項目反応理論

2022 年 11 月 16 日

3　尺度値の推定

3.4　ベイズ推定法

3.4.3　EAP 推定法

ベイズ推定法は MAP 推定法ばかりではなく EAP 法も利用される。これは、θ_i の事後分布の θ_i に関する期待値を推定値とする推定法である。式にすると、

$$\begin{aligned}\theta_i &= \int_{-\infty}^{+\infty} \theta_i \frac{g(\theta_i)L(\boldsymbol{u}_i|\theta)}{f(\boldsymbol{u}_i)} d\theta_i \\ &= \frac{\displaystyle\int_{-\infty}^{+\infty} \theta_i g(\theta_i)L(\boldsymbol{u}_i|\theta)d\theta_i}{f(\boldsymbol{u}_i)} \\ &= \frac{\displaystyle\int_{-\infty}^{+\infty} \theta_i g(\theta_i)L(\boldsymbol{u}_i|\theta)d\theta_i}{\displaystyle\int_{-\infty}^{+\infty} g(\theta_i)L(\boldsymbol{u}_i|\theta)d\theta_i} \end{aligned} \tag{3.35}$$

となる。式中の積分は区分求積法によって求めることができる。

ある区間 $[a, b]$ を N 等分した点 (幅は$\Delta\theta$) を $\theta_n(n = 1, 2, \cdots, N, N + 1)$ とすると

$$\int_b^a f(\theta)d\theta \simeq \sum_{n=1}^{N} f(\theta) \times \Delta\theta \tag{3.36}$$

によって N 個の長方形の和によって近似することができる。式 (3.35) 中の区間は $[-\infty, \infty]$ であるが、$g(\theta)$ は標準正規分布を仮定しているので、積分区間は $[-3, 3]$ や $[-4, 4]$ などで十分であると考えられる。式にすると、

$$\hat{\theta}_i \simeq \frac{\displaystyle\sum_{n=1}^{N} \theta_{in} L(\boldsymbol{u}_i|\theta_{in})g(\theta_{in})\Delta\theta}{\displaystyle\sum_{n=1}^{N} L(\boldsymbol{u}_i|\theta_{in})g(\theta_{in})\Delta\theta} \tag{3.37}$$

と近似できる。これまでの方法と違って、直接求めることができるので、反復計算の必要のない EAP 推定法は計算時間を短縮することができる。

4　項目母数の推定

第3章では項目母数は既知であると仮定したうえで進めてきた。今回からは、反応パタン $\boldsymbol{U}$ から項目母数を推定する方法について考える。

4.1　同時最尤推定法

本節では $\boldsymbol{U}$ から項目母数と被験者母数を同時に推定する方法を紹介する。第3章で導入したように、被験者母数 θ_i が与えられたとき、u_{ij} の分布は、

$$f(u_{ij}|\theta_i) = p_j(\theta_i)^{u_{ij}} q_j(\theta_i)^{1-u_{ij}} \tag{4.1}$$

であった。ただし、これまでと違って今回は項目母数も未知であるので、

$$f(u_{ij}|\theta_i, a_j, b_j, c_j) = p_j(\theta_i)^{u_{ij}} q_j(\theta_i)^{1-u_{ij}} \tag{4.2}$$

と書き直す。ここでは3母数モデルを用いて説明していく。局所独立の仮定から、n 個の項目に関する被験者 i の反応パタンベクトル $\boldsymbol{u}_i$ が観察される確率は、

$$f(\boldsymbol{u}_i|\theta_i, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \prod_{j=1}^{n} f(u_{ij}|\theta_i, a_j, b_j, c_j) \tag{4.3}$$

と表現される。ここで、

$$\boldsymbol{a} = (a_1, a_2, \cdots, a_n)' \tag{4.4}$$

$$\boldsymbol{b} = (b_1, b_2, \cdots, b_n)' \tag{4.5}$$

$$\boldsymbol{c} = (c_1, c_2, \cdots, c_n)' \tag{4.6}$$

である。被験者の反応は互いに独立であるから、被験者母数と項目母数が与えられた条件の下で、n 個の項目に対する N 人の被験者の反応パタン行列 $\boldsymbol{U}$ が観察される確率は、

$$f(\boldsymbol{U}|\theta_i, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \prod_{i=1}^{N} f(\boldsymbol{u}_i, |\theta_i, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \prod_{i=1}^{N}\prod_{j=1}^{n} f(u_{ij}|\theta_i, a_j, b_j, c_j) \tag{4.7}$$

と表現される。ここで、

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_i, \cdots, \theta_N)' \tag{4.8}$$

である。式 (4.7) は $\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ は定数であり、$\boldsymbol{U}$ が変数である。ところが、現実を考えると手元にあるのは被験者の反応パタン $\boldsymbol{U}$ だけであり、その他の項目母数は未知のままである。したがって、同じ式を定数と変数を逆にみなし、

$$L(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = f(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) \tag{4.9}$$

のように尤度とする。この尤度を最大にする母数ベクトルの値が最尤推定値である。またこのような方法を最尤推定法と言った。ここで最大値を求めるために、対数変換を行うと

$$logL(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \sum_{i=1}^{N}\sum_{j=1}^{n}[u_{ij}logp_j\theta_i + (1 - u_{ij}logq_j(\theta_i))] \tag{4.10}$$

となる。この関数の最大化を考える。尤度関数を最大化する母数の値と、対数尤度関数を最大化する母数の値は等しい。

対数尤度関数を最大化する解は、前章であったように変数で微分して 0 とおいた方程式を変数に関して解くことによって求まった。今回は変数が複数あるので、各変数 θ_i, a_j, b_j, c_j で偏微分した結果を 0 とおき、

$$\frac{\partial}{\partial\theta_i}\log L(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = D\sum_{j=1}^{n}\frac{a_j(p_j(\theta_i) - c_j)(u_ij - p_j(\theta_i))}{(1 - c_j)p_j(\theta_i)} = 0 \tag{4.11}$$

$$\frac{\partial}{\partial a_j}\log L(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \frac{D}{1 - c_j}\sum_{i=1}^{N}\frac{(\theta_i - b_j)(p_j(\theta_i) - c_j)(u_ij - p_j(\theta_i))}{p_j(\theta_i)} = 0 \tag{4.12}$$

$$\frac{\partial}{\partial a_j}\log L(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \frac{-Da_j}{1 - c_j}\sum_{i=1}^{N}\frac{(p_j(\theta_i) - c_j)(u_ij - p_j(\theta_i))}{p_j(\theta_i)} = 0 \tag{4.13}$$

$$\frac{\partial}{\partial a_j}\log L(\boldsymbol{U}|\boldsymbol{\theta}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \frac{1}{1 - c_j}\sum_{i=1}^{N}\frac{(u_ij - p_j(\theta_i))}{p_j(\theta_i)} = 0 \tag{4.13}$$

の連立方程式を解いた値が最尤推定値となる。この方法は被験者母数と項目母数を同時に推定するという意味で特に同時最尤推定法と呼ぶ。