汪文义,郑娟娟,宋丽红,等.认知诊断模型参数估计算法比较[J].江西师范大学学报(自然科学版),2022,46(3):291-299.
WANG Wenyi,ZHENG Juanjuan,SONG Lihong,et al.The comparison of parameter estimation of cognitive diagnosis models[J].Journal of Jiangxi Normal University(Natural Science),2022,46(3):291-299.

文章编号:1000-5862(2022)03-0291-09

认知诊断模型参数估计算法比较

汪文义[1],郑娟娟[1],宋丽红[2],胡海洋[1]
(1.江西师范大学计算机信息工程学院,江西 南昌 330022;2.江西师范大学教育学院,江西 南昌 330022)

摘要:该文在不同条件的组合下考查了 EM 算法和 MCMC 算法对 3 种常用的认知诊断模型(DINA 模型、DINO 模型和 G-DINA 模型)的参数估计返真性问题.借助项目参数或作答概率分布的偏差、均方根误差、平均绝对离差以及被试的平均属性判准率等指标,评价这 2 类算法的表现.模拟研究结果表明:MCMC 算法更适用于低质量题目、小样本、测验短的条件,而在其他条件下 EM 算法的表现与 MCMC 算法的表现相当.

关键词:认知诊断模型;EM 算法;MCMC 算法;参数估计
中图分类号:B 841.7 **文献标志码**:A **DOI**:10.16357/j.cnki.issn1000-5862.2022.03.12

0 引言

认知诊断评估(cognitive diagnostic assessment)是 20 世纪中后期兴起的一种新的测验方式,它以心理学、教育学为指导,同时借助现代测量学,对被试答错的项目进行深入分析与解释,能够分析被试对项目所考察的所有知识点(属性)的掌握情况,基于此可以对症下药制定补救措施,加强短板的学习.认知诊断评估通常借助于认知诊断模型(cognitive diagnostic model,CDM),根据设定好的 Q 矩阵,建立被试的属性掌握情况与项目反应之间的关系[1].

现有的认知诊断模型有 100 种以上,主要分为简化模型、饱和模型 2 个大类,其中简化模型又分为补偿模型和非补偿模型,它们的最常用模型分别为 DINO 模型[2]和 DINA 模型[3],在饱和模型中最常用的模型为 G-DINA 模型.然而要想借助于认知诊断模型实现认知诊断,就必然离不开认知诊断模型的参数估计方法,故参数估计对于模型的发展与实现有着非常重要的意义.

J. de la Torre 等[4]将马尔可夫链蒙特卡罗方法(Monte Carlo Markov chain,MCMC)用于高阶 DINA 模型参数估计.J.L. Templin 等[2]也将 MCMC 方法用于 DINO 模型参数估计.J. de la Torre[3]提出 DINA 模型 EM 参数估计算法并考查了 EM 算法的表现,但没有通过模拟研究去比较 EM 参数估计算法和 MCMC 参数估计算法的表现.S. Sen 等[5]对可用于 DINA 模型参数估计(EM 方法)的 6 个软件表现进行比较,结果显示各软件表现基本相当.J. de la Torre[6]提出 G-DINA 模型及其 EM 估计算法.有研究将变分贝叶斯推断方法用于 DINA 模型和 G-DINA模型参数估计[7-8].

Zhan Peida 等[9]为 G-DINA 模型(与对数线性认知诊断模型 LCDM 等价)提供了 MCMC 估计的 JAGS 代码.S. A. Culpepper 等[10]基于数据扩张的 MCMC 方法,并采用吉布斯方法对属性条件概率下的知识状态进行取样,用于简化的重新参数化融合模型参数估计,以提高 MCMC 效率.随后,Wang Juntao 等[11]将其推广用于 G-DINA 模型参数估计,然而也没有与 EM 算法比较.Jiang Zhehan 等[12]还使用 MCMC 的其他变形形式实现了 G-DINA 模型的参数估计.

通过对已有文献分析发现,这 2 类方法各有优劣,如 EM 算法(或变分法)收敛速度较快,但会受

收稿日期:2021-12-16
基金项目:国家自然科学基金(62067005,61967009)资助项目.
作者简介:汪文义(1983—),男,湖南衡山人,教授,博士,主要从事教育测量与信息处理的研究.E-mail:wenyiwang@jxnu.edu.cn

初值影响且可能收敛到局部最优值[6],而 MCMC 算法比较费时但不依赖于初值.于是,有必要在不同样本量、不同测验长度、不同题目质量的条件组合下,对 DINA 模型、DINO 模型、GDINA 模型下的 EM 算法和 MCMC 算法的表现进行系统比较.因为认知诊断模型常规的 EM 算法(非变分法)可借助于各种软件包(如 R 语言实现的 CDM 或 GDINA)来实现,并且 JAGS 代码实现认知诊断模型的 MCMC 十分简单,所以,本文主要考虑在认知诊断中常规的 EM 算法和 JAGS 代码实现的 MCMC 算法,不考虑它们的其他变形方法.在简要介绍了认知诊断模型下的 EM 算法和 MCMC 算法后,通过模拟研究,在不同样本量、不同测验长度、不同题目质量的条件组合下,在 DINO 模型、DINA 模型和 GDINA 模型下比较了 EM 算法和 MCMC 算法的表现,可供读者根据不同条件为各个模型选择最适合的参数估计算法.

1 模型与方法

1.1 认知诊断模型

(i) DINA 模型.确定性输入噪音与门(DINA)模型是认知诊断中运用最广的模型,同时也是一种完全非补偿的认知诊断模型,它要求被试需掌握项目所考察的所有属性,缺少其中任何一个属性都会降低答对项目的概率[6].DINA 模型的项目反应函数为

$$P_j(\boldsymbol{\alpha}_i)=P(X_{ij}=1\mid\boldsymbol{\alpha}_i)=g_j^{1-\eta_{ij}}(1-s_j)^{\eta_{ij}},$$

其中 $P_j(\boldsymbol{\alpha}_i)$ 表示在给定知识状态 $\boldsymbol{\alpha}_i$ 的条件下,被试 $i(i=1,2,\cdots,N)$ 在项目 $j(j=1,2,\cdots,M)$ 上正确作答的条件概率,$\eta_{ij}=\prod_{k=1}^{K}\alpha_{ik}^{q_{kj}}$ 表示在知识状态 $\boldsymbol{\alpha}_i$ 条件下项目 j 的理想反应(其取值为 0 或 1),s_j、g_j 和 $\boldsymbol{q}_j$ 为项目 j 失误参数、猜测参数和所测的属性向量.

(ii) DINO 模型.确定性输入噪音或门(DINO)模型恰好与 DINA 模型相反,在该模型中被试只要掌握了项目考察的任何一个属性就能正确答对该项目,即属性间可以相互替代或补偿[7].DINO 模型的项目反应函数为

$$P(X_{ij}=1\mid\boldsymbol{\alpha}_i)=(1-s_j)^{w_{ij}}g_j^{1-w_{ij}},$$

其中 $w_{ij}=1-\prod_{k=1}^{K}(1-\alpha_{ik})^{q_{jk}}$,$w_{ij}$ 表示被试 i 是否至少掌握项目 j 测量的 1 个属性,其取值为 0 或 1.$w_{ij}=1$ 表明被试 i 至少掌握了项目 j 测量的 1 个属性,$w_{ij}=0$ 表明被试未掌握项目 j 测量的所有属性.

(iii) G-DINA 模型.拓广的 DINA(G-DINA)模型与 DINA 模型不同之处在于:DINA 在每题上只能区分 2 类不同属性掌握模式,而 G-DINA 可以区分 $2^{K_j^*}$ 种属性掌握状态,K_j^* 表示第 j 题所考察的属性个数.G-DINA 假设不同的属性掌握模式具有不同的答对概率[8].G-DINA 有 3 种连接函数,由于本文只采用了其中的基于 logit 的连接函数(logit link function),所以对其他 2 种函数将不做详细介绍.基于 logit 连接函数的 G-DINA 公式为

$$f_{\text{logit}}(P(\boldsymbol{\alpha}_{lj}^*))=\delta_{j0}+\sum_{k=1}^{K_j^*}\delta_{jk}\alpha_{lk}+\sum_{k'=k+1}^{K_j^*}\sum_{k=1}^{K_j^*-1}\delta_{jkk'}\cdot\alpha_{lk}\alpha_{lk'}+\cdots+\delta_{j12\cdots K_j^*}\prod_{k=1}^{K_j^*}\alpha_{lk},$$

其中 δ_{j0} 为项目 j 的截距,即被试在未掌握项目测量的所有属性时题目的 logit 值,类似于难度参数,其取值可正可负;δ_{jk} 为在项目 j 上属性 k 的主效应,δ_{jk} 表示掌握属性 k 对答对该题的贡献值,$\delta_{jkk'}$ 为在项目 j 上属性 k 与属性 k' 的交互效应,$\delta_{j12\cdots K_j^*}$ 为项目 j 测量所有属性间的交互效应.

1.2 EM 算法

EM 算法是一种迭代优化策略,该策略在每次迭代中都分成了 2 个步骤:第 1 步是期望步,简称为 E 步;第 2 步为极大步,简称为 M 步.EM 算法是数据挖掘经典算法之一,它最初是为了解决在数据缺失情况下进行参数估计的问题而提出的,后来被用于寻找在缺失数据统计模型中参数的最大似然估计或者最大后验估计[9].

用符号化表示:记 $f(X,\boldsymbol{\alpha}\mid\xi)$ 为完全数据$(X,\boldsymbol{\alpha})$ 的联合似然函数,其中 $\xi=(\boldsymbol{\varphi},\boldsymbol{\beta})$ 表示待估计的知识状态 $\boldsymbol{\alpha}$ 的分布参数和项目参数.在 EM 算法中,第 t 步迭代主要分为 E 步(计算期望)和 M 步(将期望极大化),其中 E 步:计算 $Q(\boldsymbol{\varphi},\boldsymbol{\beta}\mid\boldsymbol{\varphi}^{(t)},\boldsymbol{\beta}^{(t)})=E((\ln f(\boldsymbol{x},\boldsymbol{\alpha}\mid\boldsymbol{\varphi},\boldsymbol{\beta})\mid\boldsymbol{x},\boldsymbol{\varphi}^{(t)},\boldsymbol{\beta}^{(t)}))$,M 步:$(\boldsymbol{\varphi}^{(t+1)},\boldsymbol{\beta}^{(t+1)})=\arg\max_{\boldsymbol{\varphi},\boldsymbol{\beta}\in\Omega}Q(\boldsymbol{\varphi},\boldsymbol{\beta}\mid\boldsymbol{\varphi}^{(t)},\boldsymbol{\beta}^{(t)})$;即在第 t 步迭代的循环中,期望步是在给定 $\xi^{(t)}$ 条件下计算联合似然函数的期望 $E((\ln f(\boldsymbol{x},\boldsymbol{\alpha}\mid\boldsymbol{\varphi},\boldsymbol{\beta})\mid\boldsymbol{x},\boldsymbol{\varphi}^{(t)},\boldsymbol{\beta}^{(t)}))$,极大化步即最大化 $Q(\boldsymbol{\varphi},\boldsymbol{\beta}\mid\boldsymbol{\varphi}^{(t)},\boldsymbol{\beta}^{(t)})$ 以估计得到 $\xi^{(t+1)}$,重复这个过程直至收敛.

1.2.1 DINA、DINO 模型的 EM 算法

通过上述介绍的 EM 算法,给定知识状态 $\boldsymbol{\alpha}_i$,在条件独立假设下(即同一被试对各个项目的作答是相互独立的),被

试在 M 个项目上的得分向量为 $X_i=(x_{i1},x_{i2},\cdots,x_{iM})$，在 N 个被试的作答模式相互独立假设下，得分矩阵 X 的条件似然函数为

$$L(X\mid\boldsymbol{\alpha})=\prod_{i=1}^{N}\prod_{j=1}^{M}P_j(\boldsymbol{\alpha}_i)^{x_{ij}}(1-P_j(\boldsymbol{\alpha}_i))^{1-x_{ij}}.$$

将所有被试看成是来自同一母体，且给定属性掌握模式的先验分布为 $P(\boldsymbol{\alpha}_c)$，$\boldsymbol{\alpha}_c\in Q_s$，其中 Q_s 为知识状态全集，如在含 K 个属性的独立结构下 Q_s 包含 2^K 列，于是可得边际似然函数(边际分布)为

$$L(X)=\prod_{i=1}^{N}L(X_i)=\prod_{i=1}^{N}\sum_{\boldsymbol{\alpha}_c\in Q_s}L(X_i\mid\boldsymbol{\alpha}_c)P(\boldsymbol{\alpha}_c),$$

即在 DINA 和 DINO 模型上边际似然函数表示为

$$L(X)=\prod_{i=1}^{N}L(X_i)=\prod_{i=1}^{N}\sum_{\boldsymbol{\alpha}_c\in Q_s}L(X_i\mid\boldsymbol{\alpha}_c)P(\boldsymbol{\alpha}_c)=$$

$$\prod_{i=1}^{N}\sum_{\boldsymbol{\alpha}_c\in Q_s}\prod_{j=1}^{M}((1-\hat{s}_j)^{\eta_{jl}}\hat{g}_j^{1-\eta_{jl}}),$$

其中 η_{jl} 在 DINA 模型上表示为 $\eta_{jl}=\prod_{k=1}^{K}\alpha_{lk}^{q_{jk}}$；在 DINO 模型上只需要将 w_{jl} 变为 η_{jl}，$\eta_{jl}=1-\prod_{k=1}^{K}(1-\alpha_{lk})^{q_{jk}}$，进而对边际似然函数里的项目参数进行求导，为了方便，令 β_{jv} 为项目参数. 当 $v=0$ 时，$\beta_{j0}=g_j$，当 $v=1$ 时，$\beta_{j1}=s_j$，即

$$\partial L(X)/\partial\beta_{jv}=\sum_{i=1}^{N}(\partial L(X_i)/\partial\beta_{jv})/L(X_i)=$$

$$\sum_{i=1}^{N}\sum_{l=1}^{2^K}p(\boldsymbol{\alpha}_l)\,\partial L(X_i\mid\boldsymbol{\alpha}_l)/\partial\beta_{jv}/L(X_i).$$

经过一系列公式推导，第 1 步给定初值 $\boldsymbol{\beta}=(s_1,g_1,s_2,g_2,\cdots,s_J,g_J)$；第 2 步根据已有的 $\boldsymbol{\beta}$ 以及用下列式子分别计算出 $I_j^{(0)}$、$R_j^{(0)}$、$I_j^{(1)}$、$R_j^{(1)}$：

$$P(\boldsymbol{\alpha}_l\mid X_i)=L(X_i\mid\boldsymbol{\alpha}_c)p(\boldsymbol{\alpha}_c)/(\sum_{\boldsymbol{\alpha}_l\in Q_s}L(X_i\mid\boldsymbol{\alpha}_l)p(\boldsymbol{\alpha}_l)),$$

$$R_{jl}=\sum_{i=1}^{N}P(\boldsymbol{\alpha}_l\mid X_i)X_{ij},I_l=\sum_{i=1}^{N}P(\boldsymbol{\alpha}_l\mid X_i),$$

其中 $R_j^{(0)}$ 为 $(1-\eta_{jl})R_{jl}$ 的行和，同时 $R_j^{(1)}$ 为 $\eta_{jl}R_{jl}$ 的行和，

$$I_j^{(0)}=\sum_{\boldsymbol{\alpha}_c\in Q_s}(1-\eta_{jl})I_l,I_j^{(1)}=\sum_{\boldsymbol{\alpha}_c\in Q_s}\eta_{jl}I;$$

第 3 步用第 2 步计算得到的 $I_j^{(0)}$、$R_j^{(0)}$、$I_j^{(1)}$、$R_j^{(1)}$ 值依据下列公式计算新的 $\boldsymbol{\beta}$ 值：

$$\hat{s}_j=(I_j^{(1)}-R_j^{(1)})/I_j^{(1)},\hat{g}_j=R_j^{(0)}/I_j^{(0)};$$

第 4 步重复第 2、第 3 步，直到 $\boldsymbol{\beta}$ 的每个分量都满足收敛准则，最后得出估计的项目参数 $\hat{s}_j$ 和 $\hat{g}_j$.

1.2.2　G-DINA 模型的 EM 算法　在 G-DINA 模型上边际似然函数表示为

$$L(X)=\log(L(X))=\log\prod_{i=1}^{N}\sum_{l=1}^{2^K}L(X_i\mid\boldsymbol{\alpha}_l)q(\boldsymbol{\alpha}_l),$$

其中似然函数 $L(X_i\mid\boldsymbol{\alpha}_l)=\prod_{j=1}^{J}P(\boldsymbol{\alpha}_{lj}^*)^{X_{ij}}(1-P(\boldsymbol{\alpha}_{ij}^*))^{(1-X_{ij})}$，$q(\boldsymbol{\alpha}_l)$ 为知识状态的先验分布，在 G-DINA 模型中答对第 j 题的概率也可表示为 $P(\boldsymbol{\alpha}_{lj}^*)$，通过数理推导的 $P(\boldsymbol{\alpha}_{lj}^*)$ 的边际极大似然估计可表示为

$$\hat{P}(\boldsymbol{\alpha}_{lj}^*)=R_{\boldsymbol{\alpha}_{lj}^*}/I_{\boldsymbol{\alpha}_{lj}^*},\tag{1}$$

其中 $I_{\boldsymbol{\alpha}_{lj}^*}=\sum_{i=1}^{N}P(\boldsymbol{\alpha}_{lj}^*\mid X_i)$ 是所有被试在 $\boldsymbol{\alpha}_{lj}^*$ 知识状态下的作答概率，$R_{\boldsymbol{\alpha}_{lj}^*}=\sum_{i=1}^{N}x_{ij}P(\boldsymbol{\alpha}_{lj}^*\mid X_i)$ 是所有被试在 $\boldsymbol{\alpha}_{lj}^*$ 知识状态下答对第 j 题的概率. 其中 $P(\boldsymbol{\alpha}_{lj}^*\mid X_i)$ 为被试 i 在 $\boldsymbol{\alpha}_{lj}^*$ 上的后验概率，$P(\boldsymbol{\alpha}_{lj}^*\mid X_i)$ 可表示为

$$P(\boldsymbol{\alpha}_{lj}^*\mid X_i)=L(X_i\mid\boldsymbol{\alpha}_{lj}^*)q(\boldsymbol{\alpha}_{lj}^*)/(\sum_{l=1}^{2^{K_j^*}}L(X_i\mid\boldsymbol{\alpha}_{lj}^*)q(\boldsymbol{\alpha}_{lj}^*)).\tag{2}$$

在 G-DINA 模型下第 j 题的饱和 Q 矩阵为 $\boldsymbol{M}_j^{(s)}$，矩阵 $\boldsymbol{M}_j^{(s)}$ 的第 1 列全为 1，第 2 列到第 K_j^*+1 列是 K_j^* 个属性所有可能的考核模式，最后一列元素由 $\prod_{k=1}^{K_j^*}\alpha_{lk}$ 获得，现在假设第 j 题考察了 2 个属性，此时的矩阵 $\boldsymbol{M}_j^{(s)}$ 为

$$\boldsymbol{M}_{j(4\times4)}^{(s)}=\begin{pmatrix}1&0&0&0\\1&1&0&0\\1&0&1&0\\1&1&1&1\end{pmatrix}.\tag{3}$$

同时定义 $\boldsymbol{P}_j=(P(\boldsymbol{\alpha}_{lj}^*))$，要想得到估计值 $\hat{\boldsymbol{P}}_j$，即需要通过最小二乘算法找到项目参数 $\boldsymbol{\delta}_j=(\delta_{j0},\delta_{j1},\cdots,\delta_{jK_j^*},\delta_{j12},\cdots,\delta_{j12\cdots K_j^*})$ 的估计值，即

$$\hat{\boldsymbol{\delta}}=(\boldsymbol{M}_j^{(s)\prime}\boldsymbol{M}_j^{(s)})^{-1}\boldsymbol{M}_j^{(s)\prime}(f_{\text{logit}}(\hat{\boldsymbol{P}}_j)).\tag{4}$$

最后将式(2)～(3)求得的数据代入式(1)来求得极大似然估计值，进而得到相对应的项目参数估计值.

1.3　MCMC 算法

MCMC 是马尔可夫链与蒙特卡罗方法的结合，即构造合适的马尔可夫链进行抽样并使用蒙特卡罗

方法进行积分计算,进而使马尔可夫链最后达到收敛平稳的状态,Metropolis 方法与 W.K. Hastings 的概括奠定了 MCMC 方法的基石[13-14].此方法在以 $\pi(x)$ 为平稳分布的马氏链上产生相互依赖的样本,即 MCMC 方法在本质上是一个蒙特卡罗综合程序,它的随机样本的产生与一条马氏链有关.基于条件分布的迭代取样是另一种重要的 MCMC 方法,其中最著名的特殊情况就是 Gibbs 抽样,它现已成为统计计算的标准工具,其最吸引人的特征是其潜在的马氏链是通过分解一系列条件分布建立起来的. W.K. Hastings[14] 推广了 Metropolis 采样法,在吉布抽样中采用 Metropolis 采样法对条件分布进行抽样,可以避免计算条件分布的正规化常数现象,被称为 MH within Gibbs[15].

本文在这 3 个模型下的 MCMC 估计均采取吉布斯的 MH 采样法,通过 Gibbs 采样法对未知参数的联合后验分布进行分解,然后应用 MH 采样法对各个参数的全条件分布进行取样,以构建未知参数后验分布为平稳分布的可逆的马尔可夫链,从而进行取样,这样就可以将取样所得到的样本分布近似为目标分布,通过对近似分布推断来实现参数估计.在 DINA 模型、DINO 模型以及 G-DINA 模型上分别采用 MCMC 算法进行项目参数估计,具体步骤如下:

(i) 确立未知参数的联合后验分布为目标分布(平稳分布).在局部独立性假设满足情况下,DINA 模型的联合似然函数为

$$L(\boldsymbol{s},\boldsymbol{g},\boldsymbol{\alpha})=\prod_{i=1}^{N}\prod_{j=1}^{M}((1-s_j)^{\eta_{ij}}g_j^{1-\eta_{ij}})^{x_{ij}}(s_j^{\eta_{ij}}(1-g_j)^{1-\eta_{ij}})^{1-x_{ij}},$$

DINO 模型的联合似然函数为

$$L(\boldsymbol{s},\boldsymbol{g},\boldsymbol{\alpha})=\prod_{i=1}^{N}\prod_{j=1}^{M}((1-s_j)^{w_{ij}}g_j^{1-w_{ij}})^{x_{ij}}(s_j^{w_{ij}}(1-g_j)^{1-w_{ij}})^{1-x_{ij}},$$

GDINA 模型的联合似然函数为

$$L(\boldsymbol{\delta},\boldsymbol{\alpha})=\prod_{i=1}^{N}\prod_{j=1}^{M}P(\boldsymbol{\alpha}_{lj}^{*})^{x_{ij}}(1-P(\boldsymbol{\alpha}_{lj}^{*}))^{(1-x_{ij})}.$$

在贝叶斯框架中,给定得分矩阵的条件,所有参数的联合后验分布(即平稳分布)为

$$P(\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g}\mid X)\propto p(X\mid\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g})P(\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g})\propto p(X\mid\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g})p(\boldsymbol{\alpha})p(\boldsymbol{s})p(\boldsymbol{g}),$$

$$P(\boldsymbol{\alpha},\boldsymbol{\delta}\mid X)\propto p(X\mid\boldsymbol{\alpha},\boldsymbol{\delta})P(\boldsymbol{\alpha},\boldsymbol{\delta})\propto p(X\mid\boldsymbol{\alpha},\boldsymbol{\delta})p(\boldsymbol{\alpha})p(\boldsymbol{\delta}).$$

在给定得分矩阵和其余参数条件下,各个参数的全条件分布为

$$p(\boldsymbol{\alpha}\mid X,\boldsymbol{s},\boldsymbol{g})\propto L(\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g})p(\boldsymbol{\alpha}),p(\boldsymbol{g}\mid X,\boldsymbol{\alpha},\boldsymbol{s})\propto L(\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g})p(\boldsymbol{g}),p(\boldsymbol{s}\mid X,\boldsymbol{\alpha},\boldsymbol{g})\propto L(\boldsymbol{\alpha},\boldsymbol{s},\boldsymbol{g})p(\boldsymbol{s}),p(\boldsymbol{\delta}\mid X,\boldsymbol{\alpha})\propto L(\boldsymbol{\alpha},\boldsymbol{\delta})p(\boldsymbol{\delta}),p(\boldsymbol{\alpha}\mid X,\boldsymbol{\delta})\propto L(\boldsymbol{\alpha},\boldsymbol{\delta})p(\boldsymbol{\alpha}),$$

其中 $p(\boldsymbol{\delta})=\prod_{j=1}^{J}\prod_{r=0}^{2^{K_j^*}-1}p(\delta_{jr})$,$\boldsymbol{\delta}_j=(\delta_{j0},\delta_{j1},\cdots,\delta_{jK_j^*},\delta_{j12},\cdots,\delta_{j12\cdots K_j^*})$,同时给定每个参数的先验分布[8]如下:

$\alpha_{ik}\sim\mathrm{Bernoulli}(0.5)$,$s\sim\mathrm{Beta}(1,1)$,$g\sim\mathrm{Beta}(1,1)\in(0,1-s)$,$\delta_0\sim N(-1.096,0.25)$,$(\delta_{j1},\cdots,\delta_{jK_j^*},\delta_{j12},\cdots,\delta_{j12\cdots K_j^*})\sim N(0,0.25)$.

(ii) 设定各参数初值,即按各参数的先验分布生成参数初值.

(iii) 从各参数的全条件分布中进行 MH 抽样,分别对知识状态、失误和猜测参数进行取样.首先,对于 $\boldsymbol{\alpha}$,$\alpha_{ik}^{(t+1)}$ 从其先验分布中随机抽取,则 $\alpha_{ik}^{(t)}$ 向 $\alpha_{ik}^{(t+1)}$ 转移概率计算如下:

$$p(\alpha_{ik}^{(t)},\alpha_{ik}^{(t+1)})=\min(L(\boldsymbol{s}^{(t)},\boldsymbol{g}^{(t)},\boldsymbol{\alpha}_{ik}^{(t+1)})/(L(\boldsymbol{s}^{(t)},\boldsymbol{g}^{(t)},\alpha_{ik}^{(t)})),1),\tag{5}$$

$$p(\alpha_{ik}^{(t)},\alpha_{ik}^{(t+1)})=\min(L(\boldsymbol{\delta}^{(t)},\alpha_{ik}^{(t+1)})/(L(\boldsymbol{\delta}^{(t)},\alpha_{ik}^{(t)})),1).\tag{6}$$

将式(5)和式(6)计算出的转移概率与随机数 $r\sim U(0,1)$ 进行比较,若转移概率大于等于 r 则接受转移,否则不转移.

其次,$\boldsymbol{g}$、$g_j^{(t+1)}$ 和 $\boldsymbol{s}$、$s_j^{(t+1)}$ 都从其先验分布中随机抽取,同理可得

$$p(g_j^{(t)},g_j^{(t+1)})=\min(L(\boldsymbol{s}^{(t)},\boldsymbol{g}^{(t+1)},\alpha_{ik}^{(t+1)})p(\boldsymbol{g}^{(t+1)})/(L(\boldsymbol{s}^{(t)},\boldsymbol{g}^{(t)},\alpha_{ik}^{(t+1)})p(\boldsymbol{g}^{(t)})),1),$$

$$p(s_j^{(t)},s_j^{(t+1)})=\min(L(\boldsymbol{s}^{(t+1)},\boldsymbol{g}^{(t+1)},\alpha_{ik}^{(t+1)})p(\boldsymbol{s}^{(t+1)})/(L(\boldsymbol{s}^{(t)},\boldsymbol{g}^{(t+1)},\alpha_{ik}^{(t+1)})p(\boldsymbol{s}^{(t)})),1),$$

$$p(\delta_{j0}^{(t)},\delta_{j0}^{(t+1)})=\min(L(\delta_{j0}^{(t+1)},\alpha_{ik}^{(t+1)})p(\delta_{j0}^{(t+1)})/(L(\delta_{j0}^{(t)},\alpha_{ik}^{(t+1)})p(\delta_{j0}^{(t)})),1).$$

由于 $\boldsymbol{\delta}_j=(\delta_{j0},\delta_{j1},\cdots,\delta_{jK_j^*},\delta_{j12},\cdots,\delta_{j12\cdots K_j^*})$,所以 $p(\delta_{j0}^{(t)},\delta_{j0}^{(t+1)})=\min(L(\delta_{j0}^{(t+1)},\alpha_{ik}^{(t+1)})p(\delta_{j0}^{(t+1)})/(L(\delta_{j0}^{(t)},\alpha_{ik}^{(t+1)})p(\delta_{j0}^{(t)})),1)$ 只是其中一个,其他的转移概率计算公式以此类推.计算出的转移概率都将与随机数 $r\sim U(0,1)$ 比较,若转移概率大于等于 r 则接受转移,否则不转移.

1.4 知识状态估计

在 EM 算法估计出项目参数后,可采用属性的边际后验概率(MPPE/EAP)估计被试知识状态(即

被试 i 掌握属性 K 的边际概率），其计算公式为

$$\hat{p}_{ik} = \sum_{\boldsymbol{\alpha}_c \in Q_s} \alpha_{ck} P(\boldsymbol{\alpha}_c \mid X_i),$$

其中 $P(\boldsymbol{\alpha}_c \mid X_i)$ 为被试 i 知识状态为 $\boldsymbol{\alpha}_c$ 的后验概率，且 $P(\boldsymbol{\alpha}_c \mid X_i) = L(X_i \mid \boldsymbol{\alpha}_c) p(\boldsymbol{\alpha}_c) / (\sum_{\boldsymbol{\alpha}_c \in Q_s} L(X_i \mid \boldsymbol{\alpha}_l) p(\boldsymbol{\alpha}_l))$.

然后，可通过确定各属性的划界分数 p_{cutscore_k} 得出被试 i 在各个属性上掌握或未掌握的状态（或多个表现水平），本文将 p_{cutscore_k} 设定为 0.5，$\hat{p}_{ik} > p_{\mathrm{cutscore}_k}$ 表示被试 i 掌握了属性 k，反之则表示被试 i 未掌握属性 k.

知识状态在 MCMC 算法中为未知的随机向量，可直接从近似平稳分布中所取的样本计算均值，然后通过上述同样的划界分数方法可得到被试在各属性上的知识状态估计，进而得到估计的知识状态.

2　模拟研究

2.1　研究目的

在不同条件组合下，探讨 DINA 模型、DINO 模型和 G-DINA 模型下的 EM 算法和 MCMC 算法估计参数的表现.

2.2　研究设计

考虑到 4 个主要的影响因素：3 种类型认知诊断模型（3 个水平）、样本量（在 DINA 模型和 DINO 模型上为 2 个水平，在 GDINA 模型上为 3 个水平）、题目参数（2 个水平）、测验长度（2 个水平）. 所采用的 Q 矩阵如图 1 所示，其中 20 题和 30 题分别对应 Q_1、Q_2，参照 J. de la Torre 等[4] 的设定：在 30 道题目上考察 5 个相互独立属性，1 ~ 10 道题考察了 1 个属性，11 ~ 20 题考察了 2 个属性，21 ~ 30 题与 1 ~ 10 题一样只考察了 1 个属性. 样本量在 DINA 模型和 DINO 模型上分别设定为 500、1 000. 考虑到 G-DINA 模型的复杂性，在该模型上的 3 个水平样本量分别设定为 500、1 000 以及 2 000. 题目质量分别设定了高质量、低质量 2 个水平的项目参数，参照文献[16] 在 DINA 模型上猜测参数与失误参数分别服从 $U(0.05, 0.25)$ 和 $U(0.05, 0.40)$ 的设定，并在此基础上进行细化，其中在 DINA 模型、DINO 模型下高质量题目参数满足失误参数 s 和猜测参数 g 都服从均匀分布 $U \sim (0.05, 0.15)$，低质量题目参数满足失误参数 s 和猜测参数 g 都服从均匀分布 $U \sim (0.25, 0.35)$. 而在 G-DINA 模型上，未掌握的概率与答对题目概率参数在高质量水平下分别服从均匀分布 $U \sim (0.05, 0.15)$ 和 $U \sim (0.85, 0.95)$，在低质量水平下分别服从均匀分布 $U \sim (0.25, 0.35)$ 和 $U \sim (0.65, 0.75)$. 程序通过 R 语言编写，其中 EM 算法使用 Ma Wenchao 等[17] 所开发的 GDINA 包，MCMC 算法使用 JAGS 来实现[9].

$$Q_1 = \begin{pmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \\ 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \\ 1\,1\,0\,0\,0 \\ 1\,0\,1\,0\,0 \\ 1\,0\,0\,1\,0 \\ 1\,0\,0\,0\,1 \\ 0\,1\,1\,0\,0 \\ 0\,1\,0\,1\,0 \\ 0\,1\,0\,0\,1 \\ 0\,0\,1\,1\,0 \\ 0\,0\,1\,0\,1 \\ 0\,0\,0\,1\,1 \end{pmatrix} \qquad Q_2 = \begin{pmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \\ 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \\ 1\,1\,0\,0\,0 \\ 1\,0\,1\,0\,0 \\ 1\,0\,0\,1\,0 \\ 1\,0\,0\,0\,1 \\ 0\,1\,1\,0\,0 \\ 0\,1\,0\,1\,0 \\ 0\,1\,0\,0\,1 \\ 0\,0\,1\,1\,0 \\ 0\,0\,1\,0\,1 \\ 0\,0\,0\,1\,1 \\ 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \\ 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \end{pmatrix}$$

图 1　测验 Q 矩阵

2.3　评价指标

（i）采用偏差（f_{BIAS}）、均方根误差（f_{RMSE}）和绝对偏差（f_{ABS}）指标来考察参数估计的精度，这 3 指标值越接近 0 说明估计精度越高，由估计精度的高低来评价参数估计的返真性.

$$f_{\mathrm{BIAS}}(x_{gsp}) = \frac{1}{J} \sum_{j=1}^{J} (\hat{x}_{gsp_j} - x_{gsp_j}),$$

$$f_{\mathrm{RMSE}}(x_{gsp}) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}(\hat{x}_{gsp_j} - x_{gsp_j})^2},$$

$$f_{\mathrm{ABS}}(x_{gsp}) = \frac{1}{J}\sum_{j=1}^{J}|\hat{x}_{gsp_j} - x_{gsp_j}|,$$

其中 $\hat{x}_{gsp_j}$ 表示第 j 题的参数估计值,x_{gsp_j} 表示参数真值,J 为总题数,在 DINA 模型和 DINO 模型上待估计的参数为 g 和 s,它们分别代表猜测参数和失误参数;p 表示在 GDINA 模型上待估计的参数,对应着所有知识状态的测验项目作答反应概率,故此处将第 j 题在 3 个模型上待估计的参数简记为 x_{gsp_j}.

(ii) 运用模拟的平均属性判准率(attribute accuracy rate)来评价知识状态判准率,其值越高模型的表现越好,其计算公式为

$$f_{\mathrm{ACR}}(\alpha) = \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N} I(\alpha_{ik} = \hat{\alpha}_{ik}).$$

2.4 实验结果

表 1 ~ 表 4 分别是 EM 算法与 MCMC 算法在 DINA 模型、DINO 模型和 G-DINA 模型上的参数估计的偏差、均方根误差以及绝对偏差.

从表 1 可以看出:在 DINA 模型上,无论哪种条件,EM 算法的失误参数都会对出现低估现象,而在 DINO 模型上却相反,这可能跟模型本身类型有关;同时在题目质量低的情况下,在 DINA 模型和 DINO 模型上,EM 算法对猜测参数也会出现低估现象.从整体来看,在题目质量较低、样本量 $N = 500$ 以及测验长度 $L = 20$ 的组合条件下,MCMC 算法在 DINA 模型和 DINO 模型上的表现更优于 EM 算法的表现.从表 4 可以看出:在 G-DINA 模型上,无论是高质量题目还是低质量题目,在样本量 $N = 1\ 000$ 与测验长度 $L = 20$ 的条件组合下,MCMC 算法均优于 EM 算法.但随着样本量的增大和测验长度变长,EM 算法估计出的项目参数越来越接近真值.

从表 2 ~ 表 4 可以看出:无论哪个模型的哪种条件,MCMC 算法下参数估计的均方根误差均优于 EM 算法下参数估计的均方根误差,尤其是在题目质量较低和小样本量的组合下,有着非常明显的效果,均方根误差差值均超过 0.020;在 DINA 模型和 DINO 模型上的参数估计偏差的差值为 0.015 左右,在 G-DINA 模型上超过0.020.这表明了 MCMC 算法对参数估计的精确度高,且更适用于题目质量较低和小样本量的组合条件.但随着样本量的增大,EM 算法在 3 个模型上的表现在大多数情况下却是都略优于 MCMC 算法在 3 个模型上的表现.

表 1 2 种算法在 DINA 模型和 DINO 模型上的参数估计偏差

题目质量	样本量	测验长度	DINA				DINO			
			g_{EM}	s_{EM}	g_{MC}	s_{MC}	g_{EM}	s_{EM}	g_{MC}	s_{MC}
高	500	20	−0.000 3	−0.012 6	0.002 4	−0.005 0	0.004 3	−0.000 8	0.010 4	0.002 9
		30	0.007 1	−0.003 1	0.010 2	0.002 2	0.000 6	−0.002 4	0.005 1	0.001 1
	1 000	20	0.000 5	−0.012 5	0.001 5	−0.009 1	−0.005 1	0.000 1	−0.002 0	0.001 5
		30	−0.000 4	−0.002 8	0.001 1	−0.000 6	0.008 2	0.001 9	0.009 7	0.004 1
低	500	20	−0.004 3	−0.017 8	0.000 7	0.008 6	−0.012 8	0.016 4	0.000 8	0.029 0
		30	−0.001 4	−0.007 5	−0.001 3	0.011 5	−0.016 0	0.004 6	−0.012 0	0.010 8
	1 000	20	−0.002 5	−0.002 9	−0.001 8	0.015 7	−0.026 0	0.007 2	−0.012 9	0.011 6
		30	−0.000 7	−0.000 5	0.000 9	0.006 3	0.020 9	−0.018 2	0.019 3	−0.010 7

注:g_{EM}、s_{EM} 和 g_{MC}、s_{MC} 分别表示 EM 算法和 MCMC 算法在 DINA 模型和 DINO 模型上的猜测参数和失误参数的估计偏差.下同.

表 2 2 种算法在 DINA 模型和 DINO 模型上的参数估计均方根误差

题目质量	样本量	测验长度	DINA				DINO			
			g_{EM}	s_{EM}	g_{MC}	s_{MC}	g_{EM}	s_{EM}	g_{MC}	s_{MC}
高	500	20	0.018 0	0.020 6	0.017 6	0.016 8	0.030 3	0.014 9	0.031 6	0.015 4
		30	0.019 2	0.020 3	0.020 3	0.020 1	0.022 5	0.020 2	0.022 0	0.020 1
	1 000	20	0.013 0	0.031 9	0.012 7	0.030 0	0.018 9	0.013 1	0.018 3	0.013 4
		30	0.012 7	0.013 7	0.012 9	0.013 6	0.017 3	0.015 5	0.018 0	0.015 4
低	500	20	0.049 8	0.076 3	0.035 2	0.058 7	0.114 4	0.071 9	0.070 6	0.052 9
		30	0.037 5	0.051 1	0.036 0	0.048 6	0.061 8	0.041 6	0.057 2	0.038 3
	1 000	20	0.031 4	0.064 6	0.025 2	0.052 2	0.064 9	0.051 4	0.049 0	0.045 1
		30	0.022 1	0.017 3	0.021 8	0.018 6	0.032 6	0.028 0	0.031 3	0.022 6

表 3　2 种算法在 DINA 模型和 DINO 模型上的参数估计平均绝对离差

题目质量	样本量	测验长度	DINA				DINO			
			g_{EM}	s_{EM}	g_{MC}	s_{MC}	g_{EM}	s_{EM}	g_{MC}	s_{MC}
高	500	20	0.013 7	0.016 5	0.012 8	0.014 9	0.024 9	0.013 0	0.025 7	0.012 3
		30	0.015 7	0.016 3	0.016 7	0.015 8	0.018 8	0.015 3	0.018 3	0.015 4
	1 000	20	0.011 5	0.024 4	0.010 8	0.022 9	0.015 5	0.010 0	0.014 7	0.010 3
		30	0.009 0	0.010 7	0.009 2	0.010 8	0.014 6	0.011 9	0.015 0	0.011 7
低	500	20	0.041 5	0.062 4	0.031 1	0.043 4	0.088 2	0.051 8	0.059 4	0.043 6
		30	0.029 3	0.037 5	0.028 0	0.039 2	0.048 9	0.033 6	0.045 9	0.031 9
	1 000	20	0.023 7	0.050 8	0.019 1	0.038 2	0.046 0	0.040 9	0.035 5	0.036 1
		30	0.017 3	0.014 9	0.016 8	0.016 3	0.026 9	0.025 2	0.025 9	0.020 1

表 4　2 种算法在 G-DINA 模型上的参数估计偏差、均方根误差以及估计平均绝对离差

题目质量	样本量	测验长度	估计偏差		均方根误差		绝对偏差	
			p_{EM}	p_{MC}	p_{EM}	p_{MC}	p_{EM}	p_{MC}
高	500	20	−0.002 2	−0.001 5	0.018 7	0.018 0	0.015 1	0.014 4
		30	−0.003 2	−0.002 6	0.012 7	0.012 8	0.010 3	0.010 2
	1 000	20	0.001 6	0.001 8	0.011 6	0.012 2	0.009 3	0.009 7
		30	−0.000 6	−0.000 3	0.008 3	0.008 2	0.006 2	0.006 1
	2 000	20	−0.000 4	0.000 7	0.008 5	0.008 6	0.006 9	0.006 6
		30	0.001 4	0.001 4	0.008 5	0.008 4	0.006 4	0.006 3
低	500	20	0.020 2	0.011 7	0.067 4	0.039 0	0.050 8	0.029 8
		30	−0.000 7	0.002 5	0.025 3	0.019 3	0.022 4	0.016 4
	1 000	20	0.006 6	0.011 6	0.053 3	0.020 5	0.046 2	0.016 4
		30	0.019 6	0.018 7	0.033 0	0.030 1	0.028 2	0.026 0
	2 000	20	0.021 1	0.024 9	0.026 7	0.028 5	0.021 8	0.025 0
		30	−0.005 7	−0.003 1	0.014 5	0.013 0	0.011 3	0.010 3

注: p_{EM}和 p_{MC}分别表示 EM 算法和 MCMC 算法在 GDINA 模型上的所有知识状态的测验作答反应的估计偏差.

表 5 和表 6 分别是 EM 算法与 MCMC 算法在 DINA 模型、DINO 模型和 GDINA 模型上得到的平均属性判准率，由表 5 和表 6 可以看出: 随着题目质量的提高、样本量的增大以及测验长度的增长，这 2 种算法的平均属性判准率增大，且越来越趋于一致. 但在题目质量较低时，MCMC 算法下得到的平均属性判准率明显优于 EM 算法下的平均属性判准率，特别是在 GDINA 模型上，当测验长度 $L=20$ 与样本量 $N=500$ 或者 $N=1\ 000$ 的条件组合时，在这 2 种算法下得到的平均属性判准率差值为 0.04 左右，这说明了 MCMC 算法在题目质量较低与短测验长度组合条件下的优越性.

表 5　2 种算法在 DINA 模型和 DINO 模型上的平均属性判准率

题目质量	样本量	测验长度	DINA		DINO	
			EM	MC	EM	MC
高	500	20	0.983	0.984	0.974	0.975
		30	0.993	0.993	0.994	0.995
	1 000	20	0.972	0.975	0.982	0.981
		30	0.996	0.996	0.990	0.990
低	500	20	0.735	0.770	0.733	0.769
		30	0.822	0.832	0.824	0.847
	1 000	20	0.738	0.768	0.769	0.783
		30	0.841	0.847	0.841	0.844

表 6 2 种算法在 GDINA 模型上的平均属性判准率

题目质量	测验长度	样本量 $N=500$		样本量 $N=1\ 000$		样本量 $N=2\ 000$	
		EM	MC	EM	MC	EM	MC
高	20	0.978	0.979	0.978	0.980	0.972	0.972
	30	0.990	0.989	0.992	0.992	0.993	0.993
低	20	0.723	0.768	0.742	0.781	0.775	0.783
	30	0.813	0.824	0.821	0.832	0.823	0.823

3 结论和讨论

本文通过引入 EM 算法和 MCMC 算法,基于认知诊断 3 种模型,在多种条件组合下进行项目参数估计,通过比较偏差、均方根误差、绝对偏差、平均属性判准率来详细分析在不同条件组合下这 2 种算法在这 3 种模型上进行参数估计的效果.

当题目质量较低、测验短以及小样本这 3 种条件的组合时,MCMC 算法得到的偏差、均方根误差、绝对偏差、平均属性判准率的值均优于 EM 算法下得到的值.但若项目质量变高或者测验变长、样本量变大,则 EM 算法下得到的数据就略优于 MCMC 算法下得到的值,同时在该条件下,考虑到 MCMC 算法在 R 程序中的运行时间过长,不如 EM 算法计算快捷,故在题目质量较高或者测验长、样本量大的条件下推荐使用 EM 算法进行参数估计,在题目质量较低、测验短以及小样本的条件下使用 MCMC 算法进行参数估计.

本文只考虑了在认知诊断中常用的 3 个模型,在其他模型上不同条件下,EM 算法和 MCMC 算法进行参数估计的效果如何,值得思考.如本文所采用的 Q 矩阵仅考察了 2 个属性,而在实际应用中可能不限于 2 个属性,在 3 个属性甚至多个属性下 EM 算法和 MCMC 算法进行参数估计的效果如何;有些研究在属性层级模型下考虑了 EM 算法的表现[18],在不同知识状态分布和 Q 矩阵设计[19]条件下,这 2 种算法表现如何.Liu Yanlou 等[20]将这 2 种算法用于在 DINA 模型下的误差估计.纵向认知诊断模型的参数估计算法大多数采用 MCMC 算法[9],也有些研究针对纵向 GDINA 模型提出了 EM 算法[21],故在纵向认知诊断模型下这 2 种算法的表现也有待进行比较研究.另外,在估计未知参数标准误方面[22],MCMC 算法表现如何,有待考虑.

4 参考文献

[1] 汪文义,高朋,宋丽红,等.带噪音预处理的改进探索性 Q 矩阵标定方法[J].江西师范大学学报(自然科学版),2020,44(2):136-141.

[2] TEMPLIN J L,HENSON R A.Measurement of psychological disorders using cognitive diagnosis models[J].Psychological Methods,2006,11(3):287-305.

[3] DE LA TORRE J.DINA model and parameter estimation: a didactic[J].Journal of Educational and Behavioral Statistics,2009,34(1):115-130.

[4] DE LA TORRE J,DOUGLAS J A.Higher-order latent trait models for cognitive diagnosis[J].Psychometrika,2004,69(3):333-353.

[5] SEN S,TERZI R.A comparison of software packages available for DINA model estimation[J].Applied Psychological Measurement,2020,44(2):150-164.

[6] DE LA TORRE J.The generalized DINA model framework[J].Psychometrika,2011,76(2):179-199.

[7] YAMAGUCHI K,OKADA K.Variational Bayes inference for the DINA model[J].Journal of Educational and Behavioral Statistics,2020,45(5):569-597.

[8] YAMAGUCHI K,OKADA K.Variational Bayes inference algorithm for the saturated diagnostic classification model[J].Psychometrika,2020,85(4):973-995.

[9] ZHAN Peida,JIAO Hong,MAN Kaiwen,et al.Using JAGS for Bayesian cognitive diagnosis modeling: a tutorial[J].Journal of Educational and Behavioral Statistics,2019,44(4):473-503.

[10] CULPEPPER S A,HUDSON A.An improved strategy for Bayesian estimation of the reduced reparameterized unified model[J].Applied Psychological Measurement,2018,42(2):99-115.

[11] WANG Juntao,SHI Ningzhong,ZHANG Xue,et al.Sequential Gibbs sampling algorithm for cognitive diagnosis models with many attributes[J].Multivariate Behavioral Research,doi: 10.1080/00273171.2021.1896352.

[12] JIANG Zhehan,CARTER R.Using Hamiltonian Monte Carlo to estimate the log-linear cognitive diagnosis model via Stan[J].Behavior Research Methods,2019,51(2):651-662.

[13] FOX C,CUI Tiangang,NEUMAYER M.Randomized reduced forward models for efficient Metropolis-Hastings MCMC,with application to subsurface fluid flow and ca-

pacitance tomography [J]. GEM-International Journal on Geomathematics, 2020, 11(1): 26.

[14] HASTINGS W K. Monte Carlo sampling methods using Markov chains and their applications [J]. Biometrika, 1970, 57(1): 97-109.

[15] 汪文义，宋丽红. 教育认知诊断评估：理论与技术研究[M]. 北京：北京师范大学出版社，2015.

[16] WANG Wenyi, SONG Lihong, CHEN Ping, et al. Attribute-level and pattern-level classification consistency and accuracy indices for cognitive diagnostic assessment [J]. Journal of Educational Measurement, 2015, 52(4): 457-476.

[17] MA Wenchao, DE LA TORRE J. GDINA: an R package for cognitive diagnosis modeling [J]. Journal of Statistical Software, 2020, 93(14): 1-26.

[18] AKBAY L, DE LA TORRE J. Estimation approaches in cognitive diagnosis modeling when attributes are hierarchically structured [J]. Psicothema, 2020, 32(1): 122-129.

[19] TU Dongbo, WANG Shiyu, CAI Yan, et al. Cognitive diagnostic models with attribute hierarchies: model estimation with a restricted Q-matrix design [J]. Applied Psychological Measurement, 2019, 43(4): 255-271.

[20] LIU Yanlou, XIN Tao, ANDERSSON B, et al. Information matrix estimation procedures for cognitive diagnostic models [J]. British Journal of Mathematical and Statistical Psychology, 2019, 72(1): 18-37.

[21] YIGIT H D, DOUGLAS J A. First-order learning models with the GDINA: estimation with the EM algorithm and applications [J]. Applied Psychological Measurement, 2021, 45(3): 143-158.

[22] LIU Yanlou, XIN Tao, JIANG Yu. Structural parameter standard error estimation method in diagnostic classification models: estimation and application [J]. Multivariate Behavioral Research, doi: 10.1080/00273171.2021.1919048.

The Comparison of Parameter Estimation of Cognitive Diagnosis Models

WANG Wenyi[1], ZHENG Juanjuan[1], SONG Lihong[2], HU Haiyang[1]

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

Abstract: The performance of the EM algorithm and MCMC algorithm under different combinations of conditions is investigated. Three cognitive diagnosis models are considered, such as the DINA (deterministic inputs, noisy "and" gate), DINO (deterministic inputs, noisy "or" gate), and the generalized DINA (G-DINA) model. The bias, root mean square error, average absolute deviation and average attribute correct classification rate are used to evaluate the performance of two parameter estimation methods applicable to the three models under different conditions. The simulation results show that the MCMC algorithm performs better than the EM algorithm for the condition combination of low-quality and short test with small sample size, while these two methods performs similarly in the other condidions.

Key words: cognitive diagnosis model; EM algorithm; MCMC algorithm; parameter estimation

（责任编辑：冉小晓）