

## 《心理学报》审稿意见与作者回应

题目：认知诊断中基于条件期望的距离判别方法

作者：汪文义 丁树良 宋丽红

### 第一轮

尊敬的评审专家：

您好，非常感谢你们提出的宝贵意见和建议。我们根据你们所提的意见进行了仔细的修改，逐条回应了专家提出的问题和意见，并在文章中作了相应的修改。主要修改部分见文章中红色标注部分。

**审稿人 1 意见：**《认知诊断中基于条件期望的距离判别方法》一文运用新方法解决问题是值得推崇的。模拟研究部分采用的模拟思路清晰，模拟方法可行，采用的指标科学，可以说为认知诊断的分类问题提供了一个新的思路。但是，该文若要在《心理学报》这样高水平的杂志上发表，还需要下功夫。建议从以下几个方面着手：

**意见 1：**摘要不能告诉读者全文概览。

**回应：**谢谢专家提出的问题。我们参照《心理学论文写作规范》，从问题提出、研究方法、研究结果和应用价值四个方面，重新撰写了摘要。修改如下：

分类是认知诊断评估的一个核心问题。考察基于观察反应模式与理想反应模式之间的距离判别法，没有充分利用总体分布信息，其类中心是未考虑统计误差的理想反应模式，且。为了更充分地利用总体分布信息、提高诊断分类效果和拓展诊断评估的适用性，本研究提出基于条件期望向量为类中心的欧氏距离判别方法，同时提出认知诊断下的项目特征曲线估计方法以获得条件期望向量。模拟研究表明：认知诊断下的项目特征曲线估计方法得到的条件期望向量返真性较高，获得的分布信息较准确；在观察反应与理想反应差异大的情形下，基于条件期望向量为中心的欧氏距离判别方法，比基于理想反应模式的分类方法更优。研究可为认知诊断分类和诊断等值设计提供一个参考。

**意见 2：**问题提出部分，逻辑不清晰。从前人研究结论到本研究问题的提出很突兀。

**回应：**谢谢专家提出的问题。我们重新撰写了引言部分，对引言进行了精简和调整。具体写作思路如下：首先，指出以期望反应模式为类中心的判别方法（广义距离判别法和非参数分类法）的不足，如以期望反应模式为类中心的判别方法未充分利用总体分布信息、在某些条件下存在限制等。然后，为了克服这种问题、更充分地利用总体分布信息和提高诊断分类效果，通过借鉴判别分析中距离判别方法，提出以项目反应模式的条件期望向量为类中心的欧氏距离判别方法，并针对实际情况中条件期望经常未知的情形，提出估计条件期望的方法。

**意见 3：**本研究的预期实验结果是怎么提出的？不是很明确。虽然后续的模拟研究印证了作者提出的预期，但是作者需要对这些预期的提出做详细的阐述。任何研究预期都不是凭空产生的。

鉴于以上三点，建议大修后再审。

**回应：**谢谢专家提出的宝贵意见。我们试图根据第 2 节的内容，主要从理论上分析预期的结果，并对此部分进行了细致的修改。修改如下：

预期的实验结果是：(1) 在定理 1 条件满足时，条件期望法与经典的极大似然估计法表现相当，而在违背定理 1 条件情况下，由定理 2 条件期望法要好于非参数分类方法。这主要是基于以下两个方面的考虑：在满足定理 1 的条件，根据条件期望法，得分 0 将与相应知识状态的答对概率小于 0.5 接近，而得分 1 将与相应知识状态的答对概率大于 0.5 接近，这与极大似然估计思路相一致。若违背定理 1 条件情况下，被试答对概率大于 0.5 时，而相应的理想反应为 0，这时非参数分类方法将产生与极大似然估计不一致的结果。(2) 认知诊断模型项目特征曲线可较好的估计。这主要是 2PLM 的 ICC 可视为认知诊断模型 ICC 的拟合曲线，某能力附近较多知识状态的答对概率较大，则此能力处的答对概率较大；反过来，某能力处项目反应概率大，则此附近能力对应的较多知识状态的答对概率应该较大；(3) ICC 估计方法能比较准确估计条件期望，则估计的条件期望法表现较好。

**审稿人 2 意见：**本研究针对非参数分类法或广义距离判别法等已有方法的不足（如以期望反应模式为类中心、未充分利用总体分布信息、在某些条件下存在限制等），提出以条件期望向量为类中心的欧氏距离判别方法，并针对实际情况中条件期望经常未知的情形，提出估计条件期望的方法（即基于 2PLM 项目特征曲线估计认知诊断模型的项目特征曲线）。研究问题比较有价值，整个思路比较新颖，整篇文章的逻辑也很清晰。文章的不足之处描述如下，希望对文章的改进有所帮助：

**意见 1：**作者在第 6 页“2.2 以条件期望向量为类中心的欧氏距离判别法”部分第 3 段对条件期望法进行了详细描述，对于被试  $i$  的观察反应模式  $\mathbf{x}_i$  与条件期望向量  $E(\mathbf{x}_i | \boldsymbol{\alpha}_i)$  的欧氏距离，有没有可能出现“ $\mathbf{x}_i$  与多个  $E(\mathbf{x}_i | \boldsymbol{\alpha}_i)$  的欧氏距离相等”的情况？如果可能，应该如何处理比较合适？

**回应：**谢谢专家提出的问题。理想反应模式中的元素只有 0 和 1 两种可能，不同知识状态，可能在多个项目上的理想反应相同，在少数项目上发生 0 和 1 置换，因此  $\mathbf{x}_i$  可能与多个理想反应模式的欧氏距离相等。条件期望并不像理想反应模式，条件期望  $E(\mathbf{x}_i | \boldsymbol{\alpha}_i)$  通常为 0-1 之间的实数，由于项目质量不同，不同项目的条件期望相同的概率非常小，即能发生置换的概率相当小，因此  $\mathbf{x}_i$  与多个知识状态对应的条件期望的欧氏距离相等的概率非常小。我们通过模拟研究发现，非参数分类方法出现“ $\mathbf{x}_i$  与多个期望反应模式的欧氏距离相等”的比例有时接近 10% 甚至更高，而条件期望法基本上未出现“ $\mathbf{x}_i$  与多个  $E(\mathbf{x}_i | \boldsymbol{\alpha}_i)$  的欧氏距离相等”的情况。

**意见 2：**作者在“2.3 认知诊断模型的项目特征曲线估计”部分提出一个非常重要的概念，即认知诊断模型的项目特征曲线，这个概念应该如何理解？根据对后文的理解，它的纵坐标代表被试正确作答题目的概率，横坐标代表知识状态对应的能力估计值，那么如何保证能力值越大，其对应的知识状态处的正确作答概率也会越大呢？换句话说，如何保证它是一条单调递增的曲线呢？另外，请作者阐述“将 2PLM 项目特征曲线视为  $(\hat{\theta}_i, P_j(\boldsymbol{\alpha}_i))$  的回归曲线”的

合理性；

**回应：**谢谢专家提出的问题。您提到“如何保证认知诊断模型的项目特征曲线是一条单调递增的曲线”，这个假设不一定会得到保证（线性结构可以保证，但是其他属性层级结构下不一定能保证）。如您所说，认知诊断模型的项目特征曲线的横坐标代表知识状态对应的能力估计值，然而能力估计值越大对应的知识状态不一定掌握某个特定的项目的属性就越多，因此，正确作答概率不一定会随着能力量尺呈现递增的趋势。一般来说，对于特定的项目，如果知识状态按照掌握项目属性的多少进行从小到大排序，所绘制的项目特征曲线会呈现递增的曲线。而根据能力排序，其实与按照知识状态掌握项目属性的多少进行排序不尽相同。

您提到“将 2PLM 项目特征曲线视为  $(\hat{\theta}_i, P_j(\alpha_i))$  的回归曲线的合理性”。认知诊断模型的项目特征曲线是指各种知识状态对应的正确作答概率，而各种知识状态根据其作答反应可以估计其对应的能力，由此可绘制出各知识状态对应的能力点与相应的正确作答概率的散点图，而 2PLM 项目特征曲线正是被试在项目上正确作答概率对潜在特质分数的回归曲线(漆书青, 戴海崎, 丁树良, 2002)，其实就相当于对散点图做了回归曲线。并且在文章相应部分进行了补充说明。

**意见 3：**“3.2 研究设计”部分第 3 段，作者针对 5 种不同的属性层级结构设计了不同的测验项目数，分别是 30、35、30、32 和 41 题。但是作者在后面的结果分析中，似乎并未强调：5 种不同属性层级结构的结果差异也有可能部分是由测验长度的不同所贡献；“3.3 评价指标”部分中，除了使用 Bias、Abs 和 RMSE 等指标评价项目特征曲线的估计精度，如果可能的话，还建议将估计得到的 ICC 以图形方式呈现出来，更为直观；

**回应：**谢谢专家提出的问题。根据您的建议，我们在结果分析中的倒数第二段最后，补充了您所提到的原因：属性结构对分类精度和特征曲线估计精度有一定的影响，可能部分是由测验长度、可达阵项目重复数量或属性结构的复杂程度的不同所贡献。图 1 和图 2 任选了 20 个项目的 ICC 图进行了呈现。

**意见 4：**在“3.4 实验结果”部分中，作者对部分结果的描述还不够准确、对部分结果也未进行解释。比如，第 10 页第 1 段第 3 行“.....差异仅在 0.01 以下”，描述不完全准确。因为对于表 3 和表 4，基于模拟参数的条件期望法与极大似然估计法两种方法的分类精度差异多数在 0.01 以上。第 10 页第 2 段第 2 至 3 行“.....四种方法的总体表现相当接近”，结果描述不完全准确。在某些属性层级结构下，四种方法的表现其实相差蛮大。

**回应：**如您所述，第 10 页第 1 段第 3 行“.....差异仅在 0.01 以下”，描述不完全准确。修改如下：在满足定理 1 条件（表 1 至表 3）和不满足定理 1 条件下（表 4）：在不同属性结构、不同样本量大小、不同认知诊断模型下，基于模拟参数的条件期望法与极大似然估计法的分类精度十分接近，绝对偏差的最小值、最大值、均值和标准差分别为：0, 0.0268, 0.0068 和 0.0061。

第 10 页第 2 段第 2 至 3 行“.....四种方法的总体表现相当接近”，结果描述不完全准确。修改如下：四种方法的总体表现比较接近，很难找出一种表现最突出的方法。

在无结构与独立结构下，负偏的程度会比较大，原因尚不明确，值得进一步深入思考。

分类精度在某些属性层级结构下不完全随着样本的增大而提高，可能是因为不同样本量下模拟的项目参数不尽相同导致。

意见 5: 另外, 稿件中还存在一些编辑或描述上的不足或错误, 罗列如下:

回应: 各条意见逐条回应, 请见意见 6-20 及其回应。

意见 6: 第 3 页中文摘要部分最后 1 句话“模拟研究表明新方法表现不错”过于简单, 能否简单描述新方法如何表现不错?

回应: 谢谢专家提出的意见。已经修改如下: 模拟研究表明: 认知诊断下的项目特征曲线估计方法得到的条件期望向量返真性较高, 获得的分布信息较准确; 基于条件期望向量为中心的欧氏距离判别方法分类准确性高, 尤其在观察反应与理想反应差异大的情形下适用。

意见 7: 第 3 页“1 引言”部分第 1 段第 2 行“为学生提供认知强项和认知弱项”, 句子含义好像没有表达完整;

回应: 谢谢专家提出的意见。由于重新撰写了引言部分, 此句已经删除。

意见 8: 第 3 页最后 1 段第 2 行至第 5 行, 突然给出一些公式, 感觉有些唐突, 而且对公式中的符号也未进行简单解释;

回应: 谢谢专家提出的意见。综合您的意见和另一位审稿专家的意见, 我们重新撰写了引言部分, 对引言进行了精简和调整, 删除了一些影响整个引言逻辑和影响问题过渡的细节。具体参见引言部分。

意见 9: 第 4 页第 2 段第 1 行到第 5 行的句子, 改为“Sun, Xin, Zhang 和 de la Torre (2013) 以及孙佳楠, 张淑梅, 辛涛和包钰 (2011) 在……, 给出了一种识别被测知识状态的认知诊断方法, 即广义距离判别法 (GDD)”可能更好一些;

回应: 谢谢专家提出的意见。已经按照您的意见对此句进行了精简, 修改如下: 近年来, 研究者提出了以期望反应模型为类中心的判别方法, 如广义距离判别法和非参数分类法 (Sun, Xin, Zhang, & de la Torre, 2013; 孙佳楠, 张淑梅, 辛涛, 包钰, 2011; Chiu & Douglas, 2013; Wang & Douglas, 2013; 丁树良, 罗芬, 汪文义, 2013)。

意见 10: 第 4 页第 2 段倒数第 3 行至倒数 2 行提到的文献“Wang 和 Douglas (2013)”并未出现在正文后的参考文献部分;

回应: 谢谢专家提出的问题。这是我们的疏忽, 我们仔细核对了参考文献和文中的引用。

意见 11: 第 5 页第 4 行的“欧式距离”含错误字, 应该是“欧氏距离”。第 6 页及第 16 页也有多处“欧式距离”;

回应: 谢谢专家提出的问题。这是我们的疏忽, 已经将“欧式距离”替换为“欧氏距离”。

意见 12: 第 5 页第“2.1 以理想反应模式为类中心的非参数分类法”部分第 1 段倒数第 2 行的

公式存在错误, 应改为  $d_{h_i} = \sum_{j=1}^m |x_{ij} - \eta_{ij}|$ ;

回应: 谢谢专家提出的问题。已经按照您的意见进行了修改。并且我们对文章的公式, 尤其是下标进行了仔细的校对。

意见 13: 第 5 页定理 1 中“设  $\alpha_i$  与  $\mathbf{x}_i$  的海明距离最小”的描述有误, 因为  $\alpha_i$  与  $\mathbf{x}_i$  两者的维度大小不同, 正确的描述应该是“设  $\eta_i$  与  $\mathbf{x}_i$  的海明距离最小”;

回应: 谢谢专家细致的审阅。已经按照您的意见进行了修改。

意见 14: 第 6 页定理 2 中“设  $\alpha_i$  与  $\mathbf{x}_i$  的欧氏距离最小”的描述有误,  $\alpha_i$  与  $\mathbf{x}_i$  两者的维度大小也不同, 正确的描述应该是“设  $E(\mathbf{x}_i | \alpha_i)$  与  $\mathbf{x}_i$  的欧氏距离最小”;

回应: 谢谢专家细致的审阅。已经按照您的意见进行了修改。

意见 15: 第 7 页第 1 行与第 3 行中的“被试答错概率大于 0.5”描述有误, 根据前后逻辑, 应该改为“被试答对概率大于 0.5”;

回应: 谢谢专家细致的审阅。已经按照您的意见进行了修改。修改如下: 非参数分类法在定理 1 条件下才与极大似然估计相一致, 若未掌握项目所考察的所有属性的被试答对概率大于 0.5 时, 易导致与极大似然估计不一致的现象。

意见 16: 第 7 页“2.3 认知诊断模型的项目特征曲线估计”部分: 第 1 段第 5 行至第 7 行, 这里描述的样本量 (如 1000、1500 和 3000) 是不是与具体的题目个数有关? 即题目越多, 是不是需要的被试量也越多? 第 1 段第 5 行至第 6 行, R-RUM 翻译为“重新参数化的统一模型”可能更好; 第 2 段第 2 行, “因为项目项目反应模型特征曲线比较容易”的描述有问题; 第 3 段第 3 行的公式“ $P_j(\theta_i)$ ”有误, 应改为“ $P_j(\theta_j)$ ”;

回应: 谢谢专家细致的审阅。由于原文中并未对不同题目数下参数估计的样本量的要求进行研究, 我们也不能作出其他的推论, 只是在文中增加了其研究中使用的项目数。

其他问题已经按照您的意见进行了修改。R-RUM(the reduced reparameterized unified model)翻译为“简化的重新参数化的统一模型”。描述有问题的句子修改为: 获得项目反应模型的项目特征曲线(ICC)相对比较容易, 样本量要求无需太大。公式“ $P_j(\theta_i)$ ”已改为“ $P_j(\theta_j)$ ”。

意见 17: 第 8 页采用最小二乘法 (MLS) 进行估计的公式中,  $\mathbf{C}$  和  $\mathbf{P}_j$  分别代表什么含义?

回应: 列向量  $\mathbf{P}_j = (p_j(\theta_l))$  元素  $p_j(\theta_l)$  表示在指定能力采样点  $\theta_l$  上项目  $j$  的 2PLM 下正确

作答概率。其中  $\mathbf{C} = (C_{lr})$  元素  $C_{lr} = \begin{cases} \frac{1}{k}, & \hat{\theta}_l \text{ 属于与 } \theta_l \text{ 距离最接近的 } k \text{ 个能力之一} \\ 0, & \text{其他} \end{cases}$ 。

意见 18: 第 8 页“2.3 认知诊断模型的项目特征曲线估计”部分最后 1 段第 1 至第 2 行, “认知诊断模型下项目反应函数的单调性假设”如何体现?

回应: 谢谢专家提出的问题。“认知诊断模型下项目反应函数的单调性假设”如何体现, 主要通过对待估参数  $P_j(\alpha_i)$  增加约束条件来实现, 文中增加了下面的例子对这一点进行了说明。

例, 对于含三个属性的线性结构, 共有四类知识状态

$[\alpha_1, \alpha_2, \alpha_3, \alpha_4] = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ 。项目  $j$  考察的属性向量为  $q_j = [1 \ 0 \ 0]$ 。现在要

估计的参数向量，即各类知识状态在项目  $j$  正确作答概率为  $Y_j = [p_j(\alpha_1) \ p_j(\alpha_2) \ p_j(\alpha_3) \ p_j(\alpha_4)]^T$ 。DINA 模型假设掌握该项目的知识状态正确

作答概率相等，可令  $Aeq = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$ ,  $beq = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$  且  $AeqY = beq$ ，即保证

$\alpha_2, \alpha_3, \alpha_4$  三类知识状态的正确作答概率相等；如果要保证 DINA 模型的单调性假设，可令  $A = [1 \ -1 \ 0 \ 0]$ ,  $b = [0]$  且  $AY \leq b$ ，即可保证  $p_j(\alpha_2) \geq p_j(\alpha_1)$ ，即  $1 - s_j \geq g_j$ 。

意见 14：第 9 页第 2 段第 4 到第 5 行“知识状态分布按祝玉芳和丁树良 (2009) 的方法模拟，使总分的分布近似满足正态分布”，与上一段已经提到“各知识状态服从离散均匀分布”矛盾；  
 回应：经核查实验程序中的知识状态模拟时所采用的方法及得到的分布为：知识状态分布按祝玉芳和丁树良 (2009) 的方法模拟，使总分的分布近似满足正态分布。

意见 19：第 9 页“3.3 实验结果”的序号不对，应改为“3.4 实验结果”；  
 回应：谢谢您，这是我们的疏忽，已经根据您的意见进行了修改。

意见 20：第 12 页第 1 段第 2 行及第 4 行的“第 4、5 和 6 列”以及“第 7、8 和 9 列”应该分别改为“第 3、4 和 5 列”以及“第 6、7 和 8 列”；  
 回应：谢谢您如此细心的建议。已经根据您的意见进行了修改。

## 第二轮

审稿人 1 意见：

意见 1：此次为复审。在前次提出的审稿意见基础上，作者做了较大修改。本人认为已经达到发表水平，此意见仅供编辑部参考。再提醒下作者，摘要部分第二行“且”字后面多了个“。”。  
 回应：谢谢专家的肯定意见和所提的修改意见。我们已经对摘要进行了修改。特别是对引言、第 2 节和结果描述部分进行了仔细的修改。

审稿人 2 意见：

意见 1：作者较好地回答了审稿专家提出的诸多问题与意见，并在稿件中进行了相应修改。但是还存在几个问题值得商榷：

回应：各条意见逐条回应，请见意见 2-4 及其回应。

意见 2：(1)对于“将 2PLM 项目特征曲线 (ICC) 视为  $(\hat{\theta}_i, P_j(\alpha_i))$  的回归曲线”的合理性，本人仍存在疑惑，不太理解“视为”所代表的含义。一方面，作者在回应中提到：“认知诊断模型的 ICC 是指各种知识状态对应的正确作答概率 (注意这句话本身存在问题，曲线并非概

率)，而各种知识状态根据其作答反应可以估计其对应的能力，由此可绘制出各知识状态对应的能力点与相应的正确作答概率的散点图”。此时各知识状态对应的正确作答概率（即  $P_j(\alpha_t)$ ）未知，如何绘制散点图（疑惑 1）？另一方面，如果可将 2PLM 的 ICC 看成是  $(\hat{\theta}_t, P_j(\alpha_t))$  的回归曲线，是否可通过如下方式而非文中提到的最小二乘法 (MLS) 来确定  $P_j(\alpha_t)$  的值：首先根据知识状态  $\alpha_t$  对应的理想反应模式  $\eta_t$  估计其对应的能力值（如  $\hat{\theta}_t$ ），然后在 2PLM 的 ICC 上找到  $\hat{\theta}_t$  处的正确作答概率  $P_j(\hat{\theta}_t)$ ，直接作为  $P_j(\alpha_t)$ （疑惑 2）？

另外，按照文中“由项目反应理论中 2PLM 的 ICC 转换得到认知诊断模型的 ICC”的做法，虽然获得项目反应理论模型的 ICC 相对比较容易且样本量无需太大，但是从项目反应理论模型 ICC 的转换也可能存在较大误差，正如作者在文中也提到“最小二乘法在无结构和独立结构下存在较大的负偏”；

回应：谢谢专家提出的问题。

如您所说“曲线并非概率”，认知诊断模型各知识状态及其对应项目反应概率构成平面上的点，并非曲线。我们将文中不当表述“认知诊断模型的 ICC”修改为更一般的术语：认知诊断模型下项目反应函数（item response function, IRF）。

针对您的疑惑 1，说明如下：认知诊断模型的项目反应函数定义了各个知识状态  $\alpha_t$  的被试在项目上的正确作答概率  $P_j(\alpha_t) (t=1, 2, \dots, T)$ 。认知诊断模型的作答数据背后隐含了知识状态对应的能力与正确作答概率的散点图。如您所说， $P_j(\alpha_t)$  未知，文中提到要用最小二乘法 (MLS) 对其进行估计，在能力和  $P_j(\alpha_t)$  未估计出来之前，是无法绘制散点图，尽管散点图是客观存在的。

采用 2PLM 分析认知诊断模型产生或拟合的作答数据，得到 2PLM 的项目参数和 ICC。2PLM 的 ICC 是被试在项目上正确作答概率对潜在特质分数的回归曲线（漆书青，戴海崎，丁树良，2002）。由统计中回归函数的定义，能力点  $\theta_t$  处对应的项目反应函数值  $p_j(\theta_t)$  反映能力为  $\theta_t$  的被试的正确作答概率平均或条件期望。由于不同知识状态的被试的总分可能相同，对应的真实能力可能比较接近。加上能力估计误差，不同知识状态的被试的观察反应模式的能力估计值可能接近同一能力点  $\theta_t$ 。因此，能力点  $\theta_t$  处 2PLM 的项目反应函数值  $p_j(\theta_t)$  是  $k$  个知识状态的正确作答概率  $P_j(\alpha_t)$  的算术平均，其中这  $k$  个知识状态对应的能力估计值接近  $\theta_t$ 。由此，基于  $k$  近邻权方法（可参见，陈希孺，柴根象，1993）的思想， $p_j(\theta_t)$  的  $k$  近邻权估计如下：……。

对于您提到的疑惑 2：“是否可通过如下方式而非文中提到的最小二乘法 (MLS) 来确定  $P_j(\alpha_t)$  的值：首先根据知识状态  $\alpha_t$  对应的理想反应模式  $\eta_t$  估计其对应的能力值（如  $\hat{\theta}_t$ ），然后在 2PLM 的 ICC 上找到  $\hat{\theta}_t$  处的正确作答概率  $P_j(\hat{\theta}_t)$ ，直接作为  $P_j(\alpha_t)$ ”。从回答您的疑惑 1 可以看出， $\hat{\theta}_t$  处的正确作答概率  $P_j(\hat{\theta}_t)$  并非单单由对应的理想反应模式  $\eta_t$  的知识状态  $\alpha_t$  的  $P_j(\alpha_t)$  所组成。这是因为多个知识状态的被试的真实能力可能比较接近，加上能力

估计存在估计误差,不同知识状态的被试的观察反应模式的能力估计值可能接近同一能力点  $\theta_i$ , 从而正确作答概率  $P_j(\hat{\theta}_i)$  也是由相应的多个知识状态的正确作答概率  $P_j(\alpha_i)$  合成的。

将“ $\hat{\theta}_i$  处的正确作答概率  $P_j(\hat{\theta}_i)$  直接作为  $P_j(\alpha_i)$ ”, 经过模拟研究发现: (1) 根据 Bias, Abs 和 RMSE 指标, 相对而言, 在线性结构和收敛型结构下 ICC 估计精度较高, 但在所有结构上的表现均不如 MLS 方法。这也符合理论预期, 对于线性结构, 所有知识状态相互可比较, 从数学上可定义为全序关系或心理学上所述的 Guttman 量表, 此时知识状态对应能力也呈现相同的序关系(掌握属性越多, 能力越高)。(2) 基于您提到的方法估计  $P_j(\alpha_i)$ , 我们还考查了相应的条件期望法的分类效果, 也没有基于 MLS 估计的条件期望法的分类效好。同样, 在线性结构和收敛型结构分类效果较好, 对于其他结构, 其分类效果仅与 RSM 分类

效果接近, 这是因为  $\sqrt{\sum_{j=1}^m [x_{ij} - P_j(\hat{\theta}_i)]^2}$  与 RSM 的警戒指标  $\zeta$  的计算公式有部分相同之处

(已经将您提的方法列入讨论中)。对于其他结构而言, 如上一次回复您提到的问题所说: “*认知诊断模型的项目特征曲线的横坐标代表知识状态对应的能力估计值, 然而能力估计值越大对应的知识状态不一定掌握某个特定的项目的属性就越多*”。因此, 正确作答概率不一定会随着能力量尺呈现递增的趋势。若将“ $\hat{\theta}_i$  处的正确作答概率  $P_j(\hat{\theta}_i)$  直接作为  $P_j(\alpha_i)$ ”, 由于能力越高, 相应 ICC 上  $P_j(\hat{\theta}_i)$  越高, 而能力越高的知识状态对应的 IRF 的  $P_j(\alpha_i)$  并不一定具有这种递增趋势(如图 2 所示)。

如您所提到的第三个问题“从项目反应理论模型 ICC 的转换也可能会存在较大误差, 最小二乘法在无结构和独立结构下存在较大的负偏”, 这也是本文值得进一步研究的一个地方。我们已经将您提到的问题列入讨论中, 并指出造成这一结果的可能原因或可改进之处: 无结构和独立结构下属性关系松散, 单维性假设更易破坏(高慧健, 辛涛, 李峰, 2011); 项目反应理论模型的 S 型曲线(S-shaped curve)形态要求太严, 可采用非参数项目反应理论下的方法估计 ICC; 不同理想反应模式出现的概率不尽相同; 构建基于其他非参数回归方法(如核回归方法)转换方法等。

意见 3: (2)根据文章的前后逻辑, “2.2……”节最后 1 段最后 1 句话中的“……被试答错概率大于 0.5……”, 改为“……被试答对概率大于 0.5”可能更为准确;

回应: 谢谢专家提出的问题。已经按照您的建议进行了修改, 并对该段进行了修改, 感谢您如此细致的审阅。

意见 4: (3) 在“3.4 实验结果”部分中, 作者对部分结果的描述仍不够准确。比如, 在表 1 上面第 3 行, 实际上对于发散型的属性层级结构, 基于估计参数的极大似然估计方法 (E-MLE) 的表现会优于基于估计参数的条件期望法 (E-CEM)。

回应: 谢谢专家提出的问题。修改如下: 在直线型、收敛型、无结构型下, 基于估计类中心的条件期望法好于基于估计参数的极大似然估计法。而在发散型和独立结构下, 基于估计参数的极大似然估计法稍好于基于估计类中心的条件期望法。

意见 5: 修改后的稿件中仍存在多处表达不清楚或不通顺、编辑错误的地方。本文如果要达

到发表的水平，作者可能还需要在文字表达的准确性、逻辑性以及流畅性等方面多下功夫。建议作者修改完稿件后再多通读几遍，以提高文章的可读性：

回应：谢谢专家提出的问题和建议。根据您的建议对摘要、引言、第2节，结果描述和讨论部分进行仔细修改，特别是引言、第2节和结果描述部分。详细修改情况，请参见下面您提到的意见6-12及其回应和正文红色标注的部分。

意见6：(1) 中文摘要第2句话表达不通顺；

回应：谢谢专家提出的意见。考虑摘要上下文的衔接，中文摘要第2句话修改如下：基于观察反应模式与理想反应模式之间的距离判别法，以确定性的理想反应模式为类中心，而这没有考虑误差，故未充分利用总体分布信息。

意见7：(2) 引言部分修改后仍存在内容表达不清楚、逻辑不清晰的问题（特别是第4段）；

回应：谢谢专家提出的问题。对引言再次进行了重新修改。具体思路如下：第一段，简要叙述认知诊断评估的意义，并引出分类问题；第二段，叙述在认知诊断评估中两种常见的分类方法，由聚类方法过渡到判别方法；第三段，结合例子，指出以期望反应模式为类中心的判别方法（广义距离判别法和非参数分类法）的不足。第四段，给出存在上述不足的根本原因，是该类判别方法以确定性的理想反应模式为类中心，而这没有考虑误差，故未充分利用总体分布信息，引出本文介绍的方法。最后一段，为更充分地利用和获得总体分布信息以提高诊断分类效果，通过借鉴判别分析中距离判别方法，提出以项目反应模式的条件期望向量为类中心的欧氏距离判别方法，但实际情况中条件期望经常未知，针对这种情形，提出估计条件期望的方法。

意见8：(3) “1 引言”部分：第1段第2句话标点有问题；第3段第3行“期望反应模型”应改为“期望反应模式”；第2段和第3段中关于聚类分析、判别分析的描述（比如，“……不知道研究的问题应分为几类……”，“……对当前所研究的问题的类别数目……”）不是特别准确。这两种多变量统计方法只关注事先有没有将变量或个案进行分类，如果分类了分为多少类、每类的特征是什么，跟将所研究的问题分为几类没有关系；第3段倒数第3句话表达不通顺，标点使用错误（i.e., 有些逗号应改为顿号，文中其他地方也存在这个问题）；第3段倒数第2句话中，正确作答概率大于0.5（只是概率大于0.5），并不一定观察的反应就一定为1；第3段最后1句话中，什么是未完全掌握的被试？应改为“未完全掌握项目所考查属性的被试”；

回应：谢谢专家提出的问题。

“1 引言”部分第1段第1、2句话修改如下：教育中的认知诊断评估(cognitive diagnostic assessment, CDA)，旨在提供学生在特定的知识结构和加工技能(简称为属性)上的认知特点，从而有针对性地进行个性化的补救。它作为一种促进学习的评估(assessment for learning)，在指导教学方面具有非常大的潜力(戴海琦，罗照盛，2010；Leighton & Gierl, 2007；Jang, 2008)。要实现认知诊断评估服务于个性化教学，一个核心问题是如何根据观察的作答反应模式对被试进行分类，即建立什么样的准则将被试分到某个潜在知识状态类。

如您所说，关于聚类分析、判别分析的描述不是特别准确。第2段第1、2句话修改如下：认知诊断是一种特殊的模式识别(Tatsuoka, 1995, 2009)。模式识别从实现方法上可分为非监督分类和监督分类(齐敏，李大健，郝重阳, 2009)，在多元统计中相应的是聚类分析与判

别分析。

原第 3 段修改如下：广义距离判别法和非参数分类法，基于观察反应模式与理想反应模式之间的距离，对被试进行归类，分类效果不错。但是，这类方法仍存在一定的不足。一般来说，在属性之间不可补偿条件下，对于项目所考查所有属性未完全掌握的被试的理想反应为 0。而在真实测验中，这类被试的项目正确作答概率大于 0.5 的情况时有发生(de la Torre, 2011; Hartz, 2002; Jang, 2009)，意味着该类被试在此类项目上观察反应多数为 1，与之最接近的理想反应自然就是 1。在这种情况下，以期望反应模式为类中心的判别方法对这类被试(观察反应模式中得分为 1 的项目数占多数)进行分类，其知识状态，即属性掌握情况很可能被高估。

意见 9: (4) “2.1……”节：第 1 段第 5 行，建议对  $T$  所代表的含义进行简单说明；知识状态为  $\alpha_t$  ( $t=1,2,\dots,T$ ) 后面增加“的被试”，语句才通顺 (文章中有多处存在这个问题)；最后 1 段倒数第 2 句话中，海明距离方法这里就不应该是“欧氏距离相等”，而应该是“海明距离相等”；

回应：谢谢专家提出的问题。

修改如下： $T$  表示所有不同的知识状态个数，若属性相互独立，则  $T=2^K$ ，否则由属性层级结构确定(Leighton, Gierl, & Hunka, 2004)； $\alpha_t$  表示第  $t$  类知识状态，其中  $t=1,2,\dots,T$ ； $\eta_t=(\eta_{t1},\eta_{t2},\dots,\eta_{tm})$  表示知识状态为  $\alpha_t$  的被试在测验 Q 矩阵下的理想反应模式； $\mathbf{X}_t=(X_{t1},X_{t2},\dots,X_{tm})$  表示知识状态为  $\alpha_t$  的被试在  $m$  个项目上的项目反应模式。另外还将主要记号说明放到了第 2 节开始处。易举例说明，海明距离方法还会出现“ $\mathbf{x}_i$  与多个理想反应模式的海明距离相等”的情况。还修改了后面结果陈述中两处。

意见 10: (5) “2.2……”节：第 1 段第 2 行，“在某些项目上发生变异”？最好换一种表达方式；第 1 段第 5 行，项目反应模式是带有随机误差的随机向量；第 1 段最后 1 句话前后连贯性差；第 2 段第 1 句话，项目反应函数并不是项目特征函数，项目反应函数所对应的曲线才称为项目特征曲线；第 2 段第 3 行，什么是各知识状态的统计类中心？第 3 段第 1 至 2 句话，知识状态  $\alpha_t$  怎么会作答呢？最后 1 段最后 1 句话“……不仅在定理 1……”，是定理 1 还是定理 2？

回应：谢谢专家提出的问题。

2.2 节前三段修改为(部分符号说明已放到第 2 节开始处)：要利用总体分布信息构建距离判别方法以改进分类效果，首先要清楚项目反应模式是随机向量，而观察反应模式只是随机向量的观察值。由于猜测或失误，观察反应模式与理想反应模式往往有所差异。两者之间的差异被认为是随机误差，因此有必要考虑其统计分布。如 Tatsuoka(1987)在建立规则空间模型时，曾考虑过失误次数服从二项分布，通常这一点难以满足。但是二项分布要求不同项目失误概率相同。若已知项目参数，认知诊断模型的项目反应函数可用来刻画各知识状态下被试的项目反应的条件分布。

项目反应模式( $\mathbf{X}_t$ )是随机向量，它依赖于给定知识状态( $\alpha_t$ )和测验 Q 矩阵下项目反应的条件分布。距离判别方法中通常是以  $\alpha_t$  对应的  $\mathbf{X}_t$  的条件期望向量为类中心。已知项目反应的条件分布，知识状态为  $\alpha_t$  的被试的项目反应模式的条件期望向量为  $E(\mathbf{X}_t | \alpha_t) =$

$[E(X_{i_1} | \alpha_i), E(X_{i_2} | \alpha_i), \dots, E(X_{i_m} | \alpha_i)]$ , 其分量  $E(X_{ij} | \alpha_i) = P(X_{ij} = 1 | \alpha_i)$  表示知识状态  $\alpha_i$  的被试在项目  $j$  上得分的条件期望(正确作答概率)。下面给出以各知识状态下项目反应模式的条件期望向量为类中心的欧氏距离判别法...

2.2 节最后一段修改为: 在定理 1 条件下, 非参数分类法的分类结果与极大似然估计相一致。若违背定理 1 条件, 如未完全掌握项目所考查的所有属性的被试答对概率大于 0.5 时, 两种分类法的分类结果就会不一致。而条件期望法并没有非参数分类法那么强的条件, 只是进行不同知识状态作答概率之间的相对比较, 在定理 1 条件下与极大似然估计相一致; 在未掌握项目所考查的所有属性的被试答对概率大于 0.5 时条件下(违背定理 1 条件), 仍能与极大似然估计保持一致。

意见 11: (6) “2.3.....”节: 第 2 段第 1 句话表达不通顺, 且为什么说认知诊断模型的项目参数或 ICC 难以置信? 是不是因为难以估计或估计得到的误差较大? 第 3 段第 2 行中, 什么是“项目反应模型”? 应改为“项目反应理论模型”; 第 3 段第 4 行描述有问题, ICC 怎么只是正确作答概率呢? 没有理解错的话, 认知诊断模型的 ICC 应该是各知识状态对应的正确作答概率对各知识状态对应能力值的回归曲线; 第 4 段第 3 行, 列向量  $\mathbf{P}_j = (p_j(\theta_i))$  的表达不太规范; 第 5 段第 2 句话表达不够准确, 是  $k$  个知识状态对应的能力值与  $\theta_i$  接近, 而不是  $k$  个知识状态与  $\theta_i$  接近;

回应: 谢谢专家提出的问题。

第 2 段第 1 句话表达修改如下: 在诊断应用中, 可能会利用不是根据认知诊断目的开发的已有测验或题库, 具有项目反应理论模型的项目参数, 并没有认知诊断模型的项目参数。认知诊断模型的项目参数可能因为需要等值或样本量不足等原因而难以获得或难以准确估计(估计误差较大)。

第 3 段第 2 行中已经按照您的建议进行了修改。

第 3 段修改请参见您所提的意见 1 (1) 的回复。

第 4 段第 3 行, 列向量修改为  $\mathbf{Y}_j = [P_j(\alpha_1), P_j(\alpha_2), \dots, P_j(\alpha_T)]^T$ 。还修改了其他列向量的表示, 如列向量  $\mathbf{P}_j = [p_j(\theta_1), p_j(\theta_2), \dots, p_j(\theta_k)]^T$

第 5 段第 2 句话修改如下: 因此, 能力点  $\theta_i$  处 2PLM 的项目反应函数值  $p_j(\theta_i)$  是  $k$  个知识状态的正确作答概率  $P_j(\alpha_i)$  的算术平均, 其中这  $k$  个知识状态对应的能力估计值接近  $\theta_i$ 。

意见 12: (7) “3.1.....”节第 2 段第 (2) 点: 第 1 句话好像没有表达完整; 第 2 句话“某能力附近较多知识状态的答对概率较大.....”想表达的含义不太理解, 是想表达“认知诊断模型的 ICC 不一定是单调曲线”的含义吗?

回应: 谢谢专家提出的问题。第 2 段第 (2) 点的叙述的确不准确, 容易引起误解故删除。

意见 13: 鉴于稿件存在的以上几点问题, 建议修后再审!

回应: 再次对专家提出的宝贵意见表示感谢。

### 第三轮

**审稿人 1 意见：**作者较好地回答了审稿专家提出的所有问题，达到发表水平。建议发表！

**回应：**谢谢专家多轮专业而细致的审阅，非常感谢。

**编委专家意见：**这篇论文针对非参数分类法或广义距离判别法等已有方法的不足 (如以期望反应模式为类中心、未充分利用总体分布信息、在某些条件下存在限制等)，提出以条件期望向量为类中心的欧氏距离判别方法，并针对实际情况中条件期望经常未知的情形，提出估计条件期望的方法 (即基于 2PLM 项目特征曲线估计认知诊断模型的项目特征曲线)。研究问题比较有价值，整个思路新颖，整篇文章的逻辑清晰，确实对给予分类的认知诊断模型的发展有所贡献。论文作者针对两位审稿人的修改意见，做了仔细地修正，并对审稿人的问题进行了认真的回答，审稿人对作者的回答表示肯定。基于上述原因，我建议发表该论文。

**回应：**谢谢编委专家的意见和肯定，非常感谢。

**主编意见：**文中最后一段，“有评审专家指出是否可将“ $\hat{\theta}_i$  处的正确作答概率  $P_j(\hat{\theta}_i)$  直接作为  $P_j(\alpha_i)$ ”，“有评审专家”指代不明，建议按照通常处理方式，不指明具体提意见的人，如：it can be argued that...

**回应：**谢谢专家提出的问题。已经将该句修改为：本研究还尝试另外一种方法，将“ $\hat{\theta}_i$  处的正确作答概率  $P_j(\hat{\theta}_i)$  直接作为  $P_j(\alpha_i)$ ”。

最后，非常感谢各位评审专家、编委和主编的宝贵意见和建议，谢谢！