R による項目反応理論

2000MM090 棚瀬　貴紀
指導教員　松田　眞一

1　はじめに

学力を把握するということは、教育関係者にとって重要な関心事である。この測定方法としてはテストを行うのが一般的あるが、その結果を分析する手段として項目反応理論がある。また、この理論を使うことで、より目的に沿ったテストを作成することも可能である。本研究では項目反応理論の概要を理解し、その分析を統計解析ソフト R にてプログラム化することを目的とする。参考として主に芝 [1]、豊田 [2] を用いた。

2　項目反応モデル

テストは普通、多くの項目から構成される。項目反応理論では、その項目の正答確率を、測定したい特性値を変数とした関数で表す。これを項目特性関数と呼び、これには単調増加関数を与える。項目反応理論では、項目特性関数にいくつかのパラメタを組み込んだ上でモデルを仮定している。このモデルを項目反応モデルと呼ぶ。ここでは項目反応モデルの一つである、ロジスティックモデルを紹介する。このモデルは、特性値 θ の分布にロジスティック分布の分布関数

$$P(\theta) = \frac{1}{1+\exp(-Df(\theta))} \tag{1}$$

を項目特性関数として用いる (D は定数 1.7)。特に (1) で $f(\theta) = a(\theta - b)$ と、a と b の 2 つのパラメタを採用したものを 2 パラメタ・ロジスティックモデルと呼ぶ。

3　情報関数

この章は豊田 [2] を参考にした。項目反応とは項目に対する回答結果を示す。それは、

$$u_j = \begin{cases} 1 & (\text{項目 } j \text{ に正解}) \\ 0 & (\text{項目 } j \text{ に不正解}) \end{cases} \tag{2}$$

である。また、被験者 i の項目 j に対する項目反応を u_{ij} と表現する。

項目反応モデルにおいて、尤度関数 $L(\boldsymbol{u_i}\,|\,\theta)$ は次式で与えられる。被験者を $i = 1, 2, \cdots, N$、項目を $j = 1, 2, \cdots, m$ とすると

$$L(\boldsymbol{u_i}\,|\,\theta_i) = \prod_{j=1}^{m} P_j(\theta_i)^{u_{ij}} Q_j(\theta_i)^{1-u_{ij}} \tag{3}$$

と与えられる。$Q_j(\theta)$ は不正答の確率で、$Q_j(\theta) = 1 - P_j(\theta)$ である。次に、テストが特性値の分布に対してどの程度の精度で測定されているかを考える。N が大であれば θ の最尤推定値 $\hat{\theta}$ の分散は、$I(\theta)^{-1}$ に限りなく近づく。$I(\theta)$ はフィッシャー情報量と呼ばれ、

$$I(\theta_i) = E\left[\left(\frac{\partial}{\partial\theta}\log L(\boldsymbol{u_i}\,|\,\theta)\right)^2_{\theta=\theta_i}\right] \tag{4}$$

と定義される。

4　パラメタの推定

4.1　被験者特性値の推定

この節は芝 [1] と豊田 [2] を参考にした。項目パラメタが既知であることが前提でテストを行い、結果として項目反応が得られた状況を仮定する。即ち、項目特性関数は θ の 1 変数関数である。その上で、被験者特性値を推定していく。推定の方法は最尤推定法とベイズ推定法の 2 つがある。

4.2　被験者母集団の推定

この節は芝 [1] を参考にした。仮定は前節と同様とする。今、特性値の母集団分布が母数 $\boldsymbol{\mu}$ で表現できるとする。例えば、母集団分布として正規分布を仮定すると $\boldsymbol{\mu} = (\text{平均}, \text{分散})$ と考えることが出来る。ここでは、この母集団分布を $h(\theta\,|\,\boldsymbol{\mu})$ とする。

この $\boldsymbol{\mu}$ を推定するのだが、被験者特性値 θ が未知であるので $L(u\,|\,\theta)$ ではなく、その周辺分布 $L(u\,|\,\boldsymbol{\mu})$ を $\boldsymbol{\mu}$ の関数とみなした尤度関数から推定を行う。被験者を $i = 1, 2, \cdots, N$、項目反応データ行列を U とすると、周辺尤度関数は

$$f(\boldsymbol{\mu}\,|\,U) = \prod_{i=1}^{N} L(\boldsymbol{u_i}\,|\,\boldsymbol{\mu}) = \prod_{i=1}^{N} \int_{-\infty}^{+\infty} L(\boldsymbol{u_i}\,|\,\theta)\, h(\theta\,|\,\boldsymbol{\mu})\, d\theta \tag{5}$$

となる。この周辺尤度関数は積分を含む性質から解析が困難なため、その方法として EM アルゴリズムと呼ばれる方法を用いる (渡辺・山口 [3] 参照)。

EM アルゴリズムとは、欠測値 (この場合は特性値 θ) を含んだ尤度関数を最大化する最尤推定値を求めるのに使われる方法である。具体的には、

E-step　対数周辺尤度関数の期待値 $Q(\boldsymbol{\mu};\, \boldsymbol{\mu}_{(n)}) = E[\log f(\boldsymbol{\mu}_{(n)}\,|\,U)]$ を求める。

M-step　$Q(\boldsymbol{\mu}_{(n+1)};\, \boldsymbol{\mu}_{(n)}) \geq Q(\boldsymbol{\mu};\, \boldsymbol{\mu}_{(n)})$ となるような $\boldsymbol{\mu}_{(n+1)}$ を求める。

M-step においての $\boldsymbol{\mu}_{(n+1)}$ の求め方としては、$\partial\, Q(\boldsymbol{\mu};\, \boldsymbol{\mu}_{(n)}) \,/\, \partial\, \boldsymbol{\mu} = 0$ の解を $\boldsymbol{\mu}_{(n+1)}$ として更新する。

以上の 2 つの step を任意の精度を満たすまで繰り返すことによって、最尤推定値 $\hat{\mu}$ を求めることが出来る。ただ、EM アルゴリズムは繰り返し回数が多く必要とされるので、母集団の規模を大きくすればそれだけ計算に時間がかかってしまう。

4.3　項目パラメタの推定

この節は、芝 [1] を参考にした。被験者母集団は既に既知であるとし、得られた項目反応データから項目パラメタを推定していくことを考える。推定すべき項目パラメタを $\Lambda = (\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \cdots, \boldsymbol{\lambda_m})$ とする。具体的に 2 パラメタ・ロジスティックモデルにおいて考えれば、

$\boldsymbol{\lambda}_j = (a_j, b_j)$ となる。Λ の推定法は、前節と同様に EM アルゴリズムを用いる。

E-step 対数周辺尤度の期待値 $Q(\Lambda\,;\Lambda_{(n)}) = E[L(\Lambda_{(n)}\,|\,U)]$ を求める。

M-step $Q(\Lambda_{(n+1)}\,;\Lambda_{(n)}) \geq Q(\Lambda\,;\Lambda_{(n)})$ となるような $\Lambda_{(n+1)}$ を求める。

M-step においての $\Lambda_{(n+1)}$ の求め方には Newton 法を用いる。以上のアルゴリズムを繰り返すことによって対数周辺尤度関数を増加させ、$\hat{\Lambda}$ を求められる。

4.4 被験者母集団と項目パラメタの同時推定

この節は芝 [1] を参考にした。被験者母集団と項目パラメタが未知であり、得られた項目反応データからその 2 つを推定していくことを考える。推定の方法は、被験者母集団の推定と項目パラメタの推定を併せた形となる。しかし項目反応モデルは特性値の尺度の原点が不定であり、ただ推定しただけでは意味がない。推定を有効にするためには何らかの制約条件を設ける必要が出てくる。それは一般的に、母集団が平均 0、分散 1 という分布になるという制約条件を設ける方法がとられている。

5 等化とは

この章は、芝 [1] と豊田 [2] を参考にした。異なる母集団に異なる版のテストを行ったとしよう。このとき各母集団において異なる尺度において特性値が推定されるので、母集団間における比較が出来ない。では、それが出来るようにするためには各母集団を共通の尺度におく必要が出てくる。そのためには、項目または被験者を共通させる必要がある。各版に推定された項目パラメタを、共通デザインを用いて全ての項目パラメタを共通の尺度で表現する。これを等化と呼ぶ。具体的に、$\theta^* = k\theta + l$、$a^* = a/k$、$b^* = kb + l$ とすることによって $P(\theta) = P(\theta^*)$ と 1 次変換することができる。これによって項目パラメタを等化することができ、k と l を等化係数という。

6 応用例

実施したテストは、漢字の読みの能力を測定することを目的として作成した。測定の対象は本大学の数理情報学部・総合政策学部の 2 学部、合計 78 名に対してテスト A、B の 2 つを行った。2 つのテストは、等化を行うために 10 問の項目を共通させた。このデータを自分で作成したプログラムにて解析を行う。

6.1 項目パラメタの推定と等化

現在得られている情報は項目反応データのみであり、項目パラメタと母集団分布は未知である。よって 78 名の母集団分布を平均 0、分散 1 の正規分布であると仮定した。この推定の結果、テスト A、B いずれにおいても通過率の低い項目の困難度パラメタは高く、通過率の高い項目の困難度パラメタは低く推定されている傾向がある。また、項目数に対して被験者数が少ないために標準誤差は大きく出てしまった。

次に 2 つのテストを共通の尺度にて考えるために、推定された項目パラメタの等化を行った。等化の結果、等化係数は $\hat{k} = 0.837$、$\hat{l} = -0.240$ と求められた。

6.2 母集団分布の推定

等化後の項目パラメタを用いて、学部別に母集団分布を推定する。分布は正規分布を仮定し、その平均と標準偏差を推定する。この推定の結果、数理情報学部は総合政策学部に比べて平均が低く標準偏差が高く推定され、両方のテストおいてそれがいえた。つまり数理情報学部の能力は、総合政策学部よりも僅かに低くかつ広く分布しているといえる。

ここで各被験者の特性値を求めて、学部別にヒストグラムにしてみた。また、推定された各被験者の特性値を、学部別に平均と分散をとってみた。その結果、数理情報学部は平均 −0.180、分散 1.165、総合政策学部は平均 0.094、分散 0.881 となった。学部によって被験者数に差が大きいのに問題はあるが、やはり数理情報学部の方がやや低く広く分布していることがわかる。

6.3 テスト情報曲線

$I(\theta)$ を曲線にして描いた。これをテスト情報曲線と呼ぶ。この結果、テスト A とテスト B の両方とも $-2 \leq \theta \leq 2$ の範囲において高い情報量を得ていた。その中でもテスト B においては最も高い情報量を得ているのは $\theta = -1$ の辺りと、平均よりやや低めの値が突出していた。つまり、テスト B においては総合政策学部よりも数理情報学部の方がやや精度良く測定出来ているということである。一方、テスト A はテスト B より情報量が小さいながらも総合政策学部に対して有用であるといる。このようにテスト情報関数を上手く使うことによって、効率よく項目を構成することも出来る。

7 おわりに

本研究を通して難航したのは、EM アルゴリズム用いた推定プログラムの作成である。前述の通り、EM アルゴリズムは計算に時間がかかってしまうという性質を備えている。この性質から、実際に正しい推定値が得られるようプログラムを組むのに時間がかかってしまった。

今後の問題点は項目パラメタの等化であろう。項目の等化を精度良く行っていくためには、多数の項目の中から選択してテストを作成し、それを繰り返していく必要がある。それには大規模なデータが必要とされるため、個人が解析を行っていくのには困難である。

参考文献

[1] 芝 祐順 (編),「項目反応理論－基礎と応用－」, 東京大学出版会, 1991.

[2] 豊田 秀樹,「項目反応理論 [入門編] －テストと測定の科学－」, 朝倉書店, 2002.

[3] 渡辺 美智子・山口 和憲 (編),「EM アルゴリズムと不完全データの諸問題」, 多賀出版, 2000.