第 27 卷第 1 期
2003 年 1 月

江西师范大学学报(自然科学版)
JOURNAL OF JIANGXI NORMAL UNIVERSITY

Vol.27 No.1
Jan.2003

文章编号:1000-5862(2003)01-0056-05

基于 IRT 若干参数估计方式的比较

罗 芬, 丁树良, 胡小松, 万宇文, 甘登文
(江西师范大学 计算机科学与技术学院,江西 南昌 330027)

摘要:在项目反应理论(IRT)框架下,就目前流行的若干能力参数和项目参数的估计方法进行分析比较,阐述了它们各自适用的范围和不足之处,为选用估计方法提供依据.

关键词:CAT;IRT;条件似然估计;联合似然估计;边际似然估计与 EM 算法;贝叶斯估计

中图分类号:O 212　　**文献标识码**:A

随着计算机技术在考试中的应用,自适应考试(Computerized adaptive test 简称 CAT)日益受到人们的关注.CAT 选择难度在被试能力范围附近的项目(题目)进行测试,施测项目少,效率高;项目的选择和评分更加灵活;测试结果能够更精确地反映被试的实际水平[4].项目反应理论(Item Response Theory,简称 IRT)是计算机化自适应考试重要的理论基础.CAT 中至关重要的问题是评估被试的能力,为达此目的必须对测验的质量(特别是项目参数)进行评价.本文基于 IRT 对被试的能力和项目参数进行估计的常用方法进行综述.文章对某些问题的处理提出了自己的看法.比如为什么参数估计不采用矩估计,一些迭代初值不同选择的根据以及通过剖析 MMLE/EM 与 MMLE 的联系,说明估计 3PLM 中项目参数时,EM 算法如何使问题变得简单的原因.

1 参数估计的由来

IRT 是一种新兴的心理与教育测验理论,它突破了经典测验理论的局限性,将被试特质水平与被试在项目上的行为关联起来并且将其参数化,模型化[2].构造模型的方法有很多种,通常采用 Logistic 模型将它们关联,对考生和试卷进行评价的指标称为参数,通常只关心能力参数(θ)和项目参数($\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$),其含义在下面论述中说明.根据参数的不同,Logistic 项目特征曲线(简记为 ICC)可分为单参数,双参数和三参数 3 种模式.

三参数模式:$p(\theta;\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=\boldsymbol{c}+(1-\boldsymbol{c})/(1+\exp(-d\boldsymbol{a}(\theta-\boldsymbol{b})))$.

其中 $d=1.702$,$\boldsymbol{\theta}$:被试能力值,$\boldsymbol{a}$:项目的区分度,$\boldsymbol{b}$:项目的难度,$\boldsymbol{c}$:项目的猜测系数,$p(\boldsymbol{\theta};\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$表示能力为 $\boldsymbol{\theta}$ 的被试答对区分度为 $\boldsymbol{a}$,难度为 $\boldsymbol{b}$,猜测度为 $\boldsymbol{c}$ 的项目的概率[2].当 $\boldsymbol{c}=0$ 为双参数模式,当 $\boldsymbol{c}=0$ 且 $\boldsymbol{a}=1$时为单参数模式.通常 $\boldsymbol{\theta},\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 都是未知的,我们要根据被试对项目的作答反应对 $\boldsymbol{\theta},\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 进行估计.为了不拘泥于细枝末节(除非另有申明),我们采用双参数 Logistic 的 0-1 评分模型进行比较.根据具体问题,可出现几种情况:第一是各项目参数已知,估计被试能力;第二是被试能力已知,估计项目参数;第三是同时估计被试能力和项目参数[2].

2 参数估计的几种方法及比较

对不同的 α,不同的 j,$u_{\alpha j}$的分布是不相同的,故无法用矩估计对 $\boldsymbol{\theta},\boldsymbol{a},\boldsymbol{b}$ 进行估计,而且在一定条件下极大似然函数有其自身优点[3],故我们估计参数最常用的方法是极大似然估计法.假定①各个被试的作答模式是相互独立的;②同一被试对各个项目的作答是相互独立的;$u_{\alpha j}$表示第 α 个被试在第 j 个项目的反应(答对 $u_{\alpha j}=1$,答错 $u_{\alpha j}=0$),则基于得分矩阵 $\boldsymbol{U}=[u_{\alpha j}]$要对 θ_α,a_j,b_j 进行推断的似然函数为:

收稿日期:2002-09-02
基金项目:国家自然科学基金资助项目(60263005),并部分受全国教育科学规划重点课题资助项目(DBB010501)
作者简介:罗 芬(1977-),女,江西宜春人,硕士研究生,主要从事智能教育软件与人工智能的研究.

$$L(\boldsymbol{\theta};\boldsymbol{a},\boldsymbol{b}\mid \boldsymbol{U})=\prod_{\alpha=1}^{N}\prod_{j=1}^{m} P_{\alpha j}^{u_{\alpha j}}Q_{\alpha j}^{1-u_{\alpha j}},\tag{1}$$

其中 $\boldsymbol{\theta}=[\theta_\alpha]$;$\boldsymbol{a}=[a_j]$;$\boldsymbol{b}=[b_j]$,$N$:被试人数;$m$:项目数 $P_{\alpha j}$:被试 α 答对第 j 个项目的概率,$Q_{\alpha j}=1-P_{\alpha j}$:被试 α 答错第 j 个项目的概率.

由于 $L(\boldsymbol{\theta};\boldsymbol{a},\boldsymbol{b}\mid \boldsymbol{U})$ 与 $\ln L(\boldsymbol{\theta};\boldsymbol{a},\boldsymbol{b}\mid \boldsymbol{U})$ 同时达到极大,我们通常讨论都是对数似然函数,简化为

$$\ln L=\sum_{\alpha=1}^{N}\sum_{j=1}^{m}\left(u_{\alpha j}\ln P_{\alpha j}+(1-u_{\alpha j})\ln Q_{\alpha j}\right).\tag{2}$$

通过求式(2)的条件极值获得该参数的极大似然估计值.

2.1 条件极大似然估计(CMLE: condition maximum likelihood estimation)

CMLE 根据已知参数的不同分为:①能力参数的条件估计;②项目参数的条件估计.

2.1.1 已知项目参数求能力值

我们估计每一位被试的能力值,令式(2)对 θ_α 的一阶导等于0,即

$$\partial\ln L/\partial\theta_\alpha=0,\alpha=1,2,\cdots,N.\tag{3}$$

上式关于能力参数表达式是非线性方程,必须用迭代法求解,一般通过 Newton - Raphson(N - R)迭代求解[2]或二分法求解[5],迭代的终止条件为 $|\theta^{(t+1)}-\theta^{(t)}|<\varepsilon$($\varepsilon$ 为指定的一个充分小的正数).由于极大似然估计是在一定范围内得到极大值,迭代初值的选取就非常重要.N - R 迭代的初值一般取该被试者在这次测试中的得分与失分之比的自然对数,令 $X_\alpha=\sum_{j=1}^{m}u_{\alpha j}$,即 $\theta^{(0)}=\ln(X_\alpha/(m-X_\alpha))$,而二分法的初值范围为 $\theta\in[-3,3]$(可放宽到 $[-10,10]$).由于二分法收敛较慢,但是全局收敛,N - R 迭代收敛较快,但对初值较敏感,故我们可以采用将 N - R 迭代与二分法相结合的混合算法求解出 θ_α.

上述估计是已知项目参数 $\boldsymbol{a},\boldsymbol{b}$ 的前提下进行求解的,在求解过程中可能有些值的估计会越界,可以将它们进行约束到 $[-3,3]$ 区间内.用 N - R 方法必须保证分母不为0,遇到分母为0时,应输出迭代失败的提示.对于所有项目全答对或全答错的被试(称之为特殊反应模式),CMLE 不能正确估计其能力,因为此时只有 $|\theta|\to\infty$ 时,似然函数才能达到最大.我们可在迭代开始前将这些特殊反应模式剔除.

2.1.2 已知能力值,求项目参数

对每个项目 j,对数似然函数(2)有两个未知参数 $\boldsymbol{a},\boldsymbol{b}$,分别对 $\boldsymbol{a},\boldsymbol{b}$ 求导:

$$\begin{cases}f_1=\partial\ln L/\partial a_j=0\\ f_2=\partial\ln L/\partial b_j=0\end{cases},\quad j=0,1,2,\cdots,m.$$

同样需迭代求解出 $\boldsymbol{a},\boldsymbol{b}$,的估计值.采用 Newton - Raphson 迭代求解,雅可比矩阵 $\boldsymbol{J}$ 由 $\ln L$ 的二阶导组成,即

$$令\ J=\begin{pmatrix}\partial^2\ln L/\partial a_j^2 & \partial^2\ln L/\partial a_j\partial b_j\\ \partial^2\ln L/\partial b_j\partial a_j & \partial^2\ln L/\partial b_j^2\end{pmatrix},则\begin{pmatrix}a_j\\ b_j\end{pmatrix}^{(t+1)}=\begin{pmatrix}a_j\\ b_j\end{pmatrix}^{(t)}+\boldsymbol{J}^{-1}\begin{pmatrix}f_1\\ f_2\end{pmatrix}_{a_j=a_j^{(t)},b_j=b_j^{(t)}}.\tag{4}$$

式(4)的迭代是对向量进行,故迭代的终止条件要用范数来表示.即

$$\left\|\begin{pmatrix}a_j^{(t+1)}-a_j^{(t)}\\ b_j^{(t+1)}-b_j^{(t)}\end{pmatrix}\right\|<\varepsilon\Rightarrow\sqrt{(a_j^{(t+1)}-a_j^{(t)})^2+(b_j^{(t+1)}-b_j^{(t)})^2}<\varepsilon.$$

我们在求解过程中,通常用 $-\boldsymbol{E}(\boldsymbol{J})$ 来代替 $\boldsymbol{J}$(称之为 Fisher - score 迭代法),对于二阶导中的未知参数 $u_{\alpha j}$,有 $\boldsymbol{E}(u_{\alpha j})=P_{\alpha j}$.起初,$-\boldsymbol{E}(\boldsymbol{J})$ 是正定阵,但随着迭代次数的增加,有可能使得 $\boldsymbol{E}(\boldsymbol{J})$ 不可逆或病态,最终导致迭代式无意义或不收敛.要消除这种影响,通常在迭代过程中对 $\boldsymbol{E}(\boldsymbol{J})$ 进行调整,即在 $\boldsymbol{E}(\boldsymbol{J})$ 的每个对角元加上一个正常数 λ(阻尼因子),使得矩阵是对角元为正数,且对角占优的正定阵.迭代中的初值同样相当重要,常用的方法有点二列相关系数与 Z 分数相结合的方法和 Berkson 提倡的极小化 χ^2 准则导出的参数 $\boldsymbol{a}$、$\boldsymbol{b}$ 的初值[1,2].这两种初值估计方法的出发点是不同的,点二列相关系数与 Z 分数相结合的方法是以教育与心理测量的专业知识为基础给出的,Berkson 的极小化 χ^2 准则导出的估计实际上是在经验 Logistic 基础上得到的,尽管人们设法找到 $\boldsymbol{a},\boldsymbol{b}$ 的较好初值,然而迭代过程中还可能出现的某个项目 $\boldsymbol{a}$ 值过大(称为 Heywood case),导致下一次迭代不收敛的情况,我们同样希望将它们拉回到指定的范围内.

与第一种情况类似,对于所有被试者都答对(或都答错)的项目,不能正确估计项目参数,也可在迭代开始前将这些项目剔除.

2.2 联合极大似然估计(JMLE: joint maximum likelihood estimation)

CMLE 是基于一定条件下估计参数.能力的条件估计特别适用于由题库项目测试被试时估计被试能力,但是在题库建立之前,被试的能力参数和项目参数均未知,欲对这些参数进行估计,CMLE 就不适用了.Birnbaum 提出了联合极大似然估计(JMLE).JMLE 含有太多的不确定信息,我们要减少这种不确定性,基于①、②假设前提可以把 JMLE 分化为能力参数的条件估计和项目参数的条件估计.能力参数的估计与项目参数的估计是一个不断互相校正的过程,将这两部分反复迭代求取稳定值[2],这是 JMLE 联合求解参数的重要思想方法.JMLE 中所有求解公式都与 CMLE 相同,只是两者的相互结合、互相校正,故可以看成是两步估计的过程,其步聚如下:

(1)计算出迭代初值向量 $\boldsymbol{a}^{(0)},\boldsymbol{b}^{(0)},\boldsymbol{\theta}^{(0)}$;(2)计算似然函数的值 $L_0=L(\boldsymbol{a}^{(0)},\boldsymbol{b}^{(0)},\boldsymbol{\theta}^{(0)})$;(3)将 $\boldsymbol{a}^{(0)},\boldsymbol{b}^{(0)}$ 看成已知,用 CMLE 中式(3)求解得 $\boldsymbol{\theta}^{(1)}$;(4)将 $\boldsymbol{\theta}^{(1)}$ 看成已知,用 CMLE 中式(4)求得 $\boldsymbol{a}^{(1)},\boldsymbol{b}^{(1)}$;(5)将 $\boldsymbol{\theta}^{(1)}$ 标准化,$\boldsymbol{a}^{(1)},\boldsymbol{b}^{(1)}$ 也随之标准化;(6)计算似然函数的值 $L_1=L(\boldsymbol{a}^{(1)},\boldsymbol{b}^{(1)},\boldsymbol{\theta}^{(1)})$;(7)判断 $|L_1-L_0|<\boldsymbol{\varepsilon}$,如果满足条件或达到给定的迭代次数则退出,否则令 $L_1\to L_0,\boldsymbol{a}^{(1)}\to\boldsymbol{a}^{(0)},\boldsymbol{b}^{(1)}\to\boldsymbol{b}^{(0)},\boldsymbol{\theta}^{(1)}\to\boldsymbol{\theta}^{(0)}$ 转至第(3)步.

为了减少偏差,我们在每一轮迭代完成都将 $\boldsymbol{\theta},\boldsymbol{a},\boldsymbol{b}$ 标准化,迭代中可能出现的问题与 CMLE 是一致的.

除此之外,JMLE 还有其固有的缺陷:在给定长度测试中(项目个数是固定的),能力参数的个数是随着被试人数的增加而增加,称之为伴随参数;项目参数是不随被试人数增加而增加,我们称之为结构参数,当能力参数和项目参数同时进行估计时,伴随参数可能不存在充分统计量(充分统计量是指压缩但不损失关于未知参数信息的统计量),则 2PLM 和 3PLM 的结构参数的 MLE 不一定是相合估计[1,2],然而相合性是对一个好的估计量的基本要求.

2.3 边际极大似然估计与 EM 算法(MMLE: marginal maximum likelihood estimation and an EM algorithm)

只有消除伴随参数的影响才能消除 JMLE 固有的缺陷,通过给出能力的先验分布和对能力参数进行积分,来消除能力参数,这便要采用能力参数边际化的手段.令

$$P(u_\alpha|\zeta)=\int L(u_\alpha|\theta_\alpha,\zeta)g(\theta_\alpha|\zeta)\mathrm{d}\theta_\alpha=\int L(u_\alpha|\theta_\alpha,\zeta)g(\theta_\alpha)\mathrm{d}\theta_\alpha,$$

其中 ζ 为项目参数向量,则 $h(\boldsymbol{\theta}|u_\alpha)=L(u_\alpha|\theta_\alpha,\zeta)g(\boldsymbol{\theta})/P(u_\alpha|\zeta)$ 为 $\boldsymbol{\theta}$ 对 u_α 的条件分布,即是 θ 的后验分布.记 $M=\prod_{\alpha=1}^{N}P(u_\alpha|\zeta)$,$M$ 是基于反应矩阵 $\boldsymbol{U}$ 的边际似然函数,则

$$\ln M=\sum_{\alpha=1}^{N}\ln P(u_\alpha|\zeta). \tag{5}$$

我们以三参数为例子,将 MMLE 与 MMLE/EM 算法作比较,更能体现出 EM 算法的优越性,求项目参数的边际似然估计.(1)求 M;(2)令 $\partial\ln M/\partial a_j=0,\partial\ln M/\partial b_j=0,\partial\ln M/\partial c_j=0,j=0,1,2,\cdots,m$,$M$ 的一阶导中都含有积分符号,通过高斯—厄米特求积公式(Gauss - Hermite quadrature)化积分为求和,可令

$$\partial\ln M/\partial a_j=(1-c_j)\sum_{k=1}^{q}(X_k-b_j)[r_{jk}-f_kP_j(X_k)]W_{kj}=0,$$

$$\partial\ln M/\partial b_j=a_j\sum_{k=1}^{q}[r_{jk}-f_kP_j(X_k)]W_{kj}=0,$$

$$\partial\ln M/\partial c_j=(1-c_j)^{-1}\sum_{k=1}^{q}[r_{jk}-f_kP_j(X_k)]/P_j(X_k)=0,$$

其中 $P_j(X_k)=c_j+(1-c_j)/(1+\exp(-a_j(X_k-b_j)))$,$L(X_k)=\prod_{j=1}^{m}P_j(X_k)^{u_{\alpha j}}Q_j(X_k)^{1-u_{\alpha j}}$,

$$f_k=\sum_{\alpha=1}^{N}h(X_k|u_\alpha,\zeta)=\sum_{\alpha=1}^{N}\frac{L(X_k)A(X_k)}{\sum_{k=1}^{q}L(X_k)A(X_k)},\quad r_{jk}=\sum_{\alpha=1}^{N}u_{\alpha j}h(X_k|u_\alpha,\zeta)=\sum_{\alpha=1}^{N}\frac{u_{\alpha j}L(X_k)A(X_k)}{\sum_{k=1}^{q}L(X_k)A(X_k)},$$

$$W_{kj}=P_j^*(X_k)Q_j^*(X_k)/[P_j(X_k)Q(X_k)],$$
$$P_j^*(X_k)=(P_j(X_k)-c_j)/(1-c_j),\ Q_j^*(X_k)=1-P_j^*(X_k).$$

X_k 为求积节点,$A(X_k)$为求积系数,f_k,r_{jk}称为人工数据,从它们表达式中可以得知其统计意义,f_k 表示容量为N 的总体中期望具有能力为X_k 的被试人数,r_{jk}表示该总体中具有能力为X_k 的被试期望答对第j 项目的人数[2],f_k、r_{jk}都依赖于项目参数$\boldsymbol{a}$,$\boldsymbol{b}$ 的值,且有 $E(r_{jk}-P_{kj}f_k)=0$.我们可以看到对于 $\ln M$ 求其二阶导是相当复杂的表达式.所以,用 MMLE 对项目参数的求解,只能对很少量项目才能得到项目参数的估计值.1981 年 Bock 和 Aiktin 提出了用 EM 算法实施 MMLE.我们设想项目参数 ζ 的值在前一次迭代中已算出了一个估计值,若设最近一次得到的值总比前一次得到的值更接近于真值,则在 MMLE 中可用最近一次获得的项目参数值作为 ζ 的值,那么人工数据 f_k,r_{jk}在下一次迭代中为已知值,且有 $E(r_{jk}-P_{kj}f_k)=0$,$\ln M$ 的二阶导变得比较简单,这种想法极大地简化了迭代的公式和求解过程.

$$E\left(\frac{\partial \ln M}{\partial a_j^2}\right)=(1-c_j)\sum_{k=1}^{q}(X_k-b_j)\left[W_{kj}\left(-f_k^{(p)}\frac{\partial P_j(X_k)}{\partial a_j}+(r_{kj}^{(p)}-P_j(X_k)f_k^{(p)})\frac{\partial W_{kj}}{\partial a_j}\right]=\right.$$
$$-\sum_{k=1}^{q}(X_k-b_j)^2f_k^{(p)}P_j(X_k)Q_j(X_k).$$

同理可得

$E\left(\frac{\partial \ln M}{\partial a_j\partial b_j}\right)$,$E\left(\frac{\partial^2 \ln M}{\partial a_j\partial b_j}\right)$,$E\left(\frac{\partial^2 \ln M}{\partial b_j^2}\right)$,$E\left(\frac{\partial^2 \ln M}{\partial b_j\partial c_j}\right)$,$E\left(\frac{\partial^2 \ln M}{\partial c_j^2}\right)$的值,则

$$\begin{pmatrix}a_j\\ b_j\\ c_j\end{pmatrix}^{(t+1)}=\begin{pmatrix}a_j\\ b_j\\ c_j\end{pmatrix}^{(t)}-\begin{pmatrix}E\left(\frac{\partial^2 \ln L}{\partial a_j^2}\right) & E\left(\frac{\partial^2 \ln L}{\partial a_j\partial b_j}\right) & E\left(\frac{\partial^2 \ln L}{\partial a_j\partial c_j}\right)\\ E\left(\frac{\partial^2 \ln L}{\partial b_j\partial a_j}\right) & E\left(\frac{\partial^2 \ln L}{\partial b_j^2}\right) & E\left(\frac{\partial^2 \ln L}{\partial b_j\partial c_j}\right)\\ E\left(\frac{\partial^2 \ln L}{\partial c_j\partial a_j}\right) & E\left(\frac{\partial^2 \ln L}{\partial c_j\partial b_j}\right) & E\left(\frac{\partial^2 \ln L}{\partial c_j^2}\right)\end{pmatrix}^{-1}\begin{pmatrix}\frac{\partial \ln L}{\partial a_j}\\ \frac{\partial \ln L}{\partial b_j}\\ \frac{\partial \ln L}{\partial c_j}\end{pmatrix}_{a_j=a_j^{(t)},\,b_j=b_j^{(t)},\,c_j=c_j^{(t)}}.$$

仿式(4)的 Fisher-score 方法求解,得到项目参数的估计值,每完成一次迭代,我们都要调整权重 $A(X_K)$ 的值[1].在边际极大似然估计中,能力的后验分布 $h(\theta\mid u,\zeta)$中的未知项目参数 ζ,如果用上一轮迭代获得的值 $\zeta^{(p)}$代替后就转化为:

$$h(\theta\mid u,\zeta^{(p)})=L(u\mid\theta,\zeta^{(p)})g(\theta)/\int L(u\mid\theta,\zeta^{(p)})g(\theta)\mathrm{d}\theta. \tag{6}$$

后验分布的这种转化后使 MMLE 恰好与 EM 算法一致,记为 MMLE/EM.简单地说 MMLE/EM 的算法分为两步[1][2]:

E 步: 计算 $E_\Theta[\ln f(u,\theta\mid\zeta)\mid u,\zeta^{(p)}]=\Omega(\zeta\mid\zeta^{(p)})$;$M$ 步:找 $\zeta^{(p+1)}$使得 $\Omega(\zeta^{(p+1)}\mid\zeta^{(p)})\geqslant\Omega(\zeta\mid\zeta^{(p)})$.

EM 算法在一般条件下可以收敛,并且计算比较简单,但是它的收敛速度较慢,并且对于特殊的反应模式(全答对,全答错)无法正确估计其项目参数,并且迭代过程中可能使得 $\boldsymbol{a}$,$\boldsymbol{b}$ 的值很大,从而使估计值越界.能力参数在 MMLE/EM 中不需要迭代,即可根据后验分布 $h(\boldsymbol{\theta}\mid u)$得到 θ_a 的 Bayes 后验期望估计,也可以通过式(6)得到 $\boldsymbol{\theta}$ 的 Bayes 后验众数估计.求 Bayes 后验众数估计又要进行迭代,因此计算较繁.

2.4　边际贝叶斯估计(marginalized Bayesian estimation)

以上若干参数估计方式中都有不能解决的问题,共同之处在于不能对特殊的反应模式进行处理,解决这个问题可用方式之一是 Bayes 方法,在统计估计中,如果除了当前样本数据外还有其它的信息,包括客观信息和经验积累的信息,那么在统计中就可以利用,以提高估计的质量,我们在参数中加入所估参数的先验信息,使得参数估计更加稳定,其值更合理、更符合实际.

在 Bayes 估计中最重要的一点是必须给出所要估计参数的先验信息,以参数先验分布的形式给出,由贝叶斯定理知:$h(\zeta\mid u,\boldsymbol{\theta})\propto L(u\mid\zeta,\boldsymbol{\theta})*g(\zeta)*g(\boldsymbol{\theta})$,其中 $g(\zeta)$:第 j 个项目参数的先验分布,$g(\boldsymbol{\theta})$:能力的先验分布,$L(u\mid\zeta,\theta)$:基于 $\boldsymbol{U}$ 的似然函数,$h(\zeta\mid u,\boldsymbol{\theta})$表示第 j 个项目参数的后验分布.

$$L(u,\zeta,\boldsymbol{\theta})=\prod_{a=1}^{N}\prod_{j=1}^{m}P_j(\theta_a)^{u_{aj}}Q_j(\theta_a)^{1-u_{aj}}g(\zeta)h(\boldsymbol{\theta})=\prod_{a=1}^{N}P(u_a\mid\theta_a,\zeta)g(\zeta)h(\boldsymbol{\theta}),$$

取上式的对数函数 $\ln L=\sum_{a=1}^{N}\ln P(u_a\mid\theta_a,\zeta)+\ln g(\zeta)+\ln h(\boldsymbol{\theta})$.

与式(5)相比,上式中的最后两项是待估参数的先验分布,如果给出参数的先验分布,这两项的表达式也可以得到,采取与边际极大似然估计相同的方法进行参数估计,则待估参数的值与先验分布有关.后验分布是贝叶斯推断的基础,边际贝叶斯估计就是极大化参数的后验分布,其他的求解方法都与边际估计相同.如同 2.3 中所述,对于能力参数也可以求取 Bayes 后验期望估计或 Bayes 后验众数估计.

引入贝叶斯方法,其最重要的作用就在于能够处理特殊的反应模式,二是对 MMLE/EM 估计中出现的越界的估计会自动调节.先验分布的作用是起收缩作用.我们希望所获得的先验分布是信息先验分布(该先验分布的方差小),这样就会将估计值压缩到先验分布的均值附近.使用边际贝叶斯估计是必须给出参数的先验分布,如果这个先验分布比较符合实际情况,用这种方式所计算的参数值是比较接近真值的,但是先验分布一旦偏离实际情况很远,那么所得到的结果则可能远离真实值.

3 分析与讨论

对于所估参数越界的情况将它们拉回到区间范围内,这样做的影响还未知.我们将那些特殊反应模式剔除,对参数估计结果的影响同样也是未知,都有待进一步研究.本文的讨论虽然大多基于双参数 Logistic 模型,但对单参数和三参数都可以作类似的讨论.选择一个合适的模型是进行这些参数评估的基础,也是 IRT 的难点.IRT 的深入发展带来了许多模型,如 评定量表模型(the rating scale model),等级反应模型(the graded response model)等,都给出了一些较好的结果,以上所介绍方法的思想和手段,原则上都可以用来解决这些模型的参数估计问题.

参考文献:

[1]Framk B.Baker Item response theory:parameter estimation techniques[M].Marcel Dekker, Inc.1992.57-62,104-105,178-179,185-186.

[2]漆书青,戴海崎,丁树良.现代教育与心理测量学[M]. 南昌:江西教育出版社,1998.92,108-109,119,124-141,150,160-163.

[3]陈希孺.概率论与数理统计[M].北京:中国科学技术大学出版社,1996.166-172.

[4]邵晨辉,陈玉泉,徐良贤.基于项目反应理论的机动自适应考试系统[J].计算机工程,2000,26(11):161-163.

[5]王能超.数值分析简明教程[M].北京:高等教育出版社,1995.140-150.

[6]漆书青, 戴海崎.项目反应理论及其应用研究[M].南昌:江西高校出版社,1992.125.

The Comparison between Different Parameters - Estimation Ways Based on Item Response Theory

LUO Fen, DING Shu-liang, HU Xiao-song, WAN Yu-wen, GAN Deng-wen

(Institute of Computer Science and Technology, Jiangxi Normal University, Nanchang 330027, China)

Abstract: In this paper, a number of ways to estimate the latent trait and item parameters was introduced and their advantages and disadvantages were analyzed to offer reliance for selecting fitting method based on item response theory.

Key words: computerized adaptive test (CAT); item response theory (IRT); condition maximum likelihood estimation; joint maximum likelihood estimation; marginal maximum likelihood estimation and an EM algorithm; Bayesian estimation