項目反応理論に関するいくつかの話題

前川眞一

東京工業大学
大学院社会理工学研究科
2012.12.08 JART

アウトライン

0. 多項分布の確率の推定

1. 項目パラメタの推定

2. 順序カテゴリの取り扱い

3. 重み付き得点からのθの推定

4. 独立に推定された項目パラメタの等化

0. 多項分布の確率の推定

1-1
数値的カテゴリを持つ多項分布の確率の推定

1-2
数値的カテゴリを持つ多項分布の和の分布

1-3
カテゴリカルデータの回帰分析

0-1 数値的カテゴリを持つ多項分布 の確率の推定

- 数値的カテゴリを持つ多項分布

Y を値域 $\{A_1, A_2, \ldots, A_m\}$ を持つ離散型確率変数とする。
$(m+1)\times 1$ のダミー変数 $\boldsymbol{Z}$ ベクトルを

$$Z_k = \begin{cases} 1 & \text{if } Y = A_k,\ k = 1, 2, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

と定義すると $\boldsymbol{Z}$ は多項分布をする。

$$\boldsymbol{Z} \sim MN(1, \boldsymbol{\pi})$$

ただし、 $\boldsymbol{\pi}$ は以下を満たすカテゴリの生起確率ベクトルである。

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_m), \quad \pi_k > 0, \quad \sum_{k=1}^{m} \pi_k = 1$$

$\boldsymbol{Z}$ の確率関数は

$$\Pr(Z_1 = z_1 \ \&\ Z_2 = z_2 \ \&\ , \ldots, \ \&\ Z_m = z_n | \boldsymbol{\pi}) = \prod_{k=1}^{m} \pi_k^{z_k}$$

0-1 数値的カテゴリを持つ多項分布 の確率の推定

n 個の観測値

$$z_i = (z_{i1}, z_{i2}, z_{i3}, \ldots, z_{im}), \quad i = 1, 2, \ldots, n$$

を与えられた時のパラメタ $\boldsymbol{\pi}$ の尤度は

$$L(\boldsymbol{\pi}) = \prod_{i=1}^{n}\prod_{k=1}^{m} \pi_k^{z_{ik}}$$

n 個の観測値から計算した度数分布

$$(A_k, f_k), \quad k = 1, 2, \ldots, m$$

ただし

$$f_k = \sum_{i=1}^{n} z_{ik}, \quad \sum_{k=1}^{m} f_k = n$$

が与えられた時のパラメタ $\boldsymbol{\pi}$ の尤度は

$$L(\boldsymbol{\pi}) = \prod_{k=1}^{m} \pi_k^{\sum_{i=1}^{n} z_{ik}} = \prod_{k=1}^{m} \pi_k^{f_k}$$

である。

0-1 数値的カテゴリを持つ多項分布 の確率の推定

尤度関数

$$L(\boldsymbol{\pi}) = \prod_{k=1}^{m} \pi_k^{f_k}$$

を最大とするパラメタの推定値、最尤解は

$$\pi_k = \frac{f_k}{\sum_{k'=1}^{m} f_{k'}} = \frac{f_k}{n}$$

である。

多項分布のカテゴリの値 $\{A_1, A_2, \ldots, A_m\}$ が数値の場合、
π_k を A_k の関数として表すことを考える。すなわち

$$\pi_k \propto g(A_k|\boldsymbol{\xi})$$

または

$$\pi_k = \pi_k(A_k) = \frac{g(A_k|\boldsymbol{\xi})}{\sum_{k=1}^{m} g(A_k|\boldsymbol{\xi})}$$

とする。

0-1 数値的カテゴリを持つ多項分布 の確率の推定

多項分布の確率

$$\Pr(Y = A_k) = \Pr(Z_k = 1) = \pi_k$$

をカテゴリの関数として求める。

$$\pi_k = \pi_k(A_k) = \frac{g(A_k|\boldsymbol{\xi})}{\sum_{k=1}^{m} g(A_k|\boldsymbol{\xi})}$$

ただし、$\boldsymbol{\xi}$ は関数 g の形を決めるパラメタで、（対数）尤度関数

$$L(\boldsymbol{\xi}) = \prod_{k=1}^{m} \pi_k^{f_k} = \prod_{k=1}^{m} \left(\frac{g(A_k|\boldsymbol{\xi})}{\sum_{k'=1}^{m} g(A_{k'}|\boldsymbol{\xi})} \right)^{f_k}$$

$$lnL(\boldsymbol{\xi}) = \sum_{k=1}^{m} f_k \log \left(\frac{g(A_k|\boldsymbol{\xi})}{\sum_{k'=1}^{m} g(A_{k'}|\boldsymbol{\xi})} \right)$$

を最大とするようにして求める。

0-1 数値的カテゴリを持つ多項分布 の確率の推定

- *g* の例

- Normal pdf: $\boldsymbol{\xi} = (\mu,\ \sigma)$

$$g(A_k|\boldsymbol{\xi}) = \exp\left(-0.5(A_k - \mu)^2/\sigma\right)$$

- Normal cdf: $\boldsymbol{\xi} = (\mu,\ \sigma)$

$$g(A_k|\boldsymbol{\xi}) = \Phi\left(\frac{(A_k + A_{k+1})/2 - \mu}{\sigma}\right) - \Phi\left(\frac{(A_{k-1} + A_k)/2 - \mu}{\sigma}\right)$$

0-1 数値的カテゴリを持つ多項分布 の確率の推定

Fisher のスコア法によるパラメタの推定

対数尤度関数の一次偏微分

$$\frac{\partial\, lnL}{\partial\, \boldsymbol{\xi}} = \sum_{k=1}^{m}(f_k - n\pi_k)\frac{\partial\, \log g(A_k|\boldsymbol{\xi})}{\partial\, \boldsymbol{\xi}} = \sum_{k=1}^{m}\frac{(f_k - n\pi_k)}{g(A_k|\boldsymbol{\xi})}\frac{\partial\, g(A_k|\boldsymbol{\xi})}{\partial\, \boldsymbol{\xi}}$$

Fisher Infomation Matrix

$$\boldsymbol{H}(\boldsymbol{\xi}) = \boldsymbol{\mathcal{E}}\left(\left(\frac{\partial\, lnL}{\partial\, \boldsymbol{\xi}}\right)\left(\frac{\partial\, lnL}{\partial\, \boldsymbol{\xi}}\right)'\right) = n\, \boldsymbol{S}(\boldsymbol{\xi})'(\mathrm{Diag}\,(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')\boldsymbol{S}(\boldsymbol{\xi})$$

ただし

$$\boldsymbol{S}(\boldsymbol{\xi}) = \frac{\partial\, \log \boldsymbol{g}(\boldsymbol{A}|\boldsymbol{\xi})}{\partial\, \boldsymbol{\xi}} = \left\{\frac{\partial\, \log g(A_k|\boldsymbol{\xi})}{\partial\, \xi_\ell}\right\} = \left\{\frac{1}{g(A_k|\boldsymbol{\xi})}\frac{\partial\, g(A_k|\boldsymbol{\xi})}{\partial\, \xi_\ell}\right\}$$

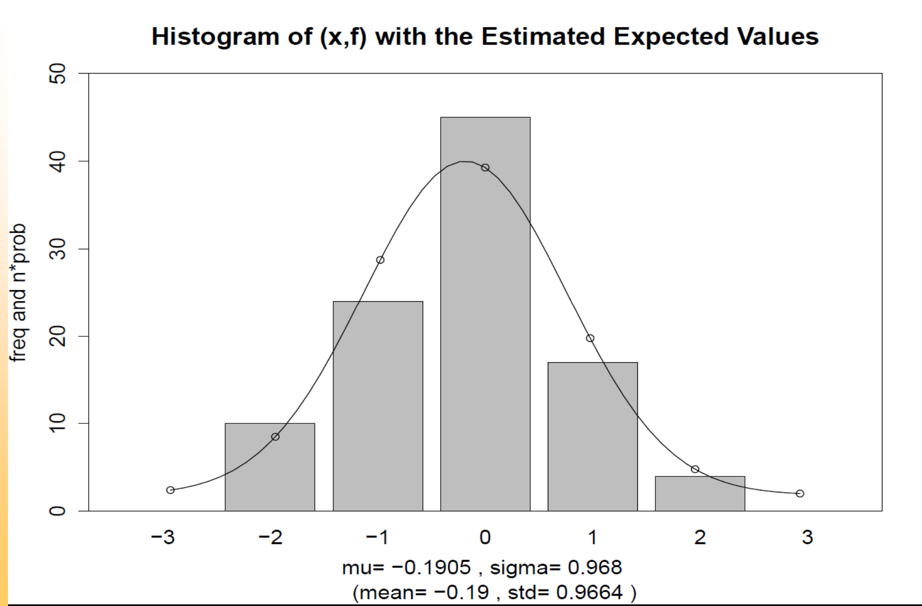
$$k = 1, 2, \ldots, m,\ \ell = 1, 2, \ldots, dim(\boldsymbol{\xi})$$

0-1 数値的カテゴリを持つ多項分布 の確率の推定

データ n=100 normal pdf model

A	f	n*pi	pi
-3	0	0.610	0.006
-2	10	7.183	0.071
-1	24	29.059	0.290
0	45	40.432	0.404
1	17	19.348	0.193
2	4	3.184	0.031
3	0	0.180	0.001

0-1 数値的カテゴリを持つ多項分布 の確率の推定



0-1 数値的カテゴリを持つ多項分布 の確率の推定

その他

平均と分散の制約を付けた確率の推定

$$平均 : mean = \sum_{k=1}^{m} A_k\pi_k$$

$$分散 : var = \sum_{k=1}^{m} A_k^2\pi_k - \left(\sum_{k=1}^{m} A_k\pi_k\right)^2$$

0-2 数値的カテゴリを持つ多項分布の和の分布

多項分布のカテゴリが $m+1$ の数値 $\{0,1,2,\ldots,m\}$
から形成されている場合、それを数値化された多項分布と呼び

$$Y \sim MN((0,1,\ldots,m),\boldsymbol{\pi}),\ 0 \le Y \le m$$

と記す。ただし、 $\boldsymbol{\pi} = \{\pi_0,\pi_1,\pi_2,\ldots,\pi_m\}$ は

$$\sum_{k=0}^{m} \pi_k = 1,$$

を満たす各カテゴリの生起確率であり、m をその分布の満点とよぶ。

0-2 数値的カテゴリを持つ多項分布の和の分布

定理　数値化された多項分布をする確率変数の和の分布

独立な二つの数値化された多項分布をする確率変数

$$\begin{aligned} Y_1 &\sim MN((0,1,\ldots,m_1),\boldsymbol{\pi}_1),\ 0 \le Y_1 \le m_1, \quad \boldsymbol{\pi}_1 = (\pi_{01},\pi_{11},\pi_{21},\ldots,\pi_{m_1 1}) \\ Y_2 &\sim MN((0,1,\ldots,m_2),\boldsymbol{\pi}_2),\ 0 \le Y_2 \le m_2, \quad \boldsymbol{\pi}_2 = (\pi_{02},\pi_{12},\pi_{22},\ldots,\pi_{m_2 2}) \end{aligned}$$

を、正の整数 w_2 を重みとして加えたもの $Y_{1+2} = Y_1 + w_2 Y_2$ は、
満点が $m_1 + w_2 m_2$ の数値化された多項分布

$$Y_{1+2} = Y_1 + w_2 Y_2, \sim MN((0,1,\ldots,m_1+m_2),\boldsymbol{\pi}_{1+2}), \quad 0 \le Y_{1+2} \le m_1 + m_2$$

となり、その生起確率 $\boldsymbol{\pi}_{1+2}$ は

$$\boldsymbol{\pi}_{1+2} = \sum_{k=0}^{m_2} \boldsymbol{z}_k$$

0-2 数値的カテゴリを持つ多項分布の和の分布

- 定理の続き

$$\boldsymbol{\pi}_{1+2} = \sum_{k=0}^{m_2} \boldsymbol{z}_k$$

ただし、

$$\boldsymbol{z}_k = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \pi_{k2}\boldsymbol{\pi}_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} w_2\ k \times 1 \\ (m_1+1)\times 1 \\ \left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} w_2(m_2-k)\times 1 \end{array}$$

であたえられる。

0-2 数値的カテゴリを持つ多項分布の和の分布

- 例　2 点満点と 3 点満点のテストの和

$m_1 = 2, \quad m_2 = 3, \quad w_2 = 2$ の場合の例
この場合、$\boldsymbol{\pi}_1 = \{\pi_{01},\pi_{11},\pi_{21}\}$ また $\boldsymbol{\pi}_2 = \{\pi_{02},\pi_{12},\pi_{22},\pi_{23}\}$ であり、$\boldsymbol{\pi}_{1+2}$ は
$(2+2\times 3+1)\times 1$ のベクトルとなる。

$$\boldsymbol{\pi}_{1+2} = \begin{pmatrix} \pi_{02}\begin{pmatrix}\pi_{01}\\ \pi_{11}\\ \pi_{21}\end{pmatrix} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \pi_{12}\begin{pmatrix}\pi_{01}\\ \pi_{11}\\ \pi_{21}\end{pmatrix} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \pi_{22}\begin{pmatrix}\pi_{01}\\ \pi_{11}\\ \pi_{21}\end{pmatrix} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \pi_{32}\begin{pmatrix}\pi_{01}\\ \pi_{11}\\ \pi_{21}\end{pmatrix} \end{pmatrix}$$

0-3カテゴリカルデータの回帰分析

データ $(y_i, \boldsymbol{x}_i), \quad , i = 1, 2, \ldots, n$

$(\boldsymbol{z}_i, \boldsymbol{x}_i), \quad , i = 1, 2, \ldots, n$

$$(\boldsymbol{r}_i, n_i, \boldsymbol{x}_i) \quad , i = 1, 2, \ldots, N, n_i = \sum_{k=0}^{m} r_{ik}$$

確率が説明変数の関数として変化する

$$\Pr(y_i = A_k | \boldsymbol{x}_i) = \Pr(z_{ik} = 1 | \boldsymbol{x}_i) = \pi_k(\boldsymbol{x}_i)$$

確率のモデル

$$\pi_k(\boldsymbol{x}_i) = \frac{g_k(\boldsymbol{x}_i|\boldsymbol{\xi})}{\sum_{k'=0}^{m} g_{k'}(\boldsymbol{x}_i|\boldsymbol{\xi})}$$

$$\pi_k = \pi_k(A_k) = \frac{g(A_k|\boldsymbol{\xi})}{\sum_{k=0}^{m} g(A_k|\boldsymbol{\xi})}$$

多項ロジットモデル

$$g_k(\boldsymbol{x}_i|\boldsymbol{\xi}) = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_k + c_k)$$

0-3カテゴリカルデータの回帰分析

確率のモデル

$$\pi_k(\boldsymbol{x}_i) = \frac{g_k(\boldsymbol{x}_i|\boldsymbol{\xi})}{\sum_{k'=0}^{m} g_{k'}(\boldsymbol{x}_i|\boldsymbol{\xi})}$$

多項ロジットモデル

$$g_k(\boldsymbol{x}_i|\boldsymbol{\xi}) = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_k + c_k)$$

尤度関数

$$L(\boldsymbol{\xi}) = \prod_{i=1}^{n}\prod_{k=0}^{m} \pi_k(\boldsymbol{x}_i|\boldsymbol{\xi})^{z_{ik}} = \prod_{i=1}^{n}\prod_{k=0}^{m} \pi_k(\boldsymbol{x}_i|\boldsymbol{\xi})^{r_{ik}}$$

対数尤度関数

$$lnL(\boldsymbol{\xi}) = \sum_{i=1}^{n}\sum_{k=0}^{m} r_{ik}\log \pi_k(\boldsymbol{x}_i|\boldsymbol{\xi})$$

0-3カテゴリカルデータの回帰分析

●Fisher のスコア法によるパラメタの推定

対数尤度関数のパラメタに関する一次偏微分

$$\frac{\partial\, lnL}{\partial\, \boldsymbol{\xi}} = \sum_{i=1}^{N}\sum_{k=0}^{m}(r_{ij} - n_i\pi_k(\boldsymbol{x}_i|\boldsymbol{\xi}))\frac{\partial\, \ln g_k(\boldsymbol{x}_i|\boldsymbol{\xi})}{\partial\, \boldsymbol{\xi}}$$

$$= \sum_{i=1}^{N}\sum_{k=0}^{m}\frac{(r_{ik} - n_i\pi_k(\boldsymbol{x}_i|\boldsymbol{\xi}))}{g_k(\boldsymbol{x}_i|\boldsymbol{\xi})}\frac{\partial\, g_k(\boldsymbol{x}_i|\boldsymbol{\xi})}{\partial\, \boldsymbol{\xi}}$$

Fisher Infomation Matrix

$$\boldsymbol{H}(\boldsymbol{\xi}) = \boldsymbol{\mathcal{E}}\left(\left(\frac{\partial\, lnL}{\partial\, \boldsymbol{\xi}}\right)\left(\frac{\partial\, lnL}{\partial\, \boldsymbol{\xi}}\right)'\right) = n\, \boldsymbol{S}(\boldsymbol{\xi})'(\mathrm{Diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}')\boldsymbol{S}(\boldsymbol{\xi})$$

ただし

$$\boldsymbol{S}(\boldsymbol{\xi}) = \frac{\partial\, \log \boldsymbol{g}(\boldsymbol{x}|\boldsymbol{\xi})}{\partial\, \boldsymbol{\xi}} = \left\{\frac{\partial\, \log g_k(\boldsymbol{x}_i|\boldsymbol{\xi})}{\partial\, \xi_\ell}\right\} = \left\{\frac{1}{g_k(\boldsymbol{x}_i|\boldsymbol{\xi})}\frac{\partial\, g_k(\boldsymbol{x}_i|\boldsymbol{\xi})}{\partial\, \xi_\ell}\right\}$$

$$i = 1, 2, \ldots, N, \ k = 0, 1, \ldots, m, \ \ell = 1, 2, \ldots, dim(\boldsymbol{\xi})$$

0-3カテゴリカルデータの回帰分析

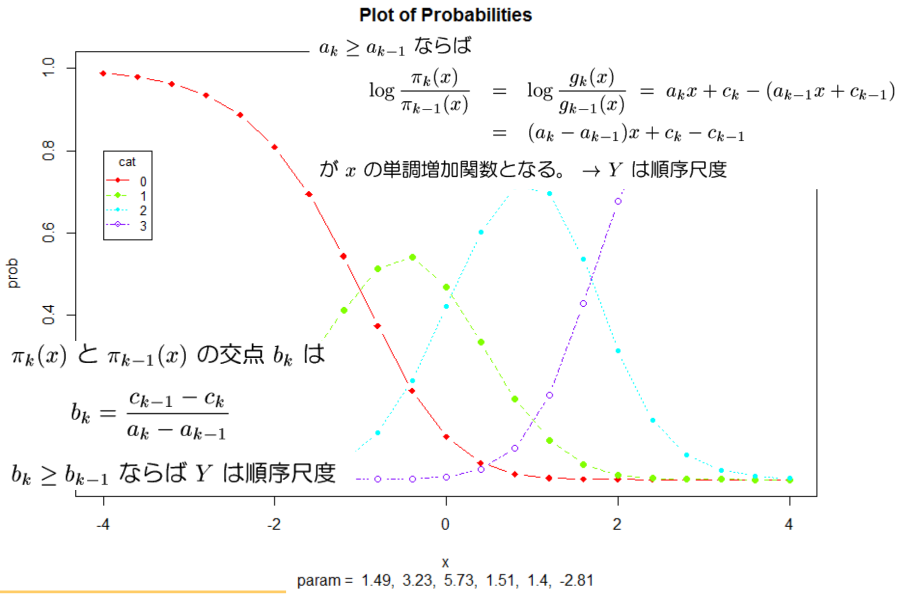
x が一次元の場合の例　*m*=3

x	A_0	A_1	A_2	A_3	n	p_0	p_1	p_2	p_3
-4.0	0	0	0	0	0	0.988	0.011	0.000	0.000
-3.6	0	0	0	0	0	0.979	0.020	0.000	0.000
-3.2	0	0	0	0	0	0.962	0.037	0.000	0.000
-2.8	1	1	0	0	2	0.934	0.065	0.000	0.000
-2.4	4	0	0	0	4	0.886	0.112	0.001	0.000
-2.0	10	1	0	0	11	0.808	0.186	0.005	0.000
-1.6	16	5	1	0	22	0.694	0.289	0.015	0.000
-1.2	22	14	3	0	39	0.543	0.411	0.045	0.000
-0.8	18	31	9	0	58	0.373	0.512	0.113	0.000
-0.4	16	36	23	0	75	0.216	0.540	0.241	0.001
0.0	6	36	36	3	81	0.103	0.468	0.421	0.006
0.4	2	28	42	2	74	0.040	0.333	0.601	0.024
0.8	0	13	39	6	58	0.013	0.196	0.713	0.077
1.2	0	6	29	4	39	0.003	0.095	0.696	0.204
1.6	0	1	14	7	22	0.000	0.036	0.535	0.427
2.0	0	0	7	4	11	0.000	0.010	0.312	0.676
2.4	0	0	1	3	4	0.000	0.002	0.145	0.852
2.8	0	0	0	2	2	0.000	0.000	0.059	0.940
3.2	0	0	0	0	0	0.000	0.000	0.022	0.977
3.6	0	0	0	0	0	0.000	0.000	0.008	0.991
4.0	0	0	0	0	0	0.000	0.000	0.003	0.996

$$g(x_i|\boldsymbol{\xi}) = \exp(a_k x_i + c_k)$$

$$a_0 = 0 \qquad c_0 = 0$$
$$a_1 = 1.489 \qquad c_1 = 1.508$$
$$a_2 = 3.234 \qquad c_2 = 1.401$$
$$a_3 = 5.728 \qquad c_3 = -2.814$$

0-3カテゴリカルデータの回帰分析



$a_k \geq a_{k-1}$ ならば

$$\log \frac{\pi_k(x)}{\pi_{k-1}(x)} = \log \frac{g_k(x)}{g_{k-1}(x)} = a_k x + c_k - (a_{k-1}x + c_{k-1}) = (a_k - a_{k-1})x + c_k - c_{k-1}$$

が x の単調増加関数となる。$\rightarrow Y$ は順序尺度

$\pi_k(x)$ と $\pi_{k-1}(x)$ の交点 b_k は

$$b_k = \frac{c_{k-1} - c_k}{a_k - a_{k-1}}$$

$b_k \geq b_{k-1}$ ならば Y は順序尺度

項目反応理論に関するいくつかの話題

1. 項目パラメタの推定

$P_{jk}(\theta)$ を θ を項目 j のカテゴリ k の項目カテゴリ反応関数 (ICRF) とする。

$$j = 1, 2, \ldots, p, \quad k = 0, 1, 2, \ldots, m_j$$

U_{ijk} を、個人 i が項目 j に於いて第 k 番目の得点を取った時に 1 となる変数とする。

θ を連続量と仮定すると、項目パラメタの周辺尤度関数は

$$mL(\Xi) = \prod_{i=1}^{n} \int_{\Theta} \prod_{j=1}^{p} \prod_{k=0}^{m_j} P_{jk}(\theta|\boldsymbol{\xi}_j)^{u_{ijk}} \quad h(\theta|\boldsymbol{\rho})\, d\theta$$

と書ける。$h(\theta|\boldsymbol{\rho})$ はパラメタ $\boldsymbol{\rho}$ を持つ正規分布の pdf

θ を $\theta_q, \quad q = 1, 2, \ldots, Q,$ の離散値を取る変数

と仮定すると、項目パラメタの周辺尤度関数は

$$mL(\Xi) = \prod_{i=1}^{n} \sum_{q=1}^{Q} h(\theta_q|\boldsymbol{\rho}) \prod_{j=1}^{p} \prod_{k=0}^{m_j} P_{jk}(\theta_q|\boldsymbol{\xi}_j)^{u_{ijk}}$$

ただし、$\Pr(\Theta = \theta_q) = h(\theta_q|\boldsymbol{\rho}) \propto \exp\left(-0.5(\theta_q - \mu)^2/\sigma^2\right)$

1. 項目パラメタの推定

θ が離散の場合の EM アルゴリズム

E-step: データ $\boldsymbol{U}$ を与えた時の

$$\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*, \ldots, \theta_Q^*)$$

の条件付き分布は多項分布

$$\boldsymbol{\theta}^*|\boldsymbol{u}_i \sim MN((\theta_1, \theta_2, \ldots, \theta_Q), \boldsymbol{h}_i)$$

ただし

$$\boldsymbol{h}_i = (h_{i1}, h_{i2}, \ldots, h_{iQ})$$

$$h_{iq} = \frac{\prod_{j=1}^{p} \prod_{k=0}^{m_j} P_{jk}(\theta_q|\boldsymbol{\xi}_j)^{u_{ijk}} \quad h(\theta_q|\boldsymbol{\rho})}{\sum_{q'=1}^{Q} \prod_{j=1}^{p} \prod_{k=0}^{m_j} P_{jk}(\theta_{q'}|\boldsymbol{\xi}_j)^{u_{ijk}} \quad h(\theta_{q'}|\boldsymbol{\rho})}$$

したがって

$$\boldsymbol{\mathcal{E}}(\boldsymbol{\theta}^*|\boldsymbol{u}_i) = \boldsymbol{h}_i, \quad \boldsymbol{\mathcal{E}}(\theta_q^*|\boldsymbol{u}_i) = h_{iq},$$

1. 項目パラメタの推定

期待対数完全データ尤度

$$\mathcal{E}\ln L(\mathbf{\Lambda},\rho) = \sum_{j=1}^{p}\sum_{q=1}^{Q}\sum_{k=0}^{m_j}\left(\sum_{i=1}^{N}h_{iq}u_{ijk}\right)\log P_{jk}(\theta_q|\boldsymbol{\xi}_j) + \sum_{q=1}^{Q}(\sum_{i=1}^{N}h_{iq})\log h(\theta_{q'}|\boldsymbol{\rho})$$

$$h(\theta_q|\mu,\sigma) = \frac{\exp\left(-0.5(\theta_q-\mu)^2/\sigma^2\right)}{\sum_{q'=1}^{Q}\exp\left(-0.5(\theta_{q'}-\mu)^2/\sigma^2\right)}$$

多項ロジットの対数尤度関数

$$lnL(\boldsymbol{\xi}) = \sum_{i=1}^{n}\sum_{k=0}^{m} r_{ik}\log\pi_k(\boldsymbol{x}_i|\boldsymbol{\xi})$$

数値カテゴリを持つ多項分布の対数尤度関数

$$lnL(\boldsymbol{\xi}) = \sum_{k=1}^{m} f_k\log\left(\frac{g(A_k|\boldsymbol{\xi})}{\sum_{k'=1}^{m}g(A_{k'}|\boldsymbol{\xi})}\right)$$

1. 項目パラメタの推定

項目カテゴリ反応関数 ICRF

3PLM: $m_j = 1$

$$P_{j1}(\theta) = c_j + \frac{1-c_j}{1+\exp\left(-1.7a_j(\theta-b_j)\right)}$$

$$P_{j0}(\theta) = 1 - P_{j1}(\theta)$$

Graded Response Model

$$P_{jk}(\theta) = P^{+}_{j,k}(\theta) - P^{+}_{j,k+1}(\theta), \quad k = 0,1,2,\ldots,m_j$$

$$P^{+}_{j,k}(\theta) = \frac{1}{1+\exp\left(-1.7a_j(\theta-b_{jk})\right)}$$

$$P^{+}_{j0}(\theta) = 1, \quad P^{+}_{j,m_j+1}(\theta) = 0$$

1. 項目パラメタの推定

項目カテゴリ反応関数 ICRF

Nominal Response Model (NRM)

$$P_{jk}(\theta) = \frac{\exp\left(a_{jk}\theta + c_{jk}\right)}{\sum_{k'=0}^{m_j}\exp\left(a_{jk'}\theta + c_{jk'}\right)}, \quad a_{j0} = c_{j0} = 0$$

$$= \frac{\exp\left(a_{jk}\theta + c_{jk}\right)}{1+\sum_{k'=1}^{m_j}\exp\left(a_{jk'}\theta + c_{jk'}\right)}, \quad P_{j0}(\theta) = 1 - \sum_{k=1}^{m_j}P_{jk}(\theta)$$

2PLM as a special case of NRM

$$m_j = 1, \quad a_{j1} = 1.7a_j, \quad c_{j1} = -1.7a_j\, b_j$$

Generalized Partial Credit Model as a special case of NRM

$$a_{jk} = 1.7a_j \times k$$

1. 項目パラメタの推定

Fisher のスコア法を用いた場合には

$$\frac{\partial\ \log P_{jk}(\theta|\boldsymbol{\xi}_j)}{\partial\ \boldsymbol{\xi}_j}$$

のみが必要。

2 順序カテゴリの取り扱い

Nominal Response Model (NRM)

$$P_{jk}(\theta) \;=\; \frac{\exp\left(a_{jk}\theta + c_{jk}\right)}{\sum_{k'=0}^{m_j}\exp\left(a_{jk'}\theta + c_{jk'}\right)}, \qquad a_{j0} = c_{j0} = 0$$

$a_{jk} \geq a_{j,k-1}$ ならば

$$\begin{aligned}\log\frac{P_{jk}(\theta)}{P_{j,k-1}(\theta)} &= \;= a_{jk}\theta + c_{jk} - (a_{j,k-1}\theta + c_{j,k-1}) \\ &= (a_{jk} - a_{j,k-1})\theta + c_{jk} - c_{j,k-1}\end{aligned}$$

が θ の単調増加関数となる。 $\rightarrow$ 項目 j は順序尺度

$P_{jk}(\theta)$ と $P_{j,k-1}(\theta)$ の交点 b_{jk} は

$$b_{jk} = \frac{c_{j,k-1} - c_{jk}}{a_{jk} - a_{j,k-1}}$$

$b_{jk} \geq b_{j,k-1}$ ならば 項目 j は順序尺度

2 順序カテゴリの取り扱い

New Ordered Model

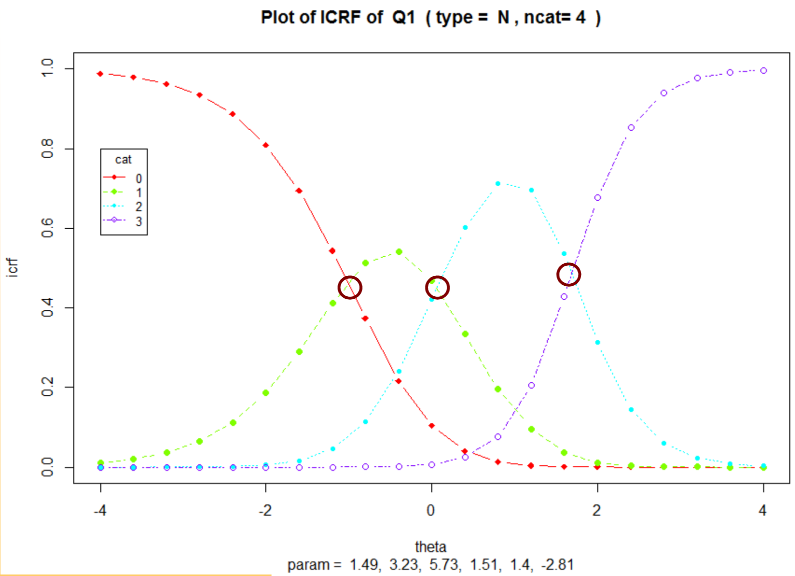
$$P_{jk}(\theta) \;=\; \frac{\exp\left(a_{jk}\theta - \sum_{h=1}^{k}(a_{jh} - a_{j,h-1})b_{jh}\right)}{\sum_{k=0}^{m_j}\exp\left(a_{jk}\theta - \sum_{h=1}^{k}(a_{jh} - a_{j,h-1})b_{jh}\right)}$$

$$a_{jk} \geq a_{j,k-1} \quad \text{and} \quad b_{jk} \geq b_{j,k-1}$$

Partial Credit Type

$$P_{jk}(\theta) \;=\; \frac{\exp\left(a_j\, k\, \theta - \sum_{h=1}^{k}\; h\, b_{jh}\right)}{\sum_{k'=0}^{m_j}\exp\left(a_j\, k'\, \theta - \sum_{h=1}^{k'}\; h\, b_{jh}\right)}$$

2 順序カテゴリの取り扱い



3. 重み付き得点からの θ の推定

重み付き得点

$$X = \sum_{j=1}^{p} w_j \sum_{k=0}^{m_j} v_{jk}U_{jk}, \quad 0 \leq X \leq \sum_{j=1}^{p} w_j v_{jm_j}$$

から θ を推定することを考える。
このためには X を与えた時の θ の事後分布

$$\Pr(\Theta = \theta | X = x) \;=\; \frac{\Pr(X = x|\Theta = \theta)\Pr(\Theta = \theta)}{\sum_\theta \Pr(X = x|\Theta = \theta)\Pr(\Theta = \theta)}$$

を計算し、その事後平均等を用いれば良い。

ここで $\Pr(X = x|\Theta = \theta)$ は、

$$\boldsymbol{U}_j|\theta \sim MN((v_{j0}, v_{j1}, v_{j2}, \ldots, v_{j,m_j}), \boldsymbol{P}_j(\theta))$$

であり、また、項目反応は局所独立であるから、
p 個の独立な数値化された多項分布の w_j による重み付きの分布となる。

3. 重み付き得点からの θ の推定

最適な重みは何か？

十分統計量で重みを付ければ良い！

NRM の十分統計量は

$$z = \sum_{j=1}^{p}\sum_{k=0}^{m_j} u_{kj} a_{kj}$$

なぜなら

$$\sum_{k=0}^{m_j} u_{kj} = 1$$

$$\begin{aligned} logL &= \sum_{j=1}^{p}\sum_{q=0}^{m_j} u_{kj} \log P_{kj}(\theta) \\ &= \sum_{j=1}^{p}\sum_{q=0}^{m_j} u_{kj}\left(a_{kj}\theta + c_{kj} - \log\sum_{q=0}^{m_j}\exp\left(a_{qj}\theta + c_{qj}\right)\right) \\ &= -\sum_{j=1}^{p}\log\left(\sum_{q=0}^{m_j}\exp(a_{qj}\theta + c_{qj})\right)\sum_{k=0}^{m_j} u_{kj} + \sum_{j=1}^{p}\sum_{k=0}^{m_j} c_{kj}u_{kj} + \theta\sum_{j=1}^{p}\sum_{k=0}^{m_j} u_{kj}a_{kj} \\ &= -\sum_{j=1}^{p}\log\left(\sum_{q=0}^{m_j}\exp(a_{qj}\theta + c_{qj})\right) + \sum_{j=1}^{p}\sum_{k=0}^{m_j} c_{kj}u_{kj} + \theta\sum_{j=1}^{p}\sum_{k=0}^{m_j} u_{kj}a_{kj} \end{aligned}$$

3. 重み付き得点からの θ の推定

最適重み

重み付き得点

$$\begin{aligned} X &= \sum_{j=1}^{p} w_j \sum_{k=0}^{m_j} v_{jk}U_{jk}, \quad 0 \le X \le \sum_{j=1}^{p} w_j v_{jm_j} \\ &= \sum_{j=1}^{p}\sum_{k=0}^{m_j} v_{jk}U_{jk}, \quad 0 \le X \le \sum_{j=1}^{p} v_{jm_j} \end{aligned}$$

2PLM

$$v_{j0} = 0, \quad v_{j1} \propto a_j$$

Generalized Partial Credit Model

$$v_{jk} \propto a_j \times k, \quad k = 0, 1, \ldots, m_j$$

重みは、加算的また乗算的不定性を持つ。

3. 重み付き得点からの θ の推定

3PLM と Graded Response Model には十分統計量がみあたらない。

重み付き得点で θ を推定する時の情報量の期待値を
最大にする重み → 大局的最適重み

重みが θ に依存しても良いのなら
局所的最適重み basic function (Samejima, 1969, p24)

$$v_{jk} = v_{jk}(\theta) = \frac{P'_{jk}(\theta)}{P_{jk}(\theta)}$$

4. 独立に推定された項目パラメタの等化

項目反応モデル

3PLM: $m_j = 1$

$$\begin{aligned} P_{j1}(\theta) &= c_j + \frac{1 - c_j}{1 + \exp\left(-1.7a_j(\theta - b_j)\right)} \\ P_{j0}(\theta) &= 1 - P_{j1}(\theta) \end{aligned}$$

Graded Response Model

$$\begin{aligned} P_{jk}(\theta) &= P^{+}_{j,k}(\theta) - P^{+}_{j,k+1}(\theta), \quad k = 0, 1, 2, \ldots, m_j \\ P^{+}_{j,k}(\theta) &= \frac{1}{1 + \exp\left(-1.7a_j(\theta - b_{jk})\right)} \\ & P^{+}_{j0}(\theta) = 1, \quad P^{+}_{j,m_j+1}(\theta) = 0 \end{aligned}$$

Generalized Partial Credit Model

$$P_{jk}(\theta) = \frac{\exp 1.7\left(\sum_{h=0}^{k}(a_j(\theta - b_{jh}))\right)}{\sum_{k'=0}^{m_j}\exp 1.7\left(\sum_{h=0}^{k'}(a_j(\theta - b_{jh}))\right)}, \quad k = 0, 1, 2, \ldots, m_j$$

4. 独立に推定された項目パラメタの等化

共通尺度上のパラメタの値

$$\boldsymbol{\xi}_j = \{a_j, b_j\}, j = 1, 2, \ldots, p, g = 1, 2, \ldots, G$$

第 g 番目の尺度上のパラメタの値

$$\boldsymbol{\xi}_j^{(g)} = \{a_j^{(g)}, b_j^{(g)}\}, j = 1, 2, \ldots, p, g = 1, 2, \ldots, G$$

個別尺度と共通尺度： 原点と単位の不定性

$$\theta^{(g)} = u_g + v_g\theta, \quad g = 1, 2, \ldots, G$$

パラメタ値の関係

$$a_j^{(g)} = \frac{1}{v_g} a_j \quad \text{and} \quad b_j^{(g)} = u_g + v_g b_j,\ j = 1, 2, \ldots, p$$

4. 独立に推定された項目パラメタの等化

b パラメタにのみ着目

$$b_j^{(g)} = u_g + v_g b_j, \quad j = 1, 2, \ldots, p$$

第 g 番目の尺度上のパラメタの推定値

$$\widehat{\boldsymbol{\xi}_j^{(g)}} = \{\widehat{a_j^{(g)}}, \widehat{b_j^{(g)}}\}, j = 1, 2, \ldots, p, \quad g = 1, 2, \ldots, G$$

因子分析

$$\widehat{b_j^{(g)}} = u_g + v_g b_j + e_j$$

4. 独立に推定された項目パラメタの等化

因子分析： cal-a

$$\widehat{b_j^{(g)}} = u_g + v_g b_j + e_j$$

ICRF メトリックの利用： cal-r

$$\begin{aligned} P_{jk}(\theta|\widehat{\boldsymbol{\xi}_j^{(g)}}) &= P_{jk}(\theta|\boldsymbol{\xi}_j^{(g)}) + e_j \\ &= P_{jk}(q_g + r_g\theta|\boldsymbol{\xi}_j) + e_j \end{aligned}$$

ただし

$$q_g = -\frac{u_g}{v_g}, \quad r_g = \frac{1}{v_g}$$

4. 独立に推定された項目パラメタの等化

● 最小２乗解

最小２乗基準： cal-a

$$RSS_A = \sum_{g=1}^{G} \sum_{j\in g} \sum_{k=1}^{m_j} \left(b_{jk}^{\hat{(g)}} - (u_g + v_g b_{jk})\right)^2$$

最小２乗基準： cal-r

$$RSS_R = \sum_{g=1}^{G} \sum_{j\in g} \sum_{k=1}^{m_j} \sum_{q=1}^{Q} \left(P_{jk}(\theta_q|\widehat{\boldsymbol{\xi}_j^{(g)}}) - P_{jk}(q_g + r_g\theta_q|\boldsymbol{\xi}_j)\right)^2 h_g(\theta_q|\mu_g, \sigma_g)$$

$$\mu_g = 0, \quad \sigma_g = 1, \quad g = 1, 2, \ldots, G$$

ご静聴、有難うございました。

mayekawa@nifty.com
www.ms.hum.titech.ac.jp