

多维项目反应理论等级反应模型

杜文久 肖涵敏

(西南大学数学与统计学院, 重庆 400715)

摘要 基于因子分析和单维项目反应理论的多维项目反应理论是测量理论的新发展方向之一。但是, 多维项目反应理论仍处于不成熟的发展阶段, 多数研究也只是以二级评分为主。本文首先介绍了逻辑斯蒂形式的多维等级反应模型, 并以二维等级反应模型为例, 分析了模型的数学函数图像及其性质。然后, 推导出了多维等级反应模型的项目信息函数, 并结合实例进行了讨论。进一步地, 本文阐述了使用联合极大似然估计和马尔科夫链蒙特卡洛方法估计多维等级反应模型参数的思想。最后, 指出了一些有待研究的问题。

关键词 多维项目反应理论; 多维等级反应模型; 项目信息函数; 参数估计

分类号 B841

1 问题的提出

自 20 世纪 60 年代至今, 测量理论经历了从经典测量理论(Classical Test Theory, CTT)到项目反应理论(Item Response Theory, IRT)的巨大变革, IRT 克服了 CTT 的种种局限, 实现了被试与项目关系间的模型化、参数化。特别是随着计算机技术的发展, 使得 IRT 得以迅速推广应用, 从而逐步取代了 CTT 而成为现代心理测量学研究的主要内容之一。IRT 通常假设测验只测量被试的一种能力水平, 然而这种单维性假设与许多测验的实际不相符, 因为测验数据的多维性与被试完成一项测验任务时需要多种能力共同支配是相符的。显然, 如果用单维模型去拟合多维数据, 势必会增大测量误差, 对被试的能力水平做出不准确的结论。因此, 建立被试反应与多种能力及项目特征之间的多维项目反应理论(Multidimensional Item Response Theory, MIRT)模型成为了 IRT 近 40 年来研究较为活跃的领域之一。

MIRT 是在因子分析和 IRT 的双重背景之下而发展起来的(Reckase, 2009), 多维项目反应模型的提出就是为了更好地对被试在完成一项测验任务时需要的多种能力、项目特征与答对概率之间的关系进行模型化(康春花, 辛涛, 2010)。IRT 研究者在

上世纪 70 年代就开始探索 MIRT 模型, Mulaik 于 1972 年率先提出了一个二级评分的 MIRT 模型, 之后, Sympon (1978)、Whitely (1981)以及 Mckinley 和 Reckase (1982)等也相继提出了各自的模型。然而, 这些早期的模型以及之后的相关研究都是针对二级评分试题的, 对于多级评分的 MIRT 模型, 直到 1993 年才由 Muraki 和 Carlson (1993)提出正态累积形式的多维等级反应模型(Multidimensional Grade Response Model, MGRM), 其它一些多级模型更是在最近几年才得以提出(Yao & Schwarz, 2006; Ferrando, 2009)。众所周知, 多级评分试题由于可以提供更多关于被试的信息而被普遍地采用, 多级评分模型特别是等级反应模型(Grade Response Model, GRM)是当前测验发展的新方向之一。鉴于 IRT 向多维发展的趋势以及等级反应模型的重要地位, 结合国内多数理论和应用研究都集中在单维 IRT 上的现状, 本文将对具有逻辑斯蒂形式的 MGRM 及其信息函数和参数估计方面进行讨论, 以期对国内 MIRT 的研究起到一定的促进作用。

2 MGRM 介绍

Muraki 和 Carlson (1993)给出了正态累积形式的 MGRM。假设被试 j 在项目 i 上正确反应共有 k_i 个等级, 并且获得第 k 个等级必需要求先达到第

$k-1$ 等级, 则被试获得第 k 个等级以上(包括第 k 个等级)的概率为

$$P_{ijk}^* = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_i \theta_j + d_{ik}} e^{-\frac{t^2}{2}} dt \quad (1)$$

并且 $P_{ij0}^* = 1$, $P_{ij,k+1}^* = 0$ 。被试 j 恰好获得第 k 个等级的概率为

$$P_{ijk} = P_{ijk}^* - P_{ij,k+1}^* = \frac{1}{\sqrt{2\pi}} \int_{a_i \theta_j + d_{i,k+1}}^{a_i \theta_j + d_{ik}} e^{-\frac{t^2}{2}} dt \quad (2)$$

这就是 Muraki 和 Carlson 提出的 MGRM。其中, a_i 为项目 i 的区分度参数向量, θ_j 为被试 j 的能力参数向量, d_{ik} 是一个标量, 它被定义为被试达到第 k 个等级的类别参数(item-category parameter)。从模型中可以看出, d_{ik} 前面的符号为正, 所以它和单维 GRM 的等级难度参数的意义正好相反, 也就是说在 MGRM 中, 等级越高, d_{ik} 的值越小。当 $k=0$ 时, $d_{i0}=\infty$, 当 $k=k_i+1$, $d_{i,k+1}=-\infty$, 所以 d_{ik} 是被试达到第 k 个等级的“易度参数”(Reckase, 2009)。

由于积分运算的复杂性以及正态函数与逻辑斯蒂函数的相似性, 可得到逻辑斯蒂形式的 MGRM, 即

$$P_{ijk}^* = \frac{e^{a_i \theta_j + d_{ik}}}{1 + e^{a_i \theta_j + d_{ik}}}, \quad P_{ijk} = P_{ijk}^* - P_{ij,k+1}^* \quad (3)$$

和单维 GRM 类似, 称(3)式中的概率曲面 P_{ijk}^* 为项目特征曲面(Item Characteristic Surface, ICS), P_{ijk} 为类别特征曲面(Category Characteristic Surface, CCS)。

进一步, 由(3)可得被试 j 在项目 i 上的期望分数为

$$E(u_j | \theta_j) = \sum_{k=0}^{k_i} k P_{ijk} = \sum_{k=1}^{k_i} P_{ijk}^* = \sum_{k=1}^{k_i} \frac{e^{a_i \theta_j + d_{ik}}}{1 + e^{a_i \theta_j + d_{ik}}} \quad (4)$$

在二维能力参数场合下, 取 $a_i = (0.7, 1.2)'$, $d_{i1}=1.5$, $d_{i2}=0.5$, $d_{i3}=-2$ 时, 采用数学软件 Matlab 2007 可得 ICS 和 CCS 如图 1、图 2 所示。

从图 1 可以看出, 当固定其中一个能力维度时, 得某个分数以上的概率值都随另一能力维度值的增加而增大。当其中一个能力值较低时, 可以通过另一个能力值的补偿而得到相同的概率值, 这就是所谓的补偿性模型(Compensatory MIRT Model, MIRT-C)(康春花, 辛涛, 2010)。图 2 反映了得各分数的概率的变化趋势。当两个能力维度值都接近-4 时, 得 0 分的概率最大, 这一概率值接近 1。随着两个能力维度值都增大, 得 0 分的概率不断减小, 而得其他分数的概率增大。当两个能力维度值都增大

到接近 4 时, 得 3 分的概率最大。

在上述参数之下, 得到期望分数曲面如图 3 所示。

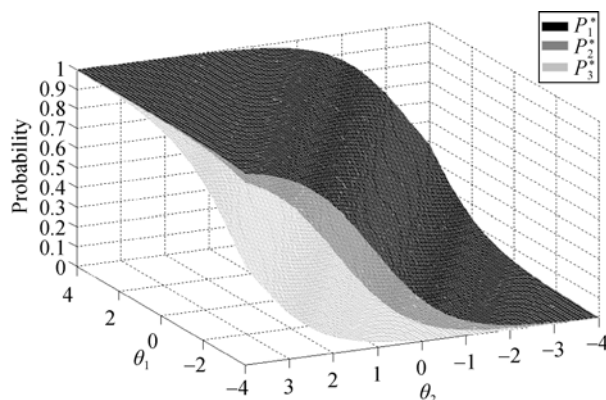


图 1 项目特征曲面

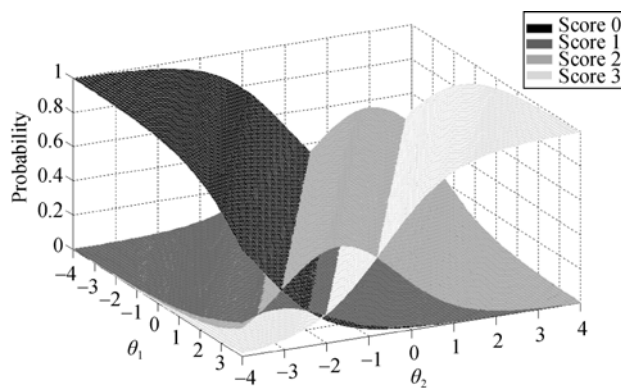


图 2 类别特征曲面

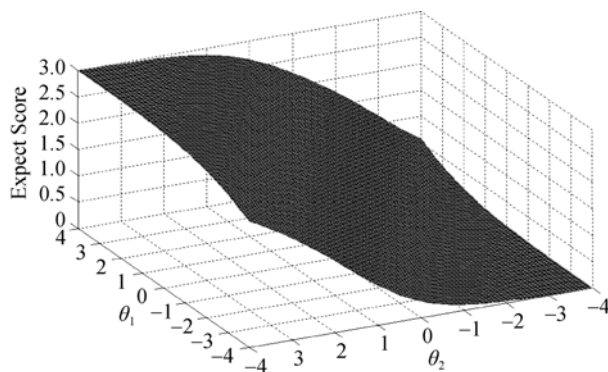


图 3 期望分数曲面

可见, 期望分数的变化范围为 0 分到 3 分之间, 且当两个能力维度都在 4 附近时, 期望分数接近最大值 3 分。

3 MGRM 的项目信息函数

信息函数作为 IRT 中的一个重要概念, 在进行

项目和测验分析以及指导测验编制中都有非常重要的应用价值,它更是当今计算机自适应测验(Computerized Adaptive Test, CAT)研究的焦点(罗照盛, 欧阳雪莲, 漆书青, 戴海琦, 丁树良, 2008)。

MIRT 的项目信息函数可以由单维场合下进行推广而得到,但是要复杂得多。Reckase (2009)介绍了 MIRT 中二级评分模型的项目信息函数,但是并没有涉及有关多级的讨论。二级评分的项目信息函数被定义为

$$I_{i\alpha}(\theta) = \frac{[\nabla_{\alpha} P_i(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (5)$$

其中, α 为方向向量满足 $\sum_{v=1}^m \cos \alpha_v^2 = 1$, m 表示能力维度数, ∇_{α} 为沿 α 的方向导数或者梯度, 且

$$\nabla_{\alpha} P_i(\theta) = \frac{\partial P_i(\theta)}{\partial \theta_1} \cos \alpha_1 + \frac{\partial P_i(\theta)}{\partial \theta_2} \cos \alpha_2 + \dots + \frac{\partial P_i(\theta)}{\partial \theta_m} \cos \alpha_m \quad (6)$$

若 $P_i(\theta) = \frac{e^{a_i\theta+d_i}}{1+e^{a_i\theta+d_i}}$, 则

$$\nabla_{\alpha} P_i(\theta) = P_i(\theta)Q_i(\theta) \sum_{v=1}^m a_{iv} \cos \alpha_v \quad (7)$$

把(7)代入(5)中, 则项目信息函数为

$$I_{i\alpha}(\theta) = P_i(\theta)Q_i(\theta) \left(\sum_{v=1}^m a_{iv} \cos \alpha_v \right)^2 \quad (8)$$

现在, 我们把单维 GRM 的项目信息函数推广到 MGRM。在单维情形下, GRM 的项目信息函数

为 $I_i(\theta) = \sum_{k=0}^{k_i} \frac{(P'_{ik})^2}{P_{ik}}$, 结合上述二级评分的讨论, 得

MGRM 的项目信息函数为

$$I_{i\alpha}(\theta) = \sum_{k=0}^{k_i} \frac{[\nabla_{\alpha} P_{ik}(\theta)]^2}{P_{ik}(\theta)} \quad (9)$$

其中 $P_{ik}(\theta)$ 的表达式如(3)所示。由于

$$\frac{\partial P_{ik}}{\partial \theta_v} = \frac{\partial}{\partial \theta_v} (P_{ik}^* - P_{i,k+1}^*) = a_{iv} (P_{ik}^* - P_{i,k+1}^*) (1 - P_{ik}^* - P_{i,k+1}^*) \quad (10)$$

再结合式(6)、(7), 则(9)变为

$$I_{i\alpha}(\theta) = \sum_{k=0}^{k_i} (P_{ik}^* - P_{i,k+1}^*) (1 - P_{ik}^* - P_{i,k+1}^*)^2 \left(\sum_{v=1}^m a_{iv} \cos \alpha_v \right)^2 \quad (11)$$

此式即为 MGRM 的项目信息函数。可以看出, 项目信息量的大小除与能力参数向量和项目参数有关外, 还与方向向量有关。

在前述所设参数之下, 分别取 α_1 等于 0° 、 45° 、 90° 得到项目信息曲面如图 4 所示。从图中可以看出, 在不同的方向向量之下, 3 个信息曲面的形状是一致的。它们都具有双峰形状, 且都在相同的能力值处取极大值或极小值。当两个能力维度值都位于 1 附近时, 项目提供最大的信息量; 当两个能力维度值都在 4 或者是 -4 附近时, 项目提供的信息量接近 0。另外, 除了两个能力维度值都在 4 或者是 -4 外, 在相同的能力处, 当 α_1 为 45° 时提供的信息量最大而 0° 时最小。 α_1 为 0° 、 45° 、 90° , 对应的信息量的最大值约为 0.18、0.55 和 0。

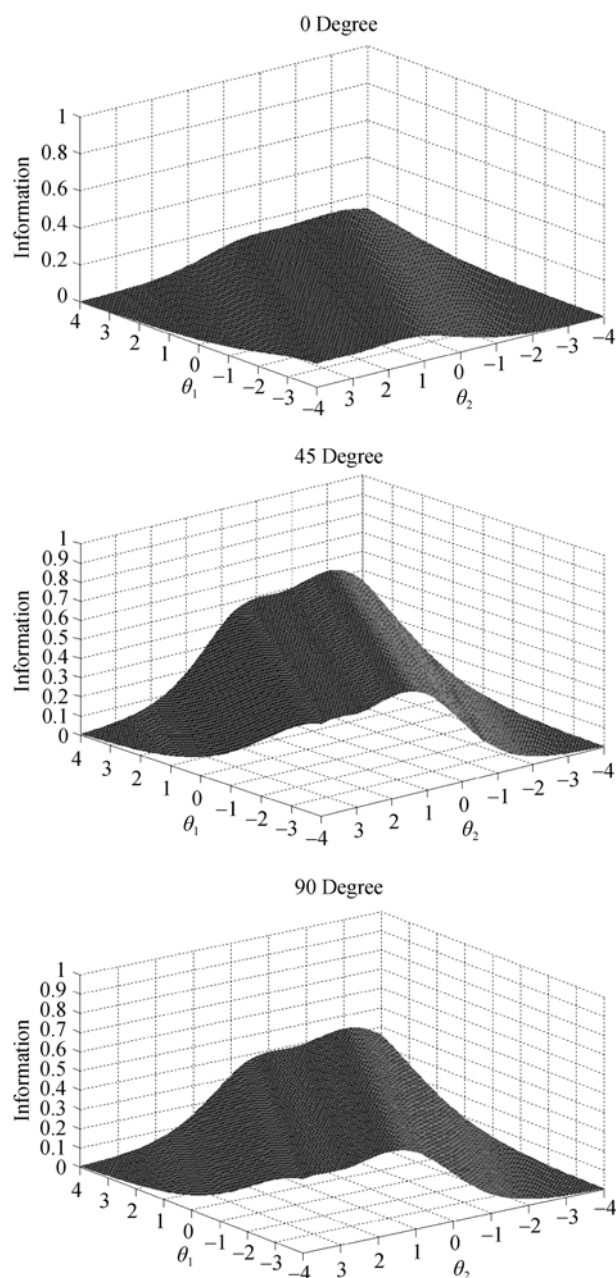


图 4 项目信息曲面

4 MGRM 的参数估计

4.1 JLME 方法

参数估计问题一直以来都被认为是 IRT 的难点之一, 特别是在 MIRT 中, 由于需要估计的参数个数较多, 处理和计算都是相当繁琐的。对于 MGRM, 在被试能力参数和项目参数真值都未知的情况下, 可以应用联合极大似然估计 (Joint Likelihood Maximum Estimation, JLME) 的方法进行参数估计。

4.1.1 MGRM 的似然方程 设 N 个被试参加由 n 个项目组成的测验, 测验反应矩阵为

$$X = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1N} & x_{2N} & \cdots & x_{nN} \end{pmatrix}_{N \times n}$$

x_{ij} 为被试 j 在项目 i 上的反应等级。建立被试 j 在项目 i 上恰好得第 k 个等级的示性函数

$$u_{ijk} = \begin{cases} 1 & x_{ij} = k \\ 0 & x_{ij} \neq k \end{cases}$$

则被试 j 在项目 i 上得某个等级的概率为 $\prod_{k=0}^{k_i} P_{ijk}^{u_{ijk}}$ 。

根据局部独立性假设, N 个被试在这 n 个项目上的联合反应概率为

$$L(\theta, a, d | X) = \prod_{i=1}^n \prod_{j=1}^N \prod_{k=0}^{k_i} P_{ijk}^{u_{ijk}} \quad (12)$$

此式即为似然函数。对(12)两边取对数, 再分别对参数向量 θ, a, d 求偏导数并令其等于 0, 得对数似然方程组为

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_{jv}} = 0 \\ \frac{\partial \ln L}{\partial a_{iv}} = 0 \\ \frac{\partial \ln L}{\partial d_{ik}} = 0 \end{cases} \quad (13)$$

其中, $i=1, 2, \dots, n, j=1, 2, \dots, N, v=1, 2, \dots, m, k=1, 2, \dots, k_i$ 。

不难发现, 方程组(13)共有 $Nm + nm + \sum_{i=1}^n k_i$ 个方程, 所以要求解此方程组是相当复杂的。通常,

$$P(\theta_{jv}^{(t)}, \theta_{jv}^{(t+1)}) = \min \left\{ \frac{L(\theta_{j1}^{(t)}, \theta_{j2}^{(t)}, \dots, \theta_{j(v-1)}^{(t)}, \theta_{jv}^{(t+1)}, \theta_{j(v+1)}^{(t)}, \dots, \theta_{jm}^{(t)}, a^{(t)}, d^{(t)} / X) f_1(\theta_{jv}^{(t+1)})}{L(\theta_{jv}^{(t)}, a^{(t)}, d^{(t)} / X) f_1(\theta_{jv}^{(t)})}, 1 \right\}$$

其中, $f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta_{jv}^{(t)})^2}{2}}$ 。将此概率与随机数 $r \sim U(0, 1)$ 比较, 若大于等于 r , 则接受转移, 否则不转移;

求解方法都是采用 Birnbaum (1968) 提出的两阶段方法 (two-stage procedure), 该方法的主要思想是把 JLME 分为项目参数的条件估计和能力参数的条件估计, 两类参数不断相互校正, 反复迭代, 最终求得参数的稳定值。

4.1.2 MGRM 的似然方程的求解 采用两阶段似然估计法求解方程组(13), 具体过程如下:

1) 给定每个被试的能力参数向量的初始值 $\theta_j^{(0)}$, 代入方程组(14)

$$\begin{cases} \frac{\partial \ln L}{\partial a_{iv}} = 0 \\ \frac{\partial \ln L}{\partial d_{ik}} = 0 \end{cases} \quad (14)$$

解出 $a_i^{(0)}$ 和 $d_i^{(0)}$, 并计算 $L_0 = L(a_i^{(0)}, d_i^{(0)}, \theta_j^{(0)})$;

2) 将求出的 $a_i^{(0)}$ 和 $d_i^{(0)}$ 代入方程组(13)的第一个方程组, 重新计算 $\theta_j^{(1)}$;

3) 将算出的 $\theta_j^{(1)}$ 的每个分量标准化, 代入方程组(14), 重新计算 $a_i^{(1)}, d_i^{(1)}$;

4) 计算 $L_1 = L(a_i^{(1)}, d_i^{(1)}, \theta_j^{(1)})$ 并判断 $|L_1 - L_0| < \varepsilon$ 是否成立。若成立, 则退出; 否则, 置 $L_1 \rightarrow L_0, a_i^{(1)} \rightarrow a_i^{(0)}, d_i^{(1)} \rightarrow d_i^{(0)}, \theta_j^{(1)} \rightarrow \theta_j^{(0)}$, 进行新一轮的计算。

4.2 MCMC 方法

除了采用两阶段似然估计法, 在实践中还可以采用马尔科夫链蒙特卡洛 (Markov Chain Monte Carlo, MCMC) 方法估计参数。MCMC 是目前十分流行的一种参数估计法, 它在估计高维参数时十分有效, MCMC 方法的基本思想是将联合似然函数视为在给定样本 X 后未知参数的后验分布, 并将其看作平稳分布, 然后再采用 M-H 方法构造一个马尔科夫链, 使其收敛于平稳分布, 建议分布均采用对称建议分布。对于 MGRM, 其具体做法如下:

1) 指定迭代周期 M 和迭代次数 M_1 ;

2) 能力参数 θ 从建议分布 $\theta_{jv}^{(t+1)} \sim N(\theta_{jv}^{(t)}, 1)$ ($j=1, 2, \dots, N, v=1, 2, \dots, m, t=0, 1, 2, \dots, M-1$) 中独立取样, θ 从 $\theta_{jv}^{(t)}$ 向 $\theta_{jv}^{(t+1)}$ 转移的接受概率为

3)项目参数 a 从建议分布 $a_{iv}^{(t+1)} \sim N(a_{iv}^{(t)}, 0.25)$ ($i=1, 2, \dots, n$, $v=1, 2, \dots, m$, $t=0, 1, 2, \dots, M-1$) 中独立取样, a 从 $a_{iv}^{(t)}$ 向 $a_{iv}^{(t+1)}$ 转移的接受概率为

$$P(a_i^{(t)}, a_i^{(t+1)}) = \min\left\{\frac{L(\theta^{(t+1)}, a_{i1}^{(t)}, a_{i2}^{(t)}, \dots, a_{i(v-1)}^{(t)}, a_i^{(t+1)}, a_{i(v+1)}^{(t)}, \dots, a_{ik}^{(t)}, d^{(t)} / X) f_2(a_i^{(t+1)})}{L(\theta^{(t+1)}, a^{(t)}, d^{(t)} / X) f_2(a_i^{(t)})}, 1\right\}$$

其中, $f_2(x) = \frac{1}{\sqrt{2\pi} \times 0.5} e^{-\frac{(x-a_{iv}^{(t)})^2}{2 \times 0.5}}$ 。将此概率与随机数 $r \sim U(0, 1)$ 比较, 若大于等于 r , 则接受转移, 否则不转移;

4) 项目参数 d 从建议分布 $d_{ik}^{(t+1)} \sim N(d_{ik}^{(t)}, 1)$ ($i=1, 2, \dots, n$, $k=1, 2, \dots, k_i$, $t=0, 1, 2, \dots, M-1$) 中独立取样, 如果 $d_{i,k+1}^{(t+1)} > d_{ik}^{(t+1)}$ 则重新抽取, 直到 $d_{i,k+1}^{(t+1)} \leq d_{ik}^{(t+1)}$ 为止。 d 从 $d_{ik}^{(t)}$ 向 $d_{ik}^{(t+1)}$ 转移的接受概率为

$$P(d_{ik}^{(t)}, d_{ik}^{(t+1)}) = \min\left\{\frac{L(\theta^{(t+1)}, a^{(t+1)}, d_{i1}^{(t)}, d_{i2}^{(t)}, \dots, d_{i(k-1)}^{(t)}, d_{ik}^{(t+1)}, d_{i(k+1)}^{(t)}, \dots, d_{ik_i}^{(t)} / X) f_3(d_{ik}^{(t+1)})}{L(\theta^{(t+1)}, a^{(t+1)}, d^{(t)} / X) f_3(d_{ik}^{(t)})}, 1\right\}$$

其中, $f_3(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-d_{ik}^{(t)})^2}{2}}$ 。将此概率与随机数 $r \sim U(0, 1)$ 比较, 若大于等于 r , 则接受转移, 否则不转移。

5) 迭代 M 次后, 再迭代 M_1 次, 并令

$$\hat{\theta}_{jv} = \frac{1}{M_1} \sum_{t=1}^{M_1} \theta_{jv}^{(M+t)} (j=1, 2, \dots, N, v=1, 2, \dots, m);$$

$$\hat{a}_{iv} = \frac{1}{M_1} \sum_{t=1}^{M_1} a_{iv}^{(M+t)} (i=1, 2, \dots, n, v=1, 2, \dots, m);$$

$$\hat{d}_{ik} = \frac{1}{M_1} \sum_{t=1}^{M_1} d_{ik}^{(M+t)} (i=1, 2, \dots, n, k=1, 2, \dots, k_i)$$

由此得到的 $\hat{\theta}$ 、 $\hat{a}$ 、 $\hat{d}$ 就可作为 θ, a, d 的估计值。实践中, M 可取为 10000, 而 M_1 可取为 1000。

5 研究展望

毋庸置疑, MIRT 将是 IRT 发展的一个主要方向之一。但是, 相比之下, MIRT 还处于早期发展阶段。特别是 MIRT 下的多级评分模型, 由于模型本身的复杂性, 无论从理论还是技术方面都是一个难题。作为使用最多的多级模型, GRM 在多维场合下还有很多问题需要解决。

第一, 关于 MGRM 的项目信息函数方面的研究。由于在 MGRM 中涉及到了方向向量, 这样, 在项目参数已知的条件下, 项目信息量的大小是方向向量和能力参数向量的函数。因此, 最大信息量以及基于最大信息量的 CAT 选题策略的问题是需要研究的问题。另外, 在固定方向向量之后, 项目难度参数与能力参数的匹配模式如何影响项目信息量的大小? 关于单维 GRM 下的相关研究(戴海崎, 陈德枝, 丁树良, 邓太萍, 2006; 陈平, 丁树良, 林海菁, 周婕, 2006; 罗照盛等, 2008)是否可以推广

或借鉴到 MGRM 中来? 这都是有待研究的。

第二, 关于参数估计以及相应参数估计软件开发方面的研究。国外对 MIRT 的大多数研究都集中在二级记分上, 已开发的多种 MIRT 的参数估计软件, 如 ConQuest (Wu, Adams, & Wilson, 1997)、NOHARM (Fraser, 1998)、TESTFACT (Bock & Schilling, 2003) 和 BMIRT (Yao, 2003), 也都只适应于二级记分的 MIRT 模型。因此, 把已有的单维 GRM 的参数估计方法(如罗照盛, 漆书青, 戴海琦, 丁树良, 2003)移植到 MGRM 中并编制相应的计算机程序是十分有意义的。

参 考 文 献

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Bock, R. D., & Schilling, S. G. (2003). IRT based item factor analysis. In M. du. Toit (ed.), *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Scientific Software International, Lincolnwood, IL.
- Chen, P., Ding, S. L., Lin, H. J., & Zhou, J. (2006). Item selection strategies of computerized adaptive testing based on graded response model (in Chinese). *Acta Psychologica Sinica*, 38(5), 461-467.
- [陈平, 丁树良, 林海菁, 周婕. (2006). 等级反应模型下计算机化自适应测验选题策略. *心理学报*, 38(5), 461-467.]
- Dai, H. Q., Chen, D. Z., Ding, S. L., & Deng, T. P. (2006). The comparison among item selection strategies of CAT with multiple-choice items. *Acta Psychologica Sinica*, 38(5), 778-783.
- [戴海崎, 陈德枝, 丁树良, 邓太萍. (2006). 多级评分题计算机自适应测验选题策略比较. *心理学报*, 38(5), 778-783.]
- Ferrando, P. J. (2009). Difficulty, discrimination, and information indices in the linear factor analysis model for continuous item responses. *Applied Psychological Measurement*, 33, 9-24.

- Fraser, C. (1998). *NOHARM: A Fortran program for fitting unidimensional and multidimensional normal ogive models in latent trait theory*. Armidale, Australia: The University of New England, Center for Behavioral Studies.
- Kang, C. H., & Xiao, T. (2010). New development in test theory: Multidimensional item response theory. *Advances in Psychological Science*, 18, 530–536.
- [康春花, 辛涛. (2010). 测验理论的新发展: 多维项目反应理论. *心理科学进展*, 18, 530–536.]
- Luo, Z. S., Ouyang, X. L., Qi, S. Q., Dai, H. Q., & Ding, S. L. (2008). IRT information function of polytomously scored items under the graded response model. *Acta Psychologica Sinica*, 40(11), 1212–1220.
- [罗照盛, 欧阳雪莲, 漆书青, 戴海崎, 丁树良. (2008). 项目反应理论等级反应模型项目信息量. *心理学报*, 40(11), 1212–1220.]
- Luo, Z. S., Qi, S. Q., Dai, H. Q., & Ding, S. L. (2003). Realization to the IRT multiple category scoring model. *Acta Psychologica Sinica*, 35(4), 555–558.
- [罗照盛, 漆书青, 戴海崎, 丁树良. (2003). 项目反应理论等级记分模型参数估计的实现. *心理学报*, 35(4), 555–558.]
- McKinley, R. L., & Reckase, M. D. (1982). *The use of the general rasch model with multidimensional item response data. (Research Report ON R82-1)*. Iowa City, IA: American College Testing.
- Mulaik, S. A. (1972). *A mathematical investigation of some multidimensional rasch models for psychological tests*. Princeton, NJ: Paper Presented at the Annual Meeting of the Psychometric Society.
- Muraki, E., & Carlson, J. E. (1993). *Full-information factor analysis for polytomous item responses*. Atlanta: Paper Presented at the Annual Meeting of the American Educational Research Association.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.
- Sympson, J. B. (1978). *A model for testing with multidimensional items*. In D. J. Weiss (Ed.), *Proceedings of the 1977 Computerized Adaptive Testing Conference*. Minneapolis: University of Minnesota.
- Whitely, S. E. (1981). Measuring aptitude processes with multicomponent latent trait models. *Journal of Educational Measurement*, 18(2), 67–84.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Generalized item response modeling software*. Victoria, Australia: ACER.
- Yao, L. (2003). *BMIRT: Bayesian multivariate item response theory*. Monterey, CA: CTB/McGraw-Hill.
- Yao, L., & Schwarz, R. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement*, 30, 469–492.

Multidimensional Grade Response Model

DU Wen-Jiu; XIAO Han-Min

(School of Mathematics and Statistics, Southwest University, Chongqing 400715, China)

Abstract

Multidimensional Item Response Theory (MIRT), which is based on factor analysis and unidimensional Item Response Theory (IRT), is one of a new development trend of IRT. It's a fact that MIRT is in an early-developing stage and most studies are mainly concentrated on MIRT models for items with two score categories. With respect to polytomous MIRT models, it's until 1993 that Muraki and Carlson produced a generalization of unidimensional Grade Response Model (GRM) and it uses response functions that have the normal ogive form. Some other models such as multidimensional Generalized Partial Credit Model (MGPCM) and Continuous Response Model (MCRM) are even developed in recent years (Yao & Schwarz, 2006; Ferrando, 2009).

In the paper, a form of logistic Multidimensional Graded Response Model (MGRM) is firstly presented. The graphics, which are plotted by Matlab 2007, and properties for a special case of two-dimensional GRM are demonstrated. Then, base on the definition of item information for a dichotomous MIRT models, the item information function for MGRM is derived and item information for a case of two-dimensional GRM discussed. Moreover, the main ideas of Joint Maximum Likelihood Estimation (JML) and Markov Chain Monte Carlo (MCMC) methods to estimate MGRM parameters are stated. Finally, some significant further researches, which include research of item and test information, developing parameter estimate program for MGRM, are illustrated in the paper.

Key words MIRT; MGRM; item information function; parameter estimate