\mathbf{a}_j, \mathbf{x}_i)$  という関数を具体的に書いてみます。

$$f_{ij}(\mathbf{a}_j, \overline{\mathbf{x}}_i) = z_{ij} - \overline{\mathbf{x}}_i \mathbf{a}_j^T$$

ここでは、 $\overline{\mathbf{x}}_i$  は  $\mathbf{x}_i$  の期待値です。  $\mathbf{x}_i$ 、 $\mathbf{a}_j$  をベクトルとして表現すると次のようになります。

$$\begin{aligned} \overline{\mathbf{x}}_i &= (\overline{x_{i1}} \quad \cdots \quad \overline{x_{ip}}) \\ \mathbf{a}_j &= (a_{j1} \quad \cdots \quad a_{jp}) \end{aligned}$$

これらが、直交座標系の座標だとして、

$$f_{ij}(\mathbf{a}_j, \mathbf{x}_i) = z_{ij} - \sum_{k=1}^p \overline{x_{ip}} a_{jp}$$

と計算できます。E ステップでは、 $\mathbf{a}_j$ は $\mathbf{A}$ という行列の中で与えられ、 $\mathbf{x}_i = (x_{i1} \ \cdots \ x_{ip})$ を期待値として求めます。 $\mathbf{X}$ を次のように縦ベクトルを横に並べた行列だとすると、

$$\mathbf{X} = (\mathbf{x}_1 \ \cdots \ \mathbf{x}_p)$$

$$\mathbf{x}_k = \begin{pmatrix} x_{1k} \\ \vdots \\ x_{pk} \end{pmatrix}$$

因子分析では、 $\mathbf{x}_k$ と $\mathbf{x}_\kappa$  ( $k \neq \kappa$ )は必ずしも直交していません。 $f_{ij}(\mathbf{a}_j, \mathbf{x}_i) = z_{ij} - \mathbf{x}_i \mathbf{a}_j^T$ と書けるのは、 $\mathbf{x}_k$ がすべて直交している時です。各 $\mathbf{x}_k$ のベクトルが作る角度は与えられませんから、実際に計算できるのは、

$$f_{ij}(\mathbf{a}_j, \mathbf{x}_i) = z_{ij} - \sum_{k=1}^p x_{ip} a_{jp}$$

です。そこで、 $\overline{\mathbf{x}_i} \mathbf{a}_j^T$ と $\sum_{k=1}^p x_{ip} a_{jp}$ の関係を考えます。結論を先に示すと、

$$\overline{\mathbf{x}_i} \mathbf{a}_j^T \geq \sum_{k=1}^p x_{ip} a_{jp}$$

で等号が成立つ条件は、すべての $\mathbf{x}_k$ が互いに直交していること、すなわち、 $\mathbf{x}_k^T \mathbf{x}_\kappa = 0$ です。この関係は、普通、イエンゼンの不等式によって説明されます。イエンゼンの不等式と、それを使った証明の方が一般的ですが、それは次に行うことにして、この場合について、まず、感覚的に分かりやすい説明をします。点 $\mathbf{z}_{ij}$ を点 $\overline{\mathbf{x}_i}$ を

$$\mathbf{z}_{ij} = (z_{ij} \ \overline{x_{i1}} a_{j1} \ \cdots \ \overline{x_{ip}} a_{jp})$$

$$\overline{\mathbf{x}_{ij}} = (L_0 \ \overline{x_{i1}} a_{j1} \ \cdots \ \overline{x_{ip}} a_{jp})$$

のように、 $p+1$ 超空間上の点だとすると、 $f_{ij}(\mathbf{a}_j, \overline{\mathbf{x}_i}) = z_{ij} - \overline{\mathbf{x}_i} \mathbf{a}_j^T$ は図 89 に示したベクトルの長さです。

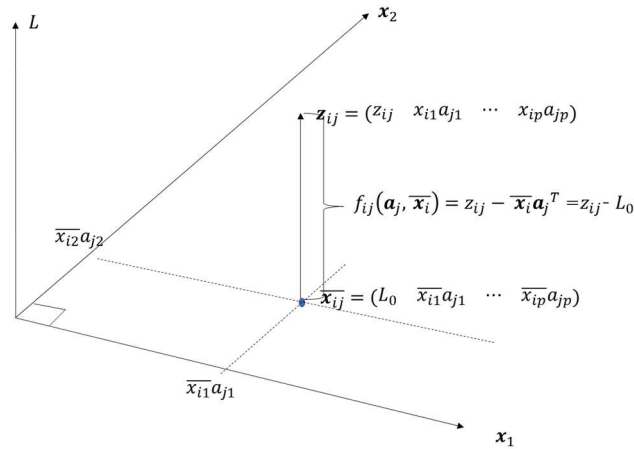


図 89.  $p+1$  次超空間上の点 $\mathbf{z}_{ij}$ と $p$ 次超平面上の点 $\overline{\mathbf{x}_{ij}}$ の関係



$\mathbf{x}_k$ がすべて直交していれば、 $\overline{\mathbf{x}}_{ij} = (L_0 \ \overline{x_{i1}}a_{j1} \ \cdots \ \overline{x_{ip}}a_{jp})$ は、 $(\overline{x_{i1}} \ \cdots \ \overline{x_{ip}})$ の線形結合として、 $\mathbf{x}_{1,\dots,\mathbf{x}_p}$ の $p$ 次超平面上の点として位置づけることができます。そのような条件で考えると、ベクトル $\overrightarrow{\overline{\mathbf{x}}_{ij}\mathbf{z}_{ij}}$ は、 $p$ 次超平面 $\mathbf{x}_{1,\dots,\mathbf{x}_p}$ から $\mathbf{z}_{ij} = (z_{ij} \ \overline{x_{i1}}a_{j1} \ \cdots \ \overline{x_{ip}}a_{jp})$ に向けた法線ベクトルで、 $\overline{\mathbf{x}}_{ij} = (L_0 \ \overline{x_{i1}}a_{j1} \ \cdots \ \overline{x_{ip}}a_{jp})$ はその脚です。平面と点の距離は法線ベクトルの長さで、その長さが、平面上の点と平面外の点の距離の最小値ですから、 $\overrightarrow{\overline{\mathbf{x}}_{ij}\mathbf{z}_{ij}}$ の長さ $|\overrightarrow{\overline{\mathbf{x}}_{ij}\mathbf{z}_{ij}}|$ は、点 $\mathbf{z}_{ij}$ と $p$ 次超平面 $\mathbf{x}_{1,\dots,\mathbf{x}_p}$ の点の距離の中で最小です。因子分析で求めている、 $(x_{i1} \ \cdots \ x_{ip})$ は $\mathbf{x}_{1,\dots,\mathbf{x}_p}$ の直交性を考慮して最適化しているのではないので、一般的に、 $(x_{i1} \ \cdots \ x_{ip}) = (\overline{x_{i1}} \ \cdots \ \overline{x_{ip}})$ ではありません。 $(x_{i1} \ \cdots \ x_{ip}) \neq (\overline{x_{i1}} \ \cdots \ \overline{x_{ip}})$ ならば

$$\mathbf{x}_{ij} = (L_0 \ x_{i1}a_{j1} \ \cdots \ x_{ip}a_{jp}) \neq (L_0 \ \overline{x_{i1}}a_{j1} \ \cdots \ \overline{x_{ip}}a_{jp}) = \overline{\mathbf{x}}_{ij}$$

です。つまり、 $\mathbf{x}_{ij} = (L_0 \ x_{i1}a_{j1} \ \cdots \ x_{ip}a_{jp})$ は $\mathbf{z}_{ij}$ からの法線の脚ではないから、

$$\begin{aligned} |\overrightarrow{\overline{\mathbf{x}}_{ij}\mathbf{z}_{ij}}| &\leq |\overrightarrow{\mathbf{x}_{ij}\mathbf{z}_{ij}}| \\ z_{ij} - \overline{\mathbf{x}}_i \mathbf{a}_j^T &\leq z_{ij} - \mathbf{x}_i \mathbf{a}_j^T \\ \overline{\mathbf{x}}_i \mathbf{a}_j^T &\geq \mathbf{x}_i \mathbf{a}_j^T \end{aligned}$$

となります。したがって、

$$\overline{\mathbf{x}}_i \mathbf{a}_j^T \geq \sum_{k=1}^p x_{ik} a_{jk}$$

です。難しいいいかたをすると、 $\sum_{k=1}^p x_{ik} a_{jk}$ は $\overline{\mathbf{x}}_i \mathbf{a}_j^T$ の変分下限（それ以上小さくならない限界）なのです。また、

$$z_{ij} - \overline{\mathbf{x}}_i \mathbf{a}_j^T \leq z_{ij} - \sum_{k=1}^p x_{ik} a_{jk}$$

なので、 $z_{ij} - \sum_{k=1}^p x_{ik} a_{jk}$ は $z_{ij} - \overline{\mathbf{x}}_i \mathbf{a}_j^T$ の変分上限で、 $z_{ij} - \overline{\mathbf{x}}_i \mathbf{a}_j^T$ は $z_{ij} - \sum_{k=1}^p x_{ik} a_{jk}$ よりも大きくなることはありません。最小二乗法では、 $z_{ij} - \overline{\mathbf{x}}_i \mathbf{a}_j^T$ の絶対値の最小値を求めています。これを直接求めることが出来なくても、変分上限である $z_{ij} - \sum_{k=1}^p x_{ik} a_{jk}$ の絶対値の最小値を求めれば、期待値として、 $(x_{i1} \ \cdots \ x_{ip})$ を求めることが出来るというのが、ここで使われる論理です。そこで、次の微分方程式を解いて、各 $x_i$ を求めます。

$$\frac{d \sum_{j=1}^m \left( f_{1j}(\mathbf{a}_j, x_1) \right)^2}{d x_1} = 0$$

$$\frac{d \sum_{j=1}^m \left( f_{1j}(\mathbf{a}_j, x_2) \right)^2}{d x_2} = 0$$

⋮

$$\frac{d \sum_{j=1}^m \left( f_{1j}(\mathbf{a}_j, x_n) \right)^2}{d x_n} = 0$$

$\mathbf{x}_i$ ベクトルによる微分を、偏微分の形で書くと、

$$\frac{\partial \sum_{j=1}^m \left( f_{1j}(\mathbf{a}_j, x_i) \right)^2}{\partial x_{i1}} = 0$$

$$\begin{aligned}\frac{\partial \sum_{j=1}^m (f_{1j}(\mathbf{a}_j, \mathbf{x}_i))^2}{\partial x_{i2}} &= 0 \\ &\vdots \\ \frac{\partial \sum_{j=1}^m (f_{1j}(\mathbf{a}_j, \mathbf{x}_i))^2}{\partial x_{ip}} &= 0\end{aligned}$$

これが、最小二乗法による E ステップです。

### イエンゼンの不等式

凸関数という関数の場合、元の数の期待値の関数と関数を先に求めて期待値を求めたものを比べると、元の数の平均値の関数の方が大きいという不等式をイエンゼンの不等式と言います。イエンゼンの不等式が成立つのは、関数が凸関数の時に限られます。平均値は期待値の一つです。単純平均、重み付きの平均なども含めて、期待値とは、ある根拠をもって、個々のデータに重みをつけて合計した総和で、すべてのデータを代表する値で、そのデータ集団と同じ母集団からランダムにサンプリングしたときに尤も得られる確率が高い値です。そういう意味で期待値と言います。平均値の場合は、すべての重みが等しいので、それぞれのデータの重みは $\frac{1}{n}$ です。

凸関数とは、変曲点を持たない関数。二階微分が正または負だけの関数です。

例えば $\log_e x$ は

$$\begin{aligned}\frac{d \log_e x}{dx} &= \frac{1}{x} \\ \frac{d^2 \log_e x}{dx^2} &= -\frac{1}{x^2} < 0\end{aligned}$$

ですから、 $\log_e x$ は上に凸の凸関数です。関数 $f(x) = \log_e x$ を使って、エンゼン不等式の

例を具体的に示してみます。

$x$ の期待値を $\text{Ex}(x)$ と表し、期待値を単純平均とすると、 $\text{Ex}(x) = \frac{\sum_{i=1}^n x_i}{n}$

$$\begin{aligned}f(\text{Ex}(x)) &= \log_e \frac{\sum_{i=1}^n x_i}{n} \\ \text{Ex}(f(x)) &= \frac{1}{n} \sum_{i=1}^n \log_e x_i\end{aligned}$$

たとえば、 $x_i = 1, 2, 4, 7$  とすると、

$$\begin{aligned}\text{Ex}(x) &= \frac{\sum_{i=1}^n x_i}{n} = \frac{1 + 2 + 4 + 7}{4} = 3.5 \\ f(\text{Ex}(x)) &= \log_e 3.5 = 1.252763\end{aligned}$$

一方

$$f(1) = \log_e 1 = 0, f(2) = \log_e 2 = 0.693147, f(4) = \log_e 4 = 1.386294, f(7) = \log_e 7 = 1.94591$$

$$Ex(f(x)) = \frac{0 + 0.693147 + 1.386294 + 1.94591}{4} = 1.006338$$

確かに

$$f(Ex(x)) > Ex(f(x))$$

になっています。

多次元で凸関数というのは、1つだけのピークを持つ、ドーム型の関数で、変曲点を持たない関数です。2次元で凸関数の場合、イエンゼンの不等式が成立つことを示します。

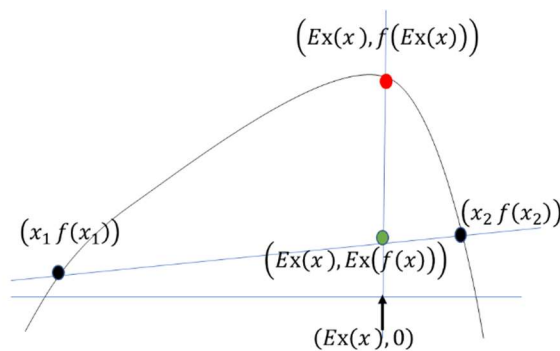


図. 90-1. 関数の期待値と期待値の関数の関係、2 データの場合

図 90 で、曲線は関数を表し、水平線が説明変数  $x$  を表しています。今、2つのデータ  $(x_1, f(x_1))$  と  $(x_2, f(x_2))$  があるとき、この2つのデータから、2つのデータが得られた母集団の期待値を求めるを考えます。もし、何らかの理由で、点  $(x_2, f(x_2))$  の方が1, 5倍母集団の特性を強く表しているとすれば、期待値は、

$$\left( \frac{1x_1 + 1.5x_2}{1 + 1.5}, \frac{1f(x_1) + 1.5f(x_2)}{1 + 1.5} \right) = (0.4x_1 + 0.6x_2, 0.4f(x_1) + 0.6f(x_2))$$

です。もし、どちらも同じように母集団の特性を表しているとすれば、単純集計で

$$\left( \frac{1x_1 + 1x_2}{1 + 1}, \frac{1f(x_1) + 1f(x_2)}{1 + 1} \right) = (0.5x_1 + 0.5x_2, 0.5f(x_1) + 0.5f(x_2))$$

となります。これがいわゆる平均です。

図形的には、 $x$  と  $f(x)$  の期待値が表す点は、重みを  $\lambda_1, \lambda_2$  とすると、線分  $(x_1, f(x_1))$  -

$(x_2, f(x_2))$  を  $\lambda_2: \lambda_1$  に内分する点です。図では緑の点がそれを表しています。図では緑の点がそれにあたります。その点は、 $(Ex(x), Ex(f(x)))$  と表せますが、 $Ex(x)$  の関数  $f(Ex(x))$  ですから、点  $(Ex(x), f(Ex(x)))$  が点  $(Ex(x), Ex(f(x)))$  より高いところにあることは明らかです。

次に点が3つある場合について考えます。

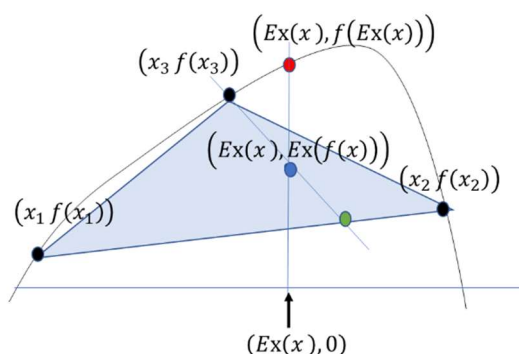


Fig. 90-2. 関数の期待値と期待値の関数の関係、3 データの場合

点がもう一つ増えると、先ほどの緑の点と、新しい点の相対的な代表性を考えて、重みをつけなおしますが、いずれにしても新しい点と緑の点を作る線分上に、期待値の点が出来ます。ですから、 $(Ex(x), f(Ex(x)))$  が点  $(Ex(x), Ex(f(x)))$  より高いところにあることは、やはり明らかです。

点をさらに増やしても、同じことです。

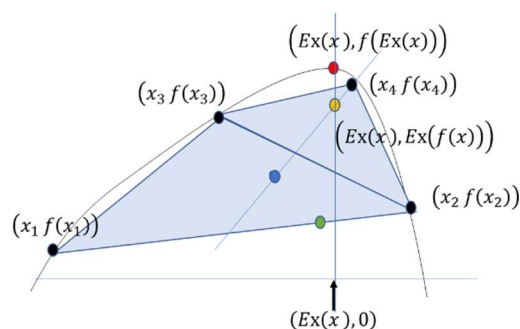


Fig. 90-3. 関数の期待値と期待値の関数の関係、4 データの場合

点を増やしていくと、頂点の数が増えていきますが、その多角形は、凸関数が作る曲線に内接しています。この多角形には特徴があって、凸包という形をしています。凸包というのは、へこんだ部分のない多角形のことです。つまり、多角形をゴムバンドで縛った時にできるゴムバンドの形です。高次元になると、多面体をゴムのラバーで包み込んだ形です。凸関数では、この多面体は必ず凸包になります。外接するドームの天井が凸包の天井より高い位置にあるのは当然のことです。イエンゼンの不等式を感覚的に表現すると、ドームの中に構造物を作れば、どの構造物の体積は、ドームの体積より小さいのは当然だと

言うのに近いのですが、これを、柱の長さにすると、少し事情が違います。もし、ドームがいくつかのピークを持つ多峰形であれば、複数のピークに集中して柱を立てれば、その平均が、全体の分布中心に立てた柱の長さよりも長いということは起こります。つまり、最尤法は、 $n$ 次元のデータ分布が単峰形であることを前提にしているのです。しかし、実際のデータには、複数の小集団が含まれることはよくあることです。そういう場合には不適解が出ることがあります。以上が、イエンゼンの不等式の一般的な説明です。

$$f(Ex(\mathbf{x})) \geq Ex(f(\mathbf{x}))$$

式 78

最小二乗法の E ステップの場合、

関数 $\mathbf{x}\mathbf{a}_j^T$ の具体的な計算の内容は、 $\sum_{k=1}^p \bar{x}_{ip} a_{jp}$ です。これは $\bar{x}_{ip}$ の重み付き平均です。また、独立に $(x_{i1} \ 0 \ \dots \ 0), (0 \ x_{i2} \ \dots \ 0), \dots, (0 \ \dots \ 0 \ x_{ip})$ を最適化したものの重み付き平均が、 $\sum_{k=1}^p x_{ip} a_{jp}$ です。つまり

$$\begin{aligned} f(Ex(\mathbf{x})) &= \mathbf{x}\mathbf{a}_j^T \\ Ex(f(\mathbf{x})) &= \sum_{k=1}^p x_{ip} a_{jp} \end{aligned}$$

となっていますから、イエンゼンの不等式を使って

$$\bar{\mathbf{x}}_i \mathbf{a}_j^T \geq \sum_{k=1}^p x_{ip} a_{jp}$$

という関係になることがわかります。

**M** ステップ

**E** ステップで $\mathbf{X}$ の期待値が得られたので、

$$\mathbf{E} = \mathbf{Z} - \mathbf{X}\mathbf{A}^T$$

$$\mathbf{Z} \cong \mathbf{X}\mathbf{A}^T$$

として、疑似逆行列、 $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}$ を両辺にかけて、 $\mathbf{A}^T$ を求めます。これで、つぎの E ステップで使う $\mathbf{A}$ が得られます。 $\mathbf{Z}$ が平均値からの差のデータであれば、 $\mathbf{Z}^T \mathbf{Z}$ は分散・共分散行列になります。また、 $\mathbf{A} \mathbf{X}^T \mathbf{X} \mathbf{A}^T$ は推定値の分散・共分散行列です。

$$\mathbf{Z}^T \mathbf{Z} = \begin{pmatrix} z_{11} & \dots & z_{n1} \\ \vdots & \ddots & \vdots \\ z_{1m} & \dots & z_{nm} \end{pmatrix} \begin{pmatrix} z_{11} & \dots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \dots & z_{nm} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n z_{i1}^2 & \sum_{i=1}^n z_{i1} z_{i2} & \dots & \sum_{i=1}^n z_{i1} z_{im} \\ \sum_{i=1}^n z_{i2} z_{i1} & \sum_{i=1}^n z_{i2}^2 & \dots & \sum_{i=1}^n z_{i2} z_{im} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n z_{in} z_{i1} & \sum_{i=1}^n z_{in} z_{i2} & \dots & \sum_{i=1}^n z_{in}^2 \end{pmatrix}$$

$$\begin{aligned}
\mathbf{XA}^T &= \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1p} & \cdots & a_{mp} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^p x_{1k} a_{1k} & \cdots & \sum_{k=1}^p x_{1k} a_{mk} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^p x_{nk} a_{1k} & \cdots & \sum_{k=1}^p x_{nk} a_{mk} \end{pmatrix} \\
\mathbf{AX}^T \mathbf{XA}^T &= \begin{pmatrix} \sum_{k=1}^p x_{1k} a_{1k} & \cdots & \sum_{k=1}^p x_{nk} a_{1k} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^p x_{1k} a_{mk} & \cdots & \sum_{k=1}^p x_{nk} a_{mk} \end{pmatrix} \begin{pmatrix} \sum_{k=1}^p x_{1k} a_{1k} & \cdots & \sum_{k=1}^p x_{1k} a_{mk} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^p x_{nk} a_{1k} & \cdots & \sum_{k=1}^p x_{nk} a_{mk} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{1k} \right)^2 \right) & \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{1k} \right) \left( \sum_{k=1}^p x_{ik} a_{2k} \right) \right) & \cdots & \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{1k} \right) \left( \sum_{k=1}^p x_{ik} a_{mk} \right) \right) \\ \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{2k} \right) \left( \sum_{k=1}^p x_{ik} a_{1k} \right) \right) & \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{2k} \right)^2 \right) & \cdots & \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{2k} \right) \left( \sum_{k=1}^p x_{ik} a_{mk} \right) \right) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{mk} \right) \left( \sum_{k=1}^p x_{ik} a_{1k} \right) \right) & \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{mk} \right) \left( \sum_{k=1}^p x_{ik} a_{2k} \right) \right) & \cdots & \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{mk} \right)^2 \right) \end{pmatrix}
\end{aligned}$$

これらの行列の対角成分は分散ですが、全分散は、誤差分散と $\mathbf{XA}^T$ で説明できる分散の和です。

$$\begin{aligned}
\sum_{i=1}^n z_{ij}^2 &= \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{jk} \right)^2 \right) + \sum_{i=1}^n e_{ij}^2 \\
\sum_{i=1}^n e_{ij}^2 &= \sum_{i=1}^n z_{ij}^2 - \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{jk} \right)^2 \right) \geq 0
\end{aligned}$$

$\sum_{i=1}^n z_{ij}^2 < \sum_{i=1}^n \left( \left( \sum_{k=1}^p x_{ik} a_{jk} \right)^2 \right)$ ならば、 $\mathbf{X}$ 、 $\mathbf{A}$ は不適解です。不適解でなければ、Eステップ

に戻り、この $\mathbf{A}$ を与えて、 $\mathbf{X}$ の期待値を求め さらにMステップに戻って、 $\mathbf{A}$ の最適値を求めます。誤差分散の正負をたしかめて、不適解でなければ、さらにEステップ、Mステップを繰り返して、誤差分散が一定の範囲で変化しなくなれば、その時の $\mathbf{A}$ 、 $\mathbf{X}$ を解とします。

### 最尤法による接近

すでに学んだように、多次元の正規分布確率は

$$P(\mathbf{Z}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{Z})^T \Sigma^{-1}(\mathbf{Z})}$$

と表せます。ここでは $\mathbf{Z}$ が、 $\mathbf{XA}^T$ によって与えられる、期待値からの差として標準化されて

いることを前提にしています。これを個々のデータが得られる確率として表すと

$$P(z_{ij}) = \frac{1}{(\sqrt{2\pi})\sigma_j} e^{-\frac{1}{2}\left(\frac{(z_{ij}-\mathbf{x}_i\mathbf{a}_j^T)}{\sigma_j}\right)^2}$$

$$\mathbf{x}_i = (x_{i1} \quad \cdots \quad x_{ip})$$

$$\mathbf{a}_j = (a_{j1} \quad \cdots \quad a_{jp})$$

$$\sigma_j^2 \text{は } \mathbf{z}_j = (z_{1j} \quad \cdots \quad z_{nj}) \text{の分散}$$

全ての $z_{ij}$ が独立しているとすれば、 $\mathbf{Z}$ が得られる確率 $P(\mathbf{Z})$ は $P(z_{ij})$ の総乗ですから、

$$P(\mathbf{Z}) = \prod_{i=1}^n \prod_{j=1}^m P(z_{ij}) = \prod_{i=1}^n \prod_{j=1}^m \frac{1}{(\sqrt{2\pi})\sigma_j} e^{-\frac{1}{2}\left(\frac{(z_{ij}-\mathbf{x}_i\mathbf{a}_j^T)}{\sigma_j}\right)^2}$$

となりますが、 $\mathbf{a}_j$ を与えて $\mathbf{x}_i$ の最適値を求めるには、 $\mathbf{x}_i$ で微分しますが、直接微分できないので、対数確率にして、 $\mathbf{x}_i$ の関数に分離します。

$$\begin{aligned} \log_e P(\mathbf{Z}) &= \sum_{i=1}^n \sum_{j=1}^m \log_e \left( \frac{1}{(\sqrt{2\pi})\sigma_j} e^{-\frac{1}{2}\left(\frac{(z_{ij}-\mathbf{x}_i\mathbf{a}_j^T)}{\sigma_j}\right)^2} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^m \left( -\log_e(\sqrt{2\pi})\sigma_j - \frac{1}{2}\left(\frac{(z_{ij}-\mathbf{x}_i\mathbf{a}_j^T)}{\sigma_j}\right)^2 \right) \end{aligned}$$

$\sum_{j=1}^m \left( -\log_e(\sqrt{2\pi})\sigma_j - \frac{1}{2}\left(\frac{(z_{ij}-\mathbf{x}_i\mathbf{a}_j^T)}{\sigma_j}\right)^2 \right)$ の中には $\mathbf{x}_i$ 以外の $\mathbf{x}_i$ を含まず、 $-\log_e(\sqrt{2\pi})\sigma_j$ は観測値 $\mathbf{Z}$ から与えられる定数ですから、

$$\frac{d \left( \sum_{j=1}^m \left( \frac{(z_{ij}-\mathbf{x}_i\mathbf{a}_j^T)}{\sigma_j} \right)^2 \right)}{d\mathbf{x}_i} = 0$$

として、連立方程式を解けば $\mathbf{x}_i$ が得られます。 $\mathbf{x}_i$ が超平面上の1点として特定できればこの計算は可能ですが、因子分析が考えているモデルでは、 $\mathbf{x}_i$ は1点として固定できません。これは、最小二乗法のEステップで説明したことと同じです。ですから、変分下限を

与えて、その確率の最大化を最尤法で計算します。最尤法では、 $\left(\frac{z_{ij}-\mathbf{x}_i\mathbf{a}_j^T}{\sigma_j}\right)$ という計算で、

標準偏差で推測値と観測値の差を標準化しています。これが最小二乗法との違いです。したがって、最尤法では、標準化したデータから計算した結果（相関行列から計算した結果）と分散共分散行列から計算した結果が同じになります。最小二乗法では、この二つの計算結果が違います。また、標準化したデータで最小二乗法によって因子抽出した結果は、最尤法で抽出した結果と同じになります。

M ステップでは、 $\Sigma$ を観測されたデータから計算して与えるので、下の微分式から連立方程式を作って $A$ を求めますが、

$$\frac{\partial LL}{\partial A} = \frac{\sum_{j=1}^m \sum_{i=1}^n (z_{ij}^2 \Sigma^{-1})}{\partial A} = 0$$

これは、重回帰の繰り返しですから、最小二乗法でも最尤法でも結果に違いはなく、簡単に計算しようと思えば、難しく考えずに、普通の重回帰のように、期待値 $X$ から疑似逆行列、 $(X^T X)^{-1} X^T$ を計算して

$$\begin{aligned} Z &\cong X A^T \\ (X^T X)^{-1} X^T Z &\cong (X^T X)^{-1} X^T X A^T \\ A^T &= (X^T X)^{-1} X^T Z \end{aligned}$$

と計算して、簡単に最適な $A$ を求めることができます。

関数が凸関数だということを前提に、変分下限を最適化しているのは、最小二乗法でも、最尤法でも同じです。確率を最大化するという意味では、最尤法は素直な考え方で論理的です。しかし、もし、実際のデータが正確に正規分布しているとすると、正規分布は変曲点が2つあるので、凸包ではありません。実際のデータの分布が凸包として扱えるかどうかは、また別の問題ですが、最尤法の方が不適解が出ることが多いはずで、実際のデータには部分的に粗密があるから、データ分布には凸凹があります。そのような凸凹があったとしても、極端でなく、それなりに滑らかで、分布がシンメトリックに近ければ最小二乗法では不定解が出にくいのです。何かの都合で、標準化しない形で因子分析をする必要があるのならば、最小二乗法を使わざるを得ないでしょう。

どちらを選ぶかをあらかじめ判断する場合、データが偏っていたら最尤法で、データが多峰形ならば、最小二乗法を試してみるというのが一般的です。とりあえず、最尤法の方が新しいし流行っているから最尤法でやり、ダメなら最小二乗法という現実的・適応的なやり方もありえます。最尤法で計算して良い結果が出た時にそれにクレームをつけるレビュアーはいません。最尤法でうまくいかない時には、データが多峰形になっていないかを疑います。多峰形ならばあきらかに凸包ではありません。データ分布が多峰形になる場合、その部分を取り除いて良いかどうかは、微妙な問題です。恣意的にデータを取り除くことが良いかどうかは、分析の目的によります。最小二乗法でも不適解が出る場合は、データの質を疑った方が良いでしょう。そもそも何らかの因子によって説明できる現象かどうかを考えます。そのために、主成分分析行って、何らかの成分のようなものが存在するか否かを確かめた方が良いでしょう。筆者は、まず主成分分析を行ってから、因子分析を行います。主成分分析と因子分析は目的が違いますし、用いている数学的なモデルが異なります。全く、違った分析ですが、主成分分析は数学的には因子分析に比べてはるかに簡単な分析で、確実に答えが出ます。因子数の決定に主成分分析の結果を利用することもあり、データの分布の概要を知るには有力な方法です。一応、主成分分析をやってお



く方が良いでしょう。

### MCMC によるベイズ的推定法

最近のコンピューターソフトの発達は著しく、その計算力に頼ってベイズ的に解に接近するという手法が様々な分野で用いられています。ある条件の中で、乱数を発生させ、これを所与のデータとして、確率分布の母数を求め、この母数を制約条件として、前の乱数とは独立に（マルコフ連鎖）制約条件にしたがう乱数を発生させてこれを所与のデータとして、次の確率分布の母数の推定に使います、これを繰り返すことによって、ある値の範囲にデータが安定したらそれを解とします。しかし、それには膨大な計算が必要です。効率よく計算するために様々なアルゴリズムが考えられていますし、そのような計算をしてくれるソフトも販売されています。しかし、現時点ではやはり開発途中であり、その全容を解説するのは筆者の手に余りますので、ソフトウェアに添付された説明を参照してください。なお、R には、MCMCpack の中に、MCMCfactnal という MCMC を使った因子分析の関数があります。現在、MCMC を含めて、様々な因子抽出法の比較がなされてその特性が議論されています。その詳細については、筆者も勉強中です。ある程度の情報がまとまれば、追加で提供します。

### VI-2-4-4. 回転

上記のような方法で因子が抽出されますが、このままだと、一つの被説明変数（測定変数）がいくつかの因子によって説明されるために、因子の解釈が難しいという問題があります。できれば、測定変数が限られた数の因子によって説明される方が、因子の内容の解釈が楽です。これを単純化と言います。回転することによって、測定変数の分布の方向が、因子の軸に一致すると、もっぱら一つの軸（因子）によって説明できることになります。それを実現するにはそれ以外の因子の因子負荷量が 0 に近くなるように、座標(因子)を回転させます。回転の計算には数学的な技術が必要ですが、回転の目的は人間の解釈のためで、数学的な必要性からではありません。解釈しやすいように回転すれば良いのです。

座標軸の回転は、一旦極座標を経れば、難しくありません。極座標変換がわからない場合は、Polar coordinate のところを読んでください。

2次元の場合、極座標を $(r, \theta)$ と表すと、その2次元の直交座標は

$$x = r \cos \theta$$

$$y = r \sin \theta$$

です。座標を反時計方向に $\theta$ 回転させると、 $\theta = 0$ になるので、

$$x = r \cos 0 = r$$

$$y = r \sin 0 = 0$$

です。

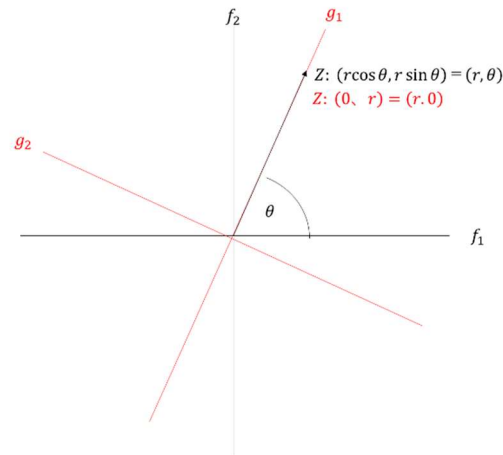


図 91. 極座標での回転

このような結果になる変換（回転行列 T）は

$$(r \ 0) = (r \cos \theta \ r \sin \theta) \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}$$

より、

$$0 = r \cos \theta t_{11} + r \sin \theta t_{21} = r(\cos \theta t_{11} + \sin \theta t_{21})$$

$$r = r \cos \theta t_{12} + r \sin \theta t_{22} = r(\cos \theta t_{12} + \sin \theta t_{22})$$

$$\cos \theta t_{11} + \sin \theta t_{21} = 1$$

$$\cos \theta t_{12} + \sin \theta t_{22} = 0$$

なので、

$$T = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

3次元の場合はいきなり回転行列を考えるよりも、Z軸を中心に回転させて、次にX軸を中心に回転させるという2工程を考えた方が易しそうです。

Z軸周りの回転は

$$\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

X軸周りの回転は

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix}$$

だから、これを連続して行くと、

$$\begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \cos \varphi & \sin \theta \sin \varphi \\ \sin \theta & \cos \theta \cos \varphi & -\cos \theta \sin \varphi \\ 0 & \sin \varphi & \cos \varphi \end{pmatrix}$$

となります。

さらに高次元の場合の回転行列も示すことはできますが、式が複雑で、その誘導まで解説すると大変なのでここでは省略します。高次元空間の回転は感覚的に理解しにくいから、線形代数的に回転行列の特性を考えてみます。

今、求めた2次と3次の回転行列には、ある特徴があります。この回転行列とその転置行列を掛けると、単位行列になるのです。つまり、転置行列が逆行列なのです。試してみます。

$$T = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

$$T^T = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

$$\begin{aligned} TT^T &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} \cos^2 \theta + \sin^2 \theta & \sin \theta \cos \theta - \sin \theta \cos \theta \\ \sin \theta \cos \theta - \sin \theta \cos \theta & \sin^2 \theta + \cos^2 \theta \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I \end{aligned}$$

3次元の場合

$$T = \begin{pmatrix} \cos \theta & -\sin \theta \cos \varphi & \sin \theta \sin \varphi \\ \sin \theta & \cos \theta \cos \varphi & -\cos \theta \sin \varphi \\ 0 & \sin \psi & \cos \psi \end{pmatrix}$$

$$T^T = \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta \cos \varphi & \cos \theta \cos \varphi & \sin \psi \\ \sin \theta \sin \varphi & -\cos \theta \sin \varphi & \cos \psi \end{pmatrix}$$

$$\begin{aligned} TT^T &= \begin{pmatrix} \cos \theta & -\sin \theta \cos \varphi & \sin \theta \sin \varphi \\ \sin \theta & \cos \theta \cos \varphi & -\cos \theta \sin \varphi \\ 0 & \sin \psi & \cos \psi \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta \cos \varphi & \cos \theta \cos \varphi & \sin \psi \\ \sin \theta \sin \varphi & -\cos \theta \sin \varphi & \cos \psi \end{pmatrix} = \\ &\begin{pmatrix} \cos^2 \theta + \sin^2 \theta \cos^2 \varphi + \sin^2 \theta \sin^2 \varphi & \sin \theta \cos \theta + \sin \theta \cos \theta \cos^2 \varphi + \sin \theta \cos \theta \sin^2 \varphi & -\sin \theta \sin \psi \cos \psi + \sin \theta \sin \psi \cos \psi \\ \cos \theta \sin \psi - \cos \theta \sin \psi \cos^2 \varphi - \cos \theta \sin \psi \sin^2 \varphi + & \sin^2 \theta + \cos^2 \theta \cos^2 \varphi + \sin^2 \theta \sin^2 \varphi + & \cos \theta \cos \psi \sin \psi - \cos \theta \sin \psi \cos \psi \\ -\sin \psi \sin \theta \cos \psi + \cos \psi \sin \theta \sin \psi & \sin \psi \cos \theta \cos \psi - \cos \psi \cos \theta \sin \psi & \sin^2 \psi + \cos^2 \psi \end{pmatrix} \\ &= \begin{pmatrix} \cos^2 \theta + \sin^2 \theta (\cos^2 \varphi + \sin^2 \varphi) & 0 & 0 \\ 0 & \sin^2 \theta + \cos^2 \theta (\cos^2 \varphi + \sin^2 \varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos^2 \theta + \sin^2 \theta & 0 & 0 \\ 0 & \sin^2 \theta + \cos^2 \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I \end{aligned}$$

考えてみれば、回転してそれを元に戻したときに、元の空間の相対的な位置関係が変わらないのだから、このような性質を持つのは当然と言えば当然です。直交性を保った回転行列では、転置行列が逆行列になるというのは、どの次元でも共通した性質です。転置行列

が逆行列になっている正方行列を掛ければ、直交性を保ったまま回できるということです。

このような性質を持った回転行列を掛けると、直交性を保ったまま回転できます。やってみます。

A という行列に回転行列 T を掛けて、直交性を保ったまま、 $\Lambda$ （ラムダ）という行列が得られたとします。

$$AT = \Lambda$$

$$T^T A^T = \Lambda^T$$

因子分析のモデルにこれを当てはめると

$$Z = FA^T + UD$$

で  $Z$  と  $UD$  に変化を与えないのだから、

$$Z = FA^T + UD$$

$$= FIA^T + UD$$

$$= FTT^T A^T + UD$$

$$= FT\Lambda^T + UD$$

ここで、 $FT = G$  とすれば、

$$Z = G\Lambda^T + UD$$

となり、 $Z$  と  $UD$  を保ったまま、 $Z$  を新しい因子  $G$  で説明できます。

直交回転では、データ間の距離（ノルム）を保ったまま、解釈しやすいように、データを回転させることが出来ます。これが直交回転の意味です。しかし、因子分析の場合、そこまで厳密に直交性を保たなければならないのかという議論があります。そもそも、因子分析は、もともとデータの直交性を無視しています。対角成分だけを考えるというのはそういうことです。非対角成分つまり共分散を無視したということと、共分散が 0 だということは全く違うことです。共分散が 0 ならば、厳格に因子は直交しています。もちろん、それぞれの因子が独立して検出されるということは、それなりに独立性が高いとは言えるでしょうが、因子間の相関が全くないということではありません。そうだとすれば、厳密に直交性を保って回転することにあまり意味がないかもしれません。

そこで、直交性を保つことを考えずに回転することを考えます。それが斜交回転です。斜交回転でもデータの分散は保たなければならないから、回転行列 T を考えると、

$$\text{diag}TT^T = I$$

転置行列を掛けた時に対角成分がすべて 1 になる行列でなければなりません。

これを、 $Z = FA^T + UD$  の式に組み込みます。

$$Z = FA^T + UD$$

$$= FIA^T + UD$$

$$= FTT^{-1}A^T + U$$

$$= FT(T^T)^{-1}A^T + UD$$

$$= FT(A(T^T)^{-1})^T + UD$$

となるので、

$$G = FT$$

$$A(T^T)^{-1} = A$$

として、

$$Z = GA^T + UD$$

と表せます。

これらが、回転の計算の仕方です。これは回転の計算方法ですから、これだけでは何を指して回転するのかがわかりません。

### 回転基準

回転法のところで、直交回転も斜交回転も回転の計算方法はわかりました。しかし、何を指して回転するのかは論じていません。回転の目的は、何がしたいのか分析の目的によって決まります。回転の目的の一つは、結果の解釈が容易になることです。つまり、構造を単純化したいのです。

例を挙げます。手元にデータがないので、ある教科書で取り上げられている例をそのまま使います（豊田秀樹(2012): 因子分析入門—R で学ぶ最新データ解析。東京図書）。

心理尺度を使った。スキー場の評価の例です。専門ではないので、全くその内容はわかりませんが、こういうものの評価（情動的意味の定量的測定：多分、人々の感じ方のことだと思います。）には、活動性、力量性というのがあるのだそうです。こういう評価の質問票調査では、一つの質問に対して、とてもそう思う—かなりそう思う—ややそう思う—どちらでもない—ややそう思わない—かなりそう思わない—全くそう思わない。のような、7つの選択肢で回答するか7件法か、そう思う—ややそう思う—どちらでもない—ややそう思わない—そう思わない—そう思わない—5つの選択肢で回答する5件法が使われます。評価にかかわる質問項目には、好き—嫌い、洒落ている—野暮ったい、特色がある—特色がないのような質問があります。つまり、7件法で言えば、好き—嫌い について、「とても好き—かなり好き—やや好き—どちらでもない—やや嫌い—かなり嫌い—とても嫌い」の7つの選択肢で回答します。活動性については、「動的—静的、明るい—暗い、強い—弱い」。力量性については、「固い—柔らかい、安定した—不安定な、小さい—大きい」などの質問があります。つまり、3因子で9つの質問（観測変数）があるということです。

下の表は、因子抽出しただけの回転前の初期解です。

初期解：回転前の因子負荷行列

|    | 因子 1 | 因子 2  | 因子 3  | 共通性  |
|----|------|-------|-------|------|
| 好き | 0.87 | -0.36 | -0.01 | 0.88 |
| 洒落 | 0.94 | -0.31 | 0.01  | 0.99 |
| 特色 | 0.86 | -0.37 | -0.02 | 0.88 |
| 動的 | 0.62 | 0.71  | 0.28  | 0.97 |

|     |       |       |       |      |
|-----|-------|-------|-------|------|
| 明るい | 0.59  | 0.69  | 0.21  | 0.88 |
| 強い  | 0.53  | 0.70  | 0.26  | 0.84 |
| 固い  | -0.37 | -0.52 | 0.72  | 0.93 |
| 安定  | -0.22 | -0.46 | 0.72  | 0.78 |
| 小さい | 0.30  | 0.29  | -0.85 | 0.90 |

それでも、何となく様子はわかります。ぼんやりしていますが、因子3は、（固い、安定、小さい）という力量性にかかわる因子のようです。因子2は、（動的、明るい。強い）という活動性にかかわっているようですが、力量性にも関わっているように見えます。因子1は、評価にかかわっているようですが、力量性にも関係しているように見えます。つまり、曖昧性があって、解釈が難しいのです。そこで、直交回転して、因子と観測変数の関係を見えやすくします。

#### 回転（バリマックス回転）後の因子負荷行列

|     | 因子1   | 因子2   | 因子3   | 共通性  |
|-----|-------|-------|-------|------|
| 好き  | 0.93  | 0.12  | -0.05 | 0.88 |
| 洒落  | 0.97  | 0.19  | -0.08 | 0.99 |
| 特色  | 0.93  | 0.11  | -0.05 | 0.88 |
| 動的  | 0.17  | 0.96  | -0.13 | 0.97 |
| 明るい | 0.16  | 0.90  | -0.18 | 0.88 |
| 強い  | 0.10  | 0.90  | -0.12 | 0.84 |
| 固い  | -0.07 | -0.28 | 0.92  | 0.93 |
| 安定  | 0.03  | -0.17 | 0.87  | 0.78 |
| 小さい | 0.13  | 0.01  | -0.94 | 0.90 |

バリマックス回転後には、因子1は評価にかかわる因子であり、因子2は活動性に、因子3は力量性にかかわることが明瞭に認識できます。これが回転の目的です。この場合、因子の直交性を維持しながら回転しているので、このくらいが限界です。もし、直交性にこだわらず、新たな軸を作るとすれば、

|     | 因子1  | 因子2  | 因子3   | 共通性 |
|-----|------|------|-------|-----|
| 好き  | 0.93 | 0    | 0     |     |
| 洒落  | 0.97 | 0    | 0     |     |
| 特色  | 0.93 | 0    | 0     |     |
| 動的  | 0    | 0.96 | 0     |     |
| 明るい | 0    | 0.90 | 0     |     |
| 強い  | 0    | 0.90 | 0     |     |
| 固い  | 0    | 0    | 0.92  |     |
| 安定  | 0    | 0    | 0.87  |     |
| 小さい | 0    | 0    | -0.94 |     |

に近い形になって、より明瞭に因子と観測変数の関係を捉えることが出来るかもしれません。斜交回転によって、実際にそのようになるとは限りませんが、このような解がえられたら、それを完全クラスター解と言います。回転にはこのように単純構造を目指す回転という考え方があります。しかし、どんな場合でも単純構造を目指すべきかどうかは、一概に言えません。一つの因子が、複数の観測変数にかかわるということも実際にあり得るからです。良く知られているのは、国語・算数・理科・社会・英語にかかわる成績を説明する因子です。これは、いわゆる論理的思考と言語能力・記憶という2つの因子がかかわっていると考えられていますが、このうち算数と理科は論理的思考能力という一つの因子によって説明され、国語と社会は言語能力・記憶という一つの因子によって説明されます。ところが、英語（語学学習力）は、論理的思考能力と言語能力・記憶力の両方の因子がかかわっているとされます。つまり、どのくらい単純化すれば良いのかは、データの特性や既往の情報、分析の目的によって異なります。その単純化の程度によって、回転基準というのが提案されています。これが、因子分析に様々な回転法がある理由です。

ここから先の説明は多くの教科書で極めて分かりにくい説明になっています。回転基準という概念が出てくるのですが、そもそも何でそんな基準が出来たのか、説明がありません。筆者流に説明します。

$n$ 次元空間内で、回転させるのだから1軸ずつ、回転角 $\theta_1, \dots, \theta_{n-1}$ を与えて、因子と説明変数の平均ベクトルの角度を最適化するのが最も素朴な考え方でしょう。目的は厳密な証明ではないから、ここでは、最も単純な因子が二つの例を考えてみます。観測変数は4つにします。

| 初期解の因子負荷行列 |      |       |
|------------|------|-------|
|            | 因子 1 | 因子 2  |
| Z1         | 0.87 | -0.36 |
| Z2         | 0.94 | -0.31 |
| Z3         | 0.62 | 0.71  |
| Z4         | 0.59 | 0.69  |

図 92 は因子負荷行列をプロットしたもので、因子 1, 2 が 1 の時の、観測変数  $Z$  をベクトルとして表しました。

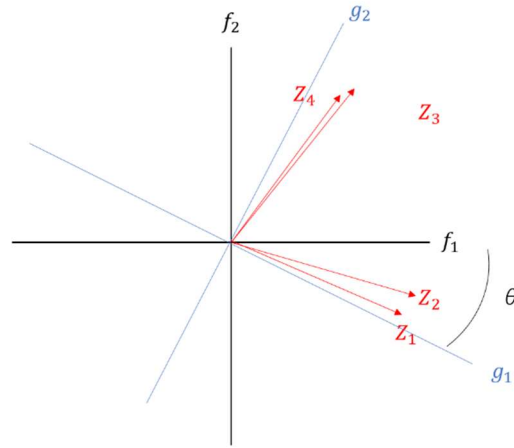


図. 92. 因子 1、因子 2 の値が 1 の時の観測変数のベクトル（期待値）

直交回転によって因子負荷行列を単純構造に近づけるとは、 $f_1 - f_2$  の軸の直交性を保ったまま、この平面を  $\theta$  回転させて、 $g_1 - g_2$  を軸とするということです。この場合、 $\theta$  はマイナスですが、 $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$  の範囲で回転させると考えれば良いでしょう。

この回転の目的は、 $g_1$  と  $Z_1, Z_2$  なす角度 ( $\psi_{11}, \varphi_{12}$ ) を 0 に近づけ、 $g_1$  と  $Z_3, Z_4$  なす角度 ( $\psi_{13}, \varphi_{14}$ ) を  $\frac{\pi}{2}$  に近づけ、 $g_2$  と  $Z_1, Z_2$  なす角度 ( $\psi_{21}, \varphi_{22}$ ) を  $-\frac{\pi}{2}$  に近づけ、 $g_2$  と  $Z_3, Z_4$  なす角度 ( $\psi_{23}, \varphi_{24}$ ) を 0 に近づけるということです。二次平面で原点を中心に回転する場合の回転行列  $T$  は

$$T = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

ですから、先ほど例に挙げた、因子負荷行列  $A$  の回転は、次の式で表せます。

$$A = \begin{pmatrix} 0.87 & -0.36 \\ 0.94 & -0.31 \\ 0.62 & 0.71 \\ 0.59 & 0.69 \end{pmatrix}$$

$$AT = \begin{pmatrix} 0.87 & -0.36 \\ 0.94 & -0.31 \\ 0.62 & 0.71 \\ 0.59 & 0.69 \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

試しに、見た目では  $\theta = -\frac{\pi}{6}$  だろうと、見当をつけて、

$$AT = \begin{pmatrix} 0.87 & -0.36 \\ 0.94 & -0.31 \\ 0.62 & 0.71 \\ 0.59 & 0.69 \end{pmatrix} \begin{pmatrix} \cos\left(-\frac{\pi}{6}\right) & -\sin\left(-\frac{\pi}{6}\right) \\ \sin\left(-\frac{\pi}{6}\right) & \cos\left(-\frac{\pi}{6}\right) \end{pmatrix} = \begin{pmatrix} 0.87 & -0.36 \\ 0.94 & -0.31 \\ 0.62 & 0.71 \\ 0.59 & 0.69 \end{pmatrix} \begin{pmatrix} \cos\left(\frac{\pi}{6}\right) & \sin\left(\frac{\pi}{6}\right) \\ -\sin\left(\frac{\pi}{6}\right) & \cos\left(\frac{\pi}{6}\right) \end{pmatrix}$$

$$AT = \begin{pmatrix} 0.87 & -0.36 \\ 0.94 & -0.31 \\ 0.62 & 0.71 \\ 0.59 & 0.69 \end{pmatrix} \begin{pmatrix} 0.866025 & 0.5 \\ -0.50 & 0.866025 \end{pmatrix} = \begin{pmatrix} 0.9247815 & 0.118231 \\ 0.9690635 & 0.20153225 \\ 0.1819355 & 0.92487775 \\ 0.16595475 & 0.89255725 \end{pmatrix}$$



$$\Lambda = \begin{pmatrix} 0.9247815 & 0.118231 \\ 0.9690635 & 0.20153225 \\ 0.1819355 & 0.92487775 \\ 0.16595475 & 0.89255725 \end{pmatrix}$$

とやってみると、適当に見当をつけて、座標軸を回転しただけで、すっきりと単純化して見やすくなります。適当に見当をつけて回転させればわかりやすくなる。それなりに有効だからそれでも良いとも言えそうです。しかし、その結果が最適解とは言えないでしょう。もう少し、理論的に考えたいくなります。ここで、因子負荷行列の列に着目します。回転前の、因子1の負荷量は0.87、0.94、0.62、0.59でしたが、回転後の因子1の負荷量は、0.92、0.97、0.18、0.17です。観測変数ごとの負荷量のバラツキが大きくなっています。因子2はあまり変わっていないような気がするかもしれませんが、実はそれなりに大きくなっています。因子負荷量は、単回帰的に考えると、因子に対する傾きで、因子1が1増加したときの観測変数の増加分で、因子の軸に対して観測変数がどのくらい離れているかを表しています。当然、マイナスの値もあり得るのですが、マイナスの距離というのはいり得ないから、その絶対値を考えるべきです。絶対値を考える場合、普通は2乗の値を考えます。つまり、 $0.92^2$ 、 $0.97^2$ 、 $0.18^2$ 、 $0.17^2$ の値の分散を考えることになります。

因子2 回転前

$$\frac{0.36^2 + 0.31^2 + 0.71^2 + 0.69^2}{4} = 0.301475$$

$$(0.36^2 - 0.301475)^2 + (0.31^2 - 0.301475)^2 + (0.71^2 - 0.301475)^2 + (0.69^2 - 0.301475)^2 = 0.143271$$

因子2 回転後

$$\frac{0.12^2 + 0.20^2 + 0.92^2 + 0.89^2}{4} = 0.423225$$

$$(0.12^2 - 0.423225)^2 + (0.20^2 - 0.423225)^2 + (0.92^2 - 0.423225)^2 + (0.89^2 - 0.423225)^2$$

と計算すると、大きく変わっています。

2 因子4 観測変数の場合の直交回転で何を最大化するのか確認してみます。

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \end{pmatrix}$$

求める因子負荷行列が上記の因子負荷行列だとすると、

$$\begin{aligned} \mu_1 &= \frac{1}{m} \sum_{j=1}^m \lambda_{j1}^2 \\ SS_1 &= \sum_{j=1}^m (\lambda_{j1}^2 - \mu_1)^2 \\ &= \sum_{j=1}^m \lambda_{j1}^4 - 2\mu_1 \sum_{j=1}^m \lambda_{j1}^2 + m\mu_1^2 \end{aligned}$$

$$\begin{aligned}
& \sum_{j=1}^m \lambda_{j1}^4 - 2 \frac{1}{m} \sum_{j=1}^m \lambda_{j1}^2 \sum_{j=1}^m \lambda_{j1}^2 + m \left( \frac{1}{m} \sum_{j=1}^m \lambda_{j1}^2 \right) \left( \frac{1}{m} \sum_{j=1}^m \lambda_{j1}^2 \right) \\
&= \sum_{j=1}^m \lambda_{j1}^4 - \frac{1}{m} \left( \sum_{j=1}^m \lambda_{j1}^2 \right) \left( \sum_{j=1}^m \lambda_{j1}^2 \right) \\
&= \sum_{j=1}^m \lambda_{j1}^4 - \frac{1}{m} \left( \sum_{j=1}^m \lambda_{j1}^2 \right)^2
\end{aligned}$$

(注：この変形は SS を計算する場合に、より省略された計算方法としてよく知られているものと同じです。)

これが、因子の数だけあるので、

$$\begin{aligned}
Q &= \sum_{k=1}^p \left( \sum_{j=1}^m \lambda_{jk}^4 - \frac{1}{m} \left( \sum_{j=1}^m \lambda_{jk}^2 \right)^2 \right) \\
&= \sum_{k=1}^p \sum_{j=1}^m \lambda_{jk}^4 - \frac{1}{m} \sum_{k=1}^p \left( \sum_{j=1}^m \lambda_{jk}^2 \right)^2
\end{aligned}$$

このQを最大化して、最大化するように回転角 $\theta$ を定めればよいということです。このQは、バリマックス回転の基準として知られているものです。このようなQはいくつかあって、回転基準と呼ばれています。因子得点によってつくられる回転基準と呼ばれる関数

$Q(\lambda)$  ( $\lambda$ は因子負荷量)を考えて、その極値を求めるという説明がされます。そのような基準をまとめて整理したものが、直交回転の場合はオーソマックス基準です。次にオーソマックス基準を示します。

因子負荷行列を

$$\Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1p} \\ \vdots & \ddots & \vdots \\ \lambda_{m1} & \cdots & \lambda_{mp} \end{pmatrix}_{m \times p}$$

として

$$Q_{or} = \sum_{k=1}^p \sum_{j=1}^m \lambda_{jk}^4 - \frac{\omega}{m} \sum_{k=1}^p \left( \sum_{j=1}^m \lambda_{jk}^2 \right)^2$$

式 79

$\omega$ は重みで、 $\omega$ を変化させることによって、クォーティマックス回転、バイクォーティマックス回転、バリマックス回転、エカマックス回転、パーシマックス回転、因子パーシモニ

一回転の、オーソマックス基準に含まれる、6つの直交回得の方法（目標）が出来ます。  
各回転の $\omega$ は、以下の表のようになっています。

#### オーソマックス基準における重みと回転法

| 回転法         | $\omega$       |
|-------------|----------------|
| クォーティマックス   | 0              |
| バイクォーティマックス | 1/2            |
| バリマックス      | 1              |
| エカマックス      | p/2            |
| パーシマックス     | m(p-1)/(m+p-2) |
| 因子パーシモニー    | m              |

$p$ は因子数、 $m$ は観測変数の数です。一般の教科書では、観測変数の数を $p$ 、因子数を $m$ と表しているのが多いと思いますが、ここでは、今までの説明に合わせて、因子数を $p$ 、観測変数の数を $m$ と表していますので気を付けてください。

私たちは、バリマックス回転の基準を数学的・理論的に導き出しました。

$$Q_v = \sum_{k=1}^p \sum_{j=1}^m \lambda_{jk}^4 - \frac{1}{m} \sum_{k=1}^p \left( \sum_{j=1}^m \lambda_{jk}^2 \right)^2$$

これは、平方和の簡略計算の形なのですが、その変形過程で明らかなように、この式の2項目は平均値で、いわば、 $\lambda_j^2$ の期待値なのです。バリマックス回転では、因子負荷行列の列の代表値（分布中心・期待値）を平均値だと考えているのです。期待値を平均値にしなければならないということはありません。たとえば、クォーティマックス回転では、期待値は0です。本来、傾きは0だと考えれば、それも一理あります。また、バイクォーティマックスでは、期待値を0と平均値の中間に置いています。つまり、 $\omega$ は計算上は重みなのですが、因子に対する観測変数の傾き全体の期待値をどこに置くかということでもあります。 $\omega$ が大きくなると、因子間の説明力の差が小さくなり、 $\omega$ が小さくなると、特定の因子の説明力が強くなります。

筆者流の解説では、 $\lambda_{kj}$ は $\theta$ の関数として表されていますから、 $Q = Q(\theta)$ のように、 $Q$ をベクトル $\theta$ の関数として表すことが出来ますから

$$\frac{dQ(\theta)}{d\theta} = 0$$

のように、 $\theta$ の各成分で微分して0と置いて、連立方程式を作ってそれを解けば $\theta$ が得られます。この説明は理論的にはほぼ完璧なのですが、弱点があります。一般的な教科書で、

筆者流の説明が使われないのはそのためです。弱点の一つは、多次元空間での回転の公式が必要になることです。これはできないことではありませんが、それを一つの回転行列にまとめるのは面倒です。これは本質的な問題ではないかもしれません。重要なのもう一つの問題です。回転角を与えるという考え方をすると、すべてがノルム（互いの距離）を保ったまま回転しますから、直交回転にしか使えないのです。直交回転でも斜交回転でも、同じ考え方をするためには、回転行列に、直交行列であれば

$$\mathbf{T}\mathbf{T}^T = \mathbf{I}$$

斜交回転であれば

$$\text{diag}\mathbf{T}\mathbf{T}^T = \mathbf{I}$$

という、制約をつけて、 $Q = Q(\mathbf{T})$ と表して

$$\frac{dQ(\mathbf{T})}{d\mathbf{T}} = 0$$

の連立方程式を解くという方法が考えられます。これならば、斜交回転にもおなじ考え方が出来ます。ただし、そのためには、制約条件付きの関数の極値問題を解く必要があります。これは、一種のトレードオフで、回転角で考えた方が、推定すべき未知数が少なく、制約条件を考えずに、 $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ の範囲で微分すれば良いのですが、回転行列を考えると、未知数が増えて、その関係を別の式で制約しなければならないから、制約条件付きの極値問題になるのです。その代わりに空間的に難しいことを考える必要がなくなり、汎用性も高いのです。制限付きの極値問題の解法には、「最大・最少」のところでやったラグランジェの未定乗数法を使います。そこでは、図を平面的に表したので、立体的なイメージがわかりにくいかもしれません。立体的に表した説明を後述します。

斜交回転の場合は、制約条件を付けて回転するという考え方と、既知の情報からある目的を与えて、回転する（というよりは調整する）という考え方がありますが、制約条件を付けて回転する場合の回転基準は、オブリン基準と言われ、下の式で表されます。

$$Q_{ob} = \sum_{k < l = 1}^p \left\{ \sum_{j=1}^m \lambda_{jk}^2 \lambda_{jl}^2 - \frac{\omega}{m} \left( \sum_{j=1}^m \lambda_{jk}^2 \right) \left( \sum_{j=1}^m \lambda_{jl}^2 \right) \right\}$$

式 80

オブリン基準における重みと回転法

|          |          |
|----------|----------|
| 回転法      | $\omega$ |
| コーティミン   | 0        |
| バイコーティミン | 1/2      |
| コバリミン    | 1        |

この式を見ると、コバリミン回転は、共分散の簡便計算の形になっています。これらの式

を使って

$$\frac{\partial Q(\mathbf{T})}{\partial \mathbf{T}} = 0$$

として、 $Q(\mathbf{T})$ の極小値を与える $\mathbf{T}$ を求めます。この時の制約条件は $\text{diag} \mathbf{T} \mathbf{T}^T = \mathbf{I}$ です。そもそも、因子分析のモデルは、共分散や因子間相関に何の関心も持っていないのです。直交回転は、初期解を直交性を保って回転したというだけのことで、初期解において、座標軸とされたものが本当に直交性を持っているかどうかはわかりません。むしろ、オプティミム基準は共分散や因子間相関を最小化しようとしているので、事後的に直交性を与えようとしているように見えます。オプティミム基準では、 $\omega$ が大きくなると、因子相関性が弱くなります。

このほかに、回転基準の極値を求める斜交回転には、Crawford-Ferguson Family 基準、ジェオミン基準、Harris-Kaiser の独立クラスター回転など、様々な基準がありますが、その理論や特性を詳細に説明するだけの知識と経験が筆者にはありません。

#### プロクラステス回転・プロマックス回転

筆者には、斜交回転を回転 (rotation) と名付ける理由が良くわかりません。Rotation というよりは適当に adjust しているだけのような気がするのです。因子分析のモデルを考えると、適当に adjust して良いような気がします。最初から因子間の相関に興味がないのですから。そうだとすれば、既往の情報があるのならば、最小二乗法でその情報に近づけるように、 $\mathbf{T}$ を決めても構わないだろうというのも、割り切った考え方です。反対に、探索的に分析しているのであれば、それではいけないというのも一理あります。ある程度確実な定説があって、それに基づいて、仮説検証的に分析するのであれば、それでもよさそうです。これが、一つの考え方です。これをプロクラステス回転と言います。私は、プロクラステス調整といった方が良いような気がします。果たして、これが実用的に使われる場面があるのかどうかは知りませんが、この考え方を使って、直交回転の結果を尊重して、これを既往の情報として、これを基準に微調整するという考え方はありそうです。これを、プロマックス回転と言います。プロクラステス+バリマックスという意味です。ただ、バリマックス回転したものに無限に近づけたら、バリマックス回転そのものになってしまいます。そこで、因子負荷行列の値を 3 乗または 4 乗した因子負荷行列に最小二乗法的に近似します。そうすることで、1 に近い、高い付加量はほとんどそのまま、小さな負荷率はより小さくなります。

実際、どのように回転するかは、迷うところです。やはり、ロジカルには、バリマックス回転が自然だから、その微調整ということで、プロマックス回転というのが主流でしょう。プロマックス回転の結果を見て、そのままその結果を受け入れるか、別の回転法を試すかを考えます。例えば、2つの因子で説明できる観測変数が存在すると判断したらジェオミン基準の

ような単純化の程度の低い回転法を試したり、因子間の直交性が高いと判断したら、分析者の恣意的な判断が入りにくいバリマックス回転に戻すというのがありそうな判断です。筆者は因子分析の経験がありません。回転法そのものや回転法を持つ特性については、様々な議論があるようですから、機械的・硬直的に考えるのではなくて、そうした議論を参考にさせていただく方が良いでしょう。

### 制約条件付き極値問題—ラグランジュの未定乗数法

ラグランジュの未定乗数法は、すでに「最大・最少」のところで説明しました。その時は、平面的な図形で説明しましたが、ラグランジュの未定乗数法は、多次元空間での制約付きの極値問題の解法として使われるのが一般的です。

制約付きの極値問題とは、

$$f(\mathbf{X}) = f(x_1, \dots, x_n)$$

の様な関数があつて、変数 $\mathbf{X}$ が $g(\mathbf{X}) = c$ のように、何らかの条件で制約されているときに、 $f(\mathbf{X})$ の最大値、最小値、極値を求めるといふ問題です。たとえば、制約として、

$$x_1^2 + \dots + x_n^2 = k$$

$$x_1^2 + \dots + x_n^2 - k = 0$$

という中で、 $f(\mathbf{X})$ の極値を求めるといふことです。これは、図形的には簡単で、

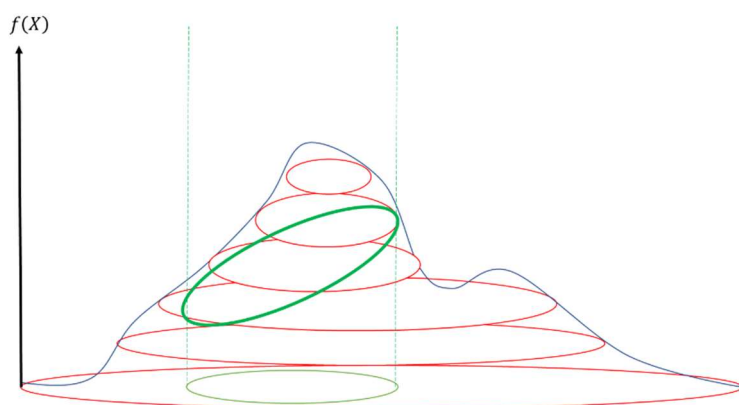


Fig. 93. 多次元空間における関数 $f(\mathbf{X})$ と制約条件 $g(\mathbf{X})$ .

この図のように、 $x_1, \dots, x_n$ を軸として作られた空間に制約条件を描き（図の細い緑の楕円）、それを、 $f(\mathbf{X})$ 軸を加えた、 $n+1$ 次元の曲面、 $f(\mathbf{X})$ の曲面に投影し（太い楕円、実際は多くの場合楕円ではない）その極値を求めればよいということです。この投影が出来れば、制約のない極値問題になります。

この図で、大切なことは、極値の部分では、赤い楕円で示した $f(\mathbf{X})$ の等高線の接線と、投影された緑の楕円の接線は、等しいということです。そうでなければ、緑の太い楕円に沿って微小区間を移動すると、等高線から離れてしまい、増加ないし、減少してしまうからです。つまり、極値を持つところで、それぞれの、接線とそれに直交する法線のベクトルの方向は

一致するのです。

$$\frac{df(\mathbf{X})}{d\mathbf{X}} = \frac{dg(\mathbf{X})}{d\mathbf{X}}$$

です。この関係を、一つの式で表したのが、ラグランジェ関数 (Lagrange function) で、以下の式で表せます。

$$L(x_1, \dots, x_n, \lambda) = f(x_1, \dots, x_n) - \lambda g(x_1, \dots, x_n)$$

この関数を  $x_1, \dots, x_n, \lambda$  で微分した時に 0 になるというのが、極値です。 $\lambda$  をラグランジェ乗数といいます。つまり、

$$\frac{\partial L}{\partial x_1} = \dots = \frac{\partial L}{\partial x_n} = \frac{\partial L}{\partial \lambda} = 0$$

が求める極値の条件です。

ラグランジェ関数はなかなか工夫された関数ですが、式を見ただけではその意味が解りません。部分に分解して解説します。

まず、 $\frac{\partial L}{\partial \lambda} = 0$  です。

$$\frac{\partial L}{\partial \lambda} = \frac{\partial f(x_1, \dots, x_n)}{\partial \lambda} - g(x_1, \dots, x_n) \frac{\partial \lambda}{\partial \lambda} = 0$$

で、 $f(x_1, \dots, x_n)$  には  $\lambda$  が含まれないから

$$\frac{\partial f(x_1, \dots, x_n)}{\partial \lambda} = 0$$

で  $\frac{\partial \lambda}{\partial \lambda} = 1$  だから

$$-g(x_1, \dots, x_n) = 0$$

$$g(x_1, \dots, x_n) = 0$$

つまり、この部分は、制約条件を表しているにすぎません。次に、 $\frac{\partial L}{\partial x_1} = \dots = \frac{\partial L}{\partial x_n}$  の部分に

ついて書くと

縦に並べて

$$\frac{\partial L}{\partial x_1} = \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} - \lambda \frac{\partial g(x_1, \dots, x_n)}{\partial x_1} = 0$$

$$\frac{\partial L}{\partial x_2} = \frac{\partial f(x_1, \dots, x_n)}{\partial x_2} - \lambda \frac{\partial g(x_1, \dots, x_n)}{\partial x_2} = 0$$

⋮

$$\frac{\partial L}{\partial x_n} = \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} - \lambda \frac{\partial g(x_1, \dots, x_n)}{\partial x_n} = 0$$

これを書き換えると

$$\begin{aligned}\frac{\partial f(x_1, \dots, x_n)}{\partial x_1} &= \lambda \frac{\partial g(x_1, \dots, x_n)}{\partial x_1} \\ \frac{\partial f(x_1, \dots, x_n)}{\partial x_2} &= \lambda \frac{\partial g(x_1, \dots, x_n)}{\partial x_2} \\ &\vdots \\ \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} &= \lambda \frac{\partial g(x_1, \dots, x_n)}{\partial x_n}\end{aligned}$$

$\frac{\partial f(x_1, \dots, x_n)}{\partial x_i}, \frac{\partial g(x_1, \dots, x_n)}{\partial x_i}$  をベクトルの成分と考えると、

$$\begin{pmatrix} \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \end{pmatrix} = \lambda \begin{pmatrix} \frac{\partial g(x_1, \dots, x_n)}{\partial x_1} \\ \vdots \\ \frac{\partial g(x_1, \dots, x_n)}{\partial x_n} \end{pmatrix}$$

$$\vec{f} = \lambda \vec{g}$$

で長さは違っていますが、方向が一致しています。つまり、ラグランジェの未定乗数法は、極値のところで、 $f(x_1, \dots, x_n)$  と  $g(x_1, \dots, x_n)$  の全微分が並行すると言っているのです。これで、ラグランジェの未定乗数法の説明は終わりなのですが、一つ蛇足を付け加えます

$$f(x_1, \dots, x_n) = c$$

ですが、これを、接線方向に微小な長さだけ移動します。接線方向つまり、等高線に沿って動かしたのだから、

$$f(x_1 + \Delta x_1, \dots, x_n + \Delta x_n) = c$$

です。一方、微小それぞれの傾き方向（接線方向）に微小だけ動かしたのだから、

$$f(x_1 + \Delta x_1, \dots, x_n + \Delta x_n) = c + \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} \Delta x_1 + \dots + \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \Delta x_n$$

ですから、

$$\frac{\partial f(x_1, \dots, x_n)}{\partial x_1} \Delta x_1 + \dots + \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \Delta x_n = 0$$

これを、行列演算の形で書くと、

$$\begin{pmatrix} \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} & \dots & \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} = 0$$

とベクトル  $\begin{pmatrix} \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} & \dots & \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \end{pmatrix}$  とベクトル  $(\Delta x_1 \quad \dots \quad \Delta x_n)$  の内積の形になります。

$(\Delta x_1 \quad \dots \quad \Delta x_n)$  は接線だから、 $\begin{pmatrix} \frac{\partial f(x_1, \dots, x_n)}{\partial x_1} & \dots & \frac{\partial f(x_1, \dots, x_n)}{\partial x_n} \end{pmatrix}$  はそれに直交する法線です。これ

を勾配 (gradient) といって、 $\nabla f$  と表します。これを使うと、ラグランジェの未定乗数法は、



$$\nabla f = \lambda \nabla g$$

と表せます。ラグランジェの未定方程式は、 $f$ と $g$ の法線の方が、その平面で一致することです。この場合、平面というのは抽象的な概念で、3次元の場合の平面ではなくて、 $f(\mathbf{X})$ を加えた、 $n+1$ 次元から見た、 $n$ 次元のことです。

#### VI-2-4-5. 因子数の決定

因子の抽出法、回転方法と並んで、因子分析で悩ましいのは、因子数をいくつにするかです。漠然と、独立した因子（直交性の強い）があると仮定して、その因子が観測側変数にどのくらい関わっているのか（因子負荷量）を求めているのですから、因子の数がある程度妥当なものでないと、分析の信頼性が損なわれます。ですから、妥当な因子数を考えることは、因子分析では本質的な問題です。

#### スクリープロット

主成分分析の考え方は、因子分析とは全く異なります。しかし、主成分分析は、データの空間的な配置を直交空間の中に位置づけているのですから、因子分析方法や因子数の決定に重要な情報を提供します。主成分分析で分散の大きな成分の方向に、因子のベクトルが向かっているだろうと想像するのは自然です。実際、初期の因子分析のソフトには、主成分分析の主成分を初期解とすることを選択できるソフトもありました。

主成分分析で主成分の数を決めるときに、寄与率を縦軸にとって、下記のように左から並べたグラフを作ります。これをスクリープロットと言います。このグラフで、崖（scree）のように、急激に寄与率が低下するところを見つけて、その崖よりも上の寄与率の主成分を選択します。（この例ではPC4まで。）。因子の数も主成分の数とほぼ同じだろうと考えるのです。

これは、1つの考え方です。寄与率とはそれぞれの成分の分散です。もともとのデータの観測変数の単位が、極端に違っていれば、分散を比較することに意味はありませんが、デ

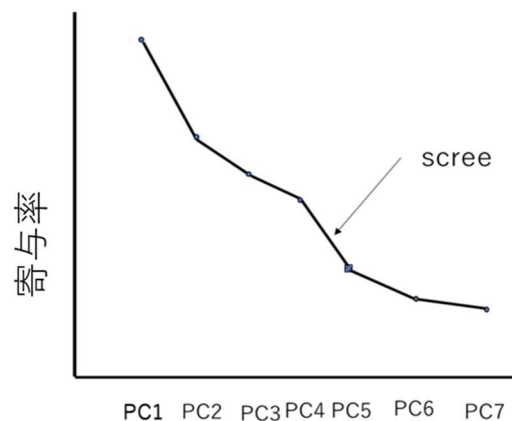


図 94.スクリープロットによる因子数の決定

ータが標準偏差で標準化されていれば、分散はその成分方向のデータの広がりですから、この比較には意味があります。つまり、分散がなだらかに変化する崖の下は、あまり意味のない成分だと考えるのです。

### ガットマン基準

さらに考えると、変数の分散は、1に統一されているのですから、主成分の分散が、1より大きいということは、主成分をとることによって、変動が大きくなったということで、それだけ、その主成分は説明力があることになります。下図のように、固有値が1を超える主成分の数だけ、因子があると考えるやり方もあります。これをガットマン基準と言います。

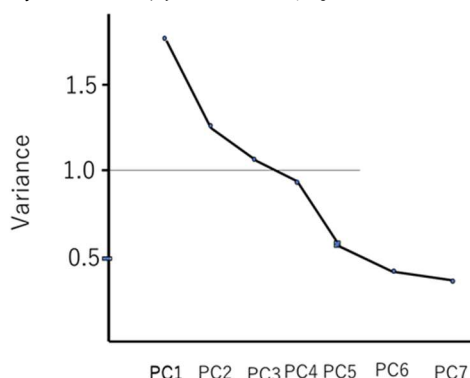


図 95. ガットマン基準による因子数の決定

### 平行分析

一方、分散が小さくても、それは他の成分の説明力が強すぎるためかもしれません。判断基準は、成分の数によって一定ではないとも考えられます。こういう場合、ランダムに選んだデータセットで実際に主成分分析を行って、固有値と成分の数の関係を描いてそれと比較するという方法も考えられます。そのような方法を平行分析と言います。特定の主成分がないようランダムに選ばれたデータでも、データの振れで、1より高い固有値になる成分もあるし、1より低い固有値になる成分もあるでしょう。これは誤差だけがつくる固有値の違いです。ランダムな変動ならばなだらかに低下するスクリープロットになります。それが図 96 の赤い曲線です。

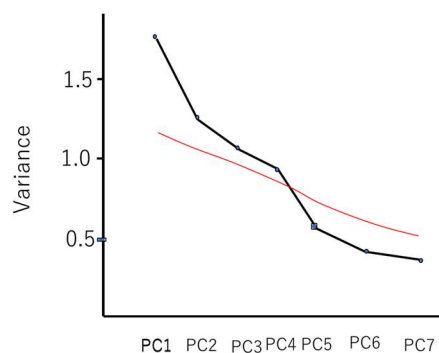


図 96. 平行分析による因子数の決定

この曲線の上に固有値がある成分を主成分として選択するということです。この主成分の数を因子の数とします。これが並行分析です。

#### MAP(Minimum Average Partial) test

スクリープロット、ガットマン基準、並行分析のいずれも、主成分のスクリープロットに依存して因子数の候補を考える方法です。この方法の最大の欠点は、因子分析と主成分分析は異なる分析だということです。事前情報として主成分分析の結果は参考になりますが、そのまま因子分析には使えません。主成分分析では、主成分の数をいくつに決めるかというのは、どこまで説明するかという、説明上の問題であって、主成分の数が違って、上位の主成分の分析結果は変わりません。しかし、因子分析では、因子の数は分析結果に大きく影響します。ですから、因子分析の結果に基づいて、因子数の妥当性を議論する必要があります。その方法の一つが、MPA です。

MPA では、因子を観測変数の統制変数だと考えます。統制変数については、偏相関分析の章を参考にしてください。簡単に言えば、複数の観測変数が相関を持っているのは、ある変数（この場合は因子）と強く相関しているためであって、その変数の影響を取り除けば、それらの観測変数には相関がないということです。もちろん、観測変数間に何らかの因果関係があるという場合もあり得ます。その関係がどのようなになっているのかを分析するのが偏相関分析で、偏相関分析の結果、観測変数の偏相関が小さければ、観測変数間には本来因果関係がないにもかかわらず、統制変数の強い影響で相関が観察されるのだと考えるのです。ある因子に対する観測変数間に偏相関がなければ、全体として、偏相関の和（二乗和）は小さくなるはずです。つまり、統制変数（因子）のみで説明できるということです。観測変数が6つあって、因子が3つの場合、観測変数を2つずつ選び出しますから、一つの因子に対して、変数の組み合わせの数は

$${}_6C_2 = \frac{6 \times 5}{2} = 15$$

でこれに因子数がかかるので、45通りの組み合わせについて、偏相関係数の二乗の総和を求めます。

これを、因子の数を変えた因子分析について行い、偏相関変数の二乗の総和を求めて、比較します。そして、最小の偏相関変数の二乗の総和を与える因子数を選びます。

事後的に因子数の妥当性を評価するので、論理的なテストとしては悪くないでしょう。しかし、ある因子に対して高い因子負荷量になっている観測変数同士に因果関係がないという仮定はいつでも成り立つわけではないだろうと筆者は思います。実際には、ある程度因果関係があるのではないかと思います。その場合、この評価は正確ではないし、多分、過小評価します。MPA も絶対というわけではないでしょう。

それでも、個々の因子について偏相関分析を行って、観測変数同士の関係を見ておくことは、結果の解釈のために有効でしょう。

結論としては、MPA は過小評価で、並行分析は多めの因子数を推薦すると言えますので、この2つの間で、適切な因子数を選ぶということになるでしょう。