文章编号:2095-6134(2018)06-0731-04

序贯最大最小距离设计的空间填充性*

滕一阳[1],武 赟[2],熊世峰[2†],杨建奎[1]

(1 北京邮电大学理学院, 北京 100876; 2 中国科学院数学与系统科学研究院, 北京 100190)

(2017 年 9 月 29 日收稿; 2017 年 11 月 16 日收修改稿)

Teng Y Y, Wu Y, Xiong S F, et al. A space-filling property of sequential maximin distance designs[J]. Journal of University of Chinese Academy of Sciences, 2018,35(6):731-734.

摘 要 最大最小距离设计是计算机试验中常用的一种空间填充设计。讨论序贯最大最小距离设计的空间填充性质,证明在样本量趋于无穷时,这种序贯设计中的点在试验区域内达到稠密.

关键词 序贯设计;空间填充设计;计算机试验

中图分类号:O212.6 **文献标志码:**A **doi:**10.7523/j.issn.2095-6134.2018.06.003

A space-filling property of sequential maximin distance designs

TENG Yiyang[1], WU Yun[2], XIONG Shifeng[2], YANG Jiankui[1]

(1 *School of Science*, *Beijing University of Posts and Telecommunications*, *Beijing* 100876, *China*;
2 *Academy of Mathmetics and Systems Science*, *Chinese Academy of Science Beijng* 100190, *China*)

Abstract Maximin distance designs, as a class of space-filling designs, are commonly used in computer experiments. In this paper we discuss the space-filling properties of sequential maximin distance design. We prove that the points in such a sequential design become dense in the experimental region as the sample size goes to infinity.

Keywords sequential design; space-filling design; computer experiment

随着计算技术的发展,计算机仿真试验在科学与工程中的应用越来越广[1-3]。由于仿真试验通常比较耗时,因此计算机试验和传统物理试验一样,需要有效地挑选少量试验点来进行仿真[4-6]。计算机试验的设计一般采用空间填充设计[7-11],其核心思想是设计点应在试验区域内分布得尽量均匀。常用的一类空间填充设计是最大最小距离设计[10-18]。其定义如下:点 $D=\{x_1,\cdots,x_n\}$ 称为空间 $[0,1]^d,d\in Z^+$ 中的最大最小距离设计如果它是优化问题 $\max\limits_{x_1,\cdots,x_n}\min\limits_{1\leqslant i<j\leqslant n}\|x_i-x_j\|$, s.t. $x_1,\cdots,x_n\in[0,1]^d,d\in Z^+$ 的解,其中 $\|\cdot\|$ 代表欧式距离[2]。最大最小距离设计也可以做成序贯类型[19],即按照最小距离准则一个个地增加试验点,这样就可以用较少的试验点得出结论,也可避免盲目加大样本数而造成浪费。文献[6]给出了其具体的定义。如今,已有不少文献讨论过最大最小距离设计的性质和构造,如文献[10-18],但目前还没有文章对序贯最大最小距离设计的性质进行讨论。本文将讨论序贯最大最小距离设计的空间填充性质。证明序贯最大最小距离设计满足最基本的空间填充性,即随着样本量的增大,该

* 国家自然科学基金(11471172, 11671386)资助
† 通信作者,E-mail:xiong@amss.ac.cn

设计在试验区域内具有稠密性,可以任意逼近试验区域内的任意一个点。

1 构造序贯最大最小距离设计

在 $[0,1]^d, d \in Z^+$ 中,按如下规则取点:使新取的点到所有已取点的最小距离最大化。引入下列记号表示:

记前 n 步取的点分别为 $s_1, s_2, \cdots, s_n$, $S_n = \{s_1, s_2, \cdots, s_n\}$ 为前 n 步取点的点集,

1) 记任意两点 $x, y \in [0,1]^d$ 间的距离为 $d(x,y) = \| x - y \|_2$。

2) 记任意一点 $x \in [0,1]^d$ 到已选点集合 S_n 的距离 $d(x,S_n) = \min_{s_i \in S_n} \{d(x,s_i)\}$。

3) 下一步要寻找的点 $s_{n+1} = \underset{x \in [0,1]^d}{\operatorname{argmax}} \{d(x,S_n)\}$,记 $q_n = \max_{x \in [0,1]^d} \{d(x,S_n)\}$。

按上述规则,在 $[0,1]^2$ 中,依次规定取点个数画图。

图 1 中,当在 $[0,1]^2$ 中第 1 次取点时,因为此时 $[0,1]^2$ 中不存在点,规定取的第 1 个点在正方形的中心。接下来就可以按照序贯最大最小距离设计在空间中依次取点。第 2 次取点时,有 4 个点同时满足规则,它们在正方形的 4 个顶点处。可以看出,按序贯最大最小距离设计取点,可以将 $[0,1]^2$ 逐步填充。

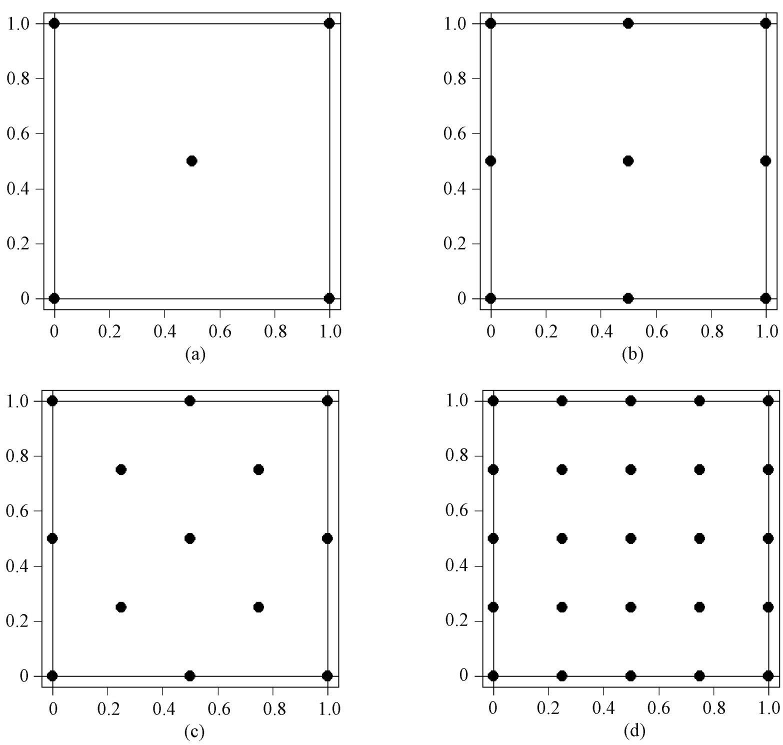

图 1 $[0,1]^2$ 中不存在点时按序贯最大最小距离设计取点

Fig. 1 Taking points by sequential maximin distance designs when there is no point in $[0,1]^2$

图 2 中,一开始 $[0,1]^2$ 中已经取了 3 个点,这 3 个点用三角形表示,再按序贯最大最小距离设计取点,取得的点用圆形表示。可以看出,这种情况下序贯最大最小距离设计也可以对 $[0,1]^2$ 进行逐步填充。

2 主要结果及证明

从第 1 节可以看出,随着样本量的增加,序贯最大最小距离设计可以对 2 维试验区域进行逐步填充。下面对任意的维数严格证明这一性质。

定理 2.1 在 $[0,1]^d, d \in Z^+$ 中,按序贯最大最小距离设计取点,设所取得的点依次为 $s_1, s_2, \cdots, s_n$,则对任意 $\varepsilon > 0$,任意 $x \in [0,1]^d$,都存在一个 $N > 0$,当 $n > N$ 时,有 $d(x,S_n) < \varepsilon$。

证明 反证方向为:存在 $\varepsilon > 0$,存在一个 $x_0 \in [0,1]^d$,对任意的 $N_1 > 0$,都存在一个 $N_2 > N_1$,有 $d(x_0, S_{N_2}) \geq \varepsilon$。

继续使用第 1 节中的符号。根据取点方式:$s_{n+1} = \arg\max_{x \in [0,1]^d} \{d(x,S_n)\}$,记 $q_n = \max_{x \in [0,1]^d} \{d(x,S_n)\}$,

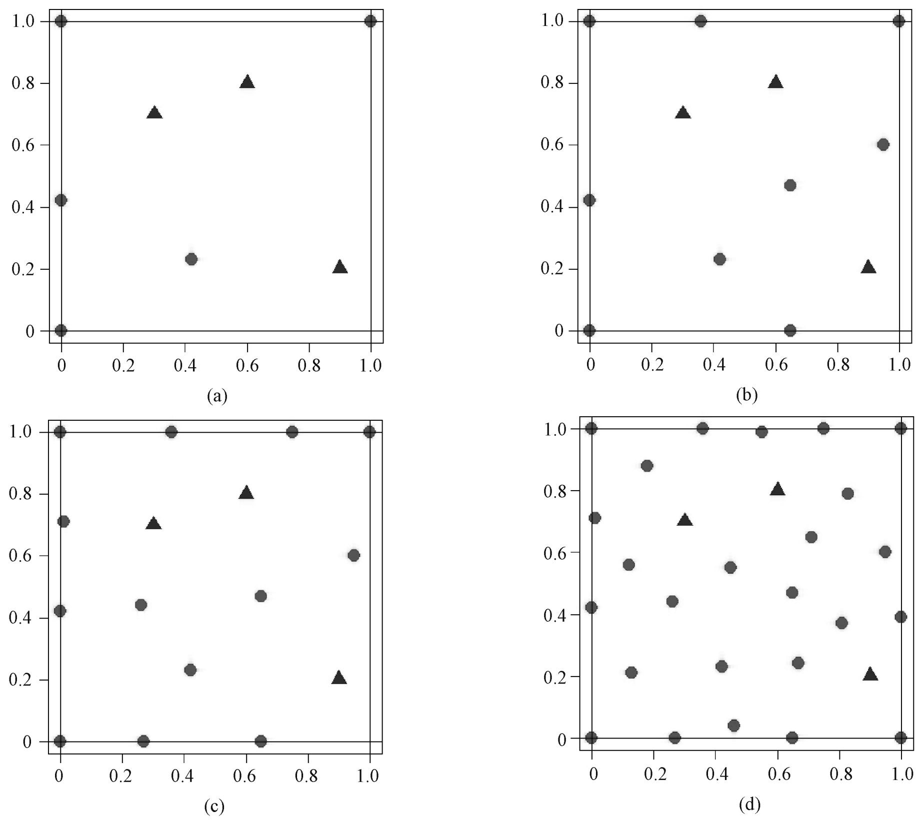

图 2　$[0,1]^2$ 中存在 3 个点时按序贯最大最小距离设计取点

Fig. 2　Taking points by sequential maximin distance designs when there are 3 points in $[0,1]^2$

则有

$$\begin{aligned} q_{n+1} &= \max_{x\in[0,1]^d}\{d(x,S_{n+1})\} \\ &= \max_{x\in[0,1]^d}\{\min_{s_i\in S_{n+1}} d(x,S_{n+1})\} \\ &= \max_{x\in[0,1]^d}\{\min_{s_i\in S_n}\{\min d(x,S_n), d(x,s_{n+1})\}\} \\ &\leqslant \max_{x\in[0,1]^d}\{d(x,S_n)\} = q_n. \end{aligned}$$

故集合 $\{q_n\}_{n=1,2,\cdots}$ 是一个单调递减的集合,且由定义有 $q_i \geqslant 0$,故集合有下确界,记 $a = \inf\limits_{n=1,2,\cdots}\{q_n\}$ 为其下确界,易得 $a \geqslant 0$。

由反证法假设,存在 $\varepsilon > 0$,存在一个 $x_0 \in [0,1]^d$,对任意的 $N_1 > 0$,都存在一个 $N_2 > N_1$,有 $d(x_0,S_{N_2}) \geqslant \varepsilon$,则 $q_{N_2} = \max\limits_{x\in[0,1]^d}\{d(x,S_{N_2})\} \geqslant \varepsilon$。

即存在 $\varepsilon > 0$,对任意的 $N_1 > 0$,都存在一个 $N_2 > N_1$,使得 $q_{N_2} \geqslant \varepsilon > 0$,且 $\{q_n\}_{n=1,2,\cdots}$ 单调递减,故可以得 $a \geqslant \varepsilon > 0$。

下面证明对于任意已取得的点 $s_i,s_j \in S_{N_2}$,均有 $d(s_i,s_j) \geqslant a > 0$,证明如下:

已知 $s_{n+1} = \operatorname*{argmax}\limits_{x\in[0,1]^d}\{d(x,S_n)\}$,$q_n = \max\limits_{x\in[0,1]^d}\{d(x,S_n)\} = d(s_{n+1},S_n) \geqslant a$,则对任意 $i \geqslant 2$,均有 $d(s_{i+1},S_i) \geqslant a$。

对于反证条件中任意的 $N_1 > 0$,取 $N_2 = N_1 + \left\lceil \dfrac{(1+d^{1/2})^d\Gamma(d/2+1)}{\pi^{d/2}\,(a/2)^d} \right\rceil + 1 > N_1$,则若存在 $s_i,s_j \in S_{N_2}$ 使得 $d(s_i,s_j) < a$,不妨设 $i > j$,则 $s_j \in S_{i-1}$,则有 $d(s_i,S_{i-1}) = \min\limits_{x\in S_{i-1}}\{d(s_i,x)\} \leqslant d(s_i,s_j) < a$。与 $d(s_i,S_{i-1}) \geqslant a$ 矛盾,所以我们证明出对于任意已取得的点 $s_i,s_j \in S_{N_2}$,均有 $d(s_i,s_j) \geqslant a > 0$,所以对 S_{N_2} 中所有的点以 $\dfrac{a}{2}$ 为半径做球,球两两不相交。这 N_2 个球的体积和为 $N_2 \cdot \pi^{d/2}(a/2)^d/\Gamma(d/2+1) > (1+d^{1/2})^d$。

但由于在空间 $[0,1]^d$ 中,任意两点间的最大距离不超过 $d^{1/2}$,故每个球的半径应满足

$$\frac{a}{2} \leqslant \frac{d^{1/2}}{2},$$

且球心均在 $[0,1]^d$ 中,所以所有的球一定被完全包含在 $\left[-\dfrac{d^{1/2}}{2}, 1+\dfrac{d^{1/2}}{2}\right]^d$ 中,且这 N_2 个球不相交,故 N_2 个球的体积和一定满足

$$V_{\mathrm{qiu}} < (1+d^{1/2})^d,$$

这与 $N_2 \cdot \pi^{d/2}(a/2)^d/\Gamma(d/2+1) > (1+d^{1/2})^d$ 矛盾。

故原假设不成立,即我们证明了:对任意 $\varepsilon > 0$,任意 $x \in [0,1]^d$,都存在一个 $N > 0$,当 $n > N$ 时,均有 $d(x,S_n) < \varepsilon$。即按序贯最大最小距离设计多次取点,可以填充空间 $[0,1]^d$。

注 1 上面证明的是 $[0,1]^d$ 中一开始没有取任何点时,按照序贯最大最小距离设计取点,可将 $[0,1]^d$ 填充。这个时候第 1 个点放在空间的中心,第 2 个点就应该取在空间 $[0,1]^d$ 的一个顶点处。

若 $[0,1]^d$ 中一开始存在 k 个点 $t_1, t_2, \cdots, t_k$,再在空间中按序贯最大最小距离设计取点时,依然可以证明定理成立。第 1 次取点时,需要使新点到这 k 个点的最小距离最大,第 2 次取点时,需要使新点到这 $k+1$ 个点的最小距离最大…… 因此在证明时,应使点集 $S_i, i = 1,2,\cdots$ 中都包含这 k 个点。取 $c = \min\{\| t_i - t_j \|, i,j = 1,2,3,\cdots,k, i \neq j\}$,$b = \min\{a, c\}$,在证明一开始将 ε 取得足够小使得 $c > \varepsilon$,就可保证 $b > \varepsilon$,再以 $\frac{b}{2}$ 为半径以 S_{N_2} 中的所有点为球心做球,且将 N_2 表达式中的 a 全部替换为 b,也可得定理结论成立。

注 2 上述定理证明的是序贯最大最小距离设计对 $[0,1]^d$ 的填充性,同样地,对任一 d 维紧集,在空间中一开始已经取了点和没取点两种情况下,仍能证明按序贯最大最小距离设计取点可将空间填充。

参考文献

[1] Morris M D, Mitchell T J, Ylvisaker D. Bayesian design and analysis of computer experiments: use of derivatives in surface prediction[J]. Technometrics, 1993, 35(3): 243-255.

[2] Morris M D. A class of three-level experimental designs for response surface modeling[J]. Technometrics, 2000, 42(2): 111-121.

[3] Moore L M, Mckay M D, Campbell K S. Combined array experiment design[J]. Reliability Engineering and System Safety, 2006, 91(10/11): 1 281-1 289.

[4] Lehman J S, Santner T J, Notz W I. Designing computer experiments to determine robust control variables[J]. Statistica Sinica, 2004, 14(2): 571-590.

[5] Mu W Y, Xiong S F. On algorithmic construction of maxmin distance designs[J]. Communications in Statistics-Simulation and Computation, 2016, 11(1): 1-14.

[6] Santner T J, Williams B J, Notz W I. The design and analysis of computer experiments[M]. New York: Springer, 2003: 138.

[7] Tang B X. A theorem for selecting oa-based latin hypercubes using a distance criterion[J]. Communication in Statistics-Theory and Methods, 2007, 23(7): 2 047-2 058.

[8] Crombecq K, Dheane T. Generating sequential space-filling designs using genetic algorithms and Monte Carlo methods[C]// International Conference on Simulated Evolution and Learning, 2010, 6 457: 80-84.

[9] Dobriban E, Fortney K. Space-filling properties of good lattice point sets[J]. Biometrika, 2015, 124(4): 959-966.

[10] Zhou Y D, Xu H. Space-filling fractional factorial designs[J]. Journel of the American Statistical Association, 2014, 109(507): 1 134-1 144.

[11] Wahl F, Mercadier C, Helbert C. A standardized distance-based index to assess the quality of space-filling designs[J]. Statistics & Computing, 2017, 27(2): 319-329.

[12] Johnson M, Moore L, Ylvisaker D. Minimax and maximin distance design[J]. Journal of Statistical Planning and Inference. 1990, 24(2): 131-148.

[13] Li Z, Nakayama S. Maximin distance-lattice hypercube design for computer experiment based on genetic algorithm[C]// International Conferences on Info-tech and Info-net, 2001, 2: 814-819.

[14] Marengo E, Todeschini R. A new algorithm for optimal, distance-based experimental design[J]. Chemometrics and Intelligent Laboratory Systems, 1992, 16(1): 37-44.

[15] Coetzer R L J, Roseouw R F, Le Roux N J. Efficient maximin distance designs for experiments in mixtures[J]. Journal of Applied Statistics, 2012, 39(9): 1 939-1 951.

[16] Xia Y, Cai T, Cai T T. Maximum Projection Designs for computer experiments[J]. Biometrika, 2015, 102(2): 371-380.

[17] Li K, Jiang B, Ai M. Sliced space-filling designs with different levels of two-dimensional uniformity[J]. Journal of Statistical Planning & Inference, 2015, 157-158: 90-99.

[18] Long T, Wu D, Guo X, et al. A deterministic sequential maximin Latin hypercube design method using successive local enumeration for metamodel-based optimization[J]. Engineering Optimization, 2016, 48(6): 1 019-1 036.

[19] Duan W, Ankenman B E, Sanchez S M, et al. Sliced full factorial-based latin hypercube designs as a framework for a batch sequential design algorithm[J]. Technometrics, 2017, 59(1): 1-39.