Lee 偏差下二三混水平因子设计的最优添加

献给方开泰教授 80 华诞

勾廷勋¹, 覃红^{2,3*}, Kashinath Chatterjee⁴

1. 西北大学数学学院, 西安 710127;

2. 中南财经政法大学统计与数学学院, 武汉 430073;

3. 华中师范大学数学与统计学学院, 武汉 430079;

4. Department of Statistics, Visva-Bharati University, Santiniketan 731235, India

E-mail: goutx@nwu.edu.cn, qinhong@mail.ccnu.edu.cn, kashinathchatterjee@gmail.com

收稿日期: 2019-04-05; 接受日期: 2019-07-21; 网络出版日期: 2020-04-29; * 通信作者

国家自然科学基金 (批准号: 11901455 和 11871237) 资助项目

摘要 根据序贯试验通过一个阶段试验接着另一个阶段试验不断扩充的特征, 新的试验点将会被添加到已经选好的设计中, 因此, 如何设计一个好的序贯试验是一个非常有意义的问题. 在 Lee 偏差度量下, 本文研究均匀的非对称拓展设计, 并用它来做序贯试验. 本文建立二三混水平拓展设计的 Lee 偏差与其初始设计和附加设计的 Lee 偏差之间的解析关系, 为构造 (近似) 均匀的拓展设计提供了理论支撑. 为了给出筛选、评价拓展设计的准则, 本文给出拓展设计 Lee 偏差的下界. 此外, 还通过一些数值例子来进一步说明、解释本文的理论结果.

关键词 均匀性 二三混水平设计 拓展设计 序贯试验

MSC (2010) 主题分类 62K15, 62K10, 62K99

1 引言

在许多领域, 如医药、制造和高新技术等行业, 新产品或者新工艺的研发常常需要进行序贯试验. 为了研究复杂的系统或者理解复杂的现象, 序贯试验 (也称为跟随试验) 的优势越来越明显. 早期, 序贯方法在质量改进的响应曲面方法中扮演了重要角色. 文献 [1] 采用序贯试验研究了集成电路的制造工艺. 文献 [2, 3] 探索了计算机仿真中的序贯试验问题. 序贯试验在生物学领域的应用可参见文献 [4–6]. 需要特别强调的是, 文献 [6] 在银杏叶提取物的指纹图谱研发中首次采用了均匀序贯试验方法.

拓展设计是在一个事先选好的设计 (称之为初始设计) 基础上通过不断添加新的子设计 (称之为附加设计) 而构成的新设计, 其中每一个子设计都可以用来进行一次试验. 对序贯试验中每个阶段的

英文引用格式: Gou T X, Qin H, Chatterjee K. Optimal additions to asymmetrical fractional factorial designs under Lee discrepancy (in Chinese). Sci Sin Math, 2020, 50: 599–612, doi: 10.1360/N012019-00095

试验, 选择某个子设计去执行试验; 当需要再进行新一阶段的试验时, 重新选择一个子设计去执行该试验. 由此可见, 设计整个序贯试验是通过设计其各个阶段的子试验的方式来开展的. 因此, 拓展设计可以有效地用于执行序贯试验. 但是, 如何评价拓展设计的好坏以及如何构造好的拓展设计是充满挑战同时也很有现实意义的问题. 文献 [7] 较早地对这个问题进行了尝试性研究, 给出了在正交表上仅添加一个新的试验点仍然是广义 I- 型最优的充分条件. 文献 [8] 限定在二水平正交表, 将文献 [7] 中的结果推广到允许加入多个试验点的情况, 并给出了添加多个新的试验点后仍然是 Schur- 最优的充分条件. 对于一般的非对称正交表, 文献 [9] 研究了添加多个试验点后仍然是 D- 最优的条件. 值得注意的是, 广义 I- 型最优 (包含了 D- 最优) 和 Schur- 最优都依赖于提前假定好的线性模型, 因为它们是基于这个线性模型的信息矩阵去评价设计好坏的. 因此, 当真实模型偏离假定好的线性模型时, 设计在这些准则下的优劣情况可能会发生极大变化, 即这些准则的稳健性较差.

广义最小低阶混杂 (generalized minimum aberration, GMA) 准则的稳健性较好. 因此, 文献 [10] 建议采用该准则选择好的拓展设计, 并研究了采用 GMA 准则通过选择由强度为 2 的二水平正交表中仅添加一个新的设计点所得的拓展设计的性质. 文献 [11] 详细分析了在强度为 2 的二水平正交表上添加两个新设计点的拓展设计的效应混杂情况. 均匀性准则是另一种重要的稳健性准则. 在实践中, 试验者事先对真实模型的知识常常比较缺乏, 因此, 设计的稳健性变得越来越重要. 从几何直观的角度出发, 均匀性准则旨在将设计点尽可能均匀地分散到给定的设计区域内, 并且对刻画自然科学以及工程实践中出现的许多复杂系统非常有效. 由于该准则使用方便并且稳健可靠, 因此, 均匀设计已经成为研究复杂系统常用的方法, 详情参见文献 [12, 13]. 文献 [14] 首先采用中心化 L_2 和可卷型 L_2 偏差研究了二水平拓展设计的均匀性问题. 三水平对称拓展设计的均匀性问题在文献 [15] 中也得到了进一步研究. 最近, 文献 [16] 在可卷型 L_2 偏差下研究了添加新试验点和添加新试验因子的问题. Lee 偏差也是一种常用的均匀性度量, 本文将利用 Lee 偏差来研究拓展设计的均匀性问题.

对于探索复杂系统的试验而言, 其实施起来成本很高并且也非常耗时, 因此, 试验次数较少的设计备受青睐, 这要求试验者对每个因子精心选择较小的水平. 为了分析和构造的简单, 平衡性在试验设计中常常被考虑. 最近, 非平衡超饱和设计的优良性质在文献 [17] 中得到了深入讨论. 本文研究二三混水平拓展设计的均匀性问题 (不论其是否为平衡设计), 并给出拓展设计的 Lee 偏差与其子设计的 Lee 偏差之间的解析线性关系, 同时给出二三混水平拓展设计 Lee 偏差的一个下界, 该下界能够有效地用于评价、选择序贯试验中好的设计.

2 Lee 偏差、正交表及其他预备知识

$\mathcal{D}_0(n, 2^{m_1} \times 3^{m_2})$ 表示由全体 n 次试验、前 m_1 个因子为二水平、其余 m_2 个因子为三水平的 U- 型 (平衡) 设计组成的集合, 其中二水平因子的水平记为 0 和 1, 三水平因子的水平记为 0、1 和 2. 此时 n 必须为 6 的倍数. 同理, 用 $\mathcal{D}_1(n, 2^{m_1} \times 3^{m_2})$ 表示全体 n 次试验、前 m_1 个因子为二水平、其余 m_2 个因子为三水平的近似 U- 型设计构成的集合, 这里近似 U- 型设计是指, 对每个因子而言, 各个水平出现的次数至多相差一次, 从而, n 可以为任意正整数. 因此, $\mathcal{D}_0(n, 2^{m_1} \times 3^{m_2}) \subseteq \mathcal{D}_1(n, 2^{m_1} \times 3^{m_2})$. 如果一个设计的 Lee 偏差在设计集合 $\mathcal{D}_0(n, 2^{m_1} \times 3^{m_2})$ (或者 $\mathcal{D}_1(n, 2^{m_1} \times 3^{m_2})$) 中取到最小值, 则称其为均匀设计.

对于任意设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} \times 3^{m_2})$, $g = 0, 1$, 其 $m (= m_1 + m_2)$ 个因子的每个水平组合都可以表示为 $x = x^{(1)}x^{(2)}$ 的形式, 其中 $x^{(1)} = x_1^{(1)}x_2^{(1)} \cdots x_{m_1}^{(1)}$ 对应它的 m_1 个二水平因子, 而 $x^{(2)} =$

$x_1^{(2)} x_2^{(2)} \cdots x_{m_2}^{(2)}$ 与它的 m_2 个三水平因子相对应, 即 $x_j^{(1)} \in \{0, 1\}$ ($1 \leq j \leq m_1$) 和 $x_j^{(2)} \in \{0, 1, 2\}$ ($1 \leq j \leq m_2$). 假设 V 是全部 $v (= 2^{m_1} 3^{m_2})$ 个水平组合构成的有序集, 其元素按水平组合 x 的字典序排列. 对每个水平组合 $x \in V$ 和任意给定的设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} \times 3^{m_2})$, $g = 0, 1$, 我们用 $n_{d_g}(x)$ 表示水平组合 x 出现在 d_g 中的次数, 用 n_{d_g} 表示由全部的 $n_{d_g}(x)$ 构成的一个 $v \times 1$ 向量, 其元素也是按照水平组合 x 的字典序排列. 对于任意设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} \times 3^{m_2})$ 的任意两行 (记为 i_1 和 i_2) 的相遇数 (即该两个水平组合中对应位置取相同水平的位置个数), 我们用 $c_{i_1 i_2}^{(g1)}$ 表示设计 d_g 中所有二水平因子 (即前 m_1 列) 在第 i_1 和 i_2 行的相遇数, 用 $c_{i_1 i_2}^{(g2)}$ 表示设计 d_g 的所有三水平因子 (即剩下的 m_2 列) 在第 i_1 和 i_2 行的相遇数.

基于编码理论中的 Lee 距离, 文献 [18] 定义了 Lee 偏差. 对 q -水平因子, 如果其水平编码为 $0, 1, 2, \dots, q-1$, 则其任意两个水平 τ_1 与 τ_2 之间的 Lee 距离为 $L(\tau_1, \tau_2) = \min\{|\tau_1 - \tau_2|, q - |\tau_1 - \tau_2|\}$. 由文献 [18] 可知, 对任意设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} \times 3^{m_2})$, $g = 0, 1$, 它的 Lee 偏差记为 $LD_2(d_g)$, 其计算公式如下:

$$[LD_2(d_g)]^2 = K + \frac{1}{n^2} \sum_{i_1=1}^n \sum_{i_2=1}^n \prod_{j=1}^{m_1} \left[1 - \frac{1}{2} L(y_{i_1 j}^{(1)}, y_{i_2 j}^{(1)}) \right] \prod_{j=m_1+1}^m \left[1 - \frac{1}{3} L(y_{i_1 j}^{(2)}, y_{i_2 j}^{(2)}) \right], \quad (2.1)$$

其中 $K = -(3/4)^{m_1} (7/9)^{m_2}$, $y_{ij}^{(1)} \in \{0, 1\}$, $y_{ij}^{(2)} \in \{0, 1, 2\}$.

本文将 Lee 偏差表示成二次型和相遇数两种不同的形式, 这两种形式将为本文后续的计算提供许多便利. 我们定义如下矩阵:

$$M_1 = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 2/3 & 2/3 \\ 2/3 & 1 & 2/3 \\ 2/3 & 2/3 & 1 \end{pmatrix},$$

$$L_1(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad L_1(0) = (1, 1), \quad L_2(1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad L_2(0) = (1, 1, 1).$$

易知, 矩阵 M_1 和 M_2 可以分别分解为

$$M_1 = \frac{1}{2} L_1(0)^T L_1(0) + \frac{1}{2} L_1(1)^T L_1(1) \quad \text{和} \quad M_2 = \frac{2}{3} L_2(0)^T L_2(0) + \frac{1}{3} L_2(1)^T L_2(1),$$

其中 A^T 表示矩阵 A 的转置. 此外, 我们用 $M_1^{m_1}$ 和 $M_2^{m_2}$ 分别表示矩阵 M_1 的 m_1 -重 Kronecker 积和矩阵 M_2 的 m_2 -重 Kronecker 积, 用 M^m 表示矩阵 $M_1^{m_1}$ 和 $M_2^{m_2}$ 的 Kronecker 积.

根据上述记号, 对于设计 d_g , $g = 0, 1$, (2.1) 可以表示为如下二次型形式:

$$[LD_2(d_g)]^2 = K + \frac{1}{n^2} n_{d_g}^T M^m n_{d_g}. \quad (2.2)$$

根据上述关于设计 d_g 相遇数的记号, (2.1) 也可以表示为如下相遇数形式:

$$[LD_2(d_g)]^2 = K + \frac{1}{n^2} \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{i_1=1}^N \sum_{i_2=1}^N 2^{c_{i_1 i_2}^{(g1)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(g2)}}. \quad (2.3)$$

文献 [19] 给出了二三混水平设计 Lee 偏差的一个下界, 这一下界对本文后续部分研究很有帮助. 其具体结论见下述引理.

引理 2.1 对于任意设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} \times 3^{m_2})$, $g = 0, 1$, (2.2) 中的 Lee 偏差 $LD_2(d_g)$ 满足下述不等式:

$$[LD_2(d_g)]^2 \geq LB_1^{(g)}(n, m_1, m_2),$$

其中

$$LB_1^{(g)}(n, m_1, m_2) = K + \frac{1}{n^2} \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{j_1=0}^{m_1} \sum_{j_2=0}^{m_2} \binom{m_1}{j_1} \binom{m_2}{j_2} \left(\frac{1}{2}\right)^{j_2} \theta_{j_1 j_2},$$

$\tau_{j_1 j_2} = \lfloor n/2^{j_1} 3^{j_2} \rfloor$, $\varsigma_{j_1 j_2} = n - 2^{j_1} 3^{j_2} \tau_{j_1 j_2}$, $\theta_{j_1 j_2} = n \tau_{j_1 j_2} + \varsigma_{j_1 j_2} (\tau_{j_1 j_2} + 1)$, $\lfloor a \rfloor$ 是不大于 a 的最大整数.

正交表在试验设计和编码理论中占有重要地位, 也是本文后续研究的基础. 为了便于理解, 我们将正交表的定义介绍如下.

定义 2.1 一个 $N \times (m_1 + m_2)$ 表, 其前 m_1 列中的每列有两个水平, 剩下的 m_2 列中的每列有三个水平, 称其为一个强度为 h 的正交表, 如果其满足下列条件: 对于其任意的 $N \times h$ 子表, 该子表恰好是一个完全因子设计或者一个完全因子设计的多次重复. 我们把这样的正交表记为 $OA(N, m_1 + m_2, 2^{m_1} \times 3^{m_2}, h)$.

3 二三混水平拓展设计的 Lee 偏差

本节主要推导二三混水平拓展设计的 Lee 偏差的计算公式, 建立拓展设计的 Lee 偏差与其初始设计和附加设计的 Lee 偏差之间的一个解析表达式. 假设在序贯试验中试验者在进行第一阶段试验时所选用的设计为 $d_0 \in \mathcal{D}_0(N, 2^{m_1} \times 3^{m_2})$, d_0 也称为初始设计; 试验者在进行第二阶段的试验时选用的设计为 $d_1 \in \mathcal{D}_1(r, 2^{m_1} \times 3^{m_2})$, d_1 也称为附加设计, 则整个试验中所使用的设计称为拓展设计, 由初始设计 d_0 和附加设计 d_1 构成, 记为 $d^E \in \mathcal{D}^E(N + r, 2^{m_1} \times 3^{m_2})$. 对于给定的拓展设计 d^E , 对水平组合 $x \in V$, 我们用 $n_{d^E}(x)$ 表示水平组合 x 出现在拓展设计 d^E 中的次数, 用 n_{d^E} 表示由所有的 $n_{d^E}(x)$ 构成的一个 $v \times 1$ 向量, 它的元素按照水平组合 x 的字典序排列.

由于拓展设计 d^E 中的试验点由初始设计 d_0 和附加设计 d_1 中的试验点构成, 其前 N 个试验点来自 d_0 , 剩下的 r 个试验点来自 d_1 . 特别地, 对于 d^E 的任意两行 (记为 i_1 和 i_2) 的相遇数, 可分为如下情形:

(i) 当 $1 \leq i_1, i_2 \leq N$ 时, $c_{i_1 i_2}^{(01)}$ 表示设计 d_0 中所有二水平因子 (即前 m_1 列) 在第 i_1 与 i_2 行的相遇数, 用 $c_{i_1 i_2}^{(02)}$ 表示设计 d_0 的所有三水平因子 (即剩下的 m_2 列) 在第 i_1 与 i_2 行的相遇数;

(ii) 当 $N + 1 \leq i_1, i_2 \leq N + r$ 时, $c_{i_1 i_2}^{(11)}$ 表示设计 d_1 的全部 m_1 个二水平因子在第 i_1 与 i_2 行的相遇数, 用 $c_{i_1 i_2}^{(12)}$ 表示 d_1 的全部 m_2 个三水平因子在第 i_1 与 i_2 行的相遇数;

(iii) 当 $1 \leq i_1 \leq N$ 且 $N + 1 \leq i_2 \leq N + r$ 时, 用 $c_{i_1 i_2}^{(011)}$ 表示 d_0 的所有 m_1 个二水平因子在第 i_1 行与 d_1 的所有 m_1 个二水平因子在第 i_2 行的相遇数, 用 $c_{i_1 i_2}^{(012)}$ 表示 d_0 的所有 m_2 个三水平因子在第 i_1 行与 d_1 的所有 m_2 个三水平因子在第 i_2 行的相遇数.

根据 (2.2) 和 (2.3), 可以得到拓展设计 d^E 的 Lee 偏差的两个不同的表达式, 它们清晰地刻画了初始设计和附加设计的 Lee 偏差与拓展设计的 Lee 偏差之间的关系, 这为构造均匀的二三混水平拓展设计提供了理论依据.

定理 3.1 给定任意拓展设计 $d^E \in \mathcal{D}_g(N+r; 2^{m_1} \times 3^{m_2})$, 其相应的初始设计为 $d_0 \in \mathcal{D}_0(N, 2^{m_1} \times 3^{m_2})$, 附加设计为 $d_1 \in \mathcal{D}_1(r, 2^{m_1} \times 3^{m_2})$, 则有

$$[LD_2(d^E)]^2 = \frac{2Nr}{(N+r)^2}K + \frac{N^2}{(N+r)^2}[LD_2(d_0)]^2 + \frac{r^2}{(N+r)^2}[LD_2(d_1)]^2 + \frac{2}{(N+r)^2}n_{d_0}^T M^m n_{d_1}, \quad (3.1)$$

以及

$$[LD_2(d^E)]^2 = \frac{2Nr}{(N+r)^2}K + \frac{N^2}{(N+r)^2}[LD_2(d_0)]^2 + \frac{r^2}{(N+r)^2}[LD_2(d_1)]^2 + \frac{2}{(N+r)^2} \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{i_1=1}^N \sum_{i_2=1}^r 2^{c_{i_1 i_2}^{(011)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(012)}}. \quad (3.2)$$

证明 注意到, d^E 中的全部试验点由 d_0 和 d_1 中的试验点构成, 故 $n_{d^E} = n_{d_0} + n_{d_1}$. 因此, 由 (2.2) 可知,

$$[LD_2(d^E)]^2 = K + \frac{1}{(N+r)^2} [n_{d_0}^T M^m n_{d_0} + 2n_{d_0}^T M^m n_{d_1} + n_{d_1}^T M^m n_{d_1}], \quad (3.3)$$

$$[LD_2(d_0)]^2 = K + \frac{n_{d_0}^T M^m n_{d_0}}{N^2} \quad \text{和} \quad [LD_2(d_1)]^2 = K + \frac{n_{d_1}^T M^m n_{d_1}}{r^2}. \quad (3.4)$$

结合 (3.3) 和 (3.4), 可知 (3.1) 成立.

同理, 根据相遇数情形 (i)–(iii), 由 (2.3), 经计算有

$$[LD_2(d^E)]^2 = K + \frac{1}{(N+r)^2} \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \left[\sum_{i_1=1}^N \sum_{i_2=1}^r 2^{c_{i_1 i_2}^{(01)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(02)}} + 2 \sum_{i_1=1}^N \sum_{i_2=1}^r 2^{c_{i_1 i_2}^{(011)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(012)}} + \sum_{i_1=1}^r \sum_{i_2=1}^r 2^{c_{i_1 i_2}^{(11)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(12)}} \right], \quad (3.5)$$

$$[LD_2(d_0)]^2 = K + \frac{1}{N^2} \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{i_1=1}^N \sum_{i_2=1}^r 2^{c_{i_1 i_2}^{(01)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(02)}} \quad (3.6)$$

和

$$[LD_2(d_1)]^2 = K + \frac{1}{r^2} \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{i_1=1}^r \sum_{i_2=1}^r 2^{c_{i_1 i_2}^{(11)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(12)}}. \quad (3.7)$$

结合 (3.5)–(3.7), 可知 (3.2) 成立. $\square$

在实践中, 试验者对正在研究的系统常常缺乏足够的先验知识. 因此, 仅通过一个试验有时很难达成预设的目标, 序贯试验是一个很好的选择. 由定理 3.1 中的 (3.1) 和 (3.2) 可知, 构造均匀的拓展设计的一个可行的方式是先选择初始设计 d_0 使得 $[LD_2(d_0)]^2$ 最小, 紧接着加入附加设计 d_1 , 使得 (3.1) 或 (3.2) 中等号右边的第 3 和 4 项尽可能小. 这也说明, 添加一个均匀的附加设计 d_1 , 即仅仅最小化 $[LD_2(d_1)]^2$ 是不够的, 这有可能会使得所得拓展设计不是最优的.

4 二三混水平拓展设计 Lee 偏差的下界

本节将给出二三混水平拓展设计 Lee 偏差的下界. 在评价拓展设计均匀性好坏时, 我们可以用这个下界作为一个有效的基准, 同时在采用算法搜索 (近似) 均匀的拓展设计时也可以用这个下界提供有用的信息, 从而提升算法的效率.

对设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} \times 3^{m_2})$, 其中 $n = 2l_1 + u_1 = 3l_2 + u_2$, l_1 和 l_2 都是非负整数并且 $u_1 \in \{0, 1\}$, $u_2 \in \{0, 1, 2\}$. 如果 $u_1 = 0$ 并且 $u_2 = 0$, 则 d_g 是一个 U -型设计, 此时 $g = 0$. 此外, 如果 $u_1 \neq 0$ 或者 $u_2 \neq 0$, 则 d_g 是一个近似 U -型设计, 其每个因子的各个水平出现的次数至多相差一次, 此时 $g = 1$. 根据文献 [12, 引理 6] 和 [13, 引理 3], 对于给定的设计, 不管这个设计是 U -型设计还是近似 U -型设计, 我们都可以建立其相遇数之间的关系. 这个关系为获得二三混水平拓展设计的 Lee 偏差的下界奠定了基础.

引理 4.1 对任意的设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} \times 3^{m_2})$, 有

$$\sum_{i_1=1}^n \sum_{i_2(\neq i_1)=1}^n c_{i_1 i_2}^{(g1)} = \frac{[n(n-2) + u_1]m_1}{2} \quad \text{且} \quad \sum_{i_1=1}^n \sum_{i_2(\neq i_1)=1}^n c_{i_1 i_2}^{(g2)} = \frac{[n(n-3) + u_2(3-u_2)]m_2}{3}.$$

为表示方便起见, 记

$$\begin{aligned} \omega_1^{(0)} &= \left\lfloor \frac{[n(n-2) + u_1]m_1}{2n(n-1)} \right\rfloor, \quad \omega_2^{(0)} = \left\lfloor \frac{[n(n-3) + u_2(3-u_2)]m_2}{3n(n-1)} \right\rfloor, \\ \nu_1^{(0)} &= \ln(2)(\omega_1^{(0)} + 1) + \ln\left(\frac{3}{2}\right)\omega_2^{(0)}, \quad \nu_2^{(0)} = \ln(2)\omega_1^{(0)} + \ln\left(\frac{3}{2}\right)(\omega_2^{(0)} + 1), \\ \nu_3^{(0)} &= \ln(2)\omega_1^{(0)} + \ln\left(\frac{3}{2}\right)\omega_2^{(0)}, \quad \nu_4^{(0)} = \ln(2)(\omega_1^{(0)} + 1) + \ln\left(\frac{3}{2}\right)(\omega_2^{(0)} + 1), \\ p_1^{(0)} &= n(n-1)(\omega_1^{(0)} + 1) - \frac{[n(n-2) + u_1]m_1}{2}, \quad q_1^{(0)} = \frac{[n(n-2) + u_1]m_1}{2} - n(n-1)\omega_1^{(0)}, \\ p_2^{(0)} &= n(n-1)(\omega_2^{(0)} + 1) - \frac{[n(n-3) + u_2(3-u_2)]m_2}{3}, \\ q_2^{(0)} &= \frac{[n(n-3) + u_2(3-u_2)]m_2}{3} - n(n-1)\omega_2^{(0)}. \end{aligned}$$

由引理 4.1 和上述记号, 根据受控 (majorization) 不等式理论, 可以得到下面的结论.

引理 4.2 对于任意的二三混水平设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} 3^{m_2})$ 和任意的正整数 l , 下述不等式成立:

$$\sum_{i_1=1}^n \sum_{i_2(\neq i_1)=1}^n \left[\ln(2)c_{i_1 i_2}^{(g1)} + \ln\left(\frac{3}{2}\right)c_{i_1 i_2}^{(g2)} \right]^l \geq \Delta_l,$$

其中

$$\begin{cases} q_1^{(0)}(\nu_1^{(0)})^l + q_2^{(0)}(\nu_2^{(0)})^l + (p_1^{(0)} - q_2^{(0)})(\nu_3^{(0)})^l, & \text{如果 } p_1^{(0)} > q_2^{(0)}, \\ p_2^{(0)}(\nu_1^{(0)})^l + p_1^{(0)}(\nu_2^{(0)})^l + (q_2^{(0)} - p_1^{(0)})(\nu_4^{(0)})^l, & \text{如果 } p_1^{(0)} \leq q_2^{(0)}. \end{cases}$$

定理 4.1 对任意的近似 U -型设计 $d_g \in \mathcal{D}_g(n, 2^{m_1} \times 3^{m_2})$, (2.3) 中 d_g 的 Lee 偏差有如下下界:

$$[LD_2(d_g)]^2 \geq LB_2^{(g)}(n, m_1, m_2),$$

其中

$$\begin{aligned} LB_2^{(g)}(n, m_1, m_2) &= K + \frac{1}{n} + \frac{1}{n^2} \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \Theta^{(0)}, \\ \Theta^{(0)} &= \begin{cases} q_1^{(0)}e^{\nu_1^{(0)}} + q_2^{(0)}e^{\nu_2^{(0)}} + (p_1^{(0)} - q_2^{(0)})e^{\nu_3^{(0)}}, & \text{如果 } p_1^{(0)} > q_2^{(0)}, \\ p_2^{(0)}e^{\nu_1^{(0)}} + p_1^{(0)}e^{\nu_2^{(0)}} + (q_2^{(0)} - p_1^{(0)})e^{\nu_4^{(0)}}, & \text{如果 } p_1^{(0)} \leq q_2^{(0)}. \end{cases} \end{aligned}$$

证明 记 (2.3) 等号右边第二项中最后部分为 Λ_1 , 并计算可得

$$\begin{aligned}\Lambda_1 &= \sum_{i_1=1}^n \sum_{i_2(\neq i_1)=1}^n 2^{c_{i_1 i_2}^{(g1)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(g2)}} \\ &= \sum_{i_1=1}^n \sum_{i_2(\neq i_1)=1}^n e^{\ln(2)c_{i_1 i_2}^{(g1)} + \ln(\frac{3}{2})c_{i_1 i_2}^{(g2)}} \\ &= \sum_{i_1=1}^n \sum_{i_2(\neq i_1)=1}^n \left\{ \sum_{l=0}^{\infty} \frac{[\ln(2)c_{i_1 i_2}^{(g1)} + \ln(\frac{3}{2})c_{i_1 i_2}^{(g2)}]^l}{l!} \right\} \\ &= \sum_{l=0}^{\infty} \frac{1}{l!} \sum_{i_1=1}^n \sum_{i_2(\neq i_1)=1}^n \left[\ln(2)c_{i_1 i_2}^{(g1)} + \ln\left(\frac{3}{2}\right)c_{i_1 i_2}^{(g2)} \right]^l.\end{aligned}$$

当 $p_1^{(0)} > q_2^{(0)}$ 时, 由引理 4.2, 可得

$$\begin{aligned}\Lambda_1 &\geq \sum_{l=0}^{\infty} \frac{1}{l!} [q_1^{(0)}(\nu_1^{(0)})^l + q_2^{(0)}(\nu_2^{(0)})^l + (p_1^{(0)} - q_2^{(0)})(\nu_3^{(0)})^l] \\ &= q_1^{(0)}e^{\nu_1^{(0)}} + q_2^{(0)}e^{\nu_2^{(0)}} + (p_1^{(0)} - q_2^{(0)})e^{\nu_3^{(0)}}.\end{aligned}$$

同理, 当 $p_1^{(0)} \leq q_2^{(0)}$ 时, 有

$$\Lambda_1 \geq p_2^{(0)}e^{\nu_1^{(0)}} + p_1^{(0)}e^{\nu_2^{(0)}} + (q_2^{(0)} - p_1^{(0)})e^{\nu_4^{(0)}}.$$

至此, 我们证明了定理中的全部结论. $\square$

注意到, $\mathcal{D}_0(n, 2^{m_1} \times 3^{m_2})$ 是 $\mathcal{D}_1(n, 2^{m_1} \times 3^{m_2})$ 的特殊情形, 因此, 定理 4.1 中给出的下界对于常见的 U -型设计也成立. 这个下界与引理 2.1 中给出的二三混水平设计的 Lee 偏差的下界相比, 当设计是近似等距设计, 即设计中任意两个水平组合的相遇数之间的差至多为 1 时, 新的下界表现更好. 当设计为饱和设计或者超饱和设计时, 新的下界表现也较好. 当设计是一个强度较高的正交表时, 引理 2.1 中的下界往往要比新的下界更紧, 但是这种情况下设计需要很多的试验次数并且所能容纳的因子数量很有限.

文献 [7–11] 在广义 I -型最优、Schur-最优和 GMA 准则下分别给出了在正交表上添加一些新试验点的许多结果. 当选择的初始设计 d_0 为正交表时, 采用 Lee 偏差度量设计的均匀性, 我们也可以得到一些有意义的结论, 参见下述定理 4.2.

引理 4.3 给定拓展设计 $d^E \in \mathcal{D}^E(N+r, 2^{m_1} \times 3^{m_2})$, 其相应的初始设计为 $d_0 \in \mathcal{D}_0(N, 2^{m_1} \times 3^{m_2})$, 附加设计为 $d_1 \in \mathcal{D}_1(r, 2^{m_1} \times 3^{m_2})$. 如果 d_0 是强度为 h 的正交表 $OA(N, m_1 + m_2, 2^{m_1} \times 3^{m_2}, h)$, 则

$$n_{d_0}^T M^m n_{d_1} \geq B_1, \quad (4.1)$$

其中

$$B_1 = r \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{j_1+j_2 \leq h} \left(\frac{1}{2}\right)^{j_2} \binom{m_1}{j_1} \binom{m_2}{j_2} \omega_{j_1 j_2},$$

且 $\omega_{j_1 j_2} = n/(2^{j_1} 3^{j_2})$.

证明 记

$$\Omega^{(i)} = \{u^{(i)} = (u_1^{(i)}, u_2^{(i)}, \dots, u_{m_i}^{(i)}) \mid u_j^{(i)} = 0, 1, 1 \leq j \leq m_i\}, \quad i = 1, 2,$$

$$\Omega_j^{(i)} = \left\{ u^{(i)} \in \Omega^{(i)} \mid \sum_{\alpha=1}^{m_i} u_\alpha^{(i)} = j, \right\}, \quad 1 \leq j \leq m_i, \quad i = 1, 2.$$

不难发现

$$M^m = \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{j_1=0}^{m_1} \sum_{j_2=0}^{m_2} \left(\frac{1}{2}\right)^{j_2} \sum_{u^{(1)} \in \Omega_{j_1}^{(1)}} \sum_{u^{(2)} \in \Omega_{j_2}^{(2)}} W^T(u^{(1)}, u^{(2)}) W(u^{(1)}, u^{(2)}),$$

其中 $W(u^{(1)}, u^{(2)}) = (\otimes_{j_1=1}^{m_1} L_1(u_{j_1}^{(1)})) \otimes (\otimes_{j_2=1}^{m_2} L_2(u_{j_2}^{(2)}))$. 因此,

$$n_{d_0}^T M^m n_{d_1} = \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{j_1=0}^{m_1} \sum_{j_2=0}^{m_2} \left(\frac{1}{2}\right)^{j_2} \sum_{u^{(0)} \in \Omega_{j_1}^{(1)}} \sum_{u^{(2)} \in \Omega_{j_2}^{(1)}} (w_{j_1 j_2}^{(0)})^T w_{j_1 j_2}^{(1)},$$

其中 $w_{j_1 j_2}^{(0)} = W(u^{(1)}, u^{(2)}) n_{d_0}$ 且 $w_{j_1 j_2}^{(1)} = W(u^{(1)}, u^{(2)}) n_{d_1}$.

注意到, $w_{j_1 j_2}^{(0)}$ 和 $w_{j_1 j_2}^{(1)}$ 都是由 $2^{j_1} 3^{j_2}$ 个非负分量构成的向量, 并且这些分量之和分别为 N 和 r . 此外, 如果 d_0 是强度为 h 的正交表 $OA(N, m_1 + m_2, 2^{m_1} \times 3^{m_2}, h)$, 则对 $0 \leq j_1 + j_2 \leq h$, 有 $w_{j_1 j_2}^{(0)} = (\frac{n}{2^{j_1} 3^{j_2}}) 1_{2^{j_1} 3^{j_2}}$, 其中 1_k 是每个分量都为 1 的 $k \times 1$ 阶向量. 因此,

$$n_{d_0}^T M^m n_{d_1} \geq r \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{j_1+j_2 \leq h} \left(\frac{1}{2}\right)^{j_2} \binom{m_1}{j_1} \binom{m_2}{j_2} \omega_{j_1 j_2}.$$

引理中不等式得证. $\square$

引理 4.4 给定任意的拓展设计 $d^E = (d_0 : d_1) \in \mathcal{D}^E(N + r, 2^{m_1} \times 3^{m_2})$, 其相应的初始设计为 $d_0 \in \mathcal{D}_0(N, 2^{m_1} \times 3^{m_2})$, 附加设计为 $d_1 \in \mathcal{D}_1(r, 2^{m_1} \times 3^{m_2})$, 则有

$$\sum_{i_1=1}^N \sum_{i_2=1}^r c_{i_1 i_2}^{(011)} = \frac{N r m_1}{2} \quad \text{且} \quad \sum_{i_1=1}^N \sum_{i_2=1}^r c_{i_1 i_2}^{(012)} = \frac{N r m_2}{3}.$$

证明 我们仅证明引理中的第一个等式, 同理可证第二个等式成立. 对设计 $d_0 \in \mathcal{D}_0(N, 2^{m_1} \times 3^{m_2})$, 任意选定 d_0 的一个二水平因子, 其每个水平都出现了 $N/2$ 次. 给定水平 α_1 , 其中 $\alpha_1 \in \{0, 1\}$, 对设计 d_1 中对应于上述 d_0 中被选到的因子而言, 记其出现的次数为 $\eta(\alpha_1)$. 因此,

$$\sum_{i_1=1}^N \sum_{i_2=1}^r c_{i_1 i_2}^{(011)} = \sum_{k_1=1}^{m_1} \sum_{\alpha_1=0}^1 \frac{N}{2} \eta(\alpha_1) = \frac{N m_1}{2} \sum_{\alpha_1=0}^1 \eta(\alpha_1) = \frac{N r m_1}{2}.$$

引理中第一个等式得证. $\square$

记 $\nu_1^{(01)} = \ln(2)(\omega_1^{(01)} + 1) + \ln(\frac{3}{2})\omega_2^{(01)}$, $\nu_2^{(01)} = \ln(2)\omega_1^{(01)} + \ln(\frac{3}{2})(\omega_2^{(01)} + 1)$, $\omega_1^{(01)} = \lfloor \frac{m_1}{2} \rfloor$, $\omega_2^{(01)} = \lfloor \frac{m_2}{3} \rfloor$, $\nu_3^{(01)} = \ln(2)\omega_1^{(01)} + \ln(\frac{3}{2})\omega_2^{(01)}$, 且 $\nu_4^{(01)} = \ln(2)(\omega_1^{(01)} + 1) + \ln(\frac{3}{2})(\omega_2^{(01)} + 1)$. 此外, 设

$$p_1^{(01)} = N r (\omega_1^{(01)} + 1) - \frac{N r m_1}{2}, \quad q_1^{(01)} = \frac{N r m_1}{2} - N r \omega_1^{(01)},$$

$$p_2^{(01)} = N r (\omega_2^{(01)} + 1) - \frac{N r m_2}{3}, \quad q_2^{(01)} = \frac{N r m_2}{3} - N r \omega_2^{(01)}.$$

从 (3.1) 和 (3.2), 不难发现

$$n_{d_0}^T M^m n_{d_1} = \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} \sum_{i_1=1}^N \sum_{i_2=1}^r 2^{c_{i_1 i_2}^{(011)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(012)}}.$$

因此, 当初始设计 d_0 仅仅是强度为 1 的正交表时, 我们可以建立 $n_{d_0}^T M^m n_{d_1}$ 的一个下界.

引理 4.5 对任意拓展设计 $d^E \in \mathcal{D}^E(N+r, 2^{m_1} \times 3^{m_2})$, 其相应的初始设计为 $d_0 \in \mathcal{D}_0(N, 2^{m_1} \times 3^{m_2})$, 附加设计为 $d_1 \in \mathcal{D}_1(r, 2^{m_1} \times 3^{m_2})$, 下述不等式成立:

$$\sum_{i_1=1}^N \sum_{i_2=1}^r 2^{c_{i_1 i_2}^{(011)}} \left(\frac{3}{2}\right)^{c_{i_1 i_2}^{(012)}} \geq B_2, \quad (4.2)$$

其中

$$B_2 = \begin{cases} q_1^{(01)} e^{\nu_1^{(01)}} + q_2^{(01)} e^{\nu_2^{(01)}} + (p_1^{(01)} - q_2^{(01)}) e^{\nu_3^{(01)}}, & \text{若 } p_1^{(01)} > q_2^{(01)}, \\ p_2^{(01)} e^{\nu_1^{(01)}} + p_1^{(01)} e^{\nu_2^{(01)}} + (q_2^{(01)} - p_1^{(01)}) e^{\nu_4^{(01)}}, & \text{若 } p_1^{(01)} \leq q_2^{(01)}. \end{cases}$$

根据引理 2.1 和定理 4.1, 记

$$LB^{(0)} = \max\{LB_1^{(0)}(N, m_1, m_2), LB_2^{(0)}(N, m_1, m_2)\}, \quad (4.3)$$

且

$$LB^{(1)} = \max\{LB_1^{(1)}(r, m_1, m_2), LB_2^{(1)}(r, m_1, m_2)\}. \quad (4.4)$$

由 (3.2)–(4.4), 我们可以给出拓展设计 d^E 的 Lee 偏差 $[LD_2(d^E)]^2$ 的下界.

定理 4.2 假设拓展设计 $d^E = (d_0 : d_1) \in \mathcal{D}^E(N+r, 2^{m_1} \times 3^{m_2})$ 由初始设计 $d_0 \in \mathcal{D}_0(N, 2^{m_1} \times 3^{m_2})$ 和附加设计 $d_1 \in \mathcal{D}_1(r, 2^{m_1} \times 3^{m_2})$ 构成, 如果 d_0 是强度为 h 的正交表 $OA(N, m_1 + m_2, 2^{m_1} \times 3^{m_2}, h)$, 那么,

$$[LD_2(d^E)]^2 \geq LB_1,$$

其中

$$LB_1 = \frac{2Nr}{(N+r)^2} K + \frac{N^2}{(N+r)^2} LB^{(0)} + \frac{r^2}{(N+r)^2} LB^{(1)} + \frac{2}{(N+r)^2} B_1.$$

定理 4.2 中给出的下界要求初始设计 d_0 是强度为 h 的正交表. 因此, 拓展设计中所能容纳的因子数量很有限, 最多能容纳的因子个数受著名的 Rao 上界控制. 更重要的是, 并不是满足这些约束条件的正交表都存在, 即使存在构造起来也不容易. 在实践中, 当需要考虑较多的因子时, 饱和设计或超饱和设计更加节省成本. 这时上述的下界可能表现较差, 下面将试图提供一个这种情形下较好的下界.

定理 4.3 对任意拓展设计 $d^E \in \mathcal{D}^E(N+r, 2^{m_1} \times 3^{m_2})$, 其相应的初始设计为 $d_0 \in \mathcal{D}_0(N, 2^{m_1} \times 3^{m_2})$, 附加设计为 $d_1 \in \mathcal{D}_1(r, 2^{m_1} \times 3^{m_2})$, 则有

$$[LD_2(d^E)]^2 \geq LB_2,$$

其中

$$LB_2 = \frac{2Nr}{(N+r)^2} K + \frac{N^2}{(N+r)^2} LB^{(0)} + \frac{r^2}{(N+r)^2} LB^{(1)} + \frac{2}{(N+r)^2} \left(\frac{1}{2}\right)^{m_1} \left(\frac{2}{3}\right)^{m_2} B_2.$$

定理 4.2 和 4.3 中的下界各有优势, 前者当初始设计的强度较高时表现更佳, 后者当设计的相遇数尽可能相等时则更紧. 因此, 注意到 U -型设计是强度为 1 的正交表, 在实践中, 我们建议取拓展设计的 Lee 偏差的下界为 $LB = \max\{LB_1, LB_2\}$. 此外, 在许多情形下, 定理 4.2 和 4.3 中的下界都不能达到, 但可以将其作为评价拓展设计均匀性好坏的有用基准. 更准确地说, 定义

$$\text{Eff}_1(d^E) = \frac{LB_1}{[LD_2(d^E)]^2} \quad \text{且} \quad \text{Eff}_2(d^E) = \frac{LB_2}{[LD_2(d^E)]^2}. \quad (4.5)$$

如果 $\text{Eff}_1 = 1$ 或者 $\text{Eff}_2 = 1$, 则拓展设计 d^E 在 Lee 偏差下的均匀性是最好的. 否则, Eff_1 或者 Eff_2 的取值越接近于 1, 拓展设计的均匀性越好. 记 $\text{Eff}(\cdot) = \max\{\text{Eff}_1(\cdot), \text{Eff}_2(\cdot)\}$, 为了便于实际应用, 我们推荐下述随机算法构造好的拓展设计 (见算法 1), 它能够有效地避免拓展设计的偏差取得局部极小值.

算法 1 (近似) 最优拓展设计的搜索算法

1. 初始化参数迭代次数 τ 、容忍门限 δ 和接受概率 μ_1 .
 2. 初始化初始设计 d_0 , 附加设计 $d_1, d_1^c := d_1, d^E := (d_0 : d_1)$.
 3. **for** $i = 1$ **to** τ **do**
 4. 生成新的附加设计 $d_1^{\text{new}}, d_1^{\text{new}} \in \mathcal{N}(d_1^c)$ (当前附加设计 d_1^c 的领域), $d_1^c := d_1^{\text{new}}, d_{\text{new}}^E := (d_0 : d_1^{\text{new}})$.
 5. 计算拓展设计的效率增量 $\nabla = \text{Eff}(d_{\text{new}}^E) - \text{Eff}(d^E)$, 并从 $[0, 1]$ 的均匀分布生成随机数 μ .
 6. **if** $\nabla < 0$ 或者 $(\nabla < \delta \text{ 且 } \mu < \mu_1)$,
 7. **then** 更新当前的附加设计 $d_1 := d_1^{\text{new}}, d^E := (d_0 : d_1)$.
 8. **end for**
 9. 输出最终的拓展设计 $d^E = (d_0 : d_1)$.
-

5 数值例子

下面的例子进一步解释了我们所建立的理论结果. 相对于折叠反转设计或者半折叠反转设计而言, 拓展设计更加灵活, 试验者可以在已经选定的设计中灵活追加新的试验点, 而不是仅仅只能添加与原有设计试验次数相同的试验点或者原有设计试验次数一半的试验点. 折叠反转设计重在解除别名, 而且受折叠反转结构的限制, 所得的拓展设计可能并不是最优的. 下面这些例子说明我们建立的下界在一些情形下是紧的.

例 5.1 假如试验者用 18 次试验的均匀 U -型设计作为初始设计 d_0 进行试验, 其设计见表 1. 在试验结束后, 试验者决定再进行 r 次新试验, 其中 $r = 2, 3, 6$. 若采用 Lee 偏差度量设计的优良性, 则下述几个设计是很好的选择.

(1) 当需要追加 2 次试验时, 可采用拓展设计 $d^{E^1} = (d_0 : d_1^{(1)}) \in \mathcal{D}^E(18+2, 2^1 \times 3^{12})$, 其中附加设计 $d_1^{(1)}$ 见表 2.

(2) 当需要追加 3 次试验时, 可采用拓展设计 $d^{E^2} = (d_0 : d_1^{(2)}) \in \mathcal{D}^E(18+3, 2^1 \times 3^{12})$, 其中附加设计 $d_1^{(2)}$ 见表 3.

(3) 当需要追加 6 次试验时, 可采用拓展设计 $d^{E^3} = (d_0 : d_1^{(3)}) \in \mathcal{D}^E(18+6, 2^1 \times 3^{12})$, 其中附加设计 $d_1^{(3)}$ 见表 4.

对于拓展设计 d^{E^1} 和 d^{E^2} , 计算可得它们的效率分别为 $\text{Eff}_2(d^{E^1}) = 1$ 和 $\text{Eff}_2(d^{E^2}) = 1$, 这说明定理 4.3 中给出的下界是可以达到的. 因此, 拓展设计 d^{E^1} 和 d^{E^2} 分别在设计类 $\mathcal{D}^E(18+2, 2^1 \times 3^{12})$ 和 $\mathcal{D}^E(18+3, 2^1 \times 3^{12})$ 是均匀的. 对于拓展设计 d^{E^3} 而言, 定理 4.3 给出的下界没有达到, 但是其效率为 $\text{Eff}_2(d^{E^3}) = 94.74\%$, 这说明拓展设计 d^{E^3} 在设计类 $\mathcal{D}^E(18+6, 2^1 \times 3^{12})$ 中具有较好的均匀性.

表 1 初始设计 d_0

0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	0	1	1	1	0	1	1	1
0	0	2	2	2	0	2	2	2	0	2	2	2
0	1	0	1	2	1	0	1	2	1	0	1	2
0	1	1	2	0	1	1	2	0	1	1	2	0
0	1	2	0	1	1	2	0	1	1	2	0	1
0	2	0	2	1	2	0	2	1	2	0	2	1
0	2	1	0	2	2	1	0	2	2	1	0	2
0	2	2	1	0	2	2	1	0	2	2	1	0
1	0	0	0	0	1	1	1	1	2	2	2	2
1	0	1	1	1	1	2	2	2	2	0	0	0
1	0	2	2	2	1	0	0	0	2	1	1	1
1	1	0	1	2	2	1	2	0	0	2	0	1
1	1	1	2	0	2	2	0	1	0	0	1	2
1	1	2	0	1	2	0	1	2	0	1	2	0
1	2	0	2	1	0	1	0	2	1	2	1	0
1	2	1	0	2	0	2	1	0	1	0	2	1
1	2	2	1	0	0	0	2	1	1	1	0	2

表 2 2 次试验的附加设计 $d_1^{(1)}$

0	1	0	0	2	0	2	2	1	2	1	1	0
1	2	1	2	0	1	0	1	2	0	2	0	1

表 3 3 次试验的附加设计 $d_1^{(2)}$

0	0	0	0	1	2	2	2	0	1	1	1	2
0	1	1	1	0	0	0	0	2	2	2	2	1
1	2	2	2	2	1	1	1	1	0	0	0	0

表 4 6 次试验的附加设计 $d_1^{(3)}$

0	0	0	1	2	2	2	0	1	1	0	2	1
0	1	2	0	0	2	2	1	0	0	1	0	2
0	2	1	2	0	0	0	1	2	1	2	1	1
1	2	1	1	1	1	1	2	0	0	2	2	2
1	1	0	2	1	1	0	0	1	2	1	0	0
1	0	2	0	2	0	1	2	2	2	0	1	0

例 5.2 假设试验者选择初始设计 $d_0 \in \mathcal{D}_0(12, 2^{11} \times 3^{11})$ 开始执行试验, 其中 d_0 见下表 5. 当试验进行的过程中或者试验结束后, 出于某种原因试验者需要追加新的试验. 设需要追加的试验次数为 r , 其中 $r = 2, 3, 6$. 若采用 Lee 偏差评价整个试验所用的拓展设计的优劣, 则选择下述拓展设计是比较理想的.

(1) 需要追加 2 次试验时, 推荐拓展设计 $d^{E^1} = (d_0 : d_1^{(1)}) \in \mathcal{D}^E(12+2, 2^{11} \times 3^{11})$, 其中附加设计 $d_1^{(1)}$ 见表 6.

(2) 需要追加 3 次试验时, 推荐拓展设计 $d^{E^2} = (d_0 : d_1^{(2)}) \in \mathcal{D}^E(12+3, 2^{11} \times 3^{11})$, 其中附加设计 $d_1^{(2)}$ 见表 7.

(3) 需要追加 6 次试验时, 推荐拓展设计 $d^{E^3} = (d_0 : d_1^{(3)}) \in \mathcal{D}^E(12+6, 2^{11} \times 3^{11})$, 其中附加设计 $d_1^{(3)}$ 见表 8.

对于拓展设计 d^{E^1} 而言, 其效率为 $\text{Eff}_2(d^{E^1}) = 1$, 这说明定理 4.3 中给出的下界是可以达到的. 因此, 拓展设计 d^{E^1} 在设计类 $\mathcal{D}^E(12+2, 2^{11} \times 3^{11})$ 是最优的. 对于二三混水平拓展设计 d^{E^2} 和 d^{E^3}

表 5 初始设计 d_0

1	1	0	1	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	
0	1	1	0	1	1	1	0	0	0	1	0	0	0	1	0	1	1	0	0	2	2
1	0	1	1	0	1	1	1	0	0	0	0	0	1	0	0	2	2	2	2	0	0
0	1	0	1	1	0	1	1	1	0	0	0	2	2	2	2	0	0	0	0	0	0
0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	2	0	1	2	2	1	0
0	0	0	1	0	1	1	0	1	1	1	1	1	0	2	1	2	2	1	0	0	2
1	0	0	0	1	0	1	1	0	1	1	1	2	1	0	2	1	0	1	2	2	2
1	1	0	0	0	1	0	1	1	0	1	1	2	1	2	0	2	1	0	1	1	1
1	1	1	0	0	0	1	0	1	1	0	2	0	2	2	2	1	2	2	1	1	2
0	1	1	1	0	0	0	1	0	1	1	2	2	1	1	1	0	2	2	0	2	1
1	0	1	1	1	0	0	0	1	0	1	2	1	2	0	1	1	1	0	2	0	1
0	0	0	0	0	0	0	0	0	0	0	2	1	2	1	0	2	0	1	1	2	0

表 6 2 次试验的附加设计 $d_1^{(1)}$

0	0	1	1	1	0	0	0	0	1	2	0	2	1	2	0	2	1	0	1	1	2
1	1	2	0	0	1	1	1	1	0	1	2	0	2	1	1	0	2	1	2	2	0

表 7 3 次试验的附加设计 $d_1^{(2)}$

1	0	1	1	1	1	0	0	0	1	0	1	2	2	2	0	2	0	0	0	2	1
1	1	0	1	0	0	1	1	0	0	1	2	1	0	1	2	0	1	1	2	1	2
0	0	0	0	1	0	0	1	1	1	1	0	0	1	0	1	1	2	2	1	0	0

表 8 6 次试验的附加设计 $d_1^{(3)}$

1	1	0	1	1	0	1	0	0	0	0	1	2	0	1	2	2	1	2	0	0	0
1	1	1	0	0	1	0	1	1	0	0	1	1	2	0	1	1	0	1	0	1	2
1	0	0	1	0	1	0	0	0	1	1	2	2	2	2	2	0	1	0	2	1	0
0	0	0	0	1	1	1	1	0	1	0	2	0	1	2	1	1	2	0	1	2	1
0	1	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	2	1	1	2	2
0	0	1	0	0	0	1	0	1	1	1	0	1	1	1	0	2	0	2	2	0	1

而言, 其效率分别为 $\text{Eff}_2(d^{E^2}) = 99.85\%$, $\text{Eff}_2(d^{E^3}) = 99.6\%$, 这表明定理 4.3 中给出的下界没有达到. 但是, 它们的效率都非常接近 1, 故设计 d^{E^2} 和 d^{E^3} 分别在设计类 $\mathcal{D}^E(12+3, 2^{11} \times 3^{11})$ 和 $\mathcal{D}^E(12+6, 2^{11} \times 3^{11})$ 上都有很好的均匀性.

6 结论

本文主要研究了二三混水平拓展设计在 Lee 偏差下的均匀性问题, 所得的结论为有效地设计序贯试验以及构造好的拓展设计提供了理论依据. 与文献中常用的折叠反转技术或者半折叠反转技术相比, 拓展设计允许试验者灵活追加试验次数. 本文建立了拓展设计的 Lee 偏差与其初始设计和附加设计的偏差之间的解析联系. 本文结论表明, 在 Lee 偏差下, 先选择一个均匀的初始设计, 紧接着再选择好的附加设计使得拓展设计的 Lee 偏差尽可能小是一个可行的构造好的拓展设计的方案. 此外, 为了比较不同二三混水平拓展设计均匀性的好坏, 本文给出了拓展设计的 Lee 偏差的下界. 需要指出的是, 在拓展设计类 $\mathcal{D}^E(N+r, 2^{m_1} \times 3^{m_2})$ 上均匀的设计, 其 Lee 偏差在设计类 $\mathcal{D}(N+r, 2^{m_1} \times 3^{m_2})$ 中不一定取最小值. 但是, 存在某些拓展设计, 它们在 $\mathcal{D}(N+r, 2^{m_1} \times 3^{m_2})$ 也是最优的. 在实践中, 当均匀的初始设计不容易获得时, 我们可以通过两阶段启发式算法来优化效率函数 $\text{Eff}_i(\cdot)$, $i = 1, 2$, 从而得到较好的拓展设计. 此外, 怎样评价和构造一般的混水平拓展设计有待进一步研究.

致谢 非常感谢编委以及审稿人提出的宝贵意见和建议.

参考文献

- 1 Bernardo M C, Buck R, Liu L, et al. Integrated circuit design optimization using a sequential strategy. *IEEE Trans Comput Aided Des Integr Circuits Syst*, 1992, 11: 361–372
- 2 Ranjan P, Bingham D, Michailidis G. Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 2008, 50: 527–541
- 3 Jala M, Levy-Leduc C, Moulines E, et al. Sequential design of computer experiments for parameter estimation with application to numerical dosimetry. In: *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. Bucharest: IEEE, 2012, 909–913
- 4 Durrieu G, Briollais L. Sequential design for microarray experiments. *J Amer Statist Assoc*, 2009, 104: 650–660
- 5 Jala M, Levy-Leduc C, Moulines E, et al. Sequential design of computer experiments for the assessment of fetus exposure to electromagnetic fields. *Technometrics*, 2016, 58: 30–42
- 6 Ji Y B, Alaerts G, Xu C J, et al. Sequential uniform designs for fingerprints development of Ginkgo biloba extracts by capillary electrophoresis. *J Chromatography A*, 2006, 1128: 273–281
- 7 Mukerjee R. On the optimality of orthogonal array plus one run plans. *Ann Statist*, 1999, 27: 82–93
- 8 Butler N A, Ramos V M. Optimal additions to and deletions from two-level orthogonal arrays. *J R Stat Soc Ser B Stat Methodol*, 2007, 69: 51–61
- 9 Bird E M, Street D J. D-optimal asymmetric orthogonal array plus p run designs. *J Statist Plann Inference*, 2016, 170: 64–76
- 10 Zhang R C, Mukerjee R. Minimum aberration designs for two-level factorials in $n = 1 \pmod{4}$ runs. *Statist Sinica*, 2013, 23: 853–872
- 11 Huda S, Mukerjee R. Two-level minimum aberration designs in $N = 2 \pmod{4}$ runs. *J Statist Plann Inference*, 2013, 143: 1872–1879
- 12 Fang K T, Li R Z, Sudjianto A. *Design and Modeling for Computer Experiments*. London: Chapman & Hall, 2005
- 13 Fang K T, Liu M Q, Qin H, et al. *Theory and Application of Uniform Experimental Designs*. Lecture Notes in Statistics, vol. 221. New York: Springer, 2018
- 14 Qin H, Gou T X, Chatterjee K. A new class of two-level optimal extended designs. *J Korean Statist Soc*, 2016, 45: 168–175
- 15 Gou T X, Qin H, Chatterjee K. A new extension strategy on three-level factorials under wrap-around L_2 -discrepancy. *Comm Statist Theory Methods*, 2017, 46: 9063–9074

- 16 Yang F, Zhou Y D, Zhang A J. Mixed-level column augmented uniform designs. *J Complexity*, 2019, 53: 23–39
- 17 Jones B, Majumdar D. Optimal supersaturated designs. *J Amer Statist Assoc*, 2014, 109: 1592–1600
- 18 Zhou Y D, Ning J H, Song X B. Lee discrepancy and its applications in experimental designs. *Statist Probab Lett*, 2008, 78: 1933–1942
- 19 Chatterjee K, Qin H, Zou N. Lee discrepancy on asymmetrical factorials with two- and three-levels. *Sci China Math*, 2012, 55: 663–670

Optimal additions to asymmetrical fractional factorial designs under Lee discrepancy

Tingxun Gou, Hong Qin & Kashinath Chatterjee

Abstract Due to the character that the sequential experiment enlarges stage by stage, the new runs will be added into the previous design that has been chosen. How to design a good sequential experiment becomes an interesting issue. This paper studies the uniform asymmetric extended designs in terms of Lee discrepancy for a possible application to sequential experimentation. An analytical relation between Lee discrepancy of two- and three-level mixed extended designs, and its initial and additional designs is established, which provides a statistical justification to construct (nearly) uniform extended designs. As a benchmark for choosing best extended designs, a lower bound of Lee discrepancy of extended designs is obtained. In addition, some examples are also presented to further support our theoretical findings.

Keywords uniformity, two- and three-level mixed design, extended design, followup experiment

MSC(2010) 62K15, 62K10, 62K99

doi: 10.1360/N012019-00095