文章编号: 1000-5862(2012)06-0636-06

初值对 2PLM 下 MCMC 参数估计影响的评价

胡 海[1], 甘登文[1], 汪文义[2], 丁树良[1*], 符华均[2]

(1. 江西师范大学计算机信息工程学院, 江西 南昌 330022; 2. 江西师范大学心理学院, 江西 南昌 330022)

摘要: 针对马尔科夫链蒙特卡洛(MCMC)做参数估计因链长太长而存在耗时太久的缺陷, 利用 Monte Carlo 模拟进行研究, 结果发现: 使用 MCMC 进行估计项目反应理论中 2PLM 的未知能力参数和项目参数, 当样本量或项目量较大, 且链长较短时, 初值作用明显, 比较准确的初值估计精度更高. 这个发现可以在一定条件下, 弥补 MCMC 耗时太久的缺陷.

关键词: 马尔科夫链蒙特卡洛; 两参数 Logistic 模型; 初值; 链长; 参数估计

中图分类号: O 626.4 文献标志码: A

0 引言

项目反应理论(IRT)是心理与教育测量学中的重要内容, 模型参数的估计是其核心的内容[1]. 参数估计方法主要有: 条件极大似然估计(CMLE)、联合极大似然估计(JMLE)、边际极大似然估计(MMLE)、条件期望-极大化算法(E-M 算法)等. 马尔科夫链蒙特卡洛(MCMC)方法是一种动态的计算机模拟技术, 它是根据贝叶斯理论中的多元后验分布来模拟随机样本. 用初等方法估计模型参数, 避开了 E-M 算法中的复杂计算, 从而提高估计的成功率[2].

20 世纪 90 年代美国统计学家 J. H. Albert[3]首先将 MCMC 算法应用到 IRT 参数估计的研究中; R. J. Patz[4]和涂冬波[5]等也应用该算法估计了 2PLM 的参数, 估计结果类似. 对于 Markov 链的初值和链长, 以上几篇文献均采用固定的初值和较长的链长来进行, 通过一定长度的迭代使得链收敛到平稳分布. 本文对 2PLM 通过 IRT 方法估计参数初值, 然后与其它几种不同的参数初值进行比较, 分别考察它们在不同链长下对于参数估计精度的影响.

1 研究方法

1.1 2PLM

2PLM 为

$$p_{ij}(x_{ij}=1\mid\theta_i)=\frac{1}{1+\exp[-1.7a_j(\theta_i-b_j)]},$$

其中 θ_i 为被试 i 的能力参数, a_j、b_j 分别为第 j 个项目的区分度和难度, p_{ij} 表示能力为 θ_i 的被试答对项目 j 的概率; 如果被试 i 答对项目 j, 则 $x_{ij}=1$, 反之 $x_{ij}=0$.

1.2 参数初值估计

(1) 能力参数初值估计

$$\theta_i=\ln\Big(\sum_{j=1}^{m}u_{ij}/\big(m-\sum_{j=1}^{m}u_{ij}\big)\Big),\ i=1,2,\cdots,N,\ j=1,2,\cdots,m,$$

其中 u_{ij} 为第 i 个被试在第 j 题上的作答反应. 对于特殊情况, 如 $u_i=(0,\cdots,0)$, 即被试 i 对 m 个题全部答错, 此时可以令 $\theta_i=-3$; 或者出现 $u_i=(1,\cdots,1)$, 即被试 i 对 m 个题全部答对, 此时令 $\theta_i=3$.

(2) 项目参数初值估计

根据反应矩阵, 令

$$p_j=\sum_{\alpha}^{N}u_{\alpha j}/N,\quad q_j=1-p_j,\ j=1,2,\cdots,m,$$

其中 p_j、q_j 分别表示 N 个被试在第 j 题上的通过率和未通过率. 项目区分度用全体被试在该项目上的得分与卷面总分的点二列相关系数 r_{pbj} 表示为

$$r_{pbi}^{(j)}=\frac{\overline{x}_{pj}-\overline{x}_{qj}}{s}\sqrt{p_jq_j},$$

其中 $\overline{x}_{pj}$ ($\overline{x}_{qj}$)表示项目 j 上正确(错误)作答的被试卷面总分的平均数, s 为被试卷面总分标准差. 通过下

收稿日期: 2012-02-26

基金项目: 国家自然科学基金(30860084, 31160203, 31100756)资助项目.

作者简介: 丁树良(1949-), 男, 江西樟树人, 教授, 博士生导师, 主要从事计算辅助教学、应用及教育和心理测量方面的研究.

述方式求难度和区分度: 先将被试在项目中的通过率 p_j 转化为标准正态分数 z_j, 即设随机变量 ξ 遵从标准正态分布 $N(0,\ 1)$, 要寻找 z_j 使得随机事件 $\xi \geq z_j$ 的概率等于 p_j, 即 $p(\xi \geq z_j)=p_j$, 亦即 $\frac{1}{\sqrt{2\pi}}\int_{z_j}^{+\infty}\mathrm{e}^{-t^2/2}\mathrm{d}t=p_j$, 由此可以计算出 z_j; 又记 $y(z_j)=\frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}z_j^2\}$, $y(z_j)$ 为 z_j 对应的正态密度函数值. 然后把点二列相关系数 r_{pbj} 转化为二列相关系数 r_b, 其计算公式为 $r_{bj}=r_{pbi}^{(j)}\sqrt{p_j q_j}/y(z_j)$, 最后取 b_j 及 a_j 的近似值为 $b_j=z_j/r_{bj}$, $a_j=r_{bj}/\sqrt{1-r_{bj}^2}$.

1.3　参数估计过程

本文采用 Gibbs 范围内的随机游动 M-H 算法[7-10]进行 IRT 模型的参数估计, 步骤如下:

(1) 设定被试参数及项目参数的初值.

(2) 已知项目参数, 估计能力参数:

①对各被试(i=1, 2,⋯, N)独立地从建议分布中抽取一状态 $\theta_i^{(k)}$. 为计算方便, 本文的建议分布取正态分布, 即 $q(\theta_{k-1},\theta_k)=q(\theta_k,\theta_{k-1})$, 其中 $q_{\theta i}\sim N(\theta_i^{(k-1)},C_\theta^2)$.

②计算从状态 $\theta_i^{(k-1)}$ 到 $\theta_i^{(k)}$ 的接受概率 $\alpha(\theta_i^{(k-1)},\theta_i^{(k)})=\min\{\frac{\pi(\theta^{(k)})}{\pi(\theta^{(k-1)})},1\}$, 其中 $\frac{\pi(\theta^{(k)})}{\pi(\theta^{(k-1)})}=$

$$\frac{\prod_{j=1}^{m}P(\theta_i^{(k)},a_j^{(k-1)},b_j^{(k-1)})^{x_{ij}}Q(\theta_i^{(k)},a_j^{(k-1)},b_j^{(k-1)})^{1-x_{ij}}}{\prod_{j=1}^{m}P(\theta_i^{(k-1)},a_j^{(k-1)},b_j^{(k-1)})^{x_{ij}}Q(\theta_i^{(k-1)},a_j^{(k-1)},b_j^{(k-1)})^{1-x_{ij}}}\times\frac{\exp\{-(\theta_i^{(k)})^2/(2\sigma_\theta^2)\}}{\exp\{-(\theta_i^{(k-1)})^2/(2\sigma_\theta^2)\}}.$$

③产生随机数 $r\sim u[0,\ 1]$, 对状态进行判断

$$\theta_i^{(k)}=\begin{cases}\theta_i^{(k)}, & r\leqslant\alpha(\theta_i^{(k-1)},\theta_i^{(k)}),\\ \theta_i^{(k-1)}, & r>\alpha(\theta_i^{(k-1)},\theta_i^{(k)}).\end{cases}$$

(3) 已知能力参数, 估计项目参数:

①对各项目(j=1, 2,⋯, m)独立地生成下一状态, 其中 $q_{a_j}\sim N(a_j^{(k-1)},C_a^2)$, $q_{b_j}\sim N(b_j^{(k-1)},C_b^2)$.

② 计算从状态($a_j^{(k-1)}$, $b_j^{(k-1)}$)到($a_j^{(k)}$, $b_j^{(k)}$)的接受概率 $\alpha((a_j^{(k-1)},b_j^{(k-1)}),(a_j^{(k)},b_j^{(k)}))=\min\{\frac{\pi(a_j^{(k)},b_j^{(k)})}{\pi(a_j^{(k-1)},b_j^{(k-1)})},1\}$, 其中

$$\frac{\pi(a_j^{(k)},b_j^{(k)})}{\pi(a_j^{(k-1)},b_j^{(k-1)})}=\frac{\prod_{i=1}^{N}P(\theta_i^{(k)},a_j^{(k)},b_j^{(k)})^{x_{ij}}Q(\theta_i^{(k)},a_j^{(k)},b_j^{(k)})^{1-x_{ij}}}{\prod_{i=1}^{N}P(\theta_i^{(k)},a_j^{(k-1)},b_j^{(k-1)})^{x_{ij}}Q(\theta_i^{(k)},a_j^{(k-1)},b_j^{(k-1)})^{1-x_{ij}}}\times$$

$$\frac{\exp\{-\frac{1}{2\sigma_b^2}(b_j^{(k)})^2\}}{\exp\{-\frac{1}{2\sigma_b^2}(b_j^{(k-1)})^2\}}\times\frac{\frac{1}{a_j^{(k)}}\exp\{-\frac{1}{2\sigma_a^2}(\log(a_j^{(k)}))^2\}}{\frac{1}{a_j^{(k-1)}}\exp\{-\frac{1}{2\sigma_a^2}(\log(a_j^{(k-1)}))^2\}}.$$

③　产生随机数 $r\sim u[0,\ 1]$, 对状态进行判断

$$(a_i^{(k)},b_i^{(k)})=\begin{cases}(a_i^{(k)},b_i^{(k)}), r\leqslant\alpha((a_i^{(k-1)},b_i^{(k-1)}),(a_i^{(k)},b_i^{(k)})),\\ (a_i^{(k-1)},b_i^{(k-1)}), r>\alpha((a_i^{(k-1)},b_i^{(k-1)}),(a_i^{(k)},b_i^{(k)})).\end{cases}$$

(4) 重复步骤(2)和(3)L 次(L 为链长, 本文分别取 L=100, 200, 300, 400, 500, 600, 700, 800, 900, 1 000), 删除为首的 D 次(本文分别删除各链长的前半段), 取最后的 W 次(本文分别取各链长的后半段)迭代的平均值作为参数的估计值.

(5) 上述 4 步产生了一条链长为 L 的 Markov 链, 本文通过多次重复试验, 取其平均值作为最终的参数估计值.

本文中的参数与 R. J. Patz[4]论文里所采用的参数相同, 即 C_θ=1.1, C_a=0.3, C_b=0.3, σ_θ=1, σ_a=0.5, σ_b=2.

2　模拟实验

2.1　实验设计

(1) 产生参数真值: 被试能力 $\theta\sim N(0,\ 1)$, $\ln a\sim N(0,\ 0.4)$, $b\sim N(0,\ 1)$.

(2) 生成得分矩阵: 将参数真值代入 IRT 模型的数学表达式中, 可得到概率反应矩阵与同维度的 0-1 随机数矩阵, 从而得到模拟得分矩阵.

(3) 计算参数: 根据得分矩阵, 采用 MCMC 算法分别估计能力参数和项目参数.

(4)比较标准: 平均绝对离差 $ABSE=\sum_{r=1}^{R}(\sum_{i=1}^{M}|x_i-\hat{x}_{ir}|/M)/R$, 其中 $\hat{x}_{ir}$ 为 x_i 的第 r 次试验的估计值, x_i 是模拟真值. 当 x 为项目参数 a、b 时, M 为对应的项目个数; 当 x 为能力参数 θ 时, M 为对应的被试个数. R 为模拟次数, 本文取 R=10. $ABSE$ 值表示估计值与真值之间的绝对偏离程度, 也将 $ABSE$ 称为估计精度, 但是要注意 $ABSE$ 值越小, 估计精度越高, $ABSE$ 值越大, 估计精度越低.

2.2　实验方案

本文的实验工具为 Matlab 7.0 语言自编的 MCMC 参数估计程序. 参数初值采用 4 种类型: (1)随机初值: $\theta\sim U(-3,\ 3)$, $a\sim U(0.2,\ 2.5)$, $b\sim U(-3,\ 3)$. (2)固

定初值. 以参数的先验分布的均值作为初值, 即 $\theta=0$, $a=1$, $b=0$.(3)IRT 初值: 采用 1.2 节的 IRT 方法估计初值. (4)真值初值: 以模拟的真值作为初值.

为了考查不同初值在不同人数和题数下对于参数估计精度的影响, 采用了以下 2 个实验方案: (1)实验 1: 人数变化, 题目固定. 固定题目为 40 题, 人数分别为 200, 500, 800, 1 100, 1 400, 1 700, 2 000 人, 且共用同批项目参数. (2)实验 2: 人数固定, 题目变化. 固定人数为 1 000 人, 题目分别为 20, 40, 60, 80, 100 题, 且共用同批能力参数.

2.3 实验结果及分析

实验结果见表 1 和表 2.

表 1 人数变化对项目参数估计精度的影响(**40** 题)

参数	链长	初值/人	200	500	800	1 100	1 400	1 700	2 000
a 值	100	随机	0.265 7	0.229 9	0.246 7	0.192 7	0.258 6	0.219 2	0.252 4
		固定	0.190 9	0.091 4	0.099 4	0.100 6	0.091 1	0.107 3	0.106 3
		IRT 初值	0.231 1	0.113 5	0.112 4	0.096 1	0.059 2	0.079 2	0.062 4
		真值	0.231 7	0.104 0	0.111 9	0.087 7	0.058 7	0.073 4	0.053 5
	200	随机	0.247 8	0.150 3	0.160 9	0.125 8	0.134 1	0.134 3	0.135 3
		固定	0.224 7	0.100 3	0.109 0	0.089 8	0.067 3	0.082 3	0.075 9
		IRT 初值	0.237 9	0.119 9	0.124 0	0.098 6	0.061 4	0.076 2	0.063 8
		真值	0.227 7	0.112 9	0.106 5	0.091 0	0.062 4	0.078 6	0.062 5
	300	随机	0.253 2	0.131 1	0.145 0	0.109 8	0.093 4	0.108 2	0.100 1
		固定	0.236 1	0.119 4	0.117 6	0.093 3	0.059 6	0.074 9	0.065 3
		IRT 初值	0.250 7	0.122 6	0.130 2	0.105 0	0.064 2	0.076 5	0.066 0
		真值	0.236 8	0.121 2	0.115 3	0.094 9	0.059 7	0.078 0	0.062 6
	400	随机	0.248 3	0.129 1	0.134 5	0.102 1	0.077 3	0.091 0	0.085 2
		固定	0.239 8	0.124 2	0.120 8	0.097 3	0.058 5	0.074 1	0.062 3
		IRT 初值	0.248 4	0.126 0	0.132 2	0.104 0	0.062 0	0.076 0	0.064 3
		真值	0.235 7	0.116 3	0.120 5	0.095 7	0.059 3	0.077 4	0.062 3
	500	随机	0.243 8	0.127 7	0.133 5	0.098 6	0.069 3	0.086 8	0.078 0
		固定	0.240 0	0.120 9	0.123 9	0.100 8	0.059 3	0.074 6	0.062 3
		IRT 初值	0.242 5	0.132 1	0.135 5	0.103 5	0.059 7	0.073 4	0.064 0
		真值	0.234 0	0.114 0	0.125 0	0.097 7	0.060 1	0.077 1	0.064 4
	600	随机	0.2408 0	0.127 2	0.131 4	0.097 2	0.065 1	0.083 7	0.073 0
		固定	0.236 9	0.122 4	0.125 8	0.102 5	0.059 1	0.076 2	0.063 0
		IRT 初值	0.239 6	0.131 8	0.133 3	0.102 6	0.059 7	0.071 9	0.062 4
		真值	0.236 9	0.113 6	0.125 3	0.098 0	0.060 6	0.078 5	0.064 4
	700	随机	0.239 6	0.124 8	0.132 5	0.096 4	0.063 2	0.081 3	0.070 2
		固定	0.237 0	0.122 9	0.125 4	0.103 0	0.059 8	0.078 3	0.064 0
		IRT 初值	0.238 5	0.127 5	0.129 1	0.103 2	0.061 2	0.071 5	0.063 4
		真值	0.238 6	0.116 2	0.124 7	0.098 3	0.062 0	0.078 5	0.064 1
	800	随机	0.242 4	0.125 1	0.132 2	0.096 8	0.061 9	0.080 5	0.068 5
		固定	0.239 5	0.124 9	0.127 4	0.102 5	0.059 8	0.078 2	0.065 0
		IRT 初值	0.236 7	0.123 7	0.128 9	0.101 1	0.061 5	0.072 2	0.063 7
		真值	0.237 3	0.118 1	0.123 7	0.099 2	0.062 1	0.078 0	0.063 9
	900	随机	0.241 9	0.124 6	0.130 7	0.096 3	0.061 6	0.079 6	0.066 0
		固定	0.237 3	0.127 4	0.129 0	0.100 9	0.060 6	0.079 0	0.065 9
		IRT 初值	0.238 0	0.121 2	0.128 4	0.100 1	0.061 7	0.073 6	0.064 2
		真值	0.236 9	0.121 0	0.124 0	0.098 8	0.061 9	0.078 6	0.064 0
	1 000	随机	0.243 8	0.124 1	0.128 7	0.096 5	0.062 3	0.077 8	0.065 0
		固定	0.235 0	0.128 7	0.127 7	0.099 3	0.061 8	0.078 8	0.065 6
		IRT 初值	0.238 5	0.119 0	0.127 0	0.098 1	0.060 7	0.074 7	0.064 5
		真值	0.238 9	0.121 1	0.125 2	0.098 2	0.061 7	0.077 8	0.064 0

续表 1

参数	链长	初值/人	200	500	800	1 100	1 400	1 700	2 000
b 值	100	随机	0.299 0	0.160 8	0.256 6	0.152 3	0.147 2	0.163 9	0.164 0
		固定	0.153 5	0.075 7	0.070 1	0.065 7	0.062 3	0.057 1	0.073 6
		IRT 初值	0.198 8	0.105 7	0.091 7	0.061 0	0.063 8	0.045 8	0.038 7
		真值	0.192 4	0.077 9	0.084 8	0.044 9	0.054 1	0.044 0	0.034 5
	200	随机	0.256 7	0.133 5	0.193 4	0.083 5	0.095 1	0.106 0	0.094 2
		固定	0.196 1	0.088 7	0.091 6	0.064 4	0.058 1	0.043 2	0.055 6
		IRT 初值	0.211 0	0.110 3	0.100 1	0.066 1	0.067 9	0.046 0	0.039 2
		真值	0.195 7	0.094 1	0.088 2	0.049 4	0.055 1	0.047 3	0.037 7
b 值	300	随机	0.248 5	0.125 1	0.160 8	0.057 7	0.076 3	0.077 3	0.060 1
		固定	0.215 7	0.100 8	0.102 1	0.067 8	0.059 6	0.038 5	0.045 2
		IRT 初值	0.221 2	0.112 7	0.102 5	0.069 9	0.071 0	0.045 4	0.040 0
		真值	0.205 7	0.098 8	0.095 4	0.053 9	0.058 6	0.047 8	0.037 5
	400	随机	0.244 4	0.126 6	0.144 7	0.054 8	0.071 5	0.061 1	0.050 7
		固定	0.222 1	0.102 2	0.104 1	0.074 7	0.060 9	0.039 3	0.040 7
		IRT 初值	0.221 9	0.115 5	0.106 0	0.071 9	0.072 4	0.043 3	0.038 2
		真值	0.205 3	0.096 9	0.103 4	0.055 7	0.058 4	0.046 8	0.037 3
	500	随机	0.233 1	0.119 1	0.135 2	0.055 3	0.069 6	0.055 1	0.046 6
		固定	0.221 6	0.103 1	0.104 7	0.080 0	0.061 9	0.041 1	0.038 2
		IRT 初值	0.214 2	0.119 5	0.112 3	0.075 3	0.071 5	0.041 5	0.037 7
		真值	0.204 4	0.097 1	0.107 2	0.060 6	0.058 8	0.047 3	0.036 9
	600	随机	0.224 4	0.113 7	0.129 9	0.056 8	0.069 0	0.050 8	0.044 6
		固定	0.214 9	0.106 2	0.107 5	0.083 9	0.064 7	0.042 8	0.035 7
		IRT 初值	0.214 7	0.118 2	0.114 6	0.075 2	0.071 0	0.040 6	0.037 2
		真值	0.211 1	0.100 7	0.107 6	0.064 9	0.058 7	0.048 7	0.036 0
	700	随机	0.217 3	0.106 4	0.126 2	0.057 6	0.068 0	0.048 1	0.042 8
		固定	0.215 2	0.107 9	0.109 3	0.083 4	0.067 5	0.044 0	0.035 4
		IRT 初值	0.216 3	0.114 6	0.113 5	0.075 4	0.071 7	0.041 5	0.036 9
		真值	0.215 9	0.103 8	0.108 2	0.066 5	0.060 4	0.048 8	0.035 2
	800	随机	0.216 8	0.105 4	0.122 0	0.059 3	0.068 3	0.046 8	0.040 6
		固定	0.220 4	0.111 7	0.111 2	0.081 6	0.070 0	0.044 1	0.035 1
		IRT 初值	0.213 5	0.111 7	0.111 3	0.073 4	0.070 6	0.043 1	0.037 3
		真值	0.220 9	0.106 5	0.108 7	0.069 6	0.061 3	0.049 2	0.035 3
	900	随机	0.216 0	0.104 0	0.117 8	0.060 9	0.068 8	0.045 7	0.040 1
		固定	0.222 0	0.114 2	0.113 6	0.077 6	0.072 1	0.045 1	0.035 2
		IRT 初值	0.215 0	0.110 0	0.109 7	0.071 2	0.069 8	0.044 3	0.037 4
		真值	0.222 6	0.107 3	0.109 1	0.070 3	0.061 3	0.049 0	0.035 6
	1 000	随机	0.221 0	0.104 4	0.114 2	0.0620 0	0.070 2	0.044 1	0.040 1
		固定	0.223 8	0.115 4	0.112 0	0.074 9	0.072 7	0.045 6	0.035 4
		IRT 初值	0.218 0	0.107 1	0.107 6	0.071 1	0.069 0	0.045 9	0.036 7
		真值	0.225 2	0.108 8	0.109 4	0.069 6	0.061 3	0.048 3	0.035 1

表 2　题数变化对被试参数估计精度的影响(1 000 人)

参数	链长	初值/题	20	40	60	80	100
		随机	0.303 1	0.262 5	0.301 6	0.361 7	0.287 8
	100	固定	0.320 2	0.221 5	0.175 5	0.169 1	0.180 0
		IRT 初值	0.302 9	0.201 9	0.171 4	0.148 2	0.135 2
		真值	0.299 7	0.197 7	0.170 7	0.145 1	0.123 7
		随机	0.297 9	0.219 2	0.233 1	0.263 0	0.205 9
	200	固定	0.307 6	0.209 9	0.170 4	0.151 3	0.157 7
		IRT 初值	0.299 3	0.199 5	0.172 8	0.148 3	0.136 3
		真值	0.295 9	0.197 8	0.169 7	0.147 1	0.125 1
		随机	0.296 6	0.206 9	0.206 4	0.220 2	0.172 0
		固定	0.302 1	0.207 5	0.170 1	0.146 7	0.145 7
θ 值	300	IRT 初值	0.298 2	0.198 5	0.173 4	0.150 6	0.137 5
		真值	0.295 8	0.197 9	0.170 8	0.148 3	0.127 4
		随机	0.295 7	0.202 8	0.190 7	0.193 2	0.156 8
	400	固定	0.298 9	0.204 7	0.171 1	0.145 9	0.140 2
		IRT 初值	0.297 9	0.197 9	0.173 9	0.151 1	0.137 5
		真值	0.295 8	0.197 5	0.172 3	0.149 1	0.129 7
		随机	0.296 1	0.200 7	0.182 2	0.176 3	0.148 9
	500	固定	0.298 1	0.202 2	0.172 5	0.147 2	0.138 8
		IRT 初值	0.297 4	0.197 4	0.174 3	0.150 9	0.137 5
		真值	0.296 1	0.197 7	0.172 6	0.149 6	0.131 6
		随机	0.296 2	0.199 8	0.177 9	0.165 9	0.144 7
	600	固定	0.297 4	0.200 4	0.173 8	0.148 9	0.140 0
		IRT 初值	0.297 4	0.197 0	0.174 9	0.151 1	0.137 4
		真值	0.296 1	0.197 6	0.173 3	0.150 3	0.133 0
		随机	0.296 4	0.199 1	0.175 4	0.159 6	0.142 2
	700	固定	0.298 0	0.199 1	0.175 2	0.150 5	0.141 5
		IRT 初值	0.296 8	0.197 3	0.175 7	0.151 8	0.137 0
		真值	0.296 1	0.197 2	0.173 8	0.150 7	0.134 4
		随机	0.296 6	0.198 4	0.174 6	0.155 8	0.140 2
		固定	0.298 2	0.198 7	0.175 8	0.151 6	0.142 2
θ 值	800	IRT 初值	0.296 5	0.197 5	0.176 3	0.151 8	0.137 3
		真值	0.296 1	0.197 5	0.173 9	0.151 8	0.135 4
		随机	0.296 7	0.198 3	0.173 9	0.153 3	0.138 4
	900	固定	0.297 5	0.198 8	0.176 0	0.152 0	0.141 5
		IRT 初值	0.296 3	0.197 6	0.176 8	0.151 5	0.137 5
		真值	0.296 0	0.197 8	0.174 0	0.152 3	0.135 9
		随机	0.296 4	0.197 8	0.173 3	0.151 4	0.137 6
	1 000	固定	0.297 5	0.198 9	0.175 8	0.152 2	0.140 4
		IRT 初值	0.296 1	0.198 0	0.176 8	0.151 6	0.137 6
		真值	0.295 6	0.197 7	0.173 8	0.152 8	0.136 2

由表 1 可知: 固定项目个数为 40 题, 当样本容量较小时(如 200 人), 初值作用不明显; 当样本容量适中时(如 1 100 人), 初值作用比较明显; 当样本容量较大时(如 2 000 人), 初值作用明显. 如在样本量为 2 000 时, 从 3 种链长情况下的结果可以明显看出: 当链长为 100 时, IRT 初值的区分度和难度与真值的平均绝对离差 *ABSE* 值分别为 0.062 4 和 0.038 7, 明显小于随机初值的区分度和难度的 *ABSE* 值 0.252 4 和 0.164 0; 当链长增加到 500 时, IRT 初值的区分度和难度的 *ABSE* 值分别为 0.064 0 和 0.037 7, 比随机

初值的区分度和难度的 *ABSE* 值 0.078 0 和 0.046 6 略小；当链长增加到 1 000 时, IRT 初值的区分度和难度的 *ABSE* 值分别为 0.064 5 和 0.036 7，和随机初值的区分度和难度的 *ABSE* 值 0.065 0 和 0.040 1 差不多.

由表 2 可知：固定被试个数为 1 000 人，当项目容量较小时(如 20 题)，初值作用不明显；当项目容量适中时(如 60 题)，初值作用比较明显；当项目容量较大时(如 100 题)，初值作用明显．如在项目量为 100 时，从 3 种链长情况下的结果可以明显看出：当链长为 100 时, IRT 初值的能力的 *ABSE* 值为 0.135 2, 明显小于随机初值的能力的 *ABSE* 值 0.287 8；当链长增加到 500 时, IRT 初值的能力的 *ABSE* 值为 0.137 5, 比随机初值的能力的 *ABSE* 值 0.148 9 略小；当链长增加到 1 000 时, IRT 初值的能力的 *ABSE* 值为 0.137 6, 和随机初值的能力的 *ABSE* 值 0.137 6 相当.

3　结果与讨论

考察了不同参数初值在不同链长下对于 MCMC 方法参数估计精度的影响．对于 2PLM，实验结果表明：在样本量或项目量增大时，对于任意初值，估计精度明显提高；当样本量或项目量固定且较小时，初值作用不明显，不同初值的估计精度差不多，且估计精度均不高；当样本量或项目量较大时，当链长较短时，初值作用明显，比较准确的初值估计精度更高，随着链长越长(约 500 以上)，各参数估计精度越来越相近．当样本量或项目量较大时通过 IRT 方法估计得到较准的初值，从而在较短的链长下得到较准的估计精度，这在一定程度上弥补了 MCMC 耗时太长的缺陷.

本文只考虑了 2PLM，对于其它模型如 3PLM、多级评分模型，参数初值等因素对于参数的估计精度是否也有影响，还有待进一步的研究.

4　参考文献

[1] 漆书青．现代测量理论在考试中的应用 [M]．武汉：华中师范大学出版社, 2003: 184-187.

[2] 王权．“马尔科夫链蒙特卡洛”(MCMC)方法估计 IRT 模型参数中的应用 [J]．考试研究, 2006(4): 45-63.

[3] Albert J H. Bayesian estimation of normal ogive item response curves using Gibbs sampling [J]. Journal of Educational Statistics, 1992(17): 251-269.

[4] Patz R J, Junker B W. A straightforward approach to Markov chain Monte Carlo methods for item response models [J]. Journal of Educational and Behavioral Statistics, 1999, 24(2): 146-178.

[5] 涂冬波，漆书青. IRT 模型参数估计的新方法: MCMC 算法 [J]. 心理科学, 2008, 31(1): 177-180.

[6] 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: 106-111.

[7] Gentle J E. Elements of computational statistics [M]. Beijing: Science Press, 2006: 39-66.

[8] 龚光鲁，钱敏平．应用随机过程教程 [M]．北京：清华大学出版社, 2003: 191-202.

[9] 茆诗松．王静龙．濮晓龙．高等数理统计 [M]．北京：高等教育出版社, 1998: 444-459.

[10] 王祖俭，黄国兵，丁树良．基于遗传算法的项目反应理论 3PLM 参数估计 [J]．江西师范大学学报：自然科学版, 2005，29(6): 475-477.

The Evaluation for the Impact of Starting Values on MCMC Parameter Estimation of 2PLM

HU Hai[1], GAN Deng-wen[1], WANG Wen-yi[2], DING Shu-liang[1*], FU Hua-jun[2]

(1. College of Computer Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China;
2. School of Psychology, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: For the Markov chain Monte Carlo (MCMC) do parameter estimation due to the chain length is too large and there are defects takes too long, using a Monte Carlo simulation study found that: use the MCMC estimated project reaction theory 2PLM unknown and project parameters, higher initial value when the sample size or project a large amount of shorter chain length, the initial value of the role, the more accurate the estimation accuracy. These findings can remedy the time-consuming defect of MCMC under a certain condition.

Key words: Markov Chain Monte Carlo; two-parameter Logistic model; starting value; chain length; parameter estimation

(责任编辑：冉小晓)