多変量解析
第 7 回: 因子分析 (1) 基礎概念

藤本 衡

2018 年 5 月 30 日

本日の内容

- 潜在的因子
- 因子分析のモデル
- 主成分分析との対比
- モデルの行列表現
- 直交解

潜在的因子 (latent factor)

- 観測可能な変量は、(潜在的な) 要因の合成ではないか？
- 観測可能な多くの変量を、より少ない潜在要因で表せれば便利

例: 科目の成績

英語の得点 (観測可能) = コミュ力 + 聞き取り能力 + 記憶力 + $\cdots$

数学の得点 (観測可能) = 論理的思考力 + 計算能力 + 読解力 + $\cdots$

国語の得点 (観測可能) = 論理的思考力 + 読解力 + 記憶力 + $\cdots$

因子分析のモデル

- 変量 X_k は潜在的因子の線型結合で表現できると仮定
- 変量はスケールが異なっている可能性がある
  → X_k を標準化する $(Z_k = (X_k - \mathbb{E}[X_k])/\sqrt{\mathbb{V}[X_k]})$

モデル

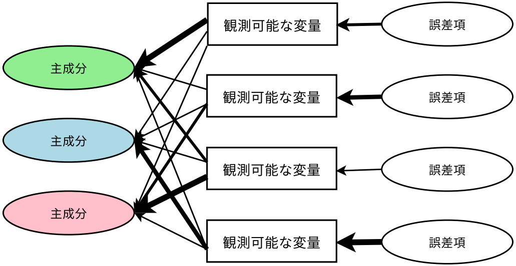
$$Z_k = a_{k1}F_1 + a_{k2}F_2 + \cdots + a_{kN}F_N + d_kU_k \quad (k = 1, 2, \ldots, m) \tag{1}$$

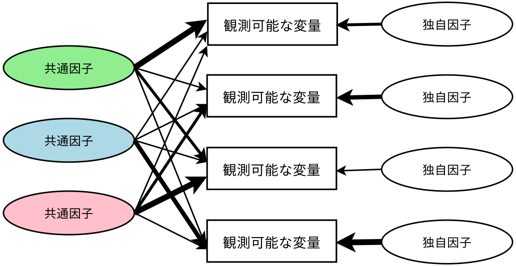
F_j: 共通因子（全変量に影響する因子）$(j = 1, 2, \ldots, N)$
U_k 独自因子（X_k のみに影響する因子） $(k = 1, 2, \ldots, m)$
a_{kj} 因子負荷量（変量 X_k に因子 f_j が与える影響）

因子分析の応用分野

- 従来は心理学分野で多用されていた
- 工学的応用は少ない？（教科書でも省略されたり）
- QoS(Quality of Service)/QoE(Quality of Experience) の重要性
    - 工業製品も客観的性能から客の主観評価の時代へ
    - Web サービスにおけるユーザ体験
    - 「カワイイ」とは？「萌え」とは？「ヤバイ」とは？

主成分分析との対比





- **観測可能な変量を要約することが目的**

- **観測可能な変量の背後にある要因を知ることが目的**

行列表現

- 個体 i に関する Z_k, F_j, U_k の実現値を z_{ik}, f_{ij}, u_{ik} とおく
- f_{ij} を<u>因子得点</u>(factor score) と呼ぶ

$$z_{ik} = a_{k1} f_{i1} + a_{k2} f_{i2} + \cdots + a_{kN} f_{iN} + d_k u_{ik} \quad (i = 1, 2, \ldots, n)$$

行列表現すると

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{A}^{\mathsf{T}} + \boxed{\text{埋めて!}} \tag{2}$$

ただし

$$\boldsymbol{Z} = (z_{ik}) \ (n \times m \text{ 行列}),$$
$$\boldsymbol{F} = (f_{ij}) \ (n \times N \text{ 行列}), \quad \boldsymbol{A} = (a_{kj}) \ (m \times N \text{ 行列}),$$
$$\boldsymbol{U} = (u_{ik}) \ (n \times m \text{ 行列}), \quad \boldsymbol{D} = \mathrm{diag}(d_k) \ (m \times m \text{ 行列})$$

直交解 (orthogonal solution)

仮定 1: $F_j,\ U_k$ **は平均** 0:

$$\mathbb{E}\left[F_j\right] = 0\ (j = 1, 2, \ldots, N),\quad \mathbb{E}\left[U_k\right] = 0\ (k = 1, 2, \ldots, m)$$

仮定 2: $F_j,\ U_k$ **は互いに無相関**:

$$\mathrm{Cov}\left[F_j,\ F_\ell\right] = 0\ (j \neq \ell),\quad \mathrm{Cov}\left[U_k,\ U_\ell\right] = 0\ (k \neq \ell),$$

$$\mathrm{Cov}\left[F_j,\ U_k\right] = 0\ (\forall j, k)$$

仮定 2 より以下が成り立つ。この仮定を満たす $\boldsymbol{F}$ を直交解という。

$$\frac{1}{n}\boldsymbol{F}^\intercal\boldsymbol{F} = \boldsymbol{I}_m,\quad \frac{1}{n}\boldsymbol{U}^\intercal\boldsymbol{U} = \boldsymbol{I}_m,\quad \boldsymbol{F}^\intercal\boldsymbol{U} = \boldsymbol{O}_{N\times m}$$

($\boldsymbol{I}_m$: m **次元単位行列**, $\boldsymbol{O}_{N\times m}$: $N \times m$ **零行列**)

Z による相関係数行列の導出

定義より $\dfrac{1}{n}Z^{\intercal}Z$ の (k,h) 要素は

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} z_{ik}z_{ih} &= \frac{1}{n}\sum_{i=1}^{n}\frac{x_{ik}-\mathbb{E}\left[X_k\right]}{\sqrt{\mathbb{V}\left[X_k\right]}}\cdot\frac{x_{ih}-\mathbb{E}\left[X_h\right]}{\sqrt{\mathbb{V}\left[X_h\right]}} \\
&= \frac{\mathrm{Cov}\left[X_k,\, X_h\right]}{\sqrt{\mathbb{V}\left[X_k\right]\mathbb{V}\left[X_h\right]}} = r_{kh} \quad (X_k \text{ と } X_h \text{ の相関係数})
\end{aligned}
$$

したがって $\dfrac{1}{n}Z^{\intercal}Z = \boldsymbol{R}$ (X_k の相関係数行列) である。

直交解における基本定理

(2) 式および仮定 2 より

$$\boldsymbol{R} = \frac{1}{n}\boldsymbol{Z}^\intercal \boldsymbol{Z} = \frac{1}{n}(\boldsymbol{F}\boldsymbol{A}^\intercal + \boldsymbol{U}\boldsymbol{D})^\intercal(\boldsymbol{F}\boldsymbol{A}^\intercal + \boldsymbol{U}\boldsymbol{D}) = \boldsymbol{A}\boldsymbol{A}^\intercal + \boldsymbol{D}^2$$

$$\boldsymbol{R}^* \equiv \boldsymbol{R} - \boldsymbol{D}^2 = \boldsymbol{A}^\intercal \boldsymbol{A} \tag{3}$$

(3) 式より、以下の関係が導かれる。

$$h_k^2 \equiv 1 - d_k^2 = \sum_{j=1}^{N} a_{jj}^2$$

$$r_{kh} = \sum_{j=1}^{N} a_{kj} a_{hj}$$

h_k^2 を共通性 (communality) と呼ぶ。

2 変量, 2 共通因子の場合

モデル: $Z_1 = a_{11}F_1 + a_{12}F_2 + d_1U_1, \quad Z_2 = a_{21}F_1 + a_{22}F_2 + d_2U_2$

$$\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}$$

3 つの式を求めて!

未知変数 6 個で式が 3 本なので、これだけでは解は求まらない！
→さらに仮定を加え、推定を行う必要