2PLM下缺失数据处理方法及其比较*

汪文义[1] 宋丽红[**2] 罗 芬[1] 丁树良[1]
（[1]江西师范大学计算机信息工程学院，南昌，330022）（[2]江西师范大学初等教育学院，南昌，330022）

摘 要 项目反应理论 (IRT) 是用于客观测量的现代教育与心理测量理论之一，广泛用于缺失数据十分常见的大尺度测验分析。IRT 中两参数逻辑斯蒂克模型 (2PLM) 下仅有完全随机缺失机制下缺失反应和缺失能力处理的 EM 算法。本研究推导 2PLM 下缺失反应忽略的 EM 算法，并提出随机缺失机制下缺失反应和缺失能力处理的 EM 算法和考虑能力估计和作答反应不确定性的多重借补法。研究显示：在各种缺失机制、缺失比例和测验设计下，缺失反应忽略的 EM 算法和多重借补法表现理想。

关键词 缺失数据 EM 算法 随机缺失机制 多重借补 项目反应理论

1 引言

项目反应理论 (IRT) 是一种现代教育与心理测量理论。基于潜在特质理论的 IRT，为揭示被试在测验项目上的反应行为与测验所测潜在特质之间的关系，建立了一系列模型。如两参数逻辑斯蒂克模型 (2PLM)，考虑项目难度和区分度因素对项目反应的影响，被试能力估计不依赖于测验项目选择，可以得到更加客观的测量 (漆书青 , 戴海崎 , 丁树良 , 2002)。

IRT 已经广泛用于分析大尺度能力测验，如 PISA，NAEP，TIMSS 和 PIRLS 等。由于测验风险、传统、设计和时间，以及被试有意识忽略等原因，缺失数据十分常见 (Cheema, 2014; Mislevy & Wu, 1988; Pohl, Gräfe, & Rose, 2014; Rose, von Davier, & Xu, 2010)。缺失数据影响数据分析结果 (Cheema, 2014)，因此缺失数据需要适当处理。在 IRT 下，处理缺失数据的方法主要分为三类 (Pohl et al., 2014)：(1) 重新记分法，包括记 0 分法 (IN)、分数记分法 (FR) 和缺失反应忽略法 (NP)；(2) 借补方法，包括均值借补法、项目反应函数借补法、EM 算法、MCMC 方法和多重借补方法；(3) 缺失机制建模方法，包括潜变量法和显变量法 (Pohl et al., 2014; Rose et al., 2010)。

有研究显示，基于缺失机制建模方法与缺失反应忽略法表现相当 (Rose et al., 2010)，因此本研究不考虑相对复杂的缺失机制建模方法。缺失反应忽略方法、多重借补方法和分数记分法较好，记 0 分法表现最差 (Finch, 2008; He & Wolfe, 2012; Rose et al., 2010)。缺失反应忽略法虽可采用商业软件 BILOG-MG 进行参数估计，由于涉及商业机密，笔者并未检索到其采用的估计算法的介绍。因此，有必要推导并编程实现缺失反应忽略算法。

基于似然函数的缺失数据处理方法通常可以产生更高效、更优良的估计量 (金勇进，邵军 , 2009)。在完全随机缺失 (MCAR) 机制下，有研究提出处理缺失反应和缺失能力的 EM 算法，当缺失比例大时该方法计算量大，并提到需要研究随机缺失机制 (MAR) 和非随机缺失机制 (NMAR) 下 EM 算法等问题 (张淑梅 , 辛涛 , 曾莉 , 孙佳楠 , 2011)。被试没有时间完成项目或有意识忽略下缺失反应，常依赖于被试作答反应或能力，分别属于 MAR 和 NMAR (Mislevy & Wu, 1988)，本研究将考虑 MAR 下缺失反应和缺失能力的 EM 算法。

Finch (2008) 基于 SAS 软件进行缺失反应多重借补，并非基于 IRT 模型的多重借补方法。项目反

* 本研究得到国家自然科学基金项目 (31500909,31360237,31160203)、全国教育科学规划教育部重点课题 (DHA150285)、教育部人文社会科学研究青年基金项目 (13YJC880060)、国家留学基金青年骨干教师出国研修项目 (201509470001)、江西省自然科学基金项目（20161BAB212044）、江西省社会科学研究“十二五”(2012 年) 规划项目 (12JY07)、江西省教育科学 2013 年度一般课题 (13YB032)、江西省教育厅科学技术研究项目 (GJJ13207, GJJ13208, GJJ13209)、江西师范大学青年成长基金和江西师范大学博士启动基金的资助。

** 通讯作者：宋丽红。E-mail: viviansong1981@163.com

DOI:10.16719/j.cnki.1671-6981.20160633

应函数借补法是一种考虑缺失反应随机性的多重借补法 (游晓锋 , 丁树良 , 刘红云 , 2011；Huisman & Molenaar, 2001)。然而能力估计误差 (或不确定性) 影响项目反应函数计算，也是带来缺失反应随机性的因素，这是本研究中多重借补方法提出的初衷。综上考虑，在 2PLM 下，本研究主要推导缺失反应忽略的 EM 算法、提出 MAR 下缺失反应和缺失能力处理的 EM 算法、提出考虑能力估计和作答反应不确定性的多重借补方法，并比较各方法的表现。

2 方法

记 N 表示被试数，M 表示项目数，$\mathbf{X}=(x_{ij})$ 表示得分矩阵，$i=1,2,\ldots,N$，$j=1,2,\ldots,M$。连续型能力分布采用离散分布近似 (Hanson & Woodruff, 1997)，即在连续型能力量尺上取 K 个离散结点，θ_k 小区间上的概率为 $p(\theta_k\mid\boldsymbol{\pi})=\pi_k, k=1,2,\ldots,K$。记 $\boldsymbol{\pi}=(\pi_1,\pi_2,\ldots,\pi_K)$ 和 β 为能力分布参数和项目参数，第 s 次估计为 $\pi^{(s)}$ 和 $\beta^{(s)}$，$s=1,2,\ldots,S$，初始值为 $\pi^{(0)}$ 和 $\beta^{(0)}$。

2PLM 下被试 i 在项目 j 上的正确作答概率为 (漆书青等 , 2002)：

$$p_j(\theta_i)=\frac{1}{1+\exp(-Da_j(\theta_i-b_j))} \tag{1}$$

其中 D=1.7，θ_i 为被试 i 的能力，a_j，b_j 为项目的区分度和难度。参数向量 $\boldsymbol{\beta}_j=(a_j,b_j)$，被试能力向量为 $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_N)$。因能力缺失，称（X, θ）为不可观察的完全数据，而 X 为可观察的不完全数据 (Baker, 1992)。以示区别，X 中的缺失数据称为缺失反应，记为 $x_{ij}=m$，若 X 中不存在缺失反应，称之为完整的观察作答反应。缺失反应和缺失能力统称为缺失数据。

2.1 缺失反应忽略的 EM 算法

在不考虑缺失反应、局部独立性和能力离散化条件下，（X, θ）的似然函数可表示为

$$\begin{aligned} f(\mathbf{X},\boldsymbol{\theta}\mid\boldsymbol{\pi},\boldsymbol{\beta}) &= \prod_{i=1}^{N}\prod_{k=1}^{K} f(\mathbf{x}_i,\theta_k\mid\boldsymbol{\pi},\boldsymbol{\beta})^{I(\theta_i=\theta_k)} \\ &= \prod_{i=1}^{N}\prod_{k=1}^{K}\pi_k^{I(\theta_i=\theta_k)}\prod_{i=1}^{M}p_j(\theta_k)^{I(x_{ij}=1)I(\theta_i=\theta_k)}(1-p_j(\theta_k))^{I(x_{ij}=0)I(\theta_i=\theta_k)}. \end{aligned} \tag{2}$$

其中 $I(.)$ 为指示函数，括号内条件满足时取 1，否则取 0。$f(\mathbf{x}_i\mid\theta_k,\boldsymbol{\beta})=L(\mathbf{x}_i\mid\theta_k,\boldsymbol{\beta})$ 表示作答反应向量 x_i 在给定能力为 θ_k 和项目参数为 β 下的似然函数。对（X, θ ）的似然函数取自然对数，可得

$$\begin{aligned} \ln f(\mathbf{X},\boldsymbol{\theta}\mid\boldsymbol{\pi},\boldsymbol{\beta}) &= \sum_{j=1}^{M}\sum_{l=1}^{K}\sum_{i=1}^{N}[I(x_{ij}=1)I(\theta_i=\theta_k)\ln p_j(\theta_k) \\ &+ I(x_{ij}=0)I(\theta_i=\theta_k)\ln(1-p_j(\theta_k))]+\sum_{k=1}^{K}\sum_{i=1}^{N}I(\theta_i=\theta_k)\ln\pi_k \end{aligned} \tag{3}$$

因为潜在能力为缺失数据，要对未知参数进行估计，即使 X 中缺失反应忽略，根据 EM 算法，在给定观察数据和上一步估计的参数值条件下，需要对缺失能力求条件期望，再通过最大化步骤估计未知参数。下面叙述期望 (E) 步和最大化 (M) 步。

(1) 缺失反应忽略的 EM 算法的 E 步。EM 的目标函数

$$Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})=E[\ln f(\mathbf{X},\boldsymbol{\theta}\mid\boldsymbol{\pi},\boldsymbol{\beta})\mid\mathbf{X},\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}] \tag{4}$$

在给定观察数据和上一步估计条件下，求期望得

$$\begin{aligned} Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}) &= E[\ln f(\mathbf{X},\boldsymbol{\theta}\mid\boldsymbol{\pi},\boldsymbol{\beta})\mid\mathbf{X},\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}] \\ &= \sum_{i=1}^{M}\sum_{k=1}^{K}[r_{kj}^{(s)}\ln p_j(\theta_k)+w_{kj}^{(s)}\ln(1-p_j(\theta_k))]+\sum_{k=1}^{K}(r_{kj}^{(s)}+w_{kj}^{(s)})\ln\pi_k \end{aligned} \tag{5}$$

其中人工数据计算公式为

$$r_{kj}^{(s)}=\sum_{i=1}^{N}\frac{I(x_{ij}=1)L(\mathbf{x}_i\mid\theta_k,\boldsymbol{\beta}^{(s)})\pi_k^{(s)}}{\sum_{k'=1}^{K}L(\mathbf{x}_i\mid\theta_{k'},\boldsymbol{\beta}^{(t)})\pi_{k'}^{(s)}} \tag{6}$$

$$w_{kj}^{(s)}=\sum_{i=1}^{N}\frac{I(x_{ij}=0)L(\mathbf{x}_i\mid\theta_k,\boldsymbol{\beta}^{(s)})\pi_k^{(s)}}{\sum_{k'=1}^{K}L(\mathbf{x}_i\mid\theta_{k'},\boldsymbol{\beta}^{(t)})\pi_{k'}^{(s)}} \tag{7}$$

$$n_{k}^{(s)}=\sum_{i=1}^{N}\frac{L(\mathbf{x}_i\mid\theta_k,\boldsymbol{\beta}^{(s)})\pi_k^{(s)}}{\sum_{k'=1}^{K}L(\mathbf{x}_i\mid\theta_{k'},\boldsymbol{\beta}^{(t)})\pi_{k'}^{(s)}} \tag{8}$$

若得分阵中不含缺失反应，则 $r_{kj}^{(s)}+w_{kj}^{(s)}=n_k^{(s)}$。若得分阵中项目 j 上含缺失数据，此时与不含缺失反应的 EM 算法相同。若被试 i 在项目 j 上的反应为缺失，在计算项目 j 上的人工数据时，其实是不包含被试 i 的部分。此时，$r_{kj}^{(s)}$ 和 $w_{kj}^{(s)}$ 分别表示作答了项目 i 的所有被试中具有能力为 θ_k 的被试答对或答错项目 θ_k 的期望人数，$n_{kj}^{(s)}$ 表示所有被试中具有能力为 θ_k 的期望人数。在缺失反应忽略时，$r_{kj}^{(s)}+w_{kj}^{(s)}\neq n_k^{(s)}$。

(2) 缺失反应忽略的 EM 算法的 M 步。最大化目标函数估计新一轮的未知参数值

$$(\boldsymbol{\pi}^{(s+1)},\boldsymbol{\beta}^{(s+1)})=\arg\max_{\boldsymbol{\pi},\boldsymbol{\beta}} Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}) \tag{9}$$

因项目参数和能力分布参数在目标函数中两个部分分离，项目参数和能力分布参数可分开进行估计。能力分布参数需考虑到概率的规范性 $\sum_{k=1}^{K}\pi_k=1$，采

用直接寻求条件极值的方法，即拉格朗日乘数法，易得

$$\pi_k^{(s+1)}=\frac{n_k^{(s)}}{\sum_{l=1}^{K}n_l^{(s)}} \tag{10}$$

对于项目参数估计，涉及非线性函数，可逐个迭代估计项目 j 的参数 (漆书青等 , 2002)

$$\boldsymbol{\beta}_j^{(s+1)}=\boldsymbol{\beta}_j^{(s)}-\left[E\left[\frac{\partial^2 Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})}{\partial\boldsymbol{\beta}_j^2}\right]\right]^{-1}\frac{\partial Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})}{\partial\boldsymbol{\beta}_j} \tag{11}$$

其中目标函数对 β_j 一阶微商向量中元素的一般式为

$$\frac{\partial Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})}{\partial\beta_{it}}=\sum_{k=1}^{Q}\frac{r_{kj}^{(s)}(1-p_j(\theta_k))-w_{kj}^{(s)}p_j(\theta_k)}{p_j(\theta_k)(1-p_j(\theta_k))}\frac{\partial p_j(\theta_k)}{\partial\beta_{it}} \tag{12}$$

由 $E[r_{kj}^{(s)}(1-p_j(\theta_k))-w_{kj}^{(s)}p_j(\theta_k)]=0$，信息矩阵中主对角线元素和非对角线元素分别为

$$\Lambda_{tt}^{j}=E\frac{\partial^2 Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})}{\partial\beta_{it}^2}=-\sum_{l=1}^{T}\frac{r_{kj}^{(s)}+w_{kj}^{(s)}}{p_i(\theta_k)(1-p_i(\theta_k))}\frac{\partial p_j(\theta_k)}{\partial\beta_{it}}\frac{\partial p_j(\theta_k)}{\partial\beta_{it}} \tag{13}$$

$$\Lambda_{tt'}^{j}=E\frac{\partial^2 Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})}{\partial\beta_{it}\partial\beta_{it'}}=-\sum_{l=1}^{T}\frac{r_{kj}^{(s)}+w_{kj}^{(s)}}{p_i(\theta_k)(1-p_i(\theta_k))}\frac{\partial p_j(\theta_k)}{\partial\beta_{it}}\frac{\partial p_j(\theta_k)}{\partial\beta_{it'}} \tag{14}$$

公式（13）和（14）中项目反应概率对区分度和难度的一阶偏导为

$$\frac{\partial p_j(\theta_k)}{\partial\beta_{j1}}=\frac{\partial p_j(\theta_k)}{\partial a_j}=D(\theta_k-b_j)p_j(\theta_k)(1-p_j(\theta_k)) \tag{15}$$

$$\frac{\partial p_j(\theta_k)}{\partial\beta_{j2}}=\frac{\partial p_j(\theta_k)}{\partial b_j}=-Da_jp_j(\theta_k)(1-p_j(\theta_k)) \tag{16}$$

根据初值或上一轮估计值计算偏导数，便可迭代估计项目参数。

2.2 MAR 下缺失反应和缺失能力处理的 EM 算法

若考虑得分阵中缺失反应，则 EM 算法中计算人工数据或期望步有所不同。下面给出 MAR 下缺失反应和缺失能力处理的 EM 算法 (记为 EE)。MAR 假设缺失能力参数和缺失反应仅依赖于观察反应。考虑 IRT 中局部独立性假设，即给定能力参数后，缺失反应仅依赖于能力参数，而不依赖于其他项目的作答反应。根据嵌套多重借补思路 (Reiter & Raghunathan, 2007)，采用以下方式对 EM 算法的目标函数求期望。

$$\begin{aligned}Q(\boldsymbol{\pi},\boldsymbol{\beta}\mid\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})&=E[\ln f(\mathbf{X},\boldsymbol{\theta}\mid\boldsymbol{\pi},\boldsymbol{\beta})\mid\mathbf{X},\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}]\\&=\sum_{i=1}^{M}\sum_{k=1}^{K}[\overline{r}_{kj}^{(s)}\ln p_j(\theta_k)+\overline{w}_{kj}^{(s)}\ln(1-p_j(\theta_k))]+\sum_{k=1}^{K}n_k^{(s)}\ln\pi_k\end{aligned} \tag{17}$$

其中 $\overline{r}_{kj}^{(s)},\overline{w}_{kj}^{(s)}$ 与上节人工数据公式不同，$\overline{r}_{kj}^{(s)},\overline{w}_{kj}^{(s)}$涉及对含缺失反应或缺失能力两个随机变量的指示函数求条件期望，即

$$E[I(x_{ij}=1)I(\theta_i=\theta_k)\mid\mathbf{x}_i,\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}] \tag{18}$$

$$E[I(x_{ij}=0)I(\theta_i=\theta_k)\mid\mathbf{x}_i,\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}] \tag{19}$$

如果 x_{ij} 为观察反应，即取值为确定值 0 或 1，只需对含缺失能力随机变量的指示函数求期望；否则，需要对含两个随机变量的指示函数求期望。人工数据计算公式为

$$\begin{aligned}\overline{r}_{kj}^{(s)}&=\sum_{i=1}^{N}E[I(x_{ij}=1)I(\theta_i=\theta_k)\mid\mathbf{x}_i,\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}]\\&=r_{kj}^{(s)}+\sum_{i=1}^{N}I(x_{ij}=m)P(x_{ij}=1\mid\theta_k,\boldsymbol{\beta}^{(s)})P(\theta_k\mid\mathbf{x}_i,\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})\end{aligned} \tag{20}$$

$$\begin{aligned}\overline{w}_{kj}^{(s)}&=\sum_{i=1}^{N}E[I(x_{ij}=0)I(\theta_i=\theta_k)\mid\mathbf{x}_i,\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)}]\\&=w_{kj}^{(s)}+\sum_{i=1}^{N}I(x_{ij}=m)P(x_{ij}=0\mid\theta_k,\boldsymbol{\beta}^{(s)})P(\theta_k\mid\mathbf{x}_i,\boldsymbol{\pi}^{(s)},\boldsymbol{\beta}^{(s)})\end{aligned} \tag{21}$$

计算人工数据并代入目标函数，然后采用 EM 算法 M 步 (与 2.1 相同) 估计未知参数。

2.3 基于项目反应函数和 EM 算法的缺失反应多重借补方法

有研究者针对 1PLM 给出了缺失反应的三种借补方法 (Huisman & Molenaar, 2001)，分别是：众数借补法、期望值借补法 (记为 PF) 和按分布随机数借补法 (记为 PR)。这些借补方法并未考虑能力估计误差。若 $x_{ij}=m$，根据 2PLM 项目参数$\hat{a}_j$，$\hat{b}_j$和被试能力值 $\hat{\theta}_i$，由公式 (1) 可计算被试 i 对项目 j 的正确作答概率$P_j(\hat{\theta}_i)$。由能力后验分布，可计算考虑被试能力估计误差的期望正确作答概率$E(P_{ij})=\sum_{k=1}^{K}P_j(\theta_k)P(\theta_k\mid\mathbf{x}_i,\boldsymbol{\pi},\boldsymbol{\beta})$。下面给出 2PLM 下考虑能力估计误差的两种借补法。

(1) 期望值借补法 (记为 EP)。该方法采用与 $E(P_{ij})$ 最接近的分值借补缺失反应

$$x_{ij}=\begin{cases}1 & 若E(P_{ij})\ge .50\\0 & 其他\end{cases} \tag{22}$$

(2) 分布随机借补法 (记为 ER)。考虑缺失反应的不确定性，均匀分布随机数 r 与 EP_{ij} 进行比较借补缺失反应

$$x_{ij}=\begin{cases}1 & 若E(P_{ij})\ge r\\0 & 其他\end{cases} \tag{23}$$

PF 和 PR 法使用$P_j(\hat{\theta}_i)$与 .50 或 r 比较。因借补方法中涉及项目和能力参数估计，下面给出基于 EM 算法的缺失反应多重借补方法的步骤（终止规则为目标函数相对变化率小于 .001）：(1) 忽略缺失

反应，采用 2.1 节算法估计项目参数与能力、能力后验分布；(2) 采用借补法 (EF, ER, PF, PR) 对缺失反应进行借补，得到借补后的完整作答反应；(3) 基于借补后的完整作答反应，通过 2.1 节算法估计项目参数; (4) 进行多重借补，即反复进行第 1 和 2 步 5 次，并使用前一轮参数进行借补，5 次平均值作为最终参数估计。

3 模拟研究

3.1 实验 1

实验设计：在不同缺失比例和测验设计下，比较忽略缺失反应下 BILOG-MG 和按 2.1 节自编程序的项目参数估计返真性。随机等组设计下考虑 5 种缺失比例 (.05, .15, .30, .40, .50)，采用 MCAR 模拟缺失反应。非等组设计下考虑锚测验（缺失比例 .40）和无锚测验（缺失比例 .50），前者缺失数据属于文件匹配缺失模式。

模拟数据：测验长度为 50，b_j和$\ln(a_j)$服从标准正态分布，且限定b_j和a_j范围 [-3，3] 和 [0.2,2.5]。根据能力和项目参数，由 2PLM 模拟完整作答反应。所有被试(总体)是由两个子总体各 500 名被试组成，子总体 A 和 B 的能力分别服从均值 -.25 和 .25，标准差均为 1 的正态分布。总体能力均值十分接近 0 和标准差只略大于 1。基于完整作答反应矩阵 (1000 行 50 列)，等组设计下所有被试的缺失反应随机产生；非等组设计下缺失反应按表 1 方式产生，两个子总体被试作答的测验项目不完全相同。

数据分析：根据能力分布抽取样本量为 1000 的 30 个样本，在项目参数固定情况下，模拟 30 个完整得分阵，再按不同缺失比例和测验设计生成带缺失反应的得分阵。然后使用 BILOG-MG(NP) 和 2.1 节公式自编程序进行参数估计。使用均方误差 RMSE，偏差 BIAS 和绝对偏差 ABS(张淑梅等 , 2011) 评价项目参数估计误差。

实验结果：从不同缺失比例、测验设计下项目参数返真性指标来看 (表 2)，自编程序和 BILOG-MG 表现基本相当。在缺失比例达到 .40 或 .50 时，多数情况下自编程序稍好于 BILOG-MG。在三种测验设计下，不含锚 – 非等组设计下均方误差最大，尤其是难度参数的均方误差更大。在缺失比例为 .40 时，随机等组设计与含锚 – 非等组设计的下各方法的表现基本相当，说明含锚 – 非等组设计下项目参数同时估计精度较好。

3.2 实验 2

实验设计：考虑三种缺失反应机制和四种缺失比例下各方法 (自编程序) 所得项目参数的返真性。

表 1 锚测验和无锚测验非等组设计

被试	锚测验 – 非等组设计			无锚测验 – 非等组设计	
	1~20	21~30	31~50	1~25	26~50
A 组	观察	观察	缺失	观察	缺失
B 组	缺失	观察	观察	缺失	观察

表 2 忽略缺失反应下 BILOG-MG 与自编程序结果比较

设计	缺失比例	估计	ABS		BIAS		RMSE	
			a	*b*	*a*	*b*	*a*	*b*
等组	.00	BM	.072	.080	–.011	–.004	.112	.114
		ZB	.085	.086	–.056	.001	.119	.119
	.05	BM	.074	.082	–.010	–.005	.116	.117
		ZB	.086	.087	–.054	.001	.122	.120
	.15	BM	.080	.086	–.005	–.005	.122	.121
		ZB	.087	.090	–.049	.001	.124	.124
	.30	BM	.088	.097	.000	–.007	.136	.141
		ZB	.092	.096	–.044	.003	.134	.135
	.40	BM	.095	.101	.006	–.002	.156	.145
		ZB	.096	.102	–.037	.007	.141	.146
	.50	BM	.111	.110	.008	–.005	.174	.162
		ZB	.110	.109	–.036	.007	.161	.156
非等组	.40	BM	.094	.122	.015	–.011	.141	.189
		ZB	.109	.136	–.056	.002	.156	.174
	.50	BM	.106	.268	.038	–.016	.161	.319
		ZB	.109	.248	–.042	.008	.155	.280

注：BM=BILOG–MG；ZB= 自编程序。

缺失反应机制含 MCAR, MAR 和 NMAR，缺失比例 p 为 .05, .15, .30 和 .40，每种条件下重复 30 次。方法涉及：2.1 节中忽略缺失反应法 NP；张淑梅等人 (2011) 提出的 EM 算法 ZS；记 0 分法 IN；2.2 节中算法 EE；2.3 节中借补方法 PR,ER,PF 和 EF。

模拟数据：能力服从标准正态分布，样本量为 1000，测验长度为 20，项目参数分布同实验 1。设定四个目标项目，区分度为 .44,1.02,.76 和 1.32，难度为 -.30, 1.28, -2.70 和 .57。基于 2PLM 模拟的完整作答反应矩阵，仿照 Finch (2008) 方法，模拟三种缺失机制和四种缺失比例下含缺失反应的得分矩阵。MCAR 下缺失反应采用随机分配方法，即对每个被试在每个目标项目产生均匀分布随机数，若随机数小于缺失比例 p，则认为反应缺失，直到满足缺失比例为止。MAR 下缺失反应依赖其他观察变量 (非目标项目上得分)，先根据非目标项目上的总分将被试划分为四类 (0~3,4~7,8~11,12~16)，统计各类频数 N_1, N_2, N_3, N_4，然后根据缺失比例模拟各类被试在目标项目的缺失反应。各类缺失比例依据总分越低目标项目上缺失比例越高的原则，并满足总缺失比例约束条件：

$$Np = pN_1 + N_2\left(p_1 - \frac{1}{3}p\right) + N_3\left(p_1 - \frac{2}{3}p\right) + N_4\left(p_1 - \frac{3}{3}p\right) \quad (24)$$

解出 0~3 分数段的缺失比例 p_1，再得出其他分数段的缺失比例 p_1-p/3，p_1-2p/3 和 p_1-p。NMAR 下缺失数据依赖缺失变量本身的值(目标项目上得分)，根据完整作答反应矩阵中目标项目上总分(0,1,2,3,4) 将被试划分为五类，类似于 MAR 方式计算各类缺失比例和模拟各类被试在目标项目的缺失反应。

实验结果：从表 3 到表 8 来看，有以下结论：从均方误差来看，与张淑梅等人 (2011) 的研究有着类似发现，缺失比例较小时 (.05)，在区分度参数上，ZS 好于 IN，甚至是最好的方法，在难度参数上，ZS 好于 IN，而在缺失比例中等或较大时 (.15,.30,.40)，ZS 比 IN 要差。NP、PR 和 ER 三种方法表现较好。在多数条件下，PR 和 ER 在偏差、绝对偏差和均方误差指标上表现均较好，ER 稍好于 PR 和 NP。对于绝对偏差和均方误差，PF 和 EF 表现不如 PR 和 ER。从偏差来看，在缺失比例较大时 (.30,.40)，PF 和 EF 倾向于高估区分度，这是因为在 PF 和 EF 借

表 3 不同缺失比例和缺失机制下区分度的偏差

缺失比例	缺失机制	NP	ZS	IN	PR	PF	ER	EF	EE
.05	MCAR	−.049	.018	.068	−.040	−.068	−.037	−.067	−.063
	MAR	−.053	.020	.007	−.045	−.068	−.042	−.068	−.065
	NMAR	−.052	.010	.025	−.046	−.069	−.042	−.069	−.065
.15	MCAR	−.055	.169	.150	−.053	−.144	−.042	−.143	−.069
	MAR	−.044	.178	.056	−.042	−.112	−.036	−.111	−.060
	NMAR	−.050	.164	.090	−.047	−.127	−.038	−.126	−.065
.30	MCAR	−.049	.370	.237	−.053	−.266	−.041	−.264	−.067
	MAR	−.049	.360	.064	−.051	−.198	−.039	−.197	−.066
	NMAR	−.034	.359	.145	−.046	−.211	−.022	−.209	−.056
.40	MCAR	−.043	.456	.280	−.058	−.362	−.029	−.359	−.064
	MAR	−.031	.474	.067	−.061	−.276	−.014	−.273	−.048
	NMAR	−.035	.457	.157	−.048	−.299	−.018	−.295	−.064

表 4 不同缺失比例和缺失机制下难度的偏差

缺失比例	缺失机制	NP	ZS	IN	PR	PF	ER	EF	EE
.05	MCAR	.072	.068	.009	.071	.074	.070	.074	.071
	MAR	.071	.073	.013	.068	.073	.068	.073	.068
	NMAR	.081	.083	.029	.077	.082	.077	.082	.079
.15	MCAR	.065	.030	−.152	.057	.069	.057	.068	.065
	MAR	.065	.069	−.253	.056	.055	.055	.055	.065
	NMAR	.090	.084	−.103	.082	.083	.084	.082	.090
.30	MCAR	.057	−.057	−.816	.051	.061	.033	.059	.058
	MAR	.053	.052	−.724	.047	.020	.035	.019	.053
	NMAR	.131	.101	−.615	.119	.106	.113	.105	.135
.40	MCAR	.057	−.127	−1.297	.041	.057	.042	.055	.060
	MAR	.050	.068	−1.011	.038	.015	.029	.014	.057
	NMAR	.170	.136	−.976	.155	.137	.145	.135	.176

表 5 不同缺失比例和缺失机制下区分度的绝对偏差

缺失比例	缺失机制	NP	ZS	IN	PR	PF	ER	EF	EE
.05	MCAR	.224	.168	.214	.222	.254	.216	.253	.234
	MAR	.226	.153	.219	.217	.263	.210	.263	.230
	NMAR	.236	.173	.239	.230	.270	.230	.269	.247
.15	MCAR	.236	.202	.209	.242	.354	.234	.352	.248
	MAR	.221	.209	.202	.214	.342	.204	.341	.217
	NMAR	.223	.198	.210	.220	.337	.212	.336	.227
.30	MCAR	.242	.362	.239	.266	.528	.233	.524	.252
	MAR	.229	.386	.188	.240	.496	.201	.494	.213
	NMAR	.231	.362	.206	.238	.494	.218	.492	.226
.40	MCAR	.211	.450	.280	.252	.609	.216	.605	.225
	MAR	.222	.474	.183	.228	.589	.192	.588	.192
	NMAR	.228	.447	.217	.245	.609	.215	.606	.226

表 6 不同缺失比例和缺失机制下难度的绝对偏差

缺失比例	缺失机制	NP	ZS	IN	PR	PF	ER	EF	EE
.05	MCAR	.098	.101	.150	.101	.104	.099	.104	.093
	MAR	.097	.096	.129	.097	.109	.098	.109	.091
	NMAR	.101	.103	.140	.101	.108	.102	.109	.096
.15	MCAR	.097	.137	.301	.094	.104	.094	.104	.094
	MAR	.093	.133	.260	.089	.117	.087	.117	.089
	NMAR	.105	.128	.217	.102	.109	.105	.108	.101
.30	MCAR	.089	.246	.816	.091	.113	.080	.114	.086
	MAR	.088	.228	.724	.088	.145	.080	.145	.084
	NMAR	.140	.234	.615	.130	.127	.128	.127	.140
.40	MCAR	.099	.363	1.297	.101	.125	.098	.126	.098
	MAR	.107	.362	1.011	.101	.147	.101	.148	.103
	NMAR	.172	.331	.976	.162	.152	.158	.152	.178

表 7 不同缺失比例和缺失机制下区分度的均方误差

缺失比例	缺失机制	NP	ZS	IN	PR	PF	ER	EF	EE
.05	MCAR	.127	.088	.234	.122	.148	.118	.148	.138
	MAR	.127	.081	.178	.123	.151	.119	.151	.137
	NMAR	.128	.085	.206	.124	.149	.120	.149	.139
.15	MCAR	.137	.205	.263	.143	.239	.129	.237	.149
	MAR	.124	.207	.204	.125	.229	.117	.229	.134
	NMAR	.130	.197	.249	.128	.229	.120	.228	.141
.30	MCAR	.141	.413	.300	.163	.399	.134	.397	.157
	MAR	.137	.402	.207	.149	.423	.128	.421	.144
	NMAR	.144	.400	.267	.157	.411	.134	.409	.151
.40	MCAR	.160	.497	.332	.191	.557	.157	.553	.178
	MAR	.138	.521	.198	.159	.533	.123	.530	.125
	NMAR	.155	.499	.271	.184	.567	.145	.563	.162

表 8 不同缺失比例和缺失机制下难度的均方误差

缺失比例	缺失机制	NP	ZS	IN	PR	PF	ER	EF	EE
.05	MCAR	.137	.142	.180	.140	.151	.139	.151	.128
	MAR	.133	.136	.158	.133	.154	.136	.154	.124
	NMAR	.139	.144	.169	.139	.153	.139	.153	.132
.15	MCAR	.134	.171	.331	.131	.155	.131	.154	.129
	MAR	.129	.168	.318	.120	.160	.118	.160	.123
	NMAR	.139	.167	.259	.134	.157	.136	.157	.134
.30	MCAR	.129	.274	.902	.125	.158	.110	.159	.125
	MAR	.121	.248	.914	.118	.180	.108	.180	.114
	NMAR	.171	.263	.725	.161	.169	.156	.169	.170
.40	MCAR	.134	.402	1.459	.132	.166	.127	.166	.133
	MAR	.145	.401	1.288	.130	.181	.135	.181	.138
	NMAR	.204	.386	1.199	.198	.191	.192	.192	.205

补缺失值时，答对概率大于 .5 或能力高者在项目上判为正确作答，而答对概率小于 .5 或能力低者在项目上判为错误作答，实际上扩大了高低分组的分数差异。IN 将低估区分度，并且高估难度 (Rose et al., 2010)。EE 与 NP 分别在难度和区分度表现有优势，整体表现基本相当。

4 结论

第一，对仅在商业软件中实现且未见资料公开报告的缺失反应忽略的 EM 算法进行推导，同时采用 Matlab 编程实现。研究显示，自编的缺失反应忽略的 EM 算法可以准确地估计项目参数，有时甚至好于 BILOG-MG 商业软件。并且该算法在较多情形下好于其他缺失反应处理方法，说明所推导的算法和自编程序具有应用价值。

第二，张淑梅等人 (2011) 提出的方法优于记 0 分法，这与其研究结果一致。特别是在缺失比例较小时 (.05)，表现好于其他方法，且对区分度的估计表现尤为突出。意味着张淑梅等人 (2011) 的方法十分适合缺失比例较小 (.05) 的测验情形，这时计算量也不会太大且估计精度理想。

第三，MCAR 下缺失反应的 EM 算法与缺失反应忽略的 EM 算法表现基本相当。这是因为缺失反应忽略的EM算法其实已经暗含了可忽略缺失机制，而 MCAR 和 MAR 均属于可忽略缺失。

第四，记 0 分法表现最差。记 0 分法将低估区分度，这是因为不管能力高低，在项目上的缺失得分均被视为 0 处理，也就意味着能力高者，本来有更多的 1 分，但被当成 0 分处理，高低分组的分数差异由此减小。该方法将更多潜在得 1 分当成了 0 处理，还将高估难度。

第五，考虑能力估计和作答反应不确定的多重借补方法表现最好。从缺失值的后验分布，即 (期望) 正确作答概率分布中取样进行多重借补方法，表现甚至稍好于缺失反应忽略的 EM 算法，明显好于按项目反应概率最近分数的借补方法。这是因为从缺失值后验预测分布中取样的多重借补方法，既考虑了项目反应自身的不确定性，还考虑了项目和能力参数 (正确作答概率) 的估计误差。

仅列出参数初值采用 2.1 节所给出的 EM 算法的多重借补方法的结果，另外还与 2PLM 项目参数初值计算方法 (漆书青等 , 2002) 进行了比较，两者结果相当。初值亦可使用双重两步迭代估计法 (丁树良，罗芬 , 2007)。缺失数据处理的 EM 算法，有待与MCMC方法(汪金晖,张淑梅,辛涛,2011;曾莉,辛涛,张淑梅,2009) 比较。认知诊断 (宋丽红 , 汪文义 , 丁树良 , 2015) 中借补方法也有待研究。在一定条件下项目参数的估计受缺失机制的影响小，但若对缺失机制进行检验，将有助于理解被试作答反应背后的机制。今后需考虑利用其他背景或辅助变量借补缺失数据 (黄慧静 , 辛涛 , 李珍 , 2012)。MNAR 下 EM 算法仍有待研究。

参考文献

丁树良 , 罗芬 . (2007). 双重两步迭代估计及其应用 . *数值计算与计算机应用 , 28(2),* 116–123.

黄慧静 , 辛涛 , 李珍 . (2012). 矩阵取样设计中的似真值能力估计方法 . *心理科学 , 35(5),* 1233–1239.

金勇进 , 邵军 . (2009). *缺失数据的统计处理* . 北京 : 中国统计出版社 .

漆书青 , 戴海崎 , 丁树良 . (2002). *现代教育与心理测量学原理* . 北京 : 高等教育出版社 .

宋丽红 , 汪文义 , 丁树良 .(2015). 测验 Q 矩阵的修正方法及其比较研究 . *江西师范大学学报 (自然科学版), 39(6),* 623–630.

汪金晖 , 张淑梅 , 辛涛 . (2011). 缺失数据下等级反应模型参数 MCMC 估计 . *北京师范大学学报 (自然科学版), 47(3),* 229–234.

游晓锋 , 丁树良 , 刘红云 . (2011). 缺失数据的估计方法及应用 . *江西师范大学学报 (自然科学版), 35(3),* 325–330.

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

张淑梅 , 辛涛 , 曾莉 , 孙佳楠 . (2011). 2PL 模型的 EM 缺失数据处理方法研究 . *应用概率统计 , 27(3)*, 241–255.

Baker, F. B. (1992). *Item response theory: Parameter estimation techniques.* New York: Marcel Dekker.

Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research, 84(4),* 487–508.

Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *Journal of Educational Measurement, 45(3),* 225–245.

Hanson, B. A., & Woodruff, D. (1997). Estimation for item response models using the EM algorithm for finite mixtures. *Paper presented at the annual meeting of the Psychometric Society, Gatlinburg, Tennessee.*

He, W., & Wolfe, E. W. (2012). Treatment of not–administered items on individually administered intelligence tests. *Educational and Psychological Measurement, 72(5),* 808–826.

Huisman, M., & Molenaar, I. W. (2001). Imputation of missing scale data with item response models. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 221–244). New York: Springer.

Mislevy, R. J., & Wu, P. K. (1988). *Inferring examinee ability when some item responses are missing (RR–88–48–ONR).* Princeton, NJ: Educational Testing Service.

Pohl, S., Gräfe, L., & Rose, N. (2014). Dealing with omitted and not–reached items in competence tests: Evaluating approaches accounting for missing responses in item response theory models. *Educational and Psychological Measurement, 74(3),* 423–452.

Reiter, J. P., & Raghunathan, T. E. (2007). The multiple adaptations of multiple imputation. *Journal of the American Statistical Association, 102(480),* 1462–1471.

Rose, N., von Davier, M., & Xu, X. L. (2010). *Modeling nonignorable missing data with item response theory (IRT)(ETS RR–10–11)*. Princeton, NJ: Educational Testing Service.

Missing Data Handling Methods Based on the 2PLM

Wang Wenyi[1], Song Lihong[2], Luo Fen[1], Ding Shuliang[1]

([1]College of Computer Information Engineering, Jiangxi Normal University, Nanchang, 330022)

([2]Elementary Educational College, Jiangxi Normal University, Nanchang, 330022)

Abstract Missing data are encountered regularly by researchers in educational research. For example, many large-scale assessments are low-stakes surveys, which typically suffer from a substantial amount of missing data. The low-stakes nature of these surveys, as well as variations in the average performance across countries and other factors such as testing traditions, test design, time limits, intentional omission, have been discussed as contributing factors to the amount of omitted responses observed in these assessments. Researchers have shown that missing data may create problems in the estimation of item parameters and subject ability parameters in the item response theory (IRT) context. A number of missing data handling methods have been developed in the IRT framework. The methods are not only involving the response function imputation, but also including treating the missing items as not presented (NP), incorrect (IN) or fractionally correct (FR), which can be carried out directly with the item parameter estimation software BILOG-MG. There have also been a number of algorithms in the context of data imputation.

The current study described several approaches to deal with missing data in the two-parameter logistic model (2PLM). Although the software BILOG-MG can handle the missing data, it is a commercial software. We needed to domestically develop an EM algorithm in which the missing responses were ignored, that is, to be treated as missing completely at random (MCAR). MCAR, which can be thought of as having no systematic cause, is only one specific type of missing data. It is noted that Zhang, Xin, Zeng, and Sun have proposed an EM algorithm (denote it as ZS) to deal with missing data under MCAR with a huge computational burden when the percent of missing data is higher. When data are missing at random (MAR), the probability of a value being missing is dependent on item response of the individual but not on the missing value itself. The estimation of item parameters and abilities may be influenced. However, to the best of our knowledge, there has been no work addressing the missing data under the assumption for MAR in 2PLM. We proposed an EM algorithm under MAR, denoted by EE. Following a general introduction of multiple imputing methods, which was based on the item response model, two new multiple imputing methods (EF and ER) were proposed by considering uncertainties of ability parameter estimates and missing item responses, compared against two original methods (PF and PR) proposed by Huisman and Molenaar, which were only based on item responses probability.

Simulation studies were provided to demonstrate the accuracy of these methods with a sample size of 1000. Various percentages of missing data were simulated: 5%, 15%, 30%, 40% and 50%. Missing data were simulated according to three different types of underlying missing data mechanism, including MCAR, MAR, and missing not at random. Missing data were imputed by NP, ZS, IN, PR, PF, ER, EF, and EE. Simulation results suggested that new multiple imputing methods and NP method worked well in various conditions; the EM algorithm under MAR had similar performance compared with the NP, because the 2PLM had the advantage of invariance of model parameters.

Key words missing data, the EM algorithm, missing at random, multiple imputing, item response theory