



Title	クラスタリングを伴う因子分析法の研究開発
Author(s)	宇野, 光平
Citation	大阪大学, 2019, 博士論文
Version Type	VoR
URL	<a href="https://doi.org/10.18910/72457">https://doi.org/10.18910/72457</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

# クラスタリングを伴う因子分析法の研究開発

行動生態学 行動統計科学研究分野

宇野 光平

21A16804

2018年12月14日

# 目 次

<b>第 1 章 数学的準備</b>	<b>4</b>
1.1 記法 . . . . .	4
1.2 線形代数と最小二乗法 . . . . .	5
1.2.1 ベクトルとユークリッドノルム . . . . .	5
1.2.2 行列と多変量データ解析とフロベニウスノルム . . . . .	8
1.3 確率分布 . . . . .	28
1.3.1 正規分布 . . . . .	28
1.3.2 条件付き正規分布 . . . . .	29
<b>第 2 章 序論</b>	<b>33</b>
2.1 因子分析の 3 つのタイプ . . . . .	35
2.1.1 確率モデル . . . . .	35
2.1.2 母数モデル . . . . .	41
2.1.3 行列モデル . . . . .	42
2.1.4 因子分析と主成分分析の違い . . . . .	43
2.2 因子分析における 2 つの不定性 . . . . .	44
2.2.1 回転の不定性 . . . . .	44
2.2.2 因子得点の不定性 . . . . .	45
<b>第 3 章 クラスタリングを伴う母数モデル因子分析</b>	<b>49</b>
3.1 導入 . . . . .	49
3.2 モデル . . . . .	51
3.3 アルゴリズム . . . . .	53
3.4 シミュレーション . . . . .	55
3.4.1 データの合成と分析 . . . . .	56

---

3.4.2 結果 . . . . .	56
3.5 類似手法 . . . . .	57
3.5.1 Reduced K-means analysis . . . . .	57
3.5.2 Factorial K-means analysis . . . . .	58
3.5.3 タンデム分析 . . . . .	59
3.6 実データ解析 . . . . .	59
3.6.1 あやめデータ . . . . .	60
3.6.2 糖尿病データ . . . . .	61
3.6.3 靈長類肩甲骨データ . . . . .	62
3.6.4 独自分散と群間分散の関係 . . . . .	63
3.7 考察 . . . . .	64
<b>第4章 行列モデル因子分析の因子得点同定</b>	<b>66</b>
4.1 導入 . . . . .	66
4.2 行列モデル因子分析 . . . . .	68
4.3 提案手法 . . . . .	70
4.4 アルゴリズム . . . . .	72
4.4.1 クラスタリングステップ . . . . .	72
4.4.2 因子得点ステップ . . . . .	73
4.4.3 アルゴリズムの全体 . . . . .	75
4.5 クラスター数 . . . . .	76
4.6 シミュレーション . . . . .	77
4.6.1 データの生成と分析 . . . . .	77
4.6.2 計算時間と局所解の頻度 . . . . .	79
4.6.3 VRCの結果 . . . . .	80
4.6.4 パラメータ復元 . . . . .	80
4.7 実データ解析 . . . . .	82
4.7.1 個体数が少ないデータでの例 . . . . .	82
4.7.2 個体数が多いデータでの例 . . . . .	86
4.8 考察 . . . . .	86

---

<b>第5章 変数クラスタリングのための因子分析</b>	<b>90</b>
5.1 完全単純構造因子分析 . . . . .	90
5.1.1 Thurstone の単純構造ルール . . . . .	90
5.1.2 モデル . . . . .	94
5.1.3 アルゴリズム . . . . .	96
5.1.4 実データ解析 . . . . .	99
5.2 スパース推定 . . . . .	103
5.2.1 lasso ペナルティとパラメータ推定 . . . . .	104
5.2.2 因子分析におけるスパース推定 . . . . .	107
5.3 考察 . . . . .	114
<b>第6章 総合考察</b>	<b>116</b>
6.1 因子分析における不定性の解消 . . . . .	116
6.1.1 回転の不定性の解消 . . . . .	116
6.1.2 因子得点の不定性の解消 . . . . .	117
6.2 多変量データ解析における本研究の位置づけ . . . . .	118
<b>付録A</b>	<b>120</b>
A.1 . . . . .	120
A.2 . . . . .	120
A.3 . . . . .	121
A.4 . . . . .	121
<b>付録B</b>	<b>122</b>
B.1 . . . . .	122
B.2 . . . . .	122
B.3 . . . . .	122
B.4 . . . . .	123
<b>参考文献</b>	<b>125</b>

# 第1章

## 数学的準備

### 1.1 記法

本学位論文では、記号を用いて説明を簡素化する場合がある。できる限り、そのページにおいても記号の説明はするものの、理解を助けるものとして、表記表を用いる。

記号	意味
$\mathbb{R}^P$	$P$ 次元実数空間
$\mathbb{R}^{N \times P}$	$N \times P$ 次元実数空間
$x \in \mathbb{R}^P$	$x$ は $P$ 次元ベクトル (太字の小文字アルファベットはベクトル)
$\mathbf{1}_P$	全ての要素が 1 の $P$ 次元のベクトル
$\mathbf{X} \in \mathbb{R}^{N \times P}$	$\mathbf{X}$ は $N \times P$ の行列 (ボールド体の大文字アルファベットは行列)
$\mathbf{I}_P$	$P$ 次元の単位行列
$\mathbf{O}_{P \times M}$	全ての要素が 0 の行列
$x^\top, \mathbf{X}^\top$	ベクトル $x$ の転置、および行列 $\mathbf{X}$ の転置行列
$\mathbf{X}^{-1}$	正則行列 $\mathbf{X}$ の逆行列
$\text{tr}(\mathbf{X})$	正方行列 $\mathbf{X}$ のトレース
$\text{rank}(\mathbf{X})$	行列 $\mathbf{X}$ のランク
$\det(\mathbf{X})$	正方行列 $\mathbf{X}$ の行列式
$[\mathbf{X}, \mathbf{Y}]$	行列 $\mathbf{X}$ と $\mathbf{Y}$ からなるブロック行列
$\ x\ _0$	ベクトル $x$ の $\ell_0$ (擬) ノルム
$\ x\ _1$	ベクトル $x$ の $\ell_1$ ノルム
$\ x\ _2$	ベクトル $x$ の $\ell_2$ (ユークリッド) ノルム
$\ \mathbf{X}\ _F$	行列 $\mathbf{X}$ のフロベニウスノルム
$\ \mathbf{X}\ _1$	行列 $\mathbf{X}$ の要素ごとの $\ell_1$ ノルム
$\log$	自然対数関数
$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	平均ベクトルが $\boldsymbol{\mu}$ 、分散共分散行列が $\boldsymbol{\Sigma}$ の多変量正規分布
$\mathbb{E}[x]$	確率変数 $x$ の期待値
$\text{Var}[x]$	確率変数 $x$ の分散
$\text{Cov}[x, y]$	確率変数 $x$ と確率変数 $y$ の共分散

## 1.2 線形代数と最小二乗法

### 1.2.1 ベクトルとユークリッドノルム

2次元ユークリッド空間上での点と点の距離を考えてみる。3点  $(2, 2)$ ,  $(-3, 1)$ ,  $(1, -2)$  があるとし、便宜上それを点1, 点2, 点3とする。これを図で表すと以下になる。

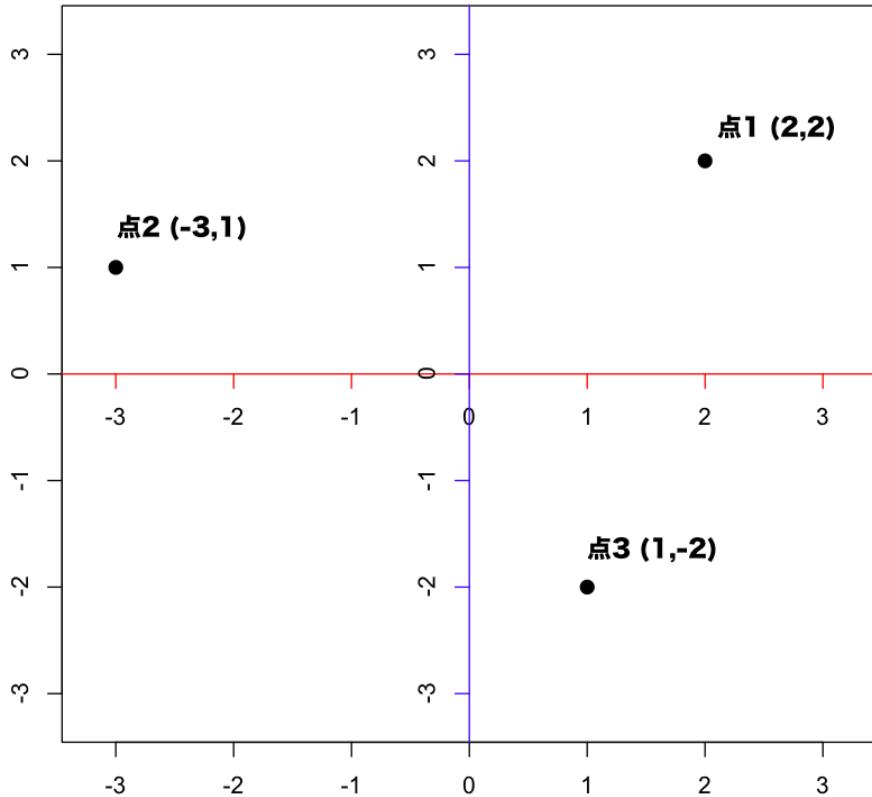


図 1.1: 3点の散布図

点1と点2の距離、そして点1と点3の距離はそれぞれ

$$\sqrt{(2 - (-3))^2 + (2 - 1)^2} = \sqrt{26} \quad (1.1)$$

$$\sqrt{(2 - 1)^2 + (2 - (-2))^2} = \sqrt{17} \quad (1.2)$$

となる。このことから、点2よりも点3のほうが点1に近いことがわかる。

実はこうした距離の概念は、統計学やデータサイエンスにおいても重要である。特に線形代数に基づく多変量解析では必須といつてもよい。そこで、まず図1.1における3点を線形代数によって再解釈する。

線形代数において3点はもはや点ではなくベクトルとして扱う。つまり

$$\mathbf{a} = [2, 2]^\top$$

$$\mathbf{b} = [-3, 1]^\top$$

$$\mathbf{c} = [1, -2]^\top$$

のように、要素が2つあることから2次元ベクトルとなる。つまり  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2$ 。このとき点間の距離(1.1)式と(1.2)式はベクトルどうしの距離と解釈でき、それぞれ

$$\begin{aligned}\|\mathbf{a} - \mathbf{b}\|_2 &= \sqrt{(\mathbf{a} - \mathbf{b})^\top (\mathbf{a} - \mathbf{b})} = \sqrt{[2 - (-3), 2 - 1]^\top [2 - (-3), 2 - 1]} = \sqrt{26} \\ \|\mathbf{a} - \mathbf{c}\|_2 &= \sqrt{(\mathbf{a} - \mathbf{c})^\top (\mathbf{a} - \mathbf{c})} = \sqrt{[2 - 1, 2 - (-2)]^\top [2 - 1, 2 - (-2)]} = \sqrt{17}\end{aligned}$$

となる。二乗根が余計な計算になるため、後述のようにベクトルのユークリッド距離は二乗して計算することが一般的である。

図1.1のように我々は2次元あるいは3次元の点を視覚的に捉えることはできるが、より高次元の空間を認知することは難しい。しかし、ベクトルを用いることで4次元以上の点を表現することは可能である。例えば100次元のベクトル

$$\mathbf{x} = [x_1, x_2, \dots, x_{100}]^\top$$

を考えれば、数式の上では100次元の世界を表現することができる。

2次元に限定せず、次元を一般化したベクトルにも距離を定義することはできる。 $N$ 次元ベクトル  $\mathbf{y}, \mathbf{z} \in \mathbb{R}^N$  のユークリッド距離の二乗は、

$$\begin{aligned}\|\mathbf{y} - \mathbf{z}\|_2^2 &= (\mathbf{y} - \mathbf{z})^\top (\mathbf{y} - \mathbf{z}) \\ &= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{z} - \mathbf{z}^\top \mathbf{y} + \mathbf{z}^\top \mathbf{z} \\ &= \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{z} + \mathbf{z}^\top \mathbf{z}\end{aligned}\tag{1.3}$$

で表すことができる。最後の等式は内積の性質である(1.5)式を用いた。ユークリッド距離はユークリッドノルム  $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\top \mathbf{x}}$  に基づいている。ベクトルの演算に関しては線形代数の書籍によって知ることができる。和書では特に新井(2006)、岩崎・吉田(2006)が統計学への応用を意識して書かれており読みやすい。洋書ではBoyd & Vandenberghe(2018)、Gentle(2017)、Harville(2008)、Magnus & Neudecker(2007)、Schott(2016)が統計学者に向けて書かれている。

**定義 1.1.** ベクトルの内積

$P$  次元ベクトル  $\mathbf{y}$  と  $\mathbf{z}$  があるとする ( $\mathbf{y}, \mathbf{z} \in \mathbb{R}^P$ ). このとき

$$\mathbf{y}^\top \mathbf{x} \quad (1.4)$$

をベクトル  $\mathbf{y}$  と  $\mathbf{z}$  の内積と呼ぶ.

**定理 1.1.** 内積の可換性

$$\mathbf{y}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{y} \quad (1.5)$$

**定義 1.2.** ユークリッドノルム

ベクトル  $\mathbf{x}$  の  $\ell_2$  ノルム (ユークリッドノルム) は,  $\mathbf{x}$  と  $\mathbf{x}$  の内積の二乗根, つまり

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^\top \mathbf{x}} \quad (1.6)$$

である.

**定義 1.3.** ユークリッド距離

$P$  次元ベクトル  $\mathbf{y}$  と  $\mathbf{z}$  があるとする. このとき  $\mathbf{y}$  と  $\mathbf{z}$  のユークリッド距離を

$$\|\mathbf{y} - \mathbf{z}\|_2 = \sqrt{(\mathbf{y} - \mathbf{z})^\top (\mathbf{y} - \mathbf{z})} \quad (1.7)$$

と定義する. ベクトル間のユークリッド距離は, 二つのベクトルの差ベクトル

$$\mathbf{e} = \mathbf{y} - \mathbf{z} \quad (1.8)$$

のユークリッドノルム

$$\|\mathbf{e}\|_2 = \sqrt{\mathbf{e}^\top \mathbf{e}} = \sqrt{(\mathbf{y} - \mathbf{z})^\top (\mathbf{y} - \mathbf{z})} \quad (1.9)$$

と捉えることもできる.

### 1.2.2 行列と多変量データ解析とフロベニウスノルム

#### 行列

行列は、ベクトルをより一般化した概念として捉えることができる。 $N$  行  $\times P$  列の行列  $\mathbf{X}$  は

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1P} \\ x_{21} & x_{22} & \dots & x_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & & & x_{NP} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{x}}_1^\top \\ \vdots \\ \tilde{\mathbf{x}}_N^\top \end{bmatrix} = [\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_N]^\top = [\mathbf{x}_1, \dots, \mathbf{x}_P]$$

のように縦だけでなく横にも要素が広がっており、行ベクトル  $\tilde{\mathbf{x}}_n$  が  $N$  個縦に連なっている、あるいは列ベクトル  $\mathbf{x}_p$  が  $P$  個横に連なっていると解釈できる。また、ベクトルの要素は  $x_n$  と添え字が 1 つであったが、行列の要素は  $x_{np}$  と添え字が 2 つであるという解釈も可能である。今後は添え字によって十分に判断できる場合、行からベクトルを抽出した場合でも  $\mathbf{x}_n$  と表記することとする。

ベクトルと行列は多変量データ解析の理解には欠かせない。例えば 100 人の生徒に 5 教科のテストを受けさせ採点したとしよう。これは  $100 \times 5$  の行列  $\mathbf{X} \in \mathbb{R}^{100 \times 5}$  によって表現することができる。生徒 1 の英語が 86 点、数学が 57 点とすれば

$$\begin{bmatrix} & \text{英語} & \text{数学} & \text{国語} & \text{理科} & \text{社会} \\ \text{生徒 1} & 86 & 57 & 71 & 82 & 91 \\ \text{生徒 2} & 81 & 86 & 88 & 75 & 69 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{生徒 100} & 59 & 43 & 62 & 62 & 41 \end{bmatrix}$$

のようにすることで、得られたデータを行列で表現できる。さらに  $n$  行目のベクトルは生徒  $n$  の各教科の得点、1 列目のベクトルは各生徒の英語の得点、というようにベクトルも重要な役割を果たす。足立・村上(2011) や Adachi(2016), ten Berge(1993) では、線形代数が多変量データ解析においてどのように使われているかを詳細に書いている。また行列のリパラメetrize としての固有値分解、特異値分解の重要性も強調されている。実際、後述のトレース最大化に関する ten Berge の定理など、特異値分解を用いた最適化は応用例が多い。

**定義 1.4.** 固有値分解

対称行列  $\mathbf{S} \in \mathbb{R}^{P \times P}$  があるとする。このとき

$$\mathbf{S} = \mathbf{LDL}^\top \quad (1.10)$$

のように  $\mathbf{S}$  をリパラメトライズすることが可能である。ここで、 $\mathbf{D} \in \mathbb{R}^{P \times P}$  は対角要素が  $\mathbf{S}$  の固有値からなる対角行列である。降順に固有値を配置することが一般的である ( $d_1 \geq d_2 \geq \dots \geq d_P$ )。 $d_P \geq 0$  となる行列  $\mathbf{S}$  を非負定値行列という。そして  $\mathbf{L} \in \mathbb{R}^{P \times P}$  は各列が固有ベクトルからなる行列であり、正規直交性を満たす。すなわち

$$\mathbf{L}^\top \mathbf{L} = \mathbf{LL}^\top = \mathbf{I}_P \quad (1.11)$$

**定義 1.5.** 特異値分解

任意のサイズの行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$  があるとする。 $\mathbf{X}$  を

$$\mathbf{X} = \mathbf{K}\Delta\mathbf{L}^\top \quad (1.12)$$

のようにリパラメトライズすることができる。ただし  $\Delta \in \mathbb{R}^{P \times P}$  は対角要素が  $\mathbf{X}$  の特異値からなる対角行列であり、降順に特異値を配置することが一般的である ( $d_1 \geq \dots \geq d_P$ )。 $\mathbf{K} \in \mathbb{R}^{N \times P}$  は正規直交性を満たす左特異行列であり、 $\mathbf{L} \in \mathbb{R}^{P \times P}$  は正規直交性を満たす右特異行列である。つまり、

$$\mathbf{K}^\top \mathbf{K} = \mathbf{L}^\top \mathbf{L} = \mathbf{LL}^\top = \mathbf{I}_P \quad (1.13)$$

である。ただし行列のサイズの問題から、 $\mathbf{K}^\top \neq \mathbf{K}$  であるため  $\mathbf{KK}^\top \neq \mathbf{I}_N$  であることに注意する必要がある。

**系 1.1.** 行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$  ( $P \leq N$ ) があり、その共分散行列  $\mathbf{X}^\top \mathbf{X}$  が非負定値行列であるとする。このとき、 $\mathbf{X}$  を特異値分解をすると、正規直交性と対角行列の性質から

$$\begin{aligned} \mathbf{X}^\top \mathbf{X} &= \mathbf{L}\Delta\mathbf{K}^\top \mathbf{K}\Delta\mathbf{L}^\top \\ &= \mathbf{L}\Delta^2\mathbf{L}^\top \end{aligned} \quad (1.14)$$

が成り立つ。 $\Delta^2 = \mathbf{D}$  とすれば、これは固有値分解の式 (1.10) 式と一致する。

## 多変量データ解析(回帰分析)

多変量データ解析の手法においては、ユークリッド距離も頻繁に見られる。たとえば重回帰分析の場合を考える。目的変数ベクトル  $\mathbf{y}$  を説明変数によって近似するのが重回帰分析だが、(1.3) 式における  $z$  を  $z = \mathbf{X}\beta$ 、(ただし  $\mathbf{X} \in \mathbb{R}^{N \times P}, \beta \in \mathbb{R}^P, P \leq N$ ) とリパラメトライズすると以下に書き換えられる。

$$f(\beta|\mathbf{y}, \mathbf{X}) = \|\mathbf{y} - \mathbf{X}\beta\|_2^2 = (\mathbf{y} - \mathbf{X}\beta)^\top(\mathbf{y} - \mathbf{X}\beta) \quad (1.15)$$

(1.15) 式における  $\beta$  は回帰分析における回帰係数ベクトルであり、値の正負や大小によって説明変数が目的変数に与える影響を表現できる。

### 回帰分析

$\|\mathbf{y} - \mathbf{X}\beta\|_2^2$ : 回帰分析の目的関数

$\mathbf{y}$ : 各個体  $n(n=1, \dots, N)$  の予測したい変数

からなる  $N \times 1$  の目的変数ベクトル

$\mathbf{X}$ : 各個体の  $P$  個の説明変数からなる  $N \times P$  の行列

$\beta$ : 各変数の重みを表現する  $P \times 1$  の回帰係数ベクトル

回帰分析において  $\mathbf{y}$  と  $\mathbf{X}$  は与えられており、推定すべきパラメータは  $\beta$  のみである。ここでは目的関数 (1.15) 式を最小化するような  $\beta$ ,  $\operatorname{argmin}_\beta f(\beta|\mathbf{y}, \mathbf{X})$  を最適パラメータとする。こうした最適化法は最小二乗法と呼ばれる。

具体的な最適化にはベクトルの偏微分 (Magnus & Neudecker, 2007) を用いて以下の方程式を得る。

$$\begin{aligned} \frac{\partial f(\beta|\mathbf{y}, \mathbf{X})}{\partial \beta} &= \frac{\partial}{\partial \beta} (\mathbf{y} - \mathbf{X}\beta)^\top(\mathbf{y} - \mathbf{X}\beta) \\ &= \frac{\partial}{\partial \beta} (\mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X}\beta - \beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta) \\ &= \frac{\partial}{\partial \beta} (\mathbf{y}^\top \mathbf{y} - 2\beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta) \\ &= 2(-\mathbf{X}^\top \mathbf{y} + \mathbf{X}^\top \mathbf{X}\beta) = \mathbf{0}_P \end{aligned} \quad (1.16)$$

(1.16) 式の 3 行目の等式には

$$\mathbf{y}^\top \mathbf{X}\beta = \beta^\top \mathbf{X}^\top \mathbf{y} \quad (1.17)$$

を用いた。これは  $\mathbf{y}^\top \mathbf{X}\beta$  がベクトル  $\mathbf{y}$  と  $\mathbf{X}\beta$  の内積であることから、内積の可換性 (1.5) 式より成り立つ。また 4 行目の等式はベクトルに関する偏微分

$$\frac{\partial \beta^\top \mathbf{X}^\top \mathbf{y}}{\partial \beta} = \mathbf{X}^\top \mathbf{y} \quad (1.18)$$

と

$$\begin{aligned} \frac{\partial \beta^\top \mathbf{X}^\top \mathbf{X}\beta}{\partial \beta} &= \frac{\partial \beta^\top (\mathbf{X}^\top \mathbf{X}\beta)}{\partial \beta} + \frac{\partial (\beta^\top \mathbf{X}^\top \mathbf{X}) \beta}{\partial \beta} \\ &= 2\mathbf{X}^\top \mathbf{X}\beta \end{aligned} \quad (1.19)$$

より従う。よって最適な  $\hat{\beta}$  は

$$\begin{aligned} \mathbf{X}^\top \mathbf{X}\hat{\beta} &= \mathbf{X}^\top \mathbf{y} \\ \hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \end{aligned} \quad (1.20)$$

によって得られる。ここで、 $(\mathbf{X}^\top \mathbf{X})^{-1}$  は正方行列  $\mathbf{X}^\top \mathbf{X}$  の逆行列である。

**定義 1.6. ベクトル偏微分**

ベクトル  $\mathbf{x} \in \mathbb{R}^P$  と行列  $\mathbf{A} \in \mathbb{R}^{N \times P}$  があるとする。このとき  $\mathbf{Ax}$  の  $\mathbf{x}$  についての偏微分は

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}^\top \quad (1.21)$$

で与えられる。またベクトル  $\mathbf{x} \in \mathbb{R}^P$  と行列  $\mathbf{B} \in \mathbb{R}^{P \times N}$  があるとして、 $\mathbf{x}^\top \mathbf{B}$  の  $\mathbf{x}$  についての偏微分は

$$\frac{\partial \mathbf{x}^\top \mathbf{B}}{\partial \mathbf{x}} = \mathbf{B} \quad (1.22)$$

である。

**定義 1.7. 逆行列**

逆行列は行列の演算における逆数に対応する概念で、正方行列  $\mathbf{X} \in \mathbb{R}^{P \times P}$  の逆行列を  $\mathbf{X}^{-1}$  と書くと、

$$\mathbf{XX}^{-1} = \mathbf{X}^{-1}\mathbf{X} = \mathbf{I}_P$$

が成り立つ。ただし、逆行列が成り立たない正方行列も存在し、その行列を非正則行列や特異行列と呼ぶ。正方行列の固有値が全て 0 より上、つまり正定値のときは正則行列である。

ここで、パラメータ  $\beta$  の役割および目的関数最小化の意味を考えてみる。簡単にするために、説明変数が 1 変数のみの単回帰分析を例とする。つまり  $P = 1$  とする。さらに切片も 0 とすると、目的関数 (1.16) 式は

$$f(\beta | \mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \beta \mathbf{x}\|_2^2 \quad (1.23)$$

となる。 $P = 1$  であるから、 $\mathbf{x} \in \mathbb{R}^N, \beta \in \mathbb{R}$  である。具体例をあげて考察するため、今  $N = 100$  人の生徒の国語を説明変数、英語の点数を目的変数としてそれらが既知とする。図 1.2 は縦軸を英語の点数、横軸を国語の点数とし、それぞれ標準化したときの散布図である。回帰分析は目的変数を説明変数を用いて近似することを目的としているが、図 1.2においては縦軸である英語の得点を、横軸である国語の得点を用いて近似することが目的である。単回帰において、回帰係数はスカラー  $\beta$  であり、この値は図 1.3 における直線の

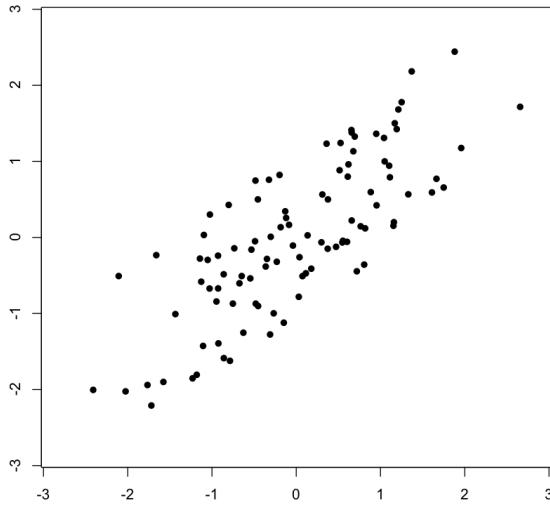


図 1.2: 各生徒の英語と国語の得点散布図

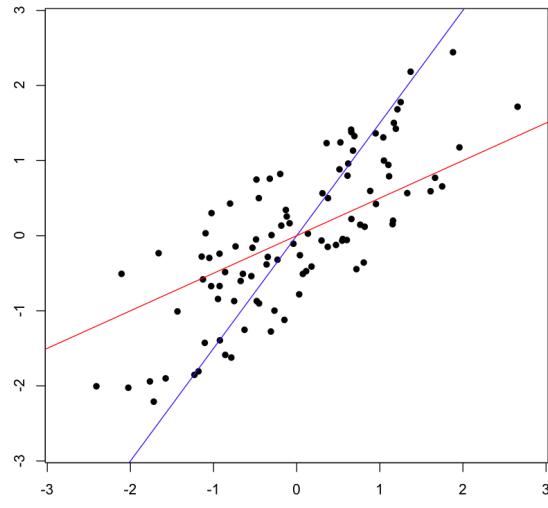


図 1.3: 得点散布図といくつかの直線

傾きを表している。しかし、図 1.3 のように様々な直線が存在可能であるため、最適な直線の傾き  $\beta$  を推定する必要がある。では、どのような基準に基づいて最適なパラメータを推定すればよいだろうか。今回は最小二乗法によって求めることとする。最小二乗法の主な構組みは以下である。

まず、得られたデータに対して何らかの仮説であるモデルを立てる。そしてそのデータとモデルとの間の誤差を最小化する。つまり

$$\text{データ} = \text{モデル} + \text{誤差}$$

としたときに誤差を最小化するようなパラメータを推定する。実際は

$$\text{誤差} = \text{データ} - \text{モデル}$$

であるから、データとモデルの距離を最小化する。

今回の例である単回帰では、データは英語の得点である  $y$ 、モデルは回帰係数と国語の得点の積  $\beta x$  であるから、

$$\|y - \beta x\|_2^2$$

が最小化すべき目的関数になる。実際これは (1.23) 式と一致する。距離として計算の都合

上からユークリッド距離を採用しているが、ユークリッドノルムの定義から

$$\|\mathbf{y} - \beta\mathbf{x}\|_2^2 = \sum_{n=1}^N (y_n - \beta x_n)^2$$

であるため、目的関数(1.23)式は図1.3における直線と各点との二乗誤差の総和とも考えられる。

### 多変量データ解析(主成分分析)

ユークリッド距離(1.7)式はベクトルの距離として定義されたが、今度は行列の距離を導入する。 $N \times P$ 次元の行列  $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N \times P}$  の距離を

$$\|\mathbf{X} - \mathbf{Y}\|_F^2 = \text{tr}(\mathbf{X} - \mathbf{Y})^\top (\mathbf{X} - \mathbf{Y}) \quad (1.24)$$

と定義する。これはフロベニウスノルム  $\|\mathbf{X}\|_F = \sqrt{\text{tr}(\mathbf{X}^\top \mathbf{X})} = \sqrt{\sum_{n=1}^N \sum_{p=1}^P |x_{np}|^2}$  に基づいている。

ここで、(1.24)式を  $\mathbf{Y} = \mathbf{FA}^\top$  ( $\mathbf{F} \in \mathbb{R}^{N \times M}, \mathbf{A} \in \mathbb{R}^{P \times M}$ ) とリパラメetrizeする。さらに  $\mathbf{A}$  に列についての正規直交制約を加える。よって  $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_M$ 。このとき(1.24)式は

$$f_{\text{PCA}}(\mathbf{F}, \mathbf{A} | \mathbf{X}) = \|\mathbf{X} - \mathbf{FA}^\top\|_F^2 \quad (1.25)$$

となる。これは主成分分析(Jolliffe, 2006)の目的関数の1つと考えることができる。主成分分析では、データ行列  $\mathbf{X}$  が与えられたときに主成分得点行列  $\mathbf{F}$  と主成分負荷量行列  $\mathbf{A}$  を推定する。

#### 主成分分析

$\|\mathbf{X} - \mathbf{FA}^\top\|_F^2$  : 主成分分析の目的関数

$\mathbf{F}$  : 各個体が主成分空間上での値を

説明する  $N \times M$  の主成分得点行列

$\mathbf{A}$  : 変数と主成分の関係を表す  $P \times M$  の主成分負荷行列

推定するパラメータが複数ある場合、一度に推定することができない場合がある。こうした場合は、各パラメータを推定し収束したと判断するまで計算を続ける方法を取る。これを交互最小二乗法(森・黒田・足立, 2017)と呼ぶ。

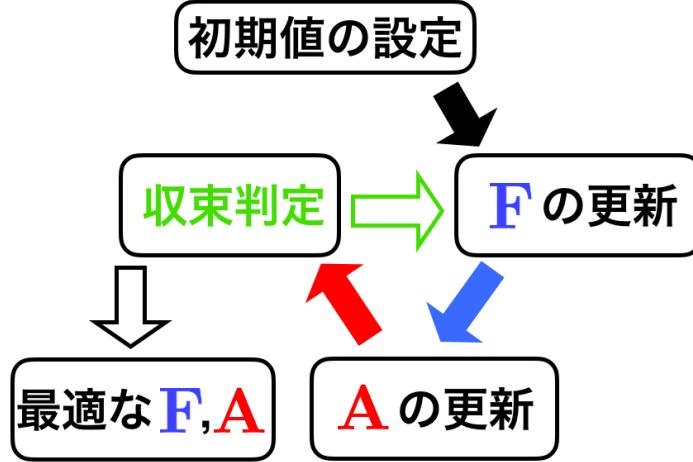


図 1.4: 交互最小二乗法の概要図

各パラメータを推定する際には、他のパラメータは既知として計算する。あるパラメータを既知として扱うことを、そのパラメータを固定するともいう。各パラメータを更新することで目的関数が逐次減少していくことを単調減少と呼び、交互最小二乗法はこの単調減少性が保証されている。また、交互最小二乗法のように収束するまでパラメータの更新を繰り返す必要がある計算を、反復計算と呼ぶことがある。

**定義 1.8.** 行列のトレース

ある  $P \times P$  の正方行列  $\mathbf{X}$  が与えられたとき, 各対角要素の総和

$$\text{tr}(\mathbf{X}) = \sum_{p=1}^P x_{pp} \quad (1.26)$$

を行列  $\mathbf{X}$  のトレースと呼ぶ.

**定理 1.2.** トレースの性質

行列の積  $\mathbf{XY}$  があり,  $\mathbf{X} \in \mathbb{R}^{N \times P}, \mathbf{Y} \in \mathbb{R}^{P \times N}$  とする. このとき,

$$\text{tr}(\mathbf{XY}) = \text{tr}(\mathbf{YX}) \quad (1.27)$$

さらに  $\mathbf{XY}$  のトレースと,  $\mathbf{XY}$  の転置行列のトレースは等しい. よって

$$\text{tr}(\mathbf{XY}) = \text{tr}((\mathbf{XY})^\top) = \text{tr}(\mathbf{Y}^\top \mathbf{X}^\top) \quad (1.28)$$

**系 1.2.** トレースとしてのベクトル内積の解釈

$P$  次元ベクトル  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^P$  の内積は  $\mathbf{x}^\top \mathbf{y}$  である. ここで,  $\mathbf{x}^\top \mathbf{y}$  を  $1 \times 1$  の行列だと考えると, 内積をトレースと捉えることができトレースの性質から

$$\mathbf{x}^\top \mathbf{y} = \text{tr}(\mathbf{x}^\top \mathbf{y}) = \text{tr}(\mathbf{y} \mathbf{x}^\top) \quad (1.29)$$

が成り立つ.

**定義 1.9.** フロベニウスノルム

$N \times P$  の行列  $\mathbf{X}$  に対して,

$$\|\mathbf{X}\|_F = \sqrt{\text{tr}(\mathbf{X}^\top \mathbf{X})} = \sqrt{\sum_{n=1}^N \sum_{p=1}^P |x_{np}|^2} \quad (1.30)$$

を  $\mathbf{X}$  のフロベニウスノルムと呼ぶ.

さて、主成分分析のパラメータを交互最小二乗法によって推定してみる。まず  $\mathbf{A}$  を固定し、 $\mathbf{F}$  を更新する。 $\mathbf{F}$  の更新式は、目的関数  $f_{\text{PCA}}(\mathbf{F}|\mathbf{X}, \mathbf{A})$  の最小化によって得られる。ただし  $\mathbf{A}$  は観測されていないが、既知として与える必要があるため、乱数によって初期値を発生させる。ただし  $\mathbf{A}$  には直交制約があるため、その制約を満たすように乱数を発生させることに注意が必要がある。

ベクトル偏微分を用いる場合は、 $\mathbf{F}$  の  $n$  行を更新するために

$$\begin{aligned} f_{\text{PCA}}(\mathbf{F}|\mathbf{X}, \mathbf{A}) &= \|\mathbf{X} - \mathbf{F}\mathbf{A}^\top\|_F^2 \\ &= \sum_{n=1}^N f_{\text{PCA}}(\mathbf{f}_n|\mathbf{x}_n, \mathbf{A}) \\ &= \sum_{n=1}^N \|\mathbf{x}_n - \mathbf{A}\mathbf{f}_n\|_2^2 \end{aligned} \quad (1.31)$$

を最小化するような  $\mathbf{f}_n$  を求める。したがって方程式

$$\begin{aligned} \frac{\partial f_{\text{PCA}}(\mathbf{f}_n|\mathbf{x}_n, \mathbf{A})}{\partial \mathbf{f}_n} &= \frac{\partial}{\partial \mathbf{f}_n} (\mathbf{x}_n^\top \mathbf{x}_n) - 2 \frac{\partial}{\partial \mathbf{f}_n} (\mathbf{x}_n^\top \mathbf{A} \mathbf{f}_n) + \frac{\partial}{\partial \mathbf{f}_n} (\mathbf{f}_n^\top \mathbf{A}^\top \mathbf{A} \mathbf{f}_n) \\ &= -2\mathbf{A}^\top \mathbf{x}_n + 2\mathbf{f}_n = \mathbf{0}_M \end{aligned} \quad (1.32)$$

を得る。よって、

$$\hat{\mathbf{f}}_n = \mathbf{A}^\top \mathbf{x}_n \quad (1.33)$$

となる。ただし (1.32) 式では、 $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_M$  を用いた。

ベクトル偏微分ではなく、トレースの偏微分を用いるとより簡潔に書くことができる。方程式

$$\begin{aligned} \frac{\partial}{\partial \mathbf{F}} f_{\text{PCA}}(\mathbf{F}|\mathbf{X}, \mathbf{A}) &= \frac{\partial}{\partial \mathbf{F}} \|\mathbf{X} - \mathbf{F}\mathbf{A}^\top\|_F^2 \\ &= \frac{\partial}{\partial \mathbf{F}} \{\text{tr}(\mathbf{X}^\top \mathbf{X}) - 2\text{tr}(\mathbf{X}^\top \mathbf{F} \mathbf{A}^\top) + \text{tr}(\mathbf{A} \mathbf{F}^\top \mathbf{F} \mathbf{A}^\top)\} \\ &= \frac{\partial}{\partial \mathbf{F}} \{-2\text{tr}(\mathbf{A}^\top \mathbf{X}^\top \mathbf{F}) + \text{tr}(\mathbf{A}^\top \mathbf{A} \mathbf{F}^\top \mathbf{F})\} \\ &= -2\mathbf{X}\mathbf{A} + 2\mathbf{F} = \mathbf{0}_M \end{aligned} \quad (1.34)$$

から、行列形式の更新式

$$\hat{\mathbf{F}} = \mathbf{X}\mathbf{A} \quad (1.35)$$

を得る。4行目の等式は、トレース  $\text{tr}(\mathbf{A}^\top \mathbf{X}^\top \mathbf{F})$  の  $\mathbf{F}$  についての偏微分

$$\frac{\partial \text{tr}(\mathbf{A}^\top \mathbf{X}^\top \mathbf{F})}{\partial \mathbf{F}} = \mathbf{XA} \quad (1.36)$$

およびトレース  $\text{tr}(\mathbf{F}^\top \mathbf{F})$  の  $\mathbf{F}$  についての偏微分

$$\frac{\partial \text{tr}(\mathbf{F}^\top \mathbf{F})}{\partial \mathbf{F}} = 2\mathbf{F} \quad (1.37)$$

を用いている。

**定義 1.10.** トレースの偏微分

非対称行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$  と行列  $\mathbf{Y} \in \mathbb{R}^{P \times N}$  があるとする。このとき  $\text{tr}(\mathbf{XY})$  の  $\mathbf{X}$  についての偏微分は

$$\frac{\partial \text{tr}(\mathbf{XY})}{\partial \mathbf{X}} = \mathbf{Y}^\top \quad (1.38)$$

で与えられる。

また、行列  $\mathbf{Z} \in \mathbb{R}^{N \times P}$  があるとして

$$\frac{\partial \text{tr}(\mathbf{X}^\top \mathbf{Z})}{\partial \mathbf{X}} = \mathbf{Z} \quad (1.39)$$

である。

**系 1.3.** 非対称行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$  と行列  $\mathbf{Y} \in \mathbb{R}^{N \times N}$  があるとする。このとき、

$$\begin{aligned} \frac{\partial \text{tr}(\mathbf{X}^\top \mathbf{Y} \mathbf{X})}{\partial \mathbf{X}} &= \frac{\partial \text{tr}(\mathbf{X}^\top (\mathbf{Y} \mathbf{X}))}{\partial \mathbf{X}} + \frac{\partial \text{tr}((\mathbf{X}^\top \mathbf{Y}) \mathbf{X})}{\partial \mathbf{X}} \\ &= \mathbf{Y} \mathbf{X} + \mathbf{Y}^\top \mathbf{X} \end{aligned} \quad (1.40)$$

さて、 $\mathbf{F}$  についての最適化は偏微分によって達成されることがわかった。続いて  $\mathbf{A}$  についての最適化を行う。ただし微分を用いた単純な最適化は使用できない。なぜなら、 $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_M$  の正規直交制約を満たしたうえで、最適化を行わなければならないためである。

$\mathbf{F}$  を固定したときの最小化すべき目的関数は

$$\begin{aligned}
 f_{PCA}(\mathbf{A}|\mathbf{X}, \mathbf{F}) &= \|\mathbf{X} - \mathbf{F}\mathbf{A}^\top\|_F^2 \\
 &= \text{tr}((\mathbf{X} - \mathbf{F}\mathbf{A}^\top)^\top(\mathbf{X} - \mathbf{F}\mathbf{A}^\top)) \\
 &= \text{tr}(\mathbf{X}^\top \mathbf{X}) - 2\text{tr}(\mathbf{X}^\top \mathbf{F}\mathbf{A}^\top) + \text{tr}(\mathbf{A}\mathbf{F}^\top \mathbf{F}\mathbf{A}^\top) \\
 &= \text{tr}(\mathbf{X}^\top \mathbf{X}) - 2\text{tr}(\mathbf{X}^\top \mathbf{F}\mathbf{A}^\top) + \text{tr}(\mathbf{A}^\top \mathbf{A}\mathbf{F}^\top \mathbf{F}) \\
 &= \text{const}_{\mathbf{A}} - 2\text{tr}(\mathbf{X}^\top \mathbf{F}\mathbf{A}^\top)
 \end{aligned} \tag{1.41}$$

と書ける。途中の等式には  $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_M$  を用いた。また  $\text{const}_{\mathbf{A}}$  は  $\mathbf{A}$  に関係のない定数で、実際  $\text{tr}(\mathbf{X}^\top \mathbf{X})$  と  $\text{tr}(\mathbf{F}^\top \mathbf{F})$  は  $\mathbf{A}$  の更新には関係しない。

(1.41) 式は、 $\mathbf{A}$  についての最適化がトレースの最大化を意味することを示している。そのため、トレース最大化のためのテクニックが必要となる。もし  $P = M$  とすると  $\mathbf{A}$  は正方行列かつ正規直交行列である。この  $\mathbf{A}^\top \mathbf{A} = \mathbf{A}\mathbf{A}^\top = \mathbf{I}_P$  という条件のもとで目的関数 (1.41) 式を最適化することを直交プロクラステス問題 (Orthogonal Procrustes Problem) と呼ぶことがある (Schönemann, 1966)。直交プロクラステス問題から  $\mathbf{A}$  の更新式

$$\hat{\mathbf{A}} = \mathbf{KL}^\top \tag{1.42}$$

が得られる。ただし  $\mathbf{K} \in \mathbb{R}^{P \times M}$  と  $\mathbf{L} \in \mathbb{R}^{M \times M}$  は、 $\mathbf{X}^\top \mathbf{F}$  の固有値分解

$$\mathbf{X}^\top \mathbf{F} = \mathbf{K} \Delta \mathbf{L}^\top \tag{1.43}$$

によって得られる左、右特異行列であり正規直交性

$$\mathbf{K}^\top \mathbf{K} = \mathbf{L}^\top \mathbf{L} = \mathbf{L} \mathbf{L}^\top = \mathbf{I}_M \tag{1.44}$$

を満たす。

**定理 1.3. 直交プロクラステス問題**

行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$ ,  $\mathbf{F} \in \mathbb{R}^{N \times P}$ ,  $\mathbf{A} \in \mathbb{R}^{P \times P}$  が与えられているとする. そして正規直交制約  $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_P$  を満たす. このとき目的関数

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{F}\mathbf{A}^\top\|_F^2 \text{ s.t. } \mathbf{A}^\top \mathbf{A} = \mathbf{I}_P$$

を最小化する  $\mathbf{A}$  は

$$\mathbf{A} = \mathbf{KL}^\top \quad (1.45)$$

によって得られる. ただし  $\mathbf{K}, \mathbf{L}$  は特異値分解  $\mathbf{X}^\top \mathbf{F} = \mathbf{K}\Delta\mathbf{L}^\top$  によって得られた左特異行列  $\mathbf{K} \in \mathbb{R}^{P \times P}$  と右特異行列  $\mathbf{L} \in \mathbb{R}^{P \times P}$  である.

証明.

$$\begin{aligned} \min_{\mathbf{A}} \|\mathbf{X} - \mathbf{F}\mathbf{A}^\top\|_F^2 &= \text{tr}((\mathbf{X} - \mathbf{F}\mathbf{A}^\top)^\top(\mathbf{X} - \mathbf{F}\mathbf{A}^\top)) \\ &= \text{const}_{\mathbf{A}} - 2\text{tr}(\mathbf{X}^\top \mathbf{F}\mathbf{A}^\top) \end{aligned} \quad (1.46)$$

となることから, 目的関数の最小化問題はトレースの最大化問題に帰着する. ここで  $\mathbf{X}^\top \mathbf{F}$  の特異値分解とトレースの性質から,

$$\text{tr}(\mathbf{X}^\top \mathbf{F}\mathbf{A}^\top) = \text{tr}(\mathbf{K}\Delta\mathbf{L}^\top \mathbf{A}^\top) = \text{tr}(\mathbf{A}\mathbf{L}\Delta\mathbf{K}^\top) = \text{tr}(\mathbf{K}^\top \mathbf{A}\mathbf{L}\Delta) = \mathbf{G}\Delta \quad (1.47)$$

を得る. ただし  $\mathbf{G} = \mathbf{K}^\top \mathbf{A}\mathbf{L}$  とした. ここで

$$\mathbf{L}^\top \mathbf{A}^\top \mathbf{K} \mathbf{K}^\top \mathbf{A} \mathbf{L} = \mathbf{I}_P \quad (1.48)$$

であるから,  $\mathbf{G}$  は正規直交行列である. 正規直交行列かつ正方な行列  $\mathbf{G}$  の対角要素は明らかに 1 以下であるから,

$$\text{tr}(\mathbf{G}\Delta) = \sum_{p=1}^P g_{pp}\delta_p \leq \sum_{p=1}^P \delta_p \quad (1.49)$$

となる. (1.49) 式が最大値になるときは,  $\mathbf{G}$  の各対角要素が 1 になるとき, つまり  $\mathbf{G} = \mathbf{I}_P$  の場合である. よって

$$\mathbf{I}_P = \mathbf{K}^\top \mathbf{A}\mathbf{L} \Leftrightarrow \mathbf{A} = \mathbf{KL}^\top \quad (1.50)$$

ここで問題がひとつある。主成分分析のパラメータ行列  $\mathbf{A}$  は正方行列ではなく矩形行列であることが多い。つまり,  $\mathbf{A} \in \mathbb{R}^{P \times M} (M \leq P)$ 。このとき, 直交プロクラステス問題の証明は用いることができない。そこで von Neumann の定理 (von Neumann, 1937) および Kristof の定理 (Kristof, 1970) を一般化した ten Berge の定理 (ten Berge, 1983; 1993) によって最適化を行う。驚くべきことに, 後述の通り ten Berge の定理により示される  $\mathbf{A}$  の更新式と直交プロクラステス問題における  $\mathbf{A}$  の更新式は同一である。すなわち,  $\mathbf{A}$  が正方行列であるか矩形行列であるかを問わず

$$\hat{\mathbf{A}} = \mathbf{KL}^\top$$

で更新される。ten Berge の定理には幾つか定義と補題が必要となる。

**定義 1.11.** 部分正規直交行列

正規直交行列の部分行列を部分正規直交行列 (sub-orthonormal matrix) と呼ぶ。

**補題 1.1.** 部分正規直交行列の積

$\mathbf{X} \in \mathbb{R}^{N \times P}$  と  $\mathbf{Y} \in \mathbb{R}^{P \times M}$  が, それぞれ部分正規直交行列とする。このとき, それらの積  $\mathbf{XY}$  も部分正規直交行列である。

**補題 1.2.** 部分正規直交行列の特異値

部分正規直交行列  $\mathbf{X} \in \mathbb{R}^{N \times P} (P \leq N)$  の特異値分解  $\mathbf{X} = \mathbf{K}\Delta\mathbf{L}^\top$  によって得られた各特異値  $\delta_p$  は 0 から 1, すなわち  $p = 1, \dots, P$  について  $0 \leq \delta_p \leq 1$  である。

**補題 1.3.** 部分正規直交行列と対角行列の積のトレース最大化

$\mathbf{G} \in \mathbb{R}^{P \times P}$  を部分正規直交行列とし,  $\mathbf{D} \in \mathbb{R}^{P \times P}$  を対角行列とする。このとき不等式

$$\text{tr}(\mathbf{GD}) \leq \text{tr}(\mathbf{D})$$

が成り立つ。また等式は  $\mathbf{G} = \mathbf{I}_P$  のとき成り立つ。

証明。補題 1.2 の部分正規直交行列の特異値の性質から  $\mathbf{G}$  の対角要素は全て 1 以下と言える。よって,

$$\text{tr}(\mathbf{GD}) = \sum_{p=1}^P g_{pp} d_{pp} \leq \sum_{p=1}^P d_{pp} = \text{tr}(\mathbf{D})$$

であり,  $\text{tr}(\mathbf{GD}) = \text{tr}(\mathbf{D})$  は  $\mathbf{G} = \mathbf{I}_P$  のとき成り立つ。  $\square$

**定理 1.4.** ten Berge の定理の特殊例

いま行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$ ,  $\mathbf{F} \in \mathbb{R}^{N \times M}$  が与えられているとする. そして直交制約  $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_M$  を満たす. このときトレース

$$\text{tr}(\mathbf{X}^\top \mathbf{F} \mathbf{A}^\top)$$

を最大化する  $\mathbf{A}$  は

$$\hat{\mathbf{A}} = \mathbf{K} \mathbf{L}^\top$$

によって得られる. ただし  $\mathbf{K}$  と  $\mathbf{L}$  は特異値分解  $\mathbf{X}^\top \mathbf{F} = \mathbf{K} \Delta \mathbf{L}^\top$  によって得られた左特異行列  $\mathbf{K} \in \mathbb{R}^{P \times M}$  と右特異行列  $\mathbf{L} \in \mathbb{R}^{M \times M}$  である.

証明.  $\mathbf{X}$  に特異値分解を行うと

$$\begin{aligned} \text{tr}(\mathbf{X}^\top \mathbf{F} \mathbf{A}^\top) &= \text{tr}(\mathbf{K} \Delta \mathbf{L}^\top \mathbf{A}^\top) \\ &= \text{tr}((\mathbf{L}^\top \mathbf{A}^\top \mathbf{K}) \Delta) \end{aligned} \quad (1.51)$$

を得る. ここで  $\mathbf{L}, \mathbf{A}, \mathbf{K}$  全てが部分正規直交行列であるため, その積である  $\mathbf{L}^\top \mathbf{A}^\top \mathbf{K}$  も部分正規直交行列である. よって補題 1.3 から,

$$\text{tr}(\mathbf{L}^\top \mathbf{A}^\top \mathbf{K} \Delta) \leq \text{tr}(\Delta)$$

かつ等号成立は  $\mathbf{L}^\top \mathbf{A}^\top \mathbf{K} = \mathbf{I}$  のときである. よって,  $\mathbf{K}$  と  $\mathbf{L}$  の列正規直交性から

$$\mathbf{A} = \mathbf{K} \mathbf{L}^\top$$

□

ten Berge の定理の特殊形である定理 1.4 から,  $M \leq P$  の場合でも,  $\mathbf{A}$  の更新式は

$$\hat{\mathbf{A}} = \mathbf{K} \mathbf{L}^\top \quad (1.52)$$

である. まとめると, (1.35) 式と (1.52) 式を収束するまで繰り返すことで, 最適なパラメータ  $\mathbf{F}$  と  $\mathbf{A}$  を得る.

ところが, 実際には主成分分析は反復計算を必要とせず, (1.35) 式と (1.52) 式を 1 回ずつ計算すれば収束する. そして偏微分やトレース最大化を用いず, 特異値分解のみでパラ

メータを推定できる。

今データ行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$  ( $P \leq N$ ) があるとする。 $\mathbf{X}$  の特異値分解は

$$\mathbf{X} = \mathbf{K}\Delta\mathbf{L}^\top \quad (1.53)$$

と書けて、 $N \times P$  の行列  $\mathbf{K} \in \mathbb{R}^{N \times P}$  は列正規直交性を満たす。つまり

$$\mathbf{K}^\top \mathbf{K} = \mathbf{I}_P$$

である。 $\mathbf{K}$  は列ごとに正規直交性が満たされたので、 $\mathbf{K}$  の左から  $M$  列のみを  $\mathbf{K}_1$ 、残りの  $P - M$  列を  $\mathbf{K}_2$  とすると

$$\mathbf{K} = [\mathbf{K}_1, \mathbf{K}_2]$$

であり、 $\mathbf{K}_1^\top \mathbf{K}_1 = \mathbf{I}_M$  かつ  $\mathbf{K}_2^\top \mathbf{K}_2 = \mathbf{I}_{P-M}$  である。つまり列について分割しても、列に関しては正規直交性は保たれる。 $P \times P$  の行列  $\mathbf{L} \in \mathbb{R}^{P \times P}$  においても同様に  $\mathbf{L} = [\mathbf{L}_1, \mathbf{L}_2]$ 、 $\mathbf{L}_1 \in \mathbb{R}^{P \times M}$  とすると  $\mathbf{L}_1^\top \mathbf{L}_1 = \mathbf{I}_M$  かつ  $\mathbf{L}_2^\top \mathbf{L}_2 = \mathbf{I}_{P-M}$  である。そして特異値分解 (1.53) 式は

$$\begin{aligned} \mathbf{X} &= \mathbf{K}\Delta\mathbf{L}^\top = [\mathbf{K}_1, \mathbf{K}_2] \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix} [\mathbf{L}_1, \mathbf{L}_2]^\top \\ &= \mathbf{K}_1 \Delta_1 \mathbf{L}_1^\top + \mathbf{K}_2 \Delta_2 \mathbf{L}_2^\top \\ &= \mathbf{X}_1 + \mathbf{X}_2 \end{aligned}$$

のように書くこともできる。ただし  $\Delta_1$  は  $M \times M$  の対角行列で、 $\Delta_2$  は  $(P - M) \times (P - M)$  の対角行列。また足し算が成立していることから明らかのように、 $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2$  は全て  $N \times P$  の行列である。いま、 $\Delta$  の対角要素が降順 ( $\delta_1 \geq \delta_2 \geq \dots \geq \delta_P \geq 0$ ) とする。すると、行列のランクが  $M$  以下という条件を満たした上で  $\mathbf{X}$  を最もよく近似する、つまり

$$\|\mathbf{X} - \mathbf{Y}\|_F^2 \text{ s.t. } \text{rank}(\mathbf{Y}) \leq M \quad (1.54)$$

を最小にするような行列  $\mathbf{Y}$  は

$$\mathbf{Y} = \mathbf{K}_1 \Delta_1 \mathbf{L}_1^\top \quad (1.55)$$

である (Eckart & Young, 1936)。行列のサイズこそ同じものの、その行列のランクは  $\text{rank}(\mathbf{X}) = P$  よりも低ランクの  $\text{rank}(\mathbf{Y}) \leq M$  であるため、(1.54) 式の問題は低ランク

近似と呼ばれる。また、論文著者の名前などから(1.54)式の問題と(1.55)式を Eckart-Young 定理(Eckart-Young-Mirsky 定理)とも呼ぶ。ちなみに  $M$  個目までの特異値が全て正( $\delta_1 \geq \dots \geq \delta_M > 0$ )であるとき  $\text{rank}(\mathbf{Y}) = M$  である。

ここで、

$$\mathbf{K}_1 \Delta_1 = \mathbf{F} \quad (1.56)$$

$$\mathbf{L}_1 = \mathbf{A} \quad (1.57)$$

$$\mathbf{K}_2 \Delta_2 \mathbf{L}_2^\top = \mathbf{E} \quad (1.58)$$

とすると、

$$\|\mathbf{X} - \mathbf{K}_1 \Delta_1 \mathbf{L}_1^\top\|_F^2 = \|\mathbf{X} - \mathbf{F} \mathbf{A}^\top\|_F^2 = \|\mathbf{E}\|_F^2$$

と書け、低ランク近似の問題と主成分分析を同一視することができる。すなわち、データ行列の特異値分解によって主成分分析のパラメータを求めることができる。

### $k$ -means クラスタリング

それでは、どのような問題では反復計算が必要となるのであろうか。例えば  $k$ -means クラスタリングは反復計算が必要な代表例である。

$k$ -means クラスタリングの目的関数は

$$\sum_{n=1}^N \sum_{k=1}^K g_{nk} \|\mathbf{x}_n - \mathbf{c}_k\|_2^2 \quad (1.59)$$

のように書く。ただし  $\mathbf{x}_n, \mathbf{c}_k \in \mathbb{R}^P$  であり、 $N$  は個体数、 $P$  は変数の数とする。そして  $\mathbf{x}_n$  は個体  $n$  のデータベクトル、 $\mathbf{c}_k$  はクラスター  $k$  のセントロイドベクトル(代表ベクトル)である。まず、 $\|\mathbf{x}_n - \mathbf{c}_k\|_2^2$  の意味について考えてみる。これは(1.7)式を見ればわかるように、 $\mathbf{x}_n$  と  $\mathbf{c}_k$  のユークリッド距離である。

それでは  $g_{nk}$  は何を表しているのだろうか。これは個体が  $n$  に所属するときのみ値が 1 となり、 $k$  に所属しない場合は 0 となる関数である。

$$g_{nk} = \begin{cases} 1 & \text{個体 } n \text{ がクラスター } k \text{ に所属する} \\ 0 & \text{個体 } n \text{ がクラスター } k \text{ に所属しない} \end{cases}$$

所属を表す関数のため,  $g_{nk}$  をメンバーシップ関数と呼ぶ. あるいは指示関数と呼ぶ場合もある. また,  $k$ -means クラスタリングでは1つの個体が複数のクラスターに所属することは許されない. よって, 任意の  $n$  について制約

$$\sum_{k=1}^K g_{nk} = 1 \quad (1.60)$$

が成り立つ.

具体例を考えてみよう. いま 100 個体 2 変数のデータが得られたとし, さらに 3 つのクラスターのセントロイドからなる散布図を図 1.5 に描写する.

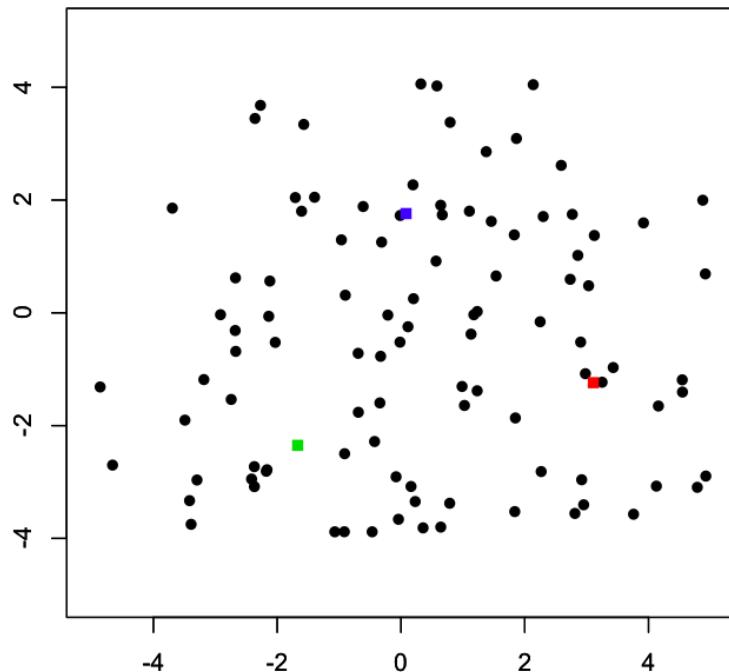


図 1.5: 100 個体 2 変数のデータと  $K = 3$  としたときのセントロイドの散布図

さらに図 1.6 のように, ある個体  $x_{50}$  のみに注目してみる.

この個体  $x_{50}$  とそれぞれのクラスターの距離が

$$(-2.03 - 3.11)^2 + (-0.52 - (-1.24))^2 \simeq 5.43$$

$$(-2.03 - 0.08)^2 + (-0.52 - 1.76)^2 \simeq 3.11$$

$$(-2.03 - (-1.67))^2 + (-0.52 - (-2.35))^2 \simeq 1.87$$

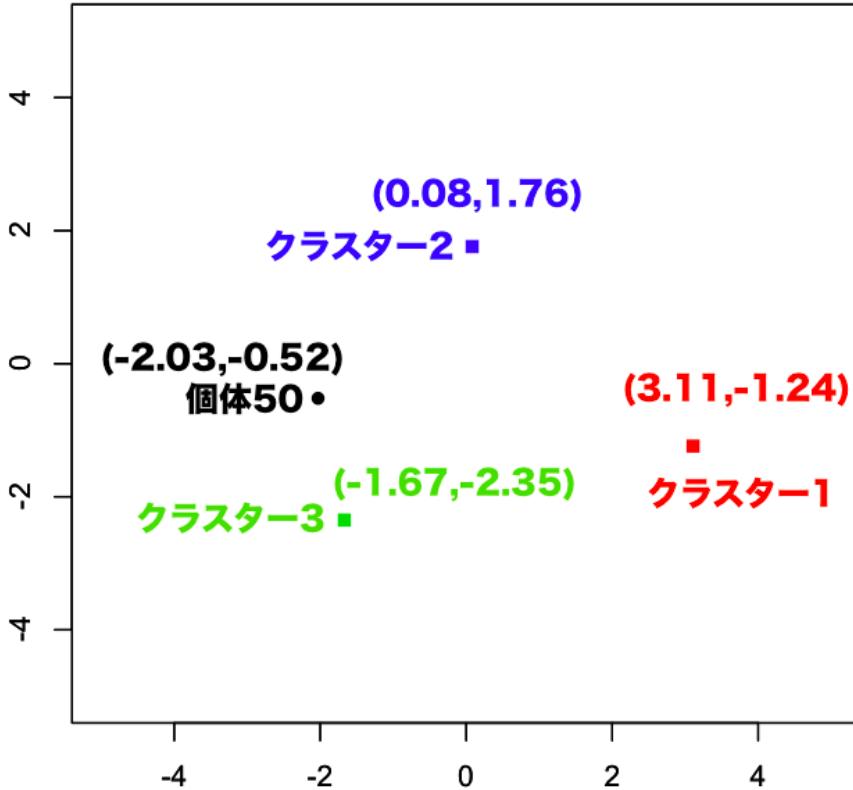


図 1.6: ある個体とセントロイドの散布図

となった. よって個体 50 と一番距離の近いクラスター 3 を個体 50 が所属するクラスターと見なし,  $g_{50,3} = 1$  となる. そして  $g_{50,1}, g_{50,2} = 0$  となることから, それらの距離は無視することとなる. よって,

$$\sum_{k=1}^K g_{50,k} \|\mathbf{x}_{50} - \mathbf{c}_k\|_2^2 = \|\mathbf{x}_{50} - \mathbf{c}_3\|_2^2$$

となる.

さて,  $k$ -means クラスタリングのパラメータ推定は, 交互最小二乗法によって行う.  $\{g_{n1}, \dots, g_{nK}\}$  は場合分けによって求めるしかない. すなわち

$$\|\mathbf{x}_n - \mathbf{c}_l\|_2^2$$

が最小となる  $l (l = 1, \dots, K)$  を探索し  $g_{nl} = 1$  とし, それ以外のメンバーシップ関数を 0

とする。つまり,  $g_{nk}$  の更新式は

$$\hat{g}_{nk} = \begin{cases} 1 & (k = \operatorname{argmin}_{1 \leq l \leq K} \|\mathbf{x}_n - \mathbf{c}_l\|_2^2) \\ 0 & (\text{それ以外}) \end{cases} \quad (1.61)$$

によって得られる。

次に  $\mathbf{c}_k$  の更新を行う。 $g_{nk}$  を固定し  $\mathbf{c}_k$  についてベクトルの偏微分を行うと, 方程式

$$\begin{aligned} & \frac{\partial}{\partial \mathbf{c}_k} \sum_{n=1}^N g_{nk} \|\mathbf{x}_n - \mathbf{c}_k\|_2^2 \\ &= \frac{\partial}{\partial \mathbf{c}_k} \sum_{n=1}^N g_{nk} (\mathbf{x}_n^\top \mathbf{x}_n - 2\mathbf{x}_n^\top \mathbf{c}_k + \mathbf{c}_k^\top \mathbf{c}_k) \\ &= \sum_{n=1}^N g_{nk} (-2\mathbf{x}_n + 2\mathbf{c}_k) \\ &= -2 \left( \sum_{n=1}^N g_{nk} \mathbf{x}_n \right) + 2\mathbf{c}_k \left( \sum_{n=1}^N g_{nk} \right) = \mathbf{0}_P \end{aligned} \quad (1.62)$$

を得る。したがって,  $\mathbf{c}_k$  の更新式は

$$\hat{\mathbf{c}}_k = \frac{\sum_{n=1}^N g_{nk} \mathbf{x}_n}{\sum_{n=1}^N g_{nk}} \quad (1.63)$$

である。

$k$ -means クラスタリングの目的関数は行列を用いて表現することも可能である。

$k$ -means クラスタリング

$\|\mathbf{X} - \mathbf{GC}\|_F^2 : k$ -means クラスタリングの目的関数

$\mathbf{X} : N$  個体  $\times P$  変数のデータ行列

$\mathbf{G} :$ 各個体の所属を示す  $N \times K$  のメンバーシップ行列

$\mathbf{C} :$ 各クラスターの中心点を表す  $K \times P$  のセントロイド行列

データ行列を  $N \times P$  の行列  $\mathbf{X}$  とすると,  $k$ -means クラスタリングの目的関数は

$$\|\mathbf{X} - \mathbf{GC}\|_F^2 \quad (1.64)$$

と書ける。ただし  $\mathbf{X} \in \mathbb{R}^{N \times P}$ ,  $\mathbf{G} \in \mathbb{R}^{N \times K}$ ,  $\mathbf{C} \in \mathbb{R}^{K \times P}$  とし,  $K$  はクラスター数とする。 $\mathbf{G}$  は  $g_{nk}$  を要素とする行列で、メンバーシップ行列あるいは指示行列と呼ばれる。具体的には  $\mathbf{G} \in \mathbb{R}^{7 \times 3}$  とすると

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

のような形になる。メンバーシップ関数の制約 (1.60) 式は,

$$\mathbf{G}\mathbf{1}_K = \mathbf{1}_N \quad (1.65)$$

であるから  $\mathbf{G}$  は制約を満たしていることがわかる。

ところで、得られた解が真の解である保証が  $k$ -means クラスタリングに関してはない。これは目的関数の非凸性と関係しており (Jain & Kar, 2017), 実際には複数の初期値から解を求め、それらの中で目的関数が最も小さいものを採用する。これはランダムスタートなどと呼ぶ。

## 1.3 確率分布

### 1.3.1 正規分布

多変量解析において最も活用されている分布の1つは、正規分布である。 $x \in \mathbb{R}$  が正規分布にしたがうとき確率密度関数は

$$p(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\} \quad (1.66)$$

である。ただし  $\mu$  は平均,  $\sigma^2$  は分散を表す。そして  $\mathbf{x} \in \mathbb{R}^P$  が多変量正規分布にしたがうとき確率密度関数は

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{P/2}} \frac{1}{\det(\boldsymbol{\Sigma})^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\} \quad (1.67)$$

である。ただし  $\boldsymbol{\mu} \in \mathbb{R}^P$  は平均ベクトル,  $\boldsymbol{\Sigma} \in \mathbb{R}^{P \times P}$  は共分散行列を表す。ここで,  $\mathbf{x}_1, \dots, \mathbf{x}_N$  が独立かつ同一の多変量正規分布に従うとする。このとき  $\mathbf{x}_1, \dots, \mathbf{x}_N$  の同時分布  $p(\mathbf{x}_1, \dots, \mathbf{x}_N)$  の確率密度関数は、その独立性から各  $\mathbf{x}_n$  の確率密度関数の積となり

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{NP/2}} \frac{1}{\det(\boldsymbol{\Sigma})^{N/2}} \exp \left\{ -\frac{1}{2} \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) \right\} \quad (1.68)$$

である。この式は  $\mathbf{S} = \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^\top$  と置くことで、

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{NP/2}} \frac{1}{\det(\boldsymbol{\Sigma})^{N/2}} \exp \left\{ -\frac{1}{2} \text{tr}(\mathbf{S} \boldsymbol{\Sigma}^{-1}) \right\} \quad (1.69)$$

のように書くことができる。これには (1.29) 式を用いて、ベクトルの内積をトレースに書き換えている。この式はグラフィカルモデルなどのように  $\boldsymbol{\mu}$  には重きをおかない、あるいは  $\boldsymbol{\mu} = \mathbf{0}_P$  を仮定し、むしろ  $\boldsymbol{\Sigma}$  の構造に关心がある場合などに用いられることが多い。

### 1.3.2 条件付き正規分布

2つの確率変数  $\mathbf{a} \in \mathbb{R}^P, \mathbf{b} \in \mathbb{R}^M$  があるとし、それぞれ異なる正規分布,  $\mathbf{a} \sim N(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$ ,  $\mathbf{b} \sim N(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$  に従うとする。すなわち

$$\mathbb{E}[\mathbf{a}] = \boldsymbol{\mu}_a$$

$$\mathbb{E}[\mathbf{b}] = \boldsymbol{\mu}_b$$

である。ここで  $\mathbf{b}$  が与えられた場合の  $\mathbf{a}$ 、つまり条件付き分布  $(\mathbf{a}|\mathbf{b})$  を考える。重要な事実だが、 $\mathbf{a}$  と  $\mathbf{b}$  それが正規分布に従うとき、条件付き分布もまた正規分布に従う。

条件付き分布のパラメータ推定を行うためには、まず  $\mathbf{c} = [\mathbf{a}, \mathbf{b}]^\top \in \mathbb{R}^{P+M}$  を考え、 $\mathbf{c} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  とする。このとき  $\mathbf{c}$  の期待値は

$$\mathbb{E}[\mathbf{c}] = [\mathbb{E}[\mathbf{a}], \mathbb{E}[\mathbf{b}]] \quad (1.70)$$

つまり

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_a, \boldsymbol{\mu}_b]$$

である.  $\mathbf{c}$  の分散は

$$\begin{aligned}\mathbb{V}\text{ar}[\mathbf{c}] &= \begin{bmatrix} \mathbb{V}\text{ar}[\mathbf{a}] & \mathbb{C}\text{ov}[\mathbf{a}, \mathbf{b}] \\ \mathbb{C}\text{ov}[\mathbf{b}, \mathbf{a}] & \mathbb{V}\text{ar}[\mathbf{b}] \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{a} - \mathbb{E}[\mathbf{a}]) (\mathbf{a} - \mathbb{E}[\mathbf{a}])^\top & (\mathbf{a} - \mathbb{E}[\mathbf{a}]) (\mathbf{b} - \mathbb{E}[\mathbf{b}])^\top \\ (\mathbf{b} - \mathbb{E}[\mathbf{b}]) (\mathbf{a} - \mathbb{E}[\mathbf{a}])^\top & (\mathbf{b} - \mathbb{E}[\mathbf{b}]) (\mathbf{b} - \mathbb{E}[\mathbf{b}])^\top \end{bmatrix} \quad (1.71)\end{aligned}$$

である. つまり

$$\begin{aligned}\boldsymbol{\Sigma} &= \begin{bmatrix} (\mathbf{a} - \boldsymbol{\mu}_a) (\mathbf{a} - \boldsymbol{\mu}_a)^\top & (\mathbf{a} - \boldsymbol{\mu}_a) (\mathbf{b} - \boldsymbol{\mu}_b)^\top \\ (\mathbf{b} - \boldsymbol{\mu}_b) (\mathbf{a} - \boldsymbol{\mu}_b)^\top & (\mathbf{b} - \boldsymbol{\mu}_b) (\mathbf{b} - \boldsymbol{\mu}_b)^\top \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ab} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}\end{aligned}$$

ここで, 共分散行列の逆行列を  $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$  とし  $\boldsymbol{\Omega}$  も  $\boldsymbol{\Sigma}$  と同様に分割を行う. つまり

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{aa} & \boldsymbol{\Omega}_{ab} \\ \boldsymbol{\Omega}_{ba} & \boldsymbol{\Omega}_{bb} \end{bmatrix}$$

とする. ただし  $\boldsymbol{\Sigma}_{aa}^{-1} \neq \boldsymbol{\Omega}_{aa}$  であることに注意が必要である. すると

$$\begin{aligned}& -\frac{1}{2}(\mathbf{c} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{c} - \boldsymbol{\mu}) \\ &= -\frac{1}{2}(\mathbf{a} - \boldsymbol{\mu}_a)^\top \boldsymbol{\Omega}_{aa}(\mathbf{a} - \boldsymbol{\mu}_a) - \frac{1}{2}(\mathbf{a} - \boldsymbol{\mu}_a)^\top \boldsymbol{\Omega}_{ab}(\mathbf{b} - \boldsymbol{\mu}_b) \\ &\quad - \frac{1}{2}(\mathbf{b} - \boldsymbol{\mu}_b)^\top \boldsymbol{\Omega}_{ba}(\mathbf{a} - \boldsymbol{\mu}_a) - \frac{1}{2}(\mathbf{b} - \boldsymbol{\mu}_b)^\top \boldsymbol{\Omega}_{bb}(\mathbf{b} - \boldsymbol{\mu}_b) \quad (1.72)\end{aligned}$$

のように分割して書くこともできる.

条件付き正規分布のパラメータの推定には, 平方完成と呼ばれるテクニックを用いることができる (Bishop, 2006). 一般に  $\mathbf{x} \in \mathbb{R}^P$  が正規分布  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  に従うとき, その確率密度関数の指数部分は

$$-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = -\frac{1}{2}\mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\mathbf{x} + \mathbf{x}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \text{const}_{\mathbf{x}} \quad (1.73)$$

と書くことができる. ただし  $\text{const}_{\mathbf{x}}$  は  $\mathbf{x}$  に関係しない項を表す. 正規分布の確率密度関数の指数部を上式のような式で表現することができれば,  $\mathbf{x}$  の二次の項の係数行列が共分散行列の逆行列  $\boldsymbol{\Sigma}^{-1}$  に等しく,  $\mathbf{x}$  の一次の項の係数ベクトルが  $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$  に等しいことから,  $\boldsymbol{\mu}$  を得ることができる.

では、この平方完成を条件付き分布  $(\mathbf{a}|\mathbf{b}) \sim N(\boldsymbol{\mu}_{\mathbf{a}|\mathbf{b}}, \boldsymbol{\Sigma}_{\mathbf{a}|\mathbf{b}})$  に適用することで、 $\boldsymbol{\mu}_{\mathbf{a}|\mathbf{b}}$ ,  $\boldsymbol{\Sigma}_{\mathbf{a}|\mathbf{b}}$  を計算してみる。

実際に平方完成を用いると、 $\mathbf{a}$  についての二次の項は

$$-\frac{1}{2}\mathbf{a}^\top \boldsymbol{\Omega}_{\mathbf{aa}} \mathbf{a} \quad (1.74)$$

となる。したがって、

$$\boldsymbol{\Sigma}_{\mathbf{a}|\mathbf{b}} = \boldsymbol{\Omega}_{\mathbf{aa}}^{-1} \quad (1.75)$$

となる。さらに一次の項も

$$\mathbf{a}^\top \boldsymbol{\Omega}_{\mathbf{aa}} \boldsymbol{\mu}_{\mathbf{a}} - \mathbf{a}^\top \boldsymbol{\Omega}_{\mathbf{ab}} (\mathbf{b} - \boldsymbol{\mu}_{\mathbf{b}}) = \mathbf{a}^\top \{\boldsymbol{\Omega}_{\mathbf{aa}} \boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\Omega}_{\mathbf{ab}} (\mathbf{b} - \boldsymbol{\mu}_{\mathbf{b}})\} \quad (1.76)$$

であることから、 $\boldsymbol{\mu}_{\mathbf{a}|\mathbf{b}}$  を計算できる。係数  $\boldsymbol{\Omega}_{\mathbf{aa}} \boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\Omega}_{\mathbf{ab}} (\mathbf{b} - \boldsymbol{\mu}_{\mathbf{b}})$  が  $\boldsymbol{\Sigma}_{\mathbf{a}|\mathbf{b}}^{-1} \boldsymbol{\mu}_{\mathbf{a}|\mathbf{b}}$  と等しくなることから

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{a}|\mathbf{b}}^{-1} \boldsymbol{\mu}_{\mathbf{a}|\mathbf{b}} &= \boldsymbol{\Omega}_{\mathbf{aa}} \boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\Omega}_{\mathbf{ab}} (\mathbf{b} - \boldsymbol{\mu}_{\mathbf{b}}) \\ \boldsymbol{\mu}_{\mathbf{a}|\mathbf{b}} &= \boldsymbol{\Sigma}_{\mathbf{a}|\mathbf{b}} \{\boldsymbol{\Omega}_{\mathbf{aa}} \boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\Omega}_{\mathbf{ab}} (\mathbf{b} - \boldsymbol{\mu}_{\mathbf{b}})\} \\ &= \boldsymbol{\mu}_{\mathbf{a}} - \boldsymbol{\Omega}_{\mathbf{aa}}^{-1} \boldsymbol{\Omega}_{\mathbf{ab}} (\mathbf{b} - \boldsymbol{\mu}_{\mathbf{b}}) \end{aligned} \quad (1.77)$$

であることがわかる。しかしこの式は  $\boldsymbol{\Omega}$  に基づくもので、共分散行列  $\boldsymbol{\Sigma}$  に書き換える必要である。そのためには以下の分割された逆行列の公式が有用である。

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{M} & -\mathbf{MBD}^{-1} \\ -\mathbf{D}^{-1}\mathbf{CM} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{CMBD}^{-1} \end{bmatrix} \quad (1.78)$$

ただし

$$\mathbf{M} = (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} \quad (1.79)$$

この公式から

$$\boldsymbol{\Omega}_{\mathbf{aa}} = (\boldsymbol{\Sigma}_{\mathbf{aa}} - \boldsymbol{\Sigma}_{\mathbf{ab}} \boldsymbol{\Sigma}_{\mathbf{bb}}^{-1} \boldsymbol{\Sigma}_{\mathbf{ba}})^{-1} \quad (1.80)$$

$$\boldsymbol{\Omega}_{\mathbf{ab}} = -(\boldsymbol{\Sigma}_{\mathbf{aa}} - \boldsymbol{\Sigma}_{\mathbf{ab}} \boldsymbol{\Sigma}_{\mathbf{bb}}^{-1} \boldsymbol{\Sigma}_{\mathbf{ba}})^{-1} \boldsymbol{\Sigma}_{\mathbf{ab}} \boldsymbol{\Sigma}_{\mathbf{bb}}^{-1} \quad (1.81)$$

を得る. よって条件付き分布  $(\mathbf{a}|\mathbf{b})$  の平均ベクトルおよび共分散行列は

$$\boldsymbol{\mu}_{\mathbf{a}|\mathbf{b}} = \boldsymbol{\mu}_{\mathbf{a}} + \boldsymbol{\Sigma}_{\mathbf{ab}} \boldsymbol{\Sigma}_{\mathbf{bb}}^{-1} (\mathbf{b} - \boldsymbol{\mu}_{\mathbf{b}}) \quad (1.82)$$

$$\boldsymbol{\Sigma}_{\mathbf{a}|\mathbf{b}} = \boldsymbol{\Sigma}_{\mathbf{aa}} - \boldsymbol{\Sigma}_{\mathbf{ab}} \boldsymbol{\Sigma}_{\mathbf{bb}}^{-1} \boldsymbol{\Sigma}_{\mathbf{ba}} \quad (1.83)$$

である. 期待値  $\mathbb{E}[\cdot]$  および分散  $\text{Var}[\cdot]$  の表記で表すと

$$\mathbb{E}[\mathbf{a}|\mathbf{b}] = \mathbb{E}[\mathbf{a}] + \text{Cov}[\mathbf{a}, \mathbf{b}] \text{Var}[\mathbf{b}]^{-1} (\mathbf{b} - \mathbb{E}[\mathbf{b}]) \quad (1.84)$$

$$\text{Var}[\mathbf{a}|\mathbf{b}] = \text{Var}[\mathbf{a}] - \text{Cov}[\mathbf{a}, \mathbf{b}] \text{Var}[\mathbf{b}]^{-1} \text{Cov}[\mathbf{b}, \mathbf{a}] \quad (1.85)$$

となる.

## 第2章

### 序論

データを集めるだけでは情報を得ることはできない。得られたデータから情報を抽出するためには、解釈が不可欠である。たとえば、2つの変数によるペアが得られたとしよう。この2つの変数の関係を解釈するに、例えば相関係数を算出することなどが考えられるが、図2.1のような散布図から関係を解釈することも重要である。

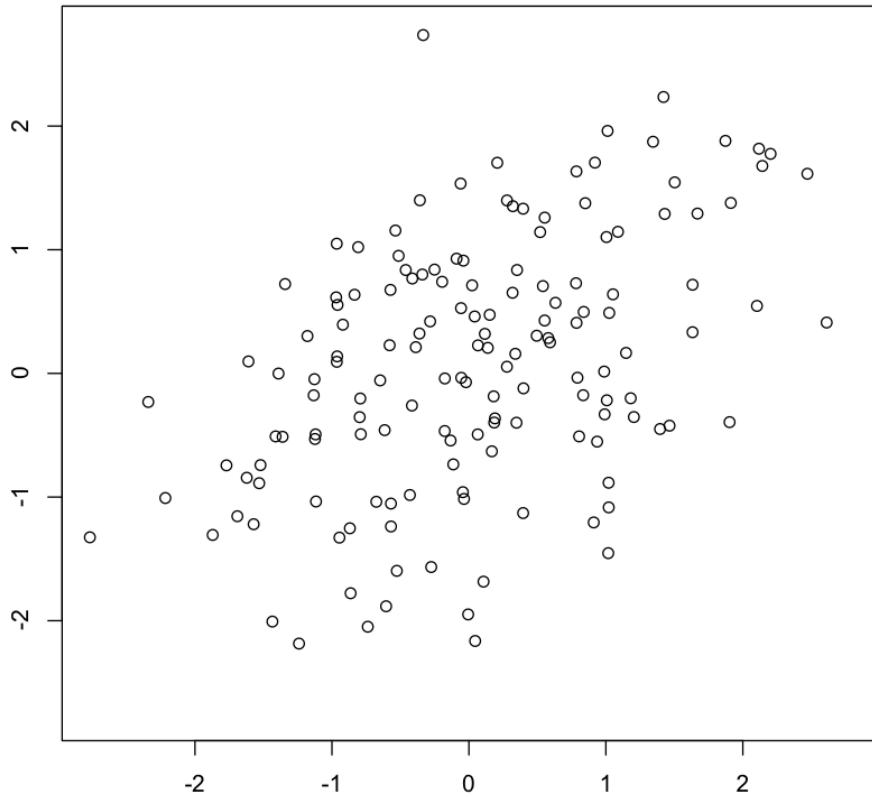


図 2.1: 2 変数の散布図

しかし、得られたデータが4変数や5変数の場合、1つの散布図でデータを表すことはできない。なんらかの方法で、多変数を2変数に変換することができればこのような問題は解決できる。こうした場合に有用な手法が次元縮約法と呼ばれる手法である。次元縮約

法は多数の変数を少ない変数によって表現する手法で、例えば図 2.2 のように 5 変数を 2 変数に縮約できるため散布図での解釈が容易になる。

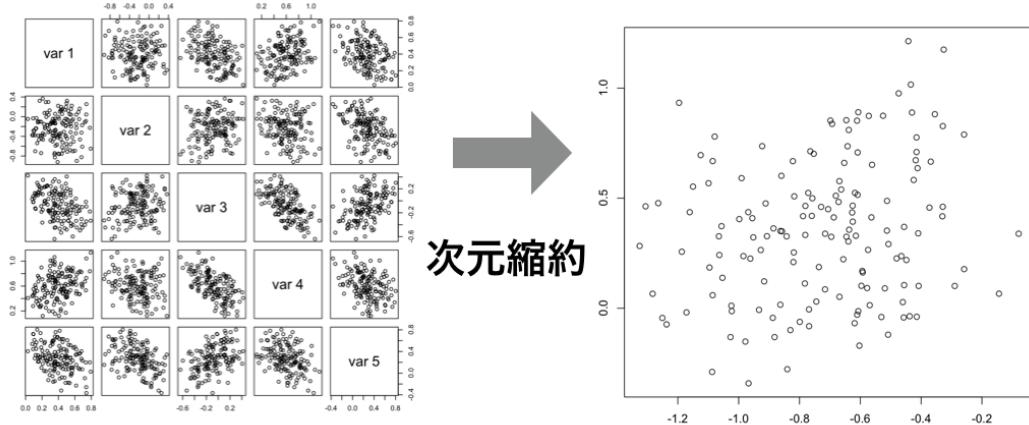


図 2.2: 5 変数の散布図および 2 変数へ次元縮約したときの散布図

次元縮約法は例を用いて説明したような視覚化はもちろん、変数の関係を解釈する上で也有用である。

古くから用いられている主成分分析や多次元尺度法などが次元縮約法の代表例にあげられるが、機械学習においても様々な手法が今なお提案されており (van der Maaten, 2009), 重要な研究テーマの一つである。心理学において活用されることの多い因子分析も、この次元縮約法の一つと考えることができる。

因子分析は 20 世紀はじめに提案されたモデルであり、提案から 100 年以上経た今日でも研究が続けられている。応用範囲も心理学にとどまらず、経済学、社会学、遺伝子解析など多岐に渡っている。

因子分析はもともと、人間の知能を表現するモデルとして提案された (Spearman, 1904)。これは知能の二因子モデルと呼ばれるもので、Spearman が考えたこの概念を視覚的に表現すると、図 2.4 のように表すことができる。音楽や数学など外国語などの習得には知能が関わっていると考えられているが、知能の二因子モデルでは全ての知的活動は一般知能と個別の知能の二つの因子によって説明できるとしている。この Spearman による知能の二因子モデルは、因子分析の一因子モデルと考えることもできる。一因子モデルの因子分析のパス図を図 2.4 と対応させて表現すると、図 2.3 のようになる。

図 2.3, 2.4において「一般知能」とした概念は、因子分析においては共通因子と呼ばれ

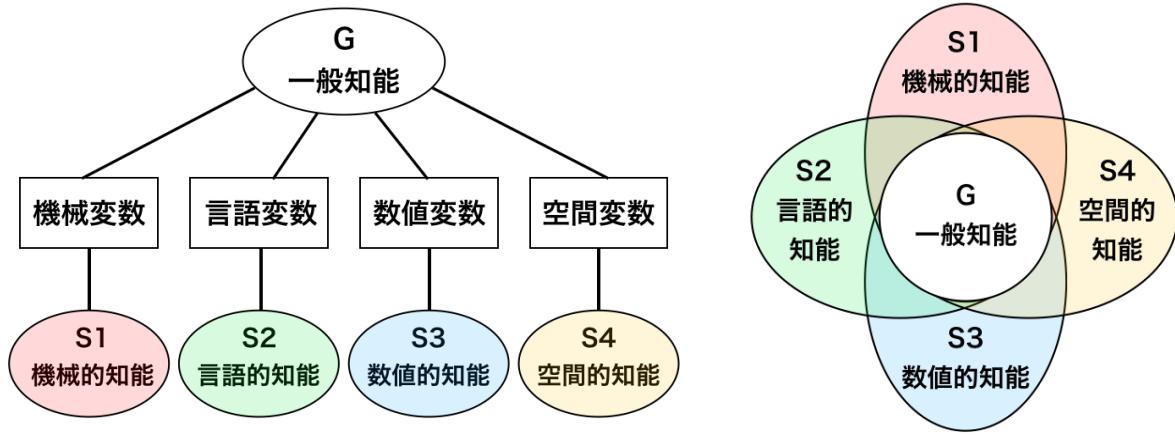


図 2.3: 因子分析のパス図

図 2.4: 知能の二因子モデルの概念図

るもので、それぞれに特有の潜在変数である独自因子とは対照的に、全ての変数と関わる潜在変数である。共通因子は1つに限定されるわけではなく、後に多因子モデルも提案されている。

## 2.1 因子分析の3つのタイプ

### 2.1.1 確率モデル

変数モデルと呼ばれることが多い確率モデル因子分析は、最も広まっている因子分析のモデルである。本章では、他のモデルとの兼ね合いもあり確率モデル因子分析と呼ぶ。確率モデル因子分析において  $P$  次元観測ベクトル  $\mathbf{x}$  は、

$$\mathbf{x} = \Lambda \mathbf{f} + \mathbf{u} \quad (2.1)$$

と表現される。ただし、 $P$  次元多変量正規分布を  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  として、 $M$  次元の共通因子得点ベクトル  $\mathbf{f}$  と  $P$  次元の独自因子得点ベクトル  $\mathbf{u}$  がそれぞれ

$$\mathbf{f} \sim N(\mathbf{0}_M, \mathbf{I}_M) \quad (2.2)$$

$$\mathbf{u} \sim N(\mathbf{0}_P, \boldsymbol{\Psi}^2) \quad (2.3)$$

に従う。さらに  $\Lambda$  は  $P \times M$  の因子負荷行列であり、 $P$  個の顕在変数と  $M$  個の潜在変数の関係を表す。 $\Lambda$ 、および各変数が因子では説明できない独自性をどの程度有するかを示

す独自分散を対角要素とする独自分散行列  $\Psi^2$  の2つが確率モデル因子分析におけるパラメータである。

ここで、重要な仮定として  $f$  と  $u$  は無相関

$$\text{Cov}[f, u] = \mathbf{O}_{M \times P} \quad (2.4)$$

と想定する。

$$\mathbb{E}[x] = \Lambda \mathbb{E}[f] + \mathbb{E}[u] = \mathbf{0}_P \quad (2.5)$$

である。さらに  $\mathbb{V}[f] = \mathbf{I}_M$  であるから

$$\begin{aligned} \mathbb{E}[(f - \mathbb{E}[f])(f - \mathbb{E}[f])^\top] &= \mathbf{I}_M \\ \mathbb{E}[ff^\top] &= \mathbf{I}_M \end{aligned} \quad (2.6)$$

かつ

$$\begin{aligned} \mathbb{E}[(u - \mathbb{E}[u])(u - \mathbb{E}[u])^\top] &= \Psi^2 \\ \mathbb{E}[uu^\top] &= \Psi^2 \end{aligned} \quad (2.7)$$

である。仮定(2.4)式および(2.5)~(2.7)式を用いて(2.1)式の分散を計算すると

$$\begin{aligned} \text{Var}[x] &= \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^\top] \\ &= \mathbb{E}[xx^\top] \\ &= \mathbb{E}[(\Lambda f + u)(\Lambda f + u)^\top] \\ &= \mathbb{E}[\Lambda ff^\top \Lambda^\top + \Lambda fu^\top + uf^\top \Lambda^\top + uu^\top] \\ &= \Lambda \mathbb{E}[ff^\top] \Lambda^\top + \Lambda \mathbb{E}[fu^\top] + \mathbb{E}[uf^\top] \Lambda^\top + \mathbb{E}[uu^\top] \\ &= \Lambda \mathbb{E}[ff^\top] \Lambda^\top + \Lambda \text{Cov}[f, u] + \text{Cov}[u, f] \Lambda^\top + \mathbb{E}[uu^\top] \\ &= \Lambda \Lambda^\top + \Psi^2 \end{aligned}$$

となる。つまりデータの持つ標本分散を

$$\text{データ分散} = \text{共通分散} + \text{独自分散}$$

のように共通因子によって説明可能な割合を示す共通分散と、共通因子では説明できない各変数の独自性を示す独自分散という2つの項に完全に分離した形で表現することが可能となる。

パラメータの推定方法はいくつか存在するが(市川, 2010), 最も汎用的な手法はExpectation - Maximization(EM) アルゴリズム(Dempster, Laird, & Rubin, 1977)による推定法である(Rubin & Thayer, 1982). EM アルゴリズムの基本的な考え方は, 観測されていない欠損データを条件付き期待値によって推定し, その推定値を用いてパラメータの推定を行うことである. 因子分析においては, 共通因子得点  $\mathbf{f}$  を欠損ベクトルとみなし, 条件付き期待値  $\mathbb{E}(\mathbf{f}|\mathbf{x})$  によって推定する.

まず条件付き多変量正規分布を考える.

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{f} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{0}_P \\ \mathbf{0}_M \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2 & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}^\top & \mathbf{I}_M \end{bmatrix} \right) \quad (2.8)$$

であるから, 正規分布の条件付き平均の(1.84)式より

$$\begin{aligned} \mathbb{E}[\mathbf{f}|\mathbf{x}] &= \mathbb{E}[\mathbf{f}] + \text{Cov}[\mathbf{f}, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1}(\mathbf{x} - \mathbb{E}[\mathbf{x}]) \\ &= \boldsymbol{\Lambda}^\top (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2)^{-1} \mathbf{x} \\ &= \mathbf{B}\mathbf{x} \end{aligned} \quad (2.9)$$

また正規分布の条件付き分散は

$$\begin{aligned} \text{Var}[\mathbf{f}|\mathbf{x}] &= \mathbb{E} \left[ (\mathbf{f} - \mathbb{E}[\mathbf{f}|\mathbf{x}]) (\mathbf{f} - \mathbb{E}[\mathbf{f}|\mathbf{x}])^\top | \mathbf{x} \right] \\ &= \mathbb{E} [\mathbf{f}\mathbf{f}^\top - 2\mathbb{E}[\mathbf{f}|\mathbf{x}]\mathbf{f}^\top + \mathbb{E}[\mathbf{f}|\mathbf{x}]\mathbb{E}[\mathbf{f}|\mathbf{x}]^\top | \mathbf{x}] \\ &= \mathbb{E} [\mathbf{f}\mathbf{f}^\top | \mathbf{x}] - \mathbb{E}[\mathbf{f}|\mathbf{x}]\mathbb{E}[\mathbf{f}|\mathbf{x}]^\top \end{aligned}$$

のようにも書けるので, 正規分布の条件付き分散の定義(1.85)式および(2.10)式から

$$\begin{aligned} \mathbb{E}[\mathbf{f}\mathbf{f}^\top | \mathbf{x}] &= \text{Var}[\mathbf{f}|\mathbf{x}] + \mathbb{E}[\mathbf{f}|\mathbf{x}]\mathbb{E}[\mathbf{f}|\mathbf{x}]^\top \\ &= \mathbf{I}_M - \boldsymbol{\Lambda}^\top (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2)^{-1} \boldsymbol{\Lambda} + \mathbf{B}\mathbf{x}\mathbf{x}^\top \mathbf{B}^\top \\ &= \mathbf{I}_M - \mathbf{B}\boldsymbol{\Lambda} + \mathbf{B}\mathbf{x}\mathbf{x}^\top \mathbf{B}^\top \end{aligned} \quad (2.11)$$

となる.

では(2.10)式と(2.11)式を用いて, 欠損パラメータを条件付き期待値によって補完した完全対数尤度であるQ関数を計算してみる. ベイズの定理より  $p(\mathbf{x})$  を  $\mathbf{x}$  の確率密度関数とすると

$$p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{f}) d\mathbf{f} = \int p(\mathbf{x}|\mathbf{f})p(\mathbf{f}) d\mathbf{f}$$

であるから  $p(\mathbf{f})$  および  $p(\mathbf{x}|\mathbf{f})$  を知っておく必要がある。しかし、

$$\prod_{n=1}^N p(\mathbf{f}_n) = \prod_{n=1}^N (2\pi)^{P/2} \det(\text{Var}[\mathbf{f}_n])^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{f}_n - \mathbb{E}[\mathbf{f}])^\top \text{Var}[\mathbf{f}]^{-1} (\mathbf{f}_n - \mathbb{E}[\mathbf{f}]) \right\}$$

であるから、 $\text{Var}[\mathbf{f}] = \mathbf{I}_M$  のとき、定数になるため、 $p(\mathbf{x}|\mathbf{f})$  のみを考えることにする。

正規分布の条件付き平均の定義 (1.84) 式から

$$\begin{aligned} \mathbb{E}[\mathbf{x}|\mathbf{f}] &= \mathbb{E}[\mathbf{x}] + \text{Cov}[\mathbf{x}, \mathbf{f}] \text{Var}[\mathbf{f}]^{-1} (\mathbf{f} - \mathbb{E}[\mathbf{f}]) \\ &= \boldsymbol{\Lambda} \mathbf{f} \end{aligned}$$

であり、正規分布の条件付き分散の定義 (1.85) 式から

$$\begin{aligned} \text{Var}[\mathbf{x}|\mathbf{f}] &= \text{Var}[\mathbf{x}] - \text{Cov}[\mathbf{x}, \mathbf{f}] \text{Var}[\mathbf{f}]^{-1} \text{Cov}[\mathbf{f}, \mathbf{x}] \\ &= \boldsymbol{\Lambda} \boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2 - \boldsymbol{\Lambda} \boldsymbol{\Lambda}^\top \\ &= \boldsymbol{\Psi}^2 \end{aligned}$$

である。よって、 $(\mathbf{x}|\mathbf{f}) \sim N(\boldsymbol{\Lambda} \mathbf{f}, \boldsymbol{\Psi}^2)$  であるから

$$\begin{aligned} Q &= \mathbb{E} \left[ \log \prod_{n=1}^N (2\pi)^{P/2} \det(\boldsymbol{\Psi}^2)^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n)^\top (\boldsymbol{\Psi}^2)^{-1} (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n) \right\} \right] \\ &= \text{const} - \frac{N}{2} \log \det(\boldsymbol{\Psi}^2) - \frac{1}{2} \sum_{n=1}^N \left\{ \mathbb{E} \left[ (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n)^\top \boldsymbol{\Psi}^{-2} (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n) \right] \right\} \\ &= \text{const}_+ - \frac{1}{2} \sum_{n=1}^N \{ \mathbf{x}_n^\top \boldsymbol{\Psi}^{-2} \mathbf{x}_n - \mathbb{E}[\mathbf{f}|\mathbf{x}_n]^\top \boldsymbol{\Lambda} \boldsymbol{\Psi}^{-2} \mathbf{x}_n \\ &\quad - \mathbf{x}_n^\top \boldsymbol{\Psi}^{-2} \boldsymbol{\Lambda} \mathbb{E}[\mathbf{f}|\mathbf{x}_n] + \mathbb{E}[\mathbf{f}_n^\top \boldsymbol{\Lambda}^\top \boldsymbol{\Psi}^{-2} \boldsymbol{\Lambda} \mathbf{f}_n] \} \\ &= \text{const}_+ - \frac{1}{2} \sum_{n=1}^N \{ \mathbf{x}_n^\top \boldsymbol{\Psi}^{-2} \mathbf{x}_n - 2\mathbf{x}_n^\top \boldsymbol{\Psi}^{-2} \boldsymbol{\Lambda} \mathbb{E}[\mathbf{f}|\mathbf{x}_n] + \text{tr} \boldsymbol{\Lambda}^\top \boldsymbol{\Psi}^{-2} \boldsymbol{\Lambda} \mathbb{E}[\mathbf{f} \mathbf{f}^\top |\mathbf{x}_n] \} \quad (2.12) \end{aligned}$$

となる。ただし  $\text{const} = \log(2\pi)^{NP/2}$ ,  $\text{const}_+ = \text{const} - \frac{N}{2} \log \det(\boldsymbol{\Psi}^2)$  である。 $(2.12)$  式の中の等式は (1.15) 式と (1.29) 式、内積の可換性とベクトル内積とトレースの性質から得られる。

$Q$  関数 (2.12) 式を  $\boldsymbol{\Lambda}$  あるいは  $\boldsymbol{\Psi}^{-2}$  について偏微分することによりパラメータ  $\boldsymbol{\Lambda}, \boldsymbol{\Psi}^2$  を得る。

まずは  $\Lambda$  について更新する。

$$\begin{aligned}\frac{\partial Q}{\partial \Lambda} &= \frac{\partial}{\partial \Lambda} \left\{ -2 \sum_{n=1}^N \mathbf{x}_n^\top \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} | \mathbf{x}_n] + \sum_{n=1}^N \text{tr} \Lambda^\top \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \right\} \\ &= \frac{\partial}{\partial \Lambda} \left\{ -2 \sum_{n=1}^N \text{tr} \mathbb{E} [\mathbf{f} | \mathbf{x}_n] \mathbf{x}_n^\top \Psi^{-2} \Lambda + \sum_{n=1}^N \text{tr} \Lambda^\top \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \right\} \quad (2.13)\end{aligned}$$

であるから、トレースの偏微分が必要となる。

$$\text{tr} \Lambda^\top \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] = \text{tr} \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \Lambda^\top \Psi^{-2} \Lambda$$

を利用すると尤度方程式

$$\frac{\partial Q}{\partial \Lambda} = -2 \sum_{n=1}^N \Psi^{-2} \mathbf{x}_n \mathbb{E} [\mathbf{f} | \mathbf{x}_n] + 2 \sum_{n=1}^N \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] = \mathbf{0}_P \quad (2.14)$$

を得る。したがって、

$$\begin{aligned}\Psi^{-2} \sum_{n=1}^N \mathbf{x}_n \mathbb{E} [\mathbf{f} | \mathbf{x}_n] &= \Psi^{-2} \Lambda \sum_{n=1}^N \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \\ \Lambda &= \left( \sum_{n=1}^N \mathbf{x}_n \mathbb{E} [\mathbf{f} | \mathbf{x}_n] \right) \left( \sum_{n=1}^N \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \right)^{-1} \quad (2.15)\end{aligned}$$

が更新式となる。

続いて、 $\Psi^2$  について更新する。 $\Psi^{-2}$  について微分するためには、トレースの偏微分および行列式の偏微分が必要になる。まず指数部分を  $Q_{exp}$  とすると、 $Q - \text{const}_+$  であるから

$$\begin{aligned}Q_{exp} &= -\frac{1}{2} \sum_{n=1}^N \{ \mathbf{x}_n^\top \Psi^{-2} \mathbf{x}_n - 2 \mathbf{x}_n^\top \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} | \mathbf{x}_n] + \text{tr} \Lambda^\top \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \} \\ &= -\frac{1}{2} \sum_{n=1}^N \{ \text{tr} \Psi^{-2} \mathbf{x}_n \mathbf{x}_n^\top - 2 \text{tr} \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} | \mathbf{x}_n] \mathbf{x}_n^\top + \text{tr} \Psi^{-2} \Lambda \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \Lambda^\top \} \quad (2.16)\end{aligned}$$

であるから、

$$\begin{aligned}\frac{\partial Q_{exp}}{\partial \Psi^{-2}} &= -\frac{1}{2} \sum_{n=1}^N \{ \mathbf{x}_n \mathbf{x}_n^\top - 2 \Lambda \mathbb{E} [\mathbf{f} | \mathbf{x}_n] \mathbf{x}_n^\top + \Lambda \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \Lambda^\top \} \\ &= -\frac{1}{2} \sum_{n=1}^N \{ \mathbf{x}_n \mathbf{x}_n^\top \} + \Lambda \sum_{n=1}^N \{ \mathbb{E} [\mathbf{f} | \mathbf{x}_n] \mathbf{x}_n^\top \} - \frac{1}{2} \Lambda \left( \sum_{n=1}^N \mathbb{E} [\mathbf{f} \mathbf{f}^\top | \mathbf{x}_n] \right) \Lambda^\top \quad (2.17)\end{aligned}$$

となる. (2.17) 式をさらに簡潔な形にすることもできる.  $\Lambda$  の更新式 (2.15) 式を最後の項の  $\Lambda^\top$  に代入すると

$$\frac{\partial Q_{exp}}{\partial \Psi^{-2}} = -\frac{1}{2} \sum_{n=1}^N \{\mathbf{x}_n \mathbf{x}_n^\top\} + \frac{1}{2} \Lambda \sum_{n=1}^N \mathbb{E} [\mathbf{f} | \mathbf{x}_n] \mathbf{x}_n^\top \quad (2.18)$$

を得る. つづいて,  $\text{const}_+$ , 特に  $\text{const}_+$  における  $-\frac{N}{2} \log \det(\Psi^2)$  を  $\Psi^{-2}$  で微分する. 逆行列の行列式の性質および行列式の偏微分から,

$$\begin{aligned} \frac{\partial \text{const}_+}{\partial \Psi^{-2}} &= -\frac{\partial}{\partial \Psi^{-2}} \frac{N}{2} \log \det(\Psi^2) \\ &= -\frac{\partial}{\partial \Psi^{-2}} \frac{N}{2} \log \det(\Psi^{-2})^{-1} \\ &= \frac{\partial}{\partial \Psi^{-2}} \frac{N}{2} \log \det(\Psi^{-2}) \\ &= \frac{N}{2} ((\Psi^{-2})^{-1})^\top = \frac{N}{2} \Psi^2 \end{aligned} \quad (2.19)$$

である. (2.18) 式と (2.19) 式から尤度方程式は

$$\frac{\partial Q}{\partial \Psi^{-2}} = \frac{N}{2} \Psi^2 - \frac{1}{2} \sum_{n=1}^N \{\mathbf{x}_n \mathbf{x}_n^\top\} + \frac{1}{2} \Lambda \sum_{n=1}^N \{\mathbb{E} [\mathbf{f} | \mathbf{x}_n] \mathbf{x}_n^\top\} = \mathbf{O}_{P \times P} \quad (2.20)$$

である. したがって,  $\Psi^2$  が対角行列であることに注意すると

$$\Psi^2 = \frac{1}{N} \text{diag} \left( \sum_{n=1}^N \{\mathbf{x}_n \mathbf{x}_n^\top\} - \Lambda \sum_{n=1}^N \mathbb{E} \{[\mathbf{f} | \mathbf{x}_n] \mathbf{x}_n^\top\} \right) \quad (2.21)$$

が  $\Psi^2$  の更新式である. ここで  $\text{diag}(\cdot)$  は非対角要素を 0 にする作用素である.

### 定義 2.1. 行列式の対数の偏微分

$P$  次元の平方行列  $\mathbf{A} \in \mathbb{R}^{P \times P}$  があるとする. このとき以下が成り立つ.

$$\frac{\partial \det(\mathbf{A})}{\partial \mathbf{A}} = (\mathbf{A}^{-1})^\top$$

### 定理 2.1. 逆行列の行列式

行列式には様々な性質がある. 特に逆行列の行列式は有用で以下が成り立つ.

$$\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$$

### 2.1.2 母数モデル

因子分析において、パラメータである因子負荷量および独自分散が解釈の材料となる。しかし、これらは変数に関するパラメータであり個体についてのパラメータではない。個体に関心がある場合は共通因子得点が主な解釈の対象となるが、確率モデル因子分析においては共通因子得点は確率変数でありパラメータではない。そのため近似手法によって共通因子得点を推定することしかできない。

母数モデル因子分析では、独自因子得点を確率変数としたまま、共通因子得点をパラメータとしたモデルである。よって独自因子得点についての分布の仮定は変わらず、

$$\mathbf{u} \sim N(\mathbf{0}_P, \boldsymbol{\Psi}^2) \quad (2.22)$$

とする。問題は共通因子得点であり、確率変数ではなくパラメータとして扱うことから正規分布を用いて共通因子得点の性質を表すことはできない。そこで、(2.2)式の性質を線形代数的に記述する。まず、平均が  $\mathbf{0}$  であることから

$$\mathbf{1}_N^\top \mathbf{F} = \mathbf{0}_M \quad (2.23)$$

とし、共通因子得点行列  $\mathbf{F}$  の各列の平均を 0 とした。また分散が  $\mathbf{I}_M$  であることから

$$N^{-1} \mathbf{F}^\top \mathbf{F} = \mathbf{I}_M \quad (2.24)$$

パラメータ推定としては、最尤推定法が考えられる。モデルは確率モデルと同様に

$$\mathbf{x} = \boldsymbol{\Lambda} \mathbf{f} + \mathbf{u} \quad (2.25)$$

とするので、(2.15)式を移項することで得られる

$$\mathbf{u} = \mathbf{x} - \boldsymbol{\Lambda} \mathbf{f} \quad (2.26)$$

が正規分布 (2.13) 式に従うことがわかる。

$\mathbf{u}$  が独立かつ同一の分布 (i.i.d.) に従うとすると、それらの同時分布の確率密度関数は

$$\begin{aligned} p(\mathbf{u}_1, \dots, \mathbf{u}_N | \mathbf{0}_P, \boldsymbol{\Psi}^2) &= \frac{1}{(2\pi)^{NP/2}} \frac{1}{\det(\boldsymbol{\Psi}^2)^{N/2}} \exp \left\{ -\frac{1}{2} \sum_{n=1}^N (\mathbf{u}_n - \mathbf{0}_P)^\top (\boldsymbol{\Psi}^2)^{-1} (\mathbf{u}_n - \mathbf{0}_P) \right\} \\ &\propto \frac{1}{\det(\boldsymbol{\Psi}^2)^{N/2}} \exp \left\{ -\frac{1}{2} \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n)^\top \boldsymbol{\Psi}^{-2} (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n) \right\} \end{aligned} \quad (2.27)$$

さらに  $\mathbf{S} = \sum_{n=1}^N (\mathbf{x}_n - \Lambda \mathbf{f}_n)(\mathbf{x}_n - \Lambda \mathbf{f}_n)^\top$  とすると (2.27) 式を

$$p(\mathbf{u}_1, \dots, \mathbf{u}_N | \mathbf{0}_P, \Psi^2) \propto \frac{1}{\det(\Psi^2)^{P/2}} \exp \left\{ -\frac{1}{2} \text{tr} (\mathbf{S} \Psi^{-2}) \right\} \quad (2.28)$$

と書き換えることもできる。

通常は最尤法では対数尤度を最大化するため、目的関数は

$$\log L = -\det(\Psi^2)^{N/2} - \frac{1}{2} \{ \text{tr}(\mathbf{x}_n - \Lambda \mathbf{f}_n)^\top \Psi^{-2} (\mathbf{x}_n - \Lambda \mathbf{f}_n) \} + \text{const} \quad (2.29)$$

であり、パラメータは目的関数を最大化することにより推定できる。

ところが、最尤法による母数モデルのパラメータ推定は、尤度が発散してしまうため不可能である。これは母数モデルを活用するうえで致命的な問題である。

### 2.1.3 行列モデル

確率モデルでは、共通因子得点と独自因子得点を確率変数として考えていた。近年、因子負荷行列と独自分散のみではなく、因子得点もパラメータとして扱う行列モデル因子分析が研究されている。この行列モデルという名称は、最適化の過程で行列分解を用いられていることに因んでいる。確率モデルや母数モデルに比べ、行列モデルは新しいモデルであるため、理論的な研究も進み始めたばかりである (Adachi & Trendafilov, 2018)。

行列モデル因子分析では、データ行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$  が与えられたとき

$$\mathbf{X} = \mathbf{F} \Lambda^\top + \mathbf{U} \Psi + \mathbf{E}_{FA} \quad (2.30)$$

というモデルを考える。 $\mathbf{F} \in \mathbb{R}^{N \times M}$  は共通因子得点行列、 $\Lambda \in \mathbb{R}^{P \times M}$  は因子負荷行列、 $\mathbf{U} \in \mathbb{R}^{N \times P}$  は独自因子得点行列、 $\Psi \in \mathbb{R}^{P \times P}$  は対角行列であり、 $\Psi^2$  の対角要素は独自分散である。これら 4 つの行列全てがパラメータであり、制約条件

$$\mathbf{1}_N^\top \mathbf{F} = \mathbf{0}_M \quad (2.31)$$

$$\mathbf{1}_N^\top \mathbf{U} = \mathbf{0}_P \quad (2.32)$$

$$N^{-1} \mathbf{F}^\top \mathbf{F} = \mathbf{I}_M \quad (2.33)$$

$$N^{-1} \mathbf{U}^\top \mathbf{U} = \mathbf{I}_P \quad (2.34)$$

$$\mathbf{F}^\top \mathbf{U} = \mathbf{O}_{M \times P} \quad (2.35)$$

を満たす。パラメータの最適解を一度に解くことはできないため、交互最小二乗法により推定する。最小化する目的関数は、

$$f(\mathbf{F}, \boldsymbol{\Lambda}, \mathbf{U}, \boldsymbol{\Psi}) = \|\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}^\top - \mathbf{U}\boldsymbol{\Psi}\|_F^2 \quad (2.36)$$

であり、最適化したいパラメータ以外を固定したうえで更新し、 $f$  が収束するまで更新を繰り返す。

#### 2.1.4 因子分析と主成分分析の違い

古くから用いられてきた次元縮約法である因子分析と主成分分析は、その考え方が似ていることから、しばしば混同されてきた。そこで本小節では、因子分析と主成分分析の違いを述べる。

2.1節で述べたように、因子分析には3つのタイプがあり確率モデルが最も普及している。それに対して主成分分析は、確率分布の仮定を用いない行列モデルが一般的である。そこで、まず確率モデルの主成分分析 (Tipping & Bishop, 1999) と確率モデル因子分析を比較する。その後、行列モデルの主成分分析と行列モデル因子分析を比較する。確率モデルの主成分分析は、確率主成分分析 (Probabilistic Principal Component Analysis) と呼ばれ、そのモデルは

$$\mathbf{x} = \mathbf{A}\mathbf{h} + \mathbf{e} \quad (2.37)$$

と確率モデル因子分析 (2.1) 式と同じ形式である。しかし、それぞれの確率変数の従う分布に違いがあり

$$\mathbf{h} \sim N(\mathbf{0}_M, \mathbf{I}_M) \quad (2.38)$$

$$\mathbf{e} \sim N(\mathbf{0}_P, \sigma \mathbf{I}_P) \quad (2.39)$$

となっている。誤差の分布 (2.39) 式と因子分析における独自因子の分布 (2.3) 式を比べると、主成分分析では誤差分散が変数間で同一なのに対して、因子分析の独自分散は変数ごとに分散が変わっている。それ以外に違いはないため、誤差の等分散性を仮定するか否かのみが違いとなる。

次に行列モデルでの違いを比較する。行列モデルの主成分分析は、データ行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$  が与えられたとき

$$\mathbf{X} = \mathbf{F}\mathbf{A}^\top + \mathbf{E}_{PCA} \quad (2.40)$$

というモデルを想定する。行列モデル因子分析のモデル (2.30) 式と比較すると、主成分分析のモデル誤差にあたる  $\mathbf{E}_{PCA}$  が、行列モデル因子分析では  $\mathbf{U}\Psi + \mathbf{E}_{FA}$  とさらにリパラメトライズされていることがわかる。

これは重要な違いであり、行列モデル因子分析では主成分分析や確率モデル因子分析では分離不可能だったモデルの独自部分と観測誤差を  $\mathbf{U}\Psi + \mathbf{E}_{FA}$  に分離可能なことを示唆している。実際、行列モデル因子分析では確率モデル因子分析に比べ  $\Psi^2$  の値が小さくなる傾向にあることが報告されている (足立, 2014)。考えられる説として、従来は観測誤差の影響によって独自性が過大評価されていた可能性がある。

## 2.2 因子分析における2つの不定性

因子分析には2つの不定性がある。1つは回転の不定性であり、もう1つは因子得点の不定性である。不定性は、パラメータを変換すれば異なるモデルをいくらでも作ることができることを意味する。したがって、不定性を解消することは因子分析を運用するうえで重要なテーマである。本節では、因子分析におけるこの2つの不定性について述べる。

### 2.2.1 回転の不定性

前節にて解説した3つのタイプの因子分析いずれも回転の不定性を有しており、なんらかの強い仮定を置かなければ不定性を避けることはできない。そして因子分析を用いる際に、分析者は回転の不定性の深刻さを知っておく必要がある。なぜなら回転の不定性は、因子分析における主目的である観測変数と因子との関係の解釈に用いられる因子負荷量の値を著しく変化させる恐れがあるからである。

確率モデル因子分析のモデルは

$$\mathbf{x} = \Lambda\mathbf{f} + \mathbf{u}$$

である。行列  $\mathbf{T} \in \mathbb{R}^{M \times M}$  が制約  $\mathbf{T}^\top \mathbf{T} = \mathbf{T} \mathbf{T}^\top = \mathbf{I}_M$  を満たすならば

$$\mathbf{x} = \Lambda \mathbf{T} \mathbf{T}^\top \mathbf{f} + \mathbf{u} = \tilde{\Lambda} \tilde{\mathbf{f}} + \mathbf{u} \quad (2.41)$$

のように書き換えることができる。しかし  $\Lambda$  と  $\tilde{\Lambda}$  はもはや別の行列であるにもかかわらず、因子分析の目的関数は変化しない。なぜなら

$$\begin{aligned} \|N^{-1} \mathbf{X}^\top \mathbf{X} - (\Lambda \Lambda^\top + \Psi^2)\|_F^2 &= \|\mathbf{S}_{\mathbf{XX}} - (\Lambda \mathbf{T} \mathbf{T}^\top \Lambda^\top + \Psi^2)\|_F^2 \\ &= \|\mathbf{S}_{\mathbf{XX}} - (\tilde{\Lambda} \tilde{\Lambda}^\top + \Psi^2)\|_F^2 \end{aligned} \quad (2.42)$$

あるいは

$$\begin{aligned} &\log(|\Lambda \Lambda^\top + \Psi^2|) + \text{tr}((\Lambda \Lambda^\top + \Psi^2)^{-1} \mathbf{S}_{\mathbf{XX}}) \\ &= \log(|\Lambda \mathbf{T} \mathbf{T}^\top \Lambda^\top + \Psi^2|) + \text{tr}((\Lambda \mathbf{T} \mathbf{T}^\top \Lambda^\top + \Psi^2)^{-1} \mathbf{S}_{\mathbf{XX}}) \\ &= \log(|\tilde{\Lambda} \tilde{\Lambda}^\top + \Psi^2|) + \text{tr}((\tilde{\Lambda} \tilde{\Lambda}^\top + \Psi^2)^{-1} \mathbf{S}_{\mathbf{XX}}) \end{aligned} \quad (2.43)$$

のように、制約  $\mathbf{T}^\top \mathbf{T} = \mathbf{T} \mathbf{T}^\top = \mathbf{I}_M$  のおかげで  $\Lambda$  が  $\tilde{\Lambda}$  に置き換えたとしても目的関数は変化しない。

実際に例を用いて  $\Lambda$  と  $\tilde{\Lambda}$  でどれくらい異なるか、検証してみる。

$$\Lambda = \begin{bmatrix} 0.80 & 0.60 \\ 0.60 & 0.50 \\ 0.60 & -0.50 \\ -0.50 & 0.70 \\ -0.60 & 0.40 \\ -0.50 & -0.60 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} 0.71 & -0.70 \\ 0.70 & 0.71 \end{bmatrix}, \quad \tilde{\Lambda} = \Lambda \mathbf{T} = \begin{bmatrix} 0.99 & -0.13 \\ 0.78 & -0.07 \\ 0.08 & -0.78 \\ 0.14 & 0.85 \\ -0.15 & 0.71 \\ -0.78 & -0.08 \end{bmatrix}$$

上式のように、 $\Lambda$  と  $\tilde{\Lambda}$  の解釈容易性は全く異なっていることがわかる。そのため、因子分析によって得られた因子負荷行列には何らかの回転法を適用した後に解釈する2段階の分析が通例となっている。

## 2.2.2 因子得点の不定性

回転の不定性、あるいは回転の不定性を利用した回転法の存在が心理学者に広く認知されているのに対して、因子得点の不定性は十分に知られているとはいえない。しかし、因

子得点の不定性は因子分析によって変数ではなく個人の関係を解釈しようとする場合、多大な影響を与える。なぜなら、回転の不定性によって因子負荷量が様々な値に変化しうるのと同様に、因子得点の不定性によって個人の関係を表す共通因子得点および独自因子得点が、大きく変化する可能性があるからである。

確率モデル因子分析における因子得点の不定性は丘本(1986)などで詳しく述べられている。今、因子モデルをもつデータベクトル  $\mathbf{x}$  に無相関な  $M$  次元確率ベクトル  $\mathbf{t}$  を考える。 $\mathbf{t}$  の平均ベクトルと共分散行列を

$$\mathbb{E}[\mathbf{t}] = \mathbf{0}_M \quad (2.44)$$

$$\mathbb{E}[\mathbf{t}\mathbf{t}^\top] = \Sigma_{tt} \quad (2.45)$$

とする。ここで

$$\tilde{\mathbf{t}} = (\mathbf{I}_M - \Lambda^\top \Sigma^{-1} \Lambda)^{1/2} \Sigma_{tt}^{-1/2} \mathbf{t} \quad (2.46)$$

とすると、

$$\begin{aligned} \mathbb{E}[\tilde{\mathbf{t}}] &= (\mathbf{I}_M - \Lambda^\top \Sigma^{-1} \Lambda)^{1/2} \Sigma_{tt}^{-1/2} \mathbb{E}[\mathbf{t}] \\ &= \mathbf{0}_M \end{aligned} \quad (2.47)$$

かつ

$$\begin{aligned} \mathbb{E}[\tilde{\mathbf{t}}\tilde{\mathbf{t}}^\top] &= \mathbb{E}\left[(\mathbf{I}_M - \Lambda^\top \Sigma^{-1} \Lambda)^{1/2} \Sigma_{tt}^{-1/2} \mathbf{t} \mathbf{t}^\top \Sigma_{tt}^{-1/2} (\mathbf{I}_M - \Lambda^\top \Sigma^{-1} \Lambda)^{1/2}\right] \\ &= (\mathbf{I}_M - \Lambda^\top \Sigma^{-1} \Lambda) \end{aligned} \quad (2.48)$$

となる。 $\tilde{\mathbf{t}}$  を用いて、共通因子得点ベクトル  $\mathbf{f}$  と独自因子得点ベクトル  $\mathbf{u}$  を

$$\tilde{\mathbf{f}} = \Lambda^\top \Sigma^{-1} \mathbf{x} + \tilde{\mathbf{t}} \quad (2.49)$$

$$\tilde{\mathbf{u}} = \Psi^2 \Sigma^{-1} \mathbf{x} - \Lambda \tilde{\mathbf{t}} \quad (2.50)$$

と書き換えると、

$$\mathbf{x} = \Lambda \tilde{\mathbf{f}} + \tilde{\mathbf{u}} \quad (2.51)$$

が成り立つ。明らかに  $\mathbb{E}[\tilde{\mathbf{f}}] = \mathbf{0}$ ,  $\mathbb{E}[\tilde{\mathbf{u}}] = \mathbf{0}$  で、 $\mathbf{t}$  と  $\mathbf{x}$  は無相関  $\mathbb{E}[\mathbf{x}\mathbf{t}^\top] = \mathbf{0}_{P \times M}$  であるから、

$$\begin{aligned} \mathbb{E}[\mathbf{x}\tilde{\mathbf{t}}^\top] &= \mathbb{E}\left[\mathbf{x}\mathbf{t}^\top \Sigma_{tt}^{-1/2} (\mathbf{I}_M - \Lambda^\top \Sigma^{-1} \Lambda)^{1/2}\right] \\ &= \mathbf{0}_{P \times M} \end{aligned} \quad (2.52)$$

となる。このことから

$$\begin{aligned}
\mathbb{E} [\tilde{\mathbf{f}} \tilde{\mathbf{f}}^\top] &= \mathbb{E} [(\Lambda^\top \Sigma^{-1} \mathbf{x} + \tilde{\mathbf{t}}) (\Lambda^\top \Sigma^{-1} \mathbf{x} + \tilde{\mathbf{t}})^\top] \\
&= \Lambda^\top \Sigma^{-1} \Lambda + \Lambda^\top \Sigma^{-1} \mathbb{E} [\mathbf{x} \tilde{\mathbf{t}}^\top] + \mathbb{E} [\tilde{\mathbf{t}} \mathbf{x}^\top] \Sigma^{-1} \Lambda + \mathbb{E} [\tilde{\mathbf{t}} \tilde{\mathbf{t}}^\top] \\
&= \Lambda^\top \Sigma^{-1} \Lambda + \mathbf{O}_{M \times M} + \mathbf{O}_{M \times M} + (\mathbf{I}_M - \Lambda^\top \Sigma^{-1} \Lambda) \\
&= \mathbf{I}_M
\end{aligned} \tag{2.53}$$

さらに

$$\begin{aligned}
\mathbb{E} [\tilde{\mathbf{u}} \tilde{\mathbf{u}}^\top] &= \mathbb{E} [(\Psi^2 \Sigma^{-1} \mathbf{x} - \Lambda \tilde{\mathbf{t}})(\Psi^2 \Sigma^{-1} \mathbf{x} - \Lambda \tilde{\mathbf{t}})^\top] \\
&= \Psi^2 \Sigma^{-1} \Psi^2 + \Psi^2 \Sigma^{-1} \mathbb{E} [\mathbf{x} \tilde{\mathbf{t}}^\top] \Lambda - \Lambda \mathbb{E} [\tilde{\mathbf{t}} \mathbf{x}^\top] \Sigma^{-1} \Psi^2 + \Lambda \mathbb{E} [\tilde{\mathbf{t}} \tilde{\mathbf{t}}^\top] \Lambda^\top \\
&= \Psi^2 \Sigma^{-1} \Psi^2 + \Lambda (\mathbf{I}_M - \Lambda^\top \Sigma^{-1} \Lambda) \Lambda^\top \\
&= \Psi^2 \Sigma^{-1} \Psi^2 + \Lambda \Lambda^\top - \Lambda \Lambda^\top \Sigma^{-1} \Lambda \Lambda^\top \\
&= \Psi^2 \Sigma^{-1} \Psi^2 + \Sigma - \Psi^2 - (\Sigma - \Psi^2) \Sigma^{-1} (\Sigma - \Psi^2) \\
&= \Psi^2 \Sigma^{-1} \Psi^2 + \Sigma - \Psi^2 - \Sigma + \Psi^2 + \Psi^2 - \Psi^2 \Sigma \Psi^2 \\
&= \Psi^2
\end{aligned} \tag{2.54}$$

であり

$$\begin{aligned}
\mathbb{E} [\tilde{\mathbf{f}} \tilde{\mathbf{u}}^\top] &= \mathbb{E} [(\Lambda^\top \Sigma^{-1} \mathbf{x} + \tilde{\mathbf{t}}) (\Psi^2 \Sigma^{-1} \mathbf{x} - \Lambda \tilde{\mathbf{t}})^\top] \\
&= \Lambda^\top \Sigma^{-1} \Psi^2 - \Lambda^\top \Sigma^{-1} \mathbb{E} [\mathbf{x} \tilde{\mathbf{t}}^\top] \Lambda^\top + \mathbb{E} [\tilde{\mathbf{t}} \mathbf{x}^\top] \Sigma^{-1} \Psi^2 - \mathbb{E} [\tilde{\mathbf{t}} \tilde{\mathbf{t}}^\top] \Lambda^\top \\
&= \Lambda^\top \Sigma^{-1} \Psi^2 - (\mathbf{I}_P - \Lambda^\top \Sigma^{-1} \Lambda) \Lambda^\top \\
&= \Lambda^\top \Sigma^{-1} \Psi^2 - \Lambda^\top + \Lambda^\top \Sigma^{-1} \Lambda \Lambda^\top \\
&= \Lambda^\top \Sigma^{-1} \Psi^2 - \Lambda^\top + \Lambda^\top \Sigma^{-1} (\Sigma - \Psi^2) \\
&= \Lambda^\top \Sigma^{-1} \Psi^2 - \Lambda^\top + \Lambda^\top - \Lambda^\top \Sigma^{-1} \Psi^2 \\
&= \mathbf{O}_{M \times P}
\end{aligned} \tag{2.55}$$

である。よって、 $\mathbf{f}$ と $\mathbf{u}$ を $\tilde{\mathbf{f}}$ と $\tilde{\mathbf{u}}$ に置き換えるても因子分析モデルが成り立つ。ただし $\tilde{\mathbf{f}}$ の第2項 $\tilde{\mathbf{t}}$ は、幾通りも考えることができる。したがって確率モデル因子分析には因子得点の不定性が存在する。

行列モデル因子分析における、因子得点の不定性は違った形で示される。目的関数が

$$\begin{aligned} & \| \mathbf{X} - \mathbf{F}\Lambda^\top + \mathbf{U}\Psi \|^2_F \\ &= \| \mathbf{X} - [\mathbf{F}, \mathbf{U}][\Lambda, \Psi]^\top \|^2_F \\ &= \| \mathbf{X} - \mathbf{Z}\Lambda^\top \|^2_F \end{aligned}$$

と書きかえる。ここで、共通因子得点行列と独自因子得点行列からなるブロック行列  $\mathbf{Z}$  の更新は、ten Berge の定理より

$$\mathbf{Z} = N^{1/2} \mathbf{KL}^\top \quad (2.56)$$

によって達成される。(Unkel & Trendafilov, 2010)。ただし、行列  $\mathbf{K}$  と行列  $\mathbf{L}$  は固有値分解

$$N^{-1/2} \mathbf{XA} = \mathbf{K}\Theta\mathbf{L}^\top \quad (2.57)$$

によって得られる。ただしこの固有値分解は

$$\begin{aligned} N^{-1/2} \mathbf{XA} &= \mathbf{K}\Theta_1\mathbf{L}^\top \\ &= [\mathbf{K}_1, \mathbf{K}_2] \begin{bmatrix} \Theta & \\ & \mathbf{O}_{M \times M} \end{bmatrix} \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{bmatrix} \end{aligned} \quad (2.58)$$

のようにブロック行列を用いて表現することができる。ただし  $\mathbf{K}_1 \in \mathbb{R}^{N \times P}$ ,  $\mathbf{K}_2 \in \mathbb{R}^{N \times M}$ ,  $\mathbf{L}_1 \in \mathbb{R}^{(P+M) \times P}$ ,  $\mathbf{L}_2 \in \mathbb{R}^{(P+M) \times M}$  であり、 $\Theta_1$  は  $P$  次元の対角行列である。そして、制約条件にに対応するように

$$\mathbf{K}_1^\top \mathbf{K}_1 = \mathbf{L}_2^\top \mathbf{L}_2 = \mathbf{I}_P, \mathbf{K}_1^\top \mathbf{K}_2 = \mathbf{L}_1^\top \mathbf{L}_2 = \mathbf{O}_{P \times M} \quad (2.59)$$

を満たす。上式は  $\text{rank}(\mathbf{XA}) = P < (P + M)$  により、ランク落ちした結果、 $M$  個の特異値が 0 であることを示唆している。その結果  $\Theta$  の中の  $\mathbf{O}_{M \times M}$  に対応する特異行列である  $\mathbf{K}_2$  と  $\mathbf{L}_2$  は、制約条件を満たす限り自由に値をとることができる。このことから、

$$\begin{aligned} \mathbf{Z} &= N^{1/2} \mathbf{KL}^\top \\ &= N^{1/2} \mathbf{K}_1 \mathbf{L}_1^\top + N^{1/2} \mathbf{K}_2 \mathbf{L}_2 \end{aligned} \quad (2.60)$$

と表すと、左辺第 2 項  $N^{1/2} \mathbf{K}_2 \mathbf{L}_2$  は一意に定まらない。これが行列モデル因子分析における因子得点の不定性である。

## 第3章

# クラスタリングを伴う母数モデル因子分析

本章は Uno, Satomura, & Adachi(2016) および宇野 (2016) を拡張したものである.

### 3.1 導入

$N$  個体× $P$  変数 の列中心化された行列を  $\mathbf{X}$  としたとき, 因子分析のモデルは次のように書くことができる.

$$\mathbf{X} = \mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{E}, \quad (3.1)$$

(Mulaik, 2010). このとき, 行列  $\mathbf{F}$  は  $N$  個体× $M$  因子の因子得点行列,  $\boldsymbol{\Lambda}$  は  $P \times M$  の因子負荷行列である. ただし,  $M < P$  とする. また, 行列  $\mathbf{E}$  の各列は,  $1 \times P$  の確率誤差ベクトル  $e^\top$  とする. 確率誤差ベクトル  $e$  の確率分布を仮定するとき, 通常は平均ベクトルを  $P \times 1$  の 0 ベクトルとする多変量正規分布に従うものとする. つまり,

$$\mathbf{e} \sim N(\mathbf{0}_P, \boldsymbol{\Psi}^2). \quad (3.2)$$

ここで, 共分散行列  $\boldsymbol{\Psi}^2$  は対角行列とし,

$$\boldsymbol{\Psi}^2 = \text{diag}\{\psi_1^2, \dots, \psi_P^2\}, \quad (3.3)$$

と書ける.  $\boldsymbol{\Psi}^2$  の対角要素,  $\psi_1^2, \dots, \psi_P^2$  は独自分散と呼ばれる. 通常, 因子分析では, 推定するパラメータ行列は  $\boldsymbol{\Lambda}$  と  $\boldsymbol{\Psi}$  である. その一方, 行列  $\mathbf{F}$  の中の, 因子得点は確率変数として扱われることが一般的である. しかし, この行列  $\mathbf{F}$  もパラメータ行列として推定する因子分析が Lawley(1942) によって提案された. このモデルは, fixed factor analysis(Unkel & Trendafilov, 2010), もしくは母数モデル因子分析と呼ばれている (丘本, 1986). 本研究では, この母数モデル因子分析に着目する. 母数モデル因子分析において, 行列  $\mathbf{F}$  は列中心

化されており、直交行列である。つまり、

$$\mathbf{1}_N^\top \mathbf{F} = \mathbf{0}_M, \quad (3.4)$$

$$\frac{1}{N} \mathbf{F}^\top \mathbf{F} = \mathbf{I}_M, \quad (3.5)$$

を満たす。ここで、 $\mathbf{1}_P$  は全ての要素が 1 である  $N \times 1$  のベクトル、 $\mathbf{I}_M$  は  $M \times M$  の単位行列を表す。したがって、母数モデル因子分析においてパラメータ行列は  $\mathbf{F}, \boldsymbol{\Lambda}$  と  $\boldsymbol{\Psi}^2$  である。ただし次節で説明する通り、これらのパラメータを同時に推定することは不可能であることが知られている (Anderson & Rubin, 1956)。本提案手法では、 $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_N]^\top$  に制約をかけることで、パラメータを同時に推定できることを目指す。ここで、 $\mathbf{f}_n$  は個体  $n$  の因子得点ベクトルを意味する。

$\mathbf{F}$  にかける制約であるが、提案手法では、 $\mathbf{F}$  中の  $N$  個体が少数のクラスターに分類され、同じクラスターに所属する個体は全て同一の因子得点を持つとした。この制約を加えた母数モデル因子分析を、本論文では **Fixed Clustered Factor Analysis** より、以後 FCFA と呼ぶ。FCFA の定式化は第 2 節で行い、なぜ FCFA においてはパラメータを同時に推定できるかを説明する。第 3 節では FCFA のアルゴリズムを述べ、第 4 節ではシミュレーションによって FCFA が正しく推定できることを確認する。

FCFA は、二つの目的を同時に達成することが可能な手法で、[1] 多数の変数を説明する因子を抽出する、[2] 個体を分類する、という目的のために有効である。同様の目的は、既に提案されている二つの手法、1994 年に De Soete と Carsoll により提案された **Reduced K-Means analysis** (RKM) と 2001 年に Vichi と Kiers により提案された **Factorial K-Means analysis** によっても達成できる。ただし、この二つの手法は因子分析ではなく主成分分析に基づいた手法である。よって、[1] における因子は、主成分に置き換わる。RKM と FKM が FCFA とどのように関連しているかは第 5 節にて述べる。さらに RKM と FKM が最小二乗法による推定なのに対し、FCFA は最尤法による推定である。しかし、式 (3.3) を単位行列  $\mathbf{I}_P$  として RKM の最尤推定と FCFA が一致することから、RKM は FCFA の特殊例と言える。そして第 6 節では、分類精度の点から FCFA が RKM と FKM を上回っていることを示す。

### 3.2 モデル

本節ではまず、なぜ母数モデル因子分析におけるパラメータを同時に推定することが不可能なのかを説明する。その後、提案手法であるFCFAを定式化する。そして最後に、なぜFCFAにおいてはパラメータの同時推定が可能であるかと、同時推定が不可能となる例外を述べる。

式(3.1)～(3.3)を用いると、母数モデル因子分析の対数尤度は正規分布の確率密度関数に  $\mathbf{E}$  を代入することで得られ、以下のように書ける。

$$\begin{aligned} LL(\mathbf{F}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}^2) &\propto -N \sum_{p=1}^P \log \psi_p^2 - \text{tr}(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}^\top)\boldsymbol{\Psi}^{-2}(\mathbf{X} - \mathbf{F}\boldsymbol{\Lambda}^\top)^\top \\ &= -\sum_{p=1}^P \left\{ \frac{1}{\psi_p^2} \|\mathbf{x}_p - \mathbf{F}\boldsymbol{\lambda}_p\|_2^2 + N \log \psi_p^2 \right\}. \end{aligned} \quad (3.6)$$

ここでベクトル  $\mathbf{x}_p$  は  $\mathbf{X}$  の第  $p$  列を、 $\boldsymbol{\lambda}_p$  は  $\boldsymbol{\Lambda}$  の第  $p$  行を、 $\psi_p^2$  は  $\boldsymbol{\Psi}^2$  の  $p$  番目の対角要素を表す。しかし、(3.6)式の最尤推定量は以下で述べるように存在しない。

独自分散の推定値は、必ず

$$\psi_p^2 = \frac{1}{N} \|\mathbf{x}_p - \mathbf{F}\boldsymbol{\lambda}_p\|_2^2, \quad (3.7)$$

を満たす。これは尤度方程式

$$\begin{aligned} \frac{\partial}{\partial \psi_p^2} \left\{ \frac{1}{\psi_p^2} \|\mathbf{x}_p - \mathbf{F}\boldsymbol{\lambda}_p\|_2^2 + N \log \psi_p^2 \right\} &= 0 \\ -\frac{1}{\psi_p^4} \|\mathbf{x}_p - \mathbf{F}\boldsymbol{\lambda}_p\|_2^2 + \psi_p^2 N &= 0 \end{aligned}$$

から得られる。しかし、(3.6)式は、 $\mathbf{F}\boldsymbol{\lambda}_p \rightarrow \mathbf{x}_p$ 、すなわち  $\mathbf{x}_p = \mathbf{F}\boldsymbol{\lambda}_p$  のように発散するので (Anderson & Rubin, 1956)，独自分散の推定値は 0 になってしまう。

それに対して、提案手法であるFCFAでは  $\mathbf{F}$  に制約を加えているため、 $\mathbf{F}, \boldsymbol{\Lambda}$  と  $\boldsymbol{\Psi}^2$  の最尤推定量が与えられ、(3.6)式が最大化される。その制約とは、前述の通り、 $\mathbf{F}$  における個体、つまり  $\mathbf{F}$  の行  $\mathbf{f}_1, \dots, \mathbf{f}_N$  が  $K$  個のクラスターに分類されるという制約である。ただし、 $K$  は  $K < N$  を満たす。この制約を式で表すと、

$$\mathbf{F} = \mathbf{G}\mathbf{C}, \quad (3.8)$$

となる。ここで、 $\mathbf{G} = \{g_{nk}\}$  は  $N$  個体× $K$  クラスターのメンバーシップ行列を表す。個体  $n$  がクラスター  $k$  に分類されるときのみ  $g_{nk} = 1$  となり、それ以外の場合は  $g_{nk} = 0$  である。式で表現すると、

$$g_{nk} = \{0, 1\}, \quad \mathbf{G}\mathbf{1}_K = \mathbf{0}_N, \quad \text{rank}(\mathbf{G}) = K, \quad (3.9)$$

となる。 $\text{rank}(\mathbf{G})$  は  $\mathbf{G}$  のランクを表し、 $\text{rank}(\mathbf{G}) = K$  は全てのクラスターに少なくとも 1 個体は含まれることを意味する。一方、行列  $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_k, \dots, \mathbf{c}_K]^\top$  は  $K \times M$  の行列であり、 $\mathbf{c}_k$  は  $\mathbf{C}$  の第  $k$  行、すなわちクラスター  $k$  の得点ベクトルを意味し、クラスター  $k$  に所属する全ての個体の因子得点は  $\mathbf{c}_k$  となる。

ここで、 $\mathbf{x}_p$  における異なる値の数  $N_p$  を考える。たとえば、 $\mathbf{x}_p = [-1, 4, 2, -1, 2]^\top$  における異なる値は 1, 2 と 4 なので、 $N_p = 3$  となる。もし  $N_p$  がクラスター数より大きいとき、つまり

$$N_p > K, \quad (3.10)$$

のとき、FCFA では (3.7) は 0 にならない。なぜなら (3.8) 式を (3.7) 式に代入すると

$$\psi_p^2 = \frac{1}{N} \|\mathbf{x}_p - \mathbf{GC}\boldsymbol{\lambda}_p\|_2^2, \quad (3.11)$$

になるからである。このとき、行列  $\mathbf{GC}\boldsymbol{\lambda}_p$  の中の異なる値は、 $\mathbf{c}_1^\top \boldsymbol{\lambda}_p, \dots, \mathbf{c}_K^\top \boldsymbol{\lambda}_p$  の  $K$  個だが、 $\mathbf{x}_p$  は  $N_p$  個の異なる値を持つ。これは、制約として (3.8) 式を加えた (3.7) 式、もしくは 3.11 式は 0 になり得ないということを意味している。よって、(3.10) 式を満たすとき、(3.6) 式は発散しない。 $\mathbf{x}_p$  が様々な実数を含み、クラスターが限られた数のとき、(3.10) 式は通常満たされると考えることができる。しかし、 $N_p \leq K$  となるとき、(3.11) 式は 0 となってしまう。例えば、 $\mathbf{x}_p = [-1, 4, 2, -1, 2]^\top, K = 3$ ,

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{C} = s \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}, \quad \text{and } \boldsymbol{\lambda}_p = \frac{1}{s} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

とすると、 $\mathbf{x}_p = \mathbf{GC}\boldsymbol{\lambda}_p$  となる。ただし  $s \neq 0$  とする。

(3.8) 式を (3.1), (3.4) と (3.5) 式に代入するとそれぞれ以下の 3 つの式に書き換えることができる.

$$\mathbf{X} = \mathbf{GC}\boldsymbol{\Lambda}^\top + \mathbf{E}, \quad (3.12)$$

$$\mathbf{1}_N^\top \mathbf{GC} = \mathbf{0}_M, \quad (3.13)$$

$$\frac{1}{N} \mathbf{C}^\top \mathbf{DC} = \mathbf{I}_M. \quad (3.14)$$

ここで, 行列  $\mathbf{D} = \mathbf{G}^\top \mathbf{G}$  は対角行列とする. さらに FCFA モデル (3.12) 式と (3.2) 式より対数尤度は

$$LL(\mathbf{G}, \mathbf{C}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}) = -N \sum_{p=1}^P \log \psi_p^2 - \text{tr}(\mathbf{X} - \mathbf{GC}\boldsymbol{\Lambda}^\top) \boldsymbol{\Psi}^{-2} (\mathbf{X} - \mathbf{GC}\boldsymbol{\Lambda}^\top)^\top, \quad (3.15)$$

となる. したがって FCFA は, 制約条件 (3.8), (3.13), (3.14) 式および,  $\boldsymbol{\Psi}^2$  は対角行列であるという制約の元, パラメータ  $\mathbf{G}, \mathbf{C}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}^2$  について (3.15) 式を最大化すると定式化される.

(3.8) 式および (3.13) 式では  $\text{rank}(\mathbf{GC}) \leq K-1$  を示しているが, (3.14) 式では  $\text{rank}(\mathbf{GC}) = M$  であることが必要であるため,

$$\text{rank}(\mathbf{GC}) = M \leq K - 1, \quad (3.16)$$

となることを明記する. この不等式 (3.16) 式より,  $K > M$  が導ける. すなわちクラスター数は因子数よりも大きくなければならない.

### 3.3 アルゴリズム

FCFA の問題を解くために, 収束するまで各パラメータ  $\mathbf{G}, \mathbf{C}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}^2$  を順番に更新するアルゴリズムを提案する. 行列  $\mathbf{G}, \mathbf{C}$  と  $\boldsymbol{\Lambda}$  を固定すると, 最適な独自分散  $\psi_p^2$  は (3.11) 式によって得られる. (3.11) 式によって得られた  $\psi_p^2$  を固定したとき,  $\mathbf{G}, \mathbf{C}, \boldsymbol{\Lambda}$  について (3.15) 式を最大化することは次のトレースを  $\mathbf{G}, \mathbf{C}, \boldsymbol{\Lambda}$  について最小化することと等しい.

$$\begin{aligned} \text{tr}(\mathbf{X} - \mathbf{GC}\boldsymbol{\Lambda}^\top) \boldsymbol{\Psi}^{-2} (\mathbf{X} - \mathbf{GC}\boldsymbol{\Lambda}^\top)^\top &= \|\mathbf{X}\boldsymbol{\Psi}^{-1} - \mathbf{GC}\boldsymbol{\Lambda}^\top\boldsymbol{\Psi}^{-1}\|_F^2 \\ &= \text{tr}\mathbf{X}^\top \mathbf{X}\boldsymbol{\Psi}^{-2} + N \text{tr}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top\boldsymbol{\Psi}^{-2} - 2\text{tr}\boldsymbol{\Psi}^{-2}\mathbf{X}^\top \mathbf{GC}\boldsymbol{\Lambda}^\top. \end{aligned} \quad (3.17)$$

したがって,  $\mathbf{G}, \mathbf{C}, \boldsymbol{\Lambda}$  を更新する際は, (3.17) 式の最小化のみを考慮すればよい.

$\boldsymbol{\Lambda}$  の更新は, 簡単に導ける.  $\mathbf{G}, \mathbf{C}$  と  $\boldsymbol{\Psi}$  を固定したとき

$$\boldsymbol{\Lambda} = \frac{1}{N} \mathbf{X}^\top \mathbf{G} \mathbf{C} \quad (3.18)$$

続いて  $\mathbf{G} = \{g_{nk}\}$  の更新するために, (3.8) 式の制約条件の元, (3.17) 式を最小化する.  $\mathbf{X}$  の第  $n$  行を  $\mathbf{x}_n^\top$ ,  $\mathbf{C}$  の第  $k$  列を  $\mathbf{c}_k^\top$  としたとき, (3.17) 式は,  $\sum_n \sum_k g_{nk} \|\mathbf{x}_n^\top \boldsymbol{\Psi}^{-1} - \mathbf{c}_k^\top \boldsymbol{\Lambda}^\top \boldsymbol{\Psi}^{-1}\|_2^2$  と書き換えられ,

$$g_{nk} = \begin{cases} 1 & \text{iff } k = \operatorname{argmin}_{1 \leq l \leq K} \|\mathbf{x}_n^\top \boldsymbol{\Psi}^{-1} - \mathbf{c}_l^\top \boldsymbol{\Lambda}^\top \boldsymbol{\Psi}^{-1}\|_2^2 \\ 0 & \text{otherwise.} \end{cases} \quad (3.19)$$

ここで制約条件 (3.13), (3.14) 式および  $\operatorname{rank}(\mathbf{G}) = K$  は, (3.9) 式では考慮されていないことを考える. これらの制約条件は, FCFA のアルゴリズム中で満たすことが保証されない. よって, あるクラスターにどの個体も所属しない,  $\operatorname{rank}(\mathbf{G}) < K$  となる場合がある. この場合, FCFA モデルがデータセットに適していないことが考えられるため, 計算を中止する. 制約条件 (3.13), (3.14) 式が満たされている場合,  $\mathbf{C}$  を更新するために  $\mathbf{G}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}^2$  を固定して,  $\mathbf{C}$  について (3.17) 式を最小化する必要がある. この最小化は, 関数  $f = \operatorname{tr} \boldsymbol{\Psi}^{-2} \mathbf{X}^\top \mathbf{G} \mathbf{C} \boldsymbol{\Lambda}^\top$  の最大化と等しい.  $\mathbf{M} = N^{\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}} \mathbf{G}^\top \mathbf{X} \boldsymbol{\Psi}^{-2} \boldsymbol{\Lambda}$  としたとき,  $f = \operatorname{tr} \mathbf{M}^\top (N^{-\frac{1}{2}} \mathbf{D}^{\frac{1}{2}} \mathbf{C})$  と書き換えることもできる. この一次形式  $f$  は, 以下の不等式を満たす.

$$f = \operatorname{tr} \mathbf{M} \mathbf{D}^{\frac{1}{2}} \mathbf{C} \leq \operatorname{tr} \boldsymbol{\Delta} \quad (3.20)$$

この不等式は, 制約条件 (3.14) 式を満たすので成り立つ (ten Berge, 1983). 行列  $\boldsymbol{\Delta}$  は特異値分解

$$\mathbf{M} = \mathbf{V} \boldsymbol{\Delta} \mathbf{W}^\top, \quad (3.21)$$

によって得られる. ここで,  $\mathbf{V}^\top \mathbf{V} = \mathbf{W}^\top \mathbf{W} = \mathbf{I}_M$  かつ  $\boldsymbol{\Delta}$  は対角行列である. (3.20) 式における  $\operatorname{tr} \boldsymbol{\Delta}$  の上限は,  $\mathbf{D}^{\frac{1}{2}} \mathbf{C} = N^{\frac{1}{2}} \mathbf{V} \mathbf{W}^\top$

$$\mathbf{C} = N^{\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}} \mathbf{V} \mathbf{W}^\top = N^{\frac{1}{2}} \mathbf{D}^{-\frac{1}{2}} \mathbf{M} \mathbf{W} \boldsymbol{\Delta}^{-1} \mathbf{W}^\top, \quad (3.22)$$

(3.21) 式の  $\mathbf{V} = \mathbf{M} \mathbf{W} \boldsymbol{\Delta}^{-1}$  を使うと, 得られた行列  $\mathbf{C}$  によって  $\mathbf{G} \mathbf{C}$  が制約条件 (3.13) 式を満たしていることについては, 付録にて述べる.

ここで、(3.18) 式を (3.17) 式に代入すると、制約条件 (3.14) 式を用いて  $N(\text{tr}\mathbf{S} - \text{tr}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)$  に書き換えることができる。ただし  $\mathbf{S}$  は共分散行列  $\mathbf{S} = N^{-1}\mathbf{X}^\top\mathbf{X}$  である。したがって、(3.18) 式の更新後の対数尤度 (3.15) 式の値は次のように表現できる。

$$LL(\mathbf{G}, \mathbf{C}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}^2) \propto -N \sum_{p=1}^P \log \psi_p^2 - N(\text{tr}\mathbf{S} - \text{tr}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)\boldsymbol{\Psi}^{-2}. \quad (3.23)$$

ここで右側項は容易に収束しているかどうかを判定できる。FCFA のアルゴリズムは以下である：

- [1]  $\mathbf{G}, \mathbf{C}$  と  $\boldsymbol{\Psi}^2$  を初期化する。
- [2] (3.18) 式によって  $\boldsymbol{\Lambda}$  で更新する。
- [3] 収束していれば終了し、収束していなければ [4] に移動する。
- [4]  $\mathbf{C}, \boldsymbol{\Lambda}$  と  $\boldsymbol{\Psi}^2$  を固定して、 $\mathbf{G}$  を (3.19) 式を更新する。
- [5]  $\mathbf{C}$  を (3.22) 式によって更新する。
- [6]  $\boldsymbol{\Psi}^2$  を (3.11) 式によって更新する。その後 [2] に移動する。

ステップ [3] の収束判定は、(3.23) 式における右側項の値を  $N\text{tr}\mathbf{S}$  で割った値が、前回の更新時と比べた増加量が  $10^{-6}$  未満だったときに収束とする。ステップ [1] の初期化は乱数によって行う。 $\mathbf{G}$  に関しては、各行で 1 となる列を無作為に選び、 $\mathbf{C}$  は各要素を一様分布  $U(-1, 1)$  から抽出する。 $\boldsymbol{\Psi}^2$  の対角要素は、一様分布  $U(0.1, 0.9)$  から抽出する。最適解が局所解となることを避けるために、200 回初期値を発生させてそれぞれ更新して (3.15) 式の値が一番大きいときの解を最適解とする。ここで、 $M \geq 2$  のとき

$$\mathbf{C}\boldsymbol{\Lambda}^\top = \mathbf{C}\mathbf{T}\mathbf{T}^\top\boldsymbol{\Lambda}^\top \quad (3.24)$$

のように得られた  $\mathbf{C}$  と  $\boldsymbol{\Lambda}$  は回転の不定性を有する。行列  $\mathbf{T}$  は  $M \times M$  の直交回転行列で、 $\mathbf{T}^\top\mathbf{T} = \mathbf{T}\mathbf{T}^\top = \mathbf{I}_M$  を満たす。このとき最適解のクラスター中心行列は  $\mathbf{CT}$ 、因子負荷行列は  $\boldsymbol{\Lambda}\mathbf{T}$  となる。ここでは既存の回転手法であるバリマックス回転を用いる。

### 3.4 シミュレーション

推定されたパラメータが真のパラメータを再現できているか、シミュレーションを行う。どのようにシミュレーションを行うかを 4.1 で記述し、その結果を 4.2 で報告する。

### 3.4.1 データの合成と分析

因子数を  $M = 2$ , クラスター数を  $K = 4$  とする. そして, 制約条件 (3.8), (3.10), (3.13), (3.14) 式を満たす FCFA モデル (3.12) 式に従うような  $200 \times 12$  の行列  $\mathbf{X}$  をつくる. (3.12) 式中の真のパラメータと誤差行列については, 以下の通りに生成した.

- [1] 各クラスターに個体数が少なくとも 10 以上となるよう無作為に  $\mathbf{G}$  を生成する.
- [2]  $\mathbf{C}$  の各要素は, 一様分布  $U(-1, 1)$  から抽出したあと,  $\mathbf{GC}$  が (3.13), (3.14) 式を満たすように正規化する.
- [3]  $\Lambda$  の各要素は, 一様分布  $U(-1, 1)$  から抽出し,  $\Psi^2$  の対角要素は  $\mathbf{I}_P - \text{diag}(\Lambda\Lambda^\top) = \Psi^2$  によって生成する. ここで,  $\text{diag}(\Lambda\Lambda^\top)$  は,  $\Lambda\Lambda^\top$  の対角要素からなる対角行列である.  $\Psi^2$  のある対角要素が 0.05 以下のときは,  $\Lambda$  を再生成する.
- [4] (3.4) 式を満たすよう, 誤差行列  $\mathbf{E}$  の各列を生成する.

$\mathbf{X}$  を 500 個作成し,  $M = 2, K = 4$  として FCFA をそれぞれの  $\mathbf{X}$  に適用する. 得られた  $\mathbf{C}$  と  $\Lambda$  には (3.24) 式のように回転の不定性が存在する. ここで,  $\mathbf{T}$  は直交プロクラステス回転によって求める. つまり  $(P + K) \times M$  の行列  $\mathbf{M} = [P^{-1}\Lambda^\top, K^{-1}\mathbf{C}^\top]^\top$  を真値からなる行列  $\mathbf{M}_{\text{true}}$  に最も近づける,  $\mathbf{T}$  について  $\|\mathbf{MT} - \mathbf{M}_{\text{true}}\|_F^2$  を最小化にする  $\mathbf{T}$  を最適解とする. 求められた  $\mathbf{T}$  による  $\mathbf{CT}$  と  $\mathbf{\Lambda T}$  をそれぞれ  $\hat{\mathbf{C}} = (\hat{c}_{kl})$  と  $\hat{\Lambda} = (\hat{\lambda}_{pl})$  のように表現する. 同様に, 得られた  $\mathbf{G}$  と  $\Psi^2$  をそれぞれ  $\hat{\mathbf{G}} = (\hat{g}_{nk})$ ,  $\hat{\Psi}^2 = \text{diag}(\hat{\psi}_1^2, \dots, \hat{\psi}_P^2)$  と表す.

### 3.4.2 結果

表 3.1: AAD の平均値とパーセンタイル

	$\Lambda$	$\Psi^2 \mathbf{1}_P$	$\mathbf{C}$	$\mathbf{G}$
平均値	0.035	0.033	0.035	0.022
中央値	0.035	0.033	0.033	0.013
75 パーセンタイル	0.039	0.039	0.043	0.030
95 パーセンタイル	0.073	0.047	0.070	0.043

真のパラメータをどの程度再現できたかの指標として, 平均絶対差 (AAD) を用いる. 因子負荷行列  $\Lambda$  の AAD は,  $\text{AAD}(\Lambda) = (PM)^{-1} \sum_p \sum_l |\hat{\lambda}_{pl} - \lambda_{pl}|$  で定義される. 同様にメン

バーシップ行列  $\mathbf{G}$ , セントロイド行列  $\mathbf{C}$ , 独自分散行列  $\Psi^2$  の AAD はそれぞれ  $AAD(\mathbf{G}) = (2N)^{-1} \sum_n \sum_k |\hat{g}_{nk} - g_{nk}|$ ,  $AAD(\mathbf{C}) = (KM)^{-1} \sum_k \sum_l |\hat{c}_{kl} - c_{kl}|$ ,  $AAD(\Psi\mathbf{1}_p) = P^{-1} \sum_p |\hat{\psi}_p^2 - \psi_p|$  で定義される。表 3.1 は、500 個のデータセットで得られた AAD の平均値、中央値と二つのパーセンタイルを表示している。表 3.1 の通り、全てのパラメーターにおいてよく再現できている。たとえば、 $AAD(\Lambda)$  が 0.07 と、95% で絶対値が 0.07 以内である。また  $AAD(\mathbf{G})$  の平均値は 0.02 であった。

### 3.5 類似手法

本節では、FCFA と Reduced K-means analysis(RKM), Factorial K-means analysis(FKM) がどのように関係しているかを述べる。RKM と FKM は、FCFA と同様の目的で用いられる。まず RKM と FCFA が非常に密接した関係にあることを示し、次に、FCFA が RKM や FKM と同様にタンデムクラスタリングと呼ばれる手法の欠点を回避できることについて説明する。

#### 3.5.1 Reduced K-means analysis

RKM は、制約つき主成分分析と考えることができる。主成分分析の最小二乗基準における目的関数は、 $\|\mathbf{X} - \mathbf{F}\Lambda^\top\|_F^2$  と書くことができる。制約条件 (3.8) 式をこの目的関数に加えると、

$$f_{RKM}(\mathbf{G}, \mathbf{C}, \Lambda) = \|\mathbf{X} - \mathbf{G}\mathbf{C}\Lambda^\top\|_F^2, \quad (3.25)$$

と表せる。RKM では、この目的関数 (3.25) 式を  $\mathbf{G}, \mathbf{C}, \Lambda$  について制約条件 (3.9), (3.13) 式と

$$\text{rank}(\mathbf{C}\Lambda^\top) \leq \min(K - 1, P) \quad (3.26)$$

の元、最小化する (DeSoete & Carroll, 1994)。 (3.26) 式は FCFA の制約条件 (3.16) 式とは異なるが、 $\text{rank}(\mathbf{G}\mathbf{C}) = M$  のとき一致する。詳細は付録 A.2 で記述する。また (3.25) 式は最小二乗の関数だが、FCFA モデルの (3.15) 式は最尤法の関数である。

よって FCFA と RKM の違いは、

- [1] RKM が最小二乗法による手法なのに対して, FCFA は最尤法による手法であること
- [2] RKM において, 独自分散  $\Psi^2$  は現れない

である. しかし, [1] は重要ではない. なぜなら, (3.2) 式という仮定をわけば,

$$\Psi = \nu \mathbf{I}_P \quad (3.27)$$

という制約を除いて, 最小二乗基準の (3.25) 式と RKM の最尤法は等しくなる. よって, 対数尤度 (3.15) 式を制約条件 (3.27) 式の元で最大化すると, 付録 A.3 で詳しく述べる通り, 最小二乗の RKM で得られた解と同じ解が得られる. これらのことから, RKM は (3.27) 式の制約を加えた FCFA といえる. (3.27) 式を加えるということは, 誤差分散が全ての変数で同一であることを意味する. 一方, FCFA における  $\Psi^2 = \psi_1^2, \dots, \psi_P^2$  は制約のない対角行列である. ある誤差は他の変数の誤差と無相関であり,  $\psi_p^2 (p = 1, \dots, P)$  は変数  $p$  に対して独自の値をとることから  $\psi_p^2$  は独自分散と呼ばれる.

それぞれの変数における分散  $\text{Var}(\mathbf{x}_p)$  は, 以下のようにクラスター因子  $\mathbf{GC}$  によって説明できる部分と説明できない部分  $\psi_p^2$  に分解することができる.

$$\text{Var}(\mathbf{x}_p) = \frac{1}{N} \|\mathbf{GC}\boldsymbol{\lambda}_p\|_2^2 + \psi_p^2. \quad (3.28)$$

これらについては付録 A.4 にて証明する. (3.28) 式より, FCFA における独自分散  $\psi_p^2$  は  $\mathbf{GC}$  によって説明される部分より, 大きな値を取りうる. 一方, RKM では  $\psi_p^2$  には制約 (3.27) が課せられこのようなことを起こり得ない.

### 3.5.2 Factorial K-means analysis

FKM は以下の目的関数

$$f_{\text{FKM}}(\mathbf{W}, \mathbf{G}, \mathbf{C}) = \|\mathbf{XW} - \mathbf{GC}\|_F^2 \quad (3.29)$$

を  $\mathbf{W}, \mathbf{G}, \mathbf{C}$  について最小化する. 制約条件は, (3.9) 式および,  $\mathbf{W}^\top \mathbf{W} = \mathbf{I}_M$  である (Vichi & Kiers, 2001). ここで,  $\mathbf{W}$  は  $P \times M$  の行列で,  $\mathbf{XW}$  は  $\mathbf{X}$  中の  $P$  個の変数が  $M$  個の成分に要約された,  $N \times M$  の成分行列とみなせる. FKM は,  $\mathbf{XW}$  の各行をクラスター化するために, 成分行列  $\mathbf{XW}$  と  $\mathbf{GC}$  を近似させる. しかし RKM と FCFA では, クラスター因子, またはクラスター成分行列  $\mathbf{GC}$  と負荷行列の転置行列  $\mathbf{A}^\top$  の積をデータ行列  $\mathbf{X}$  に近

似させる。この違いによって、RKM の場合とは異なり、FCFA と FKM の直接的に関係しているとすることは難しい。しかし、FCFA と FKM は共に  $\mathbf{GC}$  を求め、FKMにおいて変数と成分の関係性を表す  $\mathbf{W}$  の役割は、FCFAにおいては  $\Lambda$  が担っているので、比較することができる。

### 3.5.3 タンデム分析

RKM と FKM は [1] 変数を説明する成分の抽出と [2] 個体の分類を同時に行う。これら二つの目的は、データ行列  $\mathbf{X}$  に主成分分析を行い、その結果得られた成分行列に対してクラスター分析を行うことでも達成することができる。しかし、タンデム分析と呼ばれるこの手法は批判されることが多い。なぜなら、 $\mathbf{X}$  の成分は  $\mathbf{X}$  のクラスター構造を必ずしも反映しているとは言えないからである。実際、タンデム分析に比べて RKM のほうがより正しく個体を分類できることが経験的に知られている。

同様に、 $\mathbf{X}$  に因子分析を行って得られた因子得点に対してクラスター分析を行う、タンデム分析に比べて FCFA がより分類精度がよいことは自然な発想である。変数モデル因子分析が  $\Lambda$  と  $\Psi^2$  のみをパラメータとし、因子得点  $\mathbf{F}$  は確率変数として扱う。因子得点  $\mathbf{F}$  は、因子分析を行って得た  $\Lambda$  と  $\Psi^2$  を使って計算できるので、因子分析におけるタンデム分析、タンデム因子分析は、次の三つの手順からなる。

- [1] データセットに対して因子分析を行い、 $\Lambda$  と  $\Psi^2$  を得る。
- [2] [1] によって得られた  $\Lambda$  と  $\Psi^2$  により  $\mathbf{F}$  を計算する。
- [3]  $\mathbf{F}$  に対してクラスター分析を行い、 $\mathbf{G}$  と  $\mathbf{C}$  を得る。

次節では、[1] は最尤法、[2] は回帰法、[3] は  $k$ -means を用いた。

## 3.6 実データ解析

本節では、四つの手法、FCFA、RKM、FKM、タンデム因子分析を、事前に個体がどのクラスターに分類されるか知られているデータセットを用いて比較する。用いたデータセットは全て Izenman(2008) のサイト <http://astro.temple.edu/~alan/MMST/datasets.html> で得ることができる。また、全てのデータセットを平均 0、分散 1 に標準化して分析した。

加えて、クラスター数  $K$  は正しい群の数にした。FCFA と比較するために RKM の制約条件を (3.14) 式を用いた。Timmerman(2010)などでは、(3.14) 式ではなく  $\Lambda$  に直交制約  $\Lambda^\top \Lambda = \mathbf{I}_M$  を設けているが、付録で述べる通り分類結果は変わらないため問題ない。FKM の制約条件  $\mathbf{W}^\top \mathbf{W} = \mathbf{I}_M$  は、代わりとなる制約条件をみつけることができなかった。

### 3.6.1 あやめデータ

表 3.2: あやめデータに対する主成分負荷量と因子負荷量

	タンデム		FCFA		RKM	FKM
	$\Lambda$	$\psi_p^2$	$\Lambda$	$\psi_p^2$	$\Lambda$	$\mathbf{W}$
S.L	0.87	0.24	0.80	0.36	0.81	0.28
S.W	-0.42	0.82	-0.48	0.76	-0.52	-0.14
P.L	0.99	0.01	0.97	0.06	0.97	-0.81
P.W	0.96	0.07	0.96	0.07	0.95	0.50
	$\mathbf{C}$		$\mathbf{C}$		$\mathbf{C}$	$\mathbf{C}$
SE <sup>1</sup>	-1.30		-1.32		-1.34	0.01
VE <sup>2</sup>	0.30		0.26		0.30	0.18
VP <sup>3</sup>	1.07		1.12		1.06	-0.19

表 3.3: あやめデータにおける真のクラスターと結果得られたクラスターの分割表

	Cluster	タンデム			FCFA			RKM			FKM		
		SE	VE	VI	SE	VE	VI	SE	VE	VI	SE	VE	VI
True	SE	<b>50</b>	0	0	<b>50</b>	0	0	<b>50</b>	0	0	<b>31</b>	6	13
	VE	0	<b>48</b>	2	0	<b>49</b>	1	0	<b>45</b>	5	26	<b>16</b>	8
	VI	0	7	<b>43</b>	0	4	<b>46</b>	0	6	<b>44</b>	21	16	13
CCP		94			97			93			40		

第一の例は、有名な Fisher(1936) のあやめデータを用いる。あやめデータは、150 の花びらと 4 つの指標からなるデータで、それぞれの花びらは三つのグループに分けられる。このデータは判別精度を試すために広く用いられている。変数が  $P = 4$  と少数なため、 $M = 1$  とした。表には、 $\Lambda, \mathbf{W}, \mathbf{C}$  と、タンデム因子分析および FCFA における  $\psi_p^2$  が書か

<sup>1</sup>Setosa

<sup>2</sup>Versicolor

<sup>3</sup>Virginica

れている。表3.2の通り、FCFAとタンデム因子分析、RKMは似通った負荷行列を持つが、FKMの重み行列は異なる形をしている。この傾向は、クラスター得点行列  $\mathbf{C}$ においても見られる。各手法で得られたメンバーシップ行列  $\mathbf{G}$  の分類精度は、表3.3に示している。表3.3は、行が真のクラスター、列が手法によって得たクラスターの分割表である。分割表の対角マスは、正しく分類された個体の数を表す。例えば、表3.3ではFCFAは真のクラスターVIにおいて46個体が正しく分類されたが、4個体はVEに誤って分類されている。また正分類率は、 $100 \times (50 + 49 + 46) / (50 + 50 + 50) = 97\%$ となる。表3.3の通り、FCFAが最も正分類率が高く、反対に最も正分類率の低い手法はFKMであった。

### 3.6.2 糖尿病データ

表3.4: 糖尿病データに対する主成分負荷量と因子負荷量

	タンデム			FCFA			RKM		FKM	
	$\Lambda$		$\psi_p^2$	$\Lambda$		$\psi_p^2$	$\Lambda$		$\mathbf{W}$	
glucose.area	<b>0.56</b>	0.09	0.67	<b>0.46</b>	0.13	0.77	<b>0.64</b>	0.17	0.21	-0.02
Insulin.area	-0.18	<b>0.97</b>	0.02	-0.21	<b>0.89</b>	0.16	-0.20	<b>0.89</b>	<b>0.39</b>	<b>0.67</b>
SSPG	-0.09	<b>0.98</b>	0.04	-0.08	<b>0.91</b>	0.16	-0.11	<b>0.89</b>	0.26	<b>-0.73</b>
Relative.weight	<b>0.49</b>	-0.31	0.66	<b>0.54</b>	-0.22	0.65	<b>0.53</b>	-0.24	-0.07	0.02
f.p.glucose	<b>0.54</b>	<b>0.84</b>	0.01	0.31	<b>0.83</b>	0.20	0.30	<b>0.80</b>	<b>-0.86</b>	0.08
	<b>C</b>			<b>C</b>			<b>C</b>		<b>C</b>	
O <sup>4</sup>	-0.81	2.45		-0.89	2.33		-0.79	2.37	-0.60	-0.04
C <sup>5</sup>	1.08	0.13		1.46	0.13		1.28	0.01	-0.04	0.04
N <sup>6</sup>	-0.55	-0.54		-0.62	-0.67		-0.78	-0.68	0.47	-0.01

表3.5: 糖尿病データにおける真のクラスターと結果得られたクラスターの分割表

	Cluster	タンデム			FCFA			RKM			FKM		
		O	C	N	O	C	N	O	C	N	O	C	N
True	O	<b>15</b>	14	4	<b>20</b>	12	1	<b>20</b>	10	3	<b>13</b>	10	10
	C	0	<b>26</b>	10	0	<b>27</b>	9	0	<b>29</b>	7	17	<b>14</b>	5
	N	0	11	<b>65</b>	0	7	<b>69</b>	0	16	<b>60</b>	10	24	<b>42</b>
CCP		73			80			75			48		

二つめのデータは、糖尿病に関するデータで、<http://archive.ics.uci.edu/ml/datasets/>

<sup>4</sup>Overt diabetic

<sup>5</sup>Chemical diabetic

<sup>6</sup>Normal

Diabetes で手に入れることができる。データは 145 人の被験者の 5 つの指標からなる。因子数は、 $M = 2$  とし  $\Lambda, \mathbf{W}$  にはそれぞれバリマックス回転を施した。表 3.4 には、得られた  $\Lambda, \mathbf{W}, \mathbf{C}$  と、FCFA とタンデム因子分析によって得られた  $\psi_p^2$  を載せている。あやめデータと同様、FCFA とタンデム因子分析、RKM の負荷行列は近しいのに対し、FKM の重み行列は全く違う構造をしている。各手法の分類精度に関しては、表 3.5 に記しており、書き方は表 3.3 に準拠している。表 3.5 の通り、またしても FCFA が正分類率が最も高く、FKM が正分類率が最も低かった。

### 3.6.3 霊長類肩甲骨データ

表 3.6: 霊長類肩甲骨データに対する主成分負荷量と因子負荷量

	タンデム			FCFA			RKM		FKM	
	$\Lambda$		$\psi_p^2$	$\Lambda$		$\psi_p^2$	$\Lambda$		$\mathbf{W}$	
AD.BD	<b>0.82</b>	<b>0.52</b>	0.01	<b>0.75</b>	<b>0.59</b>	0.07	<b>0.72</b>	<b>0.59</b>	-0.15	<b>0.78</b>
AD.CD	<b>0.85</b>	0.04	0.28	<b>0.91</b>	-0.04	0.16	<b>0.91</b>	0.01	<b>0.63</b>	0.03
EX.CD	<b>-0.96</b>	-0.02	0.07	<b>-0.85</b>	0.16	0.24	<b>-0.84</b>	-0.15	0.04	<b>0.48</b>
Dx.CD	0.01	<b>0.54</b>	0.71	-0.06	<b>0.61</b>	0.62	-0.15	<b>0.69</b>	0.11	-0.10
SH.ACR	-0.24	-0.14	0.92	-0.26	-0.17	0.89	-0.19	-0.37	-0.19	-0.01
EAD	-0.09	<b>-0.95</b>	0.09	-0.13	<b>-0.80</b>	0.33	-0.08	<b>-0.80</b>	0.09	<b>0.37</b>
beta	<b>-0.82</b>	0.06	0.32	<b>-0.91</b>	0.13	0.15	<b>-0.90</b>	0.03	<b>0.72</b>	0.08
	<b>C</b>			<b>C</b>			<b>C</b>		<b>C</b>	
HY <sup>7</sup>	1.18	-1.15		1.68	-1.37		1.72	-1.14	-0.52	0.11
PO <sup>8</sup>	-0.66	-1.03		-0.62	-1.28		-0.38	-1.60	0.18	-0.16
PA <sup>9</sup>	0.98	0.57		0.88	1.27		0.66	1.10	-0.23	-0.02
GO <sup>10</sup>	0.59	1.31		0.43	0.85		-1.14	0.64	-0.11	0.24
HO <sup>11</sup>	-0.99	0.26		-1.01	0.24		-0.91	-0.40	0.55	-0.13

表 3.7: 霊長類肩甲骨データにおける真のクラスターと結果得られたクラスターの分割表

		タンデム					FCFA					RKM					FKM					
		HY	PO	PA	GO	HO	HY	PO	PA	GO	HO	HY	PO	PA	GO	HO	HY	PO	PA	GO	HO	
True	HY	<b>16</b>	0	0	0	0	<b>15</b>	1	0	0	0	<b>15</b>	1	0	0	0	<b>6</b>	2	2	4	2	
	PO	1	<b>12</b>	0	0	2	0	<b>15</b>	0	0	0	0	<b>8</b>	0	0	7	4	<b>3</b>	6	1	1	
	PA	0	0	<b>12</b>	8	0	0	0	<b>17</b>	3	0	0	0	<b>20</b>	0	0	0	7	0	<b>5</b>	8	0
	GO	0	0	4	<b>10</b>	0	0	0	3	<b>11</b>	0	0	0	14	<b>0</b>	0	0	2	1	3	<b>8</b>	0
	HO	0	9	0	0	<b>31</b>	0	4	0	0	<b>36</b>	0	4	0	17	<b>19</b>	0	17	3	1	<b>19</b>	
CCP		77					90					59					39					

三つ目のデータは靈長類の肩甲骨についてのデータで、105個体7変数のデータである。因子数を  $M = 2$  と  $M = 3$  で分析したが、ここでは  $M = 2$  のときの結果のみを表3.6, 3.7に記す。 $\Lambda$  と  $\mathbf{W}$  にはバリマックス回転を行った。表3.6の通り、FCFAとタンデム因子分析、RKMは似た負荷行列を持つが、FKMの重み行列は全く異なる。

表3.7は、各手法の分類精度に関する分割表であり、書き方は表3.3, 3.5と同様である。これまでのデータ同様、FCFAが最も正分類率が高く、FKMは正分類率が低かった。この傾向は  $M = 3$  の場合でも同様で、FCFAの正分類率が最も高く90%，FKMの正分類率は最も低く39%であった。タンデム因子分析とRKMの正分類率は、それぞれ77%，59%であった。

### 3.6.4 独自分散と群間分散の関係

本節で紹介した全てのデータセットで、FCFAは最も正分類率が高く、FKMは正分類率が最も低かった。FKMの正分類率が低い理由は、本論文の主題から外れるので省略する。FCFAがタンデム因子分析を上回っている理由は、RKMがタンデム主成分分析より分類精度が良い理由と同じ理由であり(De Soete & Carroll, 1994)、因子分析によって得られた因子得点は、 $\mathbf{X}$ のクラスター構造を必ずしも反映しているとは限らないからである。

ここで注目すべきは、FCFAが正分類率の高さでRKMを上回っている点である。この原因は、3.5節で述べたRKMとFCFAの違いにある。(3.28)式の通り、FCFAにおける独自分散  $\psi_p^2$  は  $\mathbf{GC}$  によって説明できない変数、すなわちクラスター構造にあまり関係していない変数において高くなる。しかし、RKMでは(3.27)式のようにそのようなことは起こらない。つまり、FCFAの高い正分類率は独自分散によるものであると導ける。FCFAの独自分散  $\psi_p^2$  と群間分散とに強い負の相関関係があることから推し量ることができる。個体の分離度を表す指標として、 $\mathbf{X}$ の各変数における比、群間分散/全体分散を用いることとする。群間分散とは、真のクラスター間の二乗和を意味し、全体分散は、二乗和の合計を意味する。図3.1は、実データでの独自分散と群間分散/全体分散からなる散布図である。図の通り、例に挙げた全てのデータで強い負の相関が確認できる。

---

<sup>7</sup>Hylobates

<sup>8</sup>Pongo

<sup>9</sup>Pan

<sup>10</sup>Gorilla

<sup>11</sup>Homo

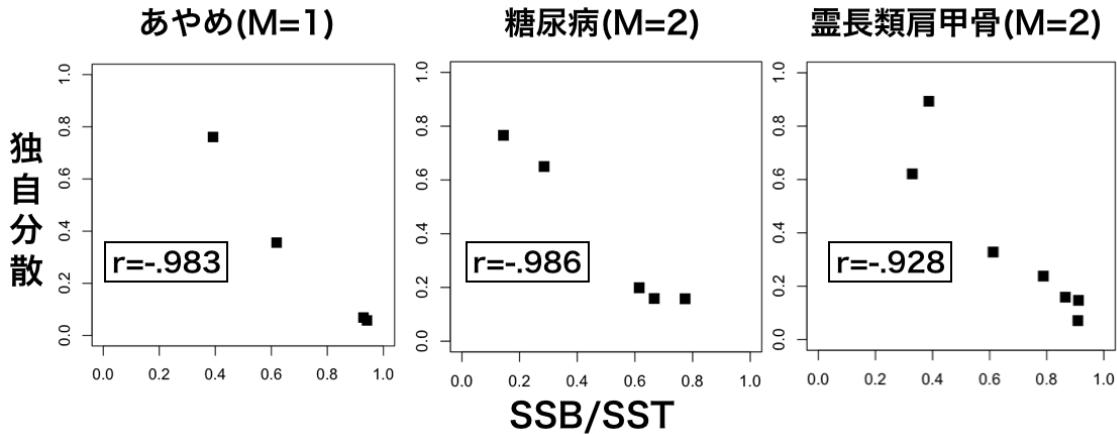


図 3.1: 分離度と独自分散による散布図

独自分散のこの役割は、タンデム因子分析が RKM より糖尿病データを除くと正分類率が上回っていることとも関係しているかもしれない。とすれば、タンデム因子分析で用いられる通常の因子分析では (3.8) 式のようなクラスター構造を仮定していないにもかかわらず、独自分散が正分類率を高くしている可能性がある。しかし、結果についての深い考察は本論文の目的からは外れるので省略する。

### 3.7 考察

母数モデル因子分析において、各個体の因子得点はパラメータとして扱われるが、他のパラメータである因子負荷行列、独自分散と同時に推定することはできない。そこで全てのパラメータを同時に推定するために、個体は少数のクラスターに所属し、同一のクラスターに属する個体は因子得点が等しいという制約を加えたモデルを提案した。この制約つき因子分析を Fixed Clustered Factor Analysis (FCFA) と名付けた。シミュレーションによって、真のパラメータをよく再現できることが確認された。

3.5 節では、FCFA がどのように RKM と FKM に関係しているかを述べた。特に RKM は、独自分散が変数に関わらず同一、つまり各変数に独自ではないという制約を加えた FCFA の特殊形とみなせることを示した。また実データの例を使って、FCFA が RKM や FKM、タンデム因子分析に分類精度の点で上回っていたことを示した。この結果から、FCFA が変数を説明する因子の抽出と個体の分類をするために有用であることがわかる。

FCFA が RKM より分類精度が良い理由には、独自分散が深く関わっていると考えられ

る。FCFA では、クラスタリングにあまり寄与していない変数に対して独自分散が高くなり、得られた計算結果からも確かめることができる。独自分散の利点に関する理論的研究は、今後の課題である。

FCFA は RKM と FKM より分類精度がよかつたが、後者の二つの手法にはない制限がある。(3.16) 式のよう FCFA では  $K > M$ 、つまりクラスター数は因子数より大きくなければならないという制約があったが、RKM と FKM にはこのような制約はない。この制約は、因子得点の直交制約(3.5)式あるいは(3.14)式が起因している。つまり、列直交の  $K \times M$  の因子得点行列  $\mathbf{F} = \mathbf{GC}$  は  $K > M$  を暗に意味している。そこで、因子得点の直交性を緩める、つまり  $K \leq M$  を許すような FCFA の拡張が、今後の課題として検討される。

もう一つの FCFA の制限は、誤差の分布に正規分布を仮定していることである。前述の RKM と FCFA の比較の際は、RKM を正規分布に従う最尤法として考えたが、最小二乗法による RKM や FKM には分布の仮定は必要ない。よって、本来は比較のためには最小二乗法によるクラスタリングを伴う因子分析を考えるべきだ。しかし最小二乗法の因子分析は、共分散行列にモデル部分  $\Lambda\Lambda^\top$  を近似するので、クラスター化されるべき  $\mathbf{F}$  は考慮されないので、最小二乗法によるクラスタリングを伴う因子分析(Adachi, 2015; 足立, 2015; Harman & Jones, 1966)は難しいことがわかる。ただし近年新しく提案された行列モデル因子分析は(Adachi, 2012; Unkel & Trendafilov, 2010)、最小二乗法の枠組みで  $\mathbf{F}$  を含むモデル部分をデータ行列  $\mathbf{X}$  に近似させる手法なので、FCFA の最小二乗モデルに役立つ可能性がある。

FCFA は因子分析における重要な不定性、因子得点の不定性の無い特殊なモデルである。したがって、推定されたパラメータを用いて解釈を行うことに特別の注意を払う必要はない。一方で、回転の不定性という問題は残したままとなっている。回転法を適用することなく因子負荷量の解釈を行うことはできず、さらにどのような回転基準を利用したかによって、因子負荷量の値は変化してしまう。近年の統計学の発展により、因子負荷量に 0 の要素が多くあることを仮定しておくことで、回転の不定性を解消できることが確認されている。そのため、FCFA における回転の不定性も同様に解消可能であることが示唆される。

## 第4章

# 行列モデル因子分析の因子得点同定

本章の内容は, Uno, Adachi, & Trendafilov (2018) を拡張したものである.

### 4.1 導入

因子分析において,  $N$  個体  $\times P$  変数の列中心化されたデータ行列は,  $\mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{U}\boldsymbol{\Psi}$  のようにモデル化される. ただし  $\mathbf{F}$  は  $N$  個体  $\times M$  因子の共通因子得点行列,  $\boldsymbol{\Lambda}$  は  $P$  変数  $\times M$  因子の因子負荷行列,  $\mathbf{U}$  は  $N$  個体  $\times P$  変数の独自因子得点行列,  $\boldsymbol{\Psi}$  は  $P$  変数  $\times P$  変数の対角行列でその対角要素の二乗は独自分散と呼ばれる. そして  $M < P < N$  を想定している (Harman, 1976; Mulaik, 2010). 不等式  $M < P$  は,  $\mathbf{F}\boldsymbol{\Lambda}^\top$  が  $P$  個の変数をより少數の共通因子によって説明していることを意味している. 一方で  $\mathbf{U}\boldsymbol{\Psi}$  の各列はそれぞれに対応する変数のみ特有に寄与している.

因子分析には, 回転の不定性と因子得点の不定性という 2 つの不定性があることが知られている (Mulaik, 2010). 本研究では因子得点の不定性を用いて, 解釈のしやすい解を得ることを目指す. ここで, 因子得点の不定性とは  $\mathbf{F}$  と  $\mathbf{U}$  の共通因子得点と独自因子得点が一意には定められないことを指す. より詳しく述べると,  $\mathbf{F}$  と  $\mathbf{U}$  からなる  $n \times (m+p)$  のブロック行列  $[\mathbf{F}, \mathbf{U}]$  は,

$$[\mathbf{F}, \mathbf{U}] = [\mathbf{F}_d, \mathbf{U}_d] + [\mathbf{F}_u, \mathbf{U}_u] \quad (4.1)$$

のように 2 つの項の和で表すことができる. ただし  $\mathbf{F}_d$  と  $\mathbf{U}_d$  はそれぞれ共通因子得点と独自因子得点の一意に定まる部分を表す一方で,  $\mathbf{F}_u$  と  $\mathbf{U}_u$  は共通因子得点と独自因子得点の一意に定まらない部分を表す (Gutmann, 1955; Mulaik, 1976; Stegeman, 2016).

従来の因子分析では,  $\mathbf{F}$  と  $\mathbf{U}$  における因子得点を確率変数として扱っている. 分布の仮定を用いることで,  $\mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{U}\boldsymbol{\Psi}$  の共分散行列は  $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2$  となる. この共分散行列を, 標本共分散行列  $\mathbf{S}_{\mathbf{XX}} = N^{-1}\mathbf{X}^\top\mathbf{X}$  に近づけることで因子分析の解を得る. 最小二乗基準で

は  $\|\mathbf{S}_{\mathbf{XX}} - (\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2)\|_F^2$  を最小化することで近づけ, 尤度基準では  $\log(|\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2|) + \text{trace}((\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2)^{-1}\mathbf{S}_{\mathbf{XX}})$  を最小化することで近づける. 共分散行列を用いることにより  $\mathbf{F}$  と  $\mathbf{U}$  は考慮する必要がなく,  $\boldsymbol{\Lambda}$  と  $\boldsymbol{\Psi}^2$  の推定のみを考慮すればよい. 因子得点の推定については,  $\boldsymbol{\Lambda}$  と  $\boldsymbol{\Psi}^2$  を推定したのち  $\mathbf{F}$  の近似値を推定する2段階の推定方法がある (Harman, 1976; Mulaik, 2010).

しかし近年提案された新しい因子分析の推定方法では,  $\boldsymbol{\Lambda}$  と  $\boldsymbol{\Psi}^2$  だけでなく共通因子得点と独自因子得点, つまり  $\mathbf{F}$  と  $\mathbf{U}$  もパラメータとして推定することができる. Sočan(2003)で述べられているように Kiers によって最初に提案された後, 様々な研究者によって研究がなされている (Adachi & Trendafilov, 2018; Stegeman, 2016, Trendafilov & Unkel, 2011; Unkel & Trendafilov, 2010). この新しい因子分析では,  $\mathbf{F}, \boldsymbol{\Lambda}, \mathbf{U}, \boldsymbol{\Psi}$  の全てがパラメータとして扱われ, モデル部分  $\mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{U}\boldsymbol{\Psi}$  をデータ行列に最小二乗基準の意味で直接近づけることができる. 本論文では, 因子分析を行列分解問題として捉えていることから, このモデルを行列モデル因子分析と呼ぶ. 行列モデル因子分析については次の節で詳しく述べるが, 行列モデル因子分析では, 最適な  $[\mathbf{F}, \mathbf{U}]$  を推定することはできるものの, 一意に定めることはできない. 正確には, (4.1) 式における  $[\mathbf{F}_d, \mathbf{U}_d]$  のみが一意に定めることができ,  $[\mathbf{F}_u, \mathbf{U}_u]$  は一意に定めることができない.

これらの問題に対処するため, 以下の2ステップによる因子得点行列  $[\mathbf{F}, \mathbf{U}]$  を同定する手法を考える:

- [1] 行列モデル因子分析によりデータから, (4.1) 式における  $[\mathbf{F}_d, \mathbf{U}_d]$  を得る.
- [2] 既知の  $[\mathbf{F}_d, \mathbf{U}_d]$  から, なにか「望ましい特徴」を持つ  $[\mathbf{F}, \mathbf{U}]$  が得られるように, (4.1) 式における一意でない部分  $[\mathbf{F}_u, \mathbf{U}_u]$  を推定する.

本論文では [2] のステップについて述べる. また「望ましい特徴」として, (4.1) 式の因子得点が解釈しやすさを考える. 特に (4.1) 式における共通因子得点  $\mathbf{F}$  の  $N$  行(つまり個体)が, 少数のグループにまとまっていることを望ましいとした. このような  $\mathbf{F}$  は, 共通因子得点の特徴を捉える際に少数のグループに着目すればよいため, 便利である. 個体のクラスタリングは重要なテーマであり, 多変量データ分析における大きな関心のうちの一つである (Adachi, 2016; Gan, Ma, & Wu, 2007). この部では, 提案手法のことを Clustered Common Factor Identification の頭文字から CCFI と呼ぶことにする.

この章における構成は以下の通りである. 次の節では, 上記の [1] に対応する行列モデル因子分析とその解について述べる. その後, [2] のための CCFI を 4.3 節で定式化し, 4.4

節でCCFIのアルゴリズムを提案する。クラスター数の選択については4.5節で考察する。その後の節では、シミュレーションデータおよび実データへのアルゴリズムの有用性を検証するために数値実験を行う。

## 4.2 行列モデル因子分析

行列モデル因子分析は、最小二乗目的関数

$$\|\mathbf{X} - (\mathbf{F}\boldsymbol{\Lambda}^\top + \mathbf{U}\boldsymbol{\Psi})\|_F^2 = \|\mathbf{X} - [\mathbf{F}, \mathbf{U}][\boldsymbol{\Lambda}, \boldsymbol{\Psi}]^\top\|_F^2 \quad (4.2)$$

を因子得点行列  $[\mathbf{F}, \mathbf{U}]$  と共変量行列  $[\boldsymbol{\Lambda}, \boldsymbol{\Psi}]$  について、以下の制約を満たしたうえで最小化する。

$$\mathbf{1}_N^\top [\mathbf{F}, \mathbf{U}] = \mathbf{0}_{M+P}^\top \quad (4.3)$$

$$N^{-1}[\mathbf{F}, \mathbf{U}]^\top [\mathbf{F}, \mathbf{U}] = \mathbf{I}_{M+P} \quad (4.4)$$

$\boldsymbol{\Psi}$  : 対角行列

$$\text{rank}(\boldsymbol{\Lambda}) = M$$

最小化について、共変量ステップと因子得点ステップを収束するまで交互に繰り返す。目的関数(4.2)式が前回の反復からの減少値が  $10^{-5}\|\mathbf{X}\|_F^2$  よりも小さいとき、収束したと判断する。

共変量ステップでは、最適な  $[\boldsymbol{\Lambda}, \boldsymbol{\Psi}]$  を得るために、 $[\mathbf{F}, \mathbf{U}]$  は固定したうえで(4.2)式を最小化する。解は  $\boldsymbol{\Lambda} = N^{-1}\mathbf{X}^\top \mathbf{F}$  と  $\boldsymbol{\Psi} = N^{-1}\text{diag}(\mathbf{X}^\top \mathbf{U})$  によって得られる。ここで  $\text{diag}(\mathbf{X}^\top \mathbf{U})$  は  $\mathbf{X}^\top \mathbf{U}$  の対角要素を要素とする対角行列を表す。

因子得点ステップでは、(4.2)式は制約条件(4.3)式と(4.4)式のもとで  $[\mathbf{F}, \mathbf{U}]$  について最小化する。これによって

$$[\mathbf{F}, \mathbf{U}] = N^{1/2} \mathbf{K} \mathbf{L}^\top + N^{1/2} \mathbf{K}_\perp \mathbf{L}_\perp^\top \quad (4.5)$$

を得る。(4.5)式における  $\mathbf{K}(N \times P)$  と  $\mathbf{L}((M+P) \times P)$  は行列  $N^{-1/2}\mathbf{X}[\boldsymbol{\Lambda}, \boldsymbol{\Psi}]$  の特異値分解

$$N^{-1/2}\mathbf{X}[\boldsymbol{\Lambda}, \boldsymbol{\Psi}] = \mathbf{K} \boldsymbol{\Delta} \mathbf{L}^\top$$

から得られる。ただし

$$\mathbf{K}^\top \mathbf{K} = \mathbf{L}^\top \mathbf{L} = \mathbf{I}_P, \quad \mathbf{1}_N^\top \mathbf{K} = \mathbf{0}_P^\top \quad (4.6)$$

を満たし、 $\Delta$ は $P \times P$ の対角行列である。 $N \times M$ の行列 $\mathbf{K}_\perp$ と $(M+P) \times M$ の行列 $\mathbf{L}_\perp$ は制約条件

$$\mathbf{K}_\perp^\top \mathbf{K}_\perp = \mathbf{L}_\perp^\top \mathbf{L}_\perp = \mathbf{I}_M, \quad \mathbf{K}^\top \mathbf{K}_\perp = \mathbf{L}^\top \mathbf{L}_\perp = \mathbf{O}_{P \times M}, \quad \mathbf{1}_N^\top \mathbf{K}_\perp = \mathbf{0}_M^\top \quad (4.7)$$

を満たす。ただし $\mathbf{O}_{P \times M}$ は要素が全て0である $P \times M$ のゼロ行列を表す。Adachi(2012)にて証明されているとおり、 $\mathbf{X}$ は列中心化されているため(4.6)式の $\mathbf{1}_N^\top \mathbf{K} = \mathbf{0}_P$ と(4.7)式の $\mathbf{1}_N^\top \mathbf{K}_\perp = \mathbf{0}_M$ は成り立つ。

この節では、 $\mathbf{X}[\Lambda, \Psi]$ のランクは $P$ と想定する、つまり $\text{rank}(\mathbf{X}[\Lambda, \Psi]) = P$ 。これは $N^{1/2}\mathbf{KL}^\top$ が一意に定まることを意味している。一方で、制約条件(4.7)式を満たす $\mathbf{K}_\perp$ と $\mathbf{L}_\perp$ が一意に定まらない。同様に $N^{1/2}\mathbf{K}_\perp \mathbf{L}_\perp^\top$ も一意に定まらない。これらの事実を考慮し(4.1)式を(4.5)式と比較すると(4.1)式の一意に定まる部分と定まらない部分をそれぞれ

$$[\mathbf{F}_d, \mathbf{U}_d] = N^{1/2}\mathbf{KL}^\top = \mathbf{X}[\Lambda, \Psi]\mathbf{L}\Delta\mathbf{L}^\top \quad (4.8)$$

$$[\mathbf{F}_u, \mathbf{U}_u] = N^{1/2}\mathbf{K}_\perp \mathbf{L}_\perp^\top \quad (4.9)$$

と表現することができる。ここで(4.8)式の最後の等式は、特異値分解 $N^{-1/2}\mathbf{X}[\Lambda, \Psi] = \mathbf{K}\Delta\mathbf{L}^\top$ から得られる。

Adachi & Trendafilov (2017)は、(4.5)式、(4.8)式、(4.9)式の関係を図4.1のように幾何表現を用いて示した。ただし図における矢印は、行列を表す。

最適な因子得点行列 $[\mathbf{F}, \mathbf{U}]$ が、 $[\mathbf{F}_d, \mathbf{U}_d]$ を中心点とした円錐を形成することが、図4.1からわかる。ここで中心点から $[\mathbf{F}, \mathbf{U}]$ までの距離は $(NM)^{1/2}$ である。これは(4.1)式と(4.7)式、および

$$\begin{aligned} \|[\mathbf{F}, \mathbf{U}] - [\mathbf{F}_d, \mathbf{U}_d]\|_F^2 &= \|[\mathbf{F}_u, \mathbf{U}_u]\|_F^2 \\ &= \|N^{1/2}\mathbf{K}_\perp \mathbf{L}_\perp^\top\|_F^2 \\ &= N\|\mathbf{L}_\perp^\top\|_F^2 = N\text{tr}(\mathbf{L}_\perp^\top \mathbf{L}_\perp) = NM \end{aligned}$$

から従う。図4.1における円周に位置する任意の $[\mathbf{F}, \mathbf{U}]$ はいずれも最適解である。

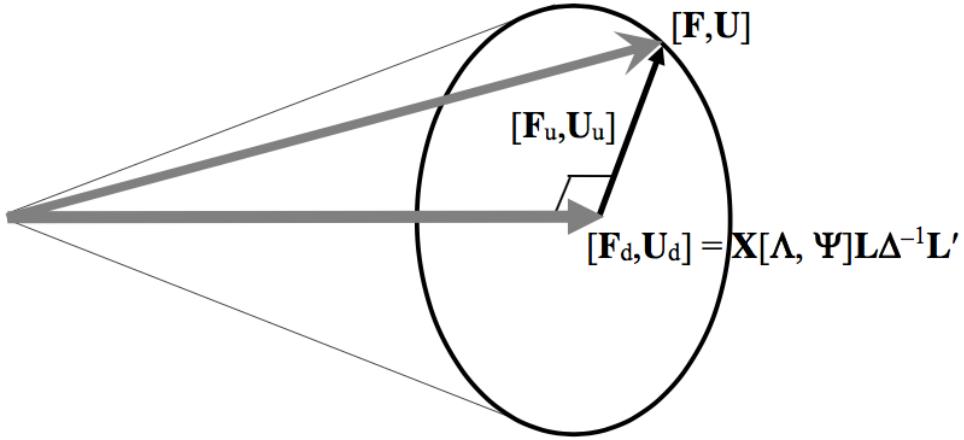


図 4.1: Adachi &amp; Trendafilov(2017) の因子得点の円錐

4.1 節で述べたように、古典的な因子分析は  $\Lambda\Lambda^\top + \Psi^2$  を  $S_{XX} = N^{-1}\mathbf{X}^\top\mathbf{X}$  に近似させており、 $\mathbf{F}$  と  $\mathbf{U}$  は計算過程から除外されている。このことは、(4.5) 式における最適な因子得点の探索を考慮していないことを意味している。代わりに、共通因子得点  $\mathbf{F}$  は  $\Lambda$  と  $\Psi^2$  が得られた後に推定される。いくつか  $\mathbf{F}$  の 2 段階推定法が提案されているが、いずれも  $\mathbf{XW}$  という形をとっている。つまり  $\mathbf{X}$  の線形関数である (Anderson & Rubin, 1956; Bartlett, 1937; Krijnen, Wansbeek, & ten Berge, 1996; Thurstone, 1935)。これは  $\mathbf{X}$  の列空間を  $Sp(\mathbf{X})$  とすると、 $Sp(\mathbf{F})$  が  $Sp(\mathbf{X})$  に必ず含まれることを意味している。ところが図 4.1 で明らかなように、 $\mathbf{F}_d$  のみが  $(\mathbf{XW})$  の形をとるため、 $Sp(\mathbf{F}) \not\subset Sp(\mathbf{X})$  である。 $Sp(\mathbf{XW})$  と表すことができる  $Sp(\mathbf{X})$  に含まれる空間は、 $Sp(\mathbf{F}_d)$  ではあるが、 $Sp(\mathbf{F})$  ではない。以上の議論から古典的な因子分析によって得られる共通因子得点行列は、 $\mathbf{F}$  というよりも一意に定まる  $\mathbf{F}_d$  の推定値ということがわかる。一方で我々は、 $\mathbf{F}_d$  を含む共通因子得点行列  $\mathbf{F}$  だけでなく独自因子得点  $\mathbf{U}$  も同時に同定することを目指したい。この目標は、次節で定式化する提案手法である CCFI によって達成される。

### 4.3 提案手法

CCFI の目的は、 $[\mathbf{F}, \mathbf{U}]$  における共通因子得点  $\mathbf{F}$  がなんらかの望ましい特徴を有しているように (4.9) 式における一意に定められない部分を同定することである。この目的を言

い換えれば、図 4.1 に描かれた円周上に位置する望ましい特徴をもつ  $[\mathbf{F}, \mathbf{U}]$  を見つけることである。本研究では望ましい特徴として、4.1 節で述べたように  $\mathbf{F}$  の  $N$  行が少数のグループにできる限りクラスター化されていることとした。

$C_k (k = 1, \dots, K)$  を  $K \ll N$  としたときのクラスターの集合とする。クラスター化された  $\mathbf{F}$  の特徴は以下で定義する:  $C_k$  に含まれる  $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_N]^\top (N \times M)$  の行は、クラスター  $C_k$  で表現される  $1 \times M$  の列ベクトル  $\mathbf{c}_k^\top$  に最も近い。これは、 $\mathbf{F}$  を  $\mathbf{GC}$  に近似することを意味している。ここで、 $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_K]^\top$  は、 $k$  列が  $\mathbf{c}_k^\top$  の  $K$  クラスター  $\times M$  因子の行列であり、 $\mathbf{G} = (g_{nk})$  は  $N \times K$  行列は  $N$  個体  $\times K$  クラスターのメンバーシップの指示行列であり、

$$g_{nk} = \begin{cases} 1 & (\text{個体 } n \text{ が } C_k \text{ に所属する}) \\ 0 & (\text{それ以外}) \end{cases} \quad (4.10)$$

である。この近似は

$$\|[\mathbf{GC}, \mathbf{U}_d] - [\mathbf{F}, \mathbf{U}]\|_F^2 = \|[\mathbf{GC}, \mathbf{U}_d] - ([\mathbf{F}_d, \mathbf{U}_d] + [\mathbf{F}_u, \mathbf{U}_u])\|_F^2 \quad (4.11)$$

の (4.8) 式において既に定まっている  $[\mathbf{F}_d, \mathbf{U}_d]$  を固定したときに、 $\mathbf{G}$ ,  $\mathbf{C}$ ,  $[\mathbf{F}_u, \mathbf{U}_u]$  について最小化することで達成される。ここで  $[\mathbf{GC}, \mathbf{U}_d]$  とした理由は  $\mathbf{U}_d$  については関心の対象外とし、一意に定まらない部分  $\mathbf{U}_u$  のみを調整することによって、 $\mathbf{F}$  の望ましい特徴を達成するためである。

このとき、(4.11) 式は

$$\begin{aligned} f(\mathbf{G}, \mathbf{C}, \mathbf{F}_u, \mathbf{U}_u) &= \|[\mathbf{GC}, \mathbf{U}_d] - [\mathbf{F}, \mathbf{U}]\|_F^2 \\ &= \|[\mathbf{GC}, \mathbf{U}_d] - ([\mathbf{F}_d, \mathbf{U}_d] + [\mathbf{F}_u, \mathbf{U}_u])\|_F^2 \\ &= \|[\mathbf{GC}, \mathbf{U}_d] - ([\mathbf{F}_d + \mathbf{F}_u, \mathbf{U}_d + \mathbf{U}_u])\|_F^2 \\ &= \|\mathbf{GC} - (\mathbf{F}_d + \mathbf{F}_u)\|_F^2 + \|\mathbf{U}_u\|_F^2 \end{aligned} \quad (4.12)$$

CCFI のアルゴリズムは、 $\mathbf{G}$ ,  $\mathbf{C}$ ,  $\mathbf{F}_u$ ,  $\mathbf{U}_u$  について、制約条件 (4.9) 式、(4.10) 式のもとで (4.12) 式を最小化する。

## 4.4 アルゴリズム

我々は、最適な  $\mathbf{G}, \mathbf{C}, [\mathbf{F}_u, \mathbf{U}_u]$  の最適な更新を交互に実行する、CCFI の交互最小二乗法を提案する。次の 4.4.1 小節と 4.4.2 小節では、

[1]  $\mathbf{G}$  と  $\mathbf{C}$  を更新するクラスタリングステップ、

[2]  $[\mathbf{F}, \mathbf{U}]$  を更新する因子得点ステップ

をそれぞれ記述している。さらに 4.4.3 小節では全体のステップをまとめている。

### 4.4.1 クラスタリングステップ

(4.12) 式における最後の等式において、 $\mathbf{G}$  と  $\mathbf{C}$  の更新に関するのは

$$h(\mathbf{G}, \mathbf{C}) = \|\mathbf{GC} - (\mathbf{F}_d + \mathbf{F}_u)\|_F^2 = \|\mathbf{GC} - \mathbf{F}\|_F^2 \quad (4.13)$$

のみである。したがって、 $\mathbf{F}$  を既知としたうえで  $\mathbf{G}$  と  $\mathbf{C}$  を最適に更新するためには、(4.13) 式を考慮するだけでよい。これらの更新のためには、 $\mathbf{F}$  の行を分類するために  $k$  平均クラスタリングアルゴリズムを使用することができる (MacQueen, 1967)。つまり、 $\mathbf{G}$  と  $\mathbf{F}$  を既知としたとき、(4.13) 式は

$$\mathbf{C} = (\mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{F} = \mathbf{D}^{-1} \mathbf{G}^\top \mathbf{F} \quad (4.14)$$

によって最小化される。ただし  $\mathbf{D} = \mathbf{G}^\top \mathbf{G}$  は、 $k$  番目の対角要素を  $C_k$  の個体数とする対角行列とする。(4.14) 式は、 $\mathbf{c}_l^\top (\mathbf{C}$  の  $l$  行) がクラスター  $C_l$  に所属する  $\mathbf{F}$  の列の平均ベクトルであることを意味している。この観点から、 $\mathbf{C}$  はクラスターセントロイド行列と呼ばれる。一方で  $\mathbf{C}$  と  $\mathbf{F}$  を既知として、 $\mathbf{G} = (g_{nk})$  の各要素を

$$g_{nk} = \begin{cases} 1 & (k = \operatorname{argmin}_{1 \leq l \leq K} \|\mathbf{f}_n - \mathbf{c}_l\|_2^2) \\ 0 & (\text{それ以外}) \end{cases} \quad (4.15)$$

で更新したとき、(4.13) 式は最小化される。ただし  $\mathbf{f}_n$  は  $\mathbf{F}$  の  $n$  行である。

#### 4.4.2 因子得点ステップ

最適な  $[\mathbf{F}_u, \mathbf{U}_u]$  の更新を考える。このブロック行列は、(4.9) 式のようにも表せるので、(4.12) 式の 2 番目の等式は以下のように書き換えることができる。

$$\begin{aligned} f([\mathbf{F}_u, \mathbf{U}_u]) &= \|[\mathbf{GC}, \mathbf{U}_d] - ([\mathbf{F}_d, \mathbf{U}_d] + [\mathbf{F}_u, \mathbf{U}_u])\|_F^2 \\ &= \|\mathbf{Y} - n^{1/2}\mathbf{KL}' - n^{1/2}\mathbf{K}_\perp\mathbf{L}_\perp^\top\|_F^2 \\ &= f(N^{1/2}\mathbf{K}_\perp\mathbf{L}_\perp^\top) \end{aligned} \quad (4.16)$$

これには  $\mathbf{Y} = [\mathbf{GC}, \mathbf{U}_d]$  と (4.8) 式、(4.9) 式を用いた。さらに (4.16) 式は

$$f(N^{1/2}\mathbf{K}_\perp\mathbf{L}_\perp^\top) = \|\mathbf{Y} - N^{1/2}\mathbf{KL}^\top\|_F^2 - 2N^{1/2}\text{tr}(\mathbf{Y}^\top\mathbf{K}_\perp\mathbf{L}_\perp^\top) + \|N^{1/2}\mathbf{K}_\perp\mathbf{L}_\perp^\top\|_F^2 \quad (4.17)$$

に書き換えることができる。ここで (4.7) 式から  $\|N^{1/2}\mathbf{K}_\perp\mathbf{L}_\perp^\top\|_F^2 = N\text{tr}\mathbf{I}_M$ 。したがって、(4.17) 式の最小化は

$$\phi(\mathbf{K}_\perp, \mathbf{L}_\perp) = \text{tr}\mathbf{Y}^\top\mathbf{K}_\perp\mathbf{L}_\perp^\top = \text{tr}\mathbf{L}_\perp\mathbf{K}_\perp^\top\mathbf{Y} \quad (4.18)$$

の最大化と等しい。ただし  $\mathbf{K}_\perp$  と  $\mathbf{L}_\perp$  は制約条件 (4.7) 式を満たす。

制約条件 (4.7) 式を満たすために、 $\mathbf{K}_\perp$  と  $\mathbf{L}_\perp$  をそれぞれ

$$\mathbf{K}_\perp = \mathbf{J}_1\mathbf{J}_K\mathbf{P}, \quad \mathbf{L}_\perp = \mathbf{J}_L\mathbf{Q} \quad (4.19)$$

とリパラメetrizeする。ただし  $\mathbf{J}_1 = \mathbf{I}_N - N^{-1}\mathbf{1}_N\mathbf{1}_N^\top$ ,  $\mathbf{J}_K = \mathbf{I}_N - \mathbf{KK}^\top$ ,  $\mathbf{J}_L = \mathbf{I}_{P+M} - \mathbf{LL}^\top$  は、既知の対称かつ冪等行列である。 $\mathbf{P}(P \times M)$  と  $\mathbf{Q}((P+M) \times M)$  は未知の行列であり、制約条件

$$\mathbf{P}'\mathbf{J}_K\mathbf{J}_1\mathbf{J}_K\mathbf{P} = \mathbf{Q}^\top\mathbf{J}_L\mathbf{Q} = \mathbf{I}_M \quad (4.20)$$

を満たす。これによって (4.19) 式における  $\mathbf{K}_\perp$  と  $\mathbf{L}_\perp$  は (4.7) 式の中の  $\mathbf{K}_\perp^\top\mathbf{K}_\perp = \mathbf{L}_\perp^\top\mathbf{L}_\perp = \mathbf{I}_m$  を満たす。付録で証明するとおり、(4.19) 式と (4.20) 式は、(4.7) 式の  $\mathbf{K}^\top\mathbf{K}_\perp = \mathbf{L}^\top\mathbf{L}_\perp = \mathbf{O}_{P \times M}$  を満たす。残りの 2 つの等式は (4.19) 式と (4.20) 式から、ただちに導ける。(4.19) 式を (4.18) 式に代入すると、(4.18) 式は

$$\phi(\mathbf{P}, \mathbf{Q}) = \text{tr}\mathbf{P}^\top\mathbf{J}_K\mathbf{J}_1\mathbf{Y}\mathbf{J}_L\mathbf{Q} = \text{tr}\mathbf{P}^\top\mathbf{J}_K\mathbf{J}_1(\mathbf{J}_1\mathbf{J}_K\mathbf{Y}\mathbf{J}_L)\mathbf{J}_L\mathbf{Q} \quad (4.21)$$

と書き換えることができる。ここで用いた  $\mathbf{J}_1 \mathbf{J}_K = \mathbf{J}_K \mathbf{J}_1$  は付録にて証明している。したがって考える問題は、(4.20) 式の制約のもとで  $\mathbf{P}$  と  $\mathbf{Q}$  について (4.21) 式を最大化することである。

これは双線型形式 (generalized bilinear form) である (4.21) 式が (4.20) 式の制約下で

$$\phi(\mathbf{P}, \mathbf{Q}) \leq \text{tr} \Phi_1 \quad (4.22)$$

を満たすことに注意することで解くことができる (ten Berge, 1983; 1993)。ここで  $\Phi_1$  は  $M \times M$  の対角行列であり、

$$\mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L = \mathbf{V} \begin{bmatrix} \Phi_1 \\ & \Phi_2 \end{bmatrix} \mathbf{W}^\top \quad (4.23)$$

で定義される  $\mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L$  の特異値分解によって得られる。ただし、 $\mathbf{V}^\top \mathbf{V} = \mathbf{W}^\top \mathbf{W} = \mathbf{I}_{M+P}$ かつ  $\begin{bmatrix} \Phi_1 \\ & \Phi_2 \end{bmatrix}$  は  $(M+P) \times (M+P)$  は対角要素が降順の非負対角行列である (ten Berge, 1993)。

$\mathbf{V}_1 (N \times M)$  と  $\mathbf{W}_1 ((M+P) \times M)$  をそれぞれ  $\mathbf{V}$  と  $\mathbf{W}$  の左から  $M$  列までの行列とする。制約 (4.20) 式がなければ、(4.22) 式の  $\text{tr} \Phi_1$  の上限は、

$$\mathbf{P} = \mathbf{V}_1, \quad \mathbf{Q} = \mathbf{W}_1 \quad (4.24)$$

によって得られる。なぜなら (4.24) 式を (4.21) 式に代入すると、

$$\begin{aligned} \phi(\mathbf{V}_1, \mathbf{W}_1) &= \mathbf{V}_1^\top \mathbf{J}_K \mathbf{J}_1 \mathbf{Y} \mathbf{J}_L \mathbf{W}_1 \\ &= \mathbf{V}_1^\top \mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L \mathbf{W}_1 \\ &= \mathbf{V}_1^\top \mathbf{V} \Phi \mathbf{W}^\top \mathbf{W}_1 \\ &= \Phi_1 \end{aligned}$$

となるからである。

したがって残った問題は、(4.24) 式が (4.20) 式を満たしているかどうかである。もし

$$R = \text{rank}(\mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L) \geq M \quad (4.25)$$

が成り立つなら,  $\Phi_1(M \times M)$  の全ての対角要素は正であり, (4.24) 式は 4.20 式を満たすことを付録にて証明した. しかし付録で証明したように,  $R \leq \min(K - 1, M)$  である. よって (4.25) 式を考慮すると,

$$K \geq M + 1 \quad (4.26)$$

とせざるをえない. したがって (4.24) 式が最適であるためには  $R = M$  が成り立つ必要がある. 今後本研究では, これが成り立つと仮定し (4.26) 式からクラスター数は因子数プラス 1 とするべきと考える.

最後に, (4.19) 式と (4.24) 式を用いて  $\mathbf{K}_\perp = \mathbf{J}_1 \mathbf{J}_K \mathbf{V}_1$  と  $\mathbf{L}_\perp = \mathbf{J}_L \mathbf{W}_1$  を得る. これらを (4.9) 式に代入すると更新式

$$[\mathbf{F}_u, \mathbf{U}_u] = N^{1/2} \mathbf{J}_1 \mathbf{J}_K \mathbf{V}_1 \mathbf{W}_1^\top \mathbf{J}_L^\top \quad (4.27)$$

を得る.

#### 4.4.3 アルゴリズムの全体

これらの結果をまとめると, CCFI のアルゴリズムは以下のようになる:

ステップ 0. (4.8) 式によって  $\mathbf{F}_d$  を計算する.

ステップ 1.  $[\mathbf{F}_u, \mathbf{U}_u]$  と  $\mathbf{G}$  を初期化する.

ステップ 2.  $\mathbf{C}$  を (4.14) 式によって更新する.

ステップ 3.  $\mathbf{G}$  を (4.15) 式によって更新する.

ステップ 4.  $[\mathbf{F}, \mathbf{U}]$  を (4.15) 式によって更新する.

ステップ 5. 収束したと判定すれば計算を終了する. 収束していないければステップ 2 に戻る.

ステップ 1 では,  $\mathbf{G}$  がメンバーシップ行列となるように乱数によって初期化する. 一方で  $[\mathbf{F}_u, \mathbf{U}_u]$  は

$$[\mathbf{F}_u, \mathbf{U}_u] = N^{1/2} \mathbf{\Gamma}_A \mathbf{\Gamma}_B \quad (4.28)$$

によって初期化される。ここで  $\Gamma_A$  と  $\Gamma_B$  は、各要素が一様分布  $U(-1, 1)$  に従う行列  $A(N \times M)$  と行列  $B((M + P) \times M)$  を用いた行列  $J_1 J_K A$  と  $J_L B$  の特異値分解

$$J_1 J_K A = \Gamma_A \Theta_A \Gamma_A^\top$$

$$J_L B = \Gamma_B \Theta_B \Gamma_B^\top$$

によって得られる。ただし

$$\Gamma_A^\top \Gamma_A = \Gamma_B^\top \Gamma_B = \Theta_A^\top \Theta_A = \Theta_B^\top \Theta_B = I_M.$$

Adachi & Trendafilov (2017) で示されたように、(4.28) 式はありうる一つの  $[F_u, U_u]$  である。ステップ 5 での収束は、(4.11) 式の前回の反復時と比べた減少量が  $10^{-5}$  以下のときと規定する。

解が局所解である場合を避けるために、上記のアルゴリズムを異なる初期値で 1000 回実行する。そこで得られる 1000 セットの解のうち、目的関数の値が最も小さい場合の解を最適解とした。

$[F, U]$  が同定されても、 $F$  と  $\Lambda$  には回転の不定性は残っている。つまり  $T$  が正規直交行列としたとき、 $F\Lambda^\top + U\Psi = FTT^\top\Lambda^\top + U\Psi$  となる。この回転行列  $T$  については既存の回転法を用いる (Browne, 2001)。

## 4.5 クラスター数

前章で述べた通り、CCFI のアルゴリズムにおけるクラスター数については (4.26) 式が成り立つとする。本章では (4.26) 式が満たされるとき、クラスター数をどのように決めるべきかについて議論する。

解釈を容易にするためには、少数のクラスターのみに着目すればよいのでクラスター数は少ない方が望ましい。このことから (4.26) 式を満たしたうえで最も小さい整数をクラスター数とすべきである、つまり

$$K = M + 1. \quad (4.29)$$

一方で、 $K$  をなんらかの適合性を測る関数に基づいて決定することもできる。目的関数 (4.11) 式による決め方が最も単純な選択法である。ここでは  $K$  についての関数  $f_K$  としよ

う.  $f_K$  の値は,  $K$  の増加に伴い単調減少する. 適切な  $K$  を選択するために, スクリープロットのような手順, すなわち変化量  $f_{K+1} - f_K$  が前回の変化量  $f_{K+1} - f_K$  とさほど変わらないような  $K$  の値を見つける方法が考えられる. しかし, この手順では (4.29) 式を選択することはできない. なぜなら,  $K = M$  としたとき CCFI アルゴリズムは適用できないため, (4.29) 式に対して  $f_{K-1}$  を得ることができないからである.

Milligan & Cooper(1985) には,  $K$  について単調に変化するとは限らないいくつかの尺度が紹介されている. この論文では

$$VRC_K = \frac{SSB_K/(K-1)}{SSW_K/(N-K)} \quad (4.30)$$

で定義される Variance Ratio Criterion(VRC) が考慮されている中で, 最も良い尺度とした. ここで  $SSB_K = \sum_{k=1}^K N_k \|\mathbf{c}_k - \bar{\mathbf{c}}\|_2^2$ ,  $SSW_r$  は (4.13) 式の値,  $N_k$  はクラスター  $C_k$  に所属する個体の数とする. (4.30) 式はどの程度クラスターが分離しているかを表しているため,  $K = M + 1 (K = 1, \dots, K_U)$  に対して (4.30) 式が最大となるような  $K$  を選択すればよい. ただし  $K_U$  は考慮する  $K$  の上限である. しかしこの尺度は次の節で示す通り CCFI ではうまくいかない. したがって,  $K$  に関するいくつかの追加/事前情報が利用可能でない限り, クラスターの数として (4.29) 式を選択する.

## 4.6 シミュレーション

本章では人工データセットを用いてシミュレーションを行う. 本研究の目的は, [1] 真のクラスタ数を VRC で検出できるかどうか, [2] 真のクラスターと因子得点を CCFI によってどれくらいうまく復元できるか, [3] CCFI の計算時間, 局所解について評価することである. .1 章ではデータの生成と分析の手順を説明する. 残りで [3], [1], [2] の結果を, それぞれこの順に報告する.

### 4.6.1 データの生成と分析

最初に, CCFI はデータ行列には適用されないことに注意する必要がある. これは, データ行列に適用された行列モデル因子分析の解に対して実行される. ただし, CCFI がデータの背後にある構造を検出するためにはデータ行列が必要である. したがって, まず我々は行列モデル因子分析によって解析されるべきデータセットを合成する.

$\mathbf{F}$  の行が正確にクラスター化できる場合,  $\mathbf{F} = \mathbf{GC}$  が保たれる. 4.2式で想定されている行列モデル因子分析  $\mathbf{X} = \mathbf{F}\Lambda^\top + \mathbf{U}\Psi + \sigma\mathbf{E}$  にこの方程式を代入することによって,

$$\mathbf{X} = \mathbf{GC}\Lambda^\top + \mathbf{U}\Psi + \sigma\mathbf{E} \quad (4.31)$$

を得る. ただし  $\mathbf{E}$  は  $N \times P$  の誤差行列,  $\sigma$  はエラーレベルを決めるスカラーである. さらに (4.3) 式と (4.4) 式に  $\mathbf{F} = \mathbf{GC}$  を代入すると, それぞれ

$$\mathbf{1}_N^\top [\mathbf{GC}, \mathbf{U}] = \mathbf{0}_{M+P}^\top \quad (4.32)$$

$$N^{-1} [\mathbf{GC}, \mathbf{U}]^\top [\mathbf{GC}, \mathbf{U}] = \mathbf{I}_{M+P} \quad (4.33)$$

となる. 得られたデータセットに行列モデル因子分析を適用し, その解に CCFI を適用する.

この手順において, CCFI の解は典型的なシミュレーションとは異なり真値に直接関係はしていない. 典型的なシミュレーションでは, そのモデルに従うデータセットを生成し单一の手法でパラメータを再現する. しかし今回のシミュレーションでは, 以下の手順に従う:

- [1] (4.31) 式に従うデータを作成する.
- [2] データに行列モデル因子分析を適用する.
- [3] 得られた解に CCFI を適用する.

そして, 手順 [1] と [3] の間に [2] を実行し, [3] の解で [1] のパラメータが復元するかどうかを評価する. したがって, このパラメータ復元は前述の典型的なシミュレーションに比べて幾分難しい. この困難さを考慮に入れて, 比較的  $\sigma$  の値は小さくする.

$P = 9, N = 100$  と固定し因子数  $M$  とクラスター数  $K$  の組合せとして,

$$[M, K] = [2, 3], [2, 4], [3, 4], [3, 5]$$

の4通りを考える。 $M = 3$ のときの真の負荷行列を Stegeman(2016) に従って

$$\Lambda = \begin{bmatrix} 0.9 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.5 \\ 0.7 & 0.6 & 0 \\ 0.7 & 0 & 0.3 \\ 0 & 0.6 & 0.3 \\ 0.3 & 0.2 & 0.1 \\ 0.6 & 0.5 & 0.3 \\ 0.6 & 0.6 & 0.4 \end{bmatrix}$$

とし、 $M = 2$ のときは上記の  $\Lambda$  から第3列を取り除いた行列を真の負荷行列とした。そして真の  $\Psi$  は  $\Lambda\Lambda^\top + \Psi^2$  の対角要素が 1 となるようにした。その他のパラメータの生成については、まず  $\mathbf{G}$  の初期化から始める。そして  $\mathbf{C}$  の各要素は  $U(-1, 1)$  から抽出し、 $\mathbf{GC}$  が制約 (4.32) 式と (4.33) 式を満たすように  $\mathbf{C}$  を変換する。 $\mathbf{U}$  の各要素は  $N(0, 1)$  から抽出し制約 (4.32) 式と (4.33) 式を満たすように  $\mathbf{U}$  を変換する。 $\mathbf{E}(N \times P)$  の各要素は  $U(-1, 1)$  から抽出し、 $\mathbf{E}$  は列中心化されているとする。最後にエラーレベルに合わせて (4.31) 式に従う2つのデータセットを作成するために、 $\sigma$  を決める。エラーレベル  $\|\sigma\mathbf{E}\|^2/\|\mathbf{X}\|$  は、0.1 もしくは 0.2 とした。

$4([M, K]) \times 2(\text{エラーレベル})$  の計 8 通りそれぞれに上記の手順を 100 回繰り返す。結果として  $8 \text{ 条件} \times 100 = 800$  データセットが得られる。これらに  $M$  を真値としたうえで行列モデル因子分析を適用する。得られた解を用いて、CCFI を  $K$  を  $M+1, \dots, M+5$  それぞれで実行した、つまり CCFI を  $800 \times 5 = 4000$  回実行した。その結果得られた  $\mathbf{F}$  と  $\mathbf{C}$  には回転の不定性が存在するため、 $\Lambda$  を回転した行列が真値の  $\Lambda$  である (4.31) 式と一致するようにプロクラス回転を行った。

#### 4.6.2 計算時間と局所解の頻度

4000 通りの CCFI の実行において、局所解を避けるためそれぞれ 1000 回ランダムスタートを行っている。本節では、この 1000 回のランダムスタートにどれだけの時間がかかる

のかを考察する。4000通り行った計算時間の平均は14.37秒であった。これは許容範囲内であると考えられる。

$f_l$  は  $l(l = 1, \dots, 1000)$  回目のランダムスタートにおける収束時の目的関数とする。 $f_{\min} = \min_{1 \leq l \leq 1000} f_l$  の場合の解を最適解とし、 $f_l - f_{\min} > 0.001$  を満たす場合の  $f_l$  の解を局所解とした。4000通りにおける  $f_1, \dots, f_{1000}$  の局所解の数を計測した。すると平均が 998.5、標準偏差 0.53 であった。これは 1000 回のランダムスタートのほとんどが局所解であることを意味している。しかも  $f_{\min}$  が大域解であることを保証しない。この結果から、 $k$  平均法と同様に CCFI は局所解が非常に多いことがわかった。しかしながら CCFIにおいて、これは深刻な問題とはならない。実際、局所解は図 4.1 における円周上に存在する最適な因子得点のうちの 1 つであり、 $f_{\min}$  においての解はよく復元されている。

### 4.6.3 VRC の結果

この小節では、CCFI で用いるクラスター数とそのクラスター数をそれぞれ  $K$  と  $K_{\text{true}}$  とする。それぞれのデータセットについて、CCFI の解が  $K = M + 1, \dots, M + 5$  の場合それについての VRC が計 5 通り得られる。ここで VRC が最大となるときの  $K$  を  $K = K_{\text{true}}$  と定義する。しかし、残念ながら  $K_{\text{true}}$  の検出は、うまくいかなかった。正しい  $K$  が検出できたデータセットの割合は、4.6% であった。また 8 条件の中で最も高いパーセンテージは 10% だった。各条件で  $K$  が大きくなるほど VRC も大きくなつた。これは VRC が有用とはいえないことを意味する。よって今後は  $K$  を (4.29) 式、つまり  $K = M + 1$  によって選ぶ。

### 4.6.4 パラメータ復元

この節では (4.29) 式の選択が成り立つ  $[M, K] = [2, 3], [3, 4]$  の場合のデータセットのみを考える。各データセットについて、(4.31) 式におけるパラメータ行列  $\mathbf{G}, \mathbf{C}, \mathbf{F}, \mathbf{U}$  の復元に関する指標、すなわち推定解  $\mathbf{G}, \mathbf{C}, \mathbf{F}, \mathbf{U}$  が真の解  $\mathbf{G}_{\text{true}}, \mathbf{C}_{\text{true}}, \mathbf{F}_{\text{true}}, \mathbf{U}_{\text{true}}$  をどれくらいうまく復元しているかを示す指標を得た。

誤分類率  $N_{MC}/N$  は  $\mathbf{G}$  の復元を示す指標で、 $N_{MC}$  は  $\mathbf{G}_{\text{true}}$  と一致しなかつた  $\mathbf{G}$  の列の数である。表 4.1 の列「MCR  $\mathbf{G}$ 」は、エラーレベルと  $[M, K]$  の各組み合わせに対する 100

表 4.1: MCR, MAE, CORR(一致係数) の平均と標準偏差

Error level	$[M, K]$	MCR		MAE			CORR	
		$\mathbf{G}$	$\mathbf{C}$	$\mathbf{F}$	$\mathbf{F}_{AR}$	$\mathbf{F}_B$	$\mathbf{U}$	$\mathbf{U}$
0.1	[2,3]	0.05	0.03 (0.04)	0.15 (0.05)	0.26 (0.02)	0.30 (0.02)	0.40 (0.02)	0.85 (0.02)
	[3,4]	0.16	0.07 (0.04)	0.31 (0.05)	0.57 (0.03)	0.75 (0.21)	0.49 (0.02)	0.77 (0.02)
0.2	[2,3]	0.08	0.05 (0.04)	0.22 (0.05)	0.32 (0.03)	0.35 (0.03)	0.52 (0.02)	0.75 (0.02)
	[3,4]	0.19	0.09 (0.05)	0.41 (0.06)	0.61 (0.04)	0.81 (0.43)	0.60 (0.02)	0.68 (0.02)

データセットにおける MCR の平均(および標準偏差)を示す。このことから MCR は十分に低く、 $\mathbf{G}_{true}$  はうまく復元されているといえる。

$\mathbf{C}_{true}$  の復元は、 $MAE(\mathbf{C}) = (NK)^{-1} \|\mathbf{C} - \mathbf{C}_{true}\|_1$  と定義される平均絶対誤差(MAE)で評価する。ここで、 $\|\cdot\|_1$  は  $\ell_1$  ノルムを表す。MAE は 2 つの対応する行列の要素の絶対差である。表 4.1 における「MAE  $\mathbf{C}$ 」は、 $MAE(\mathbf{C})$  の平均値を示す。それらは十分小さく、 $\mathbf{C}$  がよく復元されることを意味する。

「MAE  $\mathbf{F}$ 」の列は、 $MAE(\mathbf{F}) = (NK)^{-1} \|\mathbf{F} - \mathbf{F}_{true}\|_1$  である。 $\mathbf{C}$  と比較して、 $\mathbf{F}_{true}$  の復元はそれほどうまくいかないことがわかる。この結果が許容できるかどうかを検討するために、因子得点の従来の推定解と CCFI 解を比較する。つまり、古典的最尤法 FA を実行し、 $\mathbf{F}_d$  の Anderson & Rubin(1956) における推定法(AR) と Bartllet(1937) における推定法(Bar) それぞれの推定値を得る。

これらは、CCFI の解の場合と同様にプロクラステス回転を行った。表 2 の「 $\mathbf{F}_{AR}$ 」および「 $\mathbf{F}_B$ 」の欄は、Anderson-Rubin および Bartllet の推定値に対する MAE の平均値をそれぞれ示している。そこでは、CCFI の復元が古典的推定法における復元よりも実質的に優れていることがわかる。これは、 $\mathbf{F}$  が CCFI においてよく復元できているといえる。

列「MAE  $\mathbf{U}$ 」は、 $\mathbf{U}_{true}$  の復元が、他のパラメータの復元よりも悪いことを示している。このような結果になった理由の一つとして、 $\mathbf{U}_{true}$  には  $\mathbf{G}_{true}, \mathbf{C}_{true}, \mathbf{F}_{true}$  のような特殊な構造がないことが考えられます。例えば、 $\mathbf{F}_{true}$  は (4.31) 式のように  $\mathbf{G}_{true} \mathbf{C}_{true}$  に制約され、 $\mathbf{F}$  が CCFI の GC と一致するように  $\{\mathbf{G}, \mathbf{C}, \mathbf{F}\}$  が推定される。残念ながらこのような

考えは  $\mathbf{U}$  においてはない。しかしながら、表 4.1 の  $\mathbf{U}$  と  $\mathbf{U}_{\text{true}}$  との一致係数をみると、十分正の値をとっていることがわかる。よって推定した  $\mathbf{U}$  は  $\mathbf{U}_{\text{true}}$  と相関関係にあり、解釈する価値があるといえる。

## 4.7 実データ解析

本節では、2つの実データ例を用いて CCFI の有用性を示す。最初の例は、識別可能な少數の個体からなる観測データからなる。これは得られた共通因子得点と独自因子得点がどのくらい解釈可能かを説明するのに適している。一方、2つめのデータは比較的個体数の多く、クラスター化共通因子得点がどのくらい有用であるかを説明するのに適している。

### 4.7.1 個体数が少ないデータでの例

まず1つめの適用例は、表 4.2 に書かれている  $N = 14$ ,  $P = 6$  の職業印象データである。

表 4.2: 職業印象データ

職業	役に立つ	よい	誠実な	速い	騒がしい	忙しい
僧侶 (MO)	2.7	3.7	3.3	2.2	1.4	1.8
銀行員 (BC)	3.5	3.4	4.1	3.2	2.1	4.2
漫画家 (CA)	3.2	3.5	3.4	3.3	3.4	4.3
デザイナー (DE)	3.2	3.5	3.2	3.6	2.9	4.0
保育士 (NU)	4.6	4.5	4.5	2.8	3.3	4.9
大学教授 (PR)	4.0	3.8	3.7	2.4	1.5	3.0
医師 (DR)	4.8	3.9	3.7	2.4	1.5	3.0
警察官 (PM)	4.6	4.1	4.2	4.3	3.4	4.0
新聞記者 (JO)	4.3	3.7	3.9	4.7	4.2	5.0
船乗り (SA)	3.6	3.5	3.5	3.5	3.5	3.5
プロスポーツ選手 (AT)	3.2	3.7	3.7	4.9	3.5	4.1
作家 (NO)	3.7	3.5	3.3	2.3	1.8	3.3
俳優 (AC)	3.2	3.6	2.8	3.3	3.3	4.3
キャビンアテンダント (CA)	3.8	3.8	3.9	3.9	2.5	4.3

標準化したデータを  $M = 2$  として行列モデル因子分析によって分析した。得られた因

子得点に対して CCFI を  $K = 3$  で適用した。アルゴリズムを 1000 回実行した場合の 0.28 秒を要し、局所解の数は 1000 個中 986 個であった。比較のために古典的因子分析をデータセットに適用し、 $\mathbf{F}_d$  の推定値を Anderson & Rubin(1956) と Bartlett(1937) の方法によって推定した。得られた解には因子得点行列と因子負荷行列にはバリマックス回転を適用した。

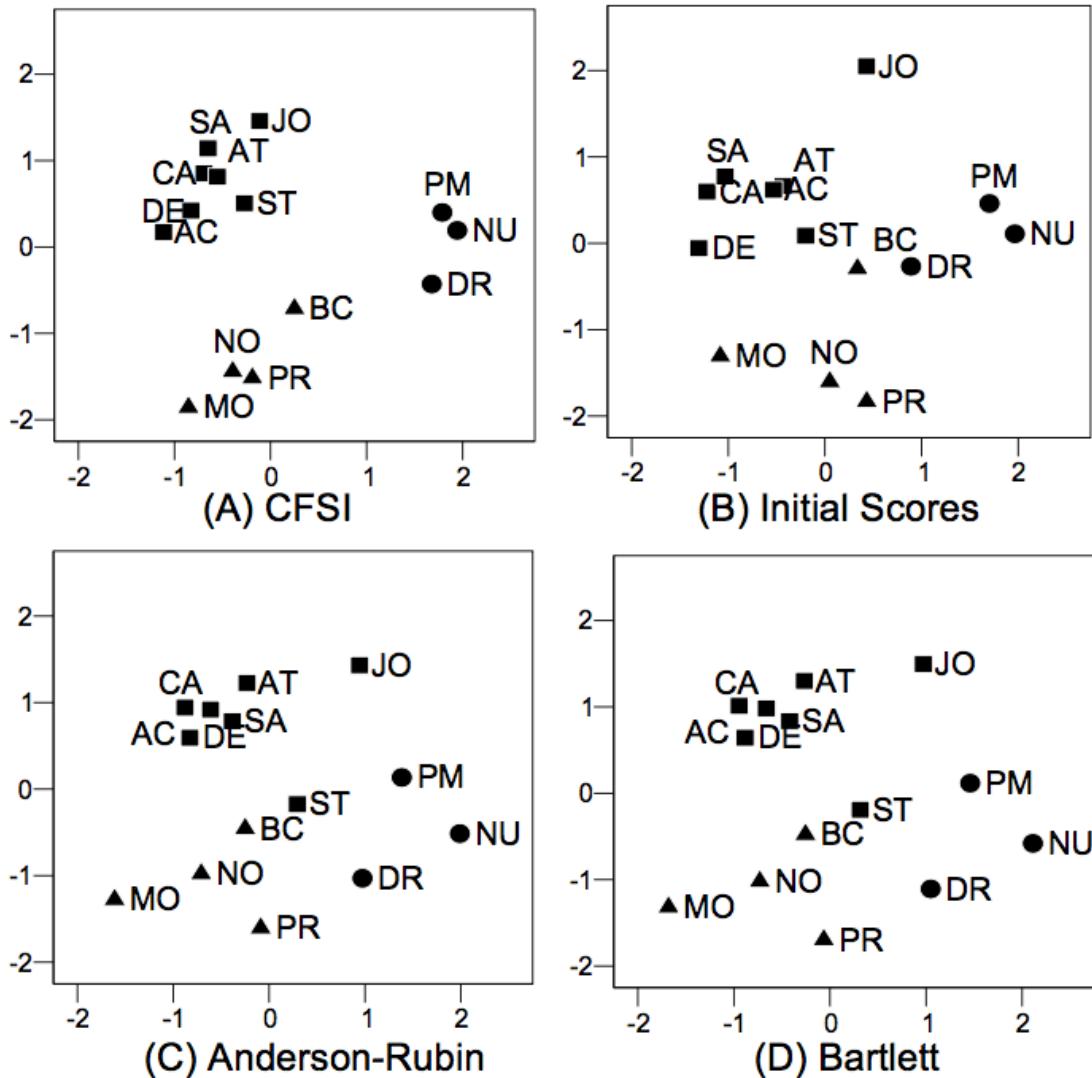


図 4.2: 職業データの共通得点散布図

図 4.2 の (A) は、共通因子得点  $\mathbf{F}$  と CCFI によって得られたクラスターのメンバーシップを表している。それに対して (B) は、(A) の解に至ったときの初期化された  $[\mathbf{F}_u, \mathbf{U}_u]$  を使った  $\mathbf{F} = \mathbf{F}_d + \mathbf{F}_u$  を表している。(B) ではクラスターがあまり分かれていないのに対

して、(B) を初期値として CCFI を適用した結果得られた得点である (A) の  $\mathbf{F}$  は十分分かれている。古典的な因子分析における  $\mathbf{F}_d$  の推定値は、図 4.2 の (C) と (D) からわかるように、クラスターは (B) のようには分かれていない。

図 4.2 から、クラスター  $C_1$  は銀行員、小説家、大学教授、僧侶という主にデスクワークから成ることがわかる。クラスター  $C_3$  は警察官、保育士、医師という人々を守る役割の職業が集まっている。クラスター  $C_2$  に分類されたキャビンアテンダントとデザイナーは芸術的な才能が必要とされるが、同じく  $C_2$  に分類されたスポーツ選手は身体的な才能が必要とされる。これらのことから、 $C_1, C_2, C_3$  をそれぞれデスクワーク、才能、人々を守る、と呼ぶことにする。これらの名称は表 4.3 のセントロイド行列  $\mathbf{C}$  の行名に用いている。

表 4.3: 因子分析によって得られた職業データのパラメータ

変数	$\mathbf{\Lambda}$		$\psi_p^2$
役に立つ	0.87	0.15	0.21
よい	0.79	0.03	0.37
誠実な	0.78	0.15	0.36
速い	0.05	0.82	0.32
騒がしい	0.05	0.94	0.11
忙しい	0.40	0.70	0.34
クラスター	$\mathbf{C}$		
デスクワーク	-0.30	-1.38	
才能	-0.60	0.70	
人々を守る	1.80	0.05	

表 4.3 には負荷行列  $\mathbf{\Lambda}$  を記載した。第 1 列、すなわち第 1 因子は職業の印象がよいかどうかが関与している一方、第 2 列である第 2 因子は、活動性を表している。この第 1 因子と第 2 因子はそれぞれ図 4.2 の (A) における縦軸と横軸を表している。因子の解釈から、表 3 のセントロイド行列  $\mathbf{C}$  から各クラスターは非活動的、活動的、高評価という特徴をもつことがわかる。表 4.3 の右側の列には、独自分散  $\psi_p^2$  が記載されている。独自分散は各変数が共通因子によって説明できない分散の割合を意味し、 $\psi_p^2$  は  $\Psi^2$  の第  $p$  番目の対角要素である。変数の「役に立つ」と「騒がしい」は  $\psi_p^2$  の値が小さく、共通因子によってよく説明できていることがわかる。

$\mathbf{U} = (u_{np})$  の独自因子得点は、CCFI によって同定される。図 4.3 には、CCFI によって得られた  $u_{np}$  に  $\psi_p$  をかけた  $\psi_p u_{np}$  を図示している。個体ごとの  $\psi_p u_{np}$  の散布図は変数  $p$

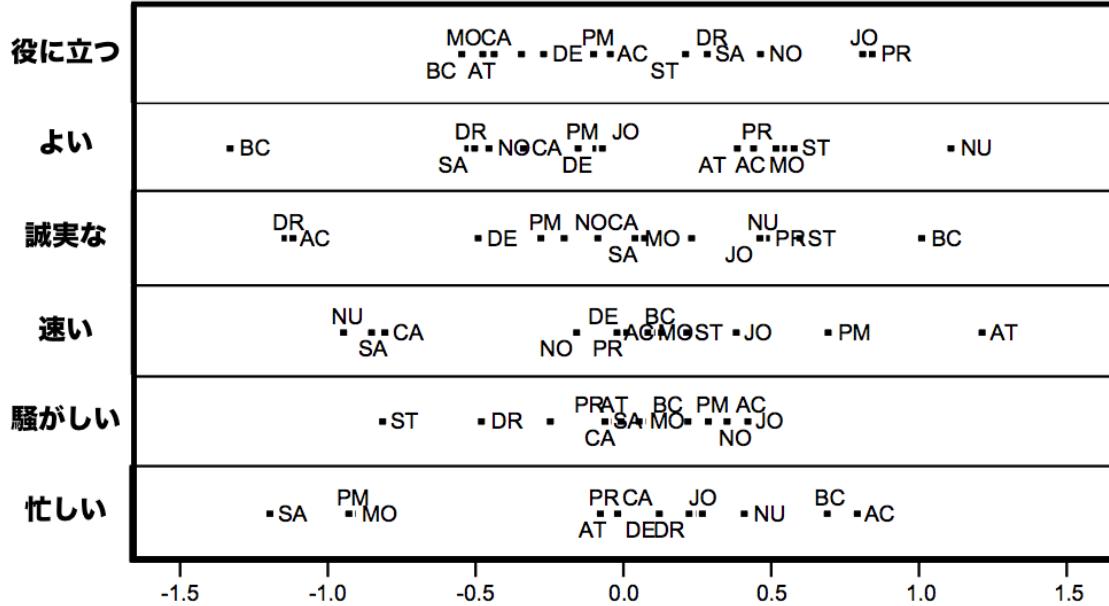


図 4.3: 職業データの独自因子得点

ごとに記載しており,  $\psi_p u_{np}$  のばらつきは独自分散を表す. 散布図を解釈すると, 行列モデル因子分析の目的関数 (4.2) 式は,  $\mathbf{X} \simeq \mathbf{F}\Lambda^\top + \mathbf{U}\Psi$  もしくは  $x_{np} \simeq \mathbf{f}_n^\top \boldsymbol{\lambda}_p + \psi_p u_{np}$  であるから,

$$\psi_p u_{np} \simeq x_{np} - \mathbf{f}_n^\top \boldsymbol{\lambda}_p \quad (4.34)$$

に書き換えられる. ここで  $x_{np}$  は  $\mathbf{X}$  の  $n$  行  $p$  列の要素であり,  $\boldsymbol{\lambda}_p$  は  $\Lambda$  の  $p$  行である. ここで  $\mathbf{f}_n^\top \boldsymbol{\lambda}_p$  は, 共通因子  $\mathbf{f}_n$  によって, その他の変数  $x_{nl}$  ( $l \neq p$ ) と関係のある  $x_{np}$  の成分を意味する. この成分は (4.34) 式を用いて,  $x_{np}$  から取り除くことができる.

この点において,  $\psi_p u_{np}$  は他の変数とは独立した  $x_{np}$  の成分, すなわち個人  $n$  が変数  $p$  のみによってどれだけ表現できるかを示すものとしてみなすことができる. たとえば, スポーツ選手は職種の中で変数「速い」で最大の  $\psi_p u_{np}$  を示している. これは, スポーツ選手は他の変数とは独立にした「速さ」によって特徴付けられると解釈することができる. 対称的に,  $\psi_p u_{np}$  が最小であることから, 他の変数とは独立の「速い」とは正反対の特徴, 言うなれば「遅い」によって保育士を特徴づけることができる. 加えて,  $\psi_p u_{np}$  の値が 0 に近い職業は特筆すべき「速い」あるいは「遅い」という特徴を持たないことがわかる.

### 4.7.2 個体数が多いデータでの例

2つめの例は Izenman(2008) のウェブページでも紹介されているワインデータの例である。データは個体数が  $N = 178$ , 変数が  $P = 13$  からなる。標準化したデータを 4.7.1 節と同様のやり方で分析した。CCFI のアルゴリズムを 1000 回実行するのに要した時間は 9.09 秒で, 得られた 1000 個の解の中で 4.6.2 節で定義した局所解の数は 998 個であった。

得られた共通因子得点を図 4.2 と同じように図 4.4 へ描写した。図 4.4(B) にある初期値の得点は, ワインクラスターがあまり分離されていない。一方で CCFI による解である図 4.4(A) ではその問題が解決されていることがわかる。実際 3 つのクラスターが明確に分かれており, 同じクラスター内のワインは中心に集中している。

図 4.4 の (C) と (D) は古典的な因子得点の推定方法によって得られたものであるが, (B) と同様クラスターが十分に分かれているとはいえない。このような散布図では, 特に個体数が多いような場合には何かの知見を得ることは難しい。それとは対称的には, 図 4.4(A) のような CCFI の解では各クラスターの個体が中心に集中しているため, 表 4.4 にあるセントロイド行列  $C$  の各行ベクトルがそのクラスターにおけるワインの特性をよく表現できている。したがって, 図 4.4(A) を見つつ表 4.3 の因子負荷行列とセントロイド行列を参照することで容易に共通因子得点の解釈ができる。

図 4.4(A)において, 縦軸と横軸はそれぞれ第 1 共通因子と第 2 共通因子を表している。図と表 4.4 を見ると, クラスター  $C_1$  と  $C_2$  は第 1 共通因子によって  $C_3$  と区別できる一方, 第 2 共通因子ではクラスター  $C_1, C_2, C_3$  はそれぞれ大きい, 中ぐらい, 小さいとして区別できることがわかる。表 4.4 が示す通り, 第 1 共通因子はアルコール, 強さ, プロリンと深く関係しているのに対し, 第 2 共通因子はフェノール, フラボノイド, OD280/OD315 が深く関係している。

## 4.8 考察

因子分析モデルは因子得点の不定性を持つ。つまり, 共通因子得点と独自因子得点は一意に定まる部分と定まらない部分の和である。本研究では, 行列分解を用いた因子分析の枠組みで一意に定まらない部分が何らかの望ましい特徴を持つように同定する手法, CCFI を提案した。本研究では共通因子得点が少数のグループにクラスター化されている

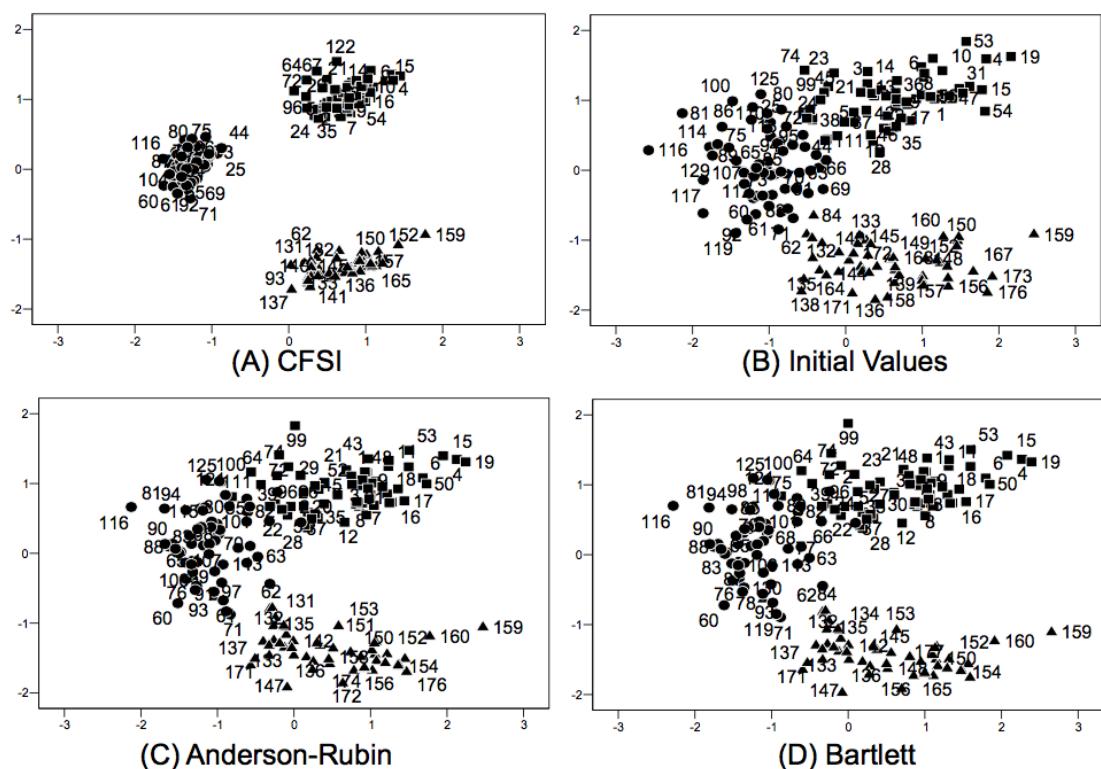


図 4.4: 職業データの共通得点散布図

表 4.4: 因子分析によって得られたワインデータのパラメータ

変数	$\Lambda$	$\psi_p^2$
アルコール	0.78	0.13
リンゴ酸	0.15	-0.51
灰	0.33	-0.04
灰のアルカリ性	-0.18	-0.42
マグネシウム	0.38	0.18
フェノール	0.28	0.82
フラバノイド	0.19	0.93
非フラバノイド	-0.10	-0.57
プロアントシアニン	0.18	0.61
色彩強度	0.79	-0.37
色調	-0.26	0.67
OD280/OD315	-0.09	0.86
プロリン	0.66	0.45
クラスター	C	
$C_1$	0.72	1.05
$C_2$	0.68	-1.40
$C_3$	1.80	0.05

ことを、望ましい特徴と位置付けた。そのような特徴を持つ因子得点行列を得るために、最小二乗基準においてのターゲット行列に近似させた。ここでターゲット行列の中の共通因子得点行列にあたる部分は、共通因子得点がよくクラスター化されているように設定した。一方で残りの独自因子得点行列にあたる部分は、一意に定まった部分の独自因子得点にあたる部分と事前に指定した。また因子得点、クラスターメンバーシップ、およびクラスターセントロイドの未決定部分に対する最小二乗関数を最小化するための反復アルゴリズムを提示した。

因子得点の不定性とはまた別に、因子分析モデルには回転の不定性がある。つまり、共通因子得点行列と因子負荷行列は回転行列によって回転することができる。心理統計学では、その回転の不定性をむしろ利用して好ましい性質を持った因子負荷量を得る回転法が数多く提案されてきた(Browne, 2001)。本研究におけるアプローチでは、回転の不定性ではなく因子得点の不定性を利用する。回転法では因子分析によって得られた因子負荷量が解釈しやすくなることを目指したが、提案手法ではクラスター化された共通因子得点を得られることを目指した。

筆者の知る限りにおいて、因子得点の不定性を利用して望ましい性質を持つ因子得点を得ることを目指した先行研究はない。さらに4.7.1節における独自因子得点の解釈というのも初めての試みと思われる。このように本研究は先進的な試みであるが、課題も残されている。例えば、本研究で用いた共通因子得点のクラスター化以外に考えられる望ましい特徴の模索である。さらに独自因子得点における有用な性質の考察も重要な課題である。

## 第5章

# 変数クラスタリングのための因子分析

因子分析は、多数の観測変数が少数の潜在変数によって説明されうるという仮定に基づく多変量モデルであり、様々な応用例がある。ところが回転の不定性と呼ばれる因子分析が抱える根本的問題があるために、推定された解をそのまま解釈することはできない。この不定性を解消するために長年研究されているのが回転法であり、むしろ不定性を利用して因子負荷量の解釈をより容易にする目的で回転法を利用する分析者も多い。

本章では、回転法とは異なるアプローチにより解釈容易な因子負荷量を得るために手法を紹介する。本章の構成は5.1節にて $k$ -meansクラスタリングを用いることにより、解釈のしやすい因子負荷量を得る手法を紹介する。その後、既存手法の欠点を補う新たな手法を紹介し、実データに適用することで有用性を検証するとともに、2つの手法を比較検証する。

次節では、 $k$ -meansタイプのクラスタリングを用いることで、解釈しやすい負荷量を得る因子分析を紹介する。

### 5.1 完全単純構造因子分析

解釈を容易にするというのは曖昧な言葉であり定義することは困難であるが、因子負荷行列に望まれるいくつかの特徴は既に提案されている。

#### 5.1.1 Thurstoneの単純構造ルール

Thurstone(1947)は、因子負荷行列が満たすべき条件を単純構造(simple structure)と呼び、明文化した。単純構造の条件は以下の5つである。

1. 各行に少なくとも 1 つ, 要素が 0 であること
2. 各列に因子数と同じ数の要素が 0 であること
3. 2 つの列に着目したとき, 一方の列では 0 でもう一方では 0 でない行が複数あること
4. もし因子数が 4 以上なら, 2 つの列に着目したとき, どちらの列も要素が 0 であること
5. 2 つの列に着目したとき, どちらの要素も 0 でない行が少ないとこと

各々の条件を満たすとき, 因子負荷行列がそれぞれどのような行列となるのかを確認する.

各行に少なくとも 1 つ, 要素が 0 であること

この条件を満たす因子負荷行列として, たとえば

$$\Lambda = \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ * & * & 0 \\ * & 0 & * \\ * & 0 & * \\ 0 & * & * \\ 0 & * & * \end{bmatrix}$$

のような行列  $\Lambda$  が考えられる. ただし, \* は非 0 の要素を表すとする. 因子負荷量が 0 であるということは, 変数と因子の間に関係がないことを表しており, 因子の解釈を行う際に有用である.

各列に因子数と同じ数の要素が 0 であること

この条件を満たす因子負荷行列として, たとえば

$$\Lambda = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & 0 & 0 \\ * & 0 & 0 \\ 0 & 0 & * \\ 0 & * & * \\ 0 & * & * \end{bmatrix}$$

のような行列  $\Lambda$  が考えられる.

2つの列に着目したとき, 一方の列では 0 でもう一方では 0 でない行が複数あること

この条件を満たす因子負荷行列として, たとえば

$$\Lambda = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & * & 0 \\ * & 0 & * \\ * & 0 & * \\ 0 & * & * \\ 0 & * & * \end{bmatrix}$$

のような行列  $\Lambda$  が考えられる.

もし因子数が 4 以上なら, 2 つの列に着目したとき, どちらの列も要素が 0 であること

この条件を満たす因子負荷行列として, たとえば

$$\Lambda = \begin{bmatrix} * & * & 0 & 0 \\ * & 0 & * & 0 \\ 0 & * & * & 0 \\ * & * & 0 & 0 \\ * & 0 & 0 & * \\ 0 & * & 0 & * \\ 0 & 0 & * & * \\ * & 0 & * & 0 \\ 0 & * & 0 & * \\ 0 & * & * & 0 \end{bmatrix}$$

のような行列  $\Lambda$  が考えられる.

2 つの列に着目したとき, どちらの要素も 0 でない行が少ないとこと

この条件を満たす因子負荷行列として, たとえば

$$\Lambda = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & 0 & 0 \\ * & 0 & * \\ 0 & 0 & * \\ 0 & * & * \\ 0 & * & 0 \end{bmatrix}$$

のような行列  $\Lambda$  が考えられる.

5 つの条件はそれぞれ重複する場合もあるが, どの条件を満たしても因子負荷量の解釈が容易になることがわかる. しかしながら, それぞれの条件を満たすかどうかを確認することは煩雑であることも事実である. そこで, これらの条件を統合した完全単純構造を利用した研究が盛んになっている.

## 完全単純構造

完全単純構造 (perfect simple structure) あるいは完全クラスター構成 (perfect cluster configuration) とも呼ばれる (Browne, 2001) 構造は、各行に非 0 の要素が多くとも 1 つという強い制約である。つまり、

$$\Lambda = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \\ 0 & * & 0 \\ 0 & * & 0 \\ 0 & 0 & * \\ 0 & 0 & * \end{bmatrix}$$

のような因子負荷行列は完全単純構造を満たす。完全単純構造の負荷行列は解釈が容易であることから、近年注目されており、様々な主成分分析などの手法に応用されるなど研究が進んでいる (Bernaards & Jennrich 2003; Jennrich 2004; Vichi & Saporta, 2009; Vichi, 2017)。

完全クラスター構成と呼ばれることからも明らかのように、完全単純構造は  $k$ -means クラスタリングにおけるメンバーシップ行列 (指示行列) の一般化と考えることができる。 $k$ -means クラスタリングでは、メンバーシップ行列の要素は値が 0 あるいは 1 であったが、完全単純構造は 0 あるいは非 0 である。また、通常  $k$ -means クラスタリングは個体をクラスタリングするが、完全単純構造では変数をクラスタリングしている点も異なる。

### 5.1.2 モデル

ある回転法は因子負荷行列を完全単純構造へと近づけることが示されているが、回転法の問題点は正確に値が 0 にはならないという点である。0.01 や -0.01 といった 0 に近い値を完全単純構造と呼べるかは疑問である。

そこで、Adachi & Trendafilov(2018) は、変数クラスタリングを用いた完全単純構造制約を満たす行列モデル因子分析を提案し、SSFA と命名した。

データ行列  $\mathbf{X}$  を  $N \times P$  の行列, としたとき, SSFA のモデルは

$$\mathbf{X} = \mathbf{Q}\mathbf{R}\boldsymbol{\Lambda}^\top + \mathbf{U}\boldsymbol{\Psi} \quad (5.1)$$

である. ただし,  $M$  は因子数であり  $\mathbf{X} \in \mathbb{R}^{N \times P}$ ,  $\mathbf{Q} \in \mathbb{R}^{N \times M}$ ,  $\mathbf{R} \in \mathbb{R}^{M \times M}$ ,  $\mathbf{U} \in \mathbb{R}^{N \times P}$ ,  $\boldsymbol{\Psi} \in \mathbb{R}^{P \times P}$  である. そして, これらのパラメータが満たすべき条件は

$$N^{-1}\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_M \quad (5.2)$$

$$\text{diag}(\mathbf{R}^\top\mathbf{R}) = \mathbf{I}_M \quad (5.3)$$

$$N^{-1}\mathbf{U}^\top\mathbf{U} = \mathbf{I}_P \quad (5.4)$$

$$\mathbf{Q}^\top\mathbf{U} = \mathbf{O}_{M \times P} \quad (5.5)$$

とする. ただし  $\mathbf{R}$  は上三角行列,  $\boldsymbol{\Psi}$  は対角行列とする. 通常の行列モデル因子分析のモデルである (2.30) 式と比較すると,  $\mathbf{F}$  が  $\mathbf{Q}\mathbf{R}$  に置き換わっていることがわかる. これは QR 分解

$$\mathbf{F} = \mathbf{Q}\mathbf{R} \quad (5.6)$$

によって達成できる. 通常の行列モデル因子分析では共通因子の直交性を仮定しているのに対して, SSFA では

$$\begin{aligned} N^{-1}\mathbf{F}^\top\mathbf{F} &= \mathbf{R}^\top(N^{-1}\mathbf{Q}^\top\mathbf{Q})\mathbf{R} \\ &= \mathbf{R}^\top\mathbf{R} = \boldsymbol{\Phi} \end{aligned} \quad (5.7)$$

と, 因子間に相関があることを仮定している.  $\boldsymbol{\Phi}$  は因子間相関行列である.

そして, SSFA の最も重要かつ強い制約として,  $\boldsymbol{\Lambda}$  は完全単純構造とする. つまり,

$$\boldsymbol{\Lambda} = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \\ 0 & * & 0 \\ 0 & * & 0 \\ 0 & 0 & * \\ 0 & 0 & * \end{bmatrix}$$

のように各行に非ゼロ要素は1つのみである。つまり

$$\sum_{m=1}^M \lambda_{pm} = \max(\lambda_{p1}, \dots, \lambda_{pM})$$

を満たす。そして、 $k$ -means クラスタリングのメンバーシップ行列がそうであるように  $\Lambda^\top \Lambda$  は対角行列である。

### 5.1.3 アルゴリズム

SSFA のアルゴリズムはモデル (5.1) 式の誤差最小化であり、目的関数は

$$f(\mathbf{Q}, \mathbf{R}, \boldsymbol{\Lambda}, \mathbf{U}, \boldsymbol{\Psi}) = \|\mathbf{X} - \mathbf{QR}\boldsymbol{\Lambda}^\top - \mathbf{U}\boldsymbol{\Psi}\|_F^2 \quad (5.8)$$

である。(5.8) 式を最小化するような  $\mathbf{Q}, \mathbf{R}, \boldsymbol{\Lambda}, \mathbf{U}, \boldsymbol{\Psi}$  を交互最小二乗法によって求める。

#### Q-U ステップ

$\mathbf{Z} = [\mathbf{Q}, \mathbf{U}]$ ,  $\mathbf{B} = [\boldsymbol{\Lambda}\mathbf{R}^\top, \boldsymbol{\Psi}]$  とブロック行列を用いると目的関数 (5.8) 式は

$$f(\mathbf{Z}, \mathbf{B}) = \|\mathbf{X} - \mathbf{Z}\mathbf{B}^\top\|_F^2 \quad (5.9)$$

と書き換えることができる。そして、 $\mathbf{Z}$  は制約条件 (5.2) 式と (5.5) 式から、

$$N^{-1}\mathbf{Z}^\top \mathbf{Z} = \mathbf{I}_{P+M} \quad (5.10)$$

が成り立つ。 $\mathbf{B}$  を固定すると、目的関数が

$$f(\mathbf{Z}, \mathbf{B}) = \text{const}_{\mathbf{Z}} - 2\text{tr}(\mathbf{X}^\top \mathbf{Z}\mathbf{B}^\top)$$

と書けることから、 $\mathbf{Z}$  の更新はトレース最大化に帰着する。ただし  $\text{const}_{\mathbf{Z}}$  は  $\mathbf{Z}$  に関係しない定数とする。ten Berge の定理から  $\mathbf{Z}$  の更新式は

$$\hat{\mathbf{Z}} = N^{1/2} \mathbf{KL}^\top = N^{1/2} \mathbf{K}_1 \mathbf{L}_1^\top + N^{1/2} \mathbf{K}_2 \mathbf{L}_2^\top \quad (5.11)$$

によって得られる。ここで  $\mathbf{K}, \mathbf{L}$  は特異値分解

$$\begin{aligned} N^{-1/2} \mathbf{XB} &= \mathbf{K} \boldsymbol{\Delta} \mathbf{L}^\top \\ &= \mathbf{K}_1 \boldsymbol{\Delta}_1 \mathbf{L}_1^\top + \mathbf{K}_2 \mathbf{O}_{M \times M} \mathbf{L}_2^\top \\ &= \mathbf{K}_1 \boldsymbol{\Delta}_1 \mathbf{L}_1^\top \end{aligned} \quad (5.12)$$

によって得られる。 $\mathbf{X}\mathbf{B}$ ではなく $N^{-1/2}\mathbf{X}\mathbf{B}$ の特異値分解を行う理由は、 $\mathbf{Z}^\top\mathbf{Z} = \mathbf{I}_{P+M}$ ではなく $N^{-1}\mathbf{Z}^\top\mathbf{Z} = \mathbf{I}_{P+M}$ であるためである。また、行列モデル因子分析における因子得点の不定性から $\mathbf{K}_2$ と $\mathbf{L}_2$ は同定することはできない。

### $\Psi$ ステップ

目的関数(5.8)式は $\Psi$ 以外のパラメータを固定すると

$$\begin{aligned} f(\Psi|\mathbf{Q}, \mathbf{R}, \Lambda, \mathbf{U}) &= \|\mathbf{X} - \mathbf{Q}\mathbf{R}\Lambda^\top - \mathbf{U}\Psi\|_F^2 \\ &= \text{tr}(\mathbf{X} - \mathbf{Q}\mathbf{R}\Lambda^\top - \mathbf{U}\Psi)^\top(\mathbf{X} - \mathbf{Q}\mathbf{R}\Lambda^\top - \mathbf{U}\Psi) \\ &= \text{const}_\Psi - \text{tr}\Psi\mathbf{U}^\top\mathbf{U}\Psi + 2\text{tr}\mathbf{X}^\top\mathbf{U}\Psi \\ &= \text{const}_\Psi - N\text{tr}\Psi^2 + 2\text{tr}\Psi\mathbf{X}^\top\mathbf{U} \\ &= \text{const}_\Psi + \|N^{1/2}\Psi - N^{-1/2}\mathbf{X}^\top\mathbf{U}\|_F^2 \end{aligned}$$

と書き換えることができる。ただし $\text{const}_\Psi$ は $\Psi$ とは関係しない定数である。 $\Psi$ が対角行列であることを踏まえると、 $\Psi$ の更新式は

$$\hat{\Psi} = N^{-1}\text{diag}(\mathbf{X}^\top\mathbf{U}) \quad (5.13)$$

となる。

### $\mathbf{R}$ ステップ

目的関数(5.8)式は $\mathbf{R}$ 以外のパラメータを固定すると

$$\begin{aligned} f(\mathbf{R}|\mathbf{Q}, \Lambda, \Psi) &= \text{tr}\mathbf{X}^\top\mathbf{X} + \text{tr}\Lambda\mathbf{R}^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{R}\Lambda^\top - 2\text{tr}\mathbf{X}^\top\mathbf{Q}\mathbf{R}\Lambda^\top - \text{tr}\Psi\mathbf{U}^\top\mathbf{U}\Psi \\ &= \text{tr}\mathbf{X}^\top\mathbf{X} + N\text{tr}\mathbf{R}^\top\mathbf{R}\Lambda^\top\Lambda - 2\text{tr}\mathbf{Y}\mathbf{R}\Lambda^\top - N\text{tr}\Psi^2 \\ &= \text{tr}\mathbf{X}^\top\mathbf{X} + N\text{tr}\Lambda^\top\Lambda - 2\text{tr}\mathbf{X}^\top\mathbf{Q}\mathbf{R}\Lambda^\top - N\text{tr}\Psi^2 \\ &= \text{const}_\mathbf{R} - 2\text{tr}\mathbf{X}^\top\mathbf{Q}\mathbf{R}\Lambda^\top \end{aligned}$$

と書き換えられる。ここで3行目の等式には $\text{diag}(\mathbf{R}^\top\mathbf{R}) = \mathbf{I}_M$ を用いている。したがって $\mathbf{R}$ について目的関数を最小化することは、トレース

$$\text{tr}\mathbf{X}^\top\mathbf{Q}\mathbf{R}\Lambda^\top = \text{tr}\Lambda^\top\mathbf{X}^\top\mathbf{Q}\mathbf{R} = \text{tr}\mathbf{Y}\mathbf{R}$$

の最大化に等しい。ただし  $\mathbf{Y} = \Lambda^\top \mathbf{X}^\top \mathbf{Q}$  とする。

$\mathbf{R}$  は上三角行列であるため、たとえば  $M = 3$  とすると

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}$$

となる。かつ制約条件として

$$r_{11} = 1 \quad (5.14)$$

$$\|\mathbf{r}_m\|_2 = 1 \quad (5.15)$$

を満たすとする。ただし  $\mathbf{r}_m$  は  $\mathbf{R}$  の  $m$  列目とする。つまり  $\mathbf{r}_m = [\mathbf{r}_m^*, \mathbf{0}_{M-m}]^\top$  のように書くことができる。

$\mathbf{R}$  の第1列の更新は必要ない。なぜなら  $\mathbf{r}_1 = [1, \mathbf{0}_{M-1}]^\top$  であるからである。2列目以降の  $\mathbf{r}_m$  の更新は、実質  $\mathbf{r}_m^*$  の更新であり

$$\mathbf{r}_m^* = \frac{\mathbf{y}_m^*}{\|\mathbf{y}_m^*\|_2} \quad (5.16)$$

によって達成される。ただし  $\mathbf{y}_m$  は行列  $\mathbf{Y}$  の第  $m$  列であり、 $\mathbf{y}_m^*$  は  $\mathbf{r}_m^*$  と同じように  $\mathbf{y}_m = [\mathbf{y}_m^*, (y_{m,m+1}, \dots, y_{m,M})]$  と分割したとする。

## Λステップ

### 目的関数

$$\text{tr} \mathbf{X}^\top \mathbf{X} + N \text{tr} \Lambda^\top \Lambda - 2 \text{tr} \mathbf{X}^\top \mathbf{Q} \mathbf{R} \Lambda^\top - N \text{tr} \Psi^2$$

を  $\Lambda$  以外のパラメータを固定したうえで最小化し、 $\Lambda$  を更新する。

$\Lambda$  は完全単純構造であるため、 $\Lambda^\top \Lambda$  は対角行列となる。したがって

$$\text{tr} \Lambda^\top \Lambda = \sum_{p=1}^P \sum_{m=1}^M \lambda_{pm}^2 \quad (5.17)$$

である。また  $N \text{tr} \mathbf{X}^\top \mathbf{Q} \mathbf{R} \boldsymbol{\Lambda} = \text{tr} \mathbf{Z} \mathbf{R} \boldsymbol{\Lambda} = \sum_{p=1}^P \sum_{m=1}^M (\mathbf{z}_p^\top \mathbf{r}_m) \lambda_{pm}$  と書ける。ただし  $\mathbf{Z} = N^{-1} \mathbf{X}^\top \mathbf{Q}$  とする。よって、 $\boldsymbol{\Lambda}$  以外のパラメータを固定すると目的関数は

$$\text{const}_{\boldsymbol{\Lambda}} + N \sum_{p=1}^P \sum_{m=1}^M \lambda_{pm}^2 - 2N \sum_{p=1}^P \sum_{m=1}^M (\mathbf{z}_p^\top \mathbf{r}_m) \lambda_{pm}$$

と書き換えられる。したがって、 $\lambda_{pm}$  の更新式は

$$\lambda_{pm} = \begin{cases} \mathbf{z}_p^\top \mathbf{r}_m & (m = \operatorname{argmin}_{1 \leq l \leq M} \{\lambda_{pl}^2 - 2(\mathbf{z}_p^\top \mathbf{r}_l) \lambda_{pl}\}) \\ 0 & (\text{それ以外}) \end{cases} \quad (5.18)$$

である。

## 全体のアルゴリズム

SSFA の全体のアルゴリズムは

1. 初期値を発生させる。
2.  $\mathbf{Z}$  を (5.11) 式で更新する。
3.  $\boldsymbol{\Psi}$  を (5.13) 式で更新する。
4.  $\mathbf{R}$  を (5.16) 式で更新する。
5.  $\boldsymbol{\Lambda}$  を (5.18) 式で更新する。
6. 収束判定をし、収束していない場合は 2. に戻る。

である。初期値を複数回発生させて、最も目的関数が小さくなった場合のパラメータを最適解とみなす。

### 5.1.4 実データ解析

本小節では、SSFA を実際のデータに適用し有用性を検証する。用いるデータセットはビッグファイブと呼ばれるパーソナリティ特性に関するデータである。

<http://bstat.jp/wp-content/uploads/2017/02/Big5.xls> にて公開されているビッグファイブデータは 190 個体 25 変数からなるデータで、ビッグファイブに関する事前知識から



図 5.1: ビッグファイブの概要図

変数である 25 項目は外向性, 神経質, 開放性, 調和性, 誠実性の 5 つの因子に分類されると推察される。

$M = 5$  として SSFA により得られた因子負荷量と独自分散と通常の因子分析にバリマックス回転を加えた因子負荷量と独自分散をそれぞれ表にした。

表 5.1: ビッグファイブデータに SSFA を適用して得たパラメータ

変数	因子 1	因子 2	因子 3	因子 4	因子 5	独自分散
Sociable	<b>0.85</b>	0	0	0	0	0.26
Talkative	<b>0.75</b>	0	0	0	0	0.38
Voluntary	<b>0.76</b>	0	0	0	0	0.39
Cheerful	<b>0.83</b>	0	0	0	0	0.27
Showy	<b>0.64</b>	0	0	0	0	0.55
Worry	0	<b>0.75</b>	0	0	0	0.40
Sensitive	0	<b>0.65</b>	0	0	0	0.52
Pessimistic	0	<b>0.79</b>	0	0	0	0.34
Unrest	0	<b>0.42</b>	0	0	0	0.71
Careful	0	<b>0.71</b>	0	0	0	0.45
Creative	0	0	<b>0.73</b>	0	0	0.41
Adventurous	0	0	<b>0.78</b>	0	0	0.36
Progressive	0	0	<b>0.68</b>	0	0	0.50
Flexible	0	0	<b>0.54</b>	0	0	0.65
Imaginative	0	0	<b>0.41</b>	0	0	0.75
Mild	0	0	0	<b>0.51</b>	0	0.68
Tenderhearted	0	0	0	<b>0.59</b>	0	0.60
Altruistic	0	0	0	<b>0.70</b>	0	0.48
Cooperative	0	0	0	<b>0.68</b>	0	0.50
Sympathetic	0	0	0	<b>0.79</b>	0	0.35
Deliberate	0	0	0	0	<b>0.61</b>	0.59
Reliable	0	0	0	0	<b>0.60</b>	0.53
Diligent	0	0	0	0	<b>0.77</b>	0.38
Systematic	0	0	0	0	<b>0.64</b>	0.55
Methodical	0	0	0	0	<b>0.77</b>	0.35

それぞれの要素の値は、小数点第 3 位で四捨五入した。また行のうちで最も大きい値は太字とした。

表 5.2: ビッグファイブデータに行列モデル因子分析を適用して得たパラメータ

変数	因子 1	因子 2	因子 3	因子 4	因子 5	独自分散
Sociable	<b>0.81</b>	-0.17	0.11	0.11	0.12	0.27
Talkative	<b>0.82</b>	0.04	-0.04	0	-0.08	0.30
Voluntary	<b>0.71</b>	-0.11	0.18	0.12	0.13	0.40
Cheerful	<b>0.79</b>	-0.23	0.08	0.18	0	0.27
Showy	<b>0.64</b>	0.05	0.23	-0.01	-0.05	0.52
Worry	-0.13	<b>0.71</b>	-0.11	0.22	0.14	0.39
Sensitive	-0.04	<b>0.75</b>	-0.02	0.01	-0.06	0.43
Pessimistic	-0.27	<b>0.67</b>	-0.12	-0.04	0.29	0.37
Unrest	0.07	<b>0.58</b>	-0.13	-0.08	-0.30	0.53
Careful	-0.08	<b>0.61</b>	-0.15	-0.10	0.33	0.47
Creative	0.02	-0.11	<b>0.85</b>	-0.05	-0.06	0.26
Adventurous	0.20	-0.23	<b>0.67</b>	-0.01	-0.19	0.40
Progressive	0.23	-0.21	<b>0.62</b>	0.09	0.09	0.50
Flexible	0.29	-0.29	<b>0.35</b>	0.20	0.03	0.66
Imaginative	0.09	0.18	<b>0.46</b>	0.12	-0.32	0.62
Mild	-0.12	-0.14	0.15	<b>0.59</b>	0.03	0.59
Tenderhearted	0.18	0.07	0.12	<b>0.60</b>	0.01	0.57
Altruistic	-0.01	0.12	0	<b>0.69</b>	0.16	0.46
Cooperative	0.25	0	-0.11	<b>0.64</b>	0.10	0.49
Sympathetic	0.11	0.10	-0.03	<b>0.72</b>	0.22	0.39
Deliberate	-0.04	0.12	0.06	0.22	<b>0.57</b>	0.59
Reliable	0.30	-0.15	0.08	0.27	<b>0.56</b>	0.49
Diligent	0.04	-0.06	-0.08	0.15	<b>0.77</b>	0.37
Systematic	-0.02	0.08	-0.10	0.01	<b>0.69</b>	0.49
Methodical	0.02	0.18	-0.19	0.04	<b>0.73</b>	0.38

それぞれの要素の値は、小数点第 3 位で四捨五入した。また行のうちで最も大きい値は太字とした。

表 5.1 と表 5.2 を比較すると、通常の因子分析の場合は因子負荷量の解釈は多くの要素を考慮する必要がある。しかし SSFA の場合は、因子負荷量の解釈は容易であることがわかる。また、SSFA の場合には回転を施す必要がないことも特徴である。これは因子負荷量に 0 の要素が多いことに起因する。

さらに、通常の行列モデル因子分析では直交制約という強い制約を設けていたが、SSFA

では因子間相関を推定可能であることも特徴である。ビッグファイブデータにSSFAを適用し得られた因子間相関行列は表5.3という結果になった。しかしSSFAの制約である完

表 5.3: ビッグファイブデータに SSFA を適用して得た因子間相関

	因子 1	因子 2	因子 3	因子 4	因子 5
因子 1	1	-0.29	0.41	0.25	0.13
因子 2	-0.29	1	-0.43	0.15	0.24
因子 3	0.41	-0.43	1	0.10	-0.18
因子 4	0.25	0.15	0.10	1	0.36
因子 5	0.13	0.24	-0.18	0.36	1

全単純構造制約は、全てのデータに対して有用というにはあまりにも制約が強すぎる。そこで完全単純構造制約を緩和し、因子負荷行列におけるいくつかの要素が0になるという制約を検討してみる。こうしたパラメータの要素のいくつかを0として推定する手法は、近年注目を集めているスパース推定を用いることで達成可能である。そこで、まずスパース推定について初步的な解説を行う。

## 5.2 スパース推定

パラメータの要素が0であるという特徴は、因子分析に限らず多変量データ解析手法では望ましいとされている。なぜならパラメータが0ということは、もはやそのパラメータを考慮する必要がないということ意味しており、変数選択という問題に関連している。変数選択をどのように行うべきかはいまだ結論の出ていない研究課題であるが、標準的な方法は情報量規準を用いて最適なモデルを採用する方法である。情報量規準は、モデルとデータの適合だけでなくモデルの複雑さも考慮に入れた指標であり、AIC(Akaike, 1974)やBIC(Schwarz, 1978)などが代表例である。一方で各モデルのパラメータ推定と情報量規準の計算という2ステップの計算を繰り返す必要があり、煩雑になることもある。

Tibshirani(1996)によって提案されたlasso(Least Absolute Shrinkage and Selection Operator)は、変数選択とパラメータ推定を同時に行うことができ、さらに通常では計算が不可能であった高次元データにも対応していることから提案から数年で統計学における一大トピックとなった。実際に様々な教科書やサーベイ論文が出版されており(Bühlmann & van de Geer, 2011; Hastie, Tibshirani & Wainwright, 2015), 日本国内でも大いに関心

を集めている(川野・廣瀬・立石・小西, 2010; 川野・松井・廣瀬, 2018).  $\ell_1$  ノルムは  $\ell_0$  ノルムの近似としては、凸性とスパース性という2つの好ましい特徴を持っている点で最適である。

### 5.2.1 lasso ペナルティとパラメータ推定

lasso の目的関数は、説明変数を表すデータ行列  $\mathbf{X} \in \mathbb{R}^{N \times P}$  と目的変数を表すベクトル  $\mathbf{y} \in \mathbb{R}^N$  が与えられたとき

$$f(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \quad (5.19)$$

である。そして最適なパラメータ  $\boldsymbol{\beta} \in \mathbb{R}^P$  は、 $\operatorname{argmin}_{\boldsymbol{\beta}} f(\boldsymbol{\beta})$  である。つまり目的関数(5.19)を最小化することで得られる。回帰分析では微分を用いて最小値をとるパラメータを推定することが一般的である。ところが、パラメータを含んだ項である  $\|\boldsymbol{\beta}\|_1$  は微分できないことがわかる。

よって通常の微分の概念を拡張した劣微分を用いる必要がある。劣微分は通常の微分を拡張した概念であり、劣勾配が複数とれる場合がある。実際、 $\boldsymbol{\beta} = [\beta_1, \dots, \beta_P]$  として各スカラー  $\beta_p$  ごとに劣微分すると

$$\frac{\partial |\beta_p|}{\partial \beta_p} \equiv d(\beta_p) = \begin{cases} 1 & (\beta_p > 0) \\ [-1, 1] & (\beta_p = 0) \\ -1 & (\beta_p < 0) \end{cases} \quad (5.20)$$

となる。(5.20)式を用いて Lasso の目的関数(5.20)式を各スカラー  $\beta_p$  について劣微分し

$$\begin{aligned} \frac{\partial f(\beta_p)}{\partial \beta_p} &= \frac{\partial}{\partial \beta_p} (\|\mathbf{y} - \mathbf{x}_p \beta_p\|_2^2 + \lambda |\beta_p|) \\ &= -2\mathbf{x}_p^\top \mathbf{y} + 2\mathbf{x}_p^\top \mathbf{x}_p \beta_p + \lambda d(\beta_p) = 0 \end{aligned} \quad (5.21)$$

とすれば,  $\beta_p$  を以下の式で更新することができる.

$$\begin{aligned}\beta_p &= \frac{\mathbf{x}_p^\top \mathbf{y}}{\mathbf{x}_p^\top \mathbf{x}_p} - \frac{\lambda d(\beta_p)}{2\mathbf{x}_p^\top \mathbf{x}_p} \\ &= \begin{cases} \beta_p^{\text{OLS}} - \lambda(2\mathbf{x}_p^\top \mathbf{x}_p)^{-1} & (\beta_p^{\text{OLS}} > 0 \text{かつ } |\beta_p^{\text{OLS}}| > \lambda(2\mathbf{x}_p^\top \mathbf{x}_p)^{-1}) \\ \beta_p^{\text{OLS}} + \lambda(2\mathbf{x}_p^\top \mathbf{x}_p)^{-1} & (\beta_p^{\text{OLS}} < 0 \text{かつ } |\beta_p^{\text{OLS}}| > \lambda(2\mathbf{x}_p^\top \mathbf{x}_p)^{-1}) \\ 0, & (|\beta_p^{\text{OLS}}| \leq \lambda(2\mathbf{x}_p^\top \mathbf{x}_p)^{-1}) \end{cases} \\ &= \text{sign}(\beta_p^{\text{OLS}}) (|\beta_p^{\text{OLS}}| - \lambda(2\mathbf{x}_p^\top \mathbf{x}_p)^{-1})_+\end{aligned}\quad (5.22)$$

ただし  $(a)_+$  は  $\min(a, 0)$  を意味する. (5.22) 式は, 通常の回帰分析で得られるパラメータ  $\beta_p^{\text{OLS}}$  から  $\lambda(2\mathbf{x}_p^\top \mathbf{x}_p)^{-1}$  だけ値の絶対値が縮小し, 一定の割合を超えると 0 になることを意味している. この場合分けは図で表すことで理解が容易になる.

実際, このずれが結果として  $\beta_p$  の推定量にバイアスを発生させてしまう. この問題が指摘されてから (Fan & Li, 2001), 値を 0 にしつつバイアスが少なくなるようなペナルティが数多く提案されている. SCAD(Smoothly Clipped Absolute Deviation) や Adaptive Lasso, MCP(Minimax Concave Penalty), Square-root Lasso などが代表例である.

ところで, lasso の目的関数は

$$f(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$$

であったが, ペナルティ項を  $P_\lambda(\boldsymbol{\beta})$  とすると

$$f(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + P_\lambda(\boldsymbol{\beta}) \quad (5.23)$$

のように一般的な形で書くことができる.  $P_\lambda(\boldsymbol{\beta})$  を lasso ペナルティ以外のペナルティにすることも可能で, それぞれのペナルティに利点がある.

## SCAD

lasso ペナルティが提案されてから初めて提案された非凸なペナルティは, SCAD(Fan & Li, 2001) であると考えられる. SCAD の考え方はシンプルで,  $\|\boldsymbol{\beta}\|_1$  が微分不可能な点は 0 の付近のみであることから, 場合分けによって 0 から大きく離れた点ではペナルティ

を加えないようにしている。SCAD のペナルティを  $P_\lambda(\beta)$  とすると実際に

$$P_\lambda(\beta) = \sum_{p=1}^P P_\lambda(\beta_p) = \sum_{p=1}^P \begin{cases} \lambda|\beta_p| & (|\beta_p| \leq \lambda) \\ -(|\beta_p|^2 - 2a\lambda|\beta_p| + \lambda^2)/2(a-1) & (\lambda < |\beta_p| \leq a\lambda) \\ (a+1)\lambda^2/2 & (a\lambda < |\beta_p|) \end{cases} \quad (5.24)$$

と場合分けしている。 $a > 2$  はチューニングパラメータであるが、Fan & Li(2001) では  $a = 3.7$  としている。一見 (5.24) 式は狙いが見えづらいが、 $P_\lambda(\beta_p)$  の  $\beta_p$  についての劣勾配を計算するとわかりやすい。劣勾配は

$$\frac{\partial P_\lambda(\beta_p)}{\partial \beta_p} = \begin{cases} \lambda d(\beta_p) & (|\beta_p| \leq \lambda) \\ -(|\beta_p| - a\lambda d(\beta_p))/(a-1) & (\lambda < |\beta_p| \leq a\lambda) \\ 0 & (a\lambda < |\beta_p|) \end{cases} \quad (5.25)$$

となるため、 $|\beta_p| \leq \lambda$  の場合は (5.20) 式、つまり lasso の場合と同様の  $\lambda d(\beta_p)$  となる。対して  $\lambda < |\beta_p| \leq a\lambda$  の場合、言い換えるなら  $\beta_p$  の推定値が 0 からは離れているものの充分に離れているとはいえない場合は、少し複雑なペナルティとなっていることがわかる。そして  $a\lambda < |\beta_p|$  の場合は、 $\beta_p$  の推定値が 0 から充分に離れているため  $\partial P_\lambda(\beta_p)/\partial \beta_p = 0$ 、つまり微分したときには推定値のペナルティがないと解釈できる。

## MCP

MCP(Zhang, 2010) は SCAD と同様に非凸かつ oracle property を有するペナルティでありベクトル  $\beta \in \mathbb{R}^P$  についての MCP を  $P_\lambda(\beta)$  とすると

$$\begin{aligned} P_\lambda(\beta) &= \sum_{p=1}^P P_\lambda(\beta_p) = \sum_{p=1}^P \lambda \int_0^{\beta_p} (1 - x/\lambda\gamma)_+ dx \\ &= \sum_{p=1}^P \begin{cases} \lambda|\beta_p| - \beta_p^2/(2\gamma) & (|\beta_p| \leq \gamma\lambda) \\ (\lambda^2\gamma)/2 & (\gamma\lambda < |\beta_p|) \end{cases} \end{aligned} \quad (5.26)$$

と書ける。ここで  $\gamma > 1$  はチューニングパラメータであり、 $\lambda \rightarrow 1$  のとき MCP は  $\ell_0$  ノルムペナルティに近似され、 $\lambda \rightarrow \infty$  のとき MCP は lasso ペナルティに近似される。

複雑な式に見えがちだが、こちらも微分することで狙いが明確になる。ペナルティ項  $P_\lambda(\beta_p)$  を  $\beta_p$  について微分すると

$$\frac{\partial P_\lambda(\beta_p)}{\partial \beta_p} = \begin{cases} \lambda d(\beta_p) - \beta_p/\gamma & (|\beta_p| \leq \gamma\lambda) \\ 0 & (\gamma\lambda < |\beta_p|) \end{cases} \quad (5.27)$$

となる。SCAD の場合の (5.25) 式と同じように、一定以上  $\beta_p$  の絶対値が大きい場合にはペナルティ項の勾配が 0 となり、バイアスがなくなることが確認できる。また閉区間の不定積分であるから、MCP は SCAD と同じく、連続な関数であることがわかる。

### 5.2.2 因子分析におけるスパース推定

これまで回帰分析におけるスパース推定について述べてきたが、因子分析にもスパース推定は活用されている。本節では、因子分析にスパース推定を用いた事例を紹介する。最初に因子分析にスパース推定を取り入れた研究は Choi, Zou, & Oehlert(2011) によるもので、その後 Hirose & Yamamoto(2014; 2015) によって望ましい特徴を持った手法に発展した。いずれもスパースにするパラメータは、因子負荷量である。

因子負荷量をスパースにするという考え方には、決して非現実的で突飛な発想ではない。本章にて述べたように、因子負荷量が満たしていることが望ましい性質はスパースモデリングが提案される前に提示されており (Thurstone, 1947), 具体的には

1. 各行に少なくとも 1 つ、要素が 0 であること
2. 各列に因子数と同じ数の要素が 0 であること
3. 2 つの列に着目したとき、一方の列では 0 でもう一方では 0 でない行が複数あること
4. もし因子数が 4 以上なら、2 つの列に着目したとき、どちらの列も要素が 0 であること
5. 2 つの列に着目したとき、どちらの要素も 0 でない行が少ないと

であった。これらを目指すべく回転法の研究が盛んに行われたが、スパース制約は Thurstone の単純構造の性質を満たすと考えられる。実際、Hirose & Yamamoto(2015) ではスパース因子分析がある回転基準の一般化であることが示されている。

2 章で述べた因子分析モデルは全て直交モデルと呼ばれるものである。つまり共通因子  $f \in \mathbb{R}^{M \times 1}$  の分散が  $\mathbf{I}_M$  であることを想定している。これは現実的とはいえない仮定であ

り, Hirose & Yamamoto(2014) では直交モデルの制約を緩めた斜交モデルのスパース因子分析を提案している.

パス図で表すと直交と斜交の違いはわかりやすい. 図 5.2 と図 5.3 における各線は変数あるいは因子の間での関係性を示しているが, 直交モデルでは因子間に関係がない. つまり各因子は無相関という強い仮定を置いている. それに対して斜交モデルでは因子間に関係があることを想定しており, より現実に即したモデルと言える.

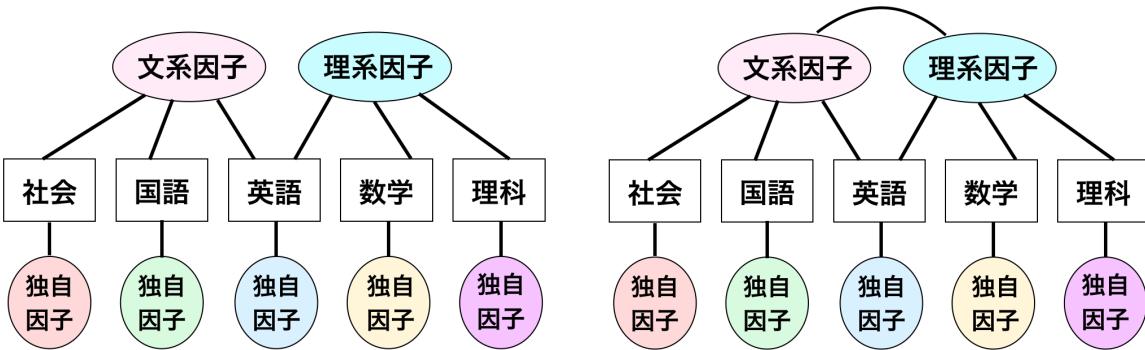


図 5.2: 直交モデルでの因子分析のパス図

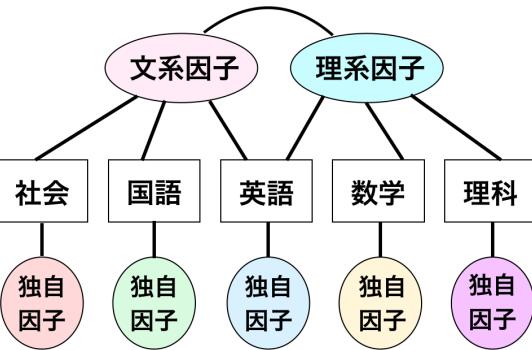


図 5.3: 斜交モデルでの因子分析のパス図

## モデルとアルゴリズム

スパース因子分析のモデルを記述する. データベクトルを  $x \in \mathbb{R}^P$  としたとき

$$x = \Lambda f + u \quad (5.28)$$

とする. ここで共通因子得点と独自因子得点はそれぞれ

$$f \sim N(\mathbf{0}_M, \Phi) \quad (5.29)$$

$$u \sim N(\mathbf{0}_P, \Psi^2) \quad (5.30)$$

に従う. ただし  $\Phi \in \mathbb{R}^{M \times M}$  は因子間相関行列である.

目的関数は, スパース制約を加えない対数尤度を  $L(\Theta)$  とすると

$$L(\Theta) - P(\Lambda, \rho) \quad (5.31)$$

である. ここで  $P(\Lambda, \rho)$  は  $\Lambda$  に関するペナルティ項である.

提案モデルのパラメータは,  $\mathbf{f}$  を欠損値としてみなすことによる EM アルゴリズムで推定する. そのためアルゴリズムは E ステップと M ステップの二つに分けられる.

### -E ステップ-

$\mathbf{x}$  の平均は

$$\mathbb{E}[\mathbf{x}] = \Lambda \mathbb{E}[\mathbf{f}] + \mathbb{E}[\mathbf{u}] = \mathbf{0}_P \quad (5.32)$$

であるから,  $\mathbf{x}$  の分散は

$$\begin{aligned} \mathbb{V}[\mathbf{x}] &= \mathbb{E} \left[ (\mathbf{x} - \mathbb{E}[\mathbf{x}]) (\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top \right] \\ &= \mathbb{E} [(\Lambda \mathbf{f} + \mathbf{u})(\Lambda \mathbf{f} + \mathbf{u})^\top] \\ &= \Lambda \mathbb{E} [\mathbf{f} \mathbf{f}^\top] \Lambda^\top + \Lambda \mathbb{E} [\mathbf{f} \mathbf{u}^\top] + \mathbb{E} [\mathbf{u}^\top \mathbf{f}] \Lambda^\top + \mathbb{E} [\mathbf{u} \mathbf{u}^\top] \\ &= \Lambda \Phi \Lambda^\top + \Psi^2 \end{aligned} \quad (5.33)$$

である.

ここで条件付き多変量正規分布を考える. まず同時分布が

$$\begin{aligned} \begin{bmatrix} \mathbf{x} \\ \mathbf{f} \end{bmatrix} &\sim N \left( \begin{bmatrix} \mathbf{0}_P \\ \mathbf{0}_M \end{bmatrix}, \begin{bmatrix} \text{Var}[\mathbf{x}] & \text{Cov}[\mathbf{x}, \mathbf{f}] \\ \text{Cov}[\mathbf{f}, \mathbf{x}] & \text{Var}[\mathbf{f}] \end{bmatrix} \right) \\ &= N \left( \begin{bmatrix} \mathbf{0}_P \\ \mathbf{0}_M \end{bmatrix}, \begin{bmatrix} \Lambda \Phi \Lambda^\top + \Psi^2 & \Lambda \Phi \\ \Phi \Lambda^\top & \Phi \end{bmatrix} \right) \end{aligned} \quad (5.34)$$

であるから,

$$\begin{aligned} \mathbb{E}[\mathbf{f}|\mathbf{x}] &= \mathbb{E}[\mathbf{f}] + \mathbb{E}[\mathbf{f}\mathbf{x}^\top] \mathbb{E}[\mathbf{x}\mathbf{x}^\top]^{-1} (\mathbf{x} - \mathbb{E}[\mathbf{x}]) \\ &= \Phi \Lambda^\top (\Lambda \Phi \Lambda^\top + \Psi^2)^{-1} \mathbf{x} \\ &= \Phi \mathbf{B} \mathbf{x} \end{aligned} \quad (5.35)$$

である. ただし

$$\mathbf{B} = \Lambda^\top (\Lambda \Phi \Lambda^\top + \Psi^2)^{-1} \quad (5.36)$$

とする。さらに条件付き分散が

$$\begin{aligned}\text{Var}[\mathbf{f}|\mathbf{x}] &= \mathbb{E} \left[ (\mathbf{f} - \mathbb{E}[\mathbf{f}|\mathbf{x}]) (\mathbf{f} - \mathbb{E}[\mathbf{f}|\mathbf{x}])^\top | \mathbf{x} \right] \\ &= \mathbb{E} [\mathbf{f}\mathbf{f}^\top | \mathbf{x}] - \mathbb{E}[\mathbf{f}|\mathbf{x}] \mathbb{E}[\mathbf{f}|\mathbf{x}]^\top\end{aligned}\quad (5.37)$$

であるから、

$$\begin{aligned}\mathbb{E} [\mathbf{f}\mathbf{f}^\top | \mathbf{x}] &= \text{Var}[\mathbf{f}|\mathbf{x}] + \mathbb{E}[\mathbf{f}|\mathbf{x}] \mathbb{E}[\mathbf{f}|\mathbf{x}]^\top \\ &= \text{Var}[\mathbf{f}] - \text{Cov}[\mathbf{f}, \mathbf{x}] \text{Var}[\mathbf{x}]^{-1} \text{Cov}[\mathbf{x}, \mathbf{f}] + \mathbb{E}[\mathbf{f}|\mathbf{x}] \mathbb{E}[\mathbf{f}|\mathbf{x}]^\top \\ &= \boldsymbol{\Phi} - \boldsymbol{\Phi} \boldsymbol{\Lambda}^\top (\boldsymbol{\Lambda} \boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}^2)^{-1} \boldsymbol{\Lambda} \boldsymbol{\Phi} + \boldsymbol{\Phi} \mathbf{B} \mathbf{x} \mathbf{x}^\top \mathbf{B}^\top \boldsymbol{\Phi} \\ &= \boldsymbol{\Phi} - \boldsymbol{\Phi} \mathbf{B} \boldsymbol{\Lambda} \boldsymbol{\Phi} + \boldsymbol{\Phi} \mathbf{B} \mathbf{x} \mathbf{x}^\top \mathbf{B}^\top \boldsymbol{\Phi}\end{aligned}\quad (5.38)$$

となる。

### -Mステップ-

ベイズの定理より  $\int p(\mathbf{x}, \mathbf{f}) d\mathbf{f} = \int p(\mathbf{x}|\mathbf{f}) p(\mathbf{f}) d\mathbf{f}$  である。そこで、最大化すべき Q 関数を得るために、条件付き分布の期待値および分散の定義から  $\mathbf{x}|\mathbf{f}$  の従う分布のパラメータ

$$\mathbb{E}[\mathbf{x}|\mathbf{f}] = \boldsymbol{\Lambda} \mathbf{f} \quad (5.39)$$

$$\text{Var}[\mathbf{x}|\mathbf{f}] = \boldsymbol{\Psi}^2 \quad (5.40)$$

を計算する。したがって Q 関数は

$$\begin{aligned}Q &= \mathbb{E} \left[ \log \prod_{n=1}^N p(\mathbf{x}_n|\mathbf{f}) p(\mathbf{f}_n) \right] \\ &= \mathbb{E} \left[ \log \prod_{n=1}^N (2\pi)^{P/2} \det(\boldsymbol{\Psi}^2)^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n)^\top \boldsymbol{\Psi}^{-2} (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n) \right\} \right. \\ &\quad \times (2\pi)^{M/2} \det(\boldsymbol{\Phi})^{-1/2} \exp \left( -\frac{1}{2} \mathbf{f}^\top \boldsymbol{\Phi} \mathbf{f} \right) \Bigg] \\ &= \text{const} - N \det(\boldsymbol{\Psi}^2) - N \det(\boldsymbol{\Phi}) - \sum_{n=1}^N \left\{ \mathbb{E}[(\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n)^\top \boldsymbol{\Psi}^{-2} (\mathbf{x}_n - \boldsymbol{\Lambda} \mathbf{f}_n) + \mathbf{f}_n^\top \boldsymbol{\Phi}^{-1} \mathbf{f}_n] \right\} \\ &\quad (5.41)\end{aligned}$$

となり, 目的関数である罰則対数尤度は以下のように書ける.

$$L^p = Q - P(\Lambda, \rho)$$

Mステップでは, 罰則対数尤度を最大化することで各パラメータを更新するが, まず  $\Lambda$  の更新を行う.  $\Lambda$  のスペース推定には, 様々な先行研究がある. ここでは, coordinate-descent による推定を行う.

$$\begin{aligned}\mathbf{A} &= \sum_{n=1}^N \{\mathbb{E}[\mathbf{f}\mathbf{f}^\top | \mathbf{x}_n]\} \\ \mathbf{B} &= \sum_{n=1}^N \{\mathbb{E}[\mathbf{f}|\mathbf{x}_n] \mathbf{x}_n^\top\}\end{aligned}$$

とすると, 目的関数は

$$\text{const}_{\Lambda} - N \left( \log \psi_p^2 - \psi_p^{-2} (s_{pp} - 2\boldsymbol{\lambda}_p^\top \mathbf{b}_p + \boldsymbol{\lambda}_p^\top \mathbf{A} \boldsymbol{\lambda}_p) \right) - 2NP(\Lambda, \rho)$$

と書き換えられる. ただし  $\text{const}_{\Lambda}$  は  $\Lambda$  に関係しない定数,  $\mathbf{b}_p$  は  $\mathbf{B}$  の  $p$  列ベクトルとする. したがって,

$$\lambda_{pm} = \operatorname{argmin}_{\lambda_{pm}} \frac{1}{2} \left\{ \lambda_{pm} - a_{mm}^{-1} \left( b_{mp} - \sum_{l \neq m} a_{lm} \tilde{\lambda}_{pl} \right) \right\}^2 + \frac{\psi_p^2}{a_{mm}} P(\lambda_{pm}, \rho)$$

によって  $\lambda_{pm}$  は求まる. ただし,  $a_{mm}$  は行列  $\mathbf{A}$  の  $m$  行  $m$  列,  $b_{mp}$  は行列  $\mathbf{B}$  の  $m$  行  $p$  列.

たとえばペナルティを lasso ペナルティとすれば,

$$\lambda_{pm} = \operatorname{sign}(\theta) \left( |\theta| - \frac{\psi_p^2 \rho}{a_{pp}} \right)_+$$

となる. ただし

$$\theta = a_{mm}^{-1} \left( b_{pm} - \sum_{l \neq m} a_{lm} \tilde{\lambda}_{pl} \right)$$

とする.

つづいて  $\Psi^2$  を更新する. 罰則対数尤度を  $\Psi^{-2}$  について偏微分し, 尤度方程式

$$N\Psi^2 - \sum_{n=1}^N \{\mathbf{x}_n \mathbf{x}_n^\top\} + 2\Lambda \sum_{n=1}^N \{\mathbb{E}[\mathbf{f}|\mathbf{x}_n] \mathbf{x}_n^\top\} - \Lambda \left( \sum_{n=1}^N \{\mathbb{E}[\mathbf{f}\mathbf{f}^\top | \mathbf{x}_n]\} \right) \Lambda^\top = \mathbf{O}_{P \times P}$$

を得る。よって  $\Psi^2$  の更新は

$$\Psi^2 = \frac{1}{N} \text{diag} \left( \sum_{n=1}^N \{\mathbf{x}_n \mathbf{x}_n^\top\} - 2\Lambda \sum_{n=1}^N \{\mathbb{E}[\mathbf{f}|\mathbf{x}_n] \mathbf{x}_n^\top\} + \Lambda \left( \sum_{n=1}^N \{\mathbb{E}[\mathbf{f}\mathbf{f}^\top|\mathbf{x}_n]\} \right) \Lambda^\top \right) \quad (5.42)$$

によって達成される。

## 実データ解析

スパース因子分析の実データの有用性を示すために、実データに適用する。

まず、SSFA の適用例でもあるビッグファイブのデータに適用した。得られた因子負荷量と独自分散を表 5.4 に記した。用いたペナルティは MCP としたためチューニングパラメータは 2 つあり、スパース性をコントロールする  $\rho > 0$  と、 $\ell_1$  ノルムから  $\ell_0$  ノルムへの調整を行う  $\gamma > 1$  である。今回は  $\rho = 0.33, \gamma = 1.5$  とした。

SSFA との結果と比較すると、完全単純構造制約を課していないにも関わらず SSFA の因子負荷量と近いことがわかる。すなわちスパース因子分析は、SSFA の制約をより一般化した因子分析と考えることができる。さらに全ての要素に当てはまるわけではないが、因子負荷量の絶対値が SSFA に比べ小さい傾向にある。これは、スパースペナルティによる縮小が影響していると考えられる。lasso に比べ MCP は縮小の影響が抑えられているが、それでも多少の影響は避けられないことが因子負荷量から推察できる。

また、独自分散に着目すると変数「Sympathetic」のように同じような因子負荷量であるにもかかわらず、スパース因子分析の独自分散が大きいことがわかる。これは、足立 (2014) にて報告されている行列モデル因子分析の傾向と一致しており、2.1.4 小節にて考察された説を支持する結果と言える。

表 5.4: ビッグファイブデータにスパース因子分析を適用して得たパラメータ

変数	因子 1	因子 2	因子 3	因子 4	因子 5	独自分散
Sociable	<b>0.73</b>	0	0	0	0	0.28
Talkative	<b>0.64</b>	0	0	0	0	0.44
Voluntary	<b>0.77</b>	0	0	0	0	0.43
Cheerful	<b>0.53</b>	0	0	0	0	0.32
Showy	<b>0.71</b>	0	0	0	0	0.60
Worry	0	<b>0.83</b>	0	0	0	0.45
Sensitive	0	<b>0.73</b>	0	0	0	0.58
Pessimistic	0	<b>0.74</b>	0	0	0	0.34
Unrest	0	<b>0.80</b>	0	0	0.31	0.67
Careful	0	<b>0.62</b>	0	0	0	0.49
Creative	0	0	<b>0.73</b>	0	0	0.44
Adventurous	0	0	<b>0.77</b>	0	0	0.36
Progressive	0	0	<b>0.64</b>	0	0	0.56
Flexible	0	0	<b>0.42</b>	0	0	0.78
Imaginative	0	0	<b>0.24</b>	0	0	0.87
Mild	0	0	0	<b>0.51</b>	0	0.74
Tenderhearted	0	0	0	<b>0.56</b>	0	0.69
Altruistic	0	0	0	<b>0.68</b>	0	0.54
Cooperative	0	0	0	<b>0.69</b>	0	0.53
Sympathetic	0	0	0	<b>0.79</b>	0	0.37
Deliberate	0	0	0	0	<b>0.60</b>	0.63
Reliable	0	0.26	0	0	<b>0.54</b>	0.56
Diligent	0	0	0	0	<b>0.77</b>	0.39
Systematic	0	0	0	0	<b>0.66</b>	0.55
Methodical	0	0	0	0	<b>0.75</b>	0.42

次に Thurstone のボックスデータに適用した。ペナルティは MCP を用い、チューニングパラメータは  $\rho = 1.0, \gamma = 1.1$  とした。得られたパラメータを表 5.5 に記した。

結果を考察すると、完全単純構造ではないが要素に 0 が多く解釈しやすく、さらに妥当な構造の因子負荷量を得ることができていることがわかる。よって、完全単純構造を仮定するには不適切と思われるデータについてもスパース因子分析は適用可能であることがボックスデータからわかる。

表 5.5: Thurstone のボックスデータにスパース因子分析を適用して得たパラメータ

変数	因子 1	因子 2	因子 3	独自分散
$x^2$	<b>0.49</b>	0	0	0.11
$y^2$	0	<b>0.50</b>	0	0.10
$z^2$	0	0	<b>0.48</b>	0.14
$xy$	<b>0.31</b>	0.29	0	0.21
$xz$	0.28	0	<b>0.29</b>	0.22
$yz$	0	<b>0.30</b>	0.29	0.19
$(x^2 + y^2)^{1/2}$	<b>0.34</b>	0.32	0	0.12
$(x^2 + z^2)^{1/2}$	<b>0.33</b>	0	<b>0.31</b>	0.13
$(y^2 + z^2)^{1/2}$	0	0.31	<b>0.33</b>	0.13
$2x + 2y$	0.34	<b>0.35</b>	0	0.09
$2x + 2z$	0.33	0	<b>0.34</b>	0.09
$2y + 2z$	0	0.32	<b>0.33</b>	0.10
$\log x$	<b>0.44</b>	0	0	0.22
$\log y$	0	<b>0.43</b>	0	0.25
$\log z$	0	0	<b>0.44</b>	0.23
$xyz$	0.16	0.18	<b>0.21</b>	0.32
$(x^2 + y^2 + z^2)^{1/2}$	<b>0.27</b>	0.24	0.25	0.12
$\exp x$	<b>0.26</b>	0	0	0.54
$\exp y$	0	<b>0.23</b>	0	0.60
$\exp z$	0	0	<b>0.25</b>	0.56

それぞれの要素の値は、小数点第 3 位で四捨五入した。また行のうちで最も大きい値は太字とした。

### 5.3 考察

スパース推定は元々回帰分析の文脈で提案された。そのため、回帰係数を 0 にするということは変数選択と同等であった。しかし因子分析の場合、因子数が  $M > 1$  のときは変数選択が行われたとはいえない。さらに、たとえある行の因子負荷量が 0 になったとしても独自因子がモデルに組み込まれているため、完全に変数の影響を排除したとはいえない。そのため、因子分析におけるスパース推定は回帰分析などにおける変数選択とは別の文脈で解釈する必要がある。実際、Hirose & Yamamoto(2015) ではスパースペナルティが回転

基準の一般化であることが述べられている。今後、因子分析に適した新たなスパースペナルティが

# 第6章

## 総合考察

因子分析は心理学だけでなく計量経済学や生物学など応用範囲の広い手法であるが、回転の不定性や因子得点の不定性といった問題も抱えている。この問題は共通因子得点あるいは独自因子得点が確率変数であることに起因するものではない。実際、共通因子をパラメータとした母数モデル因子分析、共通因子と独自因子をパラメータとした行列モデル因子分析にも問題は存在する。

### 6.1 因子分析における不定性の解消

本論文では、解釈を容易にすると考えられるクラスタリング法を因子分析法と組み合わせることで、解釈性を向上させるとともに因子分析の問題を解消する新たな手法を提案した。

#### 6.1.1 回転の不定性の解消

まず回転の不定性という大きな問題について考える。回転の不定性は確率モデル因子分析、母数モデル因子分析、行列モデル因子分析の全てに当てはまる問題であり、何らかの対処しないことにはパラメータが一意に定まらないため、解釈が目的の1つである因子分析において極めて重要な問題である。近年この不定性を解消しうる画期的な研究が進んでおり、それがスパース因子分析である。スパース推定により因子負荷行列のうちの一定数以上の要素を0とし、さらに0となる要素の場所もパラメータ推定と同時に推定することができる。

これはある意味では、事前に0の場所を分析者が指定する必要があった確認的因子分析を特殊例として含んでいるとも考えられる。つまり、従来は回転を伴う探索的因子分析と確認的因子分析を使い分けていたが、スパース因子分析により、統一的に因子分析による



図 6.1: スパース因子分析の位置づけ

分析が行える可能性が示唆される。もちろんスパース因子分析にはない確認的因子分析の利点もあるため、完全に移行することはないものの、スパース推定により心理統計学における分析手法の選択肢が広がることが期待される。

### 6.1.2 因子得点の不定性の解消

次に因子得点の不定性について考察する。因子分析が提案された初期から因子得点の不定性は指摘されていたものの、この問題を問題視し、さらには解決しようという研究は活発ではなかった。さらに因子得点を用いることで、個体の解釈を行うことについての危険性についても周知されているとはいえない。こうした現状の中、本論文の4章で提案した新しい手法は、行列モデル因子分析での因子得点の不定性をむしろ利用し、共通因子得点の解釈を容易にしつつ解を一意に定める今までにない発想の手法である。むしろ不定性を利用するという発想は、回転法で用いられてきた発想であり、この研究を機に因子負荷量だけでなく共通因子得点あるいは独自因子得点の解釈も積極的に行われる可能性が期待できる。また、4章では解釈しやすい共通因子得点の特徴をグルーピングしている、つまりグループごとに分けられていることを捉え、アルゴリズムを提案した。しかし、グループ化だけが共通因子得点に望まれる特徴ではない。また独自因子得点への応用例は未だ少ないものの、独自因子得点を理想の得点へ近似するアルゴリズムも理論上は提案可能である。そのため、今後の因子得点の推定アルゴリズム開発研究も期待できる。

## 6.2 多変量データ解析における本研究の位置づけ

本研究はクラスタリングおよび次元縮約法に関連した研究といえる。次元縮約法には、因子分析だけでなく主成分分析なども含まれる。主成分分析を拡張し、3相データに適用可能した主成分分析は3相主成分分析あるいはテンソル主成分分析と呼ばれ、これらのパラメータが解釈しやすくなるような手法も研究されている (Ikemoto & Adachi, 2016)。この3相主成分分析のパラメータを解釈するよう、完全単純構造制約を加えた手法も研究されており、因子分析や主成分分析に限らない、より広範囲の手法に対して完全単純構造制約が適用可能であることが示唆される。

完全単純構造という強い制約をより柔軟にしたともいえるスパース推定の研究については5章で言及したが、スパース推定は近年の統計学における大きなトレンドといえる。本論文では回帰分析と因子分析のスパース推定のみを紹介したが、それらに限らず様々な手法に用いられている。

例えば  $k$ -means クラスタリング (Sun, Wang, & Fang, 2012) やファジィ  $c$ -means クラスタリング (宇野, 2018)、あるいは混合モデルにもスパース推定は応用可能であり、変数選択として有用であると考えられている。

また、変数間の条件付き独立性を図示できるグラフィカルモデルにもスパース推定は積極的に活用されており、スパース推定の恩恵を受けて従来の手法では計算不可能であった遺伝子データなどにもグラフィカルモデルが適用可能であることを示している (Danaher, Wang, & Witten, 2014; Friedman, Hastie, & Tibshirani, 2008; 宇野, 2017)。

さらに因子分析を正解、不正解といった2値データを分析可能なように拡張した項目反応モデルにもスパース推定が用いられるようになってきており (Uno & Huang, 2018)，今後さらに多変量解析における一般的な概念として広がることが期待される。

さらに正規分布とは異なる分布を用いた研究も盛んに行われている。 $t$  分布のような正規分布に比べ裾の重い分布を用いることによる外れ値に対して頑健なモデルの提案は、因子分析を含めて研究されている (Zhang, Li, & Liu, 2014)。

## 要約

1章では、本論文で基礎知識となる線形代数や確率分布を多変量データ解析への応用を意識しながら解説した。

2章では、因子分析法の概念と今まで考えられている三つのタイプの因子分析モデルを紹介した。さらにそれぞれのモデルの目的関数を導出した。

3章では、三つのタイプの因子分析法のうち、母数モデル因子分析に着目した。母数モデルが最尤推定できないという問題を解消するために、パラメータである各個体の共通因子得点にクラスター制約を課した。その結果制約付き母数モデルは最尤推定可能となり、さらに類似手法を分類精度の点で上回ることが実データ分析例から示唆された。

4章では、行列モデル因子分析に着目した。行列モデル因子分析も、確率モデル因子分析と同じく因子得点の不定性を持つ。そこで、回転の不定性を解消しつつ解釈しやすい因子負荷行列を得る回転法の発想を共通因子得点に取り入れた。つまり、因子得点の不定性を解消しつつ解釈しやすい共通因子得点が得られるような手法を提案した。個体の解釈に有用と思われる制約はクラスター制約と考え、特にクラスター制約を課したときの目的関数およびアルゴリズムを提案した。

5章では、因子分析法において避けられない回転の不定性に着目した。近年統計学において盛んに研究されているスパース推定を因子分析に適用することで、回転の不定性を解消した既存の研究を紹介した。

6章では、これまでの議論を踏まえ因子分析法について考察を行った。

# 付 錄 A

## A.1

ここでは, (3.22) 式による更新後の  $\mathbf{G}\mathbf{C}$  が制約条件 (3.13) 式を満たすことについて証明する. (3.22) 式と  $\mathbf{M} = n^{\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}\mathbf{G}^\top \mathbf{X}\Psi^{-1}\Lambda$  を使うと,  $\mathbf{G}\mathbf{C}$  は

$$\begin{aligned}\mathbf{G}\mathbf{C} &= n^{\frac{1}{2}}\mathbf{G}\mathbf{D}^{-\frac{1}{2}}\mathbf{M}\mathbf{W}\Delta^{-1}\mathbf{W}^\top \\ &= n(\mathbf{G}\mathbf{D}^{-1}\mathbf{G}^\top \mathbf{X})\Psi^{-1}\Lambda\mathbf{W}\Delta^{-1}\mathbf{W}^\top,\end{aligned}\tag{A.1}$$

と書ける. ここで, (3.9) 式より  $\mathbf{1}_n^\top \mathbf{G} = \mathbf{1}_K^\top \mathbf{D}$ , また  $\mathbf{X}$  は列中心行列であることを用いると  $\mathbf{1}_n^\top \mathbf{G}\mathbf{D}^{-1}\mathbf{G}^\top \mathbf{X}$  は,

$$\mathbf{1}_n^\top \mathbf{G}\mathbf{D}^{-1}\mathbf{G}^\top \mathbf{X} = \mathbf{1}_K^\top \mathbf{D}\mathbf{D}^{-1}\mathbf{G}^\top \mathbf{X} = \mathbf{1}_K^\top \mathbf{G}^\top \mathbf{X} = \mathbf{1}_n^\top \mathbf{X} = \mathbf{0}_p\tag{A.2}$$

であると言える. よって, (3.13) 式は成り立つ.

## A.2

ここでは, 制約条件 (3.9), (3.13), (3.14) 式と制約条件 (3.9), (3.13), (3.26) 式,  $\text{rank}(\mathbf{G}\mathbf{C}) = m$  が等しいとみなせることを証明する. そのためには,

- [A] 前者の制約条件から, 後者の制約条件を示せる.
  - [B] 後者の制約条件を満たす行列は, RKM の目的関数 (3.25) 式を変化させることなく, 前者の制約条件を満たす行列に書き換えられる.
- の二つを確かめる必要がある.

まず [A] だが, (3.9), (3.13), 3.14 式から (3.16) 式が示されることから確かめられる. 加えて, 序章にて  $\Lambda$  は列についてフルランクであることが前提とされていたので,  $\text{rank}(\mathbf{G}\mathbf{C}\Lambda^\top) = m \leq \min(K - 1, p)$  となる. これに加え (3.9) 式の  $\text{rank}(\mathbf{G}) = K$  を用いると, (3.26) 式が導ける.

次に [B] だが, (3.9), (3.13), (3.26) 式において  $\mathbf{H}$  を  $m \times m$  の非特異行列として,  $\tilde{\mathbf{C}} = \mathbf{CH}$ ,  $\tilde{\Lambda} = \Lambda \mathbf{H}'^{-1}$  と書いたとき,

$$\mathbf{GC}\Lambda^\top = \mathbf{GCH}\mathbf{H}^{-1}\Lambda^\top,$$

である.  $\text{rank}(\mathbf{GC}) = m$  なので,  $\tilde{\mathbf{C}}$  が制約条件 (3.14) 式を満たすように  $\mathbf{H} = (n\mathbf{C}^\top \mathbf{DC})^{\frac{1}{2}}$  と置くことができる. このとき,  $\mathbf{G}$  が変化しない限り,  $\mathbf{H}$  もまた不变である.

### A.3

制約条件 (3.27) 式を正規分布の仮定 (3.2) 式に代入すると,  $\mathbf{e} \sim N(\mathbf{0}_p, v\mathbf{I}_p)$  が得られる. このとき対数尤度 (3.15) 式は,

$$LL(\mathbf{G}, \mathbf{C}, \Lambda, v) \propto -np \log v - \frac{1}{v} \|\mathbf{X} - \mathbf{GC}\Lambda^\top\|^2, \quad (\text{A.3})$$

と書くことができる. これは, (3.27) 式の制約を加えた (3.15) 式といえる. また,  $v$  の最尤推定量は,

$$v = \frac{1}{np} \|\mathbf{X} - \mathbf{GC}\Lambda^\top\|^2,$$

である. この最尤推定量を, (A.3) 式に代入すると,

$$-np \log \frac{1}{np} \|\mathbf{X} - \mathbf{GC}\Lambda^\top\|^2 - np,$$

が得られるが, これは RKM の最小二乗基準での目的関数と一致する.

### A.4

ここでは, (3.28) 式を証明する. 誤差行列  $\mathbf{E}$  の第  $j$  列を  $\tilde{\mathbf{e}}_j$  としたとき, FCFA のモデル (3.12) 式は,  $\mathbf{x}_j = \mathbf{GC}\lambda_j + \tilde{\mathbf{e}}_j (j = 1, \dots, p)$  と書き換えることができる.  $\mathbf{x}_j$  を中心化すると, その要素の分散は以下のように表現できる.

$$\frac{1}{n} \|\mathbf{x}_j\|^2 = \frac{1}{n} \|\mathbf{GC}\lambda_j\|^2 + \frac{2}{n} (\mathbf{GC}\lambda_j)^\top \tilde{\mathbf{e}}_j + \frac{1}{n} \|\tilde{\mathbf{e}}_j\|^2. \quad (\text{A.4})$$

ここで,  $\mathbf{GC}\lambda_j$  はパラメータ,  $\tilde{\mathbf{e}}_j$  は (3.2) 式に従う確率変数なので, (A.4) 式の母集団は

$$V(\mathbf{x}_j) = \frac{1}{n} \|\mathbf{GC}\lambda_j\|^2 + \frac{2}{n} (\mathbf{GC}\lambda_j)^\top E(\tilde{\mathbf{e}}_j) + \psi_j, \quad (\text{A.5})$$

と書ける.  $e$  の期待値  $E(\tilde{\mathbf{e}}_j)$  は, (3.2) 式より  $\mathbf{0}_p$  なので, (A.5) 式から (3.28) 式が導ける.

## 付 錄B

### B.1

(4.6) 式における等式  $\mathbf{1}_n^\top \mathbf{K} = \mathbf{0}_p$  は,  $\mathbf{1}_\mathbf{K} = \mathbf{K}$  を意味している.  $\mathbf{K}^\top \mathbf{J}_\mathbf{K} = \mathbf{O}_{p \times n}$  であるから, このことと (4.19) 式より

$$\mathbf{K}^\top \mathbf{K}_\perp = \mathbf{K}^\top \mathbf{J}_1 \mathbf{J}_\mathbf{K} \mathbf{P} = \mathbf{K}^\top \mathbf{J}_\mathbf{K} \mathbf{P} = \mathbf{O}_{p \times m}$$

が導かれる. 同様に,  $\mathbf{L}^\top \mathbf{J}_\mathbf{L} = \mathbf{O}_{p \times (m+p)}$  であるから,

$$\mathbf{L}^\top \mathbf{L}_\perp = \mathbf{L}^\top \mathbf{J}_\mathbf{L} \mathbf{Q} = \mathbf{O}_{p \times m}$$

が導かれる.

### B.2

(4.7) 式における  $\mathbf{1}_n^\top \mathbf{K} = \mathbf{0}_p^\top$  から, 等式  $\mathbf{J}_1 \mathbf{J}_\mathbf{K} = \mathbf{J}_\mathbf{K} \mathbf{J}_1$  は

$$\begin{aligned}\mathbf{J}_1 \mathbf{J}_\mathbf{K} &= (\mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^\top)(\mathbf{I}_n - \mathbf{K} \mathbf{K}^\top) = \mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^\top - \mathbf{K} \mathbf{K}^\top \\ \mathbf{J}_\mathbf{K} \mathbf{J}_1 &= (\mathbf{I}_n - \mathbf{K} \mathbf{K}^\top)(\mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^\top) = \mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^\top - \mathbf{K} \mathbf{K}^\top\end{aligned}$$

が導かれる.

### B.3

ここで (4.20) 式が成り立つとすると, このとき

$$\mathbf{V}_1^\top \mathbf{J}_\mathbf{K} \mathbf{J}_1 \mathbf{J}_\mathbf{K} \mathbf{V}_1 = \mathbf{W}_1^\top \mathbf{J}_\mathbf{L} \mathbf{W}_1 = \mathbf{I}_m \quad (\text{B.1})$$

である. これは (4.24) 式を用いたときの制約条件 (4.20) 式である.

(4.23) 式におけるSVDは、 $\mathbf{V}_1 = \mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L \mathbf{W}_1 \Phi_1^{-1}$ を意味している。 $(B.1)$ 式の左辺にこれを代入すると

$$\begin{aligned} & \Phi_1^{-1} \mathbf{W}_1^\top \mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_K \mathbf{J}_1 (\mathbf{J}_K \mathbf{J}_1 \mathbf{J}_K) \mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L \mathbf{W}_1 \Phi_1^{-1} \\ &= \Phi_1^{-1} \mathbf{W}_1^\top \mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_1 \mathbf{J}_K (\mathbf{J}_K \mathbf{J}_1 \mathbf{J}_K) \mathbf{J}_K \mathbf{J}_1 \mathbf{Y} \mathbf{J}_L \mathbf{W}_1 \Phi_1^{-1} \\ &= \Phi_1^{-1} \mathbf{W}_1^\top \mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_1 \mathbf{J}_K \mathbf{J}_1 \mathbf{J}_K \mathbf{J}_1 \mathbf{Y} \mathbf{J}_L \mathbf{W}_1 \Phi_1^{-1} \\ &= \Phi_1^{-1} \mathbf{W}_1^\top \mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_1 \mathbf{J}_K \mathbf{J}_1 \mathbf{Y} \mathbf{J}_L \mathbf{W}_1 \Phi_1^{-1} \\ &= \Phi_1^{-1} \mathbf{W}_1^\top \mathbf{W} \begin{bmatrix} \Phi_1^2 \\ \Phi_2^2 \end{bmatrix} \mathbf{W}^\top \mathbf{W}_1 \Phi_1^{-1} = \mathbf{I}_m \end{aligned}$$

これには、付録B.2で証明した $\mathbf{J}_1 \mathbf{J}_K = \mathbf{J}_K \mathbf{J}_1$ および(4.23)式から得られる $\mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_K \mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L = \mathbf{W} \begin{bmatrix} \Phi_1^2 \\ \Phi_2^2 \end{bmatrix} \mathbf{W}^\top$ を用いた。この等式は、 $\mathbf{W}_1 = \mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_K \mathbf{J}_1 \mathbf{V}_1 \Phi_1^{-1}$ であることも意味している。これを(B.1)式における $\mathbf{W}_1^\top \mathbf{J}_L \mathbf{W}_1$ に代入すると

$$\begin{aligned} & (\Phi_1^{-1} \mathbf{V}_1^\top \mathbf{J}_K \mathbf{J}_1 \mathbf{Y} \mathbf{J}_L) \mathbf{J}_L (\mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_K \mathbf{J}_1 \mathbf{V}_1 \Phi_1^{-1}) \\ &= \Phi_1^{-1} \mathbf{V}_1^\top \mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_K \mathbf{J}_1 \mathbf{V}_1 \Phi_1^{-1} \\ &= \Phi_1^{-1} \mathbf{V}_1^\top \mathbf{V} \begin{bmatrix} \Phi_1^2 \\ \Phi_2^2 \end{bmatrix} \mathbf{V}^\top \mathbf{V}_1 \Phi_1^{-1} = \mathbf{I}_m \end{aligned}$$

上記の等式には、 $\mathbf{J}_1 \mathbf{J}_K$ および(4.23)式から得られる $\mathbf{J}_L \mathbf{Y}^\top \mathbf{J}_K \mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L = \mathbf{W} \begin{bmatrix} \Phi_1^2 \\ \Phi_2^2 \end{bmatrix} \mathbf{W}^\top$ を用いた。

## B.4

ここで、(4.25)式における $R = \text{rank}(\mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L)$ について考える。 $\mathbf{Y} = [\mathbf{GC}, \mathbf{U}_d]$ を $\mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L$ へ代入すると

$$\mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L = [\mathbf{J}_1 \mathbf{J}_K \mathbf{GC}, \mathbf{J}_1 \mathbf{J}_K \mathbf{U}_d] \mathbf{J}_L = [\mathbf{J}_1 \mathbf{J}_K \mathbf{GC}, \mathbf{O}_{n \times p}] \mathbf{J}_L \quad (B.2)$$

と書き換えることができる。これは(4.8)式が示唆する $\mathbf{U}_d = n^{1/2} \mathbf{KL}^\top [\mathbf{O}_{n \times p}, \mathbf{I}_p]^\top$ とそれに従う $\mathbf{J}_K \mathbf{U}_d = (\mathbf{I}_n - \mathbf{KK}^\top \mathbf{U}_d) = \mathbf{O}_{n \times p}$ より得られる。等式(B.2)式は、 $R$ が $\text{rank}(\mathbf{J}_K \mathbf{J}_1 \mathbf{GC})$

と  $\text{rank}(\mathbf{J}_L)$  の最小値であることを示している。ここで,  $\mathbf{J}_L = \mathbf{I}_{m+p} - \mathbf{LL}^\top$  は冪等行列であり (4.6) 式から  $\text{rank}(\mathbf{LL}^\top) = p$  である (Lütkepohl, 1996) ゆえ,

$$\text{rank}(\mathbf{J}_L) = (m+p) - p = m \quad (\text{B.3})$$

である。また,  $n-p \leq r-1$  だと仮定すると

$$\text{rank}(\mathbf{J}_K \mathbf{J}_1 \mathbf{G} \mathbf{C}) = \min(n-p, r-1, m) = \min(r-1, m) \quad (\text{B.4})$$

である。これは,  $\mathbf{J}_K = \mathbf{I}_n - \mathbf{KK}^\top$  は冪等行列であること, (4.6) 式より  $\text{rank}(\mathbf{J}_K) = n-p$  であること,  $\mathbf{J}_1 \mathbf{G} \mathbf{1}_r = \mathbf{J}_1 \mathbf{1}_n = \mathbf{0}_n$  から  $\text{rank}(\mathbf{J}_1 \mathbf{G}) \leq \min(n, r-1)$  であること,  $\text{rank}(\mathbf{C}) \leq \min(r, m)$  を考えると得られる。 (B.3) 式と (B.4) 式を比べると,  $R = \text{rank}(\mathbf{J}_1 \mathbf{J}_K \mathbf{Y} \mathbf{J}_L) \leq \min(r-1, m)$  である。

## 参考文献

- 足立浩平・村上隆 (2011). 非計量多変量解析法. 朝倉書店.
- Adachi, K. (2012). Some contributions to data-fitting factor analysis with empirical comparisons to covariance-fitting factor analysis. *Journal of the Japanese Society of Computational Statistics*, **25**, 25–38.
- 足立浩平 (2014). 因子分析への行列集約アプローチ 日本統計学会, **44**, 363–382.
- Adachi, K. (2015). A new algorithm for generalized least squares factor analysis with a majorization technique. *Open Journal of Statistics*, **5**, 165–172.
- Adachi, K. (2016). *Matrix-Based Introduction to Multivariate Data Analysis*. Singapore: Springer.
- Adachi, K. & Trendafilov, N.T. (2018). Some mathematical properties of the matrix decomposition solution in factor analysis. *Psychometrika*, **83**, 407–424.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**, 716–723.
- 新井仁之 (2006). 線形代数–基礎と応用. 日本評論社.
- Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proc. Third Berkeley Symp. Math. Statist. Prob.*, vol. 5, pp. 111–150. Berkeley, CA: University of California Press.
- Anderberg, M. (1973). *Cluster Analysis for Applications*. New York: Academic Press.
- Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, **28**, 97–104.
- Bernaards, C. A., & Jennrich, R. I. (2003). Orthomax rotation and perfect simple structure. *Psychometrika*, **68**, 585–588.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Boyd, S., & Vandenberghe, L. (2018). *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge University Press.

- Browne, M. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Researches*, **36**, 111–150.
- Bühlmann, P., & van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.
- Calinski, R. B., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, **3**, 1–27.
- Choi, J., Zou, H., & Oehlert, G. (2011). A penalized maximum likelihood approach to sparse factor analysis. *Statistics and Its Interface*, **3**, 429–436.
- Danaher, P., Wang, P., & Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B*, **76**, 373–397.
- de Leeuw, J. (2004). Least squares optimal scaling of partially observed linear systems. In: van Montfort, K., Oud, J., Satorra, A. (eds.) *Recent Developments of Structural Equation Models: Theory and Applications*, pp. 121–134. Kluwer Academic, Dordrecht.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: Series B*, 1–38.
- De Soete, G., & Carroll, J. D. (1994). K-means clustering in a low-dimensional Euclidean space. In E. Diday, Y. Lechevallier, M. Schader., P. Bertrand & B. Burtschy (Eds), *New Approaches in Classification and Data Analysis*. pp. 212–219. Berlin: Springer-Verlag.
- Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211–218.
- Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, **96**, 1348–1360.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179–188.
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9**, 432–441.
- Gan, G., Ma, C., & Wu, J. (2007). *Data Clustering: Theory, algorithms, and Applications*. Philadelphia, PA, Society of Industrial and Applied Mathematics (SIAM).
- Gentle, J. E. (2017). *Matrix Algebra: Theory, Computations and Applications in Statistics*. Springer.

- Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common factor theory. *British Journal of Statistical Psychology*, **8**, 65–81.
- Harman, H. H. & Jones, W. H. (1966). Factor analysis by minimizing residuals (Minres). *Psychometrika*, **31**, 351–369.
- Harman, H. H. (1976). *Modern Factor Analysis*, 3rd Edition. Chicago: The University of Chicago Press.
- Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*. Springer Science & Business Media.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.
- Hirose, K. & Yamamoto, M. (2014). Estimation of an oblique structure via penalized likelihood factor analysis. *Computational Statistics and Data Analysis*, **79**, 120–132.
- Hirose, K. & Yamamoto, M. (2015). Sparse estimation via nonconcave penalized likelihood in a factor analysis model. *Statistics and Computing*, **25**, 863–875.
- 市川雅教 (2010). 因子分析. 朝倉書店
- 岩崎学・吉田清隆 (2006). 統計的データ解析入門 線形代数. 東京図書.
- Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. New York: Springer.
- Jain, P., & Kar, P. (2017). Non-convex optimization for machine learning. *Foundations and Trends® in Machine Learning*, **10**, 142–336.
- Jennrich, R. I. (2004). Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika*, **69**, 257–273.
- Jolliffe, I. T. (2006). *Principal Component Analysis*. Springer Science & Business Media.
- 川野秀一・廣瀬慧・立石正平・小西貞則 (2010). 回帰モデリングと  $L_1$  型正則化法の最近の展開. 日本統計学会誌, **39**, 211–242.
- 川野秀一・松井秀俊・廣瀬慧 (2018). スパース推定法による統計モデリング. 共立出版.
- Krijnen, W. P., Wansbeek, T. J., & ten Berge, J. M. F. (1996). Best linear predictors for factor scores. *Communications in Statistics: Theory and Methods*, **25**, 3013–3025.
- Kristof, W. (1970). A theorem on the trace of certain matrix products and some applications. *Journal of Mathematical Psychology*, **7**, 515–530.
- Lawley, D. N. (1942). XIV.—Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, **61**, 176–185.

- Lütkepohl, H. (1996). *Handbook of matrices*. Wiley.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium*, vol. 1, 281–297.
- Magnus, J. R. & Neudecker, H. (2007). *Matrix differential calculus with applications in statistics and econometrics*, 3rd Edition. John Wiley & Sons.
- McDonald, R. P. (1979). The simultaneous estimation of factor loadings and scores. *Brit. J. Math. Statis. Psychol.*, **32**, 212–228.
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, **50**, 159–179.
- 森裕一・黒田正博・足立浩平 (2017). 最小二乗法・交互最小二乗法. 共立出版.
- Mulaik, S. A. (1976). Comments on “The measurement of factorial indeterminacy”. *Psychometrika*, **41**, 249–262.
- Mulaik, S. A. (2010). *Foundations of Factor Analysis*, 2nd Edition. Boca Raton: CRC Press.
- 丘本正 (1986). 因子分析の基礎. 日科技連.
- Rubin, D. B., & Thayer, D. T. (1982). EM algorithms for ML factor analysis. *Psychometrika*, **47**, 69–76.
- Schott, J. R. (2016). *Matrix Analysis for Statistics*. John Wiley & Sons.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, **6**, 461–464.
- Sočan, G. (2003). *The Incremental Value of Minimum Rank Factor Analysis*. Ph.D Thesis, University of Groningen: Groningen.
- 心理学実験指導研究会編 (1985). 『実験とテスト=心理学の基礎<実習編>』. 培風館.
- Spearman, C. (1904). “General Intelligence,” objectively determined and measured. *The American Journal of Psychology*, **15**, 201–292.
- Schönemann, P. H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika*, **31**, 1–10.
- Stegeman, A. (2016). A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts. *Computational Statistics and Data Analysis*, **99**, 189–203.
- Sun, W., Wang, J., & Fang, Y. (2012). Regularized k-means clustering of high-dimensional data and its asymptotic consistency. *Electronic Journal of Statistics*, **6**, 148–167.

- ten Berge, J. M. F. (1983). A generalization of Kristof's theorem on the trace of certain matrix products. *Psychometrika*, **48**, 519–523.
- ten Berge, J. M. F. (1993). *Least Squares Optimization in Multivariate Analysis*. Leiden, The Netherlands: DSWO Press.
- Thurstone, L.L. (1935). *Vector of Mind*. Chicago: The University of Chicago Press.
- Thurstone, L.L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267–288.
- Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B*, **61**, 611–622.
- Timmerman, M. E., Ceulemans, E., Kiers, H. A. L., and Vichi, M. (2010). Factorial and reduced K-means reconsidered. *Computational Statistics and Data Analysis*, **54**, 1858–1871.
- Trendafilov, N.T. & Unkel, S. (2011). Exploratory factor analysis of data matrices with more variables than observations. *Journal of Computational and Graphical Statistics*, **20**, 874–891.
- Unkel, S., & Trendafilov, N. T. (2010). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review*, **78**, 363–382.
- Uno, K., Satomura, H., & Adachi, K. (2016). Fixed factor analysis with clustered factor score constraint *Computational Statistics and Data Analysis*, **94**, 265–274.
- 宇野光平 (2016). クラスタリングを伴う多変量データ解析法の新展開. 大阪大学大学院人間科学研究科修士論文.
- 宇野光平 (2017). クラスタリングを伴うテンソルデータに対するガウシアングラフィカルモデル. 2017年度統計関連学会連合大会講演報告集. p.230
- Uno, K. Huang, P-H. (2018). Penalized item response models for polytomous data. 日本計算機統計学会 第32回大会 講演論文集.
- 宇野光平 (2018). Group LassoによるFuzzy c-means クラスタリングの変数選択. 2018年度統計関連学会連合大会講演報告集. p.297
- Uno, K., Adachi, K., & Trendafilov, N. T. (2018). Clustered Common Factor Identification in Factor Analysis. Submitted.
- van Der Maaten, L., Postma, E., & van den Herik, J. (2009). Dimensionality reduction: a comparative. *Journal of Machine Learning Research*, **10**, 66–71.
- von Neumann, J. (1937). Some matrix-inequalities and metrization of matrix-space, *Tomsk. Univ. Rev.*, **1**, 286–300.

- Vichi, M. and Kiers, H. A. L. (2001). Factorial K-means analysis for two-way data. *Computational Statistics and Data Analysis*, **37**, 49–64.
- Vichi, M. & Saporta, G. (2009). Clustering and disjoint principal component analysis. *Computational Statistics and Data Analysis*, **53**, 3194–3208.
- Vichi, M. (2017). Disjoint factor analysis with cross-loadings. *Advances in Data Analysis and Classification*, **11**, 563–591.
- Young, G. (1941). Maximum likelihood estimation and factor analysis. *Psychometrika*, **6**, 49–53.
- Zhang, C. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*. **38**, 894-942.
- Zhang, J., Li, J., & Liu, C. (2014). Robust factor analysis using the multivariate t-distribution. *Statistica Sinica*, **24**, 291–312.

## 謝 辞

この博士論文執筆にあたって、まず指導教員である足立浩平先生に深く感謝申し上げます。研究に限らない日頃からのご助言、共同研究、さらに学会参加における渡航費宿泊費のご負担と、多岐に渡って支えていただいた足立先生のおかげでこの博士論文を書き上げることができました。

また貴重なご助言をいただき共同研究者になっていただいた里村裕紀様、Dr. Nickolay T. Trendafilov, Dr. Po-Hsien Huang に感謝の意を表します。さらに研究室の皆様との日頃の議論により、研究をより深化させることができたことに感謝いたします。学位論文の副査を承諾していただいた狩野裕先生、青野正二先生、中井宏先生にはお忙しい中、お時間をつくっていただいたことに感謝申し上げます。

本稿の内容は人間科学研究科の海外派遣支援制度の援助を受けて国際学会で発表させていただいた研究が含まれます。貴重な経験をさせていただいた人間科学研究科の支援に感謝いたします。さらにイギリスでの Dr. Nickolay T. Trendafilov との共同研究、および台湾での Dr. Po-Hsien Huang との共同研究は大阪大学数理・データ科学教育研究センター (MMDS) の全面的支援によって行うことができました。MMDS の支援に感謝いたします。

最後に、金銭面を含めて学生生活を支えてくれた家族への感謝で謝辞を締めたいと思います。

12月14日  
行動生態学講座 行動統計科学研究分野