Constructing Cascading Latin Hypercubes

C. Devon Lin, Derek Bingham, Randy Sitter, and Boxin Tang
Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

Abstract

Computer experiments commonly use space-filling designs. As the number of factors increases, the sparsity of the design points increase. Space-filling designs place all the points about the same distance (quite far) apart. If the spatial correlation length is also small relative to the spacing, there are no points close enough together to give reliable estimates of the correlation parameters. Handcock (1991) introduced cascading Latin hypercube designs (CLHD) to alleviate this issue. We develop systematic methods for constructing a rich class of CLHDs.

Introduction

Computer experiments are performed to explore complex computer codes. The goals are often:

- **Screening**
- Building an emulator of the simulator
- Optimization
- Model calibration

- We consider screening problem
- The rule of thumb for run-size is 10 times dimensions
- We use Gaussian process model with the power exponential family of correlation functions

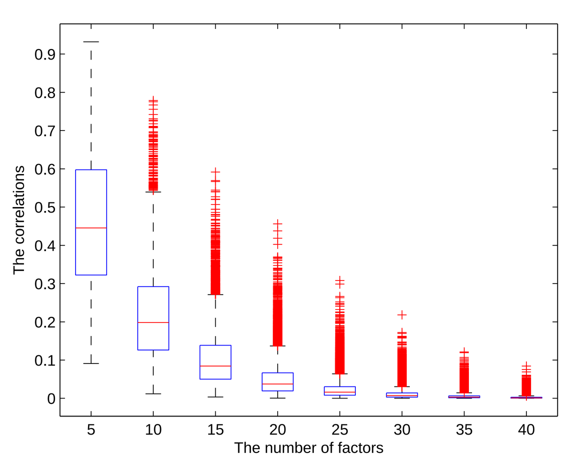

Figure 1: Boxplots of $10d$-run Maximin LHDs with d factors, correlation parameters are set to be 1.0

Problem of using space-filling designs:

- The spatial correlation length between any two design points decreases dramatically as the number of factors increases as shown in Figure 1.

Our approach

- introduce a new representation of CLHDs
- provide a basic method for constructing CLHDs
- offer a generalization construction method
- demonstrate the value of CLHDs.

Notation

- n and m represent the run size and the number of factors, respectively
- The levels are chosen to be centered, equally-spaced and integer-valued
- $-(n-1)/2, \ldots, -1, 0, \ldots, (n-1)/2$ if n is odd
- $-n+1, -n+3, \ldots, -1, 1, \ldots, n-1$ if n is even

Definitions

Definition (Handcock,1991): A CLHD of size $n = \prod_{k=1}^{p} n_k$ in m dimensions with levels $(n_1, \ldots, n_p)$ is a n_p-point LHD about each point in the $(n_1, \ldots, n_{p-1})$ centered CLHD.

New Definition: Define a matrix U of $LHD(n,m) = (L_{ij})$ to have (i,j)-th element

$$U_{ij} = \begin{cases} \lceil (L_{ij}+n)/(2n_2) \rceil, & \text{if } n \text{ is even;} \\ \lceil (L_{ij}+(n+1)/2)/(n_2) \rceil, & \text{if } n \text{ is odd,} \end{cases}$$

where $\lceil q \rceil$ be the nearest integers greater than or equal to q. A LHD is then termed a two-level CLHD of n points in m-dimensional with level (n_1, n_2) if matrix U has n_1 distinct rows and each distinct row has n_2 replicates. p-level CLHD can be defined in a similar manner.

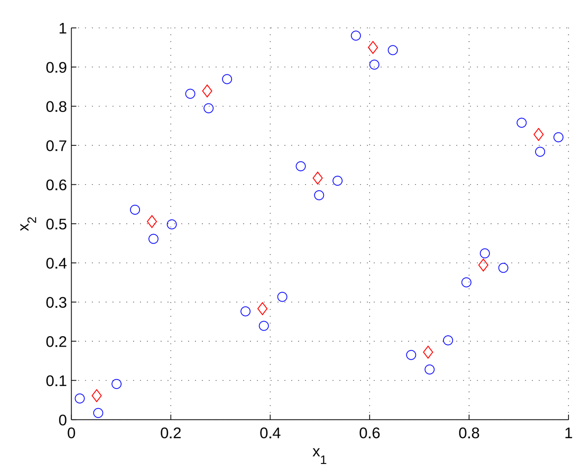

Figure 2: Cascading Latin hypercube design with 27 points in levels $(9,3)$. The 9 diamonds are a centered LHD on a 9×9 grid in $[0,1) \times [0,1)$. Around each of these diamonds is a LHD of size 3. The corresponding U matrix is 3 replicates of rows $(1,1), (2,5), (3,8), (4,3), (5,6), (6,9), (7,2), (8,4), (9,7)$.

Construction Methods

Define

- A be an $n_1 \times m_1$ design with $a_{ij} \pm 1$
- B be an $n_2 \times m_2$ Latin hypercube design
- C be an $n_1 \times m_1$ Latin hypercube design
- D be an $n_2 \times m_2$ design with $d_{ij} \pm 1$
- α and β be any positive real number
- $\otimes$ represents Kronecker product

Basic Method:

$$L = \alpha A \otimes B + \beta C \otimes D. \qquad (1)$$

Generalization Method: For each $j = 1, \ldots, m_1$, let B_j be an $n_2 \times m_2$ LHD.

$$L = (\alpha a_{ij} B_j + \beta c_{ij} D) = \begin{bmatrix} \alpha a_{11} B_1 + \beta c_{11} D & \ldots & \alpha a_{1m_1} B_{m_1} + \beta c_{1m_1} D \\ \alpha a_{21} B_1 + \beta c_{21} D & \ldots & \alpha a_{2m_1} B_{m_1} + \beta c_{2m_1} D \\ \vdots & \vdots & \vdots \\ \alpha a_{n_1 1} B_1 + \beta c_{n_1 1} D & \ldots & \alpha a_{n_1 m_1} B_{m_1} + \beta c_{n_1 m_1} D \end{bmatrix}. \quad (2)$$

Remarks:

- Method (2) is proposed for better projection property
- If design C are constructed via (1) or (2), the resulting design L in (1) or (2) will be a three-level CLHD
- $A \otimes B$ (or $a_{ij} B_j$) and C control the design points locally and globally, respectively
- The proposition below tells us the value of α, β and design D in (1) and (2) in order to obtain a two-level CLHD.

Proposition: Let D be an $n_2 \times m_2$ matrix of unit elements. A design L, formed as in (1) and (2), is a two-level CLHD of $n = n_1 n_2$ points in m-dimensional ($m = m_1 m_2$) with level (n_1, n_2) if α and β are chosen in the following way,

- $\alpha = 1$ and $\beta = n_2$ if both n_1 and n_2 are even or odd,
- $\alpha = 2$ and $\beta = n_2$ if n_1 is even and n_2 are odd,
- $\alpha = 1$ and $\beta = 2n_2$ if n_1 is odd and n_2 are even.

Example

Let $n_1 = 9, n_2 = 3, m_1 = 4, m_2 = 3$. Design D is a 3×3 matrix of unit elements. Designs A, B and C are defined as

$$A = \begin{pmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad C = \begin{pmatrix} -4 & -4 & -3 & -2 \\ -3 & 0 & 0 & 3 \\ -2 & 3 & 2 & -1 \\ -1 & -2 & -1 & 1 \\ 0 & 1 & 4 & -4 \\ 1 & 4 & -2 & 2 \\ 2 & -3 & 3 & 4 \\ 3 & -1 & -4 & 0 \\ 4 & 2 & 1 & -3 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$

Let B_j be a row permutation of B. Set $\alpha = 1$ and $\beta = 3$ based on the aforementioned proposition. The resulting design L via (2)is a two-level CLHD with 27 points in levels $(9,3)$. The first and fifth columns of design L correspond to the 27 circles in Figure 2.

Results and Conclusions

Results:

In Figure 2,

- Consider large input dimensions, $d = 20$
- CLHDs with 192 points in level $(48,4)$, $(24,8)$ and $(16,12)$ are generated
- Maximin LHD with 192 points is generated
- Power exponential correlation function with $\theta_i = 1$ is used
- The average correlation between each design point and its k-nearest neighbors is computed, $k = 4, 8, 12$
- Euclidean distance is used to find k-nearest neighbors
- We observe that when the dimension of inputs is relatively large, maximin Latin hypercube designs fail to provide close design points
- Cascading Latin hypercube designs have close design points to detect the relationship between the inputs

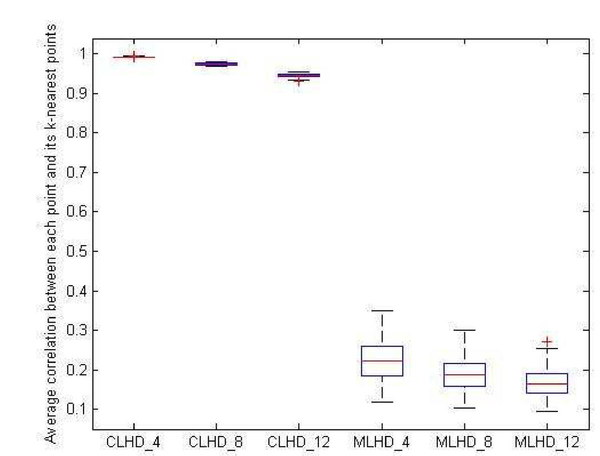

Figure 3: Comparisons of average correlation between each design point and its k-nearest points, $k = 4, 8, 12$. CLHD_4, MLHD_4 represent the boxplots with $k = 4$ for CLHD and MLHD, respectively. CLHD_8, MLHD_8 are similar for $k = 8$. CLHD_12, MLHD_12 are for $k = 12$.

Conclusion:

- When the input dimension is large, design points provided by space-filling designs are too sparse for Gaussian process to be effective
- We provide methods for systematically constructing a rich class of designs with cascading structure
- Cascading Latin hypercube designs provide local clustered points to enhance estimation of correlation parameters
- The reliable estimation of correlation parameters allows us to achieve the goal of screening factors

References

1. Handcock, M.S. (1991), "On Cascading Latin Hypercube Designs and Additive Models for Experiments." *Commun. Statist. - Theory and Method.*, 20(2), 417-439.
2. Lin C.D., Bingham, D., Sitter, R.R., and Tang, B. (2007), "Constructing Cascading Latin Hypercube Designs.", Manuscripts in Preparation.