文章编号: 1000-5862(2012)05-0456-05

概化理论偏态分布数据方差分量标准误估计

黎光明[1], 张敏强[2*], 戴海琦[3]

(1. 广州大学教育学院心理学系, 广东 广州 510006; 2. 华南师范大学心理应用研究中心, 广东 广州 510631;
3. 江西师范大学心理学院, 江西 南昌 330027)

摘要: 利用 GH 分布性质, 采用 Monte Carlo 数据模拟技术, 模拟生成一定偏度的偏态分布数据, 运用 Traditional 方法、Jackknife 方法、Bootstrap 方法和 MCMC 方法估计概化理论偏态分布数据的方差分量标准误, 探讨了数据的不同偏度对概化理论方差分量标准误估计的影响. 研究结果显示: Jackknife 方法估计偏态分布数据的方差分量标准误性能较差, Traditional和MCMC方法尚可, Bootstrap方法标准误偏差相对较小, 且偏态分布数据的偏度对概化理论方差分量标准误估计有影响, Bootstrap 方法对于偏态分布数据表现出良好的“适应性”, 偏度对其影响较小.

关键词: 概化理论; 偏态分布数据; 方差分量标准误估计; Bootstrap 方法; Monte Carlo 模拟

中图分类号: O 626.4 文献标志码: A

0 引言

概括化理论(generalizability theory, GT), 简称概化理论, 是 20 世纪 60 年代由克隆巴赫(Cronbach)和格莱塞(Gleser)等以数学形式化后引进测量领域的一种心理计量学理论[1]. 概化理论将测量情境关系分为测量目标和测量侧面 2 个部分, 测量目标是欲考察的实际特质, 而测量侧面是影响测量目标的各种因素[2]. 测量侧面在多大程度上影响了测量目标, 即在总变异中, 测量目标和测量侧面的方差分量各占多少, 是概化理论分析最为关心的问题[3].

在样本统计量研究中, 仅用一个(次)样本平均数来估计总体均值, 存在较大的风险, 因为样本平均数容易受抽样的影响. 与平均数类似, 概化理论估计的方差分量, 也受到抽样的影响, 用某个(次)样本方差分量来估计总体方差分量(参数), 存在一定误差[4]. 为了降低这种误差带来的风险, 需要报告方差分量对应的标准误, 来反映可能存在的变化程度.Gaoxiaohong 等[5]认为, 估计的方差分量、误差方差和概化系数等统计量受限于抽样, 不同的抽样样本估计的统计量可能不一样, 应该重视考察方差分量及其相关统计量的变异量.

数据分布对概化理论方差分量标准误估计可能产生影响. 特别地, 当数据为偏态分布时, 适合于正态分布数据的方差分量标准误估计方法不一定适合于偏态分布数据. R. L. Brennan 提供了正态分布下方差分量标准误估计的理论公式[6]p181 为

$$\hat{\sigma}[\hat{\sigma}^2(\alpha|M)]=\sqrt{\sum_{\beta}\frac{2[f(\beta|\alpha)MS(\beta)]^2}{df(\beta)+2}}\ , \tag{1}$$

其中, M 表示模型, α 表示分数效应, β 表示进入 $\hat{\sigma}^2(\alpha|M)$ 的均方, $f(\beta|\alpha)$ 表示 $MS(\beta)$ 的系数. 但是, 即使各侧面分数效应服从正态分布, 也无法知道相应方差分量服从何种分布(残差分布除外), 所以公式(1)也仅是一个方差分量标准误近似的估计公式. 实际上, 分数效应不一定服从正态分布, 那么这些分数的方差分量标准误就更加难以估计[6]p181.据此, 方差分量标准误估计不得不寻找其它途径.

随着社会的发展, 测量的应用领域发生了较大变化, 被测群体的知识和能力等特质在一定程度上不再服从偏度为 0 的分布[7]. A. R. Othman [8]p8 的研究表明, 许多测验数据的分布呈弱偏态, 如 CAP (california assessment program)和 UCSB (university of

收稿日期: 2012-05-16

基金项目: 教育部人文社会科学研究青年基金(12YJC190016), 全国教育科学“十二五”规划教育部重点课题(GFA111009)和广东省教育科学“十二五”规划 2011 年度研究课题(2011TJK161)资助项目.

作者简介: 张敏强(1995-), 广东河源人, 教授, 博士生导师, 主要从事心理与教育统计和测量方面的研究.

california santa barbara)，这 2 个测验数据的分布偏度值介于−0.91~0.85，如表 1 所示.

虽然 A.R.Othman 已经考虑到数据分布具有(弱)偏态，但是 A.R.Othman 并没有进行偏态分布数据的方差分量标准误估计，显得不足．本文旨在探讨偏态分布数据偏度如何影响概化理论的方差分量标准误估计.

表 1　飞行员表现性评价测验分数的偏度

实证数据			偏度
CAP	1990	飞行员评估	0.27
	1992	飞行员评估	
		5 年级	−0.06
		8 年级	0.46
		10 年级	0.14
UCSB			
	1992	飞机倾斜评估	−0.91
	1992	飞行起降评估	−0.73
	1993	飞行应激评估	0.85

1　方法

1.1　数据产生

基于 $p\times i$ 设计概化理论模型，根据 GH 分布的性质使用蒙特卡洛数据模拟技术产生偏态分布数据[9].

1.2　比较标准

比较标准为“偏差”(*bias*)，计算方法为 $bias = (\hat{\theta}_i - \theta)$，$\hat{\theta}_i$ 表示方差分量及其变异量的估计值，θ 为参数，偏差的绝对值(称为“绝对偏差”)越大，表明估计值与参数的差异越大，结果越不可靠 [10].

1.3　分析工具

分析工具为 R 软件、WinBUGS 软件、R2WinBUGS 软件包、Coda 软件包和 HyperbolicDist 软件包．借助这些软件或软件包，自编完成研究程序.

2　结果

2.1　3 种偏态分布数据估计人的方差分量标准误

对 $\beta = -2$、$\beta = -1$ 和 $\beta = 0$ 这 3 种偏态分布数据，分别计算 Traditional 方法、Jackknife 方法、Bootstrap 方法和 MCMC 方法估计人的方差分量标准误，结果如表 2 所示.

表 2　**3 种偏态分布数据估计的人的方差分量标准误**

		SE(vc.p_2)	SE(vc.p_1)	SE(vc.p_0)
Traditional	parameters	0.370 5	0.119 6	0.088 7
	trad	0.247 2	0.096 0	0.076 3
Jackknife	jack-p	0.210 3	0.085 1	0.067 3
	jack-i	0.165 7	0.064 4	0.051 7
	jack-pi	0.499 9	0.193 7	0.154 2
Bootstrap	boot-pi	0.349 0	0.120 2	0.092 0
	boot-pir	0.339 4	0.116 1	0.088 7
	boot-ir	0.074 4	0.029 1	0.023 0
	boot-i	0.078 7	0.030 6	0.024 2
	boot-pr	0.339 4	0.116 1	0.088 7
	boot-p	0.330 6	0.112 3	0.085 6
	boot-piadj	0.352 6	0.121 5	0.093 0
	boot-piradj	0.342 9	0.117 3	0.089 6
	boot-iradj	0.074 6	0.029 2	0.023 1
	boot-iadj	0.079 2	0.030 8	0.024 4
	boot-pradj	0.342 9	0.117 3	0.089 6
	boot-padj	0.333 9	0.113 5	0.086 5
MCMC	MCMC inf	0.246 6	0.095 7	0.076 1
	MCMC non-inf	0.256 0	0.099 5	0.079 0

在表 2 中，SE(vc. p_2)、SE(vc. p_1)和 SE(vc. p_0)分别表示 $\beta = -2$、$\beta = -1$ 和 $\beta = 0$ 偏态分布数据 4 种方法估计人的方差分量标准误．trad 表示 Traditional 方法．Jackknife 方法估计方差分量及其变异量时，需要考虑再抽样策略，这里仅使用 jack-p、jack-i 和 jack-pi 共 3 种策略．Bootstrap 方法也需要考虑再抽样策略，所考虑 Bootstrap 再抽样策略包括 boot-p、boot-i、boot-pi、boot-pr、boot-ir 和 boot-pir．可以对 Bootstrap 方法进行校正，用后缀 adj 来表示校正的 Bootstrap 策略．MCMC inf(MCMC with informative priors)表示有先验信息的 MCMC 方法，而 MCMC non-inf(MCMC with non-informative priors)则表示无先验信息的 MCMC 方法.

2.2　3 种偏态分布数据估计题目的方差分量标准误

对 $\beta = -2$、$\beta = -1$ 和 $\beta = 0$ 的偏态分布数据，分别使用 Traditional 方法、Jackknife 方法、Bootstrap 方法和 MCMC 方法估计题目的方差分量标准误，结果如表 3 所示.

在表 3 中，SE(vc. p_2)、SE(vc. p_1)和 SE(vc. p_0)分别表示 $\beta = -2$、$\beta = -1$ 和 $\beta = 0$ 偏态分布数据 4 种方法估计的题目的方差分量．其它表示符号及解释同表 2.

表 3　3 种偏态分布数据估计的题目的方差分量标准误

		SE(vc.i_2)	SE(vc.i_1)	SE(vc.i_0)
Traditional	parameters	0.813 0	0.272 3	0.200 3
	trad	0.554 7	0.211 5	0.167 0
Jackknife	jack-p	0.068 9	0.026 8	0.020 8
	jack-i	1.245 6	0.472 9	0.365 1
	jack-pi	0.513 6	0.195 0	0.154 1
Bootstrap	boot-pi	0.615 5	0.210 4	0.164 9
	boot-pir	0.610 5	0.208 5	0.163 4
	boot-ir	0.610 5	0.208 6	0.163 4
	boot-i	0.605 6	0.206 5	0.161 8
	boot-pr	0.073 1	0.028 5	0.022 5
	boot-p	0.074 9	0.029 2	0.023 1
	boot-piadj	0.647 9	0.221 5	0.173 6
	boot-piradj	0.642 7	0.219 5	0.172 0
	boot-iradj	0.642 7	0.219 5	0.172 0
	boot-iadj	0.637 5	0.217 3	0.170 3
	boot-pradj	0.073 1	0.028 5	0.022 5
	boot-padj	0.074 9	0.029 2	0.023 1
MCMC	MCMC inf	0.565 5	0.215 8	0.170 5
	MCMC non-inf	0.664 3	0.253 0	0.199 7

2.3　3 种偏态分布数据估计人与题目交互的方差分量标准误

对 $\beta=-2$、$\beta=-1$ 和 $\beta=0$ 的偏态分布数据，分别使用 Traditional 方法、Jackknife 方法、Bootstrap 方法和 MCMC 方法估计人与题目交互的方差分量标准误，结果如表 4 所示.

表 4　3 种偏态分布数据估计的人与题目交互的方差分量标准误

		SE(vc.pi_2)	SE(vc.pi_1)	SE(vc.pi_0)
Traditional	**parameters**	0.081 0	0.027 5	0.020 3
	trad	0.054 1	0.021 2	0.016 7
Jackknife	jack-p	0.071 2	0.024 4	0.018 1
	jack-i	0.178 6	0.059 4	0.044 6
	jack-pi	0.500 0	0.195 7	0.154 0
Bootstrap	boot-pi	0.128 9	0.043 5	0.032 3
	boot-pir	0.073 6	0.024 7	0.018 4
	boot-ir	0.073 6	0.024 7	0.018 4
	boot-i	0.076 9	0.026 3	0.019 7
	boot-pr	0.073 6	0.024 7	0.018 4
	boot-p	0.079 2	0.026 4	0.019 5
	boot-piadj	0.137 1	0.046 2	0.034 4
	boot-piradj	0.078 3	0.026 3	0.019 5
	boot-iradj	0.078 3	0.026 3	0.019 5
	boot-iadj	0.081 0	0.027 7	0.020 7
	boot-pradj	0.078 3	0.026 3	0.019 5
	boot-padj	0.080 0	0.026 6	0.019 7
MCMC	MCMC inf	0.053 0	0.020 7	0.016 3
	MCMC non-inf	0.053 7	0.021 0	0.016 5

在表 4 中，SE(vc. p_2)、SE(vc. p_1)和 SE(vc. p_0)分别表示 $\beta=-2$、$\beta=-1$ 和 $\beta=0$ 偏态分布数据 4 种方法估计的题目的方差分量. 其它表示符号及解释同表 2.

3　分析与讨论

从表 2~表 4 可以看出，Jackknife 方法严重高估或低估 3 个方差分量的标准误，例如，jack-p 估计 β 为−2、−1 和 0 偏态分布数据 vc. i 的标准误，参数分别为 0.810 3、0.272 3 和 0.200 3，而估计值分别为 0.068 9、0.026 8 和 0.020 8，明显低估，jack-i 和 jack-pi 策略情况类似. 没有任何一种 Jackknife 策略能够有效估计方差分量的标准误. 鉴于此，Jackknife 方法估计偏态分布数据方差分量标准误的结果，不再给予进一步的分析，即可认为此种方法不适合于估计偏态分布数据方差分量标准误. 因此，没有将 Jackknife 方法的结果列于图 1~图 3 中.

3.1　3 种偏态分布数据估计的 *p*、*i* 和 *pi* 的方差分量标准误偏差分析

根据表 2~表 4 中每种方法(或策略)估计的方差分量标准误与参数的差值(*bias*)，可以绘出 3 种不同偏态分布数据 4 种方法估计的方差分量标准误偏差图，如图 1~图 3 所示.

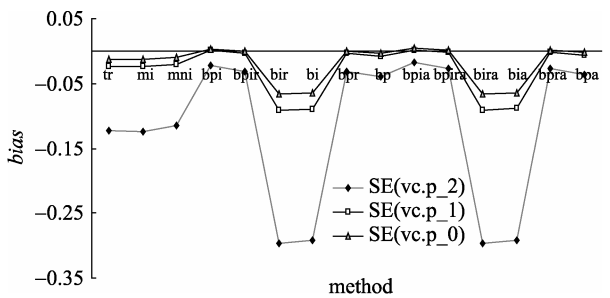

图 1　3 种偏态分布数据 3 种方法估计 *p* 的方差分量标准误偏差

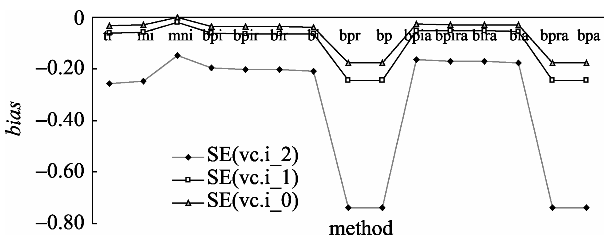

图 2　3 种偏态分布数据 3 种方法估计 *i* 的方差分量标准误偏差

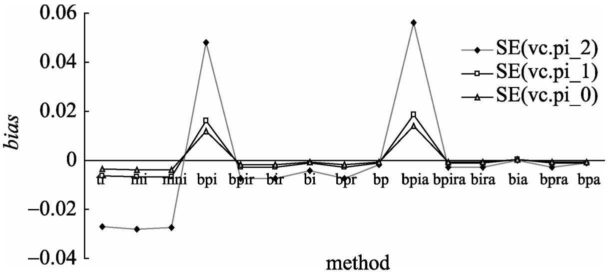

图 3　3 种偏态分布数据 3 种方法估计 ***pi*** 的方差分量标准误偏差

在图 1~图 3 中, tr 表示 trad, mi 表示 MCMC inf, mni 表示 MCMC non-inf, jp 表示 jack-p, ji 表示 jack-i, jpi 表示 jack-pi, bpi 表示 boot-pi, bpir 表示 boot-pir, bir 表示 boot-ir, bi 表示 boot-i, bpr 表示 boot-pr, bp 表示 boot-p, bpia 表示 boot-piadj, bpira 表示 boot-piradj, bira 表示 boot-iradj, bia 表示 boot-iadj, bpra 表示 boot-pradj, bpa 表示 boot-padj.

在图 1 中, SE(vc. p_2)、SE(vc. p_1)和 SE(vc. p_0)分别表示 β 为－2、－1 和 0 偏态分布数据估计人的方差分量标准误. 在图 2 中, SE(vc. i_2)、SE(vci_1)和 SE(vc. i_0)分别表示 β 为–2、–1 和 0 偏态分布数据估计题目的方差分量标准误. 在图 3 中, SE(vc. pi_2)、SE(vc. pi_1)和 SE(vc. pi_0)分别表示 β 为–2、–1 和 0 偏态分布数据估计的人与题目交互(包括残差)的方差分量标准误.

从图 1~图 3 可知, 对于偏态分布数据, 随着偏度减小, 3 种方法估计 3 个方差分量标准误的偏差趋于减小, 但减小幅度先快后慢, 偏度仅影响偏差减少的量, 并不影响偏差减小的方向, 即可认为随着偏度减小, 估计的方差分量标准误偏差减少的“步调”基本趋于一致.

分别对比 3 种偏度下的 SE(vc.p)、SE(vc.i)和 SE(vc.pi), Traditional 和 MCMC 方法结果相当, 受偏度影响较大, 当偏度较大时(如 β =–2)标准误偏差较大, 2 种方法结果不理想, 当偏度较小时结果较好. 当偏度为 0 时, Traditional、MCMC 方法估计的方差分量标准误偏差较小. 因此, 这 2 种方法界于准确与不准确之间, 即尚可, 不是十分理想, 仅为适中.

从图 1~图 3 可以看出, 校正的和未校正的 Bootstrap 方法估计偏态分布数据方差分量标准误结果相当, 2 种方法估计的标准误呈现“对称”状态. 在图 1 中, 校正的和未校正的 Bootstrap 方法有 2 个相似的“波谷”. 在图 2 中, 校正的和未校正的 Bootstrap 方法也有 2 个相似的 “波谷”, 且“波谷”之间的线几乎平行, 或在一条直线上. 在图 3 中, 校正的和未校正的 Bootstrap 方法有多个极其相似的 “波峰”, 其它 Bootstrap 策略偏差折线几乎平行, 相差较小.

3.2　Bootstrap 方法估计偏态分布数据方差分量标准误性能分析

Bootstrap 方法对于偏态分布数据表现出良好的“适应性”, 偏度对其影响较小, 这是因为不论在何种偏度下, 校正的和未校正的 Bootstrap 方法总能找到某种策略估计的方差分量标准误偏差较小. 校正的和未校正的 Bootstrap 方法估计不同偏态分布数据的方差分量标准误, 结果相当.

但是, 偏态分布数据估计方差分量标准误, Bootstrap 方法也需要使用“分而治之”策略. 这里, 仅选择校正的 Bootstrap 方法估计偏态分布数据方差分量标准误的结果来比较各种策略. 对于不同偏态分布数据, 因为校正的 Bootstrap 方法估计的方差分量标准误偏差减少的“步调”基本趋于一致, 所以可以采取综合 3 种偏态分布数据的偏差来获得“跨分布”的方差分量标准误偏差, 结果如表 5 所示.

在表 5 中, rank 表示排名, 其方法如下: 首先, 计算每个 Bootstrap 策略 β 为–2、–1 和 0 标准误绝对偏差的平均值; 其次, 比较 6 种 Bootstrap 策略在 3 个方差分量标准误上所得平均值的大小; 最后, 根

表 5　3 种偏态分布数据 **Bootstrap** 方法估计的方差分量标准误

	SE(vc.p)				SE(vc.i)				SE(vc.pi)			
	β=–2	β=–1	β=0	*rank*	β=–2	β=–1	β=0	*rank*	β=–2	β=–1	β=0	*rank*
boot-pi	–0.017 9	0.001 9	0.004 3	1	–0.165 1	–0.050 8	–0.026 7	1	0.056 1	0.018 7	0.014 1	6
boot-pir	–0.027 6	–0.002 3	0.000 9	3	–0.170 4	–0.052 8	–0.028 3	3	–0.002 7	–0.001 2	–0.000 8	5
boot-ir	–0.295 9	–0.090 4	–0.065 7	6	–0.170 3	–0.052 8	–0.028 3	2	–0.002 7	–0.001 2	–0.000 8	4
boot-i	–0.291 3	–0.088 8	–0.064 3	5	–0.175 5	–0.055 0	–0.030 0	4	0.000 0	0.000 2	0.000 4	1
boot-pr	–0.027 6	–0.002 3	0.000 9	2	–0.739 9	–0.243 8	–0.177 8	6	–0.002 7	–0.001 2	–0.000 8	3
boot-p	–0.036 6	–0.006 1	–0.002 2	4	–0.738 1	–0.243 1	–0.177 2	5	–0.001 0	–0.000 9	–0.000 6	2

据平均值越小越优的原则进行排名，最优者用“1”来表示，最差者用“6”来表示，依次类推.

根据表 5 的结果，可以得到下列结论：对于估计偏态分布数据的 SE(vc.p)，使用 boot-piadj、boot-pradj、boot-piradj 或 boot-padj 策略较好；对于估计偏态分布数据的 SE(vc.i)，使用 boot-piadj、boot-iradj、boot-piradj 或 boot-iadj 策略较好；对于估计偏态分布数据的 SE(vc.pi)，使用 boot-iadj、boot-padj、boot-pradj、boot-iradj 或 boot-piradj 策略较好.

4 结论

(1)偏态分布数据的偏度对概化理论方差分量标准误估计有影响. 随着偏度减小，标准误偏差趋于减小，“步调”趋于一致. Jackknife 方法估计偏态分布数据的方差分量标准误性能较差，Traditional、MCMC 方法尚可, Bootstrap 方法标准误偏差相对较小.

(2)Bootstrap 方法对于偏态分布数据表现出良好的“适应性”，偏度对其影响较小. 校正的和未校正的 Bootstrap 方法估计不同偏态分布数据的方差分量标准误，结果相当. 但是, 2 种方法都需要使用“分而治之”策略来估计方差分量的标准误.

5 参考文献

[1] 蔡艳, 陈抚良. 多元概化理论在教育评估信度分析中的应用研究 [J]. 江西师范大学学报: 自然科学版, 2007, 31(3): 306-310.

[2] 漆书青, 戴海崎, 丁树良. 现代教育与心理测量学原理 [J]. 北京: 高等教育出版社, 2002: 42-78.

[3] 杨志明, 张雷. 测评的概化理论及其应用 [M]. 北京: 教育科学出版社, 2003.

[4] 戴海崎, 张锋, 陈雪枫. 心理与教育测量 [M]. 3 版. 广州: 暨南大学出版社, 2011.

[5] Gao Xiaohong, Brennan R L. Variability of estimated variance components and related statistics in a performance assessment [J]. Applied Measurement in Education, 2001, 14(2): 191-203.

[6] Brennan R L. Generalizability theory [M]. New York: Springer-Verlag, 2001.

[7] 焦璨, 张敏强, 黄庆均, 等. 非正态分布测量数据对克伦巴赫信度 α 系数的影响 [J]. 应用心理学, 2008, 14(3): 276-281.

[8] Othman A R. Examining task sampling variability in science performance assessments. Unpublished doctoral dissertation [D]. Santa Barbara:University of California, 1995.

[9] Mena R H, Walker S G. On the stationary version of the generalized hyperbolic ARCH model [J]. AISM, 2007, 59: 325-348.

[10] 黎光明, 张敏强. 基于概化理论的方差分量变异量估计 [J]. 心理学报, 2009, 41(9): 889-901.

The Estimating Standard Error of Variance Component for Skewed Distribution Data in Generalizability Theory

LI Guang-ming[1], ZHANG Min-qiang[2*], DAI Hai-qi[3]

(1.Department of Psychology, School of Education in Guangzhou University, Guangzhou Guangdong 510006, China;
2.Center of Studies for Psychological Application, South China Normal University, Guangzhou Guangdong 510631, China;
3.School of Psychology, Jiangxi Normal University, Nanchang Jiangxi 330027, China)

Abstract: To explore how skew has effect on estimating standard error of variance component for Generalizability Theory. Using nature of Generalized Hyperbolic distribution, the study adopts Monte Carlo data simulation technique to simulate skewed distribution data. Traditional method, bootstrap method, jackknife method and Markov Chain Monte Carlo (MCMC)method were used to compare estimating standard error of variance component for skewed distribution data in Generalizability Theory. Jackknife method is not good to estimate standard error of variance component for skewed distribution data. Traditional method and Markov Chain Monte Carlo (MCMC)method were not very suitable, but can be accepted and bootstrap method is better. Skew of skewed distribution data have a effect on estimating standard error of variance component. Bootstrap method is a good adaptability to estimate standard error of variance component for Generalizability Theory. Skew has less effect on Bootstrap method.

Key words: generalizability theory; Skewed distribution data; estimating standard error of variance component; bootstrap method; Monte Carlo simulation

(责任编辑：冉小晓)