

DOI: 10.13718/j.cnki.xdzk.2015.05.012

多维项目反应理论二级评分模型的参数估计^①

杜文久¹, 谢荣华¹, 李洪波²

1. 西南大学 数学与统计学院, 重庆 400715; 2. 重庆市教育考试院, 重庆 401147

摘要: 利用 DSY 算法开发了多维二级评分模型的参数估计程序, 推广了该类评分模型的参数估计方法, 比较了 MCMC 和 DSY 算法, 说明了 DSY 算法能有效估计该类型评分模型的参数, 提出了一些有待研究的问题.

关键词: 多维项目反应模型; 多维二级评分模型; DSY 算法; 参数估计精度

中图分类号: O212 **文献标志码:** A **文章编号:** 1673-9868(2015)05-0078-04

在一定假设之下, 根据多维项目反应理论(multidimensional item response theory, MIRT)用一个数学函数去刻画被试在项目上可观察的作答表现与其不可观察的特质水平之间的函数关系, 这就是多维项目反应模型^[1]. 其中, 正态累积形式和逻辑斯蒂形式的多维二级评分模型应用最为广泛. 现有的多维二级评分模型参数估计程序都存在一定局限性. 而在我国, 多维二级评分模型的参数估计程序不够完善. 除了我国 MIRT 的发展相对滞后以外, 缺少一种有效的算法也是一个重要原因. 为此, 本文试图给出一种更为简便快捷的算法—DSY 算法^[2]来估计 MIRT 模型的参数, 期望本研究能对国内 MIRT 的发展起到一定推动作用.

1 模拟研究

为简便计算, 本文仅讨论三参数逻辑斯蒂形式的多维二级评分模型的参数估计. 假设被试的潜在特质为二维, 取猜测系数 0.25. 对于被试的潜在特质多于二维的情形本研究仍然适用. 采用蒙特卡洛(Monte Carlo, MC)模拟得到实验数据. 具体做法如下:

1) 随机产生各参数的真值. 被试总数和项目总数分别记为 N, m . 第 i 个被试的能力参数为 θ_i , 两个分量分别为 $\theta_{i1}, \theta_{i2} (i=1, \dots, n)$. 按照各参数的产生背景、意义及国内外研究者的使用共识来确定各参数候选值的分布^[3]. 从二维正态分布 $N(\mathbf{0}, \mathbf{\Sigma})$ 中随机产生 N 对随机数作为被试的能力真值, 其中:

$$\mathbf{0} = (0, 0), \mathbf{\Sigma} = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}$$

由于能力参数相关系数可以取 0 到 1 之间的任意实数, 在此不妨取为 0.4. 第 $a^{(k)}, b^{(k)}$ 个项目的区分度参数记为 $N(a^{(k)}, 0.2)$, 由于 $N(a^{(k)}, 0.2)$ 非负, 不妨设其分量 $a^{(k)}$ 分别从对数标准正态分布 $N(0, 0.5)$ 中随机产生, 并且假设 $a^{(k)}$ 相互独立; 第 j 个项目截距参数记为 d_j , 不妨设其从分布 $U(-1, 1)$ 中随机产生.

2) 计算得分概率. 第 i 个被试在第 j 个项目得 1 分的概率记为 p_{ij} , 即

$$p_{ij} = c_j + (1 - c_j) \frac{1}{1 + e^{-(a_{j1}\theta_{i1} + a_{j2}\theta_{i2} + d_j)}}$$

3) 生成反应矩阵. 从均匀分布($k=0, 1, 2, \dots, n$)中随机产生一个随机数 u . 如果 $u \leq p_{ij}$, 则将该被试

① 收稿日期: 2014-06-03
基金项目: 国家自然科学基金项目(11226228); 重庆市教育科学规划课题(2011-KS-035).
作者简介: 杜文久(1956-), 男, 重庆北碚人, 副教授, 主要从事项目反应理论研究.

在项目 a, b 上的作答记为 1 分; 否则, 记为 0 分, 即

$$X_{ij} = \begin{cases} 1 & u \leq p_{ij} \\ 0 & u > p_{ij} \end{cases}$$

由此得到被试在项目上的反应得分矩阵 $\mathbf{X} = (X_{ij})_{N \times m}$.

4) 误差判断指标. 设 $\mathbf{X} = (x_1, x_2, \dots, x_m)$, 以平均绝对离差 $\text{abse}(\mathbf{X}) = \frac{1}{m} \sum_{i=1}^m |x_i - \hat{x}_i|$ 为误差大小的指标, 其中: x_i 为参数真值, $\hat{x}_i$ 为参数估计值. 记 $\text{abse}(\theta_1), \text{abse}(\theta_2)$ 分别表示 N 个被试第一、二个能力维度的平均绝对离差, $\text{abse}(a_1), \text{abse}(a_2), \text{abse}(d)$ 分别表示 m 个项目 a 中 a_1, a_2 和类别截距参数的平均绝对离差.

为保证结果的准确, 每个研究结果都是取 10 次重复结果的平均值作为最后的估计值.

1.1 已知项目参数估计能力参数

设 X_{ij} 表示第 i 个被试在第 j 个项目上的反应得分 ($j = 1, 2, \dots, m$), 项目参数已知. 根据被试在测验项目集上的反应得似然函数

$$L = \prod_{j=1}^m p_{ij}^{X_{ij}} (1 - p_{ij})^{1-X_{ij}} \quad (1)$$

对(1)式两边取对数, 得

$$\ln L = \sum_{j=1}^m [X_{ij} \ln p_{ij} + (1 - X_{ij}) \ln (1 - p_{ij})]$$

用 DSY 算法求 θ_i 的极大似然估计的操作过程如下:

① 给定 $\theta_{i1}, \theta_{i2} (i = 1, \dots, N)$ 的初值: $\theta_{i1}^{(k)} = 0, \theta_{i2}^{(k)} = 0$.

② 从分布 $N(\theta_{i1}^{(k)}, 0.2)$ 中随机产生 $\theta_{i1}^{(k)}$ 的下一个候选值 $\theta_{i1}^{(*)}$, 若

$$\ln L(X, \theta_{i1}^{(*)}, \theta_{i2}^{(k)}) > \ln L(X, \theta_{i1}^{(k)}, \theta_{i2}^{(k)})$$

则 $\theta_{i1}^{(k+1)} = \theta_{i1}^{(*)}$, 否则 $\theta_{i1}^{(k+1)} = \theta_{i1}^{(k)}$.

③ 从分布 $N(\theta_{i2}^{(k)}, 0.2)$ 中随机产生 $\theta_{i2}^{(k)}$ 的下一个候选值 $\theta_{i2}^{(*)}$, 若

$$\ln L(X, \theta_{i1}^{(k+1)}, \theta_{i2}^{(*)}) > \ln L(X, \theta_{i1}^{(k+1)}, \theta_{i2}^{(k)})$$

则 $\theta_{i2}^{(k+1)} = \theta_{i2}^{(*)}$, 否则 $\theta_{i2}^{(k+1)} = \theta_{i2}^{(k)}$.

④ 重复步骤 ②, ③, 产生 $\theta_{i1}^{(k)}, \theta_{i2}^{(k)}$ 的一个迭代点列: $\theta_{i1}^{(0)}, \theta_{i2}^{(0)}, \theta_{i1}^{(1)}, \theta_{i2}^{(1)}, \theta_{i1}^{(2)}, \theta_{i2}^{(2)}, \dots, \theta_{i1}^{(n)}, \theta_{i2}^{(n)}$, 其中: $k = 0, 1, 2, \dots, n$; 迭代次数 n 取 500; 取最后一对估计值 $(\theta_{i1}^{(n)}, \theta_{i2}^{(n)})$ 作为未知参数 $\theta_{i1}, \theta_{i2} (i = 1, \dots, N)$ 的估计值. 重复 N 次, 就得到 N 个被试的能力参数估计误差.

取被试数 $N = 1\ 000$, 项目数分别取 100, 200, 500, 利用 DSY 算法可求出所有被试能力参数的估计值, 估计误差见表 1.

表 1 能力参数估计误差

m	$\text{abse}(\theta_1)$	$\text{abse}(\theta_2)$
100	0.442 9	0.479 1
200	0.326 7	0.333 2
500	0.200 4	0.212 4

由表 1 可知, 随着项目数的增大, 被试能力参数的估计精度也随之提高. 这与极大似然估计的相合性是相符的.

1.2 已知能力参数估计项目参数

在已知能力参数估计项目参数时, 似然函数为

$$L = \prod_{i=1}^N p_{ij}^{X_{ij}} (1 - p_{ij})^{1-X_{ij}} \quad (2)$$

对(2)式两边取对数, 得

$$\ln L = \sum_{i=1}^N [X_{ij} \ln p_{ij} + (1 - X_{ij}) \ln (1 - p_{ij})]$$

设项目数 m 取 100, 被试数 N 分别取 600, 1 000, 2 000, 利用 DSY 算法可得到结果如表 2.

表 2 项目参数估计误差

N	$\text{abse}(a_1)$	$\text{abse}(a_2)$	$\text{abse}(d)$
600	0.145 1	0.137 4	0.119 7
1 000	0.113 1	0.114 1	0.091 1
2 000	0.080 3	0.092 3	0.061 0

由表 2 看到, 随着被试样本容量的增大, 项目参数估计精度也随之提高, 当被试样本容量为 2 000 时, 所有参数的平均绝对离差均小于 0.1.

1.3 项目参数和能力参数的同时估计

如果能力参数和项目参数都未知, 需要同时对能力和项目参数进行估计, 在这种情况下可采用两阶段似然估计法估计项目参数与能力参数, 具体操作过程如下:

- ① 设被试 i 在测验中的得分 $X_i = \sum_{j=1}^m X_{ij}$, 则 N 个被试的平均分 $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$, 标准差 $S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}}$, 将 X 标准化, 得被试 i 在测验中的 Z 分数 $Z_i = \frac{X_i - \bar{X}}{S}$.
- ② 设第 i 个被试的能力初值为 $\theta_i = (Z_i, Z_i) (i = 1, \dots, N)$, 利用 1.2 节讨论的方法估计项目参数.
- ③ 将估计得到的项目参数作为已知, 利用 1.1 节讨论的方法重新估计能力参数.
- ④ 将估计得到的能力参数标准化, 重新估计项目参数.
- ⑤ 重复 ③, ④, 直到参数估计值不再改变为止.

通常执行完 ①, ②, ③, ④ 步骤后, 再重复 ③, ④ 步骤 5 次, 就可以得到较为理想的能力参数和项目参数的估计值. 迭代的次数越多, 估计精度也就越高, 同时耗时也较长.

表 3 是项目参数与能力参数同时估计的计算结果.

表 3 同时估计时各参数估计误差

(N, m)	$\text{abse}(a_1)$	$\text{abse}(a_2)$	$\text{abse}(d)$	$\text{abse}(\theta_1)$	$\text{abse}(\theta_1)$	计算时间/s
(2 000, 200)	0.417 4	0.370 1	0.284 6	0.460 4	0.463 4	8 425.729 9
(3 000, 300)	0.176 6	0.153 6	0.147 4	0.257 9	0.261 2	18 538.415 9
(5 000, 500)	0.119 0	0.110 0	0.094 5	0.202 0	0.210 2	53 067.811 8

表 3 表明, 随着项目数和被试数的同时增大, 估计误差均随之减小, 估计精度提高, 这符合对估计方法的一般要求.

2 DSY 算法与 MCMC 算法的比较研究

在二维条件下, 以已知能力参数估计项目参数为例说明. 设初值从对数标准正态分布中随机产生. 取链长为 15 000, 取第 5 000 次到 15 000 次的平均值作为参数估计值^[5], $m = 100$. MCMC 算法^[4]与 DSY 算法的比较结果见表 4.

表 4 MCMC 算法与 DSY 算法的比较研究结果

被试数 N	算法	$\text{abse}(a_1)$	$\text{abse}(a_2)$	$\text{abse}(d)$	运算时间/s
200	MCMC	0.551 1	0.527 8	0.207 3	1 455.871 2
	DSY	0.491 0	0.474 9	0.271 9	48.768 1
500	MCMC	0.312 6	0.319 1	0.156 5	2 887.204 9
	DSY	0.297 6	0.279 4	0.144 5	133.169 1
1 000	MCMC	0.236 6	0.222 7	0.100 3	5 603.197 8
	DSY	0.199 5	0.194 3	0.093 5	256.618 2

由表 4 可以看到,随着被试数的增加,2 种算法算得所有参数的估计误差都相应地减小.从计算精度上来看,在小样本条件下 2 种算法的估计精度大致相当,甚至 MCMC 算法会更好,而在大样本条件下 DSY 明显优于 MCMC 算法;从计算时间上来看,在同类比较中,MCMC 算法计算时间至少是 DSY 的 20 多倍.

当维数增大时,DSY 算法在估计参数时依然是非常有效的,只要样本容量足够大,DSY 算法对参数估计的精度都是比较高的,只是程序运行的时间会相应增加.

另外 2 种参数估计也类似可以得到上述结论.

3 结 论

通过本文探讨可以得到两点结论:第一,DSY 是一种十分可行的算法.DSY 算法比 MCMC 算法具有时间和精度上的优越性;第二,多维二级至多维多级评分模型的参数估计是可以很好实现的.

MIRT 的发展对指导心理测验的编制及开发具有重要意义.但是,要对 MIRT 的参数进行精确、有效的估计,还有很多问题需要解决,比如正态累积形式和逻辑斯蒂形式的多维项目反应模型有一共同的缺点:其具有的补偿性特点与实际情况有时不太相符.所以,如何改进模型,这也是下一步要研究的内容.

参考文献:

- [1] 涂冬波,蔡 艳,戴海琦.多维项目反应理论:参数估计及其在心理测验中的应用[J].心理学报,2011,43(11): 1329—1340.
- [2] 杜文久,孙胜亮,原 坤.改进的 MCMC 算法——DSY 算法及其在估计 IRT 模型参数中的应用[J].心理科学,2013, 36(3): 734—738.
- [3] 罗照盛,漆书青,戴海崎.项目反应理论等级记分模型参数估计的实现[J].心理学报,2003,35(4): 555—558.
- [4] BOLT D M, LALL V F. Estimation of Compensatory and Noncompensatory Multidimensional Item Response Models Using Markov Chain Monte Carlo [J]. Applied Psychological Measurement, 2003, 27(6): 395—414.
- [5] ZHANG B, STONE C A. Evaluating Item Fit for Multidimensional Item Response Model [J]. Educational and Psychological Measurement, 2008, 68(2): 181—196.

Parameter Estimation of MIRT Models for Dichotomous Data

DU Wen-jiu¹, XIE Rong-hua¹, LI Hong-bo²

1. School of Mathematics and Statistics, Southwest University, Chongqing 400715, China;

2. Chongqing Academy of Education and Test, Chongqing 401147, China

Abstract: In a study reported in the present paper, a procedure for estimating the parameters of MIRT models for dichotomous data is developed with the DSY algorithm, the parameter estimation methods of MIRT models for dichotomous data are generalized, and the MCMC and the DSY algorithms are compared. The results show that the DSY algorithm can effectively estimate parameters of the MIRT models for dichotomous data. Finally, a few problems to be further studied are outlined.

Key words: multidimensional item response theory (MIRT); MIRT model for dichotomous data; DSY algorithm; precision for parameter estimation

责任编辑 张 构

