順位法の為の項目反応モデルの提案

○川端一光¹

(¹ 独立行政法人国際交流基金日本語試験センター)

キーワード: Mallows-Bradley-Terry モデル, 順位法, マルコフ連鎖モンテカルロ法

An Item response model for method of rank order

Ikko KAWAHASHI¹

(¹ Center for Japanese-language Testing, The Japan Foundation)

Key words: Mallows-Bradley-Terry model, Method of rank order, Markov Chain Monte Carlo method

目 的

K 個の試料の選好度を順位法で評定させた場合に、Mallows-Bradley-Terry モデル (Mallows, 1957) を利用することで各試料の平均的選好度を推定することが可能である。このモデルは 1 セット K 個の試料に対する評定者のランキング判断をモデル化したものである。本研究では各評定者が n 個の評定課題を全て回答する状況へ Mallows-Bradley-Terry モデルを拡張する。また、モデル中に、平均的選好度の序列に対する各評定者の弁別力を潜在特性値 (latent trait) として定義する。テスト作題者の項目難易度に関する弁別力測定のように、ある基準に対する評定者の弁別力が試されるような状況での適用が想定されている。

モデルと母数推定法

K 個の試料に対して順位法で評価させた場合に、特定のランキング $l (= 1, 2, \dots, K!)$ を表現するベクトル関数を $\omega_l = [\omega_l(1), \omega_l(2), \dots, \omega_l(k), \dots, \omega_l(K)]$ と表現する。 $\omega_l(k)$ は試料 k に順位を付与する関数である。評価者 $i (= 1, 2, \dots, N)$ が順位評定課題 (項目) $j (= 1, 2, \dots, n)$ で特定のランキング l を選択する確率を次式で表現する。

$$P_j(\omega_l|\theta_i, \alpha_j) = \frac{\exp\left[\sum_{k=1}^K (K - \omega_l(k))\alpha_{jk}\theta_i\right]}{\sum_{m=1}^{K!} \exp\left[\sum_{k=1}^K (K - \omega_m(k))\alpha_{jk}\theta_i\right]} \quad (1)$$

ここで α_{jk} は項目 j の試料 k の選好度母数である。モデルの識別の為に、 $\sum_{k=1}^K \alpha_{jk} = 0$ の制約が必要となる。 θ_i は集団で決定される選好度の序列に関する評価者 i の弁別力を表現する母数であり、全項目においてチャンスレベル ($1/K!$) でしか序列を弁別できない場合には $\theta_i = 0$ 、平均的に弁別できる場合には $\theta_i = 1$ 、平均以上に弁別できる場合には $\theta_i > 1$ となる。また選好度母数の序列と逆のランキング判断を行う場合には $\theta_i < 0$ となる。

母数の推定にはマルコフ連鎖モンテカルロ法 (MCMC 法) を援用したベイズ推定法を利用する。MCMC 標本のサンプリングには Metropolis-Hasting 法を採用する。 θ と α の事前分布は $\theta_i \sim N(\mu_\theta = 1, \sigma_\theta = 1)$, $\alpha_{jk} \sim N(\mu_\alpha = 0, \sigma_\alpha = 1)$ とする。 $P(\theta_j)$ と $P(\alpha_j)$ をそれぞれ事前分布とすると、項目母数と弁別力母数の条件付き事後分布は尤度と事前分布の積に比

例する。

$$P(\alpha_j|\theta, \Omega) \propto P(\theta, \alpha_j|\Omega)P(\alpha_j) \quad (2)$$

$$P(\theta_i|A, \omega_l) \propto P(\theta_i, A|\omega_l)P(\theta_i) \quad (3)$$

ここで Ω は N 人の評定者の n 項目へのランキング評価が含まれた $N \times (n \times K)$ 行列を、 A は $K \times n$ 項目母数行列をそれぞれ表現している。M-H 法におけるアルゴリズムの更新は、次の 2 段階で行われる。a). 各項目毎に 1 セットの α_j の候補値を乱数生成し、そのセットの下で (2) 式を利用して、受容-棄却判断を 1 回のみ行う。b). 次に更新した A を (3) 式に代入して、受験者毎に θ の更新を行う。a) と b) の過程を繰り返すことで母数の事後分布を得る。

シミュレーション研究

α_j の真値を標準正規乱数、 θ_i の真値を平均値 1・分散 1 の正規乱数に設定し、 $K = 4$, $N = 100$ ・ $N = 500$, $n = 10$ ・ $n = 20$ という現実的な設定のもと、シミュレーションデータを各 20 セットずつ生成した。このデータを利用して、母数推定アルゴリズムの真値復元力を検討した。アルゴリズムの反復は 60000 回とし、最初の 30000 回を burn-in 期間とした。

結果と考察

全母数についてアルゴリズムが事後分布に収束していることを各種収束判定基準に基づき確認した。Table 1 に項目をプールした時の、推定値と真値との平均二乗誤差の平方根 (RMSE) のデータセット分の平均値を示す。 $N = 100$ の場合 n 条件に関わらず項目母数の復元力がやや低いことが明らかとなった。また $N = 500$ ならば n 条件に関わらず、項目母数について高い精度の推定値が得られることが明らかとなった。 θ については推定精度を高める為に、 $n = 20$ よりもさらに項目数を増やす必要があることが明らかとなった (より詳細なシミュレーション結果は発表日に示す)。

Table 1: 母数真値復元状況 (RMSE の平均)

N	n	α_1	α_2	α_3	α_4	θ
100	10	0.143	0.128	0.121	0.130	0.380
	20	0.126	0.127	0.126	0.118	0.299
500	10	0.071	0.069	0.073	0.064	0.383
	20	0.064	0.062	0.062	0.061	0.297

引用文献

Mallows, C.L. (1957). Non-null ranking models. I. *Biometrika*, **44**, 114-130.