

《心理学报》审稿意见与作者回应

题目：基于作答数据的模型参数和 Q 矩阵联合估计

作者：喻晓锋 罗照盛 秦春影 高椿雷 李喻骏

第一轮

审稿人 1 意见：本文在 Liu, Xu 和 Ying (2012) 研究基础上提出了基于作答反应数据对模型参数和 Q 矩阵的联合估计算法，选题具有重要意义，方法正确，结果可信，也具有实践意义。但是本文在写作上存在一些问题，给人感觉是通篇长句太多，很多地方一句话达到六、七行的长度。另外，表述啰嗦，很多内容在文中反复出现。再有，结构不清晰、重点不够突出。

回答：感谢评审专家的鼓励。我们对全文进行了详细阅读，对每句话都进行了反复的斟酌，尽量使文章能更清楚地表达我们的意思，对文章的部分结构进行了修改，以使重点能突出。

具体意见如下：

(1) 摘要部分需要陈述研究意义，研究内容，研究方法和研究结果几个部分，作者虽然按照这些内容呈现，但是表述不够简洁，需要修改。

意见 1：

回应：我们对摘要部分进行了重新改写，力图简洁的同时，也能表达本文的研究意义、内容、方法和结果。

意见 2：引言中第一段第一句太长，存在重复啰嗦之处。

回应：我们对这句话进行了重新的改写，在能充分地表达意思的同时，尽量简洁。

意见 3：引言中第二段几种认知诊断模型的表述不符合规范，倒数五行起的这句话需要修改，是意思表达更加流畅；

回应：我们按专家意见对这句话进行了修改，使表达更加符合规范。

意见 4：引言第三段存在口语化现象，且倒数第二句应该是“但是也没有涉及到”，而不应该是“但是均没有涉及到”吧；

回应：谢谢专家细致的审稿，我们对第三段进行了修改，使表达更加书面化。倒数第二句的表达不准确，按专家的意见进行了修改。

意见 5：引言最后一段表述需要更加简洁；

回应：我们对这一段进行了修改，以使表达更加简洁。

意见 6：相关符号和概念一段显得层次和结构不清楚，需作调整；

回应：我们对这一段的层次和结构进行了调整，尽量使表达更清楚。

意见 7：关于 p 向量的定义中， $\alpha \in \{0,1\}^K$ 表示是不是不恰当？，还不如直接写 $\alpha = 1,2,\dots,2^K$

回应：谢谢专家耐心的指导，这里关于 p 向量的下标，如果写成 $\alpha = 1, 2, \dots, 2^K$ 确实与通常的表示方法更加一致，由于 Liu 等人(2012)中采用的是用属性向量作为 p 的下标，本文是在 Liu 等人的文章的基础之上进行的研究，我们在符号表示上尽量与 Liu 等人的文章中保持一致，以方便阅读，我们增加了对 p 向量下标的说明，希望不至于引起混乱。

意见 8：在介绍 DINA 模型时提到 h ，但是从 (1) 及以后均没有提到 h ，是不是可以考虑去掉 h 或换种说法；

回应：这可能是由于公式编辑器的版本原因引起的，我们本来输入的是 η ，因为 η 在 DINA 模型表示理想作答，考虑到介绍 DINA 模型的项目反应函数时会使用到这个概念，我们就呈现了 η 的介绍及其计算方式，我们对版本进行了统一，应该不会出现显示不一致的问题。

意见 9：式 (5) 中 I 的含义是 $R_i = R$ 吗？它与 (8) 式中的 I 有区别吗？如果不同，需要用不同的字母加以区分；

回应：谢谢专家的提醒，这是由于我们的疏忽所导致的，式 (5) 中的 I 表示示性函数，而式 (8) 中的 I 是表示项目或项目的组合，我们将式 (8) 中的行编号直接用数字(或数字的组合)来表示，并且做了说明，这样不至于混淆。

意见 10：第四页最后一行，提到 $N_{1\wedge 2}, N_{1\wedge 3}, N_{2\wedge 3}, N_{1\wedge 2\wedge 3}$ 相等，是指在无猜测和失误情况下相等吧，需要讲清楚成立的条件；

回应：这里的相等确实是指不考虑猜测和失误情况下，我们增加了说明。

意见 11：式 (9) 左边是乘积关系，是吗？需要加个点在中间，否则令人费解；

回应：我们已经按专家意见进行了修改。

意见 12：表的标题太长，需要简洁；

回应：我们对表的标题进行了重新改写。

意见 13：第六页首段首句存在口语化现象

回应：我们对这句话进行了重新组织。

意见 14：六页第二段“本研究采用在真实 Q 矩阵的基础上修改的方法得到初始的 Q 矩阵”，是不是表述反了？另外，本文的第 m 次迭代，为什么要进行 m 次迭代，没有很明白？本文的重点在于介绍联合估计的方法，应该更加清楚地介绍该方法的具体提步骤；这里的 m 次迭代是针对不同的项目吗？是不是与项目 j 有关的？是不是每个项目都要进行迭代？ m 的范围是什么？

回应：感谢专家细致的审稿，可能是我们的表述存在问题，没有表述清楚，我们模拟包含错误的 Q 矩阵的过程是：首先模拟产生正确的 Q 矩阵，然后对 Q 矩阵进行修改，使之包含错误，用这个包含错误的 Q 矩阵作为初始的 Q 矩阵，因此，是采用“在正确 Q 矩阵的基础上修改的方法得到初始的 Q 矩阵”。

迭代的终止条件是前后两次迭代所得到的 Q 矩阵相同，如果前后两次迭代的 Q 矩阵不相同，则程序会继续执行，直到满足终止条件。

每一次迭代会将 Q 矩阵中的所有项目都进行估计(算法按顺序对每个项目进行估计, 将所有的项目都完成估计称为一次迭代), 当 Q 矩阵中包含的错误不一样时, m 的大小也不一样, m 最小是 2, 但是没有最大值, 如果初始 Q 矩阵包含的错误太多时, 可能会导致每两次估计的 Q 矩阵都会不一样, 这样程序就不会终止, 此时 m 就会很大。我们对联合估计算法的具体步骤进行了更详细描述。

意见 15: “2.3 属性个数存在错误时 Q 矩阵估计”部分的第一句话重点不突出, 重复。

回应: 我们接受专家的意见, 将第一句话删除。

意见 16: 第七页“一共有 $3 \times 4 = 12$ ”种数据, 文中没有把这个式子写出来;

回应: 感谢专家细致的审稿, 可能是由于公式编辑器版本导致的显示问题, 我们进行了修改, 希望不至于出现显示的问题。

意见 17: 表的标题和图上面的注释不简洁;

回应: 我们对表的标题和图的注释进行了重新表述, 尽量简洁一些。

意见 18: 第十页中陈述当删除一个不该删除的属性时会带来失误参数增加猜测参数减少的过程, 我没有明白, 或者说按照你的讲述, 我没能得到你的结果。特别是在倒数第四行“但实际上, 被试总体中应该正确作答并且正确作答的人数为……”开始的推导过程是不是有问题, 请检查并表述清楚;

回应: 我们对这部分内容进行了检查和修改, 尽量使表达更清楚, 更简洁。

意见 19: 既然你是联合估计模型参数和 Q 矩阵, 为什么没有呈现模型参数的估计结果;

回应: 这是我们的疏忽。当 Q 矩阵被正确估计时, 模型参数的返真性是与 de la Torre(2009) 文章中的结果是一致的; 如果 Q 矩阵估计不正确时, 会导致模型参数的估计精度下降, 并且估计精度同 Q 矩阵估计值和真实值之间的差异有关, 这方面的研究在 Rupp & Templin (2008) 年的文章中有详细和深入的研究, 我们对此在文章中增加了相应的说明。

意见 20: 需要提供原始的 matlab 程序以供检验;

回应: 我们将原始的 matlab 程序作为附件发给心理学报编辑部, 以供检验。

意见 21: 第三部分倒数第二段最后一句“模型试验也证明了这一点”, 需要提供模拟试验的方法和结果的, 不应该随便说说。

回应: 我们按专家意见对这部分内容补充了实验结果。

意见 22: 文章的总体结构显得不够清晰, 需要修改。特别是 3.2 部分, 加些适当的小标题。

回应: 谢谢专家的提醒, 我们对文章总体进行些修改。并且对 3.2 增加了部分小标题。

审稿人 2 意见: 建议从以下几个方面完善论文:

意见 1: 文中 3.2 部分, 属性个数界定错误时的 Q 矩阵估计中较多使用推理来说明估计方法的效果, 缺少足够的实验数据。特别是介绍 Q 矩阵删除一个属性时, 只用一句话“模拟实验也证明了这一点。”带过了实验结果的交待。建议可以综合使用图表, 简洁的展示一下实验结果。

回应: 我们已按专家意见增加了表格来展示实验结果。

意见 2: 文中 3.2 部分, 用图 4 展示 Q 矩阵加入一个额外属性情况的实验结果, 但并没有指出具体的实验条件。是否 15 中增加额外属性时, 4 种样本量条件下, 都会产生相同的结果呢?

回应: 谢谢专家的提醒, 我们增加了对实验条件和实验结果的说明, 在 15 种增加额外属性时, 会产生类似的结果。

意见 3: 文 3.1 部分, 用实验表明了 EM 估计方法的效果。但是否可以考虑更多的情境, 如错误的题目数量更多 (6 个及以上), 或样本量更小? 此外, 既然当 N 大于 1000 时, 所有的估计结果均正确了, 是否还有必要展示 N 为 2000 和 4000 的情况。

回应: 我们已按专家意见增加了当错误的题目数更多时和样本量更小时的估计结果。当估计结果完全正确时, 确实没有必要用表格展示估计结果, 我们删除了相应的表格, 改为用文字说明。

意见 4: 该文对公式和符号的交待很清楚, 但行文还可以改进, 减少一些长句, 以便读者阅读。例如, 摘要中有一句话长达 329 个字, 中间包含了 18 个逗号、1 个顿号, 是否可断为多句。

回应: 感谢专家细致的审稿, 我们对文章中的表述进行了修改, 对一些长句进行分解, 使之更方便阅读。

意见 5: 文章的参考文献不够丰富, 国内也有一些学者对 Q 矩阵的修正进行过研究, 如有价值, 可以考虑引用进来。

回应: 谢谢专家的提醒, 国内学者对于 Q 矩阵修正的方法也很有价值, 我们接受专家的意见, 对相关的研究在综述时进行了补充和引用。

第二轮

审稿人 1 意见: 本文在第一次修改之后有了很大进步, 特别是对主要方法的思想及步骤的叙述, 结果的呈现与解释做了更为清晰的补充说明。文章运用方法正确, 结果可信, 具有较好的实践价值。但是, 我认为还需要最多以下几点进一步修改。

意见 1: 第一: 2.4 中 Q 矩阵和项目参数联合估计算法的介绍, 在介绍第 m 步迭代步骤的时候, 特别是 (2) ~ (4) 中没有体现项目 j 是如何变化的。另外, (1) 中的目标函数 S 和后面 (2) 中目标函数的关系是什么? 感觉没有用到 (1) 中的 S。再有, 这里的 Q (m-1) 为何不用采用下标的表示方法为 Q_{m-1} ? 我建议这部分之前可以改为: “第一次迭代从 Q_0 出发, 迭代的结果记为 Q_1 , 作为第二次迭代的初始矩阵。类似地, 第 m 次迭代的……”, 这部分内容我两次审阅都没有很清楚你的意思, 需要在仔细修改, 讲述更加清楚。多请几位同行阅读, 看他们是否能明白你的意思。

回应: 谢谢专家耐心的审稿, 可能我们一直都没表达清楚。我们对这部分内容进行了修改, 希望能清楚表达我们的意思。步骤(2)到步骤(4)中是固定其它项目, 对项目 j 在所有可能的属性向量(共有 $2^k - 1$ 种情况)下计算 S 的值, 选择 S 值最小时对应的向量作为项目 j 的属性向量。其中步骤(1)中的 S 是做为基准值的, 当(2)中的 S 更小时, 就更新项目 j 的属性向量。对于符号 $Q(m-1)$ 不表示成 Q_{m-1} , 我们主要是基于 2 个方面的考虑:

第 1 个考虑是由于文中出现了 Q_j 和 Q_{j^*} ，它们的含义与 Q_{m-1} 不相同， Q_j 和 Q_{j^*} 表示算法估计过程中矩阵的某个估计值，是为了更新项目 j ，而 $Q(m-1)$ 表示第 $m-1$ 次迭代结束后的估计值。为了更容易区分它们，所以表示成 $Q(m-1)$ ；

第 2 个考虑是与 Liu 等人(2011, 2012)中的表达保持一致。不知道我们的考虑是否妥当？

意见 2: 我认为文章在第三部分的结构上可以做一些修改。第三部分完全可以改为 2 个部分，分别针对两种情况的实验（即项目属性错误界定时 Q 矩阵的估计和属性个数错误界定时 Q 矩阵的估计），每一部分又分为方法和结果两个小部分，加小标题说明。

回应: 我们接受专家的意见，对第三部分的结构进行修改，希望文章结构更清楚。

意见 3: 文章的表述还需要进一步精心修改，使之更加简洁、流畅。

上面是我提出的几个大问题。对于文中的表述问题，我还有以下几点建议：

第一，摘要部分显得有些啰嗦。你看可否从我下面的建议的基础上修改。

“Q 矩阵在认知这暖模型参数估计和诊断分类中起着重要作用。本文通过改进 Liu 等人的方法，设计了对项目参数和 Q 矩阵进行联合估计的方法。在 DINA 模型下，党项目参数未知时开展模拟研究。研究假设项目为 20 个，考察属性分别是 3,4 和 5，初始 Q 矩阵分别存在 3,4,和 5 个。结果表明，联合估计算法能在错误的初始 Q 矩阵基础上以很高的概率估计正确的 Q 矩阵。另外，当专家认定测验的属性个数存在错误时，该方法推到的 Q 矩阵和模型参数能提供很好的鉴别 Q 矩阵错误的信息。”

回应: 根据专家的意见，我们对摘要部分进行了较大的改动，希望能更简洁地表达本文的主要研究内容和结论。

意见 4: 引言第二部分参考文献的格式是不是这样写更规范“如规则空间模型（Rule Space Model, RSM）(Tatsuoka,2009)”，后面的类似。

回应: 我们接受专家的意见，对文中相应的地方进行了修改。

意见 5: 引言最后一段太罗嗦，简要说明各部分主要内容就可以了，请修改。

回应: 我们接受专家的意见，对引言最后一段进行了修改。

意见 6: 第二部分中矩阵的转置用 T 表示，为何不同 Γ 表示，避免和后面很重要的 T 矩阵混淆；

回应: 专家的意见非常好，我们接受专家的意见。

意见 7: 公式 (1) 中应该是 $\prod_{k=1}^K \alpha_{ik}^{q_{ik}}$ ，你的表述省掉了被试 i ；此外，公式 (1) 上面的两行

可以改为”在 DINA 模型中， η 表示被试 i 在项目 j 上的理想作答，即不考虑猜测和失误的情况下，被试 i 在项目 j 上的作答，其值取 1 或 0”。注意文中的变量应用斜体表示。”把后面多余的部分删掉。

回应: 我们接受专家的意见，进行了相应的修改。

意见 8: 公式 (4) (6) 中的 Q' 是什么？是指专家界定的 Q 矩阵吗？，与公式 (10) 中 Q' 的区别是什么？

回应: 公式(4)、(6)和(10)中的 Q' 都是指一般的 Q 矩阵，即该矩阵是否是正确的 Q 矩阵是未知的。由于我们的疏忽，没有介绍清楚，我们增加了相应的说明。

意见 9: 公式 (1) 中应该是 $\prod_{k=1}^K \alpha_{ik}^{q_{jk}}$, 你的表述省掉了被试 i ; 此外, 公式 (1) 上面的两行可以改为“在 DINA 模型中, η 表示被试 i 在项目 j 上的理想作答, 即不考虑猜测和失误的情况下, 被试 i 在项目 j 上的作答, 其值取 1 或 0”。注意文中的变量应用斜体表示。”把后面多余的部分删掉。

回应: 我们接受专家的意见, 进行了相应的修改。

审稿人 2 意见: 作者很好的回复了上次审稿意见, 并补充了有关数据, 是文章的质量进一步得到提升。下面还有三点意见, 请作者考虑:

意见 1: 文章模拟实验 2 重点讨论了“属性个数存在错误时的 Q 矩阵估计”, 当属性个数多于实际个数时会出现空列, 当属性个数少于实际个数时, 部分题目的参数会发生较大的变化。这两种现象是针对联合估计法才会出现的, 还是对任何 Q 矩阵估计法都如此? 如果不是联合估计法特有的情况, 一定要注明或说清, 以免引起误解。

回应: 谢谢专家细致的审稿, 对于属性个数存在错误时的 Q 矩阵估计, 目前我们似乎并未见到有文献进行详细的报导。对此, 我们增加了相应的说明。

意见 2: 文中部分语句还显得冗长、或者不够简洁通顺, 希望作者能再仔细阅读并修改, 或请自己同事、朋友多读几次。个别地方还出现了多字或缺字的情况。

回应: 为避免再次出现表达上面的错误, 我们对全文进行了通读, 也请同行进行了阅读和评价, 希望不会出现类似的错误。

意见 3: 文中存在比较多的公式和符号, 哪些符号应为斜体、哪些不需要 (例如矩阵 Q1、Q2、Q3 是否都应为斜体加粗, 目前并不统一), 还请仔细检查一下。

回应: 感谢专家的提醒, 我们对全文进行检查, 避免再次出现符号表达上的错误。

第三轮

审稿人 1 意见: 《基于作答数据的模型参数和 Q 矩阵联合估计》稿件的第三次审稿意见:

本文在第二次修改之后有了很大进步, 整个文章表述更清楚易懂。最后, 再提几个小的修改意见:

意见 1: 摘要第二句改为“...设计了同时估计项目参数和 Q 矩阵的联合估计算法。”是否更好, 第四句末尾改为“...存分别在 3,4 和 5 个属性界定错误的项目”是否更好?

回应: 专家的表述更清楚、合理, 我们进行了相应的修改。

意见 2: 引言第二段第三行 DINA 模型的表述是否应该和其它的一致, 即采用“中文 (英文, 缩写) (文献)”的格式更好?

回应: 按照专家的意见, 我们进行了相应的调整。

意见 3: 本页倒数第三行修改为“属性向量定义不准确的问题。他们通过构建...”分一下句可能更好。

回应: 根据专家的意见, 我们进行了修改, 将长句进行了分句。

意见 4: 引言倒数第二段第二行“估计出合理的项目参数”中“合理”是不是可以改为“较准确的

项目参数”？

回应：专家的表述更准确，我们进行修正。

意见 5：引言最后一段我认为一定要突出本文的联合估计算法，因此在第一句“概念和算法”改为“概念和联合估计算法”，第二、三句都加上“该算法...”

回应：根据专家的意见，我们进行了修改，重点突出本文的研究内容。

意见 6：所有公式是不是应该右对齐，更好看？

回应：我们接受专家的意见，将公式进行了调整，尽量使排版更美观。

意见 7：第 4 也公式 (7) 上面讲“Liu 等 2012 年的文章“何不改为”Liu 等 (2012) “

回应：根据专家的意见，我们将全文类所的表述都进行了修改。

意见 8：第 5 页公式 (9) 上面一段中出现矩阵转置用 T 表示的，应修改和前文一致；

回应：由于我们的疏忽所导致的，我们进行了修改，并检查全文，希望不至于出现类似的问题。

意见 9：第 6 页，既然为何 Liu 的表述一致，我上次建议的”第一次迭代从 Q_0 出发，迭代的结果结果记为 Q_1 ，作为第二次迭代的初始矩阵。类似地，第 m 次迭代的...”中就不采用下标，

而用 $Q(0)$ ， $Q(1)$ 记；步骤 (2) 到 (4) 中在有 j 出现的地方后面加上 (j=1,2, ...,J)。另外，在第 (4) 后面可以加上“重复 (2) ~ (4)，直到更新所有 J 个项目的属性向量为止。”。第 (5) 在 $Q(m) = Q(m-1)$ 后面是不是可以解释一下“即第 m 次迭代前后两次的 Q 矩阵不变。”。最后，你解释的 (1) 中的 S 的作用也应该在稳重有所说明。作者提出的方法和思想史应该让每个读者明白，而不是让别人去猜测你的做法是什么。所以，文章在表述上应该经过深思熟虑，反复检查是否有表述不清之处。

回应：专家严谨的治学态度让我们深受感动，我们对这部分内容进行了调整，使表达上一致，表述上更清楚。

意见 10：“2.5”部分的内容，我建议直接放到第九页“4.1”之前，作为过渡语。3.1 上面的那句话中“情况”改为“条件”是不是更好？

回应：按照专家的意见，我们进行了调整和修改。

意见 11：第十页上面第二段第一句出现有多余的字。

回应：感谢专家细致的审稿，我们进行了修改，并检查全文，希望不再出现类似的错误。

意见 12：最后，虽然本文经过多次修改，表述上有了大大的提升。但是，我想要作为学术论文交流，论文在表述的简洁性、清楚性和逻辑性上，请作者再下功夫，仔细修改。谢谢！

回应：谢谢专家的鼓励，我们会认真修改论文，努力提高自己的写作水平。

审稿人 2 意见：关于审稿意见四中的问题 1，作者在文中说明没有发现以前有文献报告过该问题，这是一个方面；另一方面还请从理论推理的角度看一下，使用其他参数估计方法是否也可能出现“空列”和“参数变化较大”的现象，从而发现属性个数定义错误的情况。从理论上

看，用这两种现象发现属性个数定义错误，应该不是联合估计法所独有的特点。以上建议供作者参考。

回应：专家提出的建议非常有意义，我们对 Q 矩阵中包含一个多余的属性时，联合估计算法出现“全 0 列”现象进行了解释。由于当 Q 矩阵中包含一个多余的属性时，会导致猜测参数上升(de la Torre, 2008)，进而导致目标函数 S 偏大，只有当“多余属性”对应的列为“全 0”时，此时项目参数值会更接近真值，目标函数 S 会达到最小；当 Q 矩阵中缺少必需的属性时，会导致失误参数上升(de la Torre, 2008)，并且目标函数也会变大。因此，联合估计算法对 Q 矩阵中缺少一个必须的属性和多余一个额外属性的情况可以提供很好的参考信息。而其它 Q 矩阵的估计和验证算法是单单从参数估计的角度出发，而参数估计结果总是在真值附近波动，可能会大于真值，也可能会小于真值，因此，仅仅考虑项目参数无法识别 Q 矩阵中存在多余属性和缺少必须属性的情况，而目标函数 S 是考察了所有项目和所有被试作答基础上得到的结果，只有在项目参数越接近真值时越小，并且 Liu 等(2011)已经证明，当 Q 矩阵完全正确时，对应的目标函数 S 达到最小。