

中介效应的点估计和区间估计：乘积分布法、非参数 Bootstrap 和 MCMC 法*

方 杰^{1,2} 张敏强²

(¹广东商学院人文与传播学院, 广州 510320) (²华南师范大学心理应用研究中心, 广州 510631)

摘 要 针对中介效应 ab 的抽样分布往往不是正态分布的问题, 学者近年提出了三类无需对 ab 的抽样分布进行任何限制且适用于中、小样本的方法, 包括乘积分布法、非参数 Bootstrap 和马尔科夫链蒙特卡罗 (MCMC) 方法。采用模拟技术比较了三类方法在中介效应分析中的表现。结果发现: 1) 有先验信息的 MCMC 方法的 ab 点估计最准确; 2) 有先验信息的 MCMC 方法的统计功效最高, 但付出了低估第 I 类错误率的代价, 偏差校正的非参数百分位 Bootstrap 方法的统计功效其次, 但付出了高估第 I 类错误率的代价; 3) 有先验信息的 MCMC 方法的中介效应区间估计最准确。结果表明, 当有先验信息时, 推荐使用有先验信息的 MCMC 方法; 当先验信息不可得时, 推荐使用偏差校正的非参数百分位 Bootstrap 方法。

关键词 中介效应; 乘积分布法; 非参数 Bootstrap 法; MCMC 法; 先验信息

分类号 B841

1 引言

中介 (mediation) 是社会科学研究中重要的方法学概念。如果自变量 X 通过某一变量 M 对因变量 Y 产生一定影响, 则称 M 在 X 和 Y 之间起中介作用或 M 为 X 和 Y 的中介变量 (见图 1)。中介效应分析的目的是判断自变量 X 和因变量 Y 之间的关系是部分或全部归因于中介变量 M (Baron & Kenny, 1986;

MacKinnon, 2008; Yuan & MacKinnon, 2009)。近年来, 中介效应分析大量应用在管理学 (Wood, Goodman, Cook, & Beckman, 2008)、心理学 (MacKinnon, Fairchild, & Fritz, 2007; MacKinnon, 2008; Fairchild & McQuillion, 2010) 等学科中, 因此, 中介效应分析的准确性对于研究者正确理解自变量和因变量之间的关系, 以及这种关系的构成机制都显得尤为重要。

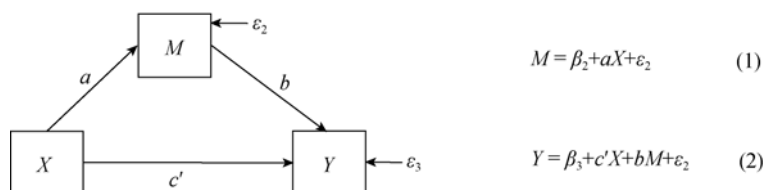


图 1 中介效应模型图

目前中介效应分析普遍使用 Sobel 检验法, 即用中介效应估计值 $\hat{a}\hat{b}$ (\hat{a} 、 \hat{b} 含义见图 1) 除以 $\hat{a}\hat{b}$ 的标准误 $\hat{\sigma}_{\hat{a}\hat{b}}$ 得到一个 z 值 ($z = \hat{a}\hat{b} / \hat{\sigma}_{\hat{a}\hat{b}}$), 将这个 z 值和基于标准正态分布的临界 z 值进行比较, 如果 z

值大于临界 z 值, 说明中介效应显著, 反之则说明中介效应不显著; 或构建一个对称的置信区间 ($\hat{a}\hat{b} - z_{\alpha/2} \times \hat{\sigma}_{\hat{a}\hat{b}}$, $\hat{a}\hat{b} + z_{\alpha/2} \times \hat{\sigma}_{\hat{a}\hat{b}}$), 如果置信区间不包括 0, 说明中介效应显著, 反之则说明中介效应不

收稿日期: 2011-05-11

* 全国教育科学“十二五”规划重点课题 (GFA111009)、广州卓越教育项目: 学生学业水平认知诊断评价的资助。

通讯作者: 张敏强, E-mail: Zhangmq1117@yahoo.com.cn

显著 (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; 温忠麟, 张雷, 侯杰泰, 刘红云, 2004)。Sobel 检验的前提假设是中介效应 $\hat{a}\hat{b}$ 是正态分布且需要大样本, 因为只有在正态分布下, 才能使用基于标准正态分布的临界 z 值。有研究者发现, 即使 \hat{a} 和 \hat{b} 都是正态分布, $\hat{a}\hat{b}$ 也不一定是正态分布, 更进一步的说, 只要 $\hat{a}\hat{b}$ 不为零, $\hat{a}\hat{b}$ 的分布就是偏态分布, 并且分布的峰值还会随着中介效应值 $\hat{a}\hat{b}$ 的变化而变化 (MacKinnon et al., 2002, 2004, 2008; Cheung & Lau, 2008; Hayes, 2009), 因此, 基于中介效应 $\hat{a}\hat{b}$ 是正态分布的中介效应分析方法 (本文称为传统法) 是不可靠的。

为了得到可靠的中介效应分析结果, 近来, 研究者们从不同角度提出了三类不需要对 $\hat{a}\hat{b}$ 的分布进行任何限制且适用于中、小样本的方法, 即乘积分布法 (distribution of the Product)、Bootstrap 方法和马尔科夫链蒙特卡罗方法 (Markov chain Monte Carlo, MCMC)。那么, 这三类方法在中介效应分析中的表现如何呢? 目前关于三类方法的比较都采用数据模拟技术进行。Mackinnon 等人 (2004) 模拟比较了传统法、乘积分布法和五种非参数 Bootstrap 方法在中介效应分析中的表现, 发现偏差校正的非参数百分位 Bootstrap 方法提供了最准确的中介效应置信区间, 统计功效最高, 但要付出在某些条件下高估第 I 类错误率的代价。Yuan 和 MacKinnon (2009) 模拟比较了乘积分布法和 MCMC 方法在中介效应分析中的表现, 发现在中、小样本时, 有先验信息的 MCMC 方法相比乘积分布法能更有效降低均方误, 而无先验信息的 MCMC 方法和乘积分布法在均方误上的表现相当; 两种 MCMC 方法的 95% 置信区间覆盖率都大于乘积分布法, 其中有先验信息的 MCMC 方法的 95% 置信区间覆盖率显著大于 95% (最小值为 96.3%)。但是, 这两个研究存在以下问题:

1) 没有将三类方法一同进行比较。尤其是, 目前大多数研究者都推荐使用 Bootstrap 方法进行中介效应分析 (MacKinnon et al., 2004; Pituch, Stapleton, & Kang, 2006; Williams & Mackinnon, 2008; Taylor, MacKinnon, & Tein, 2008; Pituch & Stapleton, 2008; Cheung & Lau, 2008; Biesanz, Falk, & Savalei, 2010), 那么, Bootstrap 方法和 MCMC 方法之间, 特别是有先验信息的 MCMC 方法之间, 谁的表现更优呢? 尚有待深入研究。

2) 两个研究的比较指标存在不足。首先, 两个研究的比较指标完全不同, 缺乏可比性。其次,

Mackinnon 等人 (2004) 的三个比较指标都是为了考察中介效应置信区间, 没有考察中介效应点估计。

第三, Yuan 和 MacKinnon (2009) 使用的 95% 置信区间覆盖率指标 (若某方法的 95% 置信区间覆盖率越接近 95%, 说明该方法的区间估计越准确) 不能有效反映 MCMC 方法和乘积分布法在区间估计上的优劣。因为 MCMC 方法构建的是可靠区间 (credible interval), 95% 的可靠区间表示在观察数据基础上得到的可靠区间中含有参数真值的概率是 95%, 而乘积分布法和 Bootstrap 方法构建的是置信区间 (confidence interval), 95% 的置信区间表示在总体中进行重复抽样, 并对每次抽样的样本计算一次置信区间, 那么在这些置信区间中, 有 95% 的置信区间会包含着参数的真值 (邱皓政, 2007; Yuan & MacKinnon, 2009; Muthén, 2010), 显然以 95% 置信区间覆盖率为比较指标将更有利于乘积分布法和 Bootstrap 方法, 因此 Yuan 等人 (2009) 发现 MCMC 方法的 95% 置信区间覆盖率高于乘积分布法, 尤其是有先验信息的 MCMC 方法的 95% 置信区间覆盖率显著大于 95%, 并不足以说明 MCMC 方法的区间估计是不准确的, 所以本研究未将 95% 置信区间覆盖率作为比较指标。第四, Yuan 等人 (2009) 还用乘积分布法和 MCMC 方法计算了某中介效应实例研究的区间宽度 (interval width) (见表 1), 发现有先验信息的 MCMC 方法的区间宽度窄于乘积分布法; 笔者更进一步用非参数 Bootstrap 方法计算了同一数据的区间宽度 (见表 1), 发现有先验信息的 MCMC 方法估计的区间宽度最窄 (0.089), 而无先验信息的 MCMC 方法和乘积分布法、非参数 Bootstrap 方法的区间估计表现相当 (0.103~0.110), 显示有先验信息的 MCMC 方法的区间估计似乎最准确。那么, 这个实例分析的结果是否具有普适性呢? 为了回答这个问题, 本研究将区间宽度设定为区间估计准确性的比较指标。

综上所述, 本文的研究目的是将三类方法一同进行点估计和区间估计的全面比较。首先, 简单介绍乘积分布法、非参数 Bootstrap 和 MCMC 方法; 然后用 5 (样本容量) \times 6 (ab 组合) \times 5 (估计方法) 的实验设计进行模拟比较; 接着对模拟比较的结果进行讨论并得出结论, 为心理学研究者在选择中介效应分析方法时, 提供有益的参考。

2 三类中介效应分析方法

2.1 乘积分布法

乘积分布法的基本原理是, 由于中介效应 $\hat{a}\hat{b}$

表 1 三类方法对同一实测数据的中介效应点估计和区间估计

方法	ab 点估计	区间估计的下限	区间估计的上限	区间宽度
乘积分布法	0.056	0.013	0.116	0.103
非参数百分位 Bootstrap 法	0.056	0.008	0.118	0.110
偏差校正的非参数百分位 Bootstrap 法	0.056	0.014	0.123	0.109
有先验信息的 MCMC 法	0.051	0.013	0.102	0.089
无先验信息的 MCMC 法	0.056	0.011	0.118	0.107

在绝大部分情况下都不是正态分布,那么在构建中介效应置信区间时,就不能使用基于标准正态分布的临界 z 值,而应该使用基于 \hat{a} 和 \hat{b} 乘积分布的临界值。乘积分布法默认 $\hat{a}\hat{b}$ 分布符合两个正态分布随机变量的乘积分布(Meeker, Cornwell, & Aroian, 1981),是不对称的偏态分布,因此乘积分布法在构建置信区间时,置信区间的上、下置信限(confidence limit)需要不同的临界值,所得到的置信区间是不对称的置信区间(asymmetric confidence interval) (MacKinnon et al., 2004; Fritz & MacKinnon, 2007; MacKinnon, Fritz, Williams, & Lockwood, 2007; Tofighi & MacKinnon, 2011; 方杰, 张敏强, 李晓鹏, 2011)。

乘积分布法分为二步。第一,利用公式(1)和(2)计算 \hat{a} 、 \hat{b} 、 $\hat{\sigma}_a$ (a 的标准误)、 $\hat{\sigma}_b$ (b 的标准误),得到中介效应的点估计值 $\hat{a}\hat{b}$ 。第二,根据第一步得到的 \hat{a} 、 \hat{b} 、 $\hat{\sigma}_a$ 、 $\hat{\sigma}_b$ 值,以及 ρ_{ab} (a 和 b 的相关)和第一类错误率 α ,利用 PRODCLIN (distribution of the PRODUCT Confidence Limits for INDIRECT effects)程序(MacKinnon, Fritz, Williams, & Lockwood, 2007)或 R 软件的 RMediation 软件包(Tofighi & MacKinnon, 2011)自动在 Meeker 等人(1981)的乘积分布表中查得上、下置信限的临界值,并构建不对称置信区间($\hat{a}\hat{b}$ - 下置信限标准化临界值 $\times \hat{\sigma}_{\hat{a}\hat{b}}$, $\hat{a}\hat{b}$ + 上置信限标准化临界值 $\times \hat{\sigma}_{\hat{a}\hat{b}}$)。

2.2 非参数 Bootstrap 法

Bootstrap 方法是由 Efron 最早提出的一种重复抽样方法(resampling methods),可分为参数 Bootstrap 和非参数 Bootstrap 两类,本文主要探讨非参数 Bootstrap 方法。非参数 Bootstrap 方法的基本思想是将原样本当做“总体”,通过有放回的重复抽样,抽取大量新的子样本并获得统计量的过程,其实质是模拟了从总体中随机抽取大量样本的过程。常用的非参数 Bootstrap 方法包括非参数百分位 Bootstrap 方法(nonparametric percentile Bootstrap method)和偏差校正的(Bias-corrected)非参数百分位 Bootstrap

方法。目前常用的各种统计软件(包括 SPSS)都能进行非参数 Bootstrap 法运算。

2.2.1 非参数百分位 Bootstrap 法 非参数百分位 Bootstrap 方法分为三步。第一,以原样本(样本容量为 n)为基础,在保证每个观察单位每次被抽到的概率相等(均为 $1/n$)的情况下进行有放回的重复抽样,得到一个样本容量为 n 的 Bootstrap 样本;第二,由步骤 1 中得到的 Bootstrap 样本计算出相应的中介效应估计值 $\hat{a}\hat{b}$;第三,重复步骤 1 和 2 若干次(记为 B ,常设 $B = 1000$),将 B 个中介效应估计值的均值作为中介效应的点估计值,将 B 个中介效应估计值 $\hat{a}\hat{b}$ 按数值大小排序,得到序列 C ,用序列 C 的第 2.5 百分位数和第 97.5 百分位数来估计 95% 的中介效应置信区间(Preacher, Rucker, & Hayes, 2007; Preacher & Hayes, 2008; Hayes, 2009; 方杰, 张敏强, 邱皓政, 2012; 温忠麟, 刘红云, 侯杰泰, 2012)。

2.2.2 偏差校正的非参数百分位 Bootstrap 法 偏差校正的非参数百分位 Bootstrap 方法和非参数百分位 Bootstrap 方法有相同的中介效应点估计值。偏差校正的非参数百分位 Bootstrap 方法校正的是非参数百分位 Bootstrap 方法默认序列 C 的中值等于原样本数据求取的中介效应估计值 $\hat{a}\hat{b}^*$ 而带来的区间估计偏差,校正的方法是调整构建置信区间的百分位点,分为以下三步。第一,求 $\hat{a}\hat{b}^*$ 在序列 C 中的百分比排位,即得到 $\hat{a}\hat{b} < \hat{a}\hat{b}^*$ 的概率 $\Phi(z_0)$;第二,在标准正态累积分布函数中,根据 $\Phi(z_0)$ 求取相应的 z_0 值;第三,求 $2Z_0 \pm Z_{\alpha/2}$ 在标准正态累积分布函数中对应的概率 $\Phi(2Z_0 \pm Z_{\alpha/2})$,用 $\Phi(2Z_0 \pm Z_{\alpha/2})$ 在序列 C 中的百分位值作为置信区间的上、下置信限,估计置信度为 $1 - \alpha$ 的中介效应置信区间(Fritz & MacKinnon, 2007; MacKinnon, 2008; Preacher & Hayes, 2008; Taylor, MacKinnon, & Tein, 2008; 方杰 等, 2011; 温忠麟 等, 2012)。

2.3 马尔科夫链蒙特卡罗(MCMC)法

马尔科夫链蒙特卡罗(MCMC)方法是在贝叶斯理论框架下,将马尔科夫链过程引入到蒙特卡罗模拟中,实现抽样分布随模拟的进行而改变的动态模

拟。MCMC 方法的基本思想可概括为如下三步。第一, 构造马尔科夫链, 使其收敛到平稳分布(stationary distribution), 平稳分布通常是待测参数的后验分布。为了使马尔科夫链收敛到平稳分布, 必须丢弃前面 t 次迭代值, 中介效应分析中, t 取 1000 次就足够了(Yuan & MacKinnon, 2009); 第二, 利用马尔科夫链进行吉布斯(Gibbs)抽样, 即利用多个一元全条件(full conditionals)分布(除一个变量外, 其他所有变量都赋予固定值的分布)进行迭代抽样, 获得 $n-t$ (约 10000)个后验样本, 得到的后验样本又被称为马尔科夫链的实现值; 第三, 由后验样本计算 10000 个中介效应估计值 $\hat{a}\hat{b}$, 将 10000 个中介效应估计值的均值作为中介效应的点估计值, 将 10000 个中介效应估计值 $\hat{a}\hat{b}$ 按数值大小排序, 用第 2.5 百分位数和第 97.5 百分位数来得到 95% 的中介效应可靠区间估计。MCMC 方法的计算常用 WinBUGS、Mplus6、SAS 和 R 软件完成(菲诗松, 王静龙, 濮晓龙, 2006; Ntzoufras, 2009; Yuan & MacKinnon, 2009; 方杰 等, 2011)。

MCMC 最有吸引力之处在于, 由贝叶斯公式可知, 待测参数后验分布与先验分布(先验信息的概率分布)和似然函数的乘积成正比, 所以 MCMC 方法允许研究者将先验信息(prior information)整合到中介效应分析中, 称为有先验信息的 MCMC (MCMC with informative priors, MCMC inf)方法, 这样能有效的利用其他相关的前人研究和导航研究的资源, 从而得到更准确的中介效应分析结果(Muthén, 2010)。有先验信息的 MCMC 方法进行中介效应分析时, 公式(1)和(2)的回归系数 a 和 b 的先验分布常设定为正态分布, 残差 ε_2 和 ε_3 的先验分布常设定为逆伽玛(inverse gamma)分布(可确保残差的方差是正值)。以回归系数 a 为例, a 的先验分布可表示为 $a \sim N(\mu, \tau)$, N 表示正态分布, μ 表示均值, τ 表示精度, 即方差 σ^2 的倒数。如果将先验分布设定为 $a \sim N(0, 10^{-6})$, 此时先验分布的均值为 0, 方差很大(为 10^6), 可以看作是无先验信息, 称为无先验信息的 MCMC (MCMC with non-informative priors, MCMC non-inf)方法。

3 模拟研究方法

3.1 模拟描述

从标准正态分布中随机产生样本量为 N 的自变量 X 、 ε_2 和 ε_3 , 然后根据公式(1)-(2)模拟产生中介变量 M 和因变量 Y 。在产生模拟数据的过程中,

系统变化以下 2 个因素: (1)样本量。遵循 Yuan 和 MacKinnon (2009)模拟研究中使用的五种样本量 $N=25, 50, 100, 200, 1000$ 。(2)参数 a 和 b 的组合, 考察三种中介效应为 0 和三种中介效应不为 0 的情况。三种中介效应为 0 的情况包括(a) $a=0, b=0$; (b) $a=0.39, b=0$; (c) $a=0, b=0.59$, 之所以选择这三种情况是因为 MacKinnon 等人(2004)的模拟研究表明, 当 a 和 b 都为 0 时, 第 I 类错误率远低于 $\alpha=0.05$; 当 a 和 b 不都为 0 时, 第 I 类错误率接近或高于 $\alpha=0.05$ 。三种中介效应不为 0 的情况包括中介效应 ab 的真值为 0.0196($a=0.14, b=0.14$), 0.1521($a=0.39, b=0.39$), 0.3481($a=0.59, b=0.59$), 对应于小, 中, 大的中介效应值(MacKinnon et al., 2002, 2004; Williams & MacKinnon, 2008; Yuan & MacKinnon, 2009)。另外一些在产生模拟数据的过程中需要的参数被设定为固定值。例如, 由于以前的中介效应的模拟研究表明, 统计功效不受 c' 变化的影响, 因此设定 $c'=0$ 以简化模型, β_2 和 β_3 也设定为 0(MacKinnon et al., 2004; Fritz & MacKinnon, 2007)。

总之, 模拟数据的产生共包括 30 种(5×6)组合, 每种组合都重复 1000 次。用 5 种方法(乘积分布法、非参数百分位 Bootstrap 法、偏差校正的非参数百分位 Bootstrap 法、有先验信息的 MCMC 法和无先验信息的 MCMC 法)对这 30,000 个($5 \times 6 \times 1000$)数据集进行中介效应分析。如果使用 Bootstrap 方法, 每种 Bootstrap 方法都需对每个数据集进行 1000 次 Bootstrap 抽样。如果使用 MCMC 方法, 则需要对每个数据集迭代 11000 次, 抛弃前 1000 次迭代值, 利用 1001 次到 11000 次的 10000 次迭代值进行中介效应分析。对于有先验信息的 MCMC 方法, 例如 $a=0.14, b=0.14$ 条件下, 先验信息为 $a \sim N(0.14, 100)$, $b \sim N(0.14, 100)$ 。

3.2 比较指标

对于中介效应的点估计, 采用偏差(Bias)和相对均方误(relative mean square error, RMSE) 指标。对于中介效应的区间估计, 采用第 I 类错误率(Type I error)、统计功效(Power)、区间宽度指标。

3.2.1 偏差和相对均方误 偏差的计算公式为 $bias = E(\hat{a}\hat{b}) - ab$, 偏差绝对值的计算公式为 $|E(\hat{a}\hat{b}) - ab|$, $E(\hat{a}\hat{b})$ 表示每种组合重复 1000 次得到的 1000 个中介效应的点估计 $\hat{a}\hat{b}$ 的均值, ab 表示中介效应真值, 偏差越小越好。

均方误(mean square error, MSE)的计算公式为 $MSE = E(\hat{a}\hat{b} - ab)^2$, 即用每种组合重复 1000 次得到

的 1000 个中介效应的点估计 \hat{ab} 与中介效应真值 ab 之差的平方求均值, 均方误差越小越好。相对均方误差的计算公式为 $RMSE = MSE/MSE_0$, MSE_0 表示用乘积分布法进行中介效应分析得到的均方误差, MSE 表示其它方法进行中介效应分析得到的均方误差, 相对均方误差越小越好(Yuan & MacKinnon, 2009)。

3.2.2 第 I 类错误率和统计功效 当中介效应 ab 为 0 时, 每种组合重复 1000 次得到的 1000 个中介效应的置信区间中, 不包含 0 的次数所占比例就是这种组合的第 I 类错误率, 第 I 类错误率越接近设定的真值 0.05 越好, Bradley (1978) 提出第 I 类错误率的合理变化范围是在 0.025 ~ 0.075 之间 (MacKinnon et al., 2004; Williams & MacKinnon, 2008)。当中介效应 ab 不为 0 时, 每种组合重复 1000 次得到的 1000 个中介效应的置信区间中, 不包含 0 的次数所占比例就是这种组合的统计功效, 统计功效越接近 1 越好。

3.2.3 区间宽度 区间宽度的计算公式 = $E(\hat{li}_{0.975}) - E(\hat{li}_{0.025})$, $E(\hat{li}_{0.975})$ 表示每种组合重复 1000 次得到的 1000 个中介效应区间估计的上限均值, $E(\hat{li}_{0.025})$ 表示每种组合重复 1000 次得到的 1000 个中介效应区间估计的下限均值, 区间宽度越小越好。

3.3 分析工具

R 软件、WinBUGS 软件、R2WinBUGS 软件包、Coda 软件包。借助这些软件或软件包, 自编完成研究程序。通过自编的 R 程序, 产生 1000 批次的模拟数据, 直接实现乘积分布法、非参数 Bootstrap 法的中介效应分析, 间接实现 MCMC 方法的中介效应分析。“间接实现”的含义是指: MCMC 方法的中介效应分析是通过自编的 R 程序触发 WinBUGS 软件“间接”地实现, 这个过程要求 R 软件先调用 R2WinBUGS 和 Coda 两个软件包, R2WinBUGS 软件包的作用在于成为 R 软件和 WinBUGS 软件的“桥梁”, Coda 软件包的作用在于输出 WinBUGS 软件生成的 MCMC 结果, 如中介效应的点估计值和 95% 的区间估计(黎光明, 张敏强, 2009)。

4 结果与分析

4.1 中介效应点估计的偏差

对相同的模拟数据, 分别计算乘积分布法、非参数 Bootstrap 法和 MCMC 法的中介效应点估计的偏差, 结果如表 2 所示。由于偏差校正的非参数百分位 Bootstrap 方法和非参数百分位 Bootstrap 方法

的中介效应点估计值相同, 因此在表 2 中统一用 Boot 表示。表 2 中的 MCMCinf 和 MCMCnoninf 分别表示有先验信息的 MCMC 方法和无先验信息的 MCMC 方法。

4×5×2 的方差分析(ANOVA)用来分析 4 种方法, 5 种样本量和 2 种中介效应情况(中介效应为 0 和不为 0)对中介效应点估计偏差的影响。方差分析显示, 只有样本量和方法的交互作用显著, $F(12, 80)=1.964, p=0.039, \eta_p^2=0.228$ 。简单主效应分析显示, 样本量为 25 时, 方法的主效应显著, $F(3, 80)=3.666, p=0.016$ 。多重比较(使用 Bonferroni 校正法, 下同)发现, 有先验信息的 MCMC 方法的偏差(3.5)与乘积分布法(-32.2)差异显著; 样本量为 50 时, 方法的主效应显著, $F(3, 80)=4.483, p=0.006$ 。多重比较发现, 无先验信息的 MCMC 方法(-23.1)与乘积分布法(17.1)差异显著; 其余样本量条件下, 方法的主效应都不显著。只有在乘积分布法下, 样本量的主效应显著, $F(4, 80)=5.528, p=0.001$, 多重比较发现, 样本量 25 与样本量 50、100、1000 差异显著, 样本量为 25 时偏差最大; 其余方法下, 样本量的主效应都不显著。

对于中介效应点估计的偏差绝对值, 4×5×2 的方差分析也显示, 只有样本量和方法的交互作用显著, $F(12, 80)=2.04, p=0.031, \eta_p^2=0.234$ 。简单主效应分析显示, 样本量为 25 时, 方法的主效应显著, $F(3, 80)=6.47, p=0.001$ 。多重比较发现, 有先验信息的 MCMC 方法的偏差绝对值(5.8)显著小于乘积分布法(34.9)和非参数 Bootstrap 法(40.7); 其余样本量条件下, 方法的主效应都不显著。在乘积分布法下, 样本量的主效应显著, $F(4, 80)=4.354, p=0.003$, 多重比较发现, 样本量 25 与样本量 100、1000 差异显著, 样本量为 25 时, 偏差绝对值最大; 在非参数 Bootstrap 法下, 样本量的主效应显著, $F(4, 80)=6.987, p<0.001$, 多重比较发现, 样本量 25 与样本量 50、200、1000 差异显著, 样本量 25 时, 偏差绝对值最大。两种 MCMC 方法下, 都没有发现样本量的主效应显著。

4.2 中介效应点估计的相对均方误差

对相同的模拟数据, 分别计算乘积分布法、非参数 Bootstrap 法和 MCMC 法的中介效应点估计的相对均方误差, 结果如表 2 所示。4×5×2 的方差分析用来分析 4 种方法, 5 种样本量和 2 种中介效应情况(中介效应为 0 和不为 0)对中介效应点估计的相对均方误差的影响。方差分析显示, 只有样本量和方法

表 2 三类方法估计的中介效应偏差(×10000)和相对均方误(×100%)

方法	$a=b=0$		$a=0.39,b=0$		$a=0,b=0.59$		$a=b=0.14$		$a=b=0.39$		$a=b=0.59$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
N=25												
乘积分布法	-32.9	100.0	-29.7	100.0	7.9	100.0	-55.1	100.0	-11.3	100.0	-72.4	100.0
Boot	-3.5	110.9	-39.3	97.1	55.0	106.4	-10.6	116.7	-99.0	107.5	-36.9	99.7
MCMCinf	-0.5	0.1	0.7	2.5	9.9	3.2	4.6	1.6	12.7	3.2	-6.4	3.4
MCMCnoninf	2.6	101.1	-15.9	107.2	-3.0	99.0	-3.0	95.6	-8.0	109.6	-76.2	91.1
N=50												
乘积分布法	5.2	100.0	29.5	100.0	23.0	100.0	10.3	100.0	44.0	100.0	-9.2	100.0
Boot	4.4	95.8	0.2	101.7	3.1	108.6	-16.8	83.2	4.4	100.3	-21.8	109.6
MCMCinf	0.9	1.0	-6.8	8.7	21.5	9.9	1.4	6.6	-2.2	9.5	-1.8	10.9
MCMCnoninf	6.0	85.7	-30.8	95.9	-43.2	100.4	4.3	97.2	-28.4	85.0	-46.2	117.1
N=100												
乘积分布法	1.4	100.0	1.2	100.0	12.5	100.0	6.1	100.0	-1.4	100.0	16.5	100.0
Boot	1.4	87.7	-0.9	105.9	-50.2	92.9	-5.8	107.8	-30.8	102.4	51.5	101.1
MCMCinf	-1.6	5.5	-6.0	23.5	13.7	9.9	0.2	20.6	-9.2	22.4	-27.8	21.7
MCMCnoninf	-3.5	77.2	-14.8	102.4	15.2	100.4	-8.8	96.1	-29.2	89.6	17.1	104.1
N=200												
乘积分布法	2.1	100.0	3.9	100.0	-23.5	100.0	5.0	100.0	-18.2	100.0	-12.0	100.0
Boot	0.5	90.2	-4.2	116.2	-1.8	99.2	-4.4	88.1	2.2	97.0	6.8	111.7
MCMCinf	-0.2	17.5	-10.3	46.1	3.4	41.7	1.5	42.3	-5.6	45.6	8.4	45.6
MCMCnoninf	1.3	80.0	4.6	107.6	9.1	92.7	-9.1	93.6	-0.8	101.0	7.2	108.8
N=1000												
乘积分布法	-0.3	100.0	3.0	100.0	-2.3	100.0	2.1	100.0	-2.6	100.0	1.5	100.0
Boot	0.1	110.0	1.0	94.4	0.8	100.6	-1.4	92.2	0.7	106.7	-4.7	92.5
MCMCinf	-0.1	69.0	-3.9	76.6	11.5	86.7	-1.2	79.1	-5.2	89.6	-1.7	75.3
MCMCnoninf	0.0	86.8	-2.8	96.5	4.7	109.9	-0.2	101.9	5.7	107.6	3.5	97.1

的交互作用显著, $F(12, 80)=11.18, p<0.001, \eta_p^2=0.626$ 。简单主效应分析显示, 五种样本量条件下, 方法的主效应都显著, 多重比较一致发现, 有先验信息的 MCMC 方法的相对均方误显著小于其它三种方法, 这和 Yuan 等人(2009)发现的有先验信息的 MCMC 方法的相对均方误显著小于乘积分布法的研究结果吻合。只有在先验信息的 MCMC 方法条件下, 样本量的主效应显著, $F(4, 80)=42.875, p<0.001$, 多重比较发现, 除了样本量 25 的相对均方误(2.4)与样本量 50 的相对均方误(7.8), 样本量 50 的相对均方误与样本量 100 的相对均方误(19.5)没有显著差异外, 其余样本量之间的相对均方误都有显著差异, 样本量增大, 相对均方误也随之增大。

4.3 中介效应分析的统计功效

对相同的模拟数据, 分别计算乘积分布法、非参数 Bootstrap 法和 MCMC 法的中介效应分析的统

计功效, 结果如表 3 所示。表 3 中的 P_Boot 表示非参数百分位 Bootstrap 方法, Bc_Boot 表示偏差校正的非参数百分位 Bootstrap 方法, MCMCinf 和 MCMCnoninf 分别表示有先验信息的 MCMC 方法和无先验信息的 MCMC 方法。

5×5 的方差分析用来分析 5 种方法, 5 种样本量对中介效应分析的统计功效的影响。方差分析显示, 只有样本量的主效应显著, 样本增加, 统计功效随之增加, $F(4, 50)=5.646, p=0.001, \eta_p^2=0.311$ 。多重比较结果显示, 样本量 1000 的统计功效(0.997)显著大于样本量 25(0.345)、50(0.548)条件。5×3 的方差分析用来分析 5 种方法, 3 种中介效应大小对中介效应分析的统计功效的影响。方差分析显示, 只有中介效应大小的主效应显著, $F(2, 60)=28.521, p<0.001, \eta_p^2=0.487$ 。多重比较结果显示, 中介效应小(0.0196)的统计功效(0.288)显著小于中介效应中等(0.1521)的统计功效(0.790)和中介效应大(0.3481)

的统计功效(0.925)。综上可知, 样本量和中介效应大小是中介效应统计功效的主要决定因素。虽然两次方差分析都没有发现方法的主效应显著, 但是有先验信息的 MCMC 方法的统计功效始终大于其他四种方法(见表 3), 即有先验信息的 MCMC 方法统计功效最高, 而以往的模拟研究由于没有纳入 MCMC 方法参与比较, 因此都认定偏差校正的

Bootstrap 方法统计功效最高(MacKinnon et al., 2004; Fritz & MacKinnon, 2007; Pituch et al., 2006, 2008; Williams & Mackinnon, 2008; Taylor et al., 2008)。表 3 的结果也显示, 偏差校正的非参数百分位 Bootstrap 方法的统计功效始终大于乘积分布法、非参数百分位 Bootstrap 和无先验信息的 MCMC 法。

表 3 三类方法估计的中介效应统计功效、第 I 类错误率和 95% 区间宽度

方法	$a=b=0$		$a=0.39, b=0$		$a=0, b=0.59$		$a=b=0.14$		$a=b=0.39$		$a=b=0.59$	
	error	width	error	width	error	width	power	width	power	width	power	width
N=25												
乘积分布法	0.001	0.237	0.035	0.393	0.053	0.526	0.010	0.287	0.209	0.504	0.611	0.713
P_Boot	0.004	0.312	0.028	0.455	0.056	0.582	0.022	0.351	0.154	0.531	0.548	0.762
Bc_Boot	0.009	0.329	0.049	0.470	0.088	0.593	0.030	0.367	0.234	0.551	0.661	0.782
MCMCinf	0.000	0.040	0.000	0.146	0.000	0.210	0.040	0.079	1.000	0.197	1.000	0.298
MCMCnoninf	0.001	0.326	0.015	0.455	0.019	0.585	0.005	0.370	0.145	0.571	0.534	0.790
N=50												
乘积分布法	0.006	0.108	0.054	0.242	0.051	0.348	0.032	0.157	0.605	0.332	0.959	0.483
P_Boot	0.001	0.144	0.041	0.260	0.058	0.362	0.012	0.182	0.525	0.334	0.932	0.494
Bc_Boot	0.004	0.151	0.070	0.265	0.089	0.366	0.049	0.190	0.643	0.346	0.951	0.503
MCMCinf	0.005	0.037	0.000	0.130	0.000	0.192	0.058	0.070	1.000	0.177	1.000	0.266
MCMCnoninf	0.001	0.150	0.032	0.262	0.041	0.367	0.019	0.181	0.505	0.339	0.933	0.492
N=100												
乘积分布法	0.002	0.053	0.050	0.166	0.060	0.242	0.087	0.092	0.941	0.224	1.000	0.333
P_Boot	0.003	0.070	0.049	0.168	0.053	0.240	0.060	0.103	0.912	0.228	1.000	0.332
Bc_Boot	0.005	0.074	0.087	0.170	0.068	0.241	0.119	0.108	0.946	0.234	1.000	0.336
MCMCinf	0.001	0.030	0.006	0.113	0.006	0.166	0.183	0.060	1.000	0.155	1.000	0.234
MCMCnoninf	0.000	0.072	0.040	0.172	0.052	0.244	0.060	0.102	0.938	0.229	1.000	0.338
N=200												
乘积分布法	0.005	0.027	0.043	0.112	0.053	0.166	0.288	0.060	1.000	0.154	1.000	0.234
P_Boot	0.003	0.036	0.055	0.114	0.060	0.165	0.222	0.063	0.999	0.156	1.000	0.233
Bc_Boot	0.009	0.038	0.079	0.114	0.064	0.165	0.312	0.066	0.999	0.158	1.000	0.235
MCMCinf	0.001	0.021	0.012	0.092	0.016	0.134	0.469	0.048	1.000	0.126	1.000	0.191
MCMCnoninf	0.003	0.036	0.057	0.116	0.041	0.166	0.203	0.063	0.999	0.158	1.000	0.235
N=1000												
乘积分布法	0.002	0.005	0.049	0.049	0.044	0.073	0.985	0.025	1.000	0.069	1.000	0.104
P_Boot	0.002	0.007	0.046	0.049	0.053	0.073	0.99	0.025	1.000	0.068	1.000	0.103
Bc_Boot	0.005	0.008	0.049	0.049	0.058	0.073	0.996	0.026	1.000	0.069	1.000	0.104
MCMCinf	0.000	0.006	0.035	0.046	0.048	0.070	0.995	0.024	1.000	0.065	1.000	0.099
MCMCnoninf	0.000	0.007	0.044	0.049	0.066	0.073	0.982	0.025	1.000	0.068	1.000	0.103

注: 加粗的数字表示第 I 类错误率在 Bradley (1978) 的合理变化范围 0.025~0.075 之外。

4.4 中介效应分析的第 I 类错误率

表 3 中的 error 表示第 I 类错误率。5×5 的方差分析用来分析 5 种方法, 5 种样本量对中介效应分析的第 I 类错误率的影响。方差分析显示, 只有

方法的主效应显著, $F(4, 50)=4.145, p=0.006, \eta_p^2=0.249$ 。多重比较发现, 有先验信息的 MCMC 方法的第 I 类错误率(0.009)显著小于偏差校正的非参数百分位 Bootstrap 方法(0.049), 其余方法之间的

第 I 类错误率无显著差异。3×5 的方差分析用来分析 3 种 a 与 b 的组合, 5 种样本量对中介效应分析的第 I 类错误率的影响。方差分析显示, 只有 ab 组合的主效应显著, $F(2, 60)=37.936$, $p<0.001$, $\eta_p^2=0.558$, 多重比较发现, 0 与 0 组合的第 I 类错误率(0.003)显著小于 0.39 和 0 的组合(0.041)、0 和 0.59 的组合(0.048), $a=b=0$ 时, 各种方法的第 I 类错误率(最大值为 0.009)都远低于真值 0.05, 也低于 a 和 b 不同时为 0 的情况(见表 3), 这和前人的研究结果一致(MacKinnon et al., 2004; Pituch et al., 2006, 2008; Williams & Mackinnon, 2008; Taylor et al., 2008)。两次方差分析都没有发现样本量的主效应显著, 因此, 方法和 ab 组合是第 I 类错误率的主要影响因素。

值得注意的是, 当 a 或 b 为 0 时, 第 I 类错误率超出 Bradley (1978)的合理变化范围上限 0.075 的情况全部发生在偏差校正的非参数百分位 Bootstrap 方法中; 第 I 类错误率低于 Bradley (1978)的合理变化范围下限 0.025 的情况全部发生在 MCMC 方法中, 特别是在有先验信息的 MCMC 方法中(见表 3)。这表明, 在 a 或 b 为 0 时, 偏差校正的非参数百分位 Bootstrap 方法会高估第 I 类错误率, 而有先验信息的 MCMC 方法会低估第 I 类错误率。

4.5 中介效应分析的区间宽度

表 3 中的 width 表示区间宽度。5×5×2 的方差分析用来分析 5 种方法, 5 种样本量和 2 种中介效应情况(中介效应为 0 和不为 0)对中介效应分析的区间宽度的影响。方差分析显示, 中介效应是否为 0 的主效应显著, $F(1, 100)=11.563$, $p=0.001$, $\eta_p^2=0.104$, 中介效应为 0 的区间宽度(0.179)显著小于中介效应不为 0 的区间宽度(0.239)。样本量和方法的交互作用显著, $F(16, 100)=1.859$, $p=0.033$, $\eta_p^2=0.229$ 。简单主效应分析显示, 只有样本量为 25 时, 方法的主效应显著, $F(4, 100)=11.458$, $p<0.001$, 多重比较发现, 有先验信息的 MCMC 方法估计的区间宽度(0.162)显著小于其他四种方法, 其他四种方法之间差异不显著; 有先验信息的 MCMC 方法条件下, 样本量的主效应不显著, $F(4, 100)=0.923$, $p=0.454$, 其余方法条件下, 都有样本量的主效应显著, 多重比较发现, 样本量为 25 的区间宽度显著大于其他样本量的区间宽度, 样本量为 50 的区间宽度显著大于样本量为 1000 的区间宽度。

5 综合讨论

5.1 三类方法的中介效应分析比较

中介效应点估计分析发现, 中介效应是否为 0 的因素不影响中介效应点估计的偏差和相对均方误, 样本量和方法两个因素对中介效应点估计偏差和相对均方误是有影响的, 样本量和方法之间存在交互作用。具体说来, 就偏差而言, 小样本($n=25$)条件下, 有先验信息的 MCMC 方法的偏差绝对值显著小于乘积分布法和非参数 Bootstrap 方法, 并且有先验信息的 MCMC 方法的偏差绝对值随样本量变化较小, 而乘积分布法和非参数 Bootstrap 方法的偏差绝对值随样本量增大而减小; 就相对均方误而言, 有先验信息的 MCMC 方法的相对均方误在所有样本量条件下都显著小于其他方法, 且有先验信息的 MCMC 方法的相对均方误会随着样本量的增大而增大。综上可知, 有先验信息的 MCMC 方法在中介效应的点估计上表现最优, 尤其是在小样本($n=25$)条件下, 有先验信息的 MCMC 方法的中介效应点估计的偏差绝对值最小, 相对均方误最小, 此时的中介效应点估计最准确。

中介效应区间估计分析也发现, 有先验信息的 MCMC 方法估计的区间宽度始终小于其他四种方法, 与 Yuan 和 MacKinnon (2009)实例分析的结果一致, 表明有先验信息的 MCMC 方法在中介效应的区间估计上的确是表现最优的, 尤其是在小样本($n=25$)条件下, 有先验信息的 MCMC 方法的区间宽度显著窄于其他四种方法。

有先验信息的 MCMC 方法在小样本条件优势显著, 可获得最准确的点估计和区间估计, 这与先验信息的作用是密切相关的。因为在小样本条件下, 从样本数据中获得的信息较少, 不足以得到准确的中介效应点估计和区间估计, 因此, 乘积分布法和非参数 Bootstrap 方法得到的中介效应点估计值 $\hat{a}\hat{b}$ 都产生了相对较大的偏差和相对均方误, 且区间估计的宽度较宽。有先验信息的 MCMC 方法由于加入了先验信息, 增加了有用的数据信息, 弥补了小样本数据获取信息不足的缺陷, 得到了更为准确的中介效应点估计和区间估计。所以, 有先验信息的 MCMC 方法更适合应用到样本量较小的特殊被试的心理学研究中, 例如, 脑损伤病人的认知研究, 聋哑、智障等特殊儿童的教育研究等。

随着样本量的增加, 从样本数据中获得的信息越来越多, 特别是大样本($n=1000$)情况下, 样本数

据提供的信息已可以基本满足中介效应点估计和区间估计分析的需要,因此先验信息的作用就不如在小样本条件下那么显著了,此时有先验信息的MCMC方法和其他方法获得的中介效应点估计和区间估计越来越接近(见表2和表3),但先验信息仍然起着一定的作用,表现在有先验信息的MCMC方法得到的中介效应点估计的相对均方误、估计的区间宽度仍然小于其他方法。无先验信息的MCMC方法由于没有先验信息,和乘积分布法、非参数Bootstrap方法的中介效应点估计和区间估计表现相当。

先验信息对MCMC方法的意义除了体现在中介效应点估计和区间估计的准确性之外,还表现在统计功效上,本研究发现,有先验信息的MCMC方法由于加入了先验信息,获得了最高的统计功效,偏差校正的非参数百分位Bootstrap方法的统计功效仅次于有先验信息的MCMC方法。

综上所述,当有先验信息时,有先验信息的MCMC方法的点估计和区间估计表现都最优,并且统计功效也最高,推荐使用有先验信息的MCMC方法进行中介效应分析。当先验信息不可得时,偏差校正的非参数百分位Bootstrap方法的统计功效最高,点估计和区间估计都与无先验信息的MCMC方法和乘积分布法相当,推荐使用偏差校正的非参数百分位Bootstrap方法进行中介效应分析。

值得一提的是,有先验信息的MCMC方法在使用时还需注意两点。首先,先验信息的选择必须谨慎,如果先验信息不合适,MCMC方法得到的点估计和区间估计将会产生较大偏差。目前获得先验信息的方法有两个,一是来自于文献中记录的相关的前人研究,二是来自于导航研究。那么,获得先验信息后,如何检验先验信息是否合适呢?一个可行的办法是用多种方法(如无先验信息和有先验信息的MCMC方法)同时进行中介效应分析(Mplus6程序见附录),如果有先验信息的MCMC方法估计的区间宽度大于其他方法,表明先验信息选取的不合适。其次,有先验信息的MCMC方法会低估第I类错误率。但有研究者认为,相对于第I类错误率而言,统计功效的表现更为重要,更具有实际意义,因此,有先验信息的MCMC方法为了得到更高的统计功效而付出低估第I类错误率的代价是值得的,同理,偏差校正的非参数百分位Bootstrap方法为了得到更高的统计功效而付出高估第I类

错误率的代价(表3有四次高估第I类错误率)也是值得的(Taylor, MacKinnon, & Tein, 2008; Williams & Mackinnon, 2008)。Cheung和Lau(2008)明确指出,可以用偏差校正的非参数百分位Bootstrap方法和非参数百分位Bootstrap方法同时进行中介效应分析(Mplus6程序见附录),利用非参数百分位Bootstrap方法良好的第I类错误率表现来控制偏差校正的非参数百分位Bootstrap方法可能高估第I类错误率的问题,Cheung将这个方称为交叉验证(cross validation)。同理,交叉验证法也可以用于控制有先验信息的MCMC方法可能低估第I类错误率的问题。

5.2 局限与展望

虽然本模拟研究同时比较了乘积分布法、非参数Bootstrap法和MCMC方法在中介效应分析点估计和区间估计中的表现,但仍然存在一些不足。

首先,本研究涉及的中介效应又被称为简单中介效应(simple mediation),当前中介效应已经发展到多重中介效应(multiple mediation)和多层中介效应(multilevel mediation)(方杰,张敏强,邱皓政,2010)。那么,三类中介效应分析方法在多重和多层中介效应分析中的表现又如何呢?目前已有的多重和多层中介效应分析的模拟研究都仅限于传统法、乘积分布法、Bootstrap方法的比较(Pituch, Stapleton, & Kang, 2006; Williams & Mackinnon, 2008; Taylor, MacKinnon, & Tein, 2008; Pituch & Stapleton, 2008),极少涉及MCMC方法,仅有Yuan等人(2009)比较了斜率随机的1-1-1多层中介效应中MCMC方法和Bauer法(一种参数Bootstrap方法, Bauer, Preacher, & Gil, 2006)的表现。但MCMC方法可以将先验信息整合入中介效应分析中,在小样本条件下优势尤为明显,还可以对缺失数据进行有效处理,且可用Mplus6软件方便求得MCMC可靠区间,计算量又显著小于Bootstrap方法(黎光明,张敏强,2009; Muthén & Muthén, 2010; Muthén, 2010; 沐守宽,周伟,2011),因此,用相同的模拟数据同时比较MCMC方法、乘积分布法和Bootstrap方法在多重和多层中介效应研究中的表现具有较大的理论和实用价值。

第二,本研究在模拟数据时,将所有变量都设定为正态分布的连续变量,但实际应用中,变量的分布往往偏离正态分布,并且常需要将自变量(例如,性别)、因变量(例如,升学与否)设置为二分变量。Mackinnon等人(1995)用模拟研究比较了自变

量为二分变量和连续变量的中介效应分析, 结果发现二者在中介效应点估计上无差异, 中介效应点估计的标准误在自变量为二分变量时略大于自变量为连续变量条件 (Mackinnon, Warsi, & Dwyer, 1995)。因此, 本研究的结果可以推广到自变量为二分变量的中介效应分析中。Biesanz 等人(2010)对比研究了正态和非正态的数据分布对简单中介效应分析的影响, Pituch 和 Stapleton (2008)对比研究了正态和非正态的数据分布对多层中介效应分析的影响, 结果一致发现, 非正态分布的 X 、 M 和 Y 数据使得比较指标的表现差于正态分布时的表现, 但不会对中介效应方法比较的结果产生影响。值得注意的是, Biesanz 等人和 Pituch 等人的两个模拟研究都只涉及乘积分布法和 Bootstrap 方法, 并未涉及本文所述的 MCMC 方法, 所以, MCMC 方法和其他方法在变量为非正态分布的中介效应分析中的表现尚待进一步深入研究。

第三, 大多中介效应的模拟研究(包括本研究)都设定所有变量为外显变量(manifest variable), 有研究指出, 外显变量的中介效应研究由于忽略了测量误差(measurement error)会低估中介效应值, 通过设置潜变量(latent variable)和多指标(multiple indicator)的方法可以得到更准确的中介效应值 (cheung & Lau, 2008; Ledgerwood & Shout, 2011; Mackinnon, Coxe & Baraldi, 2012)。Cheung 和 Lau (2008)模拟比较了传统法和四种非参数 Bootstrap 方法在含有潜变量的简单中介效应分析中的表现, 发现偏差校正的非参数百分位 Bootstrap 方法表现最优, 和 Mackinnon (2004)的研究结果一致, 但 Cheung 等人的研究并未涉及本文所述的 MCMC 方法, 所以, MCMC 方法和其他方法在含有潜变量的中介效应分析中的表现尚待进一步深入研究。需要说明的是, 在实际应用中到底选择外显变量还是潜变量进行中介效应分析, 其实是一个权衡的过程。Ledgerwood 和 Shout(2011)用相同的模拟数据, 同时使用潜变量模型和不含潜变量的模型进行简单中介效应, 发现含潜变量的中介效应分析得到更准确的中介效应值是以统计功效降低、标准误 $\hat{\sigma}_{ab}$ 估计增大, 且需要大样本为代价的。所以, Ledgerwood 等人建议在实际数据的简单中介效应分析中采用两步分析法。第一步, 用外显变量进行中介效应分析, 得到中介效应分析的显著性; 第二步, 用潜变量模型进行中介效应分析, 得到中介效应的点估计值。

第四, 本研究涉及的研究方法也有待拓展。例如, 本研究只涉及了非参数 Bootstrap 方法。MacKinnon 等人(2004)明确提出应进一步比较非参数和参数 Bootstrap 方法在中介效应分析中的表现。实际上, Pituch 等人(2006, 2008)已将参数 Bootstrap 方法引入多层中介效应分析中, 参数 Bootstrap 方法的基本思想是, 利用原样本数据计算参数 ab 和残差方差的估计值, 并假设残差为正态分布, 对残差进行 Bootstrap 抽样。参数和非参数 Bootstrap 方法在中介效应分析中的表现有待进一步深入研究。

6 结论

(1)在中介效应点估计上, 有先验信息的 MCMC 方法表现最优, 在所有样本量条件下的相对均方误都是最小, 并且在小样本($n=25$)条件下偏差绝对值也最小, 其余方法之间无显著差异。

(2)有先验信息的 MCMC 方法的统计功效最高, 但付出了低估第 I 类错误率的代价, 偏差校正的非参数百分位 Bootstrap 方法的统计功效其次, 但付出了高估第 I 类错误率的代价。

(3)在中介效应区间估计上, 有先验信息的 MCMC 方法的区间宽度最小, 表现最优。

致谢: 感谢德克萨斯州立大学安德森癌症中心的 Yuan 副教授和亚利桑那州立大学 Mackinnon 教授在本文写作中所给予的支持和帮助!

参 考 文 献

- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.
- Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 11(2), 142–163.
- Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research*, 45(4), 661–701.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31, 144–152.
- Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, 11(2), 296–325.
- Chiou, H. J. (2007). Quantitative research on the fault line: The development of statistical methodology. *αβγ Journal of Quantitative Research (Taiwan)*, 1(1), 1–51.

- [邱皓政. (2007). 断裂时代中的量化研究: 统计方法学的兴起与未来. *αβγ 量化研究学刊(台湾)*, 1(1), 1-51.]
- Fairchild, A. J., & McQuillin, S. D. (2010). Evaluating mediation and moderation effects in school psychology: A presentation of methods and review of current practice. *Journal of School Psychology*, 48(1), 53-84.
- Fang, J., Zhang, M. Q., & Chiou, H. J. (2010). Multilevel mediation based on hierarchical linear model. *Advances in Psychological Science*, 18(8), 1329-1338.
- [方杰, 张敏强, 邱皓政. (2010). 基于阶层线性理论的多层级中介效应. *心理科学进展*, 18(8), 1329-1338.]
- Fang, J., Zhang, M. Q., & Li, X. P. (2011). Estimating confidence intervals of mediating effects by using the distribution of the product, Bootstrap and Markov chain Monte Carlo methods. *Advances in Psychological Science*, 19(5), 765-774.
- [方杰, 张敏强, 李晓鹏. (2011). 中介效应的三类区间估计方法. *心理科学进展*, 19(5), 765-774.]
- Fang, J., Zhang, M. Q., & Chiou, H. J. (2012). Mediation analysis and effect size measurement: Retrospect and prospect. *Psychological Development and Education*, 28(1), 105-111.
- [方杰, 张敏强, 邱皓政. (2012). 中介效应的检验方法和效果量测量: 回顾与展望. *心理发展与教育*, 28(1), 105-111.]
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18(3), 233-239.
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408-420.
- Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, 101(6), 1174-1188.
- Li, G. M., & Zhang, M. Q. (2009). Estimating the variability of estimated variance components for Generalizability Theory. *Acta Psychologica Sinica*, 41(9), 889-901.
- [黎光明, 张敏强. (2009). 基于概化理论的方差分量变异量估计. *心理学报*, 41(9), 889-901.]
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Earlbaum.
- MacKinnon, D. P., Cox, S., & Baraldi, A. N. (2012). Guidelines for the investigation of mediating variables in business research. *Journal of Business and Psychology*, 27(1), 1-14.
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593-614.
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 39(3), 384-389.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83-104.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99-128.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30(1), 41-62.
- Mao, S. S., Wang, J. L., & Pu, X. L. (2006). *Advanced mathematical statistics* (2nd ed.). Beijing: Higher Education Press.
- [茆诗松, 王静龙, 濮晓龙. (2006). *高等数理统计* (2nd ed.). 北京: 高等教育出版社.]
- Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). The product of two normally distributed random variables. In: W J Kennedy & R E Odeh (Eds.), *Selected table in mathematical statistic* (Vol. VII, pp.129-144). Providence, RI: American Mathematical Society.
- Mu, S. K., Zhou, W. (2011). Handling missing data: Expectation-Maximization algorithm and Markov chain Monte Carlo algorithm. *Advances in Psychological Science*, 19(7), 1083-1090.
- [沐守宽, 周伟. (2011). 缺失数据处理的期望-极大化算法与马尔可夫链蒙特卡洛方法. *心理科学进展*, 19(7), 1083-1090.]
- Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction*. Retrieved May 17, 2010, from <http://www.statmodel.com>.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.
- Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Hoboken, NJ: Wiley.
- Pituch, K. A., & Stapleton, L. M. (2008). The performance of methods to test upper-level mediation in the presence of nonnormal data. *Multivariate Behavioral Research*, 43(2), 237-267.
- Pituch, K. A., Stapleton, L. M., & Kang, J. Y. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research*, 41(3), 367-400.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879-891.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42(1), 185-227.
- Taylor, A. B., MacKinnon, D. P., & Tein, J. Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods*, 11(2), 241-269.
- Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, 43(3), 692-700.
- Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling*, 15(1), 23-51.
- Wen, Z. L., Chang, L., Hau, K. T., & Liu, H. Y. (2004). Testing and application of the mediating effects. *Acta Psychologica Sinica*, 36(5), 614-620.
- [温忠麟, 张雷, 侯杰泰, 刘红云. (2004). 中介效应检验程序及其应用. *心理学报*, 36(5), 614-620.]
- Wen, Z. L., Liu, H. Y., & Hau, K. T. (2012). *Analyses of moderating and mediating effects*. Beijing: Educational Science Publishing House.
- [温忠麟, 刘红云, 侯杰泰. (2012). *调节效应和中介效应分析*. 北京: 教育科学出版社.]

- Wood, R. E., Goodman, J. S., Beckmann, N., & Cook, A. (2008). Mediation testing in management research: A review and proposals. *Organizational Research Methods*, 11(2), 270–295.
- Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, 14(4), 301–322.

附录: MCMC 和非参数 Bootstrap 中介效应分析的 Mplus6 Syntax

```
DATA:      file is firefighters.txt;      !包含原始数据的文本文件 firefighters.txt

VARIABLE:  names=x m y;                  !原始数据的变量名
           usevariables=x m y;            !选择所需变量

ANALYSIS:  estimator = bayes;             !用贝叶斯方法进行中介效应分析, 默认采用 Markov 链进行 Gibbs 抽样
           chain=2;                       !使用两条 Markov 链
           process = 2;                   !同时进行两条 Markov 链的运算
           fbiter = 10000;                !每一条 Markov 链都迭代 10000 次, 丢弃前一半迭代值确保 Markov 链收敛到平稳分布
           !bootstrap=5000;               ! bootstrap 法抽样 5000 次

MODEL PRIORS:                               !描述参数的先验分布
           a~n (0.35,0.04);               !参数 a 的先验分布假定为正态分布, 先验均值为 0.35, 先验方差为 0.04
           b~n (0.1,0.01);               !参数 b 的先验分布假定为正态分布, 先验均值为 0.1, 先验方差为 0.01

MODEL:                                           !模型设定如下
           m on x (a);                    !m 对 x 的回归, 得到回归系数 a
           y on m (b);
           y on x;                        !y 对 m 和 x 的回归, 得到回归系数 b

MODEL constraint:                             !中介效应计算部分
           new (indirect);                 !中介效应命名为 indirect
           indirect = a*b;                 !计算中介效应大小

OUTPUT:
           tech1 tech8;                   !tech1 列出模型设定的概要, 亦即提供有哪些参数被估计
                                           !tech8 列出迭代估计的过程与相关数据, 以利检查
           !cinterval (bcbootstrap);       !输出偏差校正的非参数百分位 bootstrap 计算的中介效应结果
           !cinterval (bootstrap);        !输出非参数百分位 bootstrap 计算的中介效应结果

PLOT:      type=plot2;                     !输出五种图形判断 Markov 链是否收敛到平稳分布
```

注: 1、运行此程序可以得到有先验信息的 MCMC 方法的中介效应分析结果。

2、若要得到无先验信息的 MCMC 方法的中介效应分析结果, 只需要在此程序上去掉 MODEL PRIORS 部分的语句即可。

3、若要得到偏差校正的非参数百分位 bootstrap 计算的中介效应分析结果, 在 ANALYSIS 部分仅留下 bootstrap=5000 语句, 在 OUTPUT 部分仅留下 cinterval (bcbootstrap)语句, 去掉 MODEL PRIORS 和 PLOT 部分的语句即可。

4、若要得到非参数百分位 bootstrap 计算的中介效应分析结果, 操作和 3 相似, 唯一差异在于 OUTPUT 部分仅留下 cinterval (bootstrap) 语句。

Assessing Point and Interval Estimation for the Mediating Effect: Distribution of the Product, Nonparametric Bootstrap and Markov Chain Monte Carlo Methods

FANG Jie^{1,2}; ZHANG Min-Qiang²

(¹ College of Humanities and Communication, Guangdong University of Business Studies, Guangzhou 510320, China)

(² Research Center of Psychological Application, South China Normal University, Guangzhou 510631, China)

Abstract

Because few sampling distributions of mediating effect are normally distributed, in recent years, Classic approaches to assessing mediation (Baron & Kenny, 1986; Sobel, 1982) have been supplemented by computationally intensive methods such as nonparametric bootstrap, the distribution of the product methods, and Markov chain Monte Carlo (MCMC) methods. These approaches are suitable for medium or small sample size and do not impose the assumption of normality of the sampling distribution of mediating effects. However, little is known about how these methods perform relative to each other.

This study extends Mackinnon and colleagues' (Mackinnon, Lockwood & Williams, 2004; Yuan & Mackinnon, 2009) works by conducting a simulation using R software. This simulation examines several approaches for assessing mediation. Three factors were considered in the simulation design: (a) sample size ($N=25, 50, 100, 200, 1000$); (b) parameter combinations ($a=b=0, a=0.39, b=0, a=0, b=0.59, a=b=0.14, a=b=0.39, a=b=0.59$); (c) method for assessing mediation (distribute of the product method, nonparametric percentile Bootstrap method, bias-corrected nonparametric percentile Bootstrap method, MCMC method with informative prior and MCMC method with non-informative prior). A total of 30 treatment conditions were designed in the 3-factor simulation. 1,000 replications were run for each treatment condition. For the Bootstrap method, 1,000 bootstrap samples were drawn in each replication. For the MCMC methods, 11,000 Gibbs iterate were implemented in each replication, 10,000 posterior samples of the model parameters were recorded after 1,000 burn-in iterations. The methods were compared in terms of (a) Bias (absolute of bias), (b) Relative mean square error, (c) Type I error, (d) Power, (e) Interval width.

The simulation study found the following results: 1) the performance of MCMC method with informative prior were superior to that of the other methods for Relative mean square error and Bias. 2) The Power of the MCMC method with informative prior was greatest among all the methods. However, extra power comes at the cost of underestimation of Type I error. Power of bias-corrected nonparametric percentile Bootstrap method was the second greatest, with elevated Type I error in some conditions. 3) Interval width of MCMC method with informative prior is smallest among different methods.

The simulation results indicated that 1) when informative prior was available, MCMC method with informative prior was recommended to analyze mediation. 2) If informative prior was not available, bias-corrected nonparametric percentile Bootstrap method should be adopted to analyze mediation. We also provide Mplus6 syntax to facilitate the implementation of the recommended bootstrapping and MCMC methods.

Key words mediation; Distribute of the product method; Nonparametric Bootstrap method; Markov chain Monte Carlo (MCMC) methods; prior information