

《心理学报》审稿意见与作者回应

题目：认知诊断模型的标准误与置信区间估计：并行自助法

作者：刘彦楼

第一轮

审稿人 1 意见：

标准误是统计模型中评价模型估计结果准确性的重要方面，标准误估计的准确与否，会影响测验结果的解释。论文《认知诊断模型的标准误与置信区间估计：并行自助法》提出了两种 CDM 中标准误与置信区间估计的新方法，模拟研究表明这两种方法可以较好地估计模型参数的标准误以及对应的置信区间；实证研究也进一步展示了新方法的价值和实际效果。研究具有一定的理论和实践价值。但研究还存在以下值得进一步探讨的地方：

回应：感谢审稿专家对本研究理论和实践价值的肯定，以及在审稿过程中所付出的时间与精力。以下是对审稿意见的回应。稿件正文中的修改部分用蓝色字体标注。

意见 1：使用自助法计算 CDM 中的 SE 及 CI 的理由尚未完全讲清楚。除了解析法和自助法外，还存在其它方法(如 MCMC 估计法等)也可以计算 CDM 中的 SE 及 CI ，作者为何要选择种上自助法，而不选择其它方法？作者应说明自助法相对于其它方法所能解决的独特问题。

回应：在计算 SE 与 CI 时，解析法信息矩阵与自助法是最常用的两种方法。本研究的出发点在于：尽管在多数情况下，可以有效使用 XPD, Obs 或 Sw 方法计算 CDM 的模型参数的 SE 及 CI ，但由于解析法信息矩阵可能在某些情境中存在无法计算的 SE 及 CI 的问题，需要使用另外一些方法(如，自助法)作为替代。

相对于解析法信息矩阵而言，自助法的优点在于通用性强，不需要繁琐的公式推导；不需要严格的前提假设(如，模型参数估计值渐近正态等)；不涉及矩阵求逆，受边界值问题影响较小，尤其适合 CDM 中有非允许存在结构参数情形下 SE 及 CI 的计算；在模型参数的方差—协方差矩阵中对角线元素不会出现小于 0 的情况。这是本研究选择自助法的主要原因。

正如审稿专家提到的那样，除了解析法信息矩阵、自助法外还有其他方法可以用于计算 CDM 模型参数的 SE 与 CI ，如 MCMC 方法。但是使用 MCMC 估计 CDM 的模型参数时，单次计算耗时可能会特别长。例如，本文作者曾使用 *rstan* 软件包使用 MCMC 方法估计 ECPE 数据的模型参数，发现耗时大于 1 小时。另外，这类基于贝叶斯的方法可能对于先验分布敏

感(Jing et al., 2021)。因此, 本研究没有考虑使用 MCMC 算法计算 CDM 模型参数的 SE 及 CI 。

根据审稿专家的意见, 在稿件的讨论部分增加了如下内容(第 26-27 页):

特别指出的是, 除了解析法信息矩阵、自助法外还有其他方法可以用于计算 CDM 模型参数的 SE 与 CI , 如 MCMC (Markov chain Monte Carlo)方法。MCMC 方法不仅可以用于计算模型参数估计值, 而且可以通过计算估计过程中产生的模型参数的标准差, 作为 SE 的估计。使用 MCMC 估计 CDM 的模型参数, 计算耗时可能会特别长(例如, 大于 1 小时)。对于模型参数的 SE 及 CI 进行研究时, 需要进行大量的重复(如 500 次或以上)才能获得可靠的模拟结果(刘彦楼 等, 2016; Liu, Xin et al., 2019; Philipp et al., 2018)。另外, 这类基于贝叶斯的方法可能对于先验分布敏感(Jing et al., 2021)。因此, 本研究没有探讨使用 MCMC 算法计算 CDM 模型参数的 SE 及 CI 的表现。

意见 2: 作者说其提出了新的 pNPB 和 pPB 方法。从正文第 3 章内容来看, pNPB 和 pPB 并不是在方法上对 NPB 和 PB 进行改进, 而是在运算速度上进行改进, 这应该属于技术层面的东西, 而不是方法层面的东西。换言之, pNPB 和 pPB 并不是新方法。作者在对此二者进行表述时, 需重新表述, 不应表述成新方法。

回应: 传统的 PB 以及 NPB 的计算原理是, 将观测数据作为“总体”, 根据观测数据依次循环重复进行大量的抽样及模型参数估计, 计算相应的统计量(如, SE 等)。尽管传统的 PB 以及 NPB 具有应用广泛(例如, 近期《心理学报》上的许多论文都用到了自助法)、前提假设弱(相对于解析法信息矩阵而言)、通用性强等优点; 但由于这两种方法都是计算密集型方法, 计算量大、耗时长, 导致其理论研究及实践应用受到了极大的限制(例如, 研究者无法在一个合理的时间内系统地探讨它在计算 SE 与 CI 时的表现、以及样本量问题争议等, 见 Bai et al., 2016; Efron & Tibshirani, 1993; Guo & Wind, 2021; Hayes, 2009, 2018; Lai, 2021)。

在 CDM 研究中, 尽管研究者(Guo et al., 2017; Ma & de la Torre, 2020b)认为自助法可以用于 CDM 中的 SE 及 CI 计算, 且理论上可以较好地解决解析法信息矩阵在特定条件下无法计算的问题。但是, 由于 CDM 中待估计的模型参数多, 传统 PB 以及 NPB 的单次计算耗时长, 研究者无法在一个合理的时间内系统地探讨它在计算 SE 与 CI 时的表现, 所以就这个方法的具体使用, 研究者存在一定争议(例如, Efron & Tibshirani, 1993; Hayes, 2009, 2018; Lai, 2021)。就是说, 传统 PB 以及 NPB 的优点主要在于原理简单、计算便利, 缺点主要在于耗时长。作为解析法信息矩阵(如, XPD, Obs 或 Sw)的一种潜在替代方法, 其在计算 CDM

的模型参数 SE 与 CI 时的具体表现值得进一步探索。

系统探索 PB 以及 NPB 表现的第一步，就是要解决以上提及的耗时长的问题。因此，本文在借鉴以往研究中的并行自助法计算技术(如，Bates et al., 2015; Denwood, 2016; Zhang & Wang, 2020)，首次将并行化计算技术拓展到 CDM 的 SE 及 CI 计算中，并以此为基础系统探索了 pNPB 和 pPB 的表现，获得了一些有价值的研究结果。

在参考 Khorramdel 等人(2019)，Denwood (2016)以及 Lin 等人(2019)对方法进行命名方式的基础上，为强调新方法的计算方式而将其命名为并行自助法。

并且，正如本文在 6.2 研究展望部分的陈述一样“作者认为作为通用性强的一类方法，后续研究者可以在开发并行方法的基础上，在其他统计与测量模型中深入探讨自助法的表现，以解决先前研究没有明确的结论或结论相冲突的问题(例如，Efron & Tibshirani, 1993; Hayes, 2009, 2018; Lai, 2021)”。本文使用并行自助法这一术语，强调其与传统自助法不同，希望能够使更多研究者使用这一方法解决先前研究没有明确的结论或结论相冲突的问题。

意见 3：作者可能需要阐述清楚研究者为何(或者说在什么情况下)会用饱和 CDM 去拟合含属性层级关系的测验。如果作者知道或假设属性间存在层级关系，那么，研究者理论上是不会选择饱和模型，而是会选择可处理层级关系的 CDM。而如果研究者对属性间的关系并不清楚时(该情境常出现在研究者使用 CDM 分析大规模教育测验)，通常会先做模型比较，再基于最佳模型进行后续分析，这种做法一定程度上已经避免了非允许存在的项目参数及结构参数这一问题。

回应：赞同审稿专家关于“如果作者知道层级关系，那么，研究者理论上是不会选择饱和模型，而是会选择可处理层级关系的 CDM”的论述。本文选择使用饱和 CDM 去拟合含属性层级关系的测验数据的原因在于：

首先，CDM 研究中很难完全准确地预先判断具体的属性层级关系。预先选择恰当的可处理层级关系的 CDM 是一个难题。例如图 A1(来自，Liu et al., 2021)中呈现的四种典型层级关系。图 A2 中呈现了 Liu (2018)模拟研究中使用的属性层级关系图。并且除了以上典型层级关系外，还可能经由简单层级关系而组合成较为复杂的层级关系。也就是说，实践中的属性层级关系的可能形式有很多种，研究者很难完全准确地预先判断具体的属性层级关系。

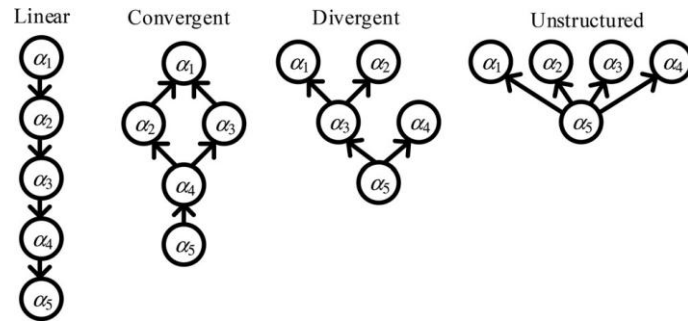


图 A1 四种典型的属性层级关系(来自, Liu et al., 2021)

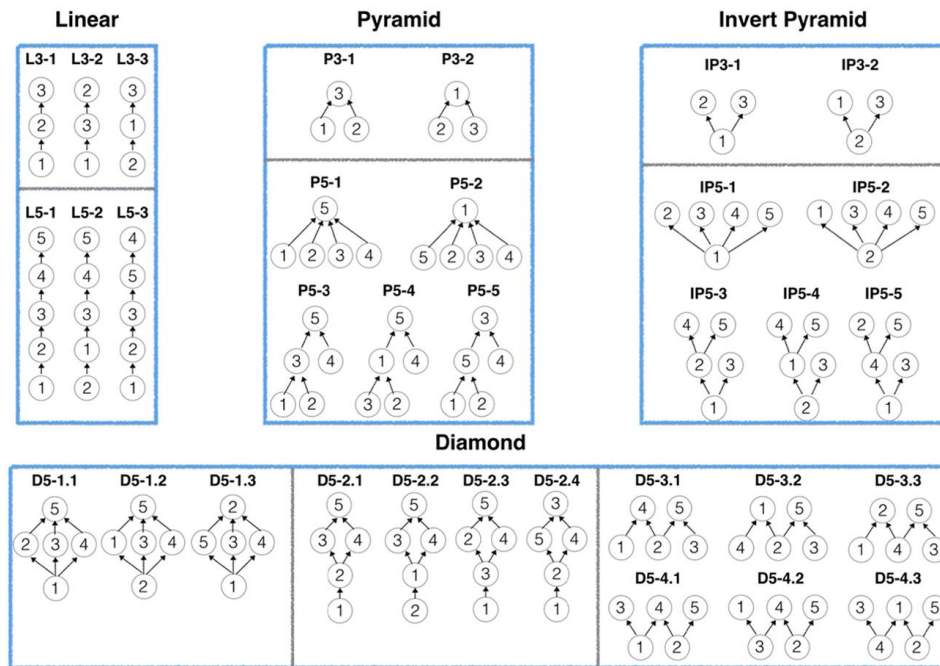


图 A2 Liu(2018)模拟研究中使用的属性层级关系图

其次, 尽管研究者(Templin & Bradshaw, 2014)提出使用似然比统计量(likelihood ratio test)进行模型比较。似然比统计量比较饱和模型与约束模型的对数似然的差, 同样需要一个预先设定了正确层级关系的 CDM。很明显, 以模型比较的方式验证属性层级关系也是一个极富挑战性的过程。

第三, 如果认知诊断测验中属性层级关系存在时, 使用饱和 CDM 拟合作答反应数据, 相应结构参数会近似等于 0, 即饱和 CDM 的结构参数能提供属性层级是否存在的证据(Liu et al., 2021; Templin & Bradshaw, 2014)。Liu 等人(2021)初步提出结构参数 SE 已知时, 可以使用 z 统计量检验饱和模型中结构参数是否显著大于 0 而探索属性层级关系。尽管在多数情况下, 可以有效使用解析法信息矩阵(如, XPD, Obs 或 Sw)计算 CDM 的模型参数 SE , 但是

使用饱和 CDM 拟合数据时可能会导致 CDM 中出现边界值问题, 使得 SE 无法计算。本文的出发点之一, 是解决当 CDM 中存在边界值时 SE 的计算问题, 因此, 本文使用饱和模型来分析数据。

根据专家的意见, 本文增加了如下内容(第 2-3 页):

准确地识别与验证 CDM 中的属性层级关系能够使研究者更深入地了解被试作答的心理过程, 具有重要的理论与实践价值(Leighton et al., 2004)。Templin 和 Bradshaw (2014)提出使用似然比统计量(likelihood ratio test)比较饱和模型与约束模型(预先设定了层级关系的模型)的对数似然的差, 即以模型比较的方式验证属性层级关系。然而, 实践中预先正确设定属性层级关系是一个非常具有挑战性的过程(Hu & Templin, 2020; Liu et al., 2021; Ma & Xu, 2021; Wang & Lu, 2021)。如果认知诊断测验中存在属性层级关系, 使用饱和 CDM 拟合作答反应数据, 相应结构参数会近似等于 0。即, 饱和 CDM 的结构参数能提供属性层级是否存在的证据(Liu et al., 2021; Templin & Bradshaw, 2014)。Liu 等人(2021)初步提出结构参数的 SE 已知时, 可以使用 z 统计量探索属性层级关系, 具体表达式为,

$$z = \frac{\hat{\eta}}{SE(\hat{\eta})} \quad (1)$$

在公式(1)中 $\hat{\eta}$ 表示结构参数估计值, $SE(\hat{\eta})$ 表示结构参数的 SE 。

意见 4: 作者在 4.1 的第二段指出“如何找到恰当的重抽样次数是需要关注的问题之一”, 如果审稿人没有理解错的话, 作者应该要解决这一问题, 不过审稿人看作者的研究设计, 作者在重抽样次数这一变量上只设定了三个水平, 那么这三个水平就能够找到恰当的重抽样次数了吗?

回应: 在对自助法文献整理过程中, 作者发现在 PB 与 NPB 的应用中重抽样的数量选取仍存在争议(例如, Bai et al., 2016; Efron & Tibshirani, 1993; Guo & Wind, 2021; Hayes, 2009, 2018; Lai, 2021)。重抽样时如果样本量过少会影响到自助法的准确性, 如果抽样过多, 其计算量会增大, 影响效率。

在 CDM 框架下本文尝试从模拟研究、实证数据分析方面来解决这个问题。

首先, 在模拟研究中发现: (1)从整体而言, 重抽样次数对于 pNPB 及 pPB 表现的影响远小于样本量、项目质量及估计方法的影响。并且重抽样次数 $B = 500$ 与 $B = 3000$ 时差异很小。也就是, 当重抽样次数 $B \geq 500$ 时各条件组合下的模拟结果开始变的稳定。(2)一些非理想情景下如项目质量低、非允许存在参数等, 重抽样次数的增加对于 pNPB 及 pPB 表现

没有明显影响。

根据审稿专家的意见，审慎起见，为进一步探讨重抽样次数的影响，本文在实验设计中新增了重抽样次数 $B = 5000$ 水平。并在图 2, 3, 6, 7, 10-17 中增加了相应的结果。结果发现，在本文的所有实验条件组合下 $B = 3000$ 与 $B = 5000$ 两者之间没有明显差异。并且在文中主要新增了如下内容：

(第 10 页) pNPB 以及 pPB 方法的重抽样次数有四个水平：200, 500, 3000 及 5000 次。

(第 13 页) 可以发现，当重抽样次数 $B \geq 500$ 时，相同条件组合下的模拟结果具有高一致性，尤其是 $B = 3000$ 与 $B = 5000$ 两者之间没有发现明显差异。

(第 19 页).....，可以发现随着重抽样次数 B 从 200 增加到 3000，项目参数的 95%CI 覆盖率略微变好，但是 $B = 3000$ 与 $B = 5000$ 两个水平下的结果高度一致。

(第 28 页).....，当重抽样次数 $B \geq 500$ 时各条件组合下的模拟结果开始变得稳定， $B = 3000$ 与 $B = 5000$ 两种重抽样次数下的结果则几乎完全相同。完全正确设定时的一些条件下的参数或模型冗余设定时允许存在参数的 95%CI 覆盖率及 BIAS 的表现随着重抽样次数 B 从 200 增加到 3000 稍有变好；

其次，本文以实证数据 ECPE 的结构参数的 SE 估计为例，计算了重复次数为 200, 500, 3000 和 10000 时的结果。发现：pNPB 在 200, 500 和 3000 下的结果与 10000 次重抽样次数下的结果相比仅有细微的差别，pPB 在 3000 次重复时结果与 10000 次重复下的结果几乎一致。也就是，实证数据分析中，即便是在 10000 次的重抽样中，也得到了与 3000 次类似的结果。

意见 5：作者指出“由于 pNPB 以及 pPB 的计算几乎不受边界值问题的影响，特别适合于 CDM 中存在属性层级关系时，模型参数 SE 及 CI 的计算。”那么，审稿人想问的是，如果事先已经知道属性间存在层级关系的话，那研究者理应会使用可处理层级关系的 CDM，或者对饱和 CDM 进行限定，让非允许存在的属性掌握模式的概率为 0，再用 Obs 等方法去估计 SE 和 CI 的话，此时也不会出现边界值问题了。因此审稿人觉得，作者可能需要增加实验条件以进一步验证 pNPB 和 pPB 是否真的比 Obs 等方法表现更优。因此，审稿人建议将解析法下的三种 SE 估计方法纳入到模拟研究中。

回应：正如作者在回答审稿专家的第 3 个问题时所提及的那样，CDM 应用实践中研究者很难完全准确地预先判断属性间的层级关系。因此，当 CDM 实践中研究者遇到属性间存在层级关系时，难以选择一个准确的 CDM 去拟合数据。一个相对可行的方法是，先使用饱和模型

去拟合数据，然后通过对结构参数进行显著性检验的方法探索可能的层级关系。相对于解析法信息矩阵 XPD, Obs, Sw, pNPB 以及 pPB 的优点是不需要有较强的前提假设以及大量的公式推导，其计算不受边界值影响。

然而，在属性层级关系存在的情形下，XPD, Obs 及 Sw 由于受到边界值的影响，难免会遇到无法求逆的问题。这也就意味着无法直接比较这两类方法的优劣。

根据审稿专家的意见，本文在模型完全正确设定条件下，将三种解析法信息矩阵 XPD, Obs 及 Sw 纳入到了模拟研究中，并系统比较了 XPD, Obs, Sw, pNPB 以及 pPB 在计算 SE 及 CI 时的表现。在文中主要新增了如下内容：

(第 9 页).....；比较 pNPB 和 pPB 方法与解析法信息矩阵 XPD, Obs 和 Sw 在估计 SE 以及 CI 时表现。

(第 9 页) 需要特别说明的是，属性间存在层级关系时 XPD, Obs 和 Sw 很容易出现无法求逆的问题(Liu et al., 2021)，因此难以在完全相同的模拟条件下比较自助法与解析法的表现。

(第 13 页) 图 4 与图 5 呈现的是呈现了 CDM 完全正确设定时，基于解析法的项目参数的 95%CI 覆盖率及 SE 的 BIAS。可以发现基于解析法 XPD, Obs 以及 Sw 矩阵的在高质量以及中等质量项目条件下计算的项目参数的 SE 有很好的表现； $N=1000$ 时，Sw 矩阵的表现略微优于 XPD 与 Obs；当样本量增加到 $N=3000$ 时，XPD, Obs 以及 Sw 矩阵的表现均在变好。对比高质量以及中等质量项目条件下 XPD, Obs 以及 Sw, pNPB 以及 pPB 的模拟结果，可以发现多数情况下 Sw 以及 Obs 矩阵的表现略微优于其他方法。低质量项目条件下，XPD, Obs 以及 Sw 矩阵计算的项目参数的 SE 的表现受到较为严重的影响； $N=1000$ 时，XPD 与 Obs 的 95%CI 覆盖率绝大部分在理论区间之下且会倾向于低估 SE，Sw 的 95%CI 覆盖率绝大部分在理论区间之上且会倾向于高估 SE； $N=3000$ 时，基于 XPD, Obs 以及 Sw 的 95%CI 覆盖率大部分在理论区间内。本研究还发现，低质量项目条件下的 BIAS 结果中，基于 XPD 及 Sw 方法的项目参数的 SE 的结果分别有 9 个及 86 个在区间[-0.2, 0.2]之外；检查发现，基于 XPD 及 Sw 方法计算的 SE 中有数值极端偏离正常值的结果(例如，SE 估计值大于 1000)。这也就是说，在低质量项目且 $N=1000$ 条件下，XPD 及 Sw 方法的表现不稳定。综合对比低质量项目条件下，XPD, Obs, Sw, pNPB 以及 pPB 的表现，可以发现 Obs 略优于其他方法。

(第 15-16 页) 图 8 与图 9 中呈现的是 CDM 完全正确设定时，基于解析法的结构参数的 95%CI 覆盖率及 SE 的 BIAS。在高和中等项目质量条件下，使用 XPD, Obs 以及 Sw 计算的结构参数的 SE 均有好的表现，几乎所有结构参数的 95%CI 覆盖率都落在图中灰线的理论

范围内或边界上, **BIAS** 几乎与 0 完全重合。低质量项目严重影响了使用 **XPD**, **Obs** 以及 **Sw** 计算的结构参数的 **SE** 的表现; $N=1000$ 时, 使用 **XPD**, **Obs** 计算的结构参数 95%**CI** 在理论范围之下且大多数 **BIAS** 小于 0, 使用 **Sw** 计算的 95%**CI** 大部分在理论范围之上且 **BIAS** 明显大于 0; $N=3000$ 时 **XPD**, **Obs** 以及 **Sw** 计算的结构参数 95%**CI** 覆盖率及 **BIAS** 的表现均在变好, 尤其是使用 **Sw** 计算的结构参数 95%**CI** 大部分在理论范围。另外, 低质量项目且 $N=1000$ 条件下, 基于 **Sw** 方法计算的结构参数 95%**CI** 覆盖率及 **BIAS** 中分别有 1 个及 3 个值在图 8 及图 9 的区间之外; 检查发现, 与先前一样, 也是由于基于 **Sw** 方法计算的 **SE** 中有数值极端偏离正常值的结果。综合对比 **XPD**, **Obs**, **Sw**, **pNPB** 以及 **pPB**, 可以发现除了低质量项目且 $N=1000$ 条件下以上方法表现均比较差之外, **Sw** 方法的表现其他条件下的表现与其他方法相当或优于其他方法。

(第 28-29 页) 本研究在 **CDM** 模型参数完全正确设定条件下对比了解析法 **XPD**, **Obs**, **Sw**, **pNPB** 及 **pPB** 的表现, 结果显示, 解析法(如, **Obs** 或 **Sw**)在一些条件下的表现要稍优于 **pNPB** 或 **pPB**。后续研究可以比较两阶段模型参数估计思路下的解析法与 **pNPB** 及 **pPB** 方法的表现。

意见 6: 第 19 页, 作者提到“如何消除非允许存在结构参数是一个非常有价值的问题”, 那么审稿人想知道, 作者是否解决了这个问题? 如果解决了, 其证据体现在哪里。

回应: 正如本文先前所述, 冗余结构参数的存在对项目参数估计值产生了影响, 进而影响到项目参数 **SE** 的表现。因此, 如何消除非允许存在结构参数是一个非常有价值的问题。先前研究(Liu et al., 2021)探讨了使用解析法计算 **SE** 进而通过 z 统计量对结构参数进行显著性检验的方法, 但是他们也指出由于解析法存在边界值及奇异矩阵问题。本研究提出的两种自助法 **pNPB** 以及 **pPB** 不存在以上两个问题, 因此使用 **pNPB** 以及 **pPB** 计算的非允许存在结构参数的 95%**CI** 覆盖率及 **BIAS** 的表现是一个重点关注的问题。

CDM 中消除非允许存在结构参数问题的实质是: 在 **CDM** 的模型参数估计中将哪些结构参数固定为 0 的问题。当已估计获得结构参数的 **SE** 时, 理论上可以通过统计量 $z = \hat{\eta} / SE(\hat{\eta})$ 检验结构参数是否显著性大于 0。通过以上 z 分数的公式可以发现, 准确的结构参数的 **SE** 估计特别重要, 是解决消除非允许存在结构参数问题的前提。因此, 本研究重点使用 **pNPB** 以及 **pPB** 计算的非允许存在结构参数的 **SE**, 以 95%**CI** 覆盖率及 **BIAS** 作为评价指标展开研究。

本文的模拟研究及实证数据分析初步显示了 pNPB 以及 pPB 在消除非允许存在结构参数方面的作用与价值。但是正如本文在 29 页研究展望中强调的那样：“本研究仅以($\alpha_1 \rightarrow \alpha_2$, $\alpha_1 \rightarrow \alpha_3$)层级关系为例，探讨了模型参数冗余设定对于 pNPB 及 pPB 表现的影响，然而不同属性层级关系条件下，模型参数的 SE 的表现，尤其是结构参数的 SE 的表现有待进一步探索。现实中不仅会存在属性层级关系，而且可能会同时存在属性之间的相关(Hu & Templin, 2020; Liu et al., 2021)，本研究中限于研究目的，没有考虑这种情景。本文认为 pNPB 及 pPB 在探索及验证属性层级关系时的表现值得进一步研究。”。

意见 7：图 12 中，为何 pNPB 和 pPB 在高质量项目和低质量项目上的 95%CI 在理论区间之外，而中等质量项目条件下，pPB 的 95%CI 却在理论区间之内？作者应该对该特殊结果进行讨论和解释。

回应：感谢审稿专家的认真与细致。本文在第 23 页中对该特殊结果增加了解释与讨论：

.....。出现这种情况的原因主要在于，在高质量项目条件下所获的 B 个 $\hat{\gamma}^*$ 的标准差 $SE(\hat{\eta})$ 大于在 R 次重复中获得的 $\hat{\gamma}$ 的标准差 $SE(\eta)$ ，即高质量项目条件下结构参数的受重抽样因素的影响而产生的变化相对更大。

.....即，中等质量项目条件下通过非参数方法获得重抽样样本 \mathbf{x}^* 而计算的 $\hat{\gamma}^*$ 的 $SE(\hat{\eta})$ 与 $SE(\eta)$ 更为接近，因此 pNPB 的表现相对较好。随着样本量的增大，除了使用 pPB 计算的第三个结构参数的 SE 外，其余均更接近理论区间。可以发现，增加重抽样次数同样没有改善这两种方法的表现。

.....。出现以上表现的原因主要是，相对于 $SE(\eta)$ 而言，非参数方法获得重抽样样本 \mathbf{x}^* 而计算的 B 个 $\hat{\gamma}^*$ 的 $SE(\hat{\eta})$ ，在多数情况下相对更大。

意见 8：作者指出，ECPE 数据中的三个属性间存在线性层级关系。这是前人的研究结论，还是作者自己提出的结论？如果是作者自己提出的结论，其理论依据是什么？从 ECPE 的 Q 矩阵来看，似乎并不支持三个属性间存在线性层级关系，否则又怎么可能出现三个属性两两组合的情况？

回应：目前至少有三个研究初步认为 ECPE 数据中的三个属性间存在线性层级关系；这不是

本研究提出的结论。第一个是 Templin 和 Bradshaw (2014)在 Templin 和 Hoffman (2013)研究基础上提出 ECPE 数据中的三个属性间存在线性层级关系，并使用 LR 统计量进行了检验，但是 Templin 和 Bradshaw (2014)发现 LR 的一类错误控制率存在问题。第二个是 Wang 和 Lu (2021)提出可以使用 latent variable selection (Xu & Shang, 2018)和 regularized latent class modeling (Chen et al., 2017)方法从数据中“学习”属性层级关系，其研究结论也认为 ECPE 数据中的三个属性间有较大可能性存在线性层级关系。第三个是 Liu 等人(2021)使用解析法信息矩阵 XPD, Obs 及 Sw 计算结构参数的 SE 之后，通过 $z = \hat{\eta}/SE(\hat{\eta})$ 统计量对结构参数进行显著性检验的方法，也初步认为 ECPE 数据中的三个属性间存在线性层级关系。

在保证模型可识别的条件下(Gu & Xu, 2020)，当三个属性之间存在线性关系时，属性之间是可以存在两两组合情况的(Liu, Huggins-Manley, & Bradshaw, 2016)。

意见 9：表 1 中，建议将各结构参数估计所对应的属性掌握模式给予呈现，以让读者更加明了地查看不同的 eta 值具体对应哪种属性掌握模式。

回应：感谢审稿专家的建议。在修改稿的第 25 页中，已经稿件正文中增加如下内容：

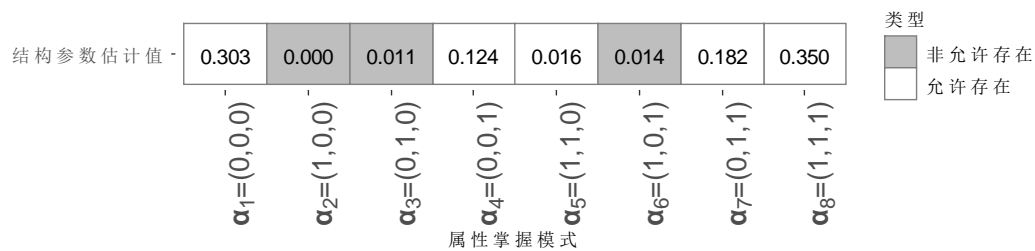


图 19 ECPE 数据集中所有可能的属性掌握模式及其对应的结构参数估计值

图 19 中呈现了使用饱和结构模型中 8 种属性掌握模式及其对应的结构参数估计值。表 1 中呈现的是使用不同方法计算的图 19 中呈现的结构参数估计值所对应的 SE。

意见 10：第 21 页第二段，作者说“使用 pNPB 及 pPB 计算第 2, 3, 6 个结构参数的 SE 的值均处于使用 XPD, Obs 以及 Sw 方法计算的 SE 的最大值与最小值之间。因此，使用这两种方法计算的 SE 同样证实了线性层级关系的存在。”审稿人不清楚这两句话之间的内在逻辑是什么，为何 pNPB 和 pPB 计算的 SE 在 XPD、Obs 和 Sw 所计算的 SE 之间时，pNPB 和 pPB 就能证实属性间存在线性层级关系？

回应：谢谢审稿专家的意见。为更清楚地说明以上问题，在稿件中增加了如下两部分内容。

第一部分是(第 3 页):

如果认知诊断测验中存在属性层级关系,使用饱和 CDM 拟合作答反应数据,相应结构参数会近似等于 0。即,饱和 CDM 的结构参数能提供属性层级是否存在的证据(Liu et al., 2021; Templin & Bradshaw, 2014)。Liu 等人(2021)初步提出结构参数的 SE 已知时,可以使用 z 统计量探索属性层级关系,具体表达式为,

$$z = \frac{\hat{\eta}}{SE(\hat{\eta})} \quad (1)$$

在公式(1)中 $\hat{\eta}$ 表示结构参数估计值, $SE(\hat{\eta})$ 表示结构参数的 SE 。

第二部分是(第 25-26 页):

当 ECPE 数据中存在线性层级关系 $\alpha_3 \rightarrow \alpha_2 \rightarrow \alpha_1$ 时,第 2, 3, 6 个结构参数(图 19 中灰色部分)应该近似等于 0 (Templin & Bradshaw, 2014),然而对于特定的结构参数而言,如 $\hat{\eta}_6=0.014$ 是否近似等于 0,需要统计检验。Liu 等人(2021)分别使用 XPD, Obs 以及 Sw 方法计算结构参数的 SE ,即 $SE(\hat{\eta})$,然后使用公式错误! 未找到引用源。中的 z 统计量检验结构参数估计值 $\hat{\eta}$ 的显著性。他们研究发现,除了 Obs 方法无法计算第 2 个参数的 SE 外,使用基于 XPD, Obs 以及 Sw 方法的 SE 计算的 z 统计量,在经过显著性水平校正后均一致地证实了存在线性层级关系的结论。需要说明的是,在 z 统计量计算公式中,结构参数估计值 $\hat{\eta}$ 在各个方法中是相同的,只有 $SE(\hat{\eta})$ 受到计算方法的影响而取值不同。本研究中,由于使用 pNPB 及 pPB 计算第 2, 3, 6 个结构参数的 SE 的值均处于使用 XPD, Obs 以及 Sw 方法计算的 SE 的最大值与最小值之间,所以使用公式(1)计算的 z 统计量的值也会位于解析法矩阵计算的 z 统计量的最大值与最小值之间。也就是,使用这两种方法计算的 SE 同样证实了线性层级关系的存在。

意见 11: 部分语句存在不通顺或口语化的情况。如“在 CDM 中非允许存在的结构参数估计值没有非常接近于 0 的情形下,结构参数估计值是偏离 0 这个真值的,即估计值是有偏的(p.7)”偏口语化;而“研究者提出使用模型参数估计值以及观察到的全部被试的作答反应矩阵 x 计算的信息矩阵(p.6)”;再如“第 2, 3, 6 个结构参数从统计上而言应该不显著大于 0 (p.21)”则过于拗口,根据审稿人的理解,这应该是想表达这三个结构参数值等于 0,即 $\eta_2=\eta_3=\eta_6=0$ 。

回应：多谢审稿专家指正。在稿件正文中已修改相关语句：

(第 8 页) 其次，如果非允许存在的结构参数的估计值偏离其真值 0，那么这个估计值是有偏的，不再符合公式(5)中的前提假设，因此对 XPD, Obs 以及 Sw 矩阵的计算会造成不良影响。

(第 6-7 页) 针对 I_e 的不足，研究者提出使用模型参数估计值 $\hat{\gamma}$ 替代真值 γ ，使用被试的观察作答反应矩阵 \mathbf{x} 代替单个被试的作答反应向量的期望，进而开发出 XPD, Obs 以及 Sw 矩阵(刘彦楼 等, 2016; Liu, Xin et al., 2019; Philipp et al., 2018)。

(第 26 页) 第 2, 3, 6 个结构参数(图 19 中灰色部分)应该近似等于 0 (Templin & Bradshaw, 2014)，然而对于特定的结构参数而言，如 $\hat{\eta}_6=0.014$ 是否近似等于 0，需要统计检验。

另外，根据审稿专家的意见，对文中其他可能出现语句存在不通顺或口语化的问题的部分也进行了修改。再次感谢审稿专家的耐心与细致。

【正文中未出现的参考文献】

Lin, Y.-S., Heathcote, A., & Holmes, W. R. (2019). Parallel probability density approximation. *Behavior Research Methods*, 51(6), 2777-2799. doi:10.3758/s13428-018-1153-1

Liu, R., Huggins-Manley, A. C., & Bradshaw, L. (2016). The Impact of Q-Matrix Designs on Diagnostic Classification Accuracy in the Presence of Attribute Hierarchies. *Educational and Psychological Measurement*, 77(2), 220-240. doi:10.1177/0013164416645636

.....

审稿人 2 意见：

本文基于已有研究，将 pNPB 以及 pPB 应用到 CDM 模型参数估计的 SE 和 CI 的计算，并充分、系统地探讨使用 pNPB 以及 pPB 在计算项目参数和结构参数的 SE 及 CI 时的表现。选题具有实践价值和研究意义。文稿行文流畅，逻辑清楚。该论文使用模拟研究系统比较了新方法在不同数据生成模型、样本量、重抽样次数、项目质量下的具体表现，使用实证数据 ECPE 展示了研究的理论与实践价值。文章选题新颖，提出了新方法、获得了新发现。以下几个问题审稿人认为需要作者进一步修改：回应：感谢审稿专家对本文选题、写作、实践价值及研究意义的肯定，以及在审稿过程中所付出的时间与精力。以下是对审稿专家意见与建议的点对点回应。稿件正文中的修改部分用蓝色字体标注。

意见 1：摘要第一句显得太空洞，不够具体。

回应：感谢审稿专家的意见。已将摘要第一句修改为(第 1 页)：

认知诊断模型的标准误(*Standard Error, SE*; 或方差—协方差矩阵)与置信区间(*Confidence Interval, CI*)在模型参数估计不确定性的度量、项目功能差异检验、项目水平上的模型比较、**Q** 矩阵检验以及探索属性层级关系等领域有重要的理论与实践价值。

并且，修改了对应的英文摘要(第 33 页)：

The model parameter standard error (*SE*; or variance–covariance matrix) that provides an estimate of the uncertainty associated with the model parameter estimate has both theoretical and practical implications in cognitive diagnostic models (CDMs), such as the computation of approximate confidence intervals (*CI*s), differential item functioning detection, item-level model comparison, **Q**-matrix estimation (validation), or exploring attribute hierarchy structure.

意见 2：引言第二段第 2-4 句话，表达上有些口语化的地方，建议修改。

回应：感谢指正，已修改。并检查及修改了文中其他可能存在口语化表达的段落。

(第 1 页)在心理统计与测量模型中，点估计值相同的两个模型参数可能由于 *SE* 不同而具有不同的置信区间(*Confidence Interval, CI*)，因此需要综合考虑模型参数的点估计值与 *CI*。例如，CDM 中两个项目的猜测参数估计值均为 0.2，但 *SE* 的估计值分别为 0.08 与 0.05，那么这两个猜测参数的估计精度不同。根据正态分布理论，第一个猜测参数的 95%*CI* 是 $[0.2-1.96*0.08, 0.2+1.96*0.08]$ ，第二个猜测参数的 95%*CI* 是 $[0.2-1.96*0.05, 0.2+1.96*0.05]$ 。

意见 3：引言第三段，最后一句“鉴于模型参数 *SE* 的重要性，研究者提出了多种基于解析法的估计方法...”建议改为“对于模型参数的 *SE* 的估计，研究者提出了多种基于解析法的估计方法...”

回应：谢谢，已修改(第 2 页)。

对于模型参数的 *SE* 的估计，研究者提出了多种基于解析法的估计方法(刘彦楼 等, 2016; Liu, Xin et al., 2019; Liu et al., 2021; Philipp et al., 2018), ……

意见 4：文章多次提到“项目参数 *SE*”或“结构参数 *SE*”，实际上是指参数的估计标准误，建议不省略“的”字，即保留“项目参数的 *SE*”。

回应：感谢审稿专家的建议，已经统一修改全文。

意见 5：引言第 4 段第 2 句，模型正确设定，模型错误设定，指的是模型参数的设定情况，还是模型本身的设定情况？根据前文，应该是参数的设定情况，请对文稿进行修正。

回应：以往研究中的模型的错误设定包括两个方面：一是项目反应模型，二是 Q 矩阵中的一些元素存在误设，这些都会导致模型中的参数出现错误设定的问题。根据审稿专家的意见，本文对第 4 段中的表达进行了修改：

(第 2 页)关于项目参数的 SE 及 CI ，研究者比较了在理想状况下(即模型与观察数据完美拟合)、在 CDM 的项目反应模型和/或 Q 矩阵错误设定条件下，XPD, Obs 或 Sw 方法的表现(Liu, Xin, et al., 2019; Philipp et al., 2018)。研究发现，当模型(包括项目反应模型与 Q 矩阵)完全正确设定时或存在较少错误设定时，这三种方法在项目参数的 SE 估计的一致性方面都有好的表现，且在这三种方法中，Obs 的表现最好；在模型存在严重错误设定时(如，项目反应模型与 Q 矩阵同时包括较多的错误)，XPD 和 Obs 无法提供一致性的项目参数的 SE 估计，只有 Sw 具有健壮性(Liu, Xin, et al., 2019)。

意见 6：请检查和说明公式 5~7 是否符合一般表达？一阶导和二阶偏导，如何用向量和向量的转置来表达？

回应：观察数据的对数似然函数关于模型参数的一阶导是一个列向量；将这个列向量与其转置(行向量)交叉相乘可获得 XPD 矩阵。观察数据的对数似然函数关于模型参数的二阶偏导本身是一个矩阵。初稿中的公式表达源自于以往关于信息矩阵的研究(Liu, Xin et al., 2019; Philipp et al., 2018)。为使表达更加清晰，对稿件进行了修改：

(第 7 页)使用观察数据对数似然函数 $l(\hat{\gamma} | \mathbf{x})$ 关于模型参数 $\gamma = (\lambda', \eta')'$ 的一阶导向量交叉相乘而计算的 XPD 矩阵的表达式为，

$$\mathbf{I}_{XPD} = \begin{bmatrix} \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \lambda_1} & \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \lambda_1} & \mathbf{K} & \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \lambda_1} & \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \eta_{L-1}} \\ & \mathbf{M} & \mathbf{O} & & \mathbf{M} \\ \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \eta_{L-1}} & \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \lambda_1} & \mathbf{L} & \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \eta_{L-1}} & \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \eta_{L-1}} \end{bmatrix} \quad (6)$$

根据观察数据对数似然函数关于模型参数的二阶偏导而计算的 Obs 矩阵可表示为(刘彦楼等, 2016; Liu, Xin et al., 2019),

$$\mathbf{I}_{Obs} = - \begin{bmatrix} \frac{\partial^2 l(\hat{\gamma} | \mathbf{x})}{\partial \lambda_1 \partial \lambda_1} & \mathbf{K} & \frac{\partial^2 l(\hat{\gamma} | \mathbf{x})}{\partial \lambda_1 \partial \eta_{L-1}} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \frac{\partial^2 l(\hat{\gamma} | \mathbf{x})}{\partial \eta_{L-1} \partial \lambda_1} & \mathbf{L} & \frac{\partial^2 l(\hat{\gamma} | \mathbf{x})}{\partial \eta_{L-1} \partial \eta_{L-1}} \end{bmatrix} \quad (7)$$

需要特别说明的是，Obs 矩阵中的元素也可以等价地表达为(Liu & Maydeu-Olivares, 2014; Liu, Xin et al., 2019),

$$\frac{\partial^2 l(\hat{\gamma} | \mathbf{x})}{\partial \gamma_1 \partial \gamma_2} = \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \gamma_1} \frac{\partial l(\hat{\gamma} | \mathbf{x})}{\partial \gamma_2} - N \sum_{v=1}^{v_0} \frac{p_v}{f(\mathbf{x}_v)} \frac{\partial^2 f(\mathbf{x}_v)}{\partial \gamma_1 \partial \gamma_2} \quad (8)$$

在公式(8)中， γ_1 与 γ_2 分别表示任意一个项目参数(λ)或结构参数(η);

意见 7: 文稿指出“针对解析法信息矩阵及传统自助法估计模型参数 *SE* 与 *CI* 时存在的不足，提出了两种新方法：并行参数化自助法(parallel Parametric Bootstrap, pPB)和并行非参数化自助法(parallel Nonparametric Bootstrap, pNPB)。”，似乎 pPB 和 pNPB 并不是直接解决解析法的不足？两者的思路完全不一样，似乎不应该表达为针对？？？的不足，提出新的方法？，因为新方法并不是直接解决这个不足的。个人理解，仅供参考。

回应：感谢审稿专家的细致与严谨，已对文中表达进行了修改(第 1 页)：

由于解析法信息矩阵及传统自助法在估计模型参数的 *SE* 与 *CI* 时存在不足，本研究提出了两种新方法：并行参数化自助法(parallel Parametric Bootstrap, pPB)和并行非参数化自助法(parallel Nonparametric Bootstrap, pNPB)。

意见 8: 第三部分，建议将 pPB 和 pNPB 的内容分成两个小标题。

回应：谢谢，已在第三部分增加小标题(第 8 页)：

3.1 并行的非参数化自助法

3.2 并行的参数化自助法

意见 9: 第 3 部分最后一段“在模型参数的方差—协方差矩阵中对角线元素不会出现小于 0 的情况。”，这句话如何解释？为什么新方法下不会出现这种情况？

回应：在 pPB 和 pNPB 方法中之所以不会出现这个问题，是因为方差—协方差矩阵是使用 *B* 个模型参数估计值向量计算的，对角线上的元素都是模型参数对应的方差，所以不会出现小

于 0 的情况。在文中已增加相应解释(第 9 页):

模型参数的方差—协方差矩阵仅需 B 个 $\hat{\gamma}^*$ 向量即可计算, 对角线元素不会出现小于 0 的情况。

意见 10: 研究提出解析法的不足, 新方法也有对应的一些缺点。那么, 模拟研究为什么不尝试比较解析法和新方法的表现呢?

回应: 谢谢。根据您的及审稿专家 1 的意见, 本文在模型完全正确设定条件下比较了解析法和 pNPB、pPB 的表现对比。另外, 由于当属性关系存在时, XPD, Obs 和 Sw 很容易出现无法求逆的问题, 导致无法计算 SE 及 CI, 因此在层级关系存在条件下无法比较解析法和新方法的表现, 对此, 本文也进行了说明。在修改稿中新增了如下内容:

(第 9 页).....; 比较 pNPB 和 pPB 方法与解析法信息矩阵 XPD, Obs 和 Sw 在估计 SE 以及 CI 时表现。

(第 9 页)需要特别说明的是, 属性间存在层级关系时 XPD, Obs 和 Sw 很容易出现无法求逆的问题(Liu et al., 2021), 因此难以在完全相同的模拟条件下比较自助法与解析法的表现。

(第 13 页) 图 4 与图 5 呈现的是呈现了 CDM 完全正确设定时, 基于解析法的项目参数的 95%CI 覆盖率及 SE 的 BIAS。可以发现基于解析法 XPD, Obs 以及 Sw 矩阵的在高质量以及中等质量项目条件下计算的项目参数的 SE 有很好的表现; $N=1000$ 时, Sw 矩阵的表现略微优于 XPD 与 Obs; 当样本量增加到 $N=3000$ 时, XPD, Obs 以及 Sw 矩阵的表现均在变好。对比高质量以及中等质量项目条件下 XPD, Obs 以及 Sw, pNPB 以及 pPB 的模拟结果, 可以发现多数情况下 Sw 以及 Obs 矩阵的表现略微优于其他方法。低质量项目条件下, XPD, Obs 以及 Sw 矩阵计算的项目参数的 SE 的表现受到较为严重的影响; $N=1000$ 时, XPD 与 Obs 的 95%CI 覆盖率绝大部分在理论区间之下且会倾向于低估 SE, Sw 的 95%CI 覆盖率绝大部分在理论区间之上且会倾向于高估 SE; $N=3000$ 时, 基于 XPD, Obs 以及 Sw 的 95%CI 覆盖率大部分在理论区间内。本研究还发现, 低质量项目条件下的 BIAS 结果中, 基于 XPD 及 Sw 方法的项目参数的 SE 的结果分别有 9 个及 86 个在区间[-0.2, 0.2]之外; 检查发现, 基于 XPD 及 Sw 方法计算的 SE 中有数值极端偏离正常值的结果(例如, SE 估计值大于 1000)。这也就是说, 在低质量项目且 $N=1000$ 条件下, XPD 及 Sw 方法的表现不稳定。综合对比低质量项目条件下, XPD, Obs, Sw, pNPB 以及 pPB 的表现, 可以发现 Obs 略优于其他方法。

(第 15-16 页) 图 8 与图 9 中呈现的是 CDM 完全正确设定时, 基于解析法的结构参数的 95%CI 覆盖率及 SE 的 BIAS。在高和中等项目质量条件下, 使用 XPD, Obs 以及 Sw 计算的结构参数的 SE 均有好的表现, 几乎所有结构参数的 95%CI 覆盖率都落在图中灰线的理论范围内或边界上, BIAS 几乎与 0 完全重合。低质量项目严重影响了使用 XPD, Obs 以及 Sw 计算的结构参数的 SE 的表现; $N=1000$ 时, 使用 XPD, Obs 计算的结构参数 95%CI 在理论范围之下且大多数 BIAS 小于 0, 使用 Sw 计算的 95%CI 大部分在理论范围之上且 BIAS 明显大于 0; $N=3000$ 时 XPD, Obs 以及 Sw 计算的结构参数 95%CI 覆盖率及 BIAS 的表现均在变好, 尤其是使用 Sw 计算的结构参数 95%CI 大部分在理论范围。另外, 低质量项目且 $N=1000$ 条件下, 基于 Sw 方法计算的结构参数 95%CI 覆盖率及 BIAS 中分别有 1 个及 3 个值在图 8 及图 9 的区间之外; 检查发现, 与先前一样, 也是由于基于 Sw 方法计算的 SE 中有数值极端偏离正常值的结果。综合对比 XPD, Obs, Sw, pNPB 以及 pPB, 可以发现除了低质量项目且 $N=1000$ 条件下以上方法表现均比较差之外, Sw 方法的表现其他条件下的表现与其他方法相当或优于其他方法。

(第 28-29 页) 本研究在 CDM 模型参数完全正确设定条件下对比了解析法 XPD, Obs, Sw, pNPB 及 pPB 的表现, 结果显示, 解析法(如, Obs 或 Sw)在一些条件下的表现要稍优于 pNPB 或 pPB。后续研究可以比较两阶段模型参数估计思路下的解析法与 pNPB 及 pPB 方法的表现。

意见 11: 文稿研究的项目参数的置信区间, 结构参数的 SE。为何这样做, 项目参数有 SE 吗? 结构参数的置信区间呢?

回应: 抱歉给审稿专家造成误解认为本文仅研究了项目的 CI, 结构参数的 SE。本文中系统研究了使用 pNPB 以及 pPB 估计项目参数的 SE 与 CI, 估计结构参数的 SE 与 CI 时的表现。为避免误解, 进行了如下修改:

(第 1 页) 在心理统计与测量模型中, 点估计值相同的两个模型参数可能由于 SE 不同而具有不同的置信区间(Confidence Interval, CI), 因此需要综合考虑模型参数的点估计值与 CI。例如, CDM 中两个项目的猜测参数估计值均为 0.2, 但 SE 的估计值分别为 0.08 与 0.05, 那么这两个猜测参数的估计精度不同。根据正态分布理论, 第一个猜测参数的 95%CI 是 $[0.2-1.96*0.08, 0.2+1.96*0.08]$, 第二个猜测参数的 95%CI 是 $[0.2-1.96*0.05, 0.2+1.96*0.05]$ 。

(第 2 页) 研究者通过数据模拟以及实证数据分析的方式探索了使用解析法信息矩阵(Liu et al., 2016; 刘彦楼 等, 2016)计算的模型参数(包括项目参数与用于描述被试分布的结构参数)

的 SE 及 CI 的表现。

(第 7 页) 在公式(8)中, γ_1 与 γ_2 分别表示任意一个项目参数(λ)或结构参数(η);

意见 12: 文中提及“然而, 在国内外的 CDM 实证研究中, 报告模型参数 SE 及 CI 的研究仍然较少。”，审稿人认为确实存在这个问题, 但造成这个问题的原因是什么, 请作者进一步解释说明。

回应: 感谢审稿专家的意见。对于这个问题, 在文中增加了如下解释及说明:

(第 2 页) 造成这种现象的原因是多方面的, 主要原因在于缺乏易用的计算方法。接下来, 本文将对两类常用的 SE 及 CI 的估计方法: 解析法以及自助法目前存在的问题展开探讨, 并提出一类简易、可行的方法。

(第 3 页) 解析法信息矩阵所存在的以上问题, 限制了其理论发展及实践应用。

(第 4 页) 作为一种计算密集型方法, 计算量大、耗时长的问题, 不仅限制了 PB 与 NPB 的理论研究, 而且也造成了实践应用的困难。

(第4页)……, pPB与pNPB是一类简易、可行的方法, 不仅能有效解决CDM中的 SE 与 CI 理论研究中的重要问题, 而且能有效提升实践应用中的计算效率。

(第29页) 同NPB及PB相比, pNPB及pPB极大地提升了计算效率, 是简易、可行的 SE 及 CI 计算方法。

意见 13: 文中提出的新方法除了可以用于计算 SE 及 CI 外, 我认为还有很多潜在的理论及实践价值, 建议作者进一步探讨。

回应: 感谢审稿专家的建议。在“6.2 研究展望”部分(第 29 页)增添了如下内容:

(5)需要特别指出的是, pNPB 及 pPB 除可以用于计算 SE 及 CI 外, 还有很多潜在的理论及实践价值。研究者可以进一步探索 pNPB 及 pPB 在项目功能差异检验、项目水平上的模型比较、Q 矩阵检验等领域中的表现。

意见 14: 模拟研究结果中对 CDM 的模型参数冗余条件下的实验条件下的实验结果的解释中提到, 极端偏离理论区间的项目参数是由于错误设定某些“非允许存在”的属性掌握模式为“存在“存在”, 为什么这样解释。

回应: 谢谢审稿专家的意见。为更清晰地说明这个问题, 在稿件中新增了如下解释:

(第 18-19 页)例如,对比公式(3)和(4),可以发现如果“真”模型是带有线性层级关系的 HCDM,但使用饱和 CDM 估计模型参数时,由于“非允许存在”的属性掌握模式 α_3 被错误地设定为“存在”,造成饱和 CDM 中结构参数 η_3 以及项目参数 $\lambda_{j,1(2)}$ 真值都等于 0。

意见 15: 写作的一些细节还是要注意,比如,(1)在 P9, 4.3 评价指标中,“使用偏差(BIAS)以及 95%CI 覆盖率评价使用 pNPB 以及 pPB 计算的项目参数以及结构 SE...”的表述中漏掉了“参数”二字,应该是“使用偏差(BIAS)以及 95%CI 覆盖率评价使用 pNPB 以及 pPB 计算的项目参数以及结构参数 SE...”; (2)在第五部分实证数据分析,其中 5.2 研究结果部分“计算机密集”应该是“计算密集”; (3)在第一部分引言,其中非参数化自助法(parallel Nonparametric Bootstrap, pNPB)应该是“并行非参数化自助法”。

回应: 感谢审稿专家指正,已在稿件中修改相关语句,并且又多次请同实验室研究者进行挑剔性阅读。

(第 11 页) 使用偏差(BIAS)以及 95%CI 覆盖率评价使用 pNPB 以及 pPB 计算的项目参数以及结构参数的 SE 的表现。

(第 26 页) 作为一类计算密集的方法,传统自助法计算速度慢,每次均需消耗大量时间。

(第 4 页) 并行非参数化自助法(parallel Nonparametric Bootstrap, pNPB),提高 CDM 中 PB 与 NPB 的计算效率。

再次感谢两位审稿专家所付出的时间与精力。

第二轮

审稿人 1 意见:

经过第一轮修改,论文《认知诊断模型的标准误与置信区间估计:并行自助法》的质量得到比较大的提升,作者也比较好地回答了审稿人提出的问题。但还存在以下几个小问题供作者参考:

回应: 感谢审稿专家的意见与建议。以下是对审稿意见的点对点回应,稿件正文中相应的修改部分用橙色字体标注。

意见 1: 结论或摘要部分可否增加一些带有指导性质的句子,如在使用并行自助抽样法时,重抽样次数大体在多少次合适?两种并行方法各自在什么条件下使用等。

回应：感谢审稿专家的建议。已经在结论及英文摘要部分增加了关于使用 pNPB 及 pPB 方法计算 CDM 模型参数的 *SE* 及 *CI* 时关于重抽样次数的建议：

(4)综合模拟研究与实证数据分析结果，本研究初步认为：在 pNPB 及 pPB 方法中为快速预览 *SE* 估计结果可以选择 200 次重抽样；为获得较为准确的估计结果，审慎起见可以选择 3000 或以上的重抽样次数。(第 28 页)

The simulation and real data studies showed that 3,000 resamples might be adequate for the bootstrap method in calculating model parameter *SEs* and *CI*s in CDMs. (第 33 页)

意见 2：引言第二段，作者指出“国内外多种心理学期刊(如，心理学报或 American Psychological Association, 2020 等)要求或建议报告 *SE* 及 95%*CI*。”审稿人想问的是 APA 可以算是心理学期刊吗？

回应：感谢审稿专家的意见。APA 不是心理学期刊，但是目前国内外绝大多数心理学期刊使用或参考 APA 中的指导或规范。本文已将相关句子修改为：

正因如此，国内外多种心理学期刊(如心理学报，或参考 American Psychological Association, 2020)要求或建议报告 *SE* 及 95%*CI*。(第 1 页)

意见 3：第 3 页第二段，是“...可以有效使用 XPD, Obs 或 Sw 方法计算 CDM 的模型参数的 *SE*”，还是“...可以使用 XPD, Obs 或 Sw 方法有效地计算 CDM 中模型参数的 *SE*”？

回应：感谢审稿专家的意见，后面的表达更恰当。已将相关句子修改为：

尽管在多数情况下，可以使用 XPD, Obs 或 Sw 方法有效地计算 CDM 模型参数的 *SE*，……。 (第 3 页)

意见 4：章节 2.2 第一段第一行，“如 G-DINA 模型”是对饱和模型的解释说明，因而应写成：对饱和模型（如 G-DINA 模型）的结构参数...

回应：感谢审稿专家的意见，已将相关句子修改为：

……，对饱和模型(如 G-DINA)的结构参数以及项目参数加以适当约束，……。 (第 5 页)

意见 5：作者将“真值为 0 的参数统称为非允许存在的参数”。非允许存在的参数比较拗口，将其改成“不允许存在的参数”是否更加合适？因为“允许存在”的反义词通常是“不允许存在”，所以审稿提出该建议，该建议仅供作者参考。

回应: CDM 中, 根据属性层级关系, 在理论上不等于 0 的模型参数, 被称为 permissible model parameter, 理论上等于 0 的模型参数被称为 impermissible model parameter。可将 impermissible model parameter 翻译为不允许存在的模型参数, 或者是非允许存在的模型参数, 意思都对。本文认为翻译成非允许存在的参数, 会更加书面语一些, 因此采用了“非允许存在”的翻译。为便于理解, 稿件中首次出现这个名词的位置进行了如下修改:

……, 对于允许存在的结构参数(permissible structural parameter), 即根据属性层级关系在理论上不等于 0 的结构参数, XPD 和 Obs 方法计算的 SE 有较好的表现; 对于非允许存在的结构参数(impermissible structural parameter), ……。(第 2 页)

意见 6: 正文第 9 页的步骤 (1) “除了执行 pNPB 中相应的步骤外”中“相应步骤”的指代并不是那么明确, 建议指出 pNPB 中的具体步骤, 如步骤 (1)。

回应: 感谢审稿专家的建议。已对稿件进行了修改:

步骤(1), 除执行 pNPB 中的步骤(1)外, 还需根据原始作答数据 \mathbf{x} 及预先指定的 CDM 估计模型的项目参数及结构参数 $\hat{\gamma} = (\hat{\lambda}, \hat{\eta})'$ 。(第 8 页)

审稿人 2 意见:

作者很好地针对评审专家的意见, 进行了深入思考和详细阐释和解答, 并对文稿进行补充和修改。另外, 还有一些小问题, 还请作者进行思考, 并对适当的地方进行修正。例如, 摘要最后一句中“大部分允许存在的模型参数的 SE 和 CI 有好的表现。”, “对于大部分允许存在的模型参数而言, 其 SE 和 CI 有好的表现。”更易于理解。又如, 对于 bias 的图 7、15、17 等是不是可以考虑缩短纵轴的范围。最后, 建议作者对文稿进行必要的精简。

总体上讲, 文稿内容丰富, 逻辑清晰, 方法正确、实验条件丰富, 写作规范, 公式、图表清楚无误。建议小修后发表。

回应: 感谢审稿专家对本文内容、方法及写作的肯定。以下是对审稿意见的回应, 稿件正文中的修改部分用橙色字体标注:

(1) 按照审稿专家的建议, 已将摘要中的相关句子修改为“……, 对于大部分允许存在的模型参数而言, 其 SE 和 CI 有好的表现。”(第 1 页)

(2) 缩短纵轴的范围, 能够更细微地呈现细节上的差异。稿件中将坐标设置为一致, 没有缩短图 7、15、17 纵轴的范围的原因是: (a) 图 7、15、17 在[-0.2, 0.2]区间已经能够非

常好地呈现 **BIAS** 结果以及不同实验条件组合下的结果差异；(b) 为了便于完整、直观地与其他图中的 **BIAS** 结果对比。

(3) 已对稿件的摘要及正文进行了删减。另外，又请同实验室的研究者进行了挑剔性阅读，对于一些可能影响理解的句子进行了修改。

再次感谢两位审稿专家对稿件提出的宝贵意见与建议。

编委意见：该论文的选题有意义。经过修改，达到发表的要求。建议发表。

主编意见：同意外审和编委意见，建议录用。