

doi:10.19920/j.cnki.jmsc.2026.04.005

# 考虑偏差修正的 Bootstrap 假设检验及大规模关键 质量因子筛选研究<sup>①</sup>

刘丽君<sup>1,2</sup>, 欧阳林寒<sup>3</sup>, 马义中<sup>2</sup>, 吴锋<sup>1</sup>

(1. 安徽工程大学经济与管理学院, 芜湖 241000; 2. 南京理工大学经济管理学院, 南京 210094;

3. 南京航空航天大学经济与管理学院, 南京 210016)

**摘要:** 随着顾客需求的不断提高, 产品已逐渐转为定制化、个性化生产模式, 加之生产过程复杂化所导致的大规模因子, 给现有的质量改进技术带来了巨大挑战. 本文针对复杂化、定制化生产过程中因子数目大、样本量小的问题, 提出了基于重抽样技术的关键因子筛选方法, 为实现持续性质量改进奠定基础. 首先, 削弱了序贯分支因子筛选方法基于特定分布类型的模型假设, 以适用小样本情形; 其次, 提出了不依赖于响应分布的序贯分支筛选步骤; 然后, 在利用自举法扩大样本量的基础上, 提出了三种改进的学生 t 假设检验过程, 用于每个分支过程中因子组效应的显著性检验; 最后, 通过蒙特卡洛仿真试验分别比较了所提的假设检验方法与经典的学生 t 检验的检验结果, 并验证了所提方法在小样本、大规模关键因子筛选问题中的有效性及稳健性.

**关键词:** 小样本; 大规模; 因子筛选; 序贯分支; 自举法; 假设检验

**中图分类号:** F273.2 **文献标识码:** A **文章编号:** 1007-9807(2026)04-0085-19

## 0 引言

随着全球化经济水平的提升与科学技术的飞速发展, 顾客对于产品的需求越来越高, 为了更好地满足顾客需求, 大批量生产过程逐渐转为定制化、个性化<sup>[1-3]</sup>. 传统的基于大批量生产的质量工程理论与方法显然难以应对新时代的客户需求<sup>[4]</sup>. 尽管随着大数据时代的到来, 通过物联网以及传感器设备可以收集到海量数据, 但数据往往呈现出较高的变异性、异质性, 针对同类产品或过程的数据仍然属于小样本<sup>[5-7]</sup>. 例如, 增材制造(也称 3D 打印)可以根据客户需求提供定制化的产品/服务, 尽管在生产过程中可以通过传感器等设备获取大量的制造过程数据, 但是产品的批量一般较小<sup>[8]</sup>, 由此带来的小样本问题给传统过程建模技术带来新的机遇和挑战. 近几年, 针对 3D

打印制造过程的小样本建模问题, 一种通常的思路是采用迁移学习方法, 利用并学习类似产品之间的信息, 实现小批量产品的在线质量控制<sup>[9]</sup>. 然而, 现有的离线质量设计以及关键因子筛选研究大多面向大批量生产过程, 鲜有讨论小样本的问题.

事实上, 关键因子筛选被认为是响应建模的第一步, 也是过程改进的基础工作, 在离线设计过程中通过少量样本尽可能地识别出显著影响系统响应的关键因子, 不仅有利于系统最优参数搭配确定, 而且可以尽量降低在线质量控制过程的波动, 提高产品性能<sup>[10-12]</sup>. 一些经典的筛选设计方法包括部分因子设计、超饱和设计、确定性筛选设计(definitive screening design)等. 然而, 随着系统的复杂程度越来越高, 系统中所涉及的变量(因子)数目越来越大, 经典的因子筛选方法一般适用于仅包含十几个变量的小规模问题.

① 收稿日期: 2021-12-15; 修订日期: 2024-06-14.

基金项目: 国家自然科学基金资助项目(72501004; 72471112; 71931006); 安徽省哲学社会科学规划项目(AHSKQ2024D028).

作者简介: 刘丽君(1991—), 女, 山东鄄城人, 博士, 讲师. Email: lijunliu@ahpu.edu.cn

近几年,仿真试验的盛行对因子筛选的效率提出了更高的要求<sup>[13]</sup>,因此 Bettonvil 提出了适用于大规模因子的序贯分支(sequential bifurcation, SB)因子筛选方法,已被广泛用于供应链、物流等大规模问题的关键因子识别<sup>[14, 15]</sup>. 与其他的因子筛选方法相比,SB 方法具有良好的序贯性及高效性,但是,经典的 SB 方法需要较为理想的模型假设(二阶多项式模型、遗传效应、效应符号已知、正态响应)<sup>[16]</sup>. 因此,SB 的相关研究主要侧重于试图削弱假设条件,提高其在实际问题中的实用性. 例如,根据部分因子试验与 SB 之间的补充关系<sup>[17]</sup>,Shen 等<sup>[18]</sup>提出一种两阶段混合因子筛选过程,用于因子效应符号未知的情形;Shi 等<sup>[19]</sup>通过结合“预分支”过程解决了 SB 在多目标因子筛选中的效应符号冲突的问题;针对变方差情形,Wan 等<sup>[20, 21]</sup>提出了可控的序贯分支(controlled sequential bifurcation)方法,通过选择合适的样本量同时控制两类统计错误;刘丽君等<sup>[22]</sup>分析了三类数据污染对经典序贯分支因子筛选可能造成的影响,并通过构造统计量改进假设检验过程,提出了稳健的序贯分支(robust sequential bifurcation)法<sup>[23]</sup>. 除此之外,Ankenman 等<sup>[24]</sup>指出因子散度效应的重要性,并提出了二阶准度的停止准则以及置信区间,在此基础上,刘丽君等<sup>[25-27]</sup>提出了同时考虑位置和散度效应的因子分类筛选方法,并通过借鉴多响应问题中的“预分支”过程,进一步提高了筛选效率.

上述研究成果促进了 SB 方法在各类应用场景中的发展,使其成为解决大规模因子筛选问题的主要研究框架. 然而,针对本文所探讨的小样本问题,经典的 SB 因子筛选法仍面临两个方面难题:首先,难以通过定性或定量方法验证小样本的分布类型,导致经典 SB 方法中响应服从正态分布的假设不合理;其次,针对未知分布类型的响应,难以获取因子的效应估计并构建相应的假设检验过程. 因此,如何利用小样本推断总体特征、设计有效的因子效应显著性检验过程将是本文要解决的关键问题.

事实上,自举法(Bootstrap)方法历史悠久,是有效解决总体分布类型未知以及小样本问题的主流方法,一直受到统计学界的关注,并且广泛用于化学、医学及工程等领域. 其核心思想是利用重抽

样扩充原始样本量,并对总体分布特征进行统计推断. 近些年,随着应用场景的复杂化,Bootstrap 的相关理论及应用研究蓬勃发展. 例如,Spokoyny 和 Zhilova<sup>[28]</sup>在样本量有限、参数不断增加、且模型存在较小偏差的情形下,验证了 Bootstrap 的有效性;张蕊晓和田茂再<sup>[29]</sup>考虑到对于经典模型 Bootstrap 置信区间的渐近理论难以推导的情况,提出了基于贝叶斯的 Bootstrap 置信区间构造方法. 因此,本文采用 Bootstrap 方法对原始样本进行重复抽样,并且结合学生 t 假设检验方法判断每个“分支”过程中组因子的显著性,最终实现小样本情形下大规模关键因子筛选. 尤其是针对本文的关键问题——小样本下两个分布类型未知总体的显著性检验,本文在经典 Bootstrap 假设检验的基础上分别考虑了偏差修正以及统计量分布的不确定性,从而提出两种改进的 Bootstrap 假设检验过程,并且通过仿真算例比较三种 Bootstrap 方法在显著性检验以及因子筛选问题中的差异性.

综上所述,本文旨在解决小样本情形下的关键因子筛选问题,以 SB 方法为作为因子筛选实施框架,将 Bootstrap 方法与学生 t 检验相结合并考虑偏差修正以及统计量分布的不确定性,提出三种因子组效应的显著性检验方法,并将其融入 SB 实施过程,从而实现小样本情形下大规模关键质量因子筛选. 要完成两个方面的工作:1) 分别从偏差修正以及统计量分布不确定两个角度考虑,改进经典 Bootstrap 的假设检验方法;2) 将上述针对两个总体显著性检验的 Bootstrap 假设检验方法融入 SB 因子筛选过程,提出适用于小样本情形的关键因子筛选实施框架. 具体实施步骤为:首先构建了分布类型未知的响应模型;其次在每个“分支”步骤中借助 Bootstrap 扩大样本量;然后,利用 Bootstrap 样本所得的估计量分别构建了三种假设检验过程,即基于经典 Bootstrap 的假设检验(basic Bootstrap for significance testing, BBT)、基于偏差修正 Bootstrap 的假设检验(adjusted bias based Bootstrap for significance testing, ABT)、以及基于稳健统计量的 Bootstrap 假设检验(robust statistics based Bootstrap for significance testing, RBT);最后通过蒙特卡洛仿真比较了所提的三种假设检验方法与经典学生 t 检验,以及其在因子筛选问题中的有效性及稳健性.

# 1 小样本情形下序贯分支关键因子筛选

## 1.1 模型假设及效应估计

为了方便描述本文所提的非参数假设检验方法,假设系统输出可由如下线性模型表述

$$y = \beta_0 + \sum_{h=1}^k \beta_h x_h + \varepsilon \quad (1)$$

其中  $y$  表示系统输出,  $\beta_h (\geq 0)$  表示编码后可控因子  $x_h (= 0, 1)$  的效应系数, 可控因子数目为  $k$ . 事实上, 线性模型包含了所有因子的主效应项, 可以基本描述响应对于因子水平变化的基本走势, 因此如果试验成本有限、且需要将更多的成本用于关于散度效应或者非正态性的探索时, 只包含主效应的线性模型假设往往是可用的选择. 在 SB 方法的框架下, 如果假设响应模型只包含主效应, 那么相应的试验成本往往会成倍减少, 降低模型复杂度以减少试验成本是筛选阶段采取的常规思路. 在随机仿真系统中,  $\varepsilon$  表示随机项, 经典的 SB 方法一般假设  $\varepsilon$  服从正态分布. 尽管正态分布的假设使得因子效应的显著性检验容易进行, 但存在一定的局限性. 显然, 错误的响应分布假设势必会增加因子效应显著性检验过程中的统计错误, 继而影响因子筛选结果. 因此, 为了保证因子筛选结果的可靠性, 需要通过定性(例如, 频率分布直方图、QQ 图)或定量(例如, KS 检验、SW 检验)的方法验证响应数据的正态性. 然而, 正态性检验对样本量具有一定的要求(一般不少于 30), 可能会造成额外的试验成本, 尤其是在小样本情形下, 正态性检验往往难以进行. 因此, 本文适当

削弱了序贯分支因子筛选模型的假设条件, 不假设响应的分布类型(distribution-free), 即假设随机项  $\varepsilon$  服从某均值为 0, 方差为  $\sigma^2$  的分布  $F(x)$ .

因此, 根据式(1)中的一阶模型假设, 只需要两个设计点即可以估计出一组因子的效应(称为因子的组效应). 例如, 假设分别在设计点  $x(k_0 - 1)$  和  $x(k_1)$ , 收集两组样本量为  $n$  的响应值  $y_i(k_0 - 1)$ 、 $y_i(k_1)$ ,  $i = 1, \dots, n$ ,  $\hat{y}(k_0 - 1)$ 、 $\hat{y}(k_1)$  分别表示这两组响应值的位置效应无偏估计, 那么因子的组效应估计可以表示为

$$\hat{\beta}_{k_0 \sim k_1} \triangleq \hat{y}(k_1) - \hat{y}(k_0 - 1) = \sum_{h=k_0}^{k_1} \beta_h \quad (2)$$

## 1.2 改进的序贯分支筛选过程

随机仿真下的 SB 因子筛选在每个“分支”阶段均要进行因子的组效应的显著性检验. 由于经典 SB 方法假设响应服从正态分布, 且正态分布具有良好的可加性质, 即两个正态分布相减(加)仍然服从正态分布, 因此, 可以采用成对的假设检验(例如, 成对的学生 t 检验). 然而, 正态响应的假设并不适用于小样本情形, 对于分布类型未知的响应模型(如式(2)所示), 因子筛选的难点根源正是因子效应的显著性, 主要体现在如下两个方面: 1) 未知的分布类型不一定具有可加性质, 无法使用成对检验的思想; 2) 现行的假设检验方法依赖于分布假设, 不适用于分布类型未知的两个总体检验.

因此, 针对小样本情形下的因子筛选问题, 本文仍然采用 SB 方法的基本步骤以保证筛选的高效性, 在进一步削弱了响应模型的假设的基础上, 提出并比较了三种非参数的假设检验方法, 用于因子效应的显著性检验. 所提出的因子筛选方法不依赖于响应分布的假设, 其具体步骤如图 1 所示.

- 
- 步骤 1 将所有因子聚合为一个组  $\{x_1, \dots, x_k\}$ , 初始化当前组 G, FIFO 队列为空.
- 步骤 2 收集当前组因子在设计点  $x(k_0 - 1)$  和  $x(k_1)$  处的响应值.
- 步骤 3 采用基于 Bootstrap 的假设检验方法(BBT/ABT/RBT)检验当前因子组效应的显著性: 如果当前组因子显著转到步骤 4; 反之, 如果当前组因子不显著转到步骤 5.
- 步骤 4 将当前组分为两个组, 并放入 FIFO 队列.
- 步骤 5 删除当前组.
- 步骤 6 如果 FIFO 非空, 则从 FIFO 队列中取出一个组, 设置为当前组 G, 并转到步骤 2. 如果 FIFO 为空, 那么终止过程.
- 

图 1 小样本情形下序贯分支因子筛选步骤

Fig. 1 The procedure of sequential bifurcation factor screening method with small sample size

- 注: 1) 三种基于 Bootstrap 的假设检验法分别为: BBT——基于经典 Bootstrap 的假设检验; ABT——基于偏差修正 Bootstrap 的假设检验; RBT——基于稳健统计量的 Bootstrap 假设检验;
- 2) FIFO 表示先入先出(First in first out).

## 2 基于 Bootstrap 的因子效应显著性检验

本文所探讨的小样本情形下的因子筛选问题,其研究重点和难点在于因子的效应显著性检验,而且,根据响应模型假设以及序贯分支方法的基本逻辑,上述问题可以进一步转化为两组样本的位置效应检验问题.由于响应的分布类型未知,且样本量较小,可以采用 Bootstrap 方法扩大样本量,然后基于 Bootstrap 样本所得的估计量进行假设检验.

Bootstrap 方法最初由 Efron 于 1992 年提出<sup>[30]</sup>,其核心思想是利用重抽样的方法扩大样本量,并且利用重抽样所得的统计量,推断初始样本的统计量,进而推断分布总体参数.根据本文的响应模型(2),假设当前组因子表示为  $\{x_{k_0}, \dots, x_{k_1}\}$ ,在两个设计点  $x(k_0 - 1)$  和  $x(k_1)$  处分别收集  $n$  个原始样本  $y_i(k_0 - 1), y_i(k_1), i = 1, \dots, n$ ,则可以通过以下步骤分别获取两组 Bootstrap 样本的参数估计量:

**步骤 1** 利用原始样本构造经验分布函数  $\hat{F}_{k_0}(x)$  和  $\hat{F}_{k_1}(x)$ ;

**步骤 2** 从经验分布函数  $\hat{F}_{k_0}(x)$  和  $\hat{F}_{k_1}(x)$  中分别抽取样本量为  $n$  的 Bootstrap 样本,并计算其均值作为位置参数估计,记为  $\hat{\mu}^{(*)}(k_0 - 1)$  和  $\hat{\mu}^{(*)}(k_1)$ ;

**步骤 3** 重复步骤(2)中的重抽样参数估计过程  $B$  次,分别得到两组基于 Bootstrap 样本的位置效应估计  $\hat{\mu}_r^{(*)}(k_0 - 1)$  和  $\hat{\mu}_r^{(*)}(k_1), r = 1, \dots, B$ .

根据 Bootstrap 的基本思想,可以利用在多次重抽样过程中所得的两组效应估计  $\hat{\mu}_r^{(*)}(k_0 - 1)$  和  $\hat{\mu}_r^{(*)}(k_1) (r = 1, \dots, B)$  分别对两个设计点下的响应分布总体进行统计推断.在因子筛选的背景下,其目的是检验当前因子组效应的显著性,即判断两个设计点下的响应分布总体是否具有明显差异,因此,可以利用上述 Bootstrap 过程所得的两组统计量进行假设检验,进而判断当前组因

子的显著性.本节分别提出三种基于 Bootstrap 样本的假设检验方法.

### 2.1 基于经典 Bootstrap 的假设检验

经典的 Bootstrap 一般假设经过重抽样所得的统计量服从以原始样本估计值为均值的正态分布,例如,  $\hat{\mu}_r^{(*)}(k_1)$  可以看作是一组服从正态分布  $N(\hat{\mu}(k_1), \sigma^2)$  的样本,其中  $\hat{\mu}(k_1)$  表示采用原始样本计算所得的估计值.如果  $\hat{\mu}(k_1)$  是响应分布总体的均值无偏估计,那么可以使用来源于两个总体  $N(\hat{\mu}(k_0 - 1), \sigma^2)$  和  $N(\hat{\mu}(k_1), \sigma^2)$  的样本  $\hat{\mu}_r^{(*)}(k_0 - 1)$  和  $\hat{\mu}_r^{(*)}(k_1)$ ,检验均值参数  $\hat{\mu}(k_0 - 1)$  与  $\hat{\mu}(k_1)$  的差异性,进而反映两个分布总体均值  $\mu(k_0 - 1)$  与  $\mu(k_1)$  的差异性.根据模型假设(2),此时两总体均值的差异即为当前因子的组效应,表示为

$$\mu(k_1) - \mu(k_0 - 1) = \sum_{h=k_0}^{k_1} \beta_h \triangleq \beta_{k_0 \sim k_1} \quad (3)$$

在正态分布假设以及 Bootstrap 样本量相同的情形下,可以针对两组样本  $\hat{\mu}^{(r)}(k_0 - 1)$  和  $\hat{\mu}^{(r)}(k_1)$  采用成对的学生 t 检验,令

$$\hat{\beta}_{k_0 \sim k_1; r} = \mu_r^{(*)}(k_1) - \mu_r^{(*)}(k_0 - 1) \quad (4)$$

则原假设与备选假设分别为

$$\mathbf{H}_0: \beta_{k_0 \sim k_1} = \Delta$$

$$\mathbf{H}_1: \beta_{k_0 \sim k_1} > \Delta$$

相应的学生 t 统计量为

$$T_{B-1} = \frac{\bar{\hat{\beta}}_{k_0 \sim k_1} - \beta_{k_0 \sim k_1}}{s(\hat{\beta}_{k_0 \sim k_1})/\sqrt{B}} \quad (5)$$

其中  $\bar{\hat{\beta}}_{k_0 \sim k_1}$  和  $s(\hat{\beta}_{k_0 \sim k_1})$  分别表示统计量  $\hat{\beta}_{k_0 \sim k_1; r} (r = 1, \dots, B)$  的均值和标准差,  $B$  表示重抽样次数,  $B - 1$  表示学生 t 分布的自由度.

### 2.2 基于偏差修正 Bootstrap 的假设检验

前述的 BBT 假设检验过程主要包含两个基本的假设:首先,假设 Bootstrap 样本的估计量服从正态分布;其次,假设原始样本的均值是响应分布位置参数的无偏估计.然而,通常情况下使用原始样本所得的估计量与响应总体的位置参数之间存在明显偏差,且此偏差可以通过 Bootstrap 样本进行估计.以设计点  $x(k_1)$  处的一组原始样本

$y_i(k_1) (i = 1, \dots, n)$  为例,原始样本均值估计与总体均值的偏差  $\hat{\mu}(k_1) - \mu(k_1)$  可以通过式(6)估计

$$B^*(k_1) = \frac{1}{R} \sum_{r=1}^B \mu_r^{(*)}(k_1) - \hat{\mu}(k_1) \\ = \bar{\mu}^{(*)}(k_1) - \hat{\mu}(k_1) \quad (6)$$

当  $\hat{\mu}(k_1) = \mu(k_1)$  且  $\hat{\mu}(k_0 - 1) = \mu(k_0 - 1)$  时,原假设  $\mu(k_1) - \mu(k_0 - 1) = \beta_{k_0-k_1} = \Delta$  与假设  $\hat{\mu}(k_1) - \hat{\mu}(k_0 - 1) = \Delta$  是等价的,因此可以采用式(5)进行成对的学生 t 检验;然而,当原始样本均值估计与总体均值的偏差存在时,即  $\hat{\mu}(k_1) - \mu(k_1) \neq 0$ ,代入式(6)对备选假设进行修正,此时

$$\beta_{k_0-k_1} = \mu(k_1) - \mu(k_0 - 1) = [\hat{\mu}(k_1) - B^*(k_1)] - [\hat{\mu}(k_0 - 1) - B^*(k_0 - 1)] \\ = [2\hat{\mu}(k_1) - \bar{\mu}^{(*)}(k_1)] - [2\hat{\mu}(k_0 - 1) - \bar{\mu}^{(*)}(k_0 - 1)] \\ = 2[\hat{\mu}(k_1) - \hat{\mu}(k_0 - 1)] - [\bar{\mu}^{(*)}(k_1) - \bar{\mu}^{(*)}(k_0 - 1)] \quad (7)$$

因此,偏差修正后的原假设和备选假设为

$$\mathbf{H}_0: \hat{\mu}_{k_0-k_1} = \Delta'$$

$$\mathbf{H}_1: \hat{\mu}_{k_0-k_1} > \Delta'$$

其中  $\hat{\mu}_{k_0-k_1} \triangleq \hat{\mu}(k_1) - \hat{\mu}(k_0 - 1)$ ,  $\Delta' = [\Delta + \bar{\mu}^{(*)}(k_1) - \bar{\mu}^{(*)}(k_0 - 1)]/2$  表示修正后的重要性阈值. 相应的学生 t 统计量为

$$T_{B-1} = \frac{\bar{\hat{\beta}}_{k_0-k_1} - \hat{\mu}_{k_0-k_1}}{s(\hat{\beta}_{k_0-k_1})/\sqrt{B}} \quad (8)$$

其中  $\bar{\hat{\beta}}_{k_0-k_1}$  和  $s(\hat{\beta}_{k_0-k_1})$  分别表示统计量  $\hat{\beta}_{k_0-k_1; r} (r = 1, \dots, B)$  的均值和标准差,  $B$  表示重抽样次数,  $B - 1$  表示学生 t 分布的自由度.

### 2.3 基于稳健统计量的 Bootstrap 假设检验

如前所述, BBT 与 ABT 假设检验过程均假设 Bootstrap 的样本所得统计量服从正态分布, 为了进一步削弱其假设条件, 本节基于 Bootstrap 样本所得的估计量, 采用稳健统计量的思想对学生 t 统计量进行改进, 提高了 Bootstrap 统计量即使不严格服从正态分布的适用性. 此外, 由于原始样本估计量与总体位置参数之间的偏差, 本文首先基于式(6), 对基于 Bootstrap 样本的统计量进行修

正, 例如, 原始样本  $y_i(k_1) (i = 1, \dots, n)$  的 Bootstrap 统计量偏差修正表示为

$$\tilde{\mu}_r^{(*)}(k_1) = \mu_r^{(*)}(k_1) - B^*(k_1) \quad (9)$$

类似地, 针对原始样本  $y_i(k_0 - 1) (i = 1, \dots, n)$  的 Bootstrap 统计量偏差修正表示为

$$\tilde{\mu}_r^{(*)}(k_0 - 1) = \mu_r^{(*)}(k_0 - 1) - B^*(k_0) \quad (10)$$

然后, 将这两组修正后的统计量(式(9)、式(10))成对相减, 并采用稳健统计量中位数和中位数绝对偏差改进经典的学生 t 统计量, 以提高其在非正态情形下的适用性. Park 等<sup>[31]</sup>指出公式均值和中位数绝对偏差统计量能够适用于分布呈现明显厚尾情形的参数估计, 因此具有良好的稳健性. 基于修正样本改进的稳健学生 t 统计量, 如式(11)所示

$$T_R = \frac{\tilde{\beta}_{k_0-k_1} - \beta_{k_0-k_1}}{\frac{\text{median}_{1 \leq r \leq n} |\tilde{\beta}_{k_0-k_1; r} - \tilde{\beta}_{k_0-k_1}|}{\sqrt{B}}} \quad (11)$$

其中

$$\tilde{\beta}_{k_0-k_1; r} = \tilde{\mu}_r^{(*)}(k_1) - \tilde{\mu}_r^{(*)}(k_0 - 1) \quad (12)$$

且

$$\tilde{\beta}_{k_0-k_1} = \text{median}_{1 \leq r \leq n} [\tilde{\beta}_{k_0-k_1; r}] \quad (13)$$

Liu 等<sup>[23]</sup>分别通过定性和定量的方法说明了公式(11)与一个常数项  $C$  相乘后是渐近服从标准正态分布的, 即

$$C \times T_R \xrightarrow{d} N(0, 1) \quad (14)$$

其中  $C = \sqrt{\frac{2}{\pi}} \Phi^{-1}\left(\frac{3}{4}\right)$ , 关于式(14)的证明可以参考 Liu 等<sup>[23]</sup>. 因此, 与 BBT 及 ABT 方法类似, 可以计算统计量所对应的参考分布分位点及  $p$  值, 以判断两个总体的显著性.

## 3 仿真试验

为了验证本文所提的三种基于重抽样假设检验方法(BBT、ABT、RBT)在假设检验以及因子筛选中的有效性, 采用蒙特卡洛仿真试验的方法, 在假设了响应分布类型、参数以及因子效应的基础

上生成响应数据观测值,然后仅利用响应数据对分布总体或者响应模型参数进行统计推断,并与假设中的真实参数或者响应模型参数进行比较,以说明统计推断的有效性.首先,3.1节假设了几类常见分布类型下两个总体的假设检验问题,分析并比较了所提的三种假设检验方法与经典学生t检验(简称为TT)的假设检验结果.然后,在3.2节中分别通过小规模以及大规模因子筛选试验说明了所提假设检验方法在序贯分支因子筛选中的有效性以及稳健性.

### 3.1 两个总体的假设检验算例

为了说明本文所提方法在小样本情形下假设检验的有效性,本文分别考察了现有研究中(例如, Park 等<sup>[31]</sup>)常见的五种对称分布以及两种偏态分布:正态分布、均匀分布、Logistic分布、Laplace分布、学生t分布(自由度为3);偏正态分布、贝塔分布.本小节的目的是检验两个同类型分布总体位置参数的差异性,因此,需要对各种分布类型的基准分布 $Y \sim F_0(y)$ 的均值和标准差进行标准化处理,即通过调整分布参数使得 $E(Y) = \mu = 0$ ,  $std(Y) = \sigma = 5$ .例如,假设正态分布为 $N(0, 25)$ ,均匀分布为 $U(-5\sqrt{3}/6, 5\sqrt{3}/6)$ ,类似地,对Logistic分布、Laplace分布以及学生t分布做上述标准化处理,统一所有基准分布的均值和标准差.此外,对于两种偏态分布类型,分别考虑了左偏(偏度系数小于0)和右偏(偏度系数大于

0)情形,因此共形成四种分布: $SN(\xi_1, \omega_1, 5)$ (右偏)、 $SN(\xi_2, \omega_2, -5)$ (左偏)、 $Beta(1,5)$ (右偏)、 $Beta(5,1)$ (左偏).对四种偏态分布进行标准化处理,可以得到两个偏正态分布的位置和尺度参数,且满足 $\xi_1 = -\xi_2$ 、 $\omega_1 = \omega_2$ .如表1所示,针对每种分布类型,假设两种试验方案,并分别采用四种假设检验方法检验每种试验方案下的两个总体的差异性.如果用 $F_1(y_1)$ 和 $F_2(y_2)$ 分别表示两个总体的分布,假设两个总体的分布仅有位置上的差异 $d$ ,那么试验方案A中设置两个总体的分布差异 $d = 2$ ;试验方案B假设两个总体的分布差异为 $d = 4$ .例如,在正态分布假设下,方案A的两个总体的分布可分别表示为 $N(0, 25)$ 和 $N(2, 25)$ ;方案B的两个总体的分布可分别表示为 $N(0, 25)$ 和 $N(4, 25)$ .在本小节的仿真中设置不重要性阈值 $\Delta = 2$ (实际案例中一般由试验人员指定),如果 $d > \Delta$ 则希望检验结果为两个总体具有显著差异;如果 $d \leq \Delta$ 则希望检验结果认为两个总体无明显差异.根据以上试验方案的假设,在每种分布类型下,分别采用四种假设检验方法对两种试验方案进行完全独立的 $N = 1\ 000$ 次重复假设检验,并记录在 $N$ 次重复检验中检验结果为具有明显差异的频率 $f_i$ ,如表1所示.此外,设定显著性水平 $\alpha = 0.05$ ,样本量 $n = 5$ ,重抽样样本量 $B = 100$ .

表1 两个总体的差异性检验结果对比

Table 1 The comparison of significance testing results for two populations

分布类型	BBT		ABT		RBT		TT	
	方案 A	方案 B	方案 A	方案 B	方案 A	方案 B	方案 A	方案 B
正态分布/%	42.00	<b>69.50</b>	<b>34.40</b>	60.90	42.20	<b>69.40</b>	5.10	13.50
均匀分布/%	42.40	<b>69.50</b>	<b>35.30</b>	60.20	43.40	<b>68.30</b>	5.60	13.60
Logistic分布/%	42.50	<b>67.90</b>	<b>35.50</b>	61.50	42.30	<b>68.60</b>	5.10	15.30
Laplace分布/%	43.20	<b>66.40</b>	<b>36.20</b>	59.60	41.80	<b>67.90</b>	4.60	14.70
学生t分布/%	43.70	<b>70.80</b>	<b>37.70</b>	65.40	43.90	<b>72.20</b>	4.60	15.00
SN分布(右)/%	39.80	<b>68.50</b>	<b>32.30</b>	61.30	45.80	59.20	4.40	13.70
SN分布(左)/%	41.20	<b>68.40</b>	<b>34.40</b>	61.40	42.20	58.40	5.10	11.30
Beta分布(右)/%	42.60	<b>68.70</b>	<b>35.10</b>	62.30	46.70	64.00	4.50	15.20
Beta分布(左)/%	41.20	<b>67.80</b>	<b>35.30</b>	61.70	44.40	61.20	4.00	13.30

注: 黑体数字表示在不同的仿真方案和分布假设下,表现较好的检验方法结果,即成功识别两分布差异的概率在60%以上.

显然,在试验方案A中两个总体的差异 $d = 2$ 恰好等于不重要阈值 $\Delta = 2$ ,因此,试验方案A

的 $f_i$ 越小越好;相反,试验方案B中两个总体的差异 $d = 4$ 大于不重要阈值 $\Delta = 2$ ,因此,试验方案B的 $f_i$ 越大越好.根据总体的分布类型,上述显著性检验结果可以从对称分布以及非对称分布两个方面进行归纳.

首先,针对对称分布类型的总体(前五种分布)可以得到如下结论.通过观察试验方案A的假设检验结果,可以发现采用经典的学生t检验(TT)的检验结果 $f_i$ 明显低于其它三种假设检验方法,而且错误判断两个总体具有显著差异的频率基本保持在预先设定的置信水平( $\alpha = 0.05$ )左右.但是,在试验方案B中,经典的学生t检验(TT)的检验结果 $f_i$ 仍然明显低于其它三种假设检验方法.也就是说,尽管TT方法几乎不会误判两个没有差异的总体为具有明显差异(第一类错误),它也很难检测出本身就具有明显差异的两个总体的差异性(第二类错误).例如,在正态分布的情形下,如果采用TT方法能够成功识别出试验方案B中两个总体存在明显差异的频率仅为13.5%,也就是说第二类错误的发生概率为86.5%.相比之下,如果采用三种改进的假设检验方法,在五种假设的分布类型下,其成功识别试验方案B中的两个总体之间差异性的频率基本可以保持在60%以上,尤其是BBT及RBT方法对于试验方案B中两个总体的差异识别频率基本保持在65%以上.可知,第二类错误在因子筛选中展现出比第一类错误更为重要的作用,因此,当样本量较小且由于成本受限等原因无法扩大样本量时,可以采用本文所提的三种假设检验方法,尽管不能将正确检验出两个总体差异性的概率控制在很高的概率范围内,但远远高于经典的学生t检验.当然,作为代价三种改进的假设检验方法更加容易错误检测出试验方案A中的差异性,但是,在因子筛选的背景中,将更多的因子包含在模型中可以在后续的建模过程中进行修正和改善.对比本文所提的三种假设检验方法,显然,BBT方法和RBT方法的第二类错误低于ABT;而ABT的第一类错误则低于BBT和RBT.

然而,通过观察四种偏态分布的假设检验结果可以发现,本文所提的三种基于Bootstrap的假设检验方法之间的差异性与在对称分布中所表现的略有不同.其中BBT与ABT方法仍然是分别在第二类错误和第一类错误方面具有优势.但是,RBT方法针对偏态分布类型的假设检验结果明显与其在对称分布中的假设检验结果相比较差,尤其是在第二类错误方面,在同种分布类型下明显高于BBT方法.

为了考察样本量、重抽样样本量、以及不同的分布类型对四种假设检验结果的影响情况,本文进一步设置不同的样本量( $n = 5, 10, 15, 20, 25, 30$ )以及重抽样样本量( $B = 100, 200, 300, 400, 500$ ),在共计30种样本量与重抽样样本量的组合方案,以及九种分布情形下,分别考察了四种假设检验方法的检验结果<sup>②</sup>.为了更直观地比较四种假设检验方法,采用折线图的方式对数据<sup>②</sup>进行可视化处理,并且得到如下结论.

1)数据量的增加基本不会影响TT方法在试验方案A中的检验结果,但是会显著改善仿真方案B的检验结果.图2所示为TT方法在不同样本量以及重抽样样本量情况下的方案A和方案B的假设检验结果,其中每个子图的横坐标被虚线分割为6个区间,从左至右的虚线条形区域所表示的样本量依次增加,而且,在每个小的条形区域中,横坐标的红色坐标点从左至右表示重抽样样本量依次增加.例如,图2(a)中的左侧第一个虚线区域表示样本量 $n = 5$ ,其中横坐标的5个子坐标点分别表示重抽样的样本点分别为 $B = 100, 200, 300, 400, 500$ .两个子图的纵坐标表示在1000次重复检验中认为两个总体之间存在显著差异的频率 $f_i$ .从图2(a)中可以看出,样本量的增加几乎不会影响TT方法检验结果的走势,随着样本量 $n$ 的增加,TT方法错误检测出方案A中两个总体存在明显差异的频率始终保持在置信水平 $\alpha = 0.05$ 左右,无明显的上升或下降趋势.与此相反,从图2(b)中可以发现,随着样本量 $n$ 的增加,TT方法正确检测出方案B中两个总体之间

② 详细数据略,有需要者可联系作者获取.

存在明显差异的频率呈现出明显的上升趋势.也就是说在样本量较小的情况下( $n \leq 30$ ),样本量的变化无法改善第一类统计错误,而且较少的样本量导致参数估计以及统计量误差较大,因此无法将第一类错误完全控制在给定的阈值水平内;但是第二类统计错误对于样本量的变化极其敏感,当样本量提升到 30 时,第二类错误( $1 - f_1$ )基本可以降低到 50% ~ 60%左右.除此之外,从图 2

中还可以观察到,TT 方法的第一类统计错误方面在所考察的五种分布类型基本没有差异,但是总体的分布类型显然会影响 TT 方法的第二类统计错误,例如,观察图 2(b)可以发现,当两个总体为均匀分布时,如果采用 TT 方法能够成功检测出两个总体差异的频率显然略低于其他几类分布,而当两个总体为学生 t 分布时,其被成功检测出差异性的频率则略高于其他四种分布.

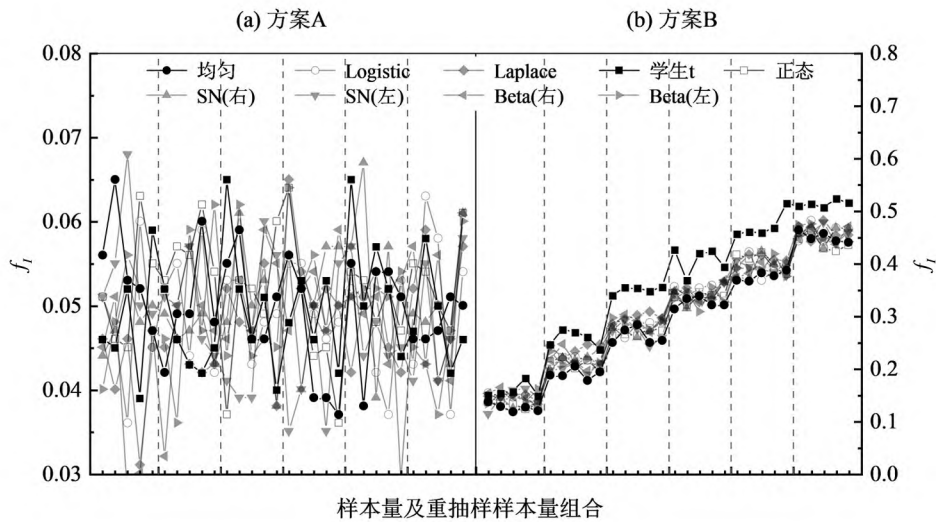


图 2 TT 方法在不同样本量及重抽样样本量下的假设检验结果

Fig. 2 The results of TT method with several combinations of sample size and the number of Bootstrap

2) 样本量对于 BBT 和 ABT 方法的影响类似,图 3 和图 4 分别是 BBT 和 ABT 方法在不同样本量及重抽样样本量下的假设检验结果.总体而言,样本量的增加会同时导致第一类统计错误的增加、第二类统计错误的降低,且对第二类统计错误影响更为明显.从两个图中看出,随着样本量的增加,方案 A 及方案 B 中的折线均呈现出上升趋势,且方案 B 中的上升趋势更为明显.另一个面,在样本量非常小的情况下,适当的增加重抽样的样本量会显著影响两类统计错误(例如,在样本量  $n = 5$  的虚线区间内),随着样本量的增加,重抽样样本量的影响逐渐趋于平缓.除此之外,通过比较图 3 和图 4 可以发现,采用 ABT 方法的第一类统计错误低于 BBT 方法,而第二类统计错误高于 BBT 方法,这与表 1 中得到的结论是一致的.随着样本量的增加,ABT 方法与 BBT 方法的检验结果越来越接近,两类统计错误的表现也更加接近.最后,观察在不同分布类型下方案 B 的假设

检验结果可以得到两点结论:一是,针对对称分布类型,BBT 和 ABT 方法能够成功检验出两个均匀分布总体的差异性的频率明显优于其它四种分布类型,通过与图 2(b)相比可知这与 TT 方法是完全相反的;二是,针对非对称分布类型,BBT 和 ABT 方法对于同种分布仅偏度系数相反情况的筛选结果差异性很大,且偏度稀疏越大筛选结果越好.

3) 样本量对于 RBT 检验方法的影响与对 BBT 和 ABT 方法的影响类似,即样本量的增加会同时影响第一类统计及第二类统计错误.从图 5 中的两个子图中可以明显看到两簇折线,因此,根据分布类型可以从两个方面(对称分布、偏态分布)对 RSB 方法的假设检验结果进行总结.首先,针对五种对称分布类型,随着样本量的变化,RBT 方法的第一类统计错误及其走势与 BBT 方法类似,尤其是在样本量较小的情形下( $n = 5$ ).然而,通过对比图 3 与图 5(b)可以发现,RBT 方法在第二类统计错误方面明显优于 BBT 及 ABT 方法.尤

其是在样本量  $15 \leq n \leq 30$  时,如果采用 BBT 或者 ABT 检验方法,除了均匀分布以外的四种分布类型随着样本量的增加其  $f_1$  的增长趋势逐渐平缓. 当样本量  $n = 30$  时,除均匀分布以外的四种分布类型的第二类统计错误基本稳定在 20% 左右. 然而,如果采用 RBT 检验方法,当两个总体为均匀分布时的假设检验结果  $f_1$  会略低于 BBT 和 ABT 方法,但是,其余四种分布的  $f_1$  基本与均匀分布类似,随着样本量的增加第二类统计错误不断降低. 除此之外,通过观测对称分布的假设检验结果发现,五种分布类型所对应的两类统计错误的折线图是互相交织在一起的,说明分布类型对 RBT 方法的检验结果无影响,具有良好的稳健

性. 另一方面,通过观察四种偏正态分布的假设检验结果可以发现,其对应的折线是与对称分布的折线分离的. 尤其是第二类错误所呈现的折线与五种对称分布第二类错误所对应的折线中间存在明显带状空白区间. 但是随着样本量的增加,折线总体仍呈现上升趋势,并且针对偏态程度的不同 RSB 方法的第二类错误差异不大,例如,与 BBT 以及 ABT 的方法相比,四种偏态分布的第二类错误所对应的折线是聚合在一起的. 同样地,四种偏态分布的第一类错误与对称分布的折线分离,但是与第二类错误相比聚合程度不高. 但是,与对称分布所呈现的走势不同,随着样本量的增加偏态分布的第一类错误呈现出略微下降的趋势.

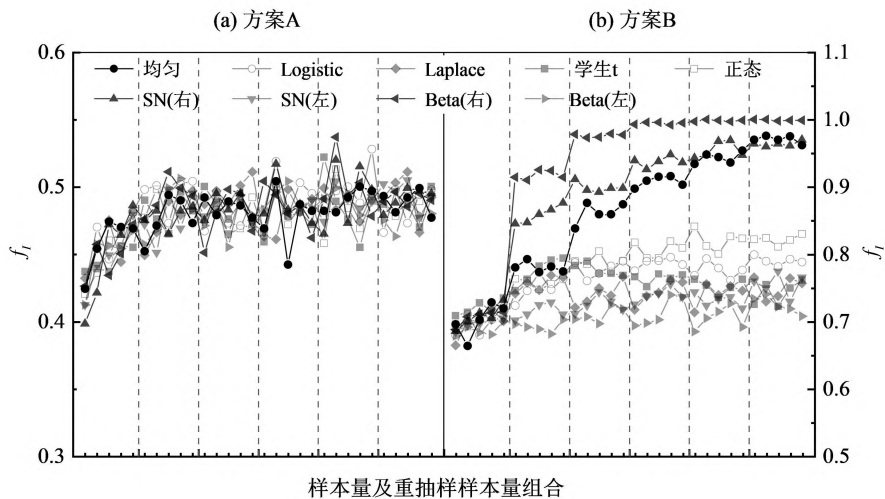


图3 BBT 方法在不同样本量及重抽样样本量下的假设检验结果

Fig. 3 The results of BBT method with several combinations of sample size and the number of Bootstrap

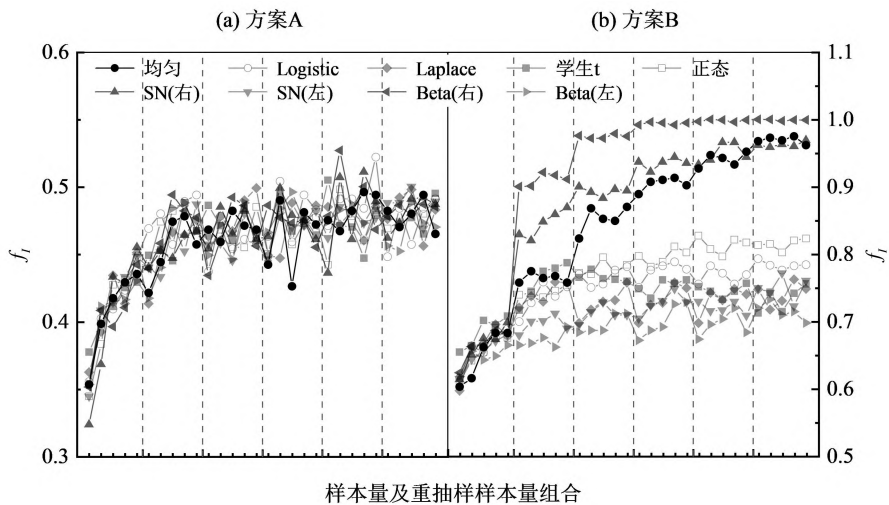


图4 ABT 方法在不同样本量及重抽样样本量下的假设检验结果

Fig. 4 The results of ABT method with several combinations of sample size and the number of Bootstrap

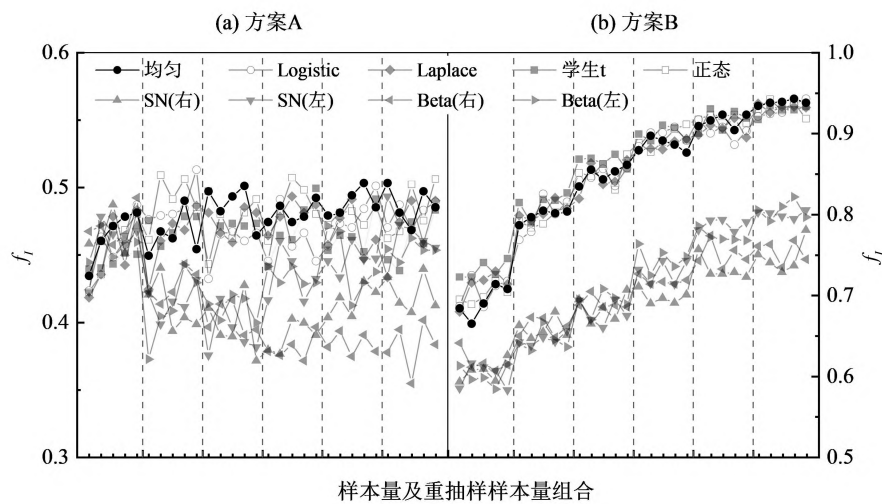


图5 RBT方法在不同样本量及重抽样样本量下的假设检验结果

Fig. 5 The results of RBT method with several combinations of sample size and the number of Bootstrap

综上所述,在本文所考察的小样本情形下,建议采用RBT方法检验两个总体的差异性,并且在无法提高样本量时可以通过提高重抽样样本量以降低第二类统计错误,尽管RSB在偏态分布类型中的第二类错误高于对称分布类型,但是其在不同偏态程度的情况下具有较好的稳健性,因此仍让可以作为一种保守的方案.但是,在实际的案例研究中,如果可以通过经验判断分布为右偏的情形,为获取更低的第二类错误,则可以采用ABT或者BBT的方法.

### 3.2 因子筛选的算例

为了进一步验证本文所提的三种假设检验方法在因子筛选过程中有效性,将4.1节中所比较的四种假设检验用于序贯分支因子筛选中每个“分支”过程的组效应显著性检验,分别记为SB-BBT,SB-ABT,SB-RBT,以及SB-TT(经典的序贯分支).首先考察了一个只包含8个因子的小规模因子筛选算例,并讨论了不同的模型标准差以及与样本量对因子筛选结果的影响.然后,通过增加92个系数为零的因子,将8个因子的仿真算例扩充为一个包含100个因子的仿真算例,进一步说明了所提因子筛选方法在稀疏效应下的有效性.

#### 3.2.1 小规模因子筛选仿真算例

在小规模仿真算例中假设8个因子的效应系数为 $\beta = (1.6, 2.0, 2.4, 2.8, 3.2, 3.6, 4.0, 4.4)$ ,不重要性阈值 $\Delta = 2$ ,重抽样样本量 $B =$

200.此外,分别考察两种分布类型(正态分布以及均匀分布)、7种过程标准差( $\sigma = 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0$ )、以及6种样本量( $n = 5, 10, 15, 20, 25, 30$ )情形下的因子筛选结果.表2为响应标准差 $\sigma = 5$ 、样本量 $n = 5$ 时的四种因子筛选方法的筛选结果.表2中第1列表示8个因子的效应系数,分别在两种分布假设下通过仿真试验得到四种因子筛选方法的筛选结果(在 $N$ 次重复筛选中因子被识别为重要因子的频率, $f_i$ ).为了方便观察随着因子效应的变化四种因子筛选方法的变化趋势,将表2采用折线图进行可视化处理,如图6所示.图6的两个子图分别表示所考察的两种分布类型,横坐标表示8个因子的效应系数,纵坐标表示筛选结果,四种图形符号分别表示四种因子筛选方法.

可以看出,在两种所考察的分布类型下,随着因子效应的增加四种因子筛选方法的筛选结果基本呈现出单调递增的趋势(除异常点A之外).本文所提的三种因子筛选方法在两种所考察的分布类型下的因子筛选结果走势类似,说明了其在不同分布类型下的稳健性.然而,观察经典的因子筛选方法,在均匀分布下的因子筛选结果略低于正态分布情形.此外,随着因子效应的增加,本文所提的三种因子筛选方法筛选结果的增长趋势明显,增长速率较高;相比之下,经典的因子筛选方法增长速率较低.事实上,随着因子效应的增加,如果因子筛选方法筛选结果增长的速率越高,说

明此因子筛选方法更容易将不同效应的因子区分开；相反，如果增长的速率低说明其容易将不同效应的因子混杂在一起。如图 6 所示，本文采用比较直接的方式用折线图中起点到终点的斜率表示每

种因子筛选方法的增长率，其计算公式  $\gamma = \frac{[f_i(\#8) - f_i(\#1)]}{(\beta_8 - \beta_1)}$ 。例如，在响应为均匀分布类型时，因子筛选方法 SB-ABT 的增长率为  $\gamma$ 。

表 2 四种因子筛选方法的筛选结果比较( $\sigma = 5, n = 5$ )

Table 2 The comparison of the factor screening results of four methods ( $\sigma = 5, n = 5$ )

$\beta$	SB-BBT		SB-ABT		SB-RBT		SB-TT	
	正态/%	均匀/%	正态/%	均匀/%	正态/%	均匀/%	正态/%	均匀/%
1.6	25.90	24.90	20.30	20.80	25.60	24.30	0.40	0.00
2.0	25.60	27.10	20.90	25.60	28.40	26.30	0.30	0.60
2.4	41.40	40.50	34.80	31.30	37.40	37.90	0.50	0.60
2.8	43.50	43.30	35.70	38.20	40.90	44.20	1.30	0.90
3.2	54.40	58.10	49.10	50.10	56.00	56.10	4.60	3.10
3.6	60.30	58.70	52.00	53.90	60.20	57.50	3.30	4.00
4.0	65.60	68.40	61.30	61.50	66.70	66.90	5.10	5.80
4.4	69.60	73.20	64.30	66.20	72.10	73.30	7.80	7.20
$\gamma$	0.156 1	0.172 5	0.157 1	0.162 1	0.166 1	0.175 0	0.026 4	0.025 7

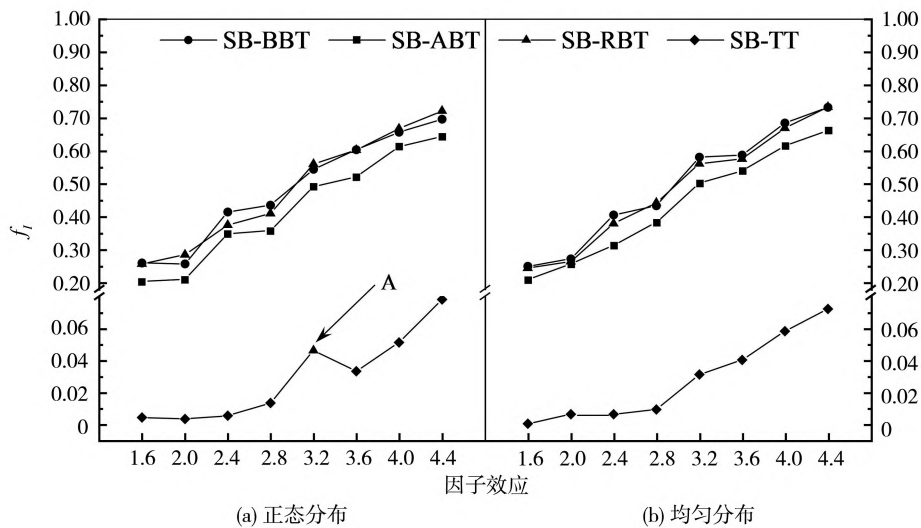


图 6 四种因子筛选方法筛选结果的变化趋势( $\sigma = 5, n = 5$ )

Fig. 6 The variation trends of the screening results from four factor screening methods( $\sigma = 5, n = 5$ )

$$\begin{aligned} \gamma &= [f_i(\#8) - f_i(\#1)] / (\beta_8 - \beta_1) \\ &= (66.2\% - 20.8\%) / (4.4 - 1.6) \\ &= 0.162 1 \end{aligned}$$

类似地，可以分别计算出两种分布假设下四种因子筛选方法的增长率，如表 2 最后一行所示。可以发现本文所提三种因子筛选方法在均匀分布下的增长率略大于正态分布，与此相反，经典因子筛选方法在均匀分布下的增长率略低于正态分布。

为了进一步考察所提因子筛选方法与经典因子筛选方法在各种响应标准差以及样本量组合下的表现情况，本文通过热力图的方式可视化所有

标准差与样本量组合下四种因子筛选方法识别因子显著频率的增长率  $\gamma$ ，如图 7 和图 8 所示。以图 7 为例，横坐标和纵坐标分别表示响应标准差以及样本量，在每组标准差以及样本量组合所对应的大方格内分为四个小方格，分别表示四种假设检验方法，小方格颜色的深浅表示所对应的因子筛选方法识别因子显著频率的增长率  $\gamma$ 。

从图 7 中可以看出，当响应标准差较小时，样本量不需要很大的情况下，经典的因子筛选方法的增长率大于本文所提的三种假设检验方法。然而，随着标准差的增大，要使得经典因子筛选方法

的增长率大于本文所提的因子筛选方法,所需要的最少样本量是不断增加的.例如,在响应标准差  $\sigma = 2$  的情况下,如果样本量  $n \geq 10$ ,那么因子筛选方法 SB-TT 的增长率优于其它三种因子筛选方法;然而,当响应标准差  $\sigma = 2.5$  时,样本量至少满足  $n \geq 15$ ,才能使得因子筛选方法 SB-TT 的增长率明显大于本文所提的三种因子筛选方法.因此,通过观察图 7 中各颜色所表示的热力值可以发现,SB-TT 方法优于其它三种因子筛选方法的方案呈现出左上角阶梯式情形;而处于对角线上标志为 (\*\*) 的 7 种情形表示 SB-TT 与本文所提的因子筛选方法无明显的颜色视觉差异.除此之外,在其余的标准差与样本量情形下(即标准差较大且样本量较小),经典的因子筛选方法 SB-TT 的增长率明显小于本文所提三种因子筛选方法.通过观察每个标准差与样本量组合下的四种因子筛选方法所代表的小方格热力值可以发现,本文所提三种方法呈现出的热力值差异较小,且在大部分情况下(除最右下角区域,即响应标准差较大但样本量极少),SB-RBT 方法的增长率高于 SB-RRT 和 SB-ABT 方法;然而,此三种因子筛选方法与 SB-TT 的差异则是随着标准差和样本量不断变化,且处在图中对角线区域时差异最小.

通过观察图 8 中均匀分布情形的增长率可以得到与正态分布情形类似的结论.然而,通过对比图 7 和图 8 可以发现,在同样的响应标准差下,如果要使得 SB-TT 方法的因子筛选结果的增长率明显优于其它三种因子筛选方法,均匀分布可能比正态分布需要更多的样本量.例如,在响应标准差  $\sigma = 4$  的情况下,如果响应服从正态分布,那么只要样本量满足  $n \geq 30$ ,图 7 中 SB-TT 的热力值明显大于其它三种因子筛选方法,然而,在图 8 中  $\sigma = 4, n = 30$  所对应的大方格内,四种因子筛选方法的差异相对较小.除此之外,图 8 中 SB-TT 对应的三个小方格热力值处在所观测热力值区间之外.也就是说当响应标准差较小时,SB-TT 方法在均匀分布下筛选结果的增长率随着样本量的增加更快;但是,在标准差较大且样本量较小时,SB-TT 方法在均匀分布下筛选结果的增长率却更容易出现较低的值.

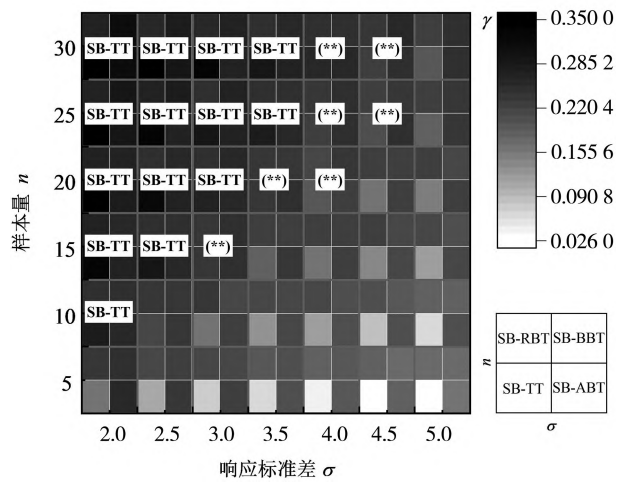


图 7 不同样本量以及响应标准差下因子筛选方法的增长率(正态分布)  
Fig. 7 The growth rate of factor screening methods under different combinations of sample size and response standard deviation (normal distribution)

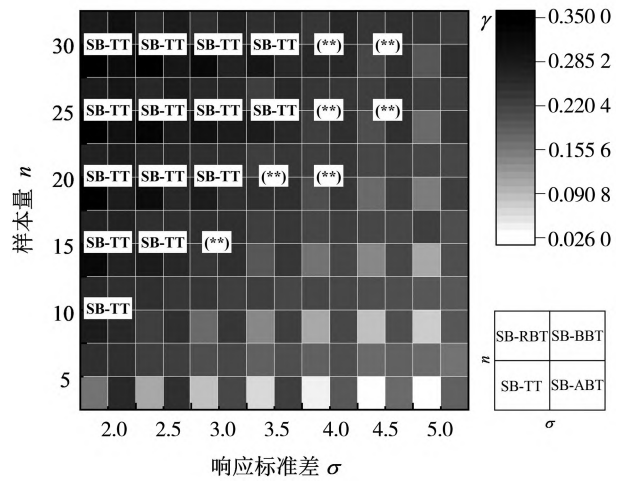


图 8 不同样本量以及响应标准差下因子筛选方法的增长率(均匀分布)  
Fig. 8 The growth rate of factor screening methods under different combinations of sample size and response standard deviation (uniform distribution)

### 3.2.2 大规模因子筛选仿真算例

序贯分支因子筛选方法因其高效性常用大规模因子且效应稀疏的筛选试验中,因此基于前述中的 8 个因子的仿真算例,通过增补 92 个效应为 0 的因子,将其扩充为一个包含 100 个因子的大规模因子筛选仿真算例.其余仿真参数与 3.2.1 小节一致,不重要性阈值  $\Delta = 2$ ,重抽样样本量  $B = 200$ .此外,7 种过程标准差( $\sigma = 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0$ )、以及 6 种样本量( $n = 5, 10, 15, 20, 25, 30$ )情形下的因子筛选结果.关于分布类型的假设,本小节除了考虑两种对称分布(正态分布、均匀分布),又增加了两种偏度系数相反的贝塔分布( $Beta(1,5)$ (右偏)、

$Beta(5,1)$ (左偏). 本文将因子筛选的结果分两个部分(非零因子筛选结果、零效应因子筛选结果)进行分析和说明. 1) 非零因子筛选结果分析.

针对 8 个因子效应非零的因子, 从因子被显著识别为重要因子的频率  $f_i$  以及筛选频率增长率  $\gamma$  两个角度, 比较四种因子筛选方法的筛选结果. 为了比较小规模仿真试验与大规模仿真试验的筛选结果的差异性, 本文分别在两种对称分布以及

两种偏度系数相反的贝塔分布的情形下, 考察仿真方案(响应标准差  $\sigma = 5$ 、样本量  $n = 5$ ) 的因子筛选结果, 如图 9 所示. 可以发现图 9 中的数据点以及折线图的大致趋势与图 6 中所反映出的信息类似, 尤其是本文所提的三种因子筛选方法在大规模仿真算例中所呈现出的线性趋势更为明显, 验证了本文所提因子筛选方法在大规模以及稀疏效应情形下的有效性.

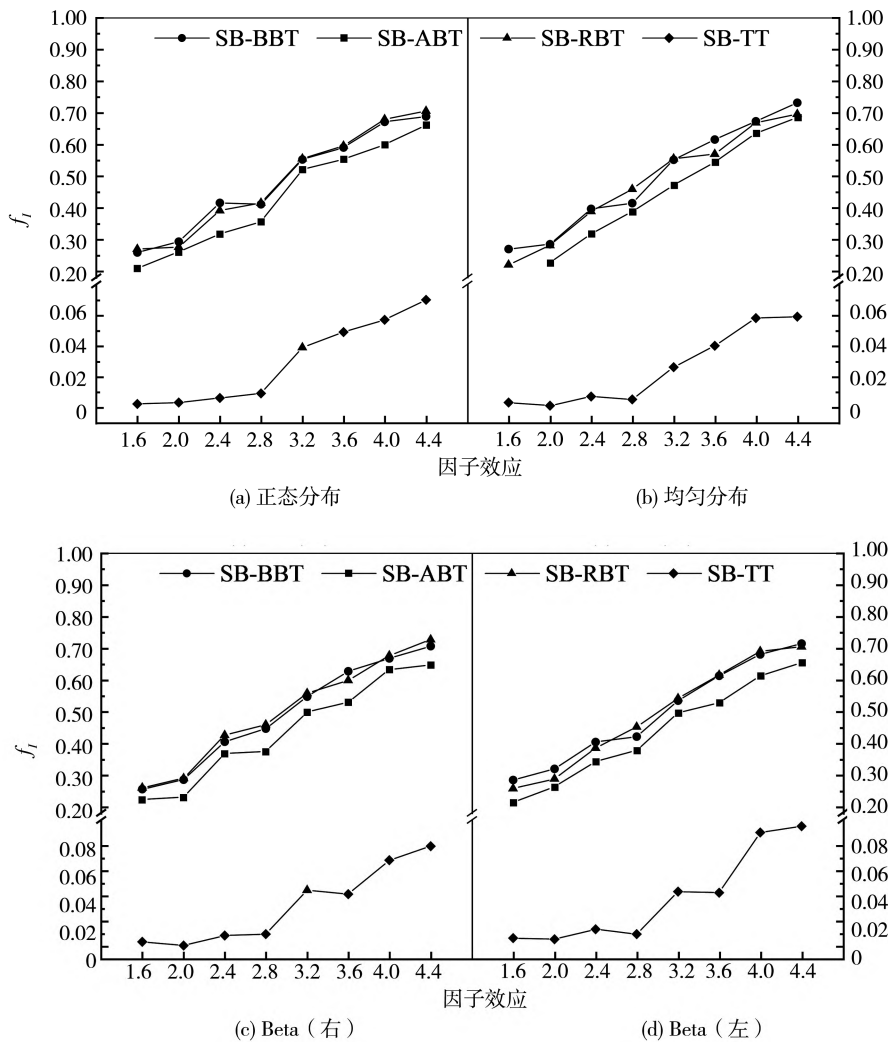


图 9 大规模筛选试验中四种因子筛选方法的非零因子筛选结果的变化趋势 ( $\sigma = 5, n = 5$ )

Fig. 9 The trend of factor screening results of four methods of non-zero factors according to effect coefficients in large-scale screening experiments ( $\sigma = 5, n = 5$ )

在此基础上, 本文在所考察的响应标准差及样本量组合下, 分别计算了四种因子筛选方法随着因子效应增加被识别为显著因子频率的增长率  $\gamma$ , 如图 10 所示. 通过比较图 7、图 8 及图 10 可以发现, 所考察的四种因子筛选方法针对大规模仿真试验中的非零因子以及小规模仿真试验的筛选

结果增长率变化情况基本类似. 首先, 在两种所假设的对称分布类型下(正态分布、均匀分布)如果响应标准差较小且样本量相对较大, 那么经典的因子筛选方法 SB-TT 的增长率大于本文所提方法; 与此相反, 当响应标准差较大且样本量相对较小时, 本文所提的三种因子筛选方法的增长率明

显大于 SB-TT. 除此之外,通过对比图 10 中的图 (a) 和图 (b) 可以发现,本文所提 SB-BBT 以及 SB-ABT 方法在均匀分布下的增长率明显大于正态分布情形,而 SB-RBT 方法的增长率受分布类型影响不明显,验证了其较好的稳健性. 另一方面,对比图 10 中的图 (c) 和图 (d) 可以发现,SB-RBT 方法和 SB-TT 方法的增长率的变化不受偏度系数的影响,并且与对称分布类型中所展现出的变化趋势基本一致,在固定响应标准差的情况下,增长率随着样本量的增加变大. 例如,在图 10 (c) 和图 10 (d) 中 SB-RBT 和 SB-TT 方法所对应

的色块热力值差异不大,且左上角区域 SB-TT 方法的热力值较大、右下角区域 SB-TT 方法的热力值较大. 与此相反,SB-BBT 和 SB-ABT 方法在偏态分布下的增长率变化趋势与对称分布差异较大,在响应标准差不变的情况下,样本量的增加反而会导致增长率的降低. 并且增长率受偏度系数的影响,例如通过对比图 10(c) 和图 10(d) 中 SB-BBT 方法的热力值可以发现,两个子图的左上角区域热力值的差异较大. 因此,比较四种因子筛选方法的综合性能,尤其是当响应分布类型了解较少的情况,处于对稳健性的考虑,更加推荐采用 SB-RSB 方法筛选关键因子.

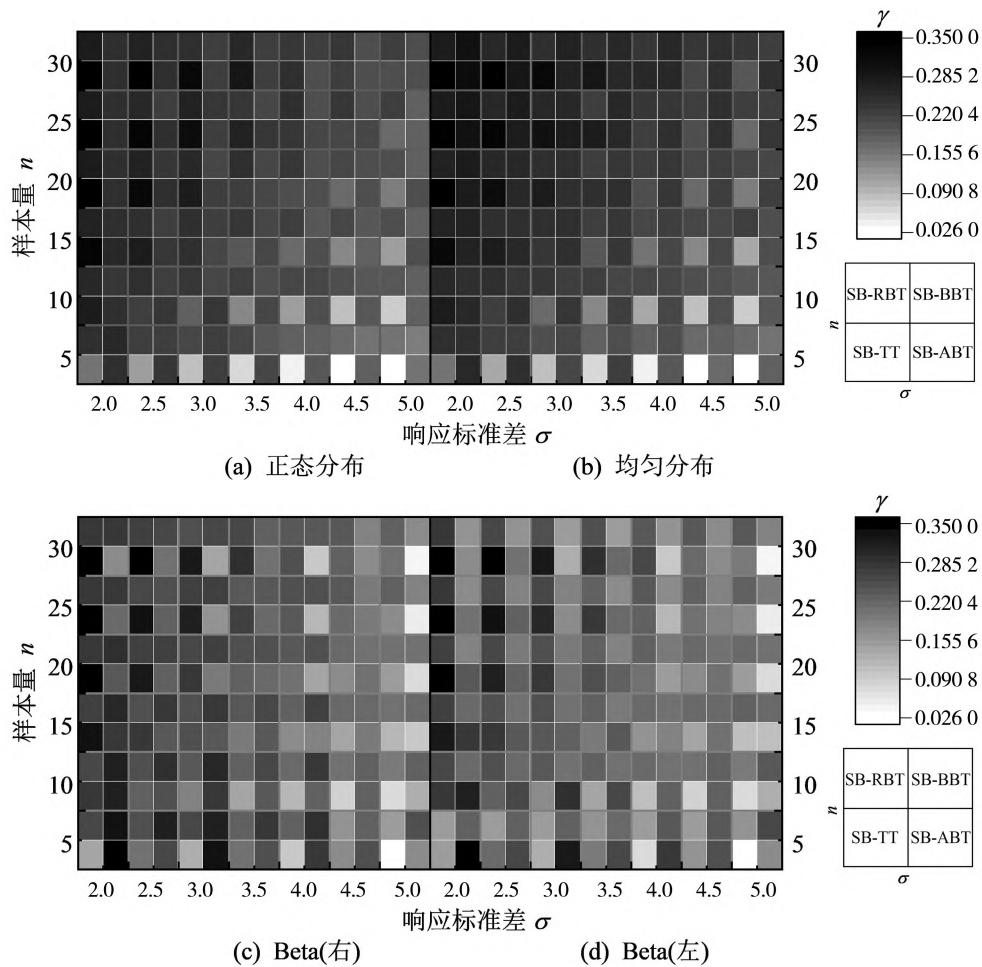


图 10 不同样本量以及响应标准差下因子筛选方法的增长率(大规模筛选试验中的非零因子)

Fig. 10 The growth rate of factor screening methods under different combinations of sample size and response standard deviation (for the non-zero factors in large-scale screening experiments)

2) 零因子筛选结果分析

根据本小节中的仿真算例假设,在 100 个因子中包含 92 个效应为零的因子(简称为,零因子),此外,分别在四种分布类型(正态分布、均匀分布、右偏及左偏贝塔分布)、42 种仿真方案

(7 种响应标准差以及 6 种样本量),统计了四种因子筛选方法在  $N(=1\ 000)$  次重复筛选中将零因子识别为重要因子的频率,记为  $f_i^0$ . 根据统计结果可以发现,在 42 种仿真方案以及四种因子筛选方法下总共可以得到共计 15 456 ( $=92 \times 42 \times$

4) 个零因子的筛选结果  $f_i^0$ , 其中大部分的  $f_i^0 = 0$ . 也就是说, 在大部分所考察的仿真情形下不会将零因子错误识别为重要因子, 例如, 在正态分布的假设下,  $f_i^0 \neq 0$  的结果共计 425 次, 占总筛选结果的 2.75%; 而在均匀分布的假设下,  $f_i^0 \neq 0$  的结果共计 247 次, 仅占总筛选结果的 1.60%. 为了进一步考察, 不同分布类型、响应标准差、样本量以及筛选方法对  $f_i^0$  的影响, 对称分布以及偏态分布类型下, 分别提取出  $f_i^0 \neq 0$  的试验方案并进行计数统计, 如图 11 和图 12 所示. 每个图又分为上下两个半区, 分别表示所考察的两种分布类型, 纵坐标表示对应分布类型假设下  $f_i^0 \neq 0$  的发生次数, 记为  $\#(f_i^0 \neq 0)$ . 虚线将横坐标分为三个区域, 从左至右分别表示 SB-BBT、SB-ABT 以及 SB-RBT 三种因子筛选方法. 每种因子筛选方法所在的虚线区间内共包含 5 个子组, 从左至右表示响应标准差依次增加 ( $\sigma = 3.0, 3.5, 4.0, 4.5, 5.0$ ), 堆叠柱状图颜色由浅至深表示样本量逐渐增加.

通过分析图 11 和图 12 中零因子筛选结果数据可以得到以下结论: 1) SB-TT 方法几乎不会将零因子错误识别为重要因子, 本文所提的三种方法在对称分布类型中的错误识别零因子概率分布在  $[0.1\%, 0.3\%]$ , 且 SB-BBT 错误频率大于 SB-ABT 大于 SB-RBT, 但是与偏态分布类型相比,

SB-RSB 方法几乎与在对称分布中表现一致, SB-BBT 以及 SB-ABT 方法在分布为右偏时略优于对称分布、为左偏时明显劣于对称分布; 2) 在对称分布类型下, SB-BBT 和 SB-ABT 方法在正态分布下的错误识别零因子为重要因子的频率大于均匀分布, SB-RBT 在两种对称分布下的差异性不明显; 类似地, 在偏态分布类型下, SB-BBT 和 SB-ABT 方法在左偏下的错误识别零因子为重要因子的频率大于右偏分布, SB-RBT 在受偏度系数影响不明显; 3) 随着响应标准差的增大, 三种因子筛选方法错误识别零因子为重要因子的频率逐渐增加. 且 SB-BBT 的增加速率大于 SB-ABT 大于 SB-RBT; 4) 在响应标准差较小且样本量较大的情况下错误识别零因子为重要因子的频率较低; 相反, 如果响应标准差较大且样本量较小则错误识别零因子为重要因子的频率较大; 5) 与 SB-BBT 以及 SB-ABT 方法相比, SB-RBT 错误识别零因子为重要因子的频率受分布类型以及偏度系数的影响不大, 稳健性较好; 而且在本文所考察的响应标准差及样本量的组合方案中, 即使响应标准差的增大会导致 SB-RBT 识别零因子为重要因子的概率增大, 但是只要可以保证样本量满足, 则错误识别零因子的频率较低且发生次数较少.

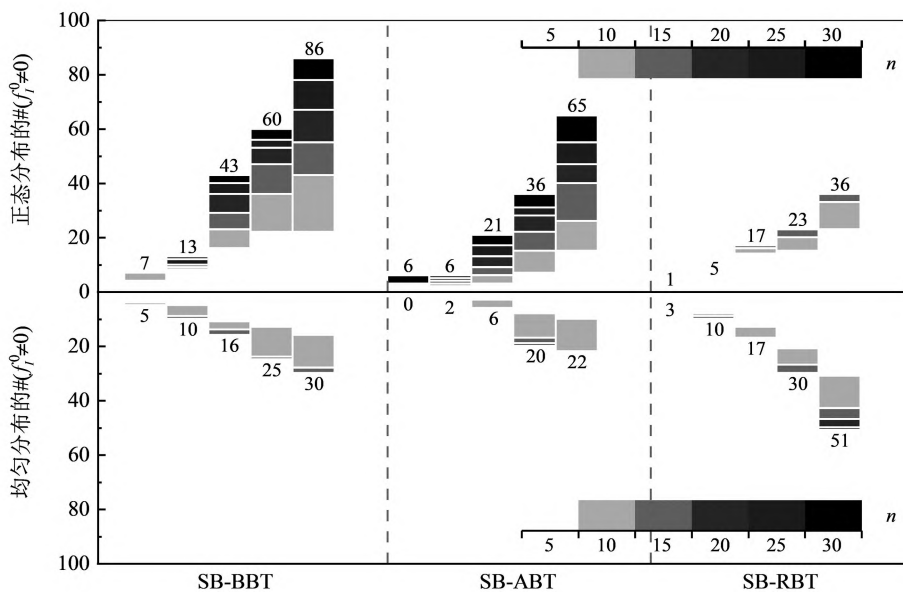


图 11 大规模筛选试验中零因子被识别为重要因子的频率 (对称分布)

Fig. 11 The frequency that zero factor identified as important factors in large-scale screening experiments (symmetric distributions)

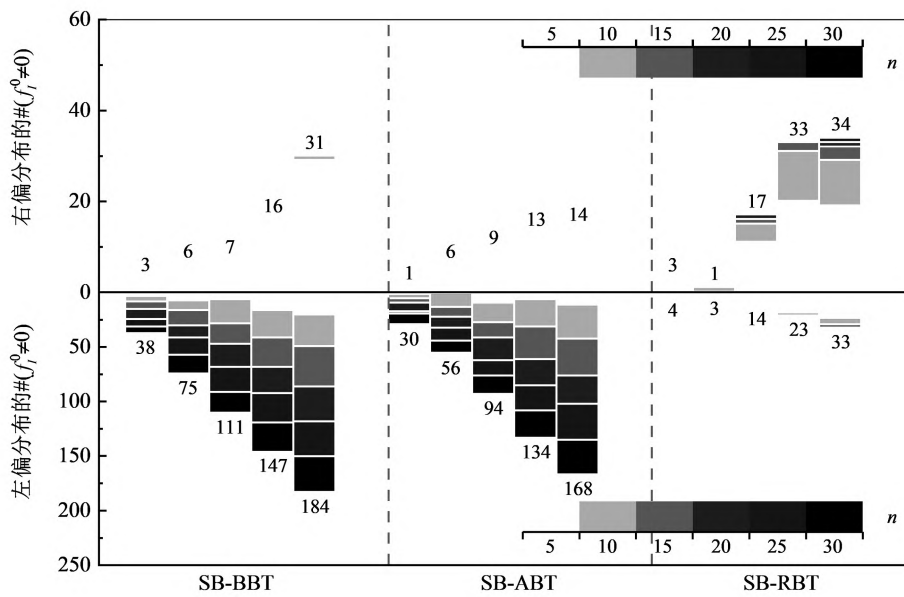


图 12 大规模筛选试验中零因子被识别为重要因子的频率 ( 偏态分布 )

Fig. 12 The frequency that zero factor identified as important factors in large-scale screening experiments ( asymmetric distributions )

### 4 结束语

客户化、定制化生产方式的普及使得关于小样本的统计建模成为突破传统批量制造模式的关键技术,也是目前质量改进领域所关注的重要问题. 本文针对定制化生产中可能出现的小样本问题,在利用自举法扩充样本量的基础上,提出了适用于小样本的三种假设检验过程,并将其融入序贯分支因子筛选方法用于因子效应的显著性检验,以实现高效的关键因子筛选,为后续的建模以及优化奠定基础. 通过蒙特卡洛仿真分别验证了本文所提的三种假设检验方法在两个总体假设检验问题以及因子筛选问题中的有效性和稳健性,并得到如下结论.

首先,经典的学生 t 检验以及相应的因子筛选方法在第一类错误方面表现良好,但是第二类错误较大,尤其是在响应标准差较大且样本量较小的情形下,几乎不能将重要因子与不重要因子明确分类,说明了经典的序贯分支方法基于学生 t 检验不适用于本文所讨论的小样本情形下的因子筛选问题. 然而,本文所提的三种假设检验以及因子筛选方法尽管会导致第一类

错误的增加,但是可以实现以较高的频率识别出重要因子,而且在因子筛选的问题中,可以实现不重要因子与重要因子的有效区分. 由于第二类错误意味着是否会丢失一些重要因子,说明了本文所提方法在小样本因子筛选问题中的必要性和有效性. 最后,通过对比本文所提的三种方法,更加推荐基于稳健统计量的 Bootstrap 假设检验方法 (robust statistics based Bootstrap for significance testing, RBT) 以及相应的因子筛选方法 SB-RBT,这是由于 SB-RST 不仅表现出与另外两种所提方法的相当的有效性,而且在不同的分布类型下具有较好的稳健性,更加适用于小样本情形下难以验证响应分布类型的问题. 本文假设样本量在每个设计点下固定且相等,不等样本量的因子筛选以及给定总试验成本的样本量分配问题将是进一步值得考虑的研究方向. 此外,在针对具体应用案例的研究中,小样本问题可能与其他复杂问题夹杂在一起,例如,定制化生产是导致小样本问题的潜在应用场景,然而在增材制造熔融沉积成型过程的质量改进中,质量的一致性、变量类型的多样性均是值得进一步研究的问题,相关的统计分析建模方法仍有待进一步探索.

## 参考文献:

- [1] 窦润亮, 郭均鹏, 田祥龙, 等. 面向客户个性化需求的交互式遗传算法[J]. 管理科学学报, 2016, 19(1): 24-34.  
Dou Runliang, Guo Junpeng, Tian Xianglong, et al. Interactive genetic algorithm based on customer demand[J]. Journal of Management Sciences in China, 2016, 19(1): 24-34. (in Chinese)
- [2] Liu L, Wu Z, Yu Y. A branch-and-price algorithm to perform single-machine scheduling for additive manufacturing[J]. Journal of Management Science and Engineering, 2023, 8(2): 273-286.
- [3] 梦炯, 张杨, 曾波. 基于个性化需求的产品竞争供应链结构选择[J]. 中国管理科学, 2019, 27(12): 67-76.  
Meng Jiong, Zhang Yang, Zeng Bo. Study on structure selection of product competition supply chain based on individualized demand[J]. Chinese Journal of Management Science, 2019, 27(12): 67-76. (in Chinese)
- [4] Zonnenshain A, Kenett R S. Quality 4.0: The challenging future of quality engineering[J]. Quality Engineering, 2020, 32(4): 614-626.
- [5] Colosimo B M, Huang Q, Dasgupta T, et al. Opportunities and challenges of quality engineering for additive manufacturing[J]. Journal of Quality Technology, 2018, 50(3): 233-252.
- [6] Olsen T L, Tomlin B. Industry 4.0: Opportunities and challenges for operations management[J]. Manufacturing and Service Operations Management, 2020, 22(1): 113-122.
- [7] Kulahci M, Frumosu F D, Khan A R, et al. Experiences with big data: Accounts from a data scientist's perspective[J]. Quality Engineering, 2020, 32(4): 529-542.
- [8] 卢秉恒, 李涤尘. 增材制造(3D打印)技术发展[J]. 机械制造与自动化, 2013, 42(4): 1-4.  
Lu Bingheng, Li Dicheng. Development of the additive manufacturing (3D) technology[J]. Machine Building & Automation, 2013, 42(4): 1-4. (in Chinese)
- [9] Tsung F, Zhang K, Cheng L, et al. Statistical transfer learning: A review and some extensions to statistical process control[J]. Quality Engineering, 2018, 30(1): 115-128.
- [10] 崔庆安, 崔楠. 试验设计及参数优化的 LS-SVR 显著性因子筛选[J]. 管理科学学报, 2023, 26(12): 42-61.  
Cui Qingan, Cui Nan. Significant factor screening of LS-SVR for experimental design and parameters optimization[J]. Journal of Management Sciences in China, 2023, 26(12): 42-61. (in Chinese)
- [11] Ouyang L, Wan L, Park C, et al. Ensemble RBF modeling technique for quality design[J]. Journal of Management Science and Engineering, 2019, 4(2): 105-118.
- [12] 汪建均, 屠雅楠, 马义中. 结合 SUR 与因子效应原则的多响应质量设计[J]. 管理科学学报, 2020, 23(12): 12-29.  
Wang Jianjun, Tu Yanan, Ma Yizhong. Multi-response quality design integrating SUR models with factorial effect principles[J]. Journal of Management Sciences in China, 2020, 23(12): 12-29. (in Chinese)
- [13] 施文, 刘志学, 刘丹. 零部件循环取货越库物流系统关键因子研究——仿真试验设计与分析方法[J]. 管理科学学报, 2013, 16(6): 22-34.  
Shi Wen, Liu Zhixue, Liu Dan. Key factors of parts logistics for milk-run and cross-docking system: Design and analysis of simulation experiments method[J]. Journal of Management Sciences in China, 2013, 16(6): 22-34. (in Chinese)
- [14] Dean A, Lewis S. Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics[M]. New York: Springer, 2006.

- [15] Shi W, Shang J, Liu Z, et al. Optimal design of the auto parts supply chain for JIT operations: Sequential bifurcation factor screening and multi-response surface methodology [J]. *European Journal of Operational Research*, 2014, 236(2): 664–676.
- [16] Kleijnen J P C. *Design and Analysis of Simulation Experiments*[M]. 2<sup>nd</sup> ed. New York: Springer, 2015.
- [17] Shen H, Wan H. Controlled sequential factorial design for simulation factor screening[J]. *European Journal of Operational Research*, 2009, 198(2): 511–519.
- [18] Shen H, Wan H, Sanchez S M. A hybrid method for simulation factor screening[J]. *Naval Research Logistics*, 2010, 57(1): 45–57.
- [19] Shi W, Kleijnen J P C, Liu Z. Factor screening for simulation with multiple responses: Sequential bifurcation[J]. *European Journal of Operational Research*, 2014, 237(1): 136–147.
- [20] Wan H, Ankenman B E, Nelson B L. Improving the efficiency and efficacy of controlled sequential bifurcation for simulation factor screening[J]. *Inform Journal on Computing*, 2010, 22(3): 482–492.
- [21] Wan H, Ankenman B E, Nelson B L. Controlled sequential bifurcation: A new factor-screening method for discrete-event simulation[J]. *Operations Research*, 2006, 54(4): 743–755.
- [22] 刘丽君, 马义中, 欧阳林寒, 等. 数据污染对于序贯分支方法的影响及应对策略[J]. *系统工程理论与实践*, 2020, 40(5): 1281–1292.  
Liu Lijun, Ma Yizhong, Ouyang Linhan, et al. The influence of data contamination on sequential bifurcation and its counter measures[J]. *System Engineering: Theory & Practice*, 2020, 40(5): 1281–1292. (in Chinese)
- [23] Liu L, Ma Y, Park C, et al. Robust sequential bifurcation for simulation factor screening under data contamination[J]. *Computers & Industrial Engineering*, 2019, (129): 102–112.
- [24] Ankenman B E, Cheng R C H, Lewis S M. Screening for dispersion effects by sequential bifurcation[J]. *ACM Transactions on Modeling and Computer Simulation*, Association for Computing Machinery, 2015, 25(1): 1–27.
- [25] 刘丽君, 马义中, 欧阳林寒, 等. 基于位置和散度效应因子分类筛选及仿真试验[J]. *系统工程理论与实践*, 2017, 37(11): 2854–2862.  
Liu Lijun, Ma Yizhong, Ouyang Linhan, et al. Screening for location and dispersion effects simultaneously by sequential bifurcation[J]. *System Engineering: Theory & Practice*, 2017, 37(11): 2854–2862. (in Chinese)
- [26] 刘丽君, 马义中, 龚本刚, 等. 基于非均衡序贯设计的大规模关键质量因子筛选研究[J]. *中国管理科学*, 2025, 33(2): 84–94.  
Liu Lijun, Ma Yizhong, Gong Bengang, et al. Key quality factor screening for large-scale simulation based on unbalanced sequential design[J]. *Chinese Journal of Management Science*, 2025, 33(2): 84–94. (in Chinese)
- [27] 刘丽君, 马义中, 欧阳林寒. 高效的因子分类筛选方法及仿真应用[J]. *系统管理学报*, 2021, 30(1): 3–13.  
Liu Lijun, Ma Yizhong, Ouyang Linhan. An efficient factor screening and classification procedure based on multi-response sequential bifurcation and applications[J]. *Journal of Systems & Management*, 2021, 30(1): 3–13. (in Chinese)
- [28] Spokoiny V, Zhilova M. Bootstrap confidence sets under model misspecification[J]. *The Annals of Statistics*, 2015, 43(6): 2653–2675.
- [29] 张尧晓, 田茂再. 基于贝叶斯的 Bootstrap 置信区间[J]. *统计与决策*, 2020, 36(1): 32–35.  
Zhang Chongliang, Tian Maozai. Bootstrap confidence interval based on Bayes[J]. *Statistics & Decision*, 2020, 36(1): 32–35. (in Chinese)
- [30] Efron B. Bootstrap methods: Another look at the Jackknife[J]. *The Annals of Statistics*, 1979, 7(1): 1–26.
- [31] Park C, Ouyang L, Byun J H, et al. Robust design under normal model departure[J]. *Computers and Industrial Engineering*, 2017, (113): 206–220.

## Bootstrap hypothesis testing with bias correction for screening large-scale quality factors

*LIU Li-jun*<sup>1,2</sup>, *OUYANG Lin-han*<sup>3</sup>, *MA Yi-zhong*<sup>2</sup>, *WU Feng*<sup>1</sup>

1. School of Economics and Management, Anhui Polytechnic University, Wuhu 241000, China;
2. School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China;
3. College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

**Abstract:** With the increasing complexity of customer demand, quality improvement faces significant challenges, as personalized and customized production patterns are frequently adopted and large-scale factors are involved in complex production. Aiming at the challenges posed by large-scale factors and small sample sizes in complex and customized production, this paper proposes a bootstrap-based factor screening method, which serves as the first stage of quality improvement. Firstly, a one-order polynomial model is adopted to fit the input-output relationship without assuming a specific distribution for the random term, accommodating the situation of small sample size. Secondly, the sequential bifurcation (SB) procedure is modified according to the distribution-free response model. Thirdly, three significance testing methods for factor effects are proposed, based on Students' t test and enhanced through bootstrapping and bias correction procedures. Finally, Monte Carlo simulations are employed to compare the proposed significance testing methods with the classic Student's t test under small sample sizes, and to verify the effectiveness and robustness of the proposed screening methods in both small-scale and large-scale factor scenarios.

**Key words:** small sample size; large scale; factor screening; sequential bifurcation; Bootstrap; hypothesis testing