Latin Hypercube Sampling

Alessandro C.

April 5, 2024

Abstract

This paper aims to define a formal form of the Latin Hypercube Sampling (LHS) and the relative expansion procedure (eLHS), finally proposing an "expansion's grade" a-priori formula . The proposed approach is expressed through Set Theory.

1 Space Binning

1.1 Range Group

We define A as a ordered sequence of N couples of real numbers between 0 and 1 that delimit a continuous interval. As follows:

$$A = \langle \ (low_q, up_q) \in [0,1]^2 : low_q < up_q \leq (low_{q+1} \ ?? \ 1), \forall q \in [1,N] \cap \mathbb{N} \ \rangle \quad (1)$$

The operator ?? is called *nullish* and it returns the right-hand if the left-hand is not defined, left-hand is returned otherwise.

N.B:

$$||A|| = N \quad (2)$$

We'll refer to a Range Set of cardinality N as an *N-Ranged Group*, also we'll call a couple $(low_q, up_q) \in A$ as a bin_q *of* A. Let's use the following format:

$$A[N]$$

1.1.1 Regularity of Range Groups

Given $A[N]$ N-Ranged Group, it is said to be *regular* iff:

$$up_q - low_q = \frac{1}{N} \ , \ \forall (low_q, up_q) \in A \quad (3)$$

It's deductible from (3) that:

$$low_1 = 0, up_N = 1$$

Furthermore, if A_1 and A_2 are two regular N-Ranged Groups then:

$$A_1 = A_2 \quad (4)$$

1.2 Binning Grid

Given P number of $\{A_j\}$ N-Ranged Groups, let's say that B is a *Binning Grid* if :

$$B = A_1 \otimes A_2 \otimes \ldots \otimes A_P \tag{5}$$

Conventionally, we'll address to the j-th Range Group of B with B_j .
If every component of $\{A_j\}$ is regular, following (4) we can simply say:

$$B = A[N]^P \ , \ A = A_1 = A_2 = \ldots = A_P$$

1.3 Sample Set Space \$

Let's define the Space \$ that contains the "*Sample Set* S of size N in P dimensions":

$$S = \{x_{ij}\} \in \$(A[N]^P) \subset M(N, P)$$

s.t.:

$$\forall i \in [1, N] \cap \mathbb{N}, \forall j \in [1, P] \cap \mathbb{N} : xij \in [0, 1] \tag{6}$$

We'll refer to each element of $\{x_i\}$ - rows of S - as "*i-th sample of S*".
We'll refer to each element of $\{x^j\}$ - columns of S - as "*projection of S on j-th axis*".

2 Latin Hypercube Sampling

2.1 LHS

Given $B = A[N]^P$ regular Binning Grid, $\{x_{ij}\} \in \$(B)$ matrix and H(x) Heaviside step function, if:

$$\forall j = [1, P] \cap \mathbb{N} : \sum_{i=1}^{N} H(x_{ij} - low) * H(up - x_{ij}) = 1 \ , \ \forall (low, up) \in B_j \tag{7}$$

then x_{ij} is a *Latin Hypercube Sample set* of size N and binning B, denoted:

$$\{x_{ij}\} \in LHS(N, B) \subset \$(B) \tag{8}$$

The property specified at (7) is called *one-projection property.*

2.2 Grade of a Sample Set

Given a Sample Set $S = \{x_{ij}\} \in M(N, P)$, we can compute an index that measures how much the S is close to achieve the one-projection property given a specific $B = A[Q]^P$ regular Binning Grid. As follows:

$$gr(S, B) = \frac{\sum_{j=1}^{P} \sum_{q=1}^{Q} min(\sum_{i=0}^{N} H(x_{ij} - low_{jq}) \cdot H(up_{jq} - x_{ij}), 1)}{Q \cdot P} \tag{9}$$

This quantity lies between 0 - when S' grade approaches 0, it tends to have less overlaps - and 1.
The S is a Latin Hypercube Sample Set on Binning B iff it has maximum grade:

$$S \in LHS(N, B) \Leftrightarrow gr(S, B) = 1 \tag{10}$$

2.3 Expanded Sample Set

Given $S = \{x_{ij}\} \in LHS(N, A[N]^P)$ and $M \in \mathbb{N}$ number of add-ons;

- Let U[N + M] be an Range Group, define:

$$C = U^P \tag{11}$$

  that represents the new Binning Grid on S by adding M new intervals.

- Introduce V_j set, composed of all intervals (low, up) which has no x_{ij} placed in it - so called *"Voids"* - for each j-th dimension:

$$V_j = \{(low_{jq}, up_{jq}) \in C_j : \sum_{i=1}^{N} H(x_{ij} - low_{jq}) \cdot H(up_{jq} - x_{ij}) = 0, \forall q \in [1, N+M] \cap \mathbb{N}\} \tag{12}$$

  The number of voids per dimension is:

$$||V_j|| \geq M$$

- Build a Binning Grid W composed of W_j subsets of V_j s.t.:

$$\forall j \in [1, P] \cap \mathbb{N} \ , \ W_j \subseteq V_j : ||W_j|| = M$$

$$W = W_1 \otimes ... \otimes W_P \tag{13}$$

  We'll say that W is the *mask* of C given S.

- Finally, set the Sample Set $E = \{y_{ij}\} \in LHS(M, W)$ called *"expansion set"*. By merging S and E we can allocate an expanded Sample Set Z:

$$Z = \{x_{1j} \, ... \, x_{Nj} \ , \ y_{1j} \, ... \, y_{Mj}\}$$

2.3.1 Expanded LHS

Given $Z = \{x_{1j} \, ... \, x_{Nj} \ , \ y_{1j} \, ... \, y_{Mj}\}$ expanded Sample Set - for sake of clarity - its grade is given by:

$$gr(\{x_{1j} \, ... \, x_{Nj} \ , \ y_{1j} \, ... \, y_{Mj}\}, C) \leq 1 \tag{14}$$

2.3.2 Perfect Expansion

Given $Z = \{x_{1j} \, ... \, x_{Nj} \, , \, y_{1j} \, ... \, y_{Mj}\}$ expanded Sample Set, let $E = \{y_{ij}\}$ be its expansion; then E is said to be a *Perfect expansion* iff:

$$gr(Z, C) = 1 \tag{15}$$

which it leads us to conclude that:

$$Z \in LHS(N + M, C) \tag{16}$$

2.3.3 Expanded Grade Prediction

Theorem Given $S = \{x_{ij}\}$ Sample Set on A^P regular Binning Grid, let Z be the expanded Sample Set of S (of M samples) on Binning Grid C.
We can compute the grade of Z on C *a-priori* using the following formula:

$$gr(Z, C) = 1 - \frac{\sum_{j=1}^{P} \sum_{i=1}^{N-1} H(\frac{\lceil x_{ij}(N+M) \rceil}{N+M} - x_{(i+1)j})}{P(N + M)} \tag{17}$$

Proof

- Let $E \in LHS(M, W)$ be the expansion set of Z, with W mask of C