4.3 空间填充设计

王正明　易泰河

系统工程学院　军事建模与仿真系

2020 年 1 月 3 日

引言

例

研究 y 与 $x \in [-1, 1]$ 之间的关系. 由最优回归设计的理论可知:

- 如果已知模型为 $y = \beta_0 + \beta_1 x + \varepsilon$, 则所有试验点都集中在 -1 和 1 处;
- 如果已知模型为 $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$, 则所有试验点都集中在 -1, 0 和 1 处;
- 如果已知模型为 $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$, 则所有试验点都集中在 -1, -0.44720, 0.4472 和 1 处.

如果模型的形式未知, 应该如何安排试验.

引言

计算机试验设计的原则:

- 不需要考虑重复、区组和随机化;
- 获得的数据应能够拟合多种响应模型, 即试验方案对模型具有稳健性.

稳健性与效率是一对均衡关系:

- 模型已知的情况下, 最优回归设计效率最高;
- 模型未知或存在模型偏差的情况下, 空间填充设计最稳健.

引言

计算机试验设计的原则:

- 不需要考虑重复、区组和随机化;
- 获得的数据应能够拟合多种响应模型, 即试验方案对模型具有稳健性.

稳健性与效率是一对均衡关系:

- 模型已知的情况下, 最优回归设计效率最高;
- 模型未知或存在模型偏差的情况下, 空间填充设计最稳健.

引言

- 考查 p 个定量因子 $\boldsymbol{x} = (x_1, \cdots, x_p)^{\mathrm{T}}$ 对响应 y 的影响, 设试验区域 $\mathcal{X} = [0,1]^p$ 为标准立方体;
- 以 $\xi_n = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 表示包含 n 个点的试验设计, 以

$$\boldsymbol{D}_{\xi_n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

表示它的设计矩阵.

假定试验的目的是估计 $f(\boldsymbol{x})$ 在 $\mathcal{X} \in \mathbb{R}^p$ 上的均值

$$\mu = \mathbb{E}[f(\boldsymbol{x})] = \int_{\mathcal{X}} f(\boldsymbol{x})\,\mathrm{d}\,\boldsymbol{x},$$

称上式为总均值模型. 设设计为 $\xi_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$, 一个自然的想法是用试验点处响应值的均值

$$\hat{\mu} = \bar{y}(\xi_n) = \frac{1}{n}\sum_{i=1}^{n} y_i$$

来估计 μ, 式中 $y_i = f(\boldsymbol{x}_i)$. 以提高估计量 $\hat{\mu}$ 的精度为目的产生了一系列空间填充设计方法.

引言

空间填充设计的分类:

- 基于抽样的设计: 把 x 看作随机变量, 从其分布中抽取 n 个样本作为试验点.
- 基于准则的最优设计: 通过优化某种准则, 得到确定的最优设计.
- 随机和确定性混合的方法: 从确定性方法中选择一个设计, 然后进行某种随机化.

4.3 空间填充设计

4.3.1 基于抽样的设计

(1) 简单随机抽样

(2) 超拉丁方抽样

(3) 随机化正交阵列

4.3.2 基于准则的最优设计

4.3.3 均匀设计

- x 是分布已知的随机变量, 如何求 $y = f(x)$ 的分布?
- 理论分布难求, 通过抽取 x 的样本 $x_1, \cdots, x_n$, 计算 y 的样本 $y_i = f(x_i)$, 用 $\{y_1, \cdots, y_n\}$ 的经验分布来估计 y 的分布;
- 用 $\{y_1, \cdots, y_n\}$ 的样本均值和样本方差作为 y 的均值和方差的估计.

- $\boldsymbol{x}$ 是分布已知的随机变量, 如何求 $y = f(\boldsymbol{x})$ 的分布?
- 理论分布难求, 通过抽取 $\boldsymbol{x}$ 的样本 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$, 计算 y 的样本 $y_i = f(\boldsymbol{x}_i)$, 用 $\{y_1, \cdots, y_n\}$ 的经验分布来估计 y 的分布;
- 用 $\{y_1, \cdots, y_n\}$ 的样本均值和样本方差作为 y 的均值和方差的估计.

- 简单随机抽样: 从 $[0,1]^d$ 上的均匀分布中抽取试验点;
- 大数定律保证了由此得到的 $\hat{\mu}$ 是 μ 的无偏估计, 方差为 $\mathrm{Var}(y(\boldsymbol{x}))/n$;
- 简单随机抽样简单易行, 但表现不稳定.

4.3 空间填充设计

4.3.1 基于抽样的设计

(1) 简单随机抽样

(2) 超拉丁方抽样

(3) 随机化正交阵列

4.3.2 基于准则的最优设计

4.3.3 均匀设计

- 超拉丁方抽样是基于拉丁方设计的一种分层随机抽样方法.
- 可以给出总体均值的无偏估计, 且渐近方差比简单随机抽样要小.
- 拉丁方设计: n 阶拉丁方设计是一个由 n 各拉丁字母组成的 $n \times n$ 的方阵, 每个字母在每行中只出现一次, 在每列也只出现一次.

回顾: 什么是左循环拉丁方设计?

超拉丁方的构造:

Step 1 取 p 个独立的 $\{1,\cdots,n\}$ 的随机置换 $\pi_j(1),\cdots,\pi_j(n)$, $j=1,\cdots,p$, 将它们作为列向量组成一个 $n\times p$ 矩阵, 记为 $\mathrm{LHD}(n,p)$, 它的第 i 行 j 列的元素为 $\pi_j(i)$;

Step 2 取 $[0,1]$ 上的 np 个均匀分布的独立抽样, $u_{ij}\sim U(0,1)$, $i=1,\cdots,n$, $j=1,\cdots,p$. 令 $\boldsymbol{x}_i=(x_{i1},\cdots,x_{ip})^{\mathrm{T}}$, 其中

$$x_{ij}=\frac{\pi_j(i)-u_{ij}}{n},\quad i=1,\cdots,n,\quad j=1,\cdots,p.$$

则设计 $\xi_n=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$ 为一个超拉丁方设计, 并记作 $\mathrm{LHS}(n,p)$.

超拉丁方的构造:

Step 1 取 p 个独立的 $\{1,\cdots,n\}$ 的随机置换 $\pi_j(1),\cdots,\pi_j(n)$, $j=1,\cdots,p$, 将它们作为列向量组成一个 $n\times p$ 矩阵, 记为 $\mathrm{LHD}(n,p)$, 它的第 i 行 j 列的元素为 $\pi_j(i)$;

Step 2 取 $[0,1]$ 上的 np 个均匀分布的独立抽样, $u_{ij}\sim U(0,1)$, $i=1,\cdots,n$, $j=1,\cdots,p$. 令 $\boldsymbol{x}_i=(x_{i1},\cdots,x_{ip})^{\mathrm{T}}$, 其中

$$x_{ij}=\frac{\pi_j(i)-u_{ij}}{n},\quad i=1,\cdots,n,\quad j=1,\cdots,p.$$

则设计 $\xi_n=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$ 为一个超拉丁方设计, 并记作 $\mathrm{LHS}(n,p)$.

- 第一步保证在 n^p 个方格中随机选取 n 个方格，使得任一行和任一列都仅有一个方格被选中;
- 第二步中选随机数的目的是使得超拉丁方抽样能够取遍整个试验区域;
- 为了降低计算复杂度，一些学者建议取消第二步中 u_{ij} 的随机性，把 LHS 定义为格结构:

$$x_{ij} = \frac{\pi_j(i) - 0.5}{n}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, p.$$

即把试验点 x_i 取到小方块的中心. 这样得到的 LHS 称为中点超拉丁方抽样，并记为 $\mathrm{MLHS}(n, p)$.

- 第一步保证在 n^p 个方格中随机选取 n 个方格，使得任一行和任一列都仅有一个方格被选中；
- 第二步中选随机数的目的是使得超拉丁方抽样能够取遍整个试验区域；
- 为了降低计算复杂度，一些学者建议取消第二步中 u_{ij} 的随机性，把 LHS 定义为格结构：

$$x_{ij} = \frac{\pi_j(i) - 0.5}{n}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, p.$$

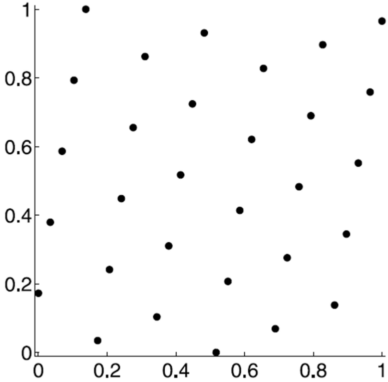
即把试验点 x_i 取到小方块的中心，这样得到的 LHS 称为中点超拉丁方抽样，并记为 $\mathrm{MLHS}(n, p)$.

- 第一步保证在 n^p 个方格中随机选取 n 个方格，使得任一行和任一列都仅有一个方格被选中；
- 第二步中选随机数的目的是使得超拉丁方抽样能够取遍整个试验区域；
- 为了降低计算复杂度，一些学者建议取消第二步中 u_{ij} 的随机性，把 LHS 定义为格结构：

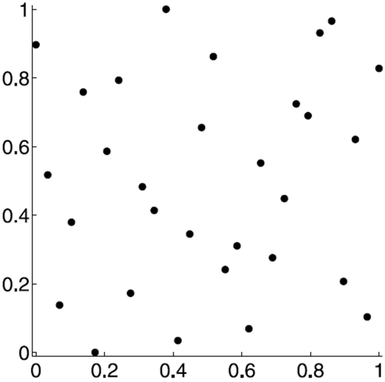
$$x_{ij} = \frac{\pi_j(i) - 0.5}{n}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, p.$$

即把试验点 $\boldsymbol{x}_i$ 取到小方块的中心。这样得到的 LHS 称为**中点超拉丁方抽样**，并记为 $\mathrm{MLHS}(n, p)$.

两因子 30 个点的超拉丁方设计:



中点超拉丁方设计



随机超拉丁方设计

- 超拉丁方抽样可以避免简单随机抽样最坏的情形, 其样本均值的方差也已证明比简单随机抽样小. 然而, 超拉丁方设计依然会出现比较极端的情形, 如所有试验点都在对角线上.

- 如果 p 太大而 n 太小, 由超拉丁方抽样的到的试验点在整个试验区域内过于稀疏, 有学者提出经验法则:
  - $n = 10p$ (Loeppky et al. 2009);
  - $n = 20p$ (Wang et al. 2014).

- 超拉丁方抽样可以避免简单随机抽样最坏的情形，其样本均值的方差也已证明比简单随机抽样小. 然而，超拉丁方设计依然会出现比较极端的情形，如所有试验点都在对角线上.
- 如果 p 太大而 n 太小，由超拉丁方抽样的到的试验点在整个试验区域内过于稀疏，有学者提出经验法则:
    - $n = 10p$ (Loeppky et al. 2009);
    - $n = 20p$ (Wang et al. 2014).

4.3 空间填充设计

4.3.1 基于抽样的设计

(1) 简单随机抽样

(2) 超拉丁方抽样

(3) 随机化正交阵列

4.3.2 基于准则的最优设计

4.3.3 均匀设计

- 随机化正交阵列是由一个正交阵列产生的超拉丁方设计，该类设计可以有效地减少估计方差.
- 一个试验次数为 n、因子个数为 p、水平数为 q 的部分因子设计称为强度为 r 的正交阵列，如果任意 $m(\leq r)$ 列都构成完全因子设计，记作 $OA(n, p, q, r)$.
- 根据定义，正交设计是强度为 ____ 的正交整列.

- 构造随机化正交阵列的步骤:

Step 1 选择合适的正交阵列 $OA(n,p,q,r)$, 记为 $\boldsymbol{A}$, 并设 $\lambda=n/q$;

Step 2 对 $\boldsymbol{A}$ 的每一列的 λ 个水平为 $k(k=1,\cdots,q)$ 的元素, 用 $\{(k-1)\lambda+1,(k-1)\lambda+2,\cdots,(k-1)\lambda+\lambda\}$ 的一个随机置换替代.

由此产生的超拉丁方设计即为随机化正交阵列, 记为 $OH(n,n^p)$.

- 正交阵列的平衡性保证了随机化正交阵列的稳定性比一般超拉丁方设计好.
- 然而, 只有部分特殊的 n, p, q 才存在强度为 $r(\geq 2)$ 的正交阵列, 因此这种方法不能构造任意试验次数 n 和任意因子个数 p 的超拉丁方.

例

从正交阵列 $OA(8,4,2,3)$ 出发构造 $OH(8,8^4)$.

No.	$OA(8,4,2,3)$				$OH(8,8^4)$			
1	1	1	1	1	2	4	3	3
2	1	1	2	2	1	1	6	7
3	1	2	1	2	3	7	4	8
4	1	2	2	1	4	5	8	4
5	2	1	1	2	6	3	1	5
6	2	1	2	1	5	2	7	1
7	2	2	1	1	7	6	2	2
8	2	2	2	2	8	8	5	6

第一列中, 4 个水平为 1 的位置替换为 $\{1,2,3,4\}$ 的随机置换 $\{2,1,3,4\}$, 4 个水平为 2 的位置替换为 $\{5,6,7,8\}$ 替换为 $\{6,5,7,8\}$; 第二列中, 4 个水平为 1 的位置替换为 $\{1,2,3,4\}$ 的随机置换 $\{4,1,3,2\}$, 4 个水平为 2 的位置替换为 $\{5,6,7,8\}$ 替换为 $\{7,5,6,8\}$; $\cdots$

4.3 空间填充设计

4.3.1 基于抽样的设计

4.3.2 基于准则的最优设计

4.3.3 均匀设计

- 首先构造某种空间填充准则, 然后设计优化算法求解最好的设计.
- 常见的准则:
  1. 基于试验点之间的距离, 如极大极小准则、极小极大准则;
  2. 基于分布之间的距离, 如均匀性准则.

基于点之间距离的准则:

- 包含 n 个点的极大极小设计$\xi_{\max\min}$ 满足

$$\min_{\boldsymbol{x}_1,\boldsymbol{x}_2\in\xi_{\min\max}} d(\boldsymbol{x}_1,\boldsymbol{x}_2)=\max_{\xi_n\subset\mathcal{X}}\min_{\boldsymbol{x}_1,\boldsymbol{x}_2\in\xi_n} d(\boldsymbol{x}_1,\boldsymbol{x}_2),$$

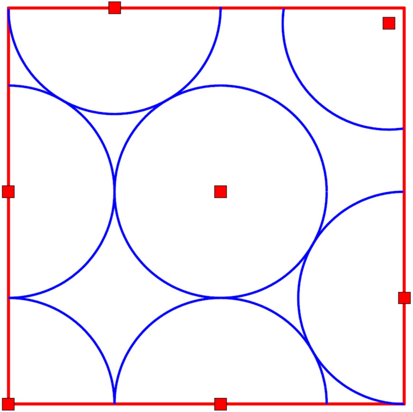
  即最大化设计点之间的最小距离;

- 包含 n 个点的极小极大设计$\xi_{\min\max}$ 满足

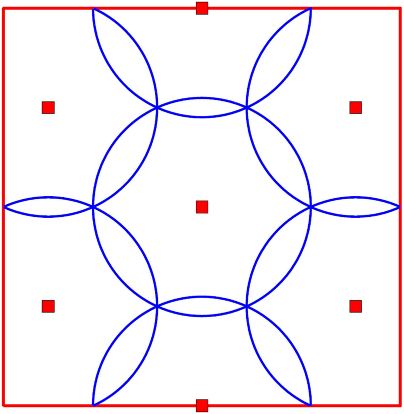
$$\max_{\boldsymbol{x}\in\mathcal{X}}\inf_{\boldsymbol{x}'\in\xi_{\min\max}} d(\boldsymbol{x},\boldsymbol{x}')=\min_{\xi_n\subset\mathcal{X}}\max_{\boldsymbol{x}\in\mathcal{X}}\inf_{\boldsymbol{x}'\in\xi_n} d(\boldsymbol{x},\boldsymbol{x}'),$$

  即最小化设计区域中的点与设计之间的最大距离.

两因子情形的极大极小设计和极小极大设计:

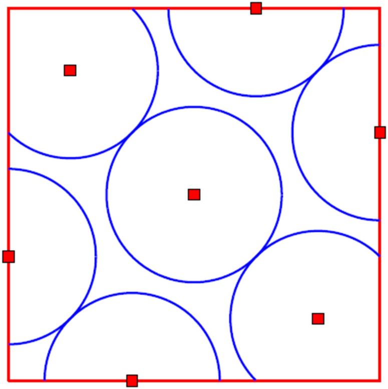


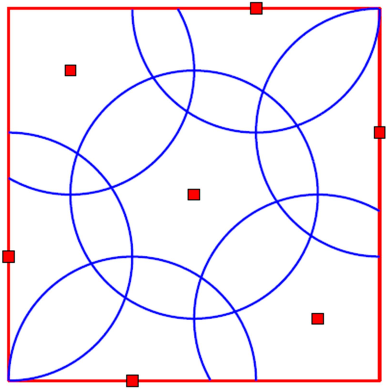
Maximin 设计



Minimax 设计

- 从所有设计中求解基于准则的最优设计解空间太大, 可限制在某些特殊设计中找最优设计;
- 如限制在超拉丁方设计中找最优设计:


极大极小-超拉丁方设计


极小极大-超拉丁方设计

4.3 空间填充设计

4.3.1 基于抽样的设计

4.3.2 基于准则的最优设计

4.3.3 均匀设计

(1) 均匀设计的定义

(2) 均匀性度量

(3) 均匀设计表

- 均匀设计的起源

"1978 年，中国七机部由于导弹设计的要求，提出了一个五因子试验，希望每个因子的水平数要多于 18 个，而试验总数又不超过 50，显然优选法和正交设计都不能用．方开泰教授在 20 世纪 70 年代中期，曾为近似计算一个多重积分问题找过我，我向他介绍了多重数值积分的方法并取得了好结果，这就使他想到是否可以将数论方法用于试验设计问题．于是我们经过几个月的共同研究，提出了一种新的试验设计，即均匀设计，将这一方法成功地用于导弹设计．"

——王元

- 均匀设计的起源
  - 方开泰教授和王元院士于 1980 年提出;
  - 给定设计 $\xi_n = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$, 从伪蒙特卡洛理论中的 Koksma-Hlawka 不等式可知

$$|\mathbb{E}[f(\boldsymbol{x})] - \bar{y}(\xi_n)| \leq V(f)\varPhi^*(\xi_n)$$

    $V(f)$ 是一个仅与 f 有关的常数; $\varPhi^*(\xi_n)$ 表示设计 ξ_n 的星偏差, 它刻画了设计 ξ_n 的均匀性. 从估计均值的角度来看, 应当使 $\varPhi^*(\xi_n)$ 最小.
  - $\varPhi^*(\xi_n)$ 与 f 无关, 因而这种设计对模型具有稳健性.

- 称 ξ_n 的均匀性度量为 ξ_n 的偏差(discrepancy), 记作 $\Phi(\xi_n)$ 或 $\Phi(\boldsymbol{D}_{\xi_n})$.
- 若一个支撑点数为 n 的设计 ξ_n^* 在一切支撑点数为 n 的设计中具有最小的偏差值, 即

$$\Phi(\xi_n^*) = \min_{\xi_n} \Phi(\xi_n)$$

则称 ξ_n^* 为在设计问题 $(n, \mathcal{X}, \Phi)$ 的均匀设计.

- 均匀设计依赖于偏差的选择, 一种偏差下的均匀设计一般不是另一种偏差下的均匀设计;
- 均匀设计与试验区域 $\mathcal{X}$ 有关, 实际问题中 $\mathcal{X}$ 一般不规则, 为简单起见假定 $\mathcal{X} = [0,1]^p$;
- 均匀设计一般不唯一.
  - 如果 ξ_n 为均匀设计, 其设计矩阵为 D_{ξ_n}, 则将 D_{ξ_n} 作行列变换获得的设计也是均匀设计.
  - 称行列变换得到的均匀设计互为等价的.
  - 等价的均匀设计至少有 $n!p!$ 个, 它们有相同的偏差值, 在实际应用中只需找到其中一个即可.

- 均匀设计依赖于偏差的选择，一种偏差下的均匀设计一般不是另一种偏差下的均匀设计；
- 均匀设计与试验区域 $\mathcal{X}$ 有关，实际问题中 $\mathcal{X}$ 一般不规则，为简单起见假定 $\mathcal{X} = [0, 1]^p$；
- 均匀设计一般不唯一.
  - 如果 ξ_n 为均匀设计，其设计矩阵为 D_{ξ_n}，则将 D_{ξ_n} 作行列变换获得的设计也是均匀设计.
  - 称行列变换得到的均匀设计互为等价的.
  - 等价的均匀设计至少有 $n!p!$ 个，它们有相同的偏差值，在实际应用中只需找到其中一个即可.

- 均匀设计依赖于偏差的选择, 一种偏差下的均匀设计一般不是另一种偏差下的均匀设计;
- 均匀设计与试验区域 $\mathcal{X}$ 有关, 实际问题中 $\mathcal{X}$ 一般不规则, 为简单起见假定 $\mathcal{X} = [0,1]^p$;
- 均匀设计一般不唯一.
  - 如果 ξ_n 为均匀设计, 其设计矩阵为 $\boldsymbol{D}_{\xi_n}$, 则将 $\boldsymbol{D}_{\xi_n}$ 作行列变换获得的设计也是均匀设计.
  - 称行列变换得到的均匀设计互为等价的.
  - 等价的均匀设计至少有 $n!p!$ 个, 它们有相同的偏差值, 在实际应用中只需找到其中一个即可.

4.3 空间填充设计

4.3.1 基于抽样的设计

4.3.2 基于准则的最优设计

4.3.3 均匀设计

(1) 均匀设计的定义
(2) 均匀性度量
(3) 均匀设计表

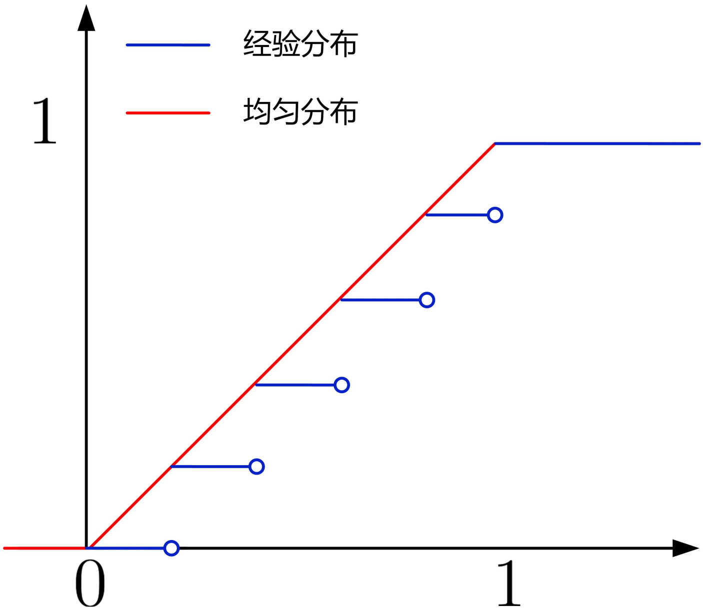
令 $F_u(\boldsymbol{x}) = x_1 \cdots x_p$ 为 $\mathcal{X}$ 上均匀分布的分布函数, 其中 $\boldsymbol{x} = (x_1, \cdots, x_p)^{\mathrm{T}}$. 以

$$F_{\xi_n}(\boldsymbol{x}) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{[\boldsymbol{x}_k, \boldsymbol{\infty})}(\boldsymbol{x})$$

表示设计 $\xi_n = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 的经验分布函数, 式中 $\boldsymbol{\infty} = (\infty, \cdots, \infty)$,

$$\mathbb{1}_A(\boldsymbol{x}) = \begin{cases} 1, & \boldsymbol{x} \in A, \\ 0, & \boldsymbol{x} \notin A. \end{cases}$$

表示集合 A 的示性函数.


经验分布
均匀分布
1
0
1

设计 ξ_n 的L_p-星偏差定义为均匀分布 $F_u(\boldsymbol{x})$ 与它的经验分布 $F_{\xi_n}(\boldsymbol{x})$ 之差的 L_p 范数:

$$\Phi_p^*(\xi_n) = \begin{cases} \left(\int_{\mathcal{X}} \left|F_{\xi_n}(\boldsymbol{x}) - F_u(\boldsymbol{x})\right|^p \mathrm{d}\,\boldsymbol{x}\right)^{\frac{1}{p}}, & 1 \le p < \infty, \\ \sup\limits_{\boldsymbol{x}\in\mathcal{X}} \left|F_u(\boldsymbol{x}) - F_{\xi_n}(\boldsymbol{x})\right|, & p = \infty. \end{cases}$$

- 当 $p=\infty$ 时, 简称星偏差, 它等价于分布拟合检验中著名的 Kormokorov-Smirnov 统计量.

- L_2-星偏差等价于 Cramer-Von Mises 统计量, Warnock(1972) 给出了计算 L_2-星偏差的简单表达式:

$$\Phi_2^*(\xi_n)=\left\{\frac{1}{3^p}-\frac{2}{n}\sum_{i=1}^{n}\prod_{j=1}^{p}\frac{1-x_{ij}^2}{2}+\frac{1}{n^2}\sum_{i,l=1}^{n}\prod_{j=1}^{p}[1-\max(x_{ij},x_{lj})]\right\}^{\frac{1}{2}},$$

当 p 给定时, 上式的计算量为 $O(n^2)$, 比星偏差的计算量大大减少.

- L_p-星偏差计算复杂, 不具反射不变性, 且低微投影均匀性不好. 统计学家们提出了好几种改进的偏差: 中心化偏差、可卷偏差、离散偏差和 Lee 偏差等等.

- 当 $p=\infty$ 时, 简称星偏差, 它等价于分布拟合检验中著名的 Kormokorov-Smirnov 统计量.

- L_2-星偏差等价于 Cramer-Von Mises 统计量, Warnock(1972) 给出了计算 L_2-星偏差的简单表达式:

$$\Phi_2^*(\xi_n)=\left\{\frac{1}{3^p}-\frac{2}{n}\sum_{i=1}^{n}\prod_{j=1}^{p}\frac{1-x_{ij}^2}{2}+\frac{1}{n^2}\sum_{i,l=1}^{n}\prod_{j=1}^{p}[1-\max(x_{ij},x_{lj})]\right\}^{\frac{1}{2}},$$

当 p 给定时, 上式的计算量为 $O(n^2)$, 比星偏差的计算量大大减少.

- L_p-星偏差计算复杂, 不具反射不变性, 且低微投影均匀性不好. 统计学家们提出了好几种改进的偏差: 中心化偏差、可卷偏差、离散偏差和 Lee 偏差等等.

- 当 $p=\infty$ 时, 简称星偏差, 它等价于分布拟合检验中著名的 Kormokorov-Smirnov 统计量.

- L_2-星偏差等价于 Cramer-Von Mises 统计量, Warnock(1972) 给出了计算 L_2-星偏差的简单表达式:

$$\Phi_2^*(\xi_n) = \left\{ \frac{1}{3^p} - \frac{2}{n}\sum_{i=1}^{n}\prod_{j=1}^{p}\frac{1-x_{ij}^2}{2} + \frac{1}{n^2}\sum_{i,l=1}^{n}\prod_{j=1}^{p}[1-\max(x_{ij},x_{lj})] \right\}^{\frac{1}{2}},$$

当 p 给定时, 上式的计算量为 $O(n^2)$, 比星偏差的计算量大大减少.

- L_p-星偏差计算复杂, 不具反射不变性, 且低微投影均匀性不好. 统计学家们提出了好几种改进的偏差: 中心化偏差、可卷偏差、离散偏差和 Lee 偏差等等.

例

对于单因子试验, 设试验区域为 $[0,1]$. 不同均匀性度量下的均匀设计不同.

- 如果利用星偏差或中心化偏差来度量均匀性, 则均匀设计为
$$\xi_n = \left\{ \frac{1}{2n}, \frac{3}{2n}, \cdots, \frac{2n-1}{2n} \right\};$$
- 如果利用可卷偏差来度量均匀性, 则有无限个均匀设计
$$\left\{ \frac{1}{2n}+a, \frac{3}{2n}+a, \cdots, \frac{2n-1}{2n}+a \right\}, \quad a \in \left[-\frac{1}{2n}, \frac{1}{2n} \right].$$

4.3 空间填充设计

4.3.1 基于抽样的设计

4.3.2 基于准则的最优设计

4.3.3 均匀设计

(1) 均匀设计的定义

(2) 均匀性度量

(3) 均匀设计表

- 给定 $(n, \mathcal{X}, \Phi)$, 求解均匀设计十分困难, 尤其是 n 和 p 比较大的时候. 统计学家把一些均匀设计编制成表, 以便查用.
- 以记号 $U_n(q^p)$ 或 $U_n^*(q^p)$ 表示均匀设计表,
  - U 表示均匀设计, n 表示要做 n 次试验, q 表示每个因子有 q 个水平, p 表示该表有 p 列.
  - 加 $*$ 的均匀设计表有更好的均匀性, 应优先选用.
  - 当 n 给定时, U_n 表比 U_n^* 表能安排更多的因子.
  - 每个均匀设计表都附有一个使用表, 它指示如何从设计表中选用适当的列, 以及由这些列所组成的试验方案的偏差.

表: $U_6^*(6^4)$

试验号	1	2	3	4
1	1	2	3	6
2	2	4	6	5
3	3	6	2	4
4	4	1	5	3
5	5	3	1	2
6	6	5	4	1

表: $U_6^*(6^4)$ 的使用表

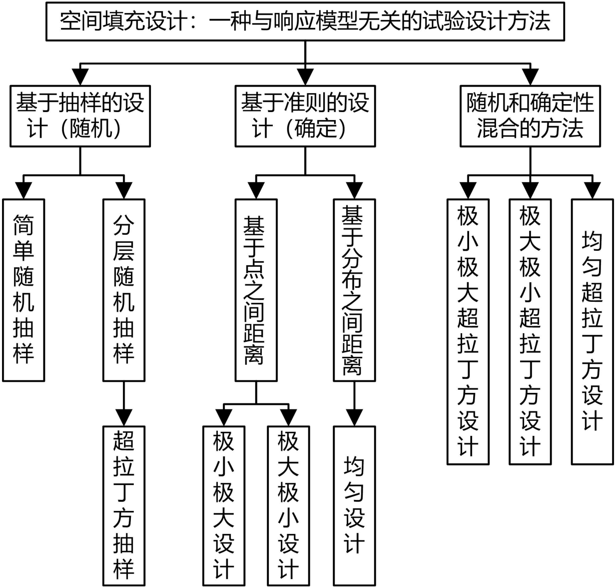
因子数	列号				Φ
2	1	3			0.1875
3	1	2	3		0.
4	1	2	3	4	0.2990

- 使用均匀设计表安排计算机试验时，需要先把表变换到试验区域 $[0,1]^p$ 上来.

均匀设计的特点:

- 每个因素的每个水平做且仅做一次试验.
- 任两个因素的试验点在平面的格子点上, 每行每列有且仅有一个试验点.
- 均匀设计表任两列组成的试验方案一般不等价, 因此, 每个均匀设计表必须有一个附加的使用表.

总结


空间填充设计：一种与响应模型无关的试验设计方法
基于抽样的设计（随机）
基于准则的设计（确定）
随机和确定性混合的方法
简单随机抽样
分层随机抽样
超拉丁方抽样
基于点之间距离
极小极大设计
极大极小设计
基于分布之间距离
均匀设计
极小极大超拉丁方设计
极大极小超拉丁方设计
均匀超拉丁方设计