

基于D-最优和A-最优设计的多维在线标定设计研究

杨森, 何引红*, 祁媛媛

南京信息工程大学, 江苏 南京

收稿日期: 2022年12月15日; 录用日期: 2023年1月8日; 发布日期: 2023年1月17日

摘要

多维计算机化自适应测验(MCAT)近年来在教育测量中受到越来越多的关注。与所有其他CAT一样,项目补充是MCAT的项目库维护和管理的一个重要组成部分。题库管理者需要定期淘汰题库中过度暴露或过时的项目,并替换为新的项目。在单维CAT(UCAT)中,在线标定技术已被用于有效地标定新项目。然而,文献中关于MCAT在线校准的讨论很少。因此本文在现有UCAT的基础,将UCAT的在线标定设计D-最优设计和A-最优设计推广至多维在线标定情境。本文进行计算机模拟实验,探究不同样本量和能力间的相关系数对D-最优设计和A-最优设计的影响。结果表明,平均后的总信息量和题目参数的信息量随着能力的相关性的增大而减少,在能力的相关性为0.2和0.5时随着新题作答的题目数量的增大呈现先增后减的趋势,而在能力的相关性为0.8时几乎不变,D-最优设计和A-最优设计对其影响较小。

关键词

计算机自适应测验, Fisher信息量, 在线标定设计, D-最优设计, A-最优设计

Research on Multidimensional Online Calibration Design Based on D-Optimal and A-Optimal Designs

Sen Yang, Yinhong He*, Yuanyuan Qi

Nanjing University of Information Science and Technology, Nanjing Jiangsu

Received: Dec. 15th, 2022; accepted: Jan. 8th, 2023; published: Jan. 17th, 2023

*通讯作者。

文章引用: 杨森, 何引红, 祁媛媛. 基于D-最优和A-最优设计的多维在线标定设计研究[J]. 应用数学进展, 2023, 12(1): 81-95. DOI: 10.12677/aam.2023.121011

Abstract

In recent years, multidimensional computerized adaptive testing (MCAT) has received more and more attention in educational measurement. As with all other CATs, item replenishment is an important component of MCAT's item bank maintenance and management. Bank managers need to regularly eliminate overexposed or outdated items in the item bank and replace them with new ones. In unidimensional CAT (UCAT), online calibration techniques have been used to calibrate new items effectively. However, there is little discussion in the literature about online calibration of MCAT. Therefore, based on the existing UCAT, this study extends the D-optimal design and A-optimal design of UCAT online calibration design to the multidimensional online calibration situation. In this study, A computer simulation experiment was conducted to explore the influence of sample size and the correlation coefficient between the ability components on D-optimal design and A-optimal design. The results showed that the average total information and the information of the item parameters decreased with the increase of the correlation of ability components; increased first and then decreased with the increase of the number of new items when the correlation of ability components was 0.2 and 0.5; and was almost unchanged when the correlation of ability components was 0.8, indicating that it was a little affected by the D- and A-optimal designs as this experimental condition.

Keywords

Computer Adaptive Testing, Fisher Information, Online Calibration Design, D-Optimal Design, A-Optimal Design

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着计算机编程技术和网络通讯技术的不断完善以及进一步发展,在大型项目评估测验中,基于项目反应理论(item response theory, IRT)而构建的计算机自适应测验(computerized adaptive testing, CAT)逐渐取代传统笔试而成为主流[1]。CAT的核心机制是通过被试当前的能力水平,以项目反应理论为指导,依赖于大型题库,智能化地通过计算机选取适合被试水平的题目并给予其作答,从而达到“量体裁衣”、“因人施测”的自适应化的测量方式[2]。CAT的主要优点是其测试精度明显高于传统笔试。并且相比于传统笔试,计算机自适应测验因人施测的测验方式可以极大减少测试题量和提高测试效率,同时可以有效防止作弊等行为。CAT其多媒体式的测验方式,可以为被试提供更加全方位的测试方式,如影像,语音,图像等,并且呈现出更为公平,公正,高效的测验结果。目前CAT在各种大型测试中得到了广泛的应用,例如美国研究生入学考试(GRE),企业管理研究生入学考试(GMAT),注册护士执照考试(NCLEX-RN)等[3]。

随着CAT的广泛使用,研究人员往往面临着一个新的问题,即题库中部分题目存在着过度曝光,过时等问题而不适于继续使用[4]。因此为了满足CAT的实施,我们需要对题库中已经不再使用的题目进行删减或休眠,同时还要为题库补充新题,即对题库进行题目增补[5]。这时往往需要对新题进行标定,才能将合适的题目归入到题库中[6]。对新题的标定主要包括两方面的内容:一是估计新题的题目参数;

二是将新题参数置于旧题的参数量尺上[7]。

目前,相比于单维计算机自适应测验(unidimensional computerized adaptive testing, UCAT),多维计算机自适应测验(multidimensional computerized adaptive testing, MCAT)不仅在一次测验中可以反应被试在多个维度上的能力水平,还极大的提高了测验精度,提供了更多的被试信息,且可以自动平衡满足各个领域的测量要求[8]。因此,MCAT兼具了多维项目反应理论和计算机自适应测验的优点,近年来得到国内外学者的巨大推崇[9]。但其作为计算机自适应测验,仍然面临着题库更新的问题,那么仍需对其进行在线标定过程,以确定题目参数及新题插入位置。虽然UCAT的在线标定设计的研究有很多,但是对MCAT的在线标定设计的研究仍处于初步探索阶段。为了实现题库的更新,使得计算机自适应测验得以长久应用,在线标定是其重要环节。

针对MCAT的在线标定设计策略,通过基于不同项目参数的信息量对其进行研究具有重大意义。在线标定设计即新题的施测方式,其核心问题是如何将被试与新题合理搭配以优化题目参数标定的效率。Wainer和Mislevy最早提出了随机和自适应两种在线标定的设计方式[4]。其中随机标定是从新题库中随机选取固定数量的新题,并且随机嵌入测验给被试作答。游晓峰等人研究发现随机设计中新题的作答次数越多,题目参数估计的越准确[5]。虽然随机标定操作简单,但嵌入新题的难度与相邻题目的难度可能存在明显差异导致考生容易察觉,导致题目参数估计的不准确,同时也没有体现出计算机自适应的特点。自适应标定设计按照标准选取最能反映项目特征的被试,或者选取最适合被试作答的新题施测。其中最重要的是基于项目信息标准的最优设计。

基于项目信息标准的最优设计的基本思路是通过项目参数的信息量来反映参数估计误差。基于项目信息标准的最优设计包括D-优化,序贯D-优化,两点D-优化和贝叶斯D-优化以及优秀度指标和D-c设计等[2]。D-优化设计通过最大化项目参数Fisher信息矩阵的行列式来最小化项目参数的广义协方差,它是以优化项目参数估计效率为目标的统计指标。在D-优化设计的基础上,Berger选取能力估计值与真值最接近的被试施测新题,即两点D-优化设计[10](two-point D-optimal design, D-Tp)。Chang和Lu基于最优能力准则,在不等长CAT中直接应用两点D-优化设计法,并按序贯的方式选取被试作答新题,称其为序贯D-优化设计[11]。Hassan和Miller基于限制性最优设计的思想,提出按照在最优能力区间而不是最优设计点进行取样,称其为局部限制性最优设计[12]。Ren等人认为从单点能力抽样被试不能产生稳定的参数估计值[13],同时基于D-优化和A-优化视角将最优能力替换为最优作答概率,提出D-Tp1、D-Tp2和D-Tp3方法[14]。Kang等人还从丰富被试信息的视角,利用被试的反应时信息,在联合反应和反应时模型下考察D-优化、A-优化和随机方法的表现[15]。van der Linden和Ren在D-优化基础上根据当前考生 p 的能力估计值,选择使当前累加考生样本与之前累加考生样本相比,能提供项目参数Fisher信息增量最多的项目给考生 p 作答,提出贝叶斯版本的D-优化设计,称为D-VR设计[16]。由于D-VR设计中到达嵌入位置的当前考生不一定是最优考生样本,He等人结合D-优化的思路对D-VR改进,他们以D-VR设计表示的最优考生在新题 j 上提供的信息增量为基准,通过衡量当前考生的能力相对于最优考生的能力在标定新题 j 上的优秀程度,从而选取优秀程度最高的题目给当前考生作答,将其称为优秀度指标[17](excellence degree, ED)方法。He和Chen还从项目参数估计误差的角度提出了D-c设计[18]。

以往关于在线标定的研究很少关注MCAT中项目的标定,本文将D-最优设计和A-最优设计从单维推广到多维,并且基于Fisher信息的视角,评估不同在线标定设计的表现,并探讨样本大小和能力间的相关系数对不同标定设计的影响,为研究者在MCAT中选择最优设计策略提供参考。

本文剩余部分按如下方式进行组织:下一节将对方法进行描述(包括IRT模型、Fisher信息矩阵的推导以及D-最优和A-最优设计等)。第三节进行了MCAT和在线标定的模拟研究,第四节对其研究结果进行了总结,第五节讨论了本文的不足以及未来的研究方向。

2. 研究方法

2.1. 多维项目反应理论模型

多维项目反应理论模型是表征被试多维潜在能力水平与其在题目上正确作答概率关系的非线性数学函数, 通过该模型可以对被试多个维度的潜在能力水平进行推断。多维项目反应理论模型还可被分为补偿模型和非补偿模型。补偿模型是指被试在某一维度上能力不足时可以由其他维度上的能力得到补偿。本文采用补偿的多维两参数 logistic 模型[19]。对于一个二分得分的项目 j , 用补偿多维两参数 logistic 模型(M2PL)定义的能力向量为 θ_i 的被试 i 正确回答该项目的概率见式(1):

$$P_j(\theta_i) = \text{Prob}(y_{ij} = 1_j | \theta_i, a_j, b_j) = \frac{1}{1 + \exp[-(\mathbf{a}_j^T \theta_i - b_j)]} \quad (1)$$

其中, y_{ij} 是一个二进制随机变量取 0 或 1。 $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})^T$ 表示被试 i 的 p 维能力向量。 $\mathbf{a}_j^T = (a_{j1}, a_{j2}, \dots, a_{jp})$ 表示与能力向量相对应的项目区分度向量, 反应了被试正确作答项目时每个维度的相对重要性, 其中 $\mathbf{a}_j^T \theta_i = a_{j1} \theta_{i1} + a_{j2} \theta_{i2} + \dots + a_{jp} \theta_{ip}$ 。 b_j 为项目 j 的难度参数, 反应项目对于被试作答的一个难易程度。此处, 当维数 p 等于 1 时, 能力向量和区分度向量退化成一个数值, 此时多维两参数 logistic 模型退化为单维的两参数 logistic 模型。

在多维计算机自适应测验和新题的在线标定过程中, 多维项目反应理论模型中参数的已知情况是不同的。多维计算机自适应测验的目的是为了对被试的能力做出估计, 评价被试的能力水平。其过程是在题库的已有题目中抽取合适的题目给予被试作答, 所以题库中题目的参数是已知的, 即项目区分度向量 $\mathbf{a}_j^T = (a_{j1}, a_{j2}, \dots, a_{jp})$ 和难度向量 b_j 已知, 而被试的能力向量 $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})^T$ 是未知的, 需要在测试中进行估计。然而在在线标定过程中, 我们的目的是为了对题目的参数进行估计, 并将新题与旧题至于同一尺度上。这一过程是通过分配能力已知的被试作答新题, 从而获得新题的题目参数信息。因此在此过程中, 需要基于被试的能力向量 $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})^T$ (能力一般也是未知的, 可以用估计值替代真值, 或者加入先验信息), 对新题参数进行估计和研究, 而新题的题目参数 $\mathbf{a}_j^T = (a_{j1}, a_{j2})$ 和 b_j 都是未知, 需要我们要进行估计和研究。

在本文中, 我们主要讨论当被试的能力向量为二维向量时的情况, 因此我们假设 $p = 2$, 此时被试 i 的能力向量 $\theta_i = (\theta_{i1}, \theta_{i2})^T$, 与之对应的项目区分度向量 $\mathbf{a}_j^T = (a_{j1}, a_{j2})$, $\mathbf{a}_j^T \theta_i = a_{j1} \theta_{i1} + a_{j2} \theta_{i2}$ 。 b_j 依旧为项目 j 的难度参数, 是一个一维数值。接下来本文所描述的模型和模拟实验研究都基于 $p = 2$ 的情况。

2.2. 自适应在线标定设计

在线标定设计的内涵即在被试作答过程中, 在被试到达新题位置时, 以一定的准则函数为依据, 将待标定的新题合理地分配给被试作答, 通过被试作答得到题目信息, 达到准确有效地估计新题的未知参数的目的。

在多维计算机自适应测验中, 在线标定新题的设计空间不是固定的, 因此针对不同的被试, 需要在未标定的新题中合理选择合适的新题给予被试作答。对此, van der Linden 和 Ren 研究出 UCAT 下基于观测信息矩阵的在线标定设计[16]。本文将在线标定设计中的 D-最优和 A-最优设计推广至适用于标定 MCAT 的情境。

2.2.1. Fisher 信息量矩阵的推导

D-最优和 A-最优设计基于 Fisher 信息矩阵, 而 Fisher 信息量矩阵是由被试对新题作答所提供的 Fisher 信息量所构成的信息矩阵[10]。令 $I(\boldsymbol{\eta}_j; \theta_i)$ 表示被试 i 对第道新题 j 提供的 Fisher 信息量, 则:

$$\begin{aligned}
 I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) &= -E \left[\frac{\partial^2}{\partial \boldsymbol{\eta}_j \partial \boldsymbol{\eta}_j'} \log L(u_{ij}; \boldsymbol{\theta}_i, \boldsymbol{\eta}_j) \right] \\
 &= -E \left\{ \frac{\partial^2}{\partial \boldsymbol{\eta}_j \partial \boldsymbol{\eta}_j'} \log \left[(P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j))^{u_{ij}} (Q(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j))^{1-u_{ij}} \right] \right\} \\
 &= -E \left\{ \frac{\partial^2}{\partial \boldsymbol{\eta}_j \partial \boldsymbol{\eta}_j'} \left[u_{ij} \log P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) + (1-u_{ij}) \log Q(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) \right] \right\}
 \end{aligned} \tag{2}$$

其中, $L(u_{ij}; \boldsymbol{\theta}_i, \boldsymbol{\eta}_j)$ 表示作答反应 u_{ij} 的似然函数, $\boldsymbol{\eta}_j$ 表示第 j 道新题的题目参数, $Q(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) = 1 - P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)$ 。在 M2PL 模型下, Fisher 信息矩阵可以表示为:

$$I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) = \begin{bmatrix} I_{a_{j1}a_{j1}} & I_{a_{j1}a_{j2}} & I_{a_{j2}b_j} \\ I_{a_{j1}a_{j2}} & I_{a_{j2}a_{j2}} & I_{a_{j2}b_j} \\ I_{a_{j2}b_j} & I_{a_{j2}b_j} & I_{a_{j2}b_j} \end{bmatrix} \tag{3}$$

下面给出 $I_{a_{j1}a_{j1}}$ 的计算公式, 其余分量的计算可类比 $I_{a_{j1}a_{j1}}$ 的计算得到。

$$\begin{aligned}
 I_{a_{j1}a_{j1}} &= -E \left[\frac{\partial^2}{\partial a_{j1} \partial a_{j1}} \log L(u_{ij}; \boldsymbol{\theta}_i, \boldsymbol{\eta}_j) \right] \\
 &= -E \left\{ \frac{\partial^2}{\partial a_{j1} \partial a_{j1}} \log \left[(P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j))^{u_{ij}} (Q(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j))^{1-u_{ij}} \right] \right\} \\
 &= -E \left\{ \frac{\partial^2}{\partial a_{j1} \partial a_{j1}} \left[u_{ij} \log P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) + (1-u_{ij}) \log Q(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) \right] \right\} \\
 &= -E \left\{ \frac{\partial}{\partial a_{j1}} \left[\frac{\partial}{\partial a_{j1}} \left[u_{ij} \log P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) + (1-u_{ij}) \log Q(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) \right] \right] \right\} \\
 &= -E \left\{ \frac{\partial}{\partial a_{j1}} \left[u_{ij} \frac{1}{P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)} \frac{\theta_{i1} \exp(-a_{j1}\theta_{i1} - a_{j2}\theta_{i2} + b_j)}{[1 + \exp(-a_{j1}\theta_{i1} - a_{j2}\theta_{i2} + b_j)]^2} \right. \right. \\
 &\quad \left. \left. + (1-u_{ij}) \frac{1}{1-P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)} \frac{-\theta_{i1} \exp(-a_{j1}\theta_{i1} - a_{j2}\theta_{i2} + b_j)}{[1 + \exp(-a_{j1}\theta_{i1} - a_{j2}\theta_{i2} + b_j)]^2} \right] \right\} \\
 &= -E \left\{ \frac{\partial}{\partial a_{j1}} \left[\theta_{i1} u_{ij} (1-P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)) - \theta_{i1} (1-u_{ij}) P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) \right] \right\} \\
 &= -E \left\{ \frac{\partial}{\partial a_{j1}} \left[\theta_{i1} (u_{ij} - P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)) \right] \right\} \\
 &= -E \left[\theta_{i1} \frac{-\theta_{i1} \exp(-a_{j1}\theta_{i1} - a_{j2}\theta_{i2} + b_j)}{[1 + \exp(-a_{j1}\theta_{i1} - a_{j2}\theta_{i2} + b_j)]^2} \right] \\
 &= -E \left[-\theta_{i1}^2 P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) (1-P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)) \right] \\
 &= \theta_{i1}^2 P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) (1-P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j))
 \end{aligned} \tag{4}$$

同理可计算得到其他分量:

$$\begin{aligned} I_{a_1a_2} &= \theta_{i1}\theta_{i2}P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)(1-P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)) \\ I_{a_1b_j} &= -\theta_{i1}P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)(1-P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)) \\ I_{a_2a_2} &= \theta_{i2}^2P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)(1-P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)) \\ I_{a_2b_j} &= -\theta_{i2}P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)(1-P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)) \\ I_{b_jb_j} &= P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)(1-P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j)) \end{aligned}$$

由此可得到, 在 M2PL 模型下, Fisher 信息量矩阵的具体表达形式为:

$$I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_i) = (1-P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j))P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j) \begin{bmatrix} \theta_{i1}^2 & \theta_{i1}\theta_{i2} & -\theta_{i1} \\ \theta_{i1}\theta_{i2} & \theta_{i2}^2 & -\theta_{i2} \\ -\theta_{i1} & -\theta_{i2} & 1 \end{bmatrix} \quad (5)$$

在局部独立性的条件下, n 名被试为题目 j 提供的 Fisher 信息量 $I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_n^*)$ 等于每名被试提供的 Fisher 信息量之和, $I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_n^*) = \sum_{i=1}^n I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_i)$, 其中 $\boldsymbol{\theta}_n^* = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n)^\top$ 是 n 名被试的能力向量矩阵, θ_{i1} 和 θ_{i2} 表示第 i 名被试的能力的第 1 和第 2 维。

2.2.2. D-最优设计

假设被试 i 达到新题位置时, 第 j 道新题已经被 n_j 名被试所作答。则基于 Fisher 信息量矩阵的 D-最优设计应从新题集合中选择满足条件:

$$k = \arg \max_j \left\{ \det \left[I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j}) + I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_i) \right] - \det \left[I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j}) \right] \right\} \quad (6)$$

的新题 k 分配给被试作答[16]。将旧题的 Fisher 信息量矩阵设为:

$$I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j}) = \left[1 - P(\boldsymbol{\theta}_{n_j};\boldsymbol{\eta}_j) \right] P(\boldsymbol{\theta}_{n_j};\boldsymbol{\eta}_j) \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{12} & C_{22} & C_{23} \\ C_{13} & C_{23} & C_{33} \end{bmatrix}, \quad (7)$$

结合(7)得:

$$I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j}) + I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_i) = (1 - P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j))P(\boldsymbol{\theta}_i;\boldsymbol{\eta}_j) \begin{bmatrix} C_{11} + \theta_{i1}^2 & C_{12} + \theta_{i1}\theta_{i2} & C_{13} - \theta_{i1} \\ C_{12} + \theta_{i1}\theta_{i2} & C_{22} + \theta_{i2}^2 & C_{23} - \theta_{i2} \\ C_{13} - \theta_{i1} & C_{23} - \theta_{i2} & C_{33} + 1 \end{bmatrix}. \quad (8)$$

将式(8)代入式(6)得到因式分解后的 D-最优设计的表达式为:

$$\begin{aligned} \arg \max_j \left\{ \left[1 - P(\boldsymbol{\theta}_{n_j};\boldsymbol{\eta}_j) \right] P(\boldsymbol{\theta}_{n_j};\boldsymbol{\eta}_j) \left[-2\theta_{i1} \left(I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j})_{[1,3]} \right) + 2\theta_{i2} \left(I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j})_{[2,3]} \right) \right. \right. \\ \left. \left. + 2\theta_{i1}\theta_{i2} \left(I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j})_{[1,2]} \right) + \theta_{i1}^2 \left(I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j})_{[1,1]} \right) + \theta_{i2}^2 \left(I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j})_{[2,2]} \right) \right] \right\}. \quad (9) \end{aligned}$$

式(9)中, $\boldsymbol{\theta}_i$ 表示第 i 名被试的能力向量, $\boldsymbol{\theta}_{n_j}$ 表示 n_j 名被试的能力向量矩阵, $I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j}) = \sum_{i=1}^{n_j} I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_i)$ 表示前 n_j 名被试为 j 道新题提供的总的 Fisher 信息量。 $I(\boldsymbol{\eta}_i;\boldsymbol{\theta}_i)$ 表示被试 i 对第 j 道新题提供的 Fisher 信息量。 $I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j})_{[l_1,l_2]}$ 表示 $I(\boldsymbol{\eta}_j;\boldsymbol{\theta}_{n_j})$ 省略第 l_1 行和第 l_2 列的子矩阵。D-最优设计的含义是被试 i 最有利于

增加新题 k 的参数向量 $\boldsymbol{\eta}_k$ 的广义方差的倒数。

2.2.3. A-最优设计

A-最优设计应从新题集合中选择满足条件:

$$k = \arg \max_j \left\{ \text{tr} \left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j})^{-1} - \left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) + I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j})^{-1} \right] \right] \right\} \quad (10)$$

的新题 k 分配给被试作答[16]。新题的 Fisher 信息量矩阵的逆矩阵为:

$$\text{tr} \left(I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i)^{-1} \right) = \frac{\sum_{l=1}^3 \det \left(\left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) \right]_{[l,l]} \right)}{\det \left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) \right]}, \quad (11)$$

其中新题和旧题的 Fisher 信息量矩阵和的逆矩阵的迹为:

$$\text{tr} \left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) + I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j})^{-1} \right] = \frac{\sum_{l=1}^3 \det \left(\left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) + I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) \right]_{[l,l]} \right)}{\det \left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) + I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) \right]}, \quad (12)$$

将式(11)和式(12)代入式(10)得:

$$\arg \max_j \left\{ \frac{\sum_{l=1}^3 \det \left(\left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) \right]_{[l,l]} \right)}{\det \left(I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) \right)} - \frac{\sum_{l=1}^3 \det \left(\left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) + I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) \right]_{[l,l]} \right)}{\det \left(I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) + I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) \right)} \right\}. \quad (13)$$

式(13)中, $I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j})^{-1}$ 为总的 Fisher 信息量矩阵的逆矩阵。 $\left(I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) \right)_{[l,l]}$ 表示 $I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j})$ 省略第 l_1 行和第 l_2 列的子矩阵。 $\left[I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) + I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) \right]_{[l,l]}$ 表示 $I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j}) + I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i)$ 省略第 l_1 行和第 l_2 列的子矩阵。A-最优设计的含义是被试 i 最有利于降低新题 k 的被估计参数向量 $\boldsymbol{\eta}_k$ 的方差之和。

2.2.4. 图形示例

为了对两种最优性准则的选题标准有一个初步印象, 本文绘制了两个二维项目的曲面。本文主要考虑以下两名被试:

$$\text{被试 1: } \boldsymbol{\theta}_1 = (-1, 1),$$

$$\text{被试 2: } \boldsymbol{\theta}_2 = (0, 2)。$$

第一名被试的两维能力分别是 -1 和 1, 第二名被试的两维能力分别是 0 和 2, 两名被试的相同点是两维能力都相差 2, 不同点是被试 1 的两维能力以 0 为对称点, 而被试 2 的能力以 1 为对称点。对于所有的新题, 将前 n_j 名被试对第 j 道新题提供的总的 Fisher 信息量 $I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_{n_j})$ 设置为 $\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$, 可以较为显著地观察到不同最优设计的曲面的区别。

下图显示两名能力固定的被试在不同区分度下的 D-最优设计的 Fisher 信息量, 更高的曲面代表了更高的信息量。由于被试的能力一样, 曲面的信息完全由公因数 $(1 - P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j))P(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)$ 决定, D-最优设计和 A-最优设计之间的差异仅仅是由于它们将 $\boldsymbol{\theta}_{ii}$ 映射到一维空间的不同方式造成的。

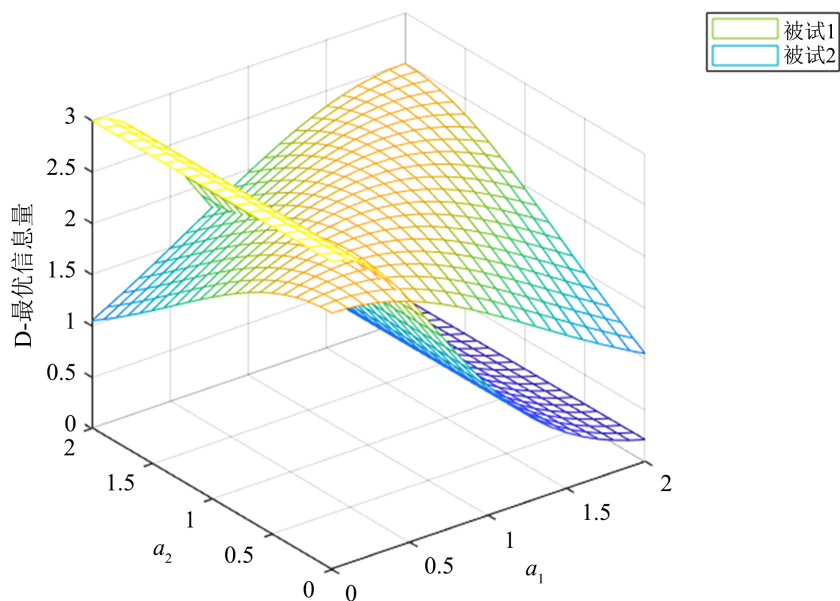


Figure 1. Parametric surface diagram of differentiation degree of D-optimal design for two examinees

图 1. D-最优设计在两名被试上的区分度参数曲面图

图 1 的曲面说明了被试 1 在两维区分度相等时 D-最优设计的信息量比较大，而在单一区分度较大而另一维区分度较小时，D-最优设计的信息量比较小。对于被试 2，D-最优设计的信息量随着 a_1 的增大而减小，而不随着 a_2 的变化而变化。综合来看，被试 1 的曲面要比被试 2 的曲面高，说明被试 1 较被试 2 的 D-最优设计的信息量大。图 2 显示了两名能力固定的被试在不同区分度下的 A-最优设计的 Fisher 信息量。

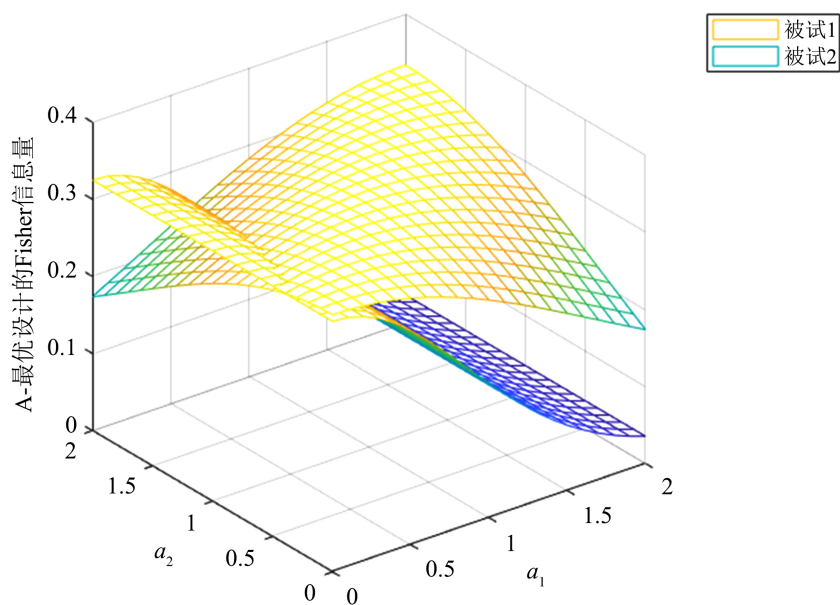


Figure 2. Parametric surface diagram of differentiation degree of A-optimal design for two examinees

图 2. A-最优设计在两名被试上的区分度参数曲面图

图 1 的曲面和图 2 的曲面的相同点是形状相同，不同点是对比 D-最优设计，在 A-最优设计中被试 1 要比被试 2 的曲面高的更加明显，说明 A-最优设计对被试 1 的偏好更强。

3. 模拟研究

本文考察基于可补偿的 2PL 模型上的 MCAT，探究在线标定过程中的 D-最优设计和 A-最优设计两种新题分配策略，并且基于 Fisher 信息量比较不同样本量和能力间的相关系数对这两种策略的影响。

3.1. 数据的生成

本文参考 Chen 等人的方法，随机生成 1000 道基于 M2PL 模型的题目的题库[7]，同前文相同，我们将多维模型设定为二维进而对其适当简化，以方便我们研究。其中题目区分度参数 $a_j^T = (a_{j1}, a_{j2})$ 中的两个分量 a_{j1}, a_{j2} 分别从服从[0, 1.3]的对数正态分布，且在[0.25, 2]的区间之中抽取，题目难度参数 b_j 从服从标准正态分布的总体之中抽取。

本文模拟 20 道新题进行在线标定过程。其中，这 20 道新题的参数模拟同题库中旧题的参数模拟方式相同，题目区分度参数 $a_j^T = (a_{j1}, a_{j2})$ 服从对数正态分布，且在[0.2, 2.5]的区间之中抽取，题目难度参数 b_j 从服从标准正态分布的总体之中抽取。题库与新题的项目参数的描述性统计见表 1。

借鉴以往 MCAT 的研究[20]，为了更为准确的选择新题进行在线标定过程，随机生成 1000、2000、3000 名被试，对应的每道新题分别由 200、400、600 名被试作答。其中，被试的二维能力向量 $\theta_i = (\theta_{i1}, \theta_{i2})^T$ 从服从均值为 0，方差为 1，相关系数分别为 0.2、0.5、0.8 的二维正态分布的总体中抽取。

Table 1. Item description

表 1. 题目描述

	a_1	a_2	b
均值	0.9986	1.0266	0.0446
标准差	0.6056	0.5954	0.9939
最小值	0.2007	0.2047	-3.4915
最大值	2.4871	2.4964	3.2662

3.2. 新题的在线标定过程

新题的在线标定过程包括对新题的标定和 MCAT 两个部分，两个部分同时进行，被试在进行 MCAT 的过程中同时植入新题，对新题进行标定。首先我们对被试的 MCAT 的过程进行描述。

CAT 的模拟主要由五个部分组成，分别为题库、初始题目选择方法、能力估计方法、选题策略和终止规则[21]。具体来说，1) 以 M2PL 模型为指导选择题库。2) 从题库中随机选取一道题目作为试题的初值。3) 采用极大似然估计方法对能力进行更新。4) 采用 D-最优选题策略。5) 采用固定测验长度作为终止规则。

本文将旧题的题目参数和被试的能力值代入两维的 2PL 模型中，可以计算被试 i 对于题目 j 的作答概率 p_{ij} ，作答概率 p_{ij} 是一个介于 0 到 1 之间的数。同时我们在 0 到 1 内随机生成一个随机数 ε_{ij} ，若作答概率 p_{ij} 大于这个随机数 ε_{ij} ，则认为被试 i 答对了题目 j ，被试的作答反应 u_{ij} 则记为 1。反之，若作答概率 p_{ij} 小于这个随机数 ε_{ij} ，则认为被试 i 没有答对题目 j ，被试的作答反应 u_{ij} 则记为 0。随后依据上文所提到的 D-最优选题策略，在题库中选择合适的题目给予被试作答。

同时，本文实验的主要目的是研究对新题的在线标定，因此在被试作答时需要同时植入新题，植入新题时，新题的分配方法通过上文提到的 D-最优设计和 A-最优设计分配。而在何时植入新题的问题上，本文参考 van der Linden 等人模拟实验的研究[9]，当每个被试作答 5 道新题时，我们在被试作答的所有题目的后八个位置中随机选取五个位置植入新题。在被试作答完成 20 道题目时，则停止作答，结束该被试的 MCAT 过程。

在本文的实验中，作答长度为 20 题，其中被试作答 5 道新题和 15 道旧题。本文设置了每道新题被作答次数的对比实验，假设每道新题分别被作答 200, 400, 600 次后则表示标定完成。所以在此设定自适应在线标定过程的终止规则，即每名被试作答 5 道新题，在这两种作答情况下，当每一道新题都被作答过 200, 400, 600 次后，则完成新题的在线标定过程，整个实验结束。

3.3. 评价标准

为了观察不同条件下两种在线标定设计策略的表现，我们需要构建信息函数，作为评价两种不同分配策略的准则。在此处我们构建 4 个信息函数，分别为被试为题目提供的总信息函数，以及题目参数向量 $\boldsymbol{\eta} = (a_1, a_2, b)$ 中三个题目参数分量各自的信息函数。信息量越大说明被试为新题提供的信息量越多，越有利于在线标定。

前文提到，Fisher 信息量矩阵的含义是被试对新题作答后为新题所提供的总的信息量，因此 Fisher 信息量矩阵可以作为衡量被试为新题提供信息量的大小的准则。本文不单独探究每一位被试对新题提供的信息量，而是研究所有参与作答新题的被试所提供的总的信息量，因此本篇文章将所有作答新题的被试的 Fisher 信息量矩阵进行加和，从而得到一个总的 Fisher 信息量矩阵，这个总的 Fisher 信息量矩阵我们用 $I(\boldsymbol{\eta}_j)$ 来表示，表示所有被试对新题所提供的 Fisher 信息量矩阵。

$$I(\boldsymbol{\eta}_j) = \sum_{i=1}^n I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) = \sum_{i=1}^n (1 - p(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)) p(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) \begin{bmatrix} \theta_{i1}^2 & \theta_{i1}\theta_{i2} & -\theta_{i1} \\ \theta_{i1}\theta_{i2} & \theta_{i2}^2 & -\theta_{i2} \\ -\theta_{i1} & -\theta_{i2} & 1 \end{bmatrix}. \quad (14)$$

其中 i 表示被试， n 表示作答新题的被试总数， j 表示第 j 道新题。我们对这个 Fisher 信息量矩阵取行列式，即可将其数值化从而便于我们进行比较。因此我们可以设新题的总信息函数 $I(F)$ ，其内涵是将作答新题的所有被试为新题提供的 Fisher 信息量矩阵相加，再取其行列式，则可以得到被试为新题提供的总的信息量。具体表达形式为：

$$I(F) = \det \left[\sum_{i=1}^n I(\boldsymbol{\eta}_j; \boldsymbol{\theta}_i) \right] = \det \left[\sum_{i=1}^n (1 - p(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j)) p(\boldsymbol{\theta}_i; \boldsymbol{\eta}_j) \begin{bmatrix} \theta_{i1}^2 & \theta_{i1}\theta_{i2} & -\theta_{i1} \\ \theta_{i1}\theta_{i2} & \theta_{i2}^2 & -\theta_{i2} \\ -\theta_{i1} & -\theta_{i2} & 1 \end{bmatrix} \right]. \quad (15)$$

其中加和对矩阵中的每一个分量都单独相加，可以得到：

$$\begin{aligned} I(a_{j1}, a_{j1}) &= \sum_{i=1}^n I_{a_{j1}a_{j2}}, & I(a_{j1}, a_{j2}) &= \sum_{i=1}^n I_{a_{j1}a_{j2}}, \\ I(a_{j1}, b_j) &= \sum_{i=1}^n I_{a_{j1}b_j}, & I(a_{j2}, a_{j2}) &= \sum_{i=1}^n I_{a_{j2}a_{j2}}, \\ I(a_{j2}, b_j) &= \sum_{i=1}^n I_{a_{j2}b_j}, & I(b_j, b_j) &= \sum_{i=1}^n I_{b_jb_j}, \end{aligned}$$

则

$$I(\boldsymbol{\eta}_j) = \begin{bmatrix} I(a_{j1}, a_{j1}) & I(a_{j1}, a_{j2}) & I(a_{j1}, b_j) \\ I(a_{j1}, a_{j2}) & I(a_{j2}, a_{j2}) & I(a_{j2}, b_j) \\ I(a_{j1}, b_j) & I(a_{j2}, b_j) & I(b_j, b_j) \end{bmatrix}. \quad (16)$$

杜文久等人曾提出, 当作答被试的数量很大时, 题目参数分布近似于正态分布, 即 $\boldsymbol{\eta}_j = (a_{j1}, a_{j2}, b)$ 服从三维正态分布, 其中各个题目参数均服从一维的正态分布[22]。令:

$$I^{-1}(\boldsymbol{\eta}_j) = \frac{I^*(\boldsymbol{\eta}_j)}{|I(\boldsymbol{\eta}_j)|}. \quad (17)$$

其中 $I^*(\boldsymbol{\eta}_j)$ 表示 Fisher 信息量矩阵的伴随矩阵, $|I(\boldsymbol{\eta}_j)|$ 表示 Fisher 信息量矩阵的行列式的值, Fisher 信息量矩阵的逆矩阵的对角线元素:

$$I_{11} = \frac{I(a_{j2}, a_{j2})I(b_j, b_j) - I(a_{j2}, b_j)^2}{|I(\boldsymbol{\eta}_j)|} \quad (18)$$

$$I_{22} = \frac{I(a_{j1}, a_{j1})I(a_{j2}, b_j) - I(a_{j1}, a_{j2})I(a_{j1}, b_j)}{|I(\boldsymbol{\eta}_j)|} \quad (19)$$

$$I_{33} = \frac{I(a_{j1}, a_{j1})I(a_{j2}, a_{j2}) - I(a_{j1}, a_{j2})^2}{|I(\boldsymbol{\eta}_j)|} \quad (20)$$

分别为题目参数 (a_{j1}, a_{j2}, b_j) 的方差。则:

$$D(a_{j1}) = \frac{I(a_{j2}, a_{j2})I(b_j, b_j) - I(a_{j2}, b_j)^2}{|I(\boldsymbol{\eta}_j)|} \quad (21)$$

$$D(a_{j2}) = \frac{I(a_{j1}, a_{j1})I(a_{j2}, b_j) - I(a_{j1}, a_{j2})I(a_{j1}, b_j)}{|I(\boldsymbol{\eta}_j)|} \quad (22)$$

$$D(a_{j3}) = \frac{I(a_{j1}, a_{j1})I(a_{j2}, a_{j2}) - I(a_{j1}, a_{j2})^2}{|I(\boldsymbol{\eta}_j)|} \quad (23)$$

由此可分别构造三个题目参数的信息函数 $I(a_1), I(a_{j2}), I(b_j)$ 。

令:

$$I(a_{j1}) = \frac{1}{D(a_{j1})} \quad (24)$$

$$I(a_{j2}) = \frac{1}{D(a_{j2})} \quad (25)$$

$$I(b_j) = \frac{1}{D(b_j)} \quad (26)$$

三个题目参数的信息函数分别表示所有被试对新题作答为新题的三个题目参数所分别提供的信息量, 该信息函数的数值越大, 说明被试的作答为不同的题目参数所提供的信息越多。

由此, 经过上文所构造的四个信息函数, 根据被试的作答结果分别计算信息函数, 从而来判断在不

同实验条件下两种不同策略能够给予新题的信息量的大小。

4. 实验结果与分析

表 2、表 3 和表 4 分别呈现了在能力相关性分别为 0.2, 0.5, 0.8 时, 不同数量被试下 D-最优设计和 A-最优设计的总信息量 $I(F)$ 、区分度参数的信息量 $I(a_1)$ 和 $I(a_2)$ 以及难度参数的信息量 $I(b)$ 。本文将总信息量和题目参数的信息量分别除以相对应的作答人数, 对平均后的信息量进行比较分析。

Table 2. The D-optimal and A-optimal Fisher information of different number of examinees when the ability correlation is 0.2
表 2. 能力相关性为 0.2 时不同数量被试下 D-最优和 A-最优的 Fisher 信息量

样本量	200		400		600	
	D-最优	A-最优	D-最优	A-最优	D-最优	A-最优
$I(F)$	112.64	110.05	423.18	415.51	942.98	967.45
$I(a_1)$	0.1259	0.1249	0.1252	0.1243	0.1199	0.1194
$I(a_2)$	0.1127	0.1115	0.1058	0.1053	0.1112	0.1130
$I(b)$	0.1365	0.1373	0.1391	0.1386	0.1372	0.1369

Table 3. The D-optimal and A-optimal Fisher information of different number of examinees when the ability correlation is 0.5
表 3. 能力相关性为 0.5 时不同数量被试下 D-最优和 A-最优的 Fisher 信息量

样本量	200		400		600	
	D-最优	A-最优	D-最优	A-最优	D-最优	A-最优
$I(F)$	74.80	73.64	280.76	282.36	597.93	593.57
$I(a_1)$	0.0980	0.0969	0.0954	0.0957	0.0931	0.0930
$I(a_2)$	0.0914	0.0915	0.0872	0.0881	0.0852	0.0842
$I(b)$	0.1301	0.1297	0.1292	0.1291	0.1328	0.1330

Table 4. The D-optimal and A-optimal Fisher information of different number of examinees when the ability correlation is 0.8
表 4. 能力相关性为 0.8 时不同数量被试下 D-最优和 A-最优的 Fisher 信息量

样本量	200		400		600	
	D-最优	A-最优	D-最优	A-最优	D-最优	A-最优
$I(F)$	28.41	27.70	118.52	118.00	271.36	271.61
$I(a_1)$	0.0487	0.0481	0.0490	0.0490	0.0489	0.0490
$I(a_2)$	0.0459	0.0461	0.0473	0.0474	0.0469	0.0471
$I(b)$	0.1272	0.1272	0.1278	0.1280	0.1258	0.1261

将表 2、表 3 和表 4 结合来看, D-最优设计和 A-最优设计的总信息量 $I(F)$ 以及题目参数的信息量 $I(a_1)$ 、 $I(a_2)$ 和 $I(b)$ 并没有显著的差别, 而作答新题的不同数量的被试以及能力系数的相关性对在线标定的信息量影响较大。随着作答人数的增多, 无论哪种能力参数的相关性, 平均后的总信息量都逐渐变大, 而题目参数的信息量都是先减小后增大。随着两维能力相关系数的增大, 总信息量和题目参数的

信息量都逐渐变小,这是因为能力的相关性越强,被试提供给题目的 Fisher 信息量越小。当两维能力的相关性等于 0.2 或 0.5 时,区分度参数和难度参数的 Fisher 信息量并没有显著差别,随着作答人数的增多,题目参数的信息量都呈现先减少后增加的趋势。而当两维能力的相关性等于 0.8 时,总信息量和题目参数的信息量几乎不随着作答人数的变化而变化。

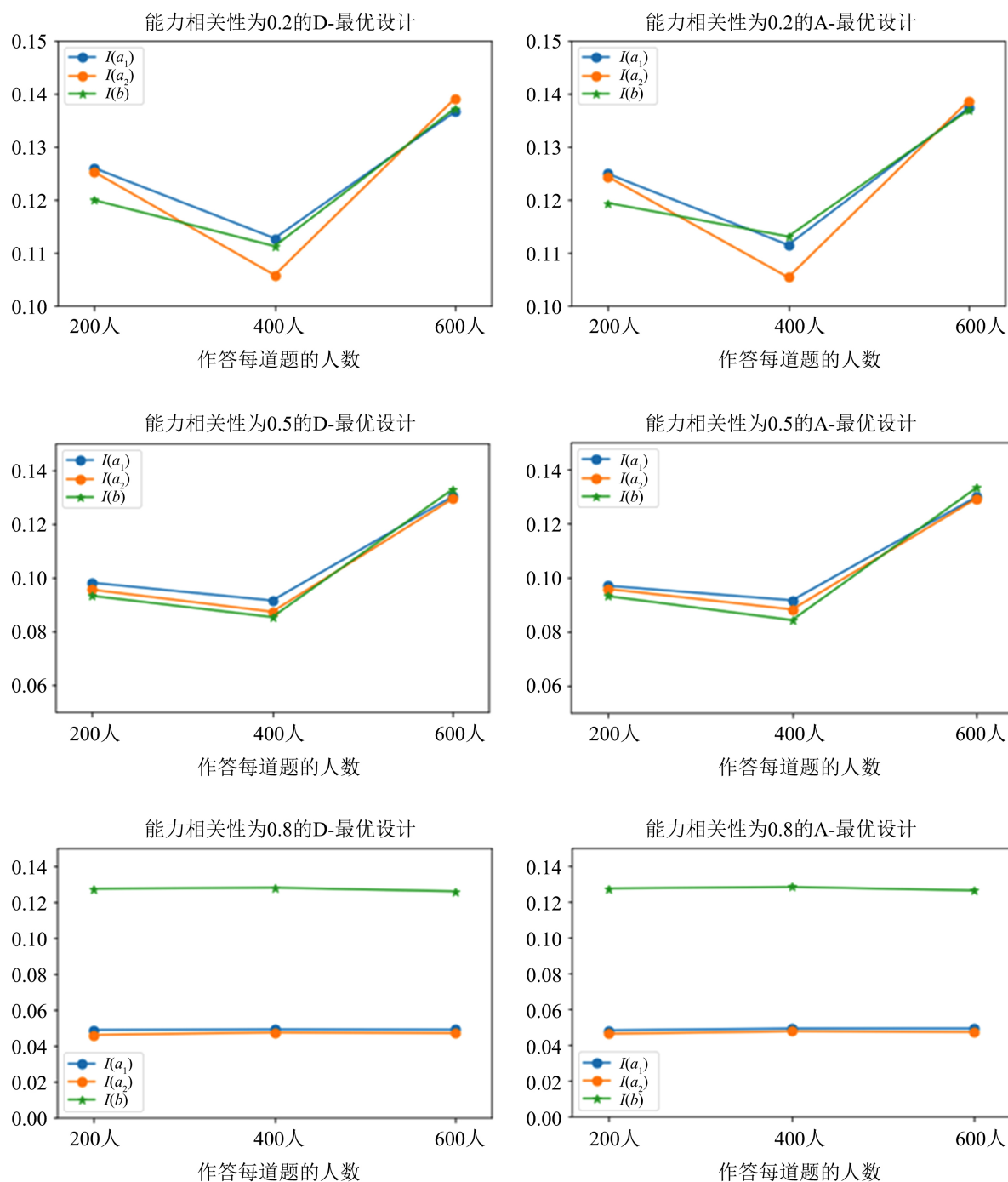


Figure 3. Information values for item parameters under different sample size and correlations in ability component generated by the D- and A-optimal designs

图 3. 不同样本量和能力相关性条件下 D-最优和 A-最优设计的题目参数信息量

为了更直观的看出不同的能力相关性和作答人数对两个最优设计的影响, 本文绘制了 D-最优设计和 A-最优设计在不同的能力相关性和作答人数的题目参数信息量的图形。

图 3 展示了在相关性为 0.2、0.5 和 0.8, 每道新题分别由 200 人、400 人和 600 人作答时, 在 D-最优设计和 A-最优设计下被试为新题提供的两维区分度参数与难度参数的信息量。对于最优设计对题目参数信息量的影响, 图中可以看出 D-最优设计和 A-最优设计在能力相关性为 0.2, 作答人数为 400 时, D-最优设计的 $I(a_1)$ 稍大于 $I(b)$, 而 A-最优设计则相反。对于能力的相关性对题目参数信息量的影响, 随着能力相关性的增大, 题目参数信息量显著降低。对于新题的作答人数对题目参数信息量的影响, 能力相关性为 0.8 时, 在不同作答人数的情况新题的下, 题目参数的信息量几乎持平, 且 $I(b)$ 显著大于 $I(a_1)$ 和 $I(a_2)$ 。

5. 结论与讨论

理论上, 本文将 D-最优设计和 A-最优设计从单维推广到了二维。首先可以弥补 UCAT 只能考察被试的单维能力以及题目的单一区分度, 其次还为未来对二维以上的在线标定的 MCAT 提供了理论依据。对于其他相关文献基本都是以题目的偏差和均方误差作为在线标定的评价指标, 本文基于 Fisher 信息量进行在线标定研究, 因此本文对评价指标总信息量 $I(F)$ 、题目参数信息量 $I(a_1)$ 和 $I(a_2)$ 以及难度参数 $I(b)$ 的 Fisher 信息量矩阵都进行了推导。

实践上, 本文以 Fisher 信息量为评价指标, 在不同实验条件下对 D-最优设计和 A-最优设计进行评价。首先本文对题库、新题以及被试进行了模拟, 使用极大似然估计方法得到能力的估计值。然后通过模拟 MCAT 以及在线标定的过程来完成对新题的在线标定。最后以总信息量 $I(F)$ 、题目参数信息量 $I(a_1)$ 和 $I(a_2)$ 为评价指标, 在不同能力的相关性以及样本量的实验条件下, 研究不同样本量和能力间的相关系数对 D-最优设计和 A-最优设计的影响。

并且本文还可以得出以下结论: 1) 随着作答人数的增多, 平均后的总信息量逐渐增大, 当能力间的相关性为 0.2 和 0.5 时, 平均后的题目参数的信息量呈现先减少后增加的趋势; 而当能力间的相关性为 0.8 时, 平均后的题目参数的信息量几乎不会改变。2) 随着能力间相关性的增大, 平均后的总信息量和题目参数的信息量都呈现逐渐减少的趋势。3) 当能力间的相关性太大时, 作答新题的人数对 Fisher 信息量几乎没有影响。

对于第一道题目的选择本文采用随机方法, 由于题目的初值一定程度上会影响到后面的 CAT 过程, 从而对新题的在线标定过程产生影响。张雪琴(2021)使用两阶段方法来确定题目的初值, 一定程度上减少了它对新题标定的影响。因此在后续的研究中, 将两阶段方法或其他方法对第一道题目进行选择是必要的。

本文研究的结果表明在不同样本量和能力间的相关系数的条件下, D-最优和 A-最优设计的结果相差很小。这可能是由于题目长度较小导致的, 未来可以在不同实验条件下探究 D-最优和 A-最优设计的区别, 或者是使用其他评价指标, 如均方差、KL 信息等对新题的在线标定进行评价, 这对于探究两种最优设计之间的联系与区别是十分必要的。

基金项目

本研究获得国家自然科学基金资助, 项目编号为 32000778; 和南京信息工程大学科研启动经费资助, 项目编号为 2018r041。

参考文献

- [1] 陈平, 张佳慧, 辛涛. 在线标定技术在计算机化自适应测验中的应用[J]. 心理科学进展, 2013(10): 1883-1892.

- [2] 张雪琴, 毛秀珍, 李佳. 基于 CAT 的在线标定: 设计与方法[J]. 心理科学进展, 2020, 28(11): 1970-1978.
- [3] Lu, H.Y. (2014) Application of Optimal Designs to Item Calibration. *PLoS ONE*, **9**, e106747. <https://doi.org/10.1371/journal.pone.0106747>
- [4] Wainer, H. and Mislevy, R.J. (1990) Item Response Theory, Item Calibration, and Proficiency Estimation. In: Wainer, H., Dorans, N.J., Green, B.F., Steinberg, L., Flaugher, R., Mislevy, R.J. and Thissen, D., Eds., *Computerized Adaptive Testing: A Primer*, Lawrence Erlbaum Associates, Inc., Mahwah, 65-102.
- [5] 游晓锋, 丁树良, 刘红云. 计算机化自适应测验中原始题项目参数的估计[J]. 心理学报, 2010, 42(7): 813-820.
- [6] Chen, P., Wang, C., Xin, T. and Chang, H.H. (2017) Developing New Online Calibration Methods for Multidimensional Computerized Adaptive Testing. *British Journal of Mathematical and Statistical Psychology*, **70**, 81-117. <https://doi.org/10.1111/bmsp.12083>
- [7] Chen, P. and Wang, C. (2016) A New Online Calibration Method for Multidimensional Computerized Adaptive Testing. *Psychometrika*, **81**, 674-701. <https://doi.org/10.1007/s11336-015-9482-9>
- [8] Wang, C., Chang, H.H. and Boughton, K.A. (2011) Kullback-Leibler Information and Its Applications in Multi-Dimensional Adaptive Testing. *Psychometrika*, **76**, 13-39. <https://doi.org/10.1007/s11336-010-9186-0>
- [9] Mulder, J. and Van der Linden, W.J. (2009) Multidimensional Adaptive Testing with Optimal Design Criteria for Item Selection. *Psychometrika*, **74**, 273-296. <https://doi.org/10.1007/s11336-008-9097-5>
- [10] Berger, M.P. (1992) Sequential Sampling Designs for the Two-Parameter Item Response Theory Model. *Psychometrika*, **57**, 521-538. <https://doi.org/10.1007/BF02294418>
- [11] Chang, Y.C.I. and Lu, H.Y. (2010) Online Calibration via Variable Length Computerized Adaptive Testing. *Psychometrika*, **75**, 140-157. <https://doi.org/10.1007/s11336-009-9133-0>
- [12] Hassan UI, M. and Miller, F. (2019) Optimal Item Calibration for Computerized Achievement Tests. *Psychometrika*, **84**, 1101-1128. <https://doi.org/10.1007/s11336-019-09673-6>
- [13] Ren, H., van der Linden, W.J. and Diao, Q. (2017) Continuous Online Item Calibration: Parameter Recovery and Item Utilization. *Psychometrika*, **82**, 498-522. <https://doi.org/10.1007/s11336-017-9553-1>
- [14] Buyske, S.G. (1998) Optimal Design for Item Calibration in Computerized Adaptive Testing: The 2PL Case. Lecture Notes-Monograph Series, 115-125. <https://doi.org/10.1214/lnms/1215456191>
- [15] Kang, H.A., Zheng, Y. and Chang, H.H. (2020) Online Calibration of a Joint Model of Item Responses and Response Times in Computerized Adaptive Testing. *Journal of Educational and Behavioral Statistics*, **45**, 175-208. <https://doi.org/10.3102/1076998619879040>
- [16] van der Linden, W.J. and Ren, H. (2015) Optimal Bayesian Adaptive Design for Test-Item Calibration. *Psychometrika*, **80**, 263-288. <https://doi.org/10.1007/s11336-013-9391-8>
- [17] He, Y., Chen, P. and Li, Y. (2020) New Efficient and Practicable Adaptive Designs for Calibrating Items Online. *Applied Psychological Measurement*, **44**, 3-16. <https://doi.org/10.1177/0146621618824854>
- [18] He, Y. and Chen, P. (2020) Optimal Online Calibration Designs for Item Replenishment in Adaptive Testing. *Psychometrika*, **85**, 35-55. <https://doi.org/10.1007/s11336-019-09687-0>
- [19] Reckase, M.D. (2009) Multidimensional Item Response Theory. Springer, New York. <https://doi.org/10.1007/978-0-387-89976-3>
- [20] 陈平. 两种新的计算机化自适应测验在线标定方法[J]. 心理学报, 2016, 48(9): 1184-1198.
- [21] 张雪琴. CD-CAT 在线标定设计的研究[D]: [硕士学位论文]. 成都: 四川师范大学, 2021.
- [22] 杜文久, 周娟, 李洪波. 二参数逻辑斯蒂模型项目参数的估计精度[J]. 心理学报, 2013(10): 1179-1186.