

# 拓展后的 GGUM 在 LOT-R 中的适用性分析\*

邓稳根<sup>1</sup>, 戴海琦<sup>2</sup>, 刘铁川<sup>2</sup>, 戴慧群<sup>1</sup>

(1. 赣南师范学院 教育科学学院, 赣州 341000; 2. 江西师范大学 心理学院, 南昌 330022)

**摘 要:** 该研究对拓广等级展开模型(GGUM)进行了拓展, 取消 GGUM 中关于主观反应类别阈限对称的假设, 并将拓展之后的新模型和 GGUM 同时用于生活取向测验修订版(LOT-R)的被试反应数据分析, 采用新编的单项目、两项目对和三项目组  $\chi^2/df$  计算程序计算和比较新模型和 GGUM 在该测验数据上的拟合差异。结果显示, 新编程序与 Stark 等人开发的 MODFIT 程序具有同样的有效性, 新模型在这些指标上的值显著小于 GGUM, 并且均小于 3, 表明新模型较 GGUM 更适合于分析 LOT-R 的反应数据, 说明新模型更适用于分析具有多个评定等级的人格测验数据。根据以上结果, 该研究认为, 未来人格测验的数据分析应该使用没有对主观反应类别阈限进行对称限定的新拓展的模型更合理。

**关键词:** 拓广等级展开模型; 主观反应类别阈限; 对称; 模型—数据拟合

**中图分类号:** B841.2

**文献标识码:** A

**文章编号:** 1003-5184(2011)01-0070-04

## 1 引言

Roberts, Donoghue 和 Laughlin 提出了拓广等

级展开模型(GGUM), 他们根据客观反应类别的概率定义 GGUM 为公式 1<sup>[1]</sup>:

$$P[Z_i = z | \theta_j] = \frac{\exp \{ \alpha_i [z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}] \} + \exp \{ \alpha_i [(M-z)(\theta_j - \delta_i) - \sum_{k=0}^{M-z} \tau_{ik}] \}}{\sum_{w=0}^C \{ \exp \{ \alpha_i [w(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik}] \} + \exp \{ \alpha_i [(M-w)(\theta_j - \delta_i) - \sum_{k=0}^{M-w} \tau_{ik}] \} \}} \quad (1)$$

其中:  $Z_i$  是第  $i$  个项目的观察反应,  $z = 0, 1, \dots, C$ ,  $z = 0$  代表最强烈的不赞成水平,  $z = C$  代表最强烈的赞成水平;  $M = 2 \times C + 1$ ;

$\theta_j$  = 第  $j$  个个体在潜在连续体上的位置;

$\delta_i$  = 第  $i$  个项目在潜在连续体上的位置;

$\alpha_i$  = 第  $i$  个项目的区分度;

$\tau_{ik}$  = 第  $i$  个项目的第  $k$  个主观反应类别阈限。

上述函数的建立是基于项目反应的四个假设。第一, 与特质正向关联的项目反应概率随反应者( $j$ )和刺激( $i$ )在潜在反应连续体上的距离的增加而减少, 即假设观察反应服从一个理想点过程; 第二, 每种观察反应和两种主观反应相关联, 因此, 反应者可能赞成或不赞成在同一潜在连续体上位置位于其特质水平之上或之下的一个刺激。第三, 主观反应服

从优势过程, 主观反应被折迭(folded)以便产生一个服从理想点过程的观察反应。第四, 主观反应类别阈限( $\tau_{ik}$ )是关于点  $\theta_j - \delta_i$  (即,  $\theta - \delta$  连续体上的原点)对称的。

Roberts, Donoghue 和 Laughlin 对  $\tau_{ik}$  的限定(即第四个假设)可能过于严格了。因为没有任何证据显示主观反应类别是关于点  $\theta_j - \delta_i$  对称的, 当反应类别较多时, 对  $\tau_{ik}$  所做的对称限定可能会导致反应类别的阈限参数不被正确地估计, 从而影响到其它项目参数估计值的准确性。目前已有研究者发现, GGUM 拟合 EPQ 人格问卷, 但却不拟合大五人格问卷<sup>[2]</sup>, 这很可能是由于 GGUM 对  $\tau_{ik}$  的限定过于严格导致的。去除  $\tau_{ik}$  对称的限定之后, 公式 1 可以被重写为:

$$\Pr [Z_{ji} = z] = \frac{\exp \{ \alpha_i [z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}] \} + \exp \{ \alpha_i [(M-z)(\theta_j - \delta_i) - \sum_{k=0}^{M-z} \tau_{ik}] \}}{\sum_{w=0}^C \{ \exp \{ \alpha_i [w(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik}] \} + \exp \{ \alpha_i [(M-w)(\theta_j - \delta_i) - \sum_{k=0}^{M-w} \tau_{ik}] \} \}} \quad (2)$$

也可记为:

$$\Pr [Z_{ji} = z] = \frac{\exp \{ \alpha_i \sum_{k=0}^z [\theta_j - (\delta_i + \tau_{ik})] \} + \exp \{ \alpha_i \sum_{k=0}^{M-z} [\theta_j - (\delta_i + \tau_{ik})] \}}{\sum_{w=0}^C \{ \exp \{ \alpha_i \sum_{k=0}^w [\theta_j - (\delta_i + \tau_{ik})] \} + \exp \{ \alpha_i \sum_{k=0}^{M-w} [\theta_j - (\delta_i + \tau_{ik})] \} \}} \quad (3)$$

\* 基金项目: 江西省高校人文社会科学研究规划基金项目(XL1001)。

通讯作者: 戴海琦, E-mail: daihaiqi@yahoo.com.cn。

公式 2 或公式 3 中各参数的定义同公式 1。这个拓展的模型在此研究中暂时被命名为拓广分部展开模型,简记为 GPUM。GPUM 与 GGUM 的唯一不同之处就在于从上端赞成或不赞成的反应阈限与从下端赞成或不赞成的反应阈限并不限定为对称。

尽管将 GGUM 拓展为 GPUM 会导致待估参数的数量成倍增加,但在项目数较少、反应者数量较多的情况下,有可能获得较好的估计值,而且由于它更充分地建模了反应数据的信息,因而可能获得更准确的估计值,产生更好的模型—数据拟合。由于上述参数较多,同时 MCMC 方法更有利于复杂模型的参数估计,因而,研究中将采用 MCMC 方法来估计 GPUM 中的参数。

2 研究方法

2.1 对象

随机抽取江西两所高校共计 3864 名大学生进行施测,所得数据随机分成两半,一半为 1929 人,记为样本 A;另一半为 1935 人,记为样本 B。采用样本 A 的反应数据估计项目参数,用样本 B 的反应数据进行模型—数据拟合检验。

2.2 工具

采用 Scheier 等人编制的生活取向测验修订版 (LOT-R) 作为测验材料<sup>[3]</sup>。LOT-R 是一个自我报告问卷,共包括 10 个项目,其中 6 个是正式项目,用来测量气质性乐观;其余 4 个是附加项目,用来掩饰问卷的真实目的,后者不用于计算总分。6 个正式项目中有 3 个正性词项目的内容与乐观正向关联,例如“在不确定的时候,我经常期望最好”;3 个负性词项目的内容与乐观反向关联,例如“如果事情有出错的可能,那么一定就会出错”。采用 5 点利克特量表进行评定,从“强烈不赞成”到“强烈赞成”分

别记为 0~4 个等级。Scheier 等人认为 LOT-R 只包括一个维度,维度的两端分别是乐观和悲观<sup>[3]</sup>。邓稳根等人采用理想点方法验证了该测验的中译本也只包含一个维度<sup>[4]</sup>。

2.3 数据分析方法

采用 WinBUGS14.3 程序估计项目参数<sup>[5]</sup>。由于可获得的 MODFIT 程序不能用于新模型<sup>[6]</sup>,所以研究者们开发了一个模型—数据拟合程序用于计算单项目、两项目对和三项目组的  $\chi^2/df$ <sup>[7]</sup>,很遗憾的是,由于研究者对校正的  $\chi^2/df$  的计算原理不太了解,而指标的提出者 (Drasgow 等人) 又没有公布计算的细节,所以新程序目前还不能计算该值。尽管未校正的  $\chi^2/df$  比校正的  $\chi^2/df$  略微不同,总体上可能会略微高估模型的拟合,但这种高估并不会太大,而且关键是并不会影响模型比较的结果。

研究者在 WinBUGS14.3 程序中设定区分度参数服从对数正态分布 (平均数为 0,标准差为 2) 之外,设定其它参数都服从标准正态分布。项目位置参数的初始值全部设定为 0。个人参数从标准正态分布中随机抽取得到。项目区分度参数的初始值全设定为 1。项目阈限参数的初始值从 (-3,+3) 的均匀分布中随机抽样。在迭代次数上,研究者先进行了 50000 次可靠性试验,然后陆续再迭代 200000 次,基本上达到了收敛标准。

3 新编程序与 MODFIT 的比较

新编程序在用于新模型之前,研究者先比较了它与 MODFIT 程序用于 GGUM 时计算到的  $\chi^2/df$  的差异。表 1 列出了自编程序和 MODFIT 计算的单项目、两项目对和三项目组的  $\chi^2/df$  结果。

表 1 MODFIT 和自编程序计算的平均的  $\chi^2$  和  $\chi^2/df$  的值

		单项目		两项目对		三项目组	
		M	SD	M	SD	M	SD
MODFIT	$\chi^2$	11.10	7.19	93.28	28.64	201.54	45.96
	$\chi^2/df$	2.78	1.80	4.21	1.29	3.66	1.13
自编程序	$\chi^2$	11.14	7.20	93.43	28.68	201.88	46.00
	$\chi^2/df$	2.78	1.80	4.21	1.29	3.67	1.13

从表 1 中可以看出,采用新编的程序计算得到的平均的  $\chi^2$  和  $\chi^2/df$  与 MODFIT 计算得到的值相差很小,两个程序的计算结果几乎是一致的。进一步考查每个项目、每个项目对和每个项目组的  $\chi^2$  和  $\chi^2/df$ ,发现两个程序计算的结果差异也很小,限于篇幅,这里不再一一列出计算结果,感兴趣的研究者可以参阅邓稳根的博士论文<sup>[8]</sup>。这些结果表明新编的程序与 MODFIT 程序在计算  $\chi^2/df$  上是同样有

效的。

4 研究结果

4.1 项目参数估计值及其标准误

表 2 呈现了 GPUM 中的项目参数及其标准误的 MCMC 估计值。从表 2 中可以看出,在参数估计值方面,项目位置参数估计值都趋向于位于潜在连续体的中立位置,项目 4 和项目 5 的阈限参数估计值的绝对值相对于其它项目更小,同时这两个项目

的区分度参数估计值比其它项目的区分度参数估计值更大,似乎表现出项目区分度越大,项目的阈限值越小的规律。在项目参数估计值的标准误方面,区分度参数估计值的标准误最低,项目位置参数和阈限参数估计值的标准误略高一些。

4.2 模型—数据拟合

表 3 呈现了采用样本 *B* 进行交叉验证时,分别得到的单项目、两项目对和三项目组的  $\chi^2/\text{df}$  结果。从表 3 中可以看出,采用 GPUM 分析数时得到的每个项目的单项目、两项目对和三项目组的  $\chi^2/\text{df}$  一

般都比采用 GGUM 时得到的相应结果(见表 1)低,所有单项目,或两项目对,或三项目组平均的  $\chi^2/\text{df}$  也显著低于 GGUM 时相应的值。进一步研究发现,每个单项目  $\chi^2/\text{df}$  都低于 3,大多数两项目对的  $\chi^2/\text{df}$  大于 3,但除一个项目对的  $\chi^2/\text{df}$  略高于 4 以外,其它的项目对  $\chi^2/\text{df}$  都小于 4。除 3 个项目组的  $\chi^2/\text{df}$  略高于 3 之外,其它项目组的  $\chi^2/\text{df}$  基本上都不超过 3,详细结果也请参阅邓稳根的博士论文<sup>[8]</sup>。这些结果合在一起表明 GPUM 相对 GGUM 有更好的拟合。

表 2 项目参数估计值及其标准误

	$\alpha$	$\delta$	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$	$\tau_5$	$\tau_6$	$\tau_7$	$\tau_8$	$\tau_9$
1	0.51 (0.03)	0.79 (0.30)	-4.84 (0.40)	-2.00 (0.33)	-3.37 (0.35)	2.10 (0.44)	1.36 (0.92)	2.31 (0.84)	2.23 (0.83)	0.49 (0.70)	2.10 (0.69)
2	0.48 (0.04)	-0.49 (0.37)	-3.78 (0.50)	0.43 (0.48)	1.24 (0.52)	3.30 (0.78)	-0.02 (0.83)	-1.97 (0.73)	-1.42 (0.62)	-0.25 (0.44)	2.41 (0.46)
3	0.57 (0.04)	0.62 (0.31)	-4.62 (0.42)	-2.65 (0.35)	-2.69 (0.35)	0.66 (0.47)	1.27 (1.03)	3.47 (0.78)	2.13 (0.81)	1.41 (0.72)	1.68 (0.67)
4	1.17 (0.11)	-0.76 (0.34)	-2.00 (0.40)	0.48 (0.37)	0.19 (0.43)	1.52 (0.57)	0.10 (0.81)	-0.83 (0.56)	-0.31 (0.43)	-0.26 (0.36)	1.47 (0.36)
5	0.98 (0.10)	-0.81 (0.28)	-2.80 (0.42)	-0.17 (0.33)	0.28 (0.37)	1.63 (0.59)	-0.05 (0.77)	-1.15 (0.42)	0.83 (0.31)	0.11 (0.32)	1.94 (0.30)
6	0.67 (0.05)	0.50 (0.33)	-2.85 (0.39)	-2.29 (0.37)	-3.81 (0.38)	0.38 (0.49)	0.81 (0.97)	3.63 (0.73)	2.80 (0.72)	0.64 (0.67)	0.77 (0.56)

表 3 模型—数据拟合结果

	单项目		两项目对		三项目组	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
$\chi^2$	6.37	3.58	69.59	16.27	145.05	25.84
$\chi^2/\text{df}$	1.59	0.90	3.15	0.73	2.51	0.55

5 讨论

研究者将 GGUM 拓展到 GPUM,不对项目主观类别阈限参数作对称的限制,采用 MCMC 方法估计项目的参数,结果显示采用单项目、两项目对和三项目组  $\chi^2/\text{df}$  比较 GPUM 和 GGUM 的模型—数据拟合时,GPUM 显得更好地拟合反应数据。同时,研究者发现,当区分度参数估计值越高时,阈限参数估计值显得越低。过去研究者常常讨论项目位置参数和阈限参数之间的关系,忽视项目区分度与阈限参数之间的关系。这里的研究结果提示人们也要重视项目区分度参数与其它参数之间的关系。

研究中得到的项目区分度参数估计精度略高一些,但项目位置参数和阈限参数估计值的精度显得更低。可能是由于模型的复杂性导致的,模型复杂时需要估计更多的项目参数,因而可能增大参数估计的标准误,该结果与一般研究者的结果是一致的。

为了研究 GPUM 的模型—数据拟合,研究者开发了一个程序计算单项目、两项目对和三项目组  $\chi^2/\text{df}$ 。结果显示新编的程序与 MODFIT 一样可靠。这既重要而又有意义。因为所获得的 MODFIT 程序有许多功能的局限,例如最多

只能用于 3000 个被试样本,项目数不超过 40 个,而且仅能用于几个有限的模型。而新编程序可以突破这些限制,特别能用于考查新模型的模型—数据拟合,从而有利于人们对新模型的探索。非常遗憾的是研究者没有编制出校正的  $\chi^2/\text{df}$  计算程序,主要原因在前面已经介绍过了,他们曾尝试按自己的理解去编制该程序,但结果都不是很理想,即与 MODFIT 计算的校正  $\chi^2/\text{df}$  值不太一致。

研究结果显示 GPUM 较 GGUM 更拟合反应数据。但估计值的精度略低一些,这可能主要是由于参数较多,样本量不够大导致的,在样本容量足够大的情况下,相信 GPUM 的项目参数及其标准误估计值会更好。由于人格测验的项目数量是较少的,因此采用更复杂的模型是可以考虑的,更复杂的模型可能更好地建构了项目反应过程的复杂性,从而有利于更正确地估计各项目的参数,取得更好的模型—数据拟合。在大样本施测的情景下,GPUM 有望取代 GGUM 用于参数估计。

GGUM 除了可以拓展到 GPUM 之外,还可以做进一步的拓展,例如,拓展为下列模型:

$$Pr [Z_{ji} = z] = \frac{\exp \left\{ \sum_{k=0}^z \alpha_{ik} [\theta_j - (\delta_i + \tau_{ik})] \right\} + \exp \left\{ \sum_{k=0}^{M-z} \alpha_{ik} [\theta_j - (\delta_i + \tau_{ik})] \right\}}{\sum_{w=0}^C \left\{ \exp \left\{ \sum_{k=0}^w \alpha_{ik} [\theta_j - (\delta_i + \tau_{ik})] \right\} + \exp \left\{ \sum_{k=0}^{M-w} \alpha_{ik} [\theta_j - (\delta_i + \tau_{ik})] \right\} \right\}} \quad (4)$$

公式 4 与公式 2 或公式 3 不同之处在于区分度参数定义在每个项目的每个反应类别上。这个模型相对于 GPUM 更复杂,可能比 GGUM 和 GPUM 能更好地建构项目反应的复杂性,但是要估计的参数更多,参数的估计值可能更不精确。因而可能不太适合反应类别较多的测验。不过对于反应类别较少(例如只有两种反应类别),项目数较合适的人格测验可能是适合的,在大样本情况下可能能够取得合理的估计值。

由于 GGUM 的拓展模型较 GGUM 更复杂,因此采用传统的 MML 方法可能难以估计它们的参数,相反地,采用 MCMC 方法却能实现更复杂模型的参数估计。但 MCMC 方法用于这类模型的参数估计能力,即在多大样本容量下,采用多少种反应类别和多少数量的项目、以及在什么样的先验分布情况下才能取得合理的估计值目前还不太清楚。对 GGUM 模型进行拓展的研究只是一个初步性的研究,未来需要深入开展这些新拓展模型的各种模拟和实际研究,考查新模型的适用性。

#### 参考文献

- 1 Roberts J S, Donoghue J R, Laughlin J E. A general item response theory model for unfolding unidimensional

polytomous responses. *Applied Psychological Measurement*, 2000, 24(1): 3—32.

- 2 郭庆科, 苗金凤, 王昭. IRT 展开模型及对非累积反应机制的检测. *心理学探新*, 2006, 26(1): 66—69.
- 3 Scheier M F, Carver C S, Bridges M W. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the life orientation test. *Journal of Personality and Social Psychology*, 1994, 67(6): 1063—1078.
- 4 邓稳根, 戴海琦. 采用理想点方法检验生活取向测验及其修订版的单维性. *心理学探新*, 2009, 29(2): 90—96.
- 5 Lunn D J, Thomas A, Best N, et al.. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 2000, 10(4): 325—337.
- 6 Stark S. MODFIT (computer software). <http://work.psych.uiuc.edu>, 2003.
- 7 Drasgow F, Levine M V, Tsien S, et al.. Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*, 1995, 19(2): 143—165.
- 8 邓稳根. 人格测验的项目反应过程与数据分析方法研究. 博士学位论文. 南昌: 江西师范大学, 2009.

## Analysis of the Fit of Extended Generalized Graded Unfolding Model to the Revised Life Orientation Test

Deng Wengen<sup>1</sup>, Dai Haiqi<sup>2</sup>, Liu Tiechuan<sup>2</sup>, Dai Huiqun<sup>1</sup>

(1. Department of Educational Science, Gannan Teachers' College, Ganzhou 341000;

2. School of Psychology, Jiangxi Normal University, Nanchang 330022)

**Abstract:** This research extended generalized graded unfolding model (GGUM) to a new model, subjective category thresholds of which were not symmetrical. The response data of the revised life orientation test were analyzed by this new model and GGUM. The fit of these two model to the data was compared. Results displayed that the new model fit the data better than GGUM. By these results, the authors of this study suggested that the new model was more fit to analyze the graded data of personality, and GGUM can be replaced by this model and other more complex models.

**Key words:** generalized graded unfolding model; subjective response category threshold; symmetry; model—data fit