

一対比較型心理測定における因子構造の探索的推定

神戸大学 分寺杏介

株式会社リクルートマネジメントソリューションズ 杉山剛

東京大学 岡田謙介

1. 問題と目的

リッカート尺度におけるバイアス等の問題への対処法として、1つの項目内で2つの要素を選択肢として同時に提示し、どちらの要素が自分をよく表しているかを評価させる一対比較型 (paired comparison) の心理測定が近年注目を集めている。一対比較型を含む多肢強制選択型の心理測定では、事前にリッカート型の測定に基づいて推定された因子構造を既知としたうえで尺度が構成されているケースが多い (e.g., Tsaousis & Al-Owaidha, 2024; Watrin et al., 2019; 丹他, 2025)。こうした尺度構成法の背後には、提示される項目要素と潜在変数 (因子) の対応関係が、提示方法 (リッカート型・強制選択型) にかかわらず不変である、という暗黙の仮定があると考えられる。しかし、強制選択型心理測定への回答は、その項目要素と同時に提示される別の要素との相対的な比較に基づいて行われるため、比較対象となる要素の有無によって、当該項目の位置づけが変化し、結果として因子構造も変化する可能性は十分に考えられる。だがこの点について、強制選択型の測定データに対するデファクトスタンダードである Thurstonian IRT モデル (Brown & Maydeu-Olivares, 2011, 以下では TIRT モデルと表す) は、因子構造を既知のものとして推定を行う、数理的には検証的因子分析モデルと等価のモデルであることから、そもそも提示方法に対する因子構造の不変性を検討することは難しい。

本研究では、一部の項目要素の因子構造を探索的に推定する方法を提案する。これにより、因子構造の不変性を検討できるだけでなく、既存の尺度に新規の項目要素を追加する際にも、比較型測定の中で自然に因子構造を推定できるようになることが期待される。

2. モデル

一対比較型の心理測定項目 j では、2つの項目要素 (j_1, j_2) が提示される。そして回答者 i は、 C 個の順序性を持つ選択肢 (1: j_1 のほうが非常に強く当てはまる, C : j_2 のほうが非常に強く当てはまる) から1つを選ぶ。このとき、回答者 i が項目 j において選択肢 c ($c = 1, \dots, C$) を選ぶ確率は

$$P(x_{ij} = c) = \begin{cases} 1 - \Phi(\alpha_{j,2} + \mathbf{a}'_j \mathbf{\Lambda}_j \boldsymbol{\eta}_i) & c = 1 \\ \Phi(\alpha_{j,c} + \mathbf{a}'_j \mathbf{\Lambda}_j \boldsymbol{\eta}_i) - \Phi(\alpha_{j,c+1} + \mathbf{a}'_j \mathbf{\Lambda}_j \boldsymbol{\eta}_i) & 1 < c < C \\ \Phi(\alpha_{j,C} + \mathbf{a}'_j \mathbf{\Lambda}_j \boldsymbol{\eta}_i) & c = C \end{cases} \quad (1)$$

と表現できる。ただし $\boldsymbol{\eta}_i$ は回答者 i の M 次元因子得点ベクトルを、 $\boldsymbol{\alpha}_j = (\alpha_{j,1}, \dots, \alpha_{j,C-1})$ は順序制約を持つ項目 j の切片パラメータをそれぞれ表す。また $\Phi(\cdot)$ は標準正規分布の累積確率を表しており、 $\Phi(\alpha_{j,c} + \mathbf{a}'_j \mathbf{\Lambda}_j \boldsymbol{\eta}_i)$ によって「カテゴリ c 以上の選択肢を選ぶ確率」を意味することとなる。また $\mathbf{\Lambda}_j$ は、項目 j に含まれる2つの項目要素の因子負荷行列を表しており、一般的な TIRT

モデルでは各要素が負荷量を持つ因子は既知であることから、例えば ($M = 6$) 因子構造において j_1 が因子 1, j_2 が因子 3 に負荷量を持つ場合には

$$\Lambda_j = \begin{bmatrix} \lambda_{j1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{j2} & 0 & 0 & 0 \end{bmatrix} \quad (2)$$

と表すことができる。TIRT モデルでは、2つの項目要素に対する効用 (utility) の差に基づいて反応確率が決定すると考えている。これを表現するため、 \mathbf{a}'_j は $[1 \quad -1]$ という横ベクトルの形をとる。したがって、一般的な TIRT モデルでは、 $\mathbf{a}'_j \Lambda_j \boldsymbol{\eta}_i = \lambda_{j1} \eta_{i1} - \lambda_{j2} \eta_{i3}$ と要素の積の差に書き下すことも可能である。

本研究で提案するモデルは、Partially CFA (Chen et al., 2021) の考え方をを用いて Λ_j の各要素を自由推定にする。具体的には、例えば(2)式の設定において j_2 の因子負荷を探索的に求める場合、

$$\Lambda_j = \begin{bmatrix} \lambda_{j1} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{j2}^{(1)} & \lambda_{j2}^{(2)} & \lambda_{j2}^{(3)} & \lambda_{j2}^{(4)} & \lambda_{j2}^{(5)} & \lambda_{j2}^{(6)} \end{bmatrix} \quad (3)$$

と表し、すべてのパラメータを自由推定する。以後では λ について、上付き添字の有無によって探索的か検証的かを区別するものとする。

パラメータ推定は stan (Stan Development Team, 2025) を用いたベイズ推定により行う。事前分布はそれぞれ

$$\begin{aligned} |\lambda_{j\cdot}| &\sim \text{normal}_+(1, 3), & \lambda_{j\cdot}^{(\cdot)} &\sim \text{DoubleExponential}(0, 0.5), & \alpha_{jc} &\sim \text{normal}(0, 5), \\ \boldsymbol{\eta} &\sim \text{multi_normal}(0, \boldsymbol{\Sigma}), & \boldsymbol{\Sigma} &\sim \text{lkj_corr}(1), \end{aligned} \quad (4)$$

と設定した。因子負荷のうち、検証的に推定される $\lambda_{j\cdot}$ については、因子の向き (符号) も既知であると考えため、推定時には非負の制約を置いて絶対値の推定を行い、事後的に 1 または -1 をかける。一方探索的に推定される $\lambda_{j\cdot}^{(\cdot)}$ については、推定結果をできるだけ単純構造に近づけるため、Bayesian Lasso (Park & Casella, 2008) の事前分布である二重指数分布 (ラプラス分布) を置く (Chen et al., 2021)。

3. 実データ分析

実データによる分析例を示すため、HEXACO 形容詞版 (Romano et al., 2023) を使用した。この尺度は、6次元の性格特性を各 10 個の形容詞によって測定するリッカート型尺度である。これを項目要素とした一対比較型に再構成するため、Li et al. (2025) と丹他 (2025) に従い、(1) 必ず異なる因子の形容詞を組み合わせる、(2) 因子負荷が異なる形容詞の組み合わせが 20% 以上になる、という条件を満たす中で、社会的望ましさ (予備調査を行い測定した値) ができるだけ近い形容詞がペアになるように、遺伝的アルゴリズムによって 30 ペアを作成した。

本調査では、クラウドソーシングサービス (Prolific) を介して募集した英語話者に、上記 30 ペアをランダムな順序で画面の左右に提示した。選択肢は、左右の端から順に “Much more”, “More”, “Slightly more” を配置し、参加者はこの中から 1 つをクリックして回答した (図1)。

以後の分析では、210 人分の回答データを使用した。はじめに、すべての形容詞について因子構造が既知である (Romano et al., 2023 に示された因子：以後「正解」と表記：にのみ負荷量を持つ) と仮定した通常の TIRT によってパラメータの推定を行った。続いて、60 個の形容詞のうち 1 個のみについて因子構造を探索的に推定するという設定でのパラメータ推定を、対象となる形容詞

Which statement describes you better, and to what extent?

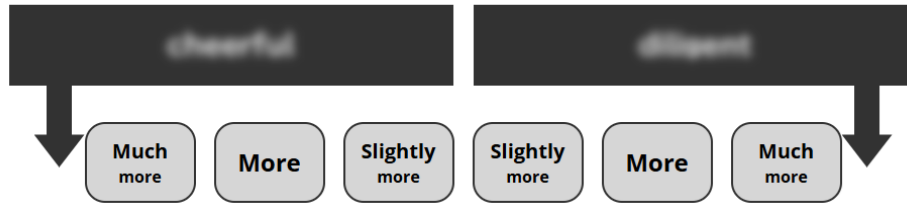


図1 調査画面

を変えて計 60 回実施した。こうして得られた推定結果について、通常の TIRT による推定値との比較を行った。

4. 結果

図2は、通常の TIRT によって得られた λ_j の推定値を X 軸に置き、Y 軸には探索的推定における「正解」の因子の推定値をおいた散布図である。直線 $y = x$ と比べると、事前分布による Shrinkage が発生しているためやや 0 に近づいた推定が行われているものの、相関係数を計算すると $r = .906$ であり、概ね通常の TIRT に近づいた値を推定していることが分かる。

一方図3では Y 軸に、探索的に推定された λ の絶対値全因子での合計に占める「正解」の因子の絶対値の割合を示している。もし「正解」の因子に対して完全な単純構造が見られるならば 1 に近い値であることが期待される値であるが、図3に示されているように最大でも 0.5 程度となっており、かなりの割合で cross-loading が発生していることが見てとれる。

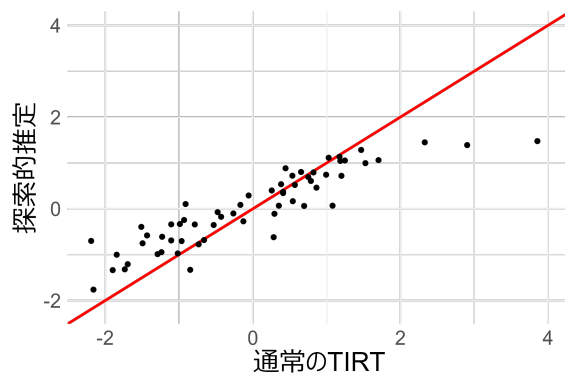


図2 λ の推定値の散布図

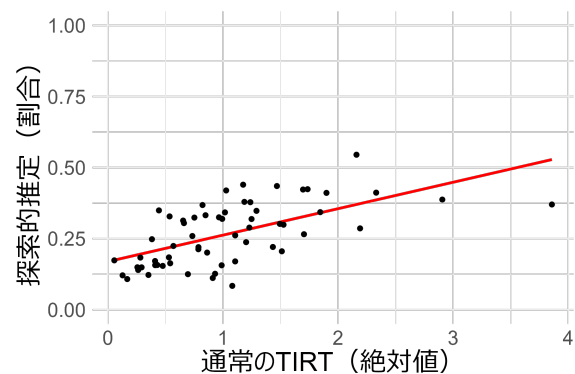


図3 λ の合計に占める割合

5. まとめ

上記に示した結果からは、1 項目であれば、cross-loading を許容した場合であっても、まず「正解」の因子に関しては通常の TIRT（検証的な状況）に近い因子負荷を推定できることが分かる。この点に関しては、複数の項目要素の因子構造を同時に推定する場合についてもその性能を評価する必要があるだろう。

一方で、事前分布による正則化を行ったうえでも大きな cross-loading が発生しており、純粋な単純構造に近い推定値を得ることができなかった。Romano et al. (2023) においても、一部の項目において 0.3 を超える cross-loading が報告されていることから、本研究で使用した尺度の性質として、そもそも純粋な単純構造とは言えない可能性もある。一方で、「正解」の因子で因子負荷の

絶対値が最大値となった形容詞は 33/60 個であり、およそ半分の形容詞では「正解」以外の因子において最も負荷量が大きく推定された。これらの結果がデータおよび尺度の性質によるものなのか、モデルのアーティファクトなのかについては、今後のシミュレーションによって明らかにする必要がある。仮にリッカート型測定では十分な単純構造が見られた尺度であっても、一対比較または強制選択型にした場合には cross-loading が増大するとしたら、TIRT モデルの仮定自体についてもさらなる検討が必要となるかもしれない。

参考文献

- Brown, A. & Maydeu-Olivares, A. (2011). Item Response Modeling of Forced-Choice Questionnaires. *Educational and Psychological Measurement*, 71 (3), 460–502. <https://doi.org/10.1177/0013164410375112>
- Chen, J., Guo, Z., Zhang, L., & Pan, J. (2021). A Partially Confirmatory Approach to Scale Development with the Bayesian Lasso. *Psychological Methods*, 26 (2), 210–235. <https://doi.org/10.1037/met0000293>
- Li, M., Zhang, B., Li, L., Sun, T., & Brown, A. (2025). Mixed-Keying or Desirability-Matching in the Construction of Forced-Choice Measures? An Empirical Investigation and Practical Recommendations. *Organizational Research Methods*, 28 (2), 296–329. <https://doi.org/10.1177/10944281241229784>
- Park, T. & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103 (482), 681–686. <https://doi.org/10.1198/016214508000000337>
- Romano, D., Costantini, G., Richetin, J., & Perugini, M. (2023). The HEXACO Adjective Scales and Its Psychometric Properties. *Assessment*, 30 (8), 2510–2532. <https://doi.org/10.1177/10731911231153833>
- Stan Development Team (2025). *Stan Reference Manual*, 2.37.0. <https://mc-stan.org>.
- Tsaousis, I. & Al-Owaidha, A. (2024). Development of a Forced-Choice Personality Inventory via Thurstonian Item Response Theory (TIRT). *Behavioral Sciences*, 14 (12), 1118. <https://doi.org/10.3390/bs14121118>
- Watrin, L., Geiger, M., Spengler, M., & Wilhelm, O. (2019). Forced-Choice versus Likert Responses on an Occupational Big Five Questionnaire. *Journal of Individual Differences*, 40 (3), 134–148. <https://doi.org/10.1027/1614-0001/a000285>
- 丹 亮人・分寺 杏介・野村 圭史・岡田 謙介 (2025). 社会的望ましきバイアスに頑健な比較型尺度の構築法——総説とチュートリアル PsyArXiv, (preprint). https://doi.org/10.31234/osf.io/7ugez_v1