改进的 MCMC 算法—DSY 算法及其在估计 IRT 模型参数中的应用

杜文久* 孙胜亮 原 坤

(西南大学数学与统计学院,重庆,400715)

摘 要 本文首先简要的阐述了 MCMC 算法的思想及在 IRT 参数估计中的操作过程;其次,针对该算法存在的一些问题,提出相应的改进建议;然后,分别运用传统的和改进的 MCMC 算法进行模拟数据分析和比较,结果显示新的方法表现更好;最后总结新方法的优点所在,并指出下一步的研究方向。

关键词 项目反应理论 MCMC 算法 DSY 算法 参数估计 平稳分布

1 问题的提出

项目反应理论(Item Response Theory, IRT)的参数估计问题是项目反应理论面临的一个核心问题,通常被用于估计 IRT 参数的算法主要有 N—R 算法、E—M 算法等等。由于这些算法的特定性质,使得在使用它们解决实际问题时总是存在这样或那样的缺点,特别是在估计多维项目反应模型的参数问题时,上述方法具有一定的局限性。因此,寻找一种新的、有效的算法一直是测验学者十分关注的问题。近年来,随着马尔科夫链蒙特卡洛(Markov Chain Monte Carlo,MCMC)算法在 IRT 参数估计中的成功应用,为解决 IRT 参数估计问题提供了一种新的思路。

MCMC 算法在统计物理学中的应用已有 50 多年的历史,除此之外,MCMC 方法在 Bayes 统计、显著性检验、极大似然估计中也得到了广泛应用。最先将此方法引入到 IRT 参数估计中的是统计学家 Albert(1992),他使用一种特殊的 MCMC 方法—Gibbs 抽样法对正态肩型曲线模型中的参数进行估计,降低了参数估计的难度,并且估计精度较高;Patz 和 Junker(1999a,1999b)应用 MCMC 方法对 IRT 中 Logistic、分部评分等模型中的参数进行了估计,并将此方法拓广应用到含有缺失反应数据、相关反应数据的项目参数估计中;Jiang(2005)在其博士论文中首次将 MCMC 算法的应用领域从单维拓广到了多维,实现了对多维项目反应模型参数的估计;Kim 和 Daniel(2007)对 MCMC 算法给出了更为详细的介绍,包括先验分布的确定、评价迭代链收敛的方法等等,并用 WINBUGS1.4 软件将文中所述方法应用于实证研究;王权(2006)介绍了国外学者关于 MCMC 算法在参数估计中的应用;涂冬波、漆树青、蔡艳、戴海琦和丁树良(2008)应用 MCMC 算法对二、三参数逻辑斯蒂模型的项目参数与能力参数进行了估计,并与 E—M 算法进行了比较;曾莉、辛涛、张淑梅(2009)使用两种 MCMC 缺失数据方法对二参数逻辑斯蒂模型中参数的估计进行了比较等。

MCMC 算法的优点在于它充分发挥了计算机模拟技术的优势,通过计算模拟采集充分大的状态样本,用初等方法估计模型参数,避开了 E－M 算法中的复杂计算,提高了估计的成功率。

然而, MCMC 算法并不是完美无缺的。尽管 MCMC 算法具有算法简单、编程方便等优点,但在实践中人们也发现,当采用 MCMC 算法估计项目参数时,其估计精度较多的受到项目参数先验分布的影响,如果项目参数的先验分布与实际情形不相符或者分布的方差较大,项目参数的估计误差会较大。另外,由于不知道 MCMC 算法产生的马尔科夫链何时收敛于平稳分布,因此链的长度通常都取得很大,一般取为 5000 到 10000,甚至 15000,从而使得算法耗时很长。为了克服 MCMC 算法的这些缺陷,本文试图对 MCMC 算法进行改进,并希望通过本文的讨论,能够为项目反应理论的参数估计提供一种新的解决方法。

2 MCMC 方法的基本思想

MCMC 方法的基本思想是通过建立一个以 $\pi(x)$ 为平稳分布的 Markov 链来得到 $\pi(x)$ 的样本,基于这些样本就可以作出各种统计推断(茆诗松,王静龙,濮晓龙,2006)。MCMC 方法的操作过程大致可

* 通讯作者:杜文久。E-mail:dwj@swu.edu.cn

分为如下三步：

1)在未知参数 x 的状态空间上选一个“合适”的转移核 p,并用这个转移核构造一个 Markov 链,这里“合适”的意思是指 $\pi(x)$ 应是这个链的平稳分布；

2)从状态空间上某一点 $x^{(0)}$ 出发,用 1)中的转移核 p 产生一个马尔科夫序列：$x^{(1)}, x^{(2)}, \cdots, x^{(n)}$；

3)对某个 m 和 n $(n>m)$,任意函数 的期望估计值为

$$\hat{E}_{\pi}f(x)=\frac{1}{n-m}\sum_{i=m+1}^{n}f(x^{(i)})$$

m 通常叫做调试周期, n 叫做链的长度。

为了构造一个以 $\pi(x)$ 为平稳分布的马尔科夫链,转移核 p 的确定至关重要,目前常用的构造马尔科夫链转移核的方法有 Metropolis - Hastings (M - H)、Gibbs 方法等等。在实际中, M - H 方法更普遍。

用 M - H 方法产生未知参数 x 的马尔科夫链的思路如下：

①任意选择一个不可约转移概率 q 以及一个函数 α, $0<\alpha\leq 1$ 。对任一组合 $(x^{(k+1)}, x^{(k)})$ 定义：$p(x^{(k+1)}, x^{(k)})=q(x^{(k+1)}, x^{(k)})\alpha$,则 p 形成一个转移核。

根据建议性转移核产生 Markov 链的下一候选值 $x^{(*)}$,其接受概率为：

$$\alpha=\min\left\{\frac{\pi(x^{(*)})q(x^{(*)}, x^{(k)})}{\pi(x^{(k)})q(x^{(k)}, x^{(*)})}, 1\right\}$$

②在[0,1] 分布上产生一个随机数 u,则：$x^{(k+1)}=\begin{cases}x^{(*)}, u\leq\alpha\\ x^{(k)}, u>\alpha\end{cases}$。

③重复步骤①和②,就得到一条马尔科夫链。

其中建议分布 q 通常取为对称分布。

3 MCMC 算法的改进—DSY 算法

用 MCMC 方法估计 IRT 参数,理论上经过足够多次迭代以后,MCMC 方法总可以找到一条或者多条收敛的 Markov 链,但是在实际应用中有时并不一定能得到。再者,由于不知道 MCMC 算法产生的马尔科夫链何时收敛于平稳分布,因此链的长度通常取 5000 以上,从而使得算法耗时很长。Richard 和 Brian(1997)运用此方法对来自 NAEP 的 6 个短结构反应项目进行参数估计,在迭代运行 7400 次时共花费了 5.7 个小时。因此提高 MCMC 方法的估计效率是 MCMC 算法有待进一步研究的问题。为了解决上述问题,本文提出以下两种改进方法：

改进一：用 MCMC 方法对 IRT 模型中的项目参数进行估计时,首先要确定平稳分布 $\pi(x)$, MCMC 算法选取的 $\pi(x)$ 为参数的联合后验分布,它由所有参数的先验分布及被试作答的似然函数所决定。然而在实践中参数的先验分布有时并不知道,即使知道其先验分布,也常常不知道先验分布中具体的参数值,实践者经常靠经验来确定先验分布和具体的参数值。因此难免会出现根据经验得到的假设与实际情况不符的情况,可见这种取法有一定的局限性。实际上,从广泛意义上说,对于任意一个非负函数 $f(x)$,只要其积分值是个有限值,就可以将 $f(x)$ 作为平稳分布 $\pi(x)$ (龚光鲁,钱敏平,2003)。因此 $\pi(x)$ 可以选取为被试作答的似然函数 $L(X/a,b)$,这种取法可以减少确定先验分布和其参数值的麻烦。

改进二：在用 MCMC 生成迭代链时,需要计算从一个迭代值到下一个迭代值的接受概率 α 。以 x 表示待估参数,在计算其接受概率 $\alpha=\min\left\{\frac{\pi(x')}{\pi(x)}, 1\right\}$ 时,首先需要比较 $\frac{\pi(x')}{\pi(x)}$ 与 1 的大小,当 $\frac{\pi(x')}{\pi(x)}>1$ 时,此时 ,则新的 x' 取代 x 成为迭代链的下一个值；若 $\frac{\pi(x')}{\pi(x)}<1$ 时,此时 $\alpha=\frac{\pi(x')}{\pi(x)}$,则还需要和随机产生的[0,1]均匀分布上数值 u 比较,若 $u<\alpha$ 则 x 转移至 x' ,若 $u>\alpha$ 则不转移,继续从原来的分布中抽取下一个值,直至找到下一个迭代值。上述 MCMC 方法会导致生成整个马尔科夫链需要较长时间,影响了 MCMC 方法的效率。因此建议在选取下一个迭代值时只比较 $\pi(x')$ 与 $\pi(x)$ 的大小。当 $\pi(x')>\pi(x)$ 时则转移,反之,则不转移。这样便不会出现表现比 x 差的点,这样就使得整个链到达最优点的时间大大减少,从而提高 MCMC 方法的使用效率。

根据上述讨论,改进型 MCMC 算法在 IRT 项目参数估计中的操作过程如下：

假设 N 个能力分别为 $\theta_1, \theta_2, \cdots, \theta_N$ 的被试在某个二级评分项目上作答的反应向量为 $X=(X_1, X_2, \cdots, X_N)$,如果被试 j 在项目上答对,则 $X_j=1$,否则 $X_j=0$,$j=1,2,\cdots,N$。

于是根据被试在项目上的反应可得似然函数

$$L(X/a,b)=\prod_{j=1}^{N}p_j^{X_j}(1-p_j)^{1-X_j}$$

其中，$p_j=\frac{1}{1+e^{-Da(\theta_j-b)}}$，被试的能力参数已知，项目参数 a,b 未知。

① 给定 a,b 的初值 $a^{(k)},b^{(k)}$，通常 a 的初值取 1，b 的初值取零；

② 从分布 $N(a^{(k)},0.3)$ 中随机产生 $a^{(k)}$ 的下一个候选值 $a^{(*)}$，若

$L(X/a^{(*)},b^{(k)})>L(X/a^{(k)},b^{(k)})$，则转移，即 $a^{(k+1)}=a^{(*)}$，否则不转移，即 $a^{(k+1)}=a^{(k)}$；

③ 从分布 $N(b^{(k)},0.3)$ 中随机产生 $b^{(k)}$ 的下一个候选值 $b^{(*)}$，若

$L(X/a^{(k+1)},b^{(*)})>L(X/a^{(k+1)},b^{(k)})$，则转移，即 $b^{(k+1)}=b^{(*)}$，否则不转移，即 $b^{(k+1)}=b^{(k)}(k=0,1,2,\cdots,n)$ n 为迭代次数；

④ 重复步骤②、③，于是产生 a,b 的一个迭代点列：

$a^{(0)},b^{(0)},a^{(1)},b^{(1)},a^{(2)},b^{(2)},\cdots,a^{(n)},b^{(n)}$，

更新点的选取来自正态分布，在迭代点列中参数 a,b 交替出现，以点在似然函数上的表现为准则，出现更新点反复抽取三次以上的可能性已非常低，迭代次数 n 取 500，并取最后一对估计值（$a^{(n)},b^{(n)}$），作为未知参数 (a,b) 的估计值。

上述方法的理论原理是：单调有界序列必有极限，由（$a^{(0)},b^{(0)}$），$(a^{(1)},b^{(1)})$，$(a^{(2)},b^{(2)})$，…，$(a^{(n)},b^{(n)})$ 的构造过程知，$L(X/a^{(0)},b^{(0)})\leq L(X/a^{(1)},b^{(1)})\leq,\cdots,\leq L(X/a^{(n)},b^{(n)})$，设 $L_k=L(X/a^{(k)},b^{(k)})$，则序列 $\{L_k\}$ 单调递增，且以似然函数 $L(X/a,b)$ 的最大值为其上确界，因此序列 $\{L_k\}$ 将收敛于似然函数 $L(X/a,b)$ 的最大值，其所对应的项目参数值就是似然函数 $L(X/a,b)$ 的极大似然估计值。

从上述讨讨看到，改进 MCMC 算法其实已经不能叫做 MCMC 算法了，因为采用上述方法产生的迭代点列并不收敛于平稳分布，为了应用上的方便，我们给上述方法取一个名称：DSY 算法。

4 实证研究及结果分析

本文利用 Matlab(2010a) 软件，自编关于文中所述传统的 MCMC 算法和 DSY 算法的程序，比较参数估计值与参数真值间的差异，两值之间的差异越小说明估计方法越有效，反之越无效。以平均绝对离差 $abse(x_j)=\frac{1}{m}\sum_{i=1}^{m}|x_{ij}-\hat{x}_{ij}|$ 为评价指标比较两种算法估计参数的优劣，其中 x_{ij} 为参数真值，$\hat{x}_{ij}$ 为参数估计值，m 为参数个数。

自编 DSY 算法的程序过程简介如下：

(1)随机模拟生成服从先验分布的能力参数、区分度参数、难度参数，运用二参数逻辑斯蒂模型和服从[0,1]均匀分布的随机数产生被试在项目上的反应矩阵；

(2)根据(1)中的反应矩阵运用上文中所述方法得出项目的似然函数；

(3)运用上文中所述改进方法生成项目参数的 Markov 迭代点列；

(4)计算迭代点列的均值并计算估计误差。

MCMC 算法中参数 (a,b) 的先验分布分别取为：$Lna\sim N(0,1)$，$b\sim N(0,1)$，实验假定能力参数已知，且能力参数服从 $\theta\sim N(0,1)$，对每一组项目参数值迭代 5000 次，取后 3000 次迭代的平均值作为初步参数估计值。具体操作过程按照涂冬波等 (2008) 所述方法进行。

为减少随机误差，每种试验均重复 10 次。试验分别考察测验项目数、被试数变化，以及参数先验分布已知、未知时，分别使用两种估计方法估计项目参数时其估计精度的变化，并对两种方法下估计精度的优劣进行比较。具体试验结果如下：

4.1 项目数、被试数变化对两种算法结果的影响

试验一：项目数固定，被试数变化。二参数逻辑斯蒂模型下，固定项目参数，观察被试数变化对估计结果的影响。项目数固定为 60，被试数分别取 200、500、1000，两种算法的试验结果如表一所示：

表一：被试数变化对试验结果的影响(n=60)

估计方法	被试数(m)	参数 a 估计误差	参数 b 估计误差
MCMC 算法	200	.1570	.1851
	500	.1053	.1381
	1000	.0894	.0891
DSY 算法	200	.1448	.2225
	500	.1117	.1198
	1000	.0806	.0827

试验二：被试数固定，项目数变化。二参数逻辑斯蒂模型下，固定被试数，观察项目数变化对估计结果的影响。被试数固定为 1000，项目数分别取 60、80、100，得两种算法的试验结果如表二所示：

表二：项目数变化对试验结果的影响(m=1000)

估计方法	项目数(n)	参数 a 估计误差	参数 b 估计误差
MCMC 算法	60	.0894	.0891
	80	.0826	.0860
	100	.0767	.0822
DSY 算法	60	.0806	.0827
	80	.0802	.0755
	100	.0737	.0710

表一、表二表明，随着项目数或者被试数的增

大，两种估计方法下的估计误差均随之减小，估计精度均有所提高，这符合对估计方法的一般要求。表一显示：项目数固定，被试数较小时，两种估计方法的表现互有优劣，随着被试数的增加，DSY 算法在两个参数估计上的表现均好于 MCMC 算法。表二显示：被试数固定，在项目数相等的情况下，DSY 算法在两个参数估计上的表现均优于 MCMC 方法。

4.2 参数的先验分布对两种 MCMC 算法结果的影响

试验三：参数的的先验分布分别取为：$Lna \sim N(0,0.5)$，$b \sim N(0,1)$，先验分布与试验的模拟分布相符合时，得两种算法下的如下试验结果：

表三：分布一致时对试验结果的影响（$m=1000$ $n=60$）

估计方法	参数 a 估计误差	参数 b 估计误差	所用时间 T(s)
MCMC 算法	.0780	.0874	368.833
DSY 算法	.0764	.0863	42.425

试验四：假设参数 的先验分布为：$Lna \sim N(0,1)$，$b \sim N(0,1)$，实际模拟的先验分布分别取为：$\log(a) \sim U(0,5)$，$b \sim U(-3,3)$。假设先验分布与试验的模拟分布不符合，得两种算法下的试验结果：

表四：分布不一致时对试验结果的影响（$m=1000$ $n=60$）

估计方法	参数 a 估计误差	参数 b 估计误差	所用时间 T(s)
MCMC 算法	.2922	.2139	384.433
DSY 算法	.2676	.1502	46.325

表三和表四表明，无论参数的先验分布与实际分布是否相符，DSY 算法的估计精度都要优于 MCMC 算法，特别是先验分布与实际情形不相符时，DSY 算法的优势更加明显。

表三和表四的最大差别在于实现算法所需的时间，DSY 算法所用时间远小于 MCMC 算法，这也实现了改进算法设计的初衷。原因在于改进算法在选取 Markov 链上的迭代值时，只选取相对于前一个迭代值而言表现更好的值，因此可以更快的达到或接近极大似然函数值点，试验表明采用改进算法达到或接近极大值点，所需次数远小于 5000 次。虽然所用时间较短，但从两表中可以看出，新算法的参数估计误差反而更小，这是改进方法的又一个优势。

总体看，项目数或被试数较大，以及待估参数的先验分布或其参数值未知时，在试验所需时间大大减少的情况下，使用 DSY 算法所得到的参数估计值更好。

为探明 DSY 算法所需的合适的迭代次数，作者曾取迭代次数分别为 100、500、1000 进行试验。结果发现：当迭代次数取 100 时，DSY 算法的估计精度即已优于 MCMC 算法，而迭代次数取 500 时，DSY 算法的精度进一步提高，但是当将迭代次数取 1000 时，DSY 算法的精度几乎已不再改善。根据上述实验，我们认为 DSY 算法的迭代次数取 500 较为合适。

5 总结与讨论

文章主要讨论了 MCMC 算法的改进及其与传统 MCMC 算法在参数估计效果上的比较。试验结果表明，DSY 算法相较于 MCMC 方法不仅所需运算时间大大减少，而且估计精度也更高。

与 MCMC 算法相比，DSY 算法具有下述优点：

（1）编程简单

DSY 算法充分保留了 MCMC 算法编程简单的优点。实际上，由于使用 DSY 算法时不需要知道未知参数的先验分布，并且减少了产生随机数的过程，因此其编程比 MCMC 算法更简单。

（2）DSY 算法的估计值将收敛于真值

DSY 算法求得的估计值是未知参数的极大似然估计值，因此该算法求得的估计值具有极大似然估计值的所有性质，如相合性、渐近正态性、渐近有效性等等。

（3）计算速度快

DSY 算法迭代链的长度远比 MCMC 算法要短，因此参数估计的的计算时间也大为缩短。但项目参数的估计误差相比 MCMC 算法却更小，这为 DSY 算法在大范围的教育测评中得以应用创造了条件，同时也为项目反应理论的参数估计提供了一种新的选择。

本文只研究了在能力参数已知时，两种算法在项目参数估计精度上的表现。当项目参数已知时，两种算法对能力参数估计精度的影响可以使用同样的方法予以检验，结论是一样的，限于篇幅及避免方法雷同，本文在此就不加以探讨了。

影响 MCMC 方法估计效率的因素除了平稳分布的形式外，还有迭代所选择的初始值。初始值对使用传统 MCMC 方法进行参数估计的影响国内已有相关研究。将选择初始值的方法与 DSY 算法相结合预计会收到更好的估计结果，这方面的工作连同迭代链何时收敛、如何检测迭代链已收敛，对于进一步缩短算法的运算时间尤为重要，这也是本研究之后要解决的问题。

参考文献

龚光鲁，钱敏平.（2003）. *应用随机过程* 北京：清华大学出版社
茆诗松，王静龙，濮晓龙.（2006）. *高等数理统计（第二版）*. 北京：

高等教育出版社.

涂冬波，漆树青，蔡艳，戴海琦，丁树良.（2008）. *IRT 模型参数估计的新方法—MCMC 算法* 心理科学，31(1)，177 - 180.

王 权.（2006）. “马尔科夫链蒙特卡洛”(MCMC)*方法在估计IRT 模型参数中的应用* *考试研究*，2 (4)，45 - 63.

曾莉，辛涛，张淑梅.（2009）. 2PL 模型的两种马尔科夫蒙特卡洛缺失数据处理方法比较，*心理学报*，*41* (3)，276 - 282.

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling . *Journal of Education Statistics*, *17*, 251 - 269.

Jiang Y. L. (2005). *Estimating parameters for multidimensional item response theory models by MCMC methods.* Unpublished Doctoral Dissertation, Michigan State University.

Kim J. S., & Daniel M. B. (2007). Estimating item response theory models using Markov Chain Monte Carlo Methods. *Educational Measurement: Issues and Practice*, *26*(4), 38 - 51.

Richard, J. P., & Bria, W. J. (1999a). A straightforward approach to Markov Chain Monte Carlo Methods for Item Response Models. *Journal of Educational and Behaviorial Statistics*, *24*(2), 146 - 178.

Richard, J. P., & Brian, W. J. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of Educational and Behaviorial Statistics*, *24*(4), 342 - 366.

Improvement of MCMC Algorithm - DSY Algorithm and its Application in Parameter Estimation under the IRT Model

Du Wenjiu, Sun Shengliang, Yuan Kun

(School of Mathematics and Statistics, Southwest University, Changqing, 400715)

Abstract The core issue in IRT is how to estimate the item and person parameters. The common methods used often were the N - R algorithm and the E - M algorithm. Because of their special characteristics in themselves, there were always certain shortcomings to hinder their development. With more and more complex models appearing, it is also difficult to estimate the parameters using those methods. Then, the algorithm called Markov Chain Monte Carlo (MCMC) appeared. The emergence of MCMC algorithm provides the new solution.

The MCMC algorithm has been used in statistical physics for more than 50 years. In recent 20 years, it was also widely used in Bayesian estimation, test of significance and maximum likelihood estimation. Albert (1992) is the first statistician to apply the algorithm in IRT parameter estimations. Many experts including Albert (1992), Patz and Junker (1999a, 1999b), Kim, Seon and Daniel, Bolt (2007) provide the information about the MCMC algorithm and how to use it in detail.

The character of the MCMC algorithm is that it gives full play to the advantages of computer simulation technology, collects a sufficiently large sample of state by simulating, uses the elementary method to estimate the model parameters, and thus bypasses the complex calculation of the EM algorithm to improve the success rate of estimation.

No algorithm is perfect. Although the traditional MCMC algorithm has been widely used, its shortcomings, such as the serious dependence on the prior distribution of the parameters and the extremely long time spent in performing the procedure, still exist. It is the main purpose of this paper to solve the problems.

In the paper, the idea of the MCMC algorithm is briefly introduced.

Then two suggestions are made to improve the algorithm and solve the existing problems. The first suggestion is about the stationary distribution: the traditional MCMC algorithm is largely dependent on the prior distribution of the model parameters. However, in practice, researchers often do not know the prior distribution. So this influences the accuracy of the estimation results. In this paper, we provide another method to avoid the above situation. The second suggestion is about the acceptance probability. In this paper, we believe that in order to reduce the run time of the procedure, only when the stationary distribution value of the new iteration value is greater than that of the original iteration value, does the new iteration value replace the old one to be the value of the iterative chain.

Then, the traditional and improved versions of MCMC algorithm are used to simulate and analyze the data. Through the comparison of the results from the two methods, it shows that the new method performs better. Finally, the advantage of the new algorithm is pointed out and the future research direction is suggested.

Key words item response theory, MCMC algorithm, DSY algorithin parameter estimation, stationary distribution