文章编号:1000-5862(2019)02-0142-05

计算机化自适应测验中能力估计新方法

李 佳,丁树良

(江西师范大学计算机信息工程学院,江西 南昌 330022)

摘要:能力估计的极大似然估计方法(MLE)不能处理全 0 或全 1 的被试反应模式,若事先设置好能力估计的上下界,则会导致能力估计的有效范围缩小的后果;而贝叶斯估计方法需要选择先验分布,先验分布的选择必须很慎重.在原有似然函数的基础上,构建 2 个新的项目,提出了改进的 MLE 方法(NMLE).NMLE 既不需要能力先验分布,也不会缩小能力估计范围,而且可以处理各种反应模式.蒙特卡洛实验结果表明新方法表现良好.

关键词: 贝叶斯众数估计方法;期望后验估计方法;改进的极大似然估计方法;能力估计效率

中图分类号:B 841 文献标志码:A DOI:10.16357/j.cnki.issn1000-5862.2019.02.05

0 引言

计算机化自适应测验(computerized adaptive testing,CAT)具有测验精度高、长度短、成本低、实时反馈考试成绩、题型灵活多样、考试时间灵活等优点,是项目反应理论(Item response theory,IRT)最成功的应用之一,被广泛应用于美国医生护士资格考试、美国研究生入学考试和中国汉语水平考试中[1].国内外学者主要研究 CAT 的选题策略,具有大量的研究成果,但在 CAT 中对能力估计方法的研究较少.而事实上,CAT 自适应选题是建立在对被试能力准确估计的基础上的,这关系到测验结果的准确性、测验的安全性和测验的可信度.因此,能力估计的准确性决定了 CAT 的使用效果[2].

目前,国际上流行的 CAT 能力估计方法主要有极大似然估计法(MLE)[3]、贝叶斯众数估计法(MAP)[4]和贝叶斯期望后验估计法(EAP)[5].这些方法各有特点:MLE 方法的主要缺点是参数估计中需要不断迭代估计以及无法处理被试全对或全错的反应模式,其优点是 MLE 估计是能力参数的充分统计量,是一种渐近无偏的能力估计方法;EAP 方法不需要迭代;但是 EAP 和 MAP 等贝叶斯方法的主要缺点是需要选择能力的先验分布,且当先验分布方差比较小时,估计会收敛到先验分布期望附近,有可能会缩小能力估计的范围.

0.1 极大似然能力估计方法

在 IRT 中假定同一被试对各个项目的作答是相互独立的(局部独立性假设),各个被试的作答模式是相互独立的,则被试反应向量(即为被试作答反应的得分阵)为 $\boldsymbol{U}=(u_{\alpha 1},u_{\alpha 2},\cdots,u_{\alpha m})$,对应的似然函数为 $L(\boldsymbol{U}|\theta_\alpha=\prod_{j=1}^{m}P_{\alpha j}^{u_{\alpha j}}(1-P_{\alpha j})^{1-u_{\alpha j}}$,$u_{\alpha j}$ 表示被试 α 对项目 j 的反应,取值为 0 或 1,分别表示答对或答错该项目,m 为施测项目数.在 IRT 框架下,$P_{\alpha j}$ 可以取不同的形式,表示能力为 θ_α 的被试正确作答项目 j 的概率,比较常见的是 3 参数 Logistic 模型(3PLM):$P_{\alpha j}=c_j+(1-c_j)/(1+\exp(-Da_j(\theta_\alpha-b_j)))$,其中 $D=1.7$.若猜测度 $c_j=0$,则化为双参数 Logistic 模型(2PLM);若 $c_j=0$ 且区分度 $a_j=1$ 则化为单参数 Logistic 模型(1PLM),b_j 表示项目 j 的难度.因为 $L(\boldsymbol{U}|\theta_\alpha)$ 表示得分向量 $\boldsymbol{U}$ 与能力 θ_α 之间的关系,所以称使 $L(\boldsymbol{U}|\hat{\theta}_\alpha)=\max\limits_{\theta_\alpha\in\Theta}L(\boldsymbol{U}|\theta_\alpha)$ 成立的 $\hat{\theta}_\alpha$ 为 θ_α 的极大似然估计值.又因为对数似然函数 $\ln L(\theta_\alpha)=\sum_{j=1}^{m}(u_{\alpha j}\ln P_{\alpha j}+(1-u_{\alpha j})\ln(1-P_{\alpha j}))$ 和似然函数 $L(\theta_\alpha)$ 在同一个 $\hat{\theta}_\alpha$ 处达到最大.求 θ_α 的极大似然估计值 $\hat{\theta}_\alpha$,可令

收稿日期:2018-07-19

基金项目:国家自然科学基金(31500909,31360237,31160203,30860084,11401271)和江西省教育厅科学技术(GJJ170212)资助项目.

作者简介:李 佳(1979-),女,江西南昌人,讲师,主要从事计算机辅助教学和心理测量方面的研究.E-mail:1276676143@qq.com

$$\partial \ln L(\theta_\alpha)/\partial\theta_\alpha = 0, \tag{1}$$

因为(1)式是非线性方程，需使用牛顿-拉夫逊迭代算法对其求解，$\hat{\theta}_\alpha$ 的第 $t+1$ 次估计值为 $[\hat{\theta}_\alpha]_{t+1} = [\hat{\theta}_\alpha]_t - [\partial^2 \ln L/\partial\theta^2]_t^{-1}[\partial \ln L/\partial\theta_\alpha]_t$，直到达到终止条件为止.

3PLM 对数似然函数 1 阶和 2 阶偏导数为

$$\frac{\partial \ln L}{\partial\theta_\alpha} = D\sum_{j=1}^{m}\frac{a_j(P_{\alpha j}-c_j)(u_{\alpha j}-P_{\alpha j})}{P_{\alpha j}(1-c_j)},$$

$$\frac{\partial^2 \ln L}{\partial\theta_\alpha^2} = D^2\sum_{j=1}^{m}\frac{a_j^2(P_{\alpha j}-c_j)(1-P_{\alpha j})(u_{\alpha j}c_j-P_{\alpha j}^2)}{P_{\alpha j}^2(1-c_j)^2}.$$

因为模型参数的似然函数包含了观察数据值所能反应的所有信息，MLE 估计是能力参数的充分统计量，具有渐近一致性和渐近正态性等优良特性[6]. 在理想条件 CAT 下，当测验较长时，MLE 是一种渐近无偏的能力估计方法[7]；但是 MLE 方法有一个明显的缺点，即当被试作答全对或者全错时似然方程会出现没有有限解的情况. 为了解决这个问题，通常人为设定一个最小和最大的能力估计值对 MLE 估计值的界限加以约束[8]，这是一种强行拉回的处理方式，会破坏 MLE 计算过程中的不连续性，从而缩小能力估计的有效范围. 虽然如此，但因为 MLE 对被试能力分布不作要求，带界限的 MLE 方法(MLET)还是被广泛地应用于 CAT 实测中.

0.2　贝叶斯众数估计方法

F. Samejima 认为若在测验之前知道被试总体的能力分布信息，则应充分利用这种信息，以提高测验的估计准确度. MAP 方法直接将先验概率密度(一般取标准正态概率密度函数)乘以似然函数构建后验分布并求极大值，似然函数 $L_{MAP}(\theta_\alpha) = f(\theta_\alpha)\prod_{j=1}^{m}P_{\alpha j}^{u_{\alpha j}}(1-P_{\alpha j})^{1-u_{\alpha j}}$，其中 $f(\theta_\alpha)$ 是 θ_α 的先验分布，其对数似然函数 $\ln L_{MAP}(\theta_\alpha) = \ln f(\theta_\alpha) + \sum_{j=1}^{m}(u_{\alpha j}\ln P_{\alpha j} + (1-u_{\alpha j})\ln(1-P_{\alpha j}))$，令

$$\partial \ln L_{MAP}(\theta_\alpha)/\partial\theta_\alpha = 0, \tag{2}$$

求得 $\hat{\theta}_\alpha$ 为 θ_α 的极大似然估计值. 同理方程(2)也是非线性方程，需要进行牛顿-拉夫逊迭代.

MAP 会出现估计向先验均值回归的现象，即有偏估计. 事实上，MAP 的先验分布不一定是标准正态分布，还可以是一般正态分布、均匀分布或者是其它先验分布.

0.3　期望后验估计方法

被试能力的 EAP 估计的理论依据是贝叶斯定理 $h(\theta_\alpha | \boldsymbol{U}, \xi) = P(\boldsymbol{U} | \theta_\alpha)g(\theta)/(P(\boldsymbol{U}))$，其中 $P(\boldsymbol{U} | \theta_\alpha) = \prod_{j=1}^{m}P_j(\theta_\alpha)^{u_{\alpha j}}(1-P_j(\theta_\alpha))^{1-u_{\alpha j}}$. 设被试后验分布为 $g(\theta_\alpha)$，其均值可以表示为

$$E(\theta_\alpha) = \int\theta_\alpha g(\theta_\alpha)\prod_{j=1}^{m}P_j(\theta_\alpha)^{u_{\alpha j}}(1-P_j(\theta_\alpha))^{1-u_{\alpha j}}\mathrm{d}\theta_\alpha \Big/ \left(\int g(\theta_\alpha)\prod_{j=1}^{m}P_j(\theta_\alpha)^{u_{\alpha j}}(1-P_j(\theta_\alpha))^{1-u_{\alpha j}}\mathrm{d}\theta_\alpha\right).$$

由于该式含有积分，R. Bock 等[5] 使用高斯-厄尔米特积分公式给出了它的数值积分形式 $\hat{\theta}_\alpha = \sum_{k=1}^{q}X_k L(X_k)A(X_k) \Big/ \left(\sum_{k=1}^{q}L(X_k)A(X_k)\right)$，其中 $X_k = -3.5 + 7(k-1)/(q-1)$ 为数值积分节点，$k = 1, 2, \cdots, q$，q 为等距点，$L(X_k) = \prod_{j=1}^{m}P_j(X_k)^{u_{\alpha j}}(1-P_j(X_k))^{1-u_{\alpha j}}$，$A(X_k) = 8\mathrm{e}^{-X_k^2/2}/(q\sqrt{2\pi})$. EAP 方法不需要迭代.

0.4　改进 MLE 的能力估计新方法

在 MLE 方法的基础上，设计 2 个有固定反应的项目来限制能力估计值. 具体而言，改造 MLE 方法中的对数似然函数，新的似然函数为 $\ln L^*(\theta_\alpha) = \ln P_{\min} + \ln(1-P_{\max}) + \sum_{j=1}^{m}(u_{\alpha j}\ln P_{\alpha j} + (1-u_{\alpha j})\ln(1-P_{\alpha j}))$.

在题库中，记所有题目中的最大难度为 $b_{\max}$，最小难度为 $b_{\min}$，最大区分度为 $a_{\max}$. 构造 2 个虚拟题目，一个是具有大区分度且特别容易的题目：难度为 $b_{\min}$，区分度为 $a_{\max}$，猜测度为 0，在 3PLM 下 $P_{\min} = 1/(1+\exp(-Da_{\max}(\theta_\alpha - b_{\min})))$，并且假设被试一定能做对；另一个是大区分度且特别难的题目，难度为 $b_{\max}$，区分度为 $a_{\max}$，猜测度为 0，在 3PLM 下 $P_{\max} = 1/(1+\exp(-Da_{\max}(\theta_\alpha - b_{\max})))$，并且假设被试一定会做错. 再令

$$\partial \ln L^*(\theta_\alpha)/\partial\theta_\alpha = 0, \tag{3}$$

求得 $\hat{\theta}_\alpha$ 为 θ_α 的极大似然估计值.

新方法在任何被试反应模式下均存在估计值，可适用于各种反应模式；其先验信息仅由 $P_{\min}$ 和 $P_{\max}$ 给出，这仅涉及 2 个项目，不会影响能力估计的整个过程，所以不会缩小被试能力估计范围. 和能力估计的 MLE 相比，NMLE 仅仅增加了 2 个“新的”项目，所以 NMLE 具有 MLE 的基本性质. 比如 NMLE 仍然是能力参数的充分统计量，也具有渐近一致性和渐近正态性等优良特性. 当测验较长时，NMLE 像 MLE 方法一样是一种渐近无偏的能力估计方法.

0.5 新方法的合理性和可行性

为了检验新方法的合理性和可行性,共有 4 种能力估计方法参与比较:(i) MLET 方法,用牛顿-拉夫逊迭代方法对方程(1)求根,迭代更新 30 次后或者在更新值误差小于 0.001 时迭代结束,并且被试能力估计值限制在 $-3.5\sim3.5$ 之间;(ii) MAP 方法,设能力的先验分布为正态分布,用牛顿-拉夫逊迭代方法对方程(2)求根,迭代更新 30 次后或者在更新值误差小于 0.001 时迭代结束;(iii) EAP 方法,设能力的先验分布为正态分布,从 $-3.5\sim3.5$ 中共取 35 个积分点;(iv) NMLE 方法,用牛顿-拉夫逊迭代方法对方程(3)求根,迭代更新 30 次后或者在更新值误差小于 0.001 时迭代结束.

1 模拟实验

1.1 被试及题库模拟

为了考察能力的先验分布对各种能力估计方法的影响,共设计 3 组被试:(i) 被试组 1,模拟产生 1 000个被试,被试能力真值均服从均值为 0、方差为 1 的标准正态分布;(ii) 被试组 2,模拟产生 1 000 个被试,被试能力真值均服从均值为 -1、方差为 1 的正态分布;(iii) 被试组 3,模拟产生 1 000 个被试,被试能力真值均服从均值为 1、方差为 1 的正态分布;后续内容中被试组 a 简称为组 a,$a=1,2,3$.

本文在 3PLM 模型下设计题库,所有试验模拟条件同文献[2].题库结构如下:模拟生成 520 个项目且满足条件 $\ln a\sim N(0,1)$,$b\sim N(0,1)$,$c\sim \mathrm{Beta}(5,17)$,$0.2<a<2.5$,$-3.5<b<3.5$,$|a-b|<4$,$c<0.4$.题库的项目数据见表 1.

表 1 题库的项目数据

项目数据	区分度 a	难度 b	猜测度 c
平均值	1.001 30	−0.006 464 7	0.223 380
标准差	0.608 37	0.979 380 0	0.807 610

1.2 模拟 CAT 的施测过程

本文不考虑内容平衡,项目曝光控制以及机会红利对 CAT 的影响,简化 CAT 设计为:(i) 取被试的能力初值为 0;(ii) 采用最大 Fisher 信息量选题策略,信息量计算公式[1]为

$$I_j(\theta_\alpha)=\frac{D^2a_j^2(1-c_j)}{(c_j+\mathrm{e}^{Da_j(\theta_\alpha-b_j)})(1+\mathrm{e}^{-Da_j(\theta_\alpha-b_j)})^2};$$

(iii) 分定长和不定长 2 种测验.定长测验的测验长度分别为 10 和 40,取测验长度为 10 是为了考察 NMLE 方法是否适用于短测验和 CAT 测验初期的能力估计,取测验长度为 40 是为了考察 NMLE 是否和 MLE 一样在长测验中是一种渐近无偏的能力估计方法;不定长测验在被试累积信息量达到 16 时结束.

1.3 评价指标

评价指标有:测验偏差(Bias) $B_{ias}=\sum_{i=1}^{N}(\hat{\theta}_i-\theta_i)/N$,测验平均绝对离差(ABS) $A_{BS}=\sum_{i=1}^{N}|\hat{\theta}_i-\theta_i|/N$;测验均方根误差(RMSE) $R_{MSE}=\sqrt{\sum_{i=1}^{N}(\hat{\theta}_i-\theta_i)^2/(N-1)}$,能力估计效率(ability estimation efficiency,AEE) $A_{EE}=(\hat{\theta}_{\max}-\theta_{\min})/(\hat{\theta}_{\max}-\theta_{\min})$,不定长测验的测验平均长度(average test length,ATL) $A_{TL}=\sum_{i=1}^{N}t_{est\_length}(i)/N$,其中 N 为被试总人数,θ_i 为第 i 个被试的能力真值,$\hat{\theta}_i$ 为第 i 个被试的能力估计值,$\theta_{\max}$ 为 N 个被试中能力的最大值,$\theta_{\min}$ 为 N 个被试中能力的最小值,$\hat{\theta}_{\max}$ 为 N 个被试中能力估计的最大值,$\hat{\theta}_{\min}$ 为 N 个被试中能力估计的最小值,$t_{est\_length}(i)$ 为被试 i 的测验长度.

测验偏差(Bias)表示能力估计的无偏性,测验平均绝对离差(ABS)和测验均方根误差(RMSE)表明了能力估计的准确性.Bias 和 ABS 反映了能力估计的系统偏差,RMSE 反映了能力估计值和真实值的随机误差,它们都是评价测验准确性的常用指标,它们越接近 0,表示能力估计越接近无偏,即能力估计越准确.能力估计效率(AEE)是本文提出的一个新的评价指标,用来评价能力估计方法对能力估计范围的影响,A_{EE}取值越接近 1 表明该能力估计方法受外界影响越小,不会缩小能力估计范围.因为不定长测验中每个被试的测量精度类似,所以早达到测验精度的被试所需测验长度更短,而晚达到测验精度的被试所需测验长度就更长,这项指标体现了测验效率[10].

1.4 实验结果及其分析

3 种测验条件下的测验偏差(Bias)值见表 2,当测验长度为 10 时,结果见表 3,当测验长度为 40 时,结果见表 4,当测验为不定长时,结果见表 5.

能力估计的无偏性对项目反应模型的应用非常重要,若参数估计的偏差较大,则会给更深入的测评带来严重的误差[11-13],测验偏差(Bias)用于评价无偏性和偏差的方向性(正偏或负偏).在长测验中,较小的 Bias 值体现了 MLET 方法和 NMLE 方法一样,具有能力估计的渐近无偏性.

表 2　3 种测验条件下测验偏差(Bias) 值

能力估计方法	测验长度为 10			测验长度为 40			不定长测验		
	组 1	组 2	组 3	组 1	组 2	组 3	组 1	组 2	组 3
MLET	0.024	0.035	0.028	0.002	0.001	0.002	0.017	0.014	0.015
MAP	0.033	0.036	−0.028	0.001	0.007	−0.006	0.012	0.020	−0.019
EAP	0.030	0.032	−0.023	0.000	0.009	−0.008	0.011	0.019	−0.017
NMLE	0.022	0.031	0.026	0.002	0.001	0.001	0.016	0.011	0.012

表 3　当测验长度为 10 时 4 种能力估计方法的表现

能力估计方法	ABS			RMSE			AEE		
	组 1	组 2	组 3	组 1	组 2	组 3	组 1	组 2	组 3
MLET	0.251	0.267	0.273	0.311	0.364	0.277	0.900	0.882	0.896
MAP	0.239	0.306	0.298	0.253	0.337	0.308	0.883	0.877	0.884
EAP	0.248	0.246	0.250	0.292	0.346	0.310	0.874	0.885	0.878
NMLE	0.252	0.250	0.262	0.315	0.296	0.274	0.991	0.990	0.992

在短测验中,被试能力分布对 MLET 方法和 NMLE 方法没有太大影响,但在被试服从标准正态分布时正好和假设的 MAP 和 EAP 的先验分布一致,此时 MAP 和 EAP 的能力估计精度更高; NMLE 方法无论是 ABS 还是 REMS 都小于 MLET 方法,这表明新方法参数估计的精度优于传统的 MLE 方法,新方法具有更小的估计误差. 又因为 NMLE 方法可以处理各种被试反应模式,所以 NMLE 方法比 MLET 方法更适用于短测验和 CAT 测验初期的能力估计. 因为测验太短,4 种能力估计的 AEE 指标差别不大,NMLE 方法表现稍好一点.

表 4　当测验长度为 40 时 4 种能力估计方法的表现

能力估计方法	ABS			RMSE			AEE		
	组 1	组 2	组 3	组 1	组 2	组 3	组 1	组 2	组 3
MLET	0.119	0.114	0.118	0.122	0.135	0.137	0.912	0.893	0.897
MAP	0.076	0.133	0.132	0.131	0.144	0.143	0.731	0.765	0.780
EAP	0.087	0.123	0.131	0.140	0.139	0.136	0.835	0.866	0.868
NMLE	0.105	0.110	0.119	0.122	0.136	0.134	0.995	0.997	0.992

在长测验中,较小的 ABS 值和 RMSE 值体现 MLET 方法和 NMLE 方法能力估计具有良好的返真性,这表明 2 种方法受测试条件的影响较小. 能力的先验分布对 MAP 方法和 EAP 方法的影响也逐渐体现出来,因为先验分布信息的作用,估计值会倾向于先验中心(在本文中先验分布标准正态分布的均值为 0,先验中心为 0) ,所以当能力估计值大于 0 时,就会被低估,当能力估计值小于 0 时,就会被高估,这样会缩小能力估计范围. 又因为 MLET 方法中设定了能力估计的上界和下界,超出界限的不同被试,尽管有不同的反应模式但得到的却是相同的能力估计值,所以这也缩小了能力估计的范围; NMLE 的 AEE 值均大于其它 3 种方法的 AEE 值,并且随着测验长度的增加,NMLE 的 AEE 值接近 1,这表明新方法不会缩小能力估计的有效范围.

表 5　不定长测验 4 种能力估计方法的表现

能力估计方法	ABS			RMSE			AEE			ATL		
	组 1	组 2	组 3	组 1	组 2	组 3	组 1	组 2	组 3	组 1	组 2	组 3
MLET	0.216	0.214	0.205	0.218	0.243	0.217	0.918	0.910	0.915	33.12	33.71	33.25
MAP	0.187	0.223	0.231	0.197	0.248	0.270	0.772	0.738	0.756	29.85	31.14	32.64
EAP	0.172	0.218	0.234	0.176	0.255	0.272	0.859	0.861	0.887	30.76	31.20	33.53
NMLE	0.201	0.211	0.204	0.203	0.231	0.209	0.993	0.993	0.994	29.34	30.70	30.20

从表 5 可以看出,当测验为不定长时,实验结果和定长测验类似,先验分布对 MAP 和 EAP 的影响依然存在. 各种方法测验精度比定长测验更低一些,但测验平均长度都短于定长测验的测验长度,这也说明了不定长测验更有利于提高测验效率. NMLE 方法的被试平均使用项目数少于其它 3 种能力估计方法,这说明 NMLE 方法比其它方法具有更高的测验效率.

通过这 4 组实验表明,新方法 NMLE 具有如下优点: (i) 对 MLE 方法而言,似然函数没有太大的改动,但易于实现; (ii) 不需要先验分布信息且还可以处理 MLE 处理不了的各种反应模式,适用于短测验和 CAT 能力估计初期; (iii) 在长测验中和 MLE 方法一样都具有能力估计的无偏性; (iv) 该方法迭代

计算过程是连续的,不会缩小能力估计有效范围;(v)具有更小的估计误差,测验精度更高.

2 讨论

能力估计的准确性影响了选题策略的自适应性,也影响了 CAT 测试结果.本文是在最简单的 CAT 模式下讨论的,仅考虑了测验精度,一般 CAT 还需要考虑如何提高题库利用率、降低机会红利、满足内容平衡等要求,这些都可能影响分析结果.能力估计新方法 NMLE,虽然它的似然函数有所改变,但是使被试对所有反应模式都有确定的能力估计值,且不像 MLET 方法对能力估计是绝对的限制,NMLE 方法依据题库参数,能力估计值是弹性变化的;它只要在似然函数中增加 2 个"新题"所以是一种相当简单的方法,而模拟实验表明它又有效.当然增加高区分度高难度的"新题"比较合理,而增加高区分度低难度的题目有一点勉强.

MAP 和 EAP 的先验分布可以是各种可能形式,理想的先验分布是像标准正态分布那样的单峰对称钟形曲线.但是在现实中,往往是非正态或者是不知名的分布,在实测中选择先验分布是一个比较困难的问题,而 NMLE 中的先验信息仅来自相应的题库,不需要被试本身的先验能力分布.所以,NMLE 方法在理论上是可行的,可以直接应用到多级评分模型中.当然,把 NMLE 方法应用于基于多维项目反应理论(multidimensional item response theory, MIRT)背景下的多维 CAT(multidimensional CAT, MCAT)[14-15]中还需要进一步讨论.

3 参考文献

[1] 漆书青,戴海崎,丁树良.现代教育与心理测量学原理[M].北京:高等教育出版社,2002.

[2] 张心,涂冬波.计算机化自适应测验中几种常用能力估计方法的特性与评价[J].中国考试,2014(5):18-25.

[3] Lord F M,Novick M R. Statistical theories of mental test scores[M]. New Jersey: Addison-Wesley,1968:392-449.

[4] Samejima F. Estimation of latent ability using a response pattern of graded scores[J]. Psychometrika,1969,34(1):1-97.

[5] Bock R,Mislevy R. Adaptive EAP estimation of ability in a microcomputer environment[J]. Applied Psychological Measurement,1982,6(4):431-444.

[6] Hambleton R K,Swaminathan H. Item response theory: Principles and application[M]. Boston: Kluwe-Nijhoff,1985.

[7] Wang Tianyou,Walter P Vispoel. Properties of ability estimation methods in computerized adaptive testing[J]. Journal of Educational Measurement,1998,35(3):109-135.

[8] Warm T A. Weighted likelihood estimation of ability in term response theory[J]. Psychmetrika,1989,54(3):427-450.

[9] 李佳,丁树良.多种分层方法在 CAT 校准误差中的应用研究[J].江西师范大学学报:自然科学版,2016,39(1):69-72.

[10] 李佳,丁树良,方剑英.基于平均数形式的选题策略比较[J].江西师范大学学报:自然科学版,2015,39(1):69-72.

[11] 孟祥斌,陶剑,陈莎莉.四参数 Logistic 模型潜在特质参数的 Warm 加权极大似然估计[J].心理学报,2016,48(8):1047-1056.

[12] Baker F B,Kim S H. Item response theory: parameter estimation techniques[M]. New York: Marcel Dekker,2004.

[13] Magis D A. Accuracy of asymptotic standard errors of the maximum and weighted likelihood estimators of proficiency levels with short tests[J]. Applied Psychology Measurement,2014,38(2):105-121.

[14] 毛秀珍,辛涛.多维计算机化自适应测验:模型、技术和方法[J].心理科学进展,2015,23(8):907-918.

[15] 韩雨婷,涂冬波,王潇潆,等.多维计算机化自适应测验选题策略的开发及比较[J].心理学报,2017,40(4):997-1004.

The New Method of Ability Estimation in CAT

LI Jia,DING Shuliang

(College of Computer Information Engineering Jiangxi Normal University,Nanchang Jiangxi 330022,China)

Abstract: The maximum likelihood estimation method (MLE) of the ability estimation does not work with special response patterns,such as all elements of the response patter are 0s or all 1s. If setting lower and upper bounds of ability estimation,the ability estimation scale will shorten. Bayesian-based estimation methods need a prior distribution,the choice of prior distribution must be careful. A new ability estimated method (NMLE) is introduced,adding two new items to establish a new likelihood function based on the existing item bank. New method not only need not ability prior distribution,but also does not shorten the ability estimation scale,and can deal with all kinds of response patterns. New method has better performance through the Monte Carlo simulation method on 3PLM.

Key words: MAP; EAP; NMLE; ability estimation efficiency

(责任编辑:冉小晓)