

# A comparison of latin hypercube sampling techniques for a supply chain network design problem

Kantapit Kaewsuwan<sup>1</sup>, Chumpol Yuangyai<sup>1,\*</sup>, Udom Janjarassuk<sup>1</sup> and Kanokporn Rienkhemaniyom<sup>2</sup>

<sup>1</sup>Industrial Engineering Department, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

<sup>2</sup>Graduate School of Management and Innovation, King Mongkut's University of Technology Thonburi, Bangkok, Thailand

**Abstract.** Currently, supply chain network design becomes more complex. In designing a supply chain network to withstand changing events, it is necessary to consider the uncertainties and risks that cause network disruptions from unexpected events. The current research related to the designing problem considers network disruptions using Monte Carlo Sampling (MCS) or Latin Hypercube Sampling (LHS) techniques. Both have a disadvantage that sample points or disruption locations are not scattered entirely sample space leading to high variation in objective function values. The purpose of this study is to apply a modified LHS or Improved Distributed Hypercube Sampling (IHS) techniques to reduce the variation. The results show that IHS techniques provide smaller standard deviation than that of the LHS technique. In addition, IHS can reduce not only the number of sample size but also and the computational time.

## 1 Introduction

Supply chain network design is a critical decision-making process that affects the efficiency of enterprise management. Especially in high volatility business environment. This is due to the risk and the uncertainty of the raw materials of suppliers. A disaster occurs in one country will affect the global supply chain, resulting in disruptions, such as in 2011, a major flood in Thailand, one of global supply chain disruptions [1].

Therefore, the proactive supply chain network design is important for all organizations. It is necessary to consider the risks that occur in countries where the suppliers, manufacturers, and warehouses are located. And also, the ability to recover from disruptions. In many supply chain network design studies.

In this paper, we use a two-stage stochastic programming model[2] for supply chain design. Our supply chain network consists of suppliers, manufacturing, warehouses, and retailers. A single product is considered. The objective function is to maximize profits under disruptions.

The two stages involve 1) to find an initial solution by a heuristic method. Then, 2) evaluation of those solutions from step 1 with difference disruptions by using Monte Carlo Sampling (MCS)

In general, MCS techniques are commonly used. However, it requires many sample size to increase the precision of the solution leading longer computation time. Latin Hypercube Sampling (LHS) and its variants can reduce the number of sample sizes and use less time to evaluate a solution. In this paper, we compare two type of LHS to improve the efficiency of the two-stage stochastic programming.

## 2 Literature review

Since our main is on the sample techniques for stochastic programming, therefore only MCS, LHS, and its variants are discussed.

MCS techniques[3] is generated by using a random number that independent of each other and uniformly distributed on the interval [0, 1]. MCS is a matrix ( $M$ ) with dimensions  $N \times S$  where  $N$  is the number of samples and  $S$  is the number of independent variables

McKay et al., 1979 proposed one of the sampling techniques used for computer-aided design, The LHS techniques provide a more uniform random sample distribution. LHS is a matrix ( $L$ ) with dimensions  $N \times S$  The design steps is shown below:

**Step 1** The matrix  $P(N, S)$  is consists of random shuffles of integers ranging from 1 to  $N$ .

**Step 2** The matrix  $M(N, S)$  is generated from a random number in Monte Carlo method, where  $M(N, S) \in U[0, 1]$  which are independent of each other.

**Step 3** The matrix  $L(N, S)$  is constructed from equation (1), and this matrix is used to simulate the event.

$$L(N, S) = \frac{1}{N} [P(N, S) - M(N, S)] \quad (1)$$

Brian K. B., 2002 [4][5] developed Improved Distributed Hypercube Sampling (IHS) from the LHS techniques, which adds conditions for finding the distance between points by using the Euclidean distances. The design process of the IHS techniques is similar to the design of the LHS techniques. It is different to construct the  $P(N, S)$  matrix with the distance between the points that are described in Section 3.3.2.

\* Corresponding author: [chumpol.yu@kmitl.ac.th](mailto:chumpol.yu@kmitl.ac.th)

### 3. Supply Chain Network Model Description

#### 3.1 Supply chain network design

In this paper, we consider a network with four echelons that consists of suppliers, manufacturing, warehouses, and retailers. As shown in Figure 1, the location of supplier, manufacturing, and warehouses depends on the probability of occurrence of disruptions within the supply chain network. The probability in each location of suppliers, manufacturing, and warehouses are varying. The goal is to decide the location of the supplier, and warehouses while maximizing profit.

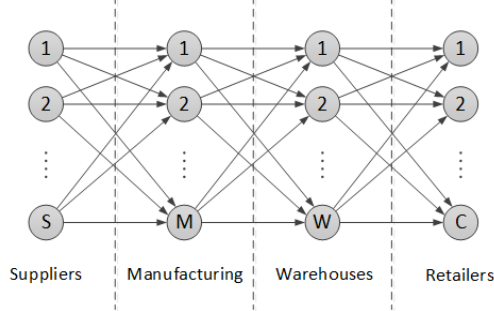


Figure 1. Supply chain network [6]

#### 3.2 Mathematical model and solution approach

Our mathematical model is developed from Kanokporn et al., 2015 [6]. The model is shown below:

*Sets:*  $S$ : Set of suppliers;  $M$ : Set of manufacturing plants;  $W$ : Set of warehouses;  $C$ : Set of retailers;  $L$ : Set of warehouse capacities;  $K$ : Set of disruptive scenarios

*Indices:*  $s$ : Index of suppliers  $s \in S$ ;  $m$ : Index of manufacturing plants  $m \in M$ ;  $w$ : Index of warehouses  $w \in W$ ;  $c$ : Index of retailer  $c \in C$ ;  $l$ : Index of warehouse capacities  $l \in L$ ;  $k$ : Index of disruptive scenarios  $k \in K$

##### Parameters

- $cap_m$  Production capacity at manufacturing plant  $m$
- $cap_s$  Capacity at supplier  $s$
- $cap_w^l$  Capacity at warehouse  $w$  of size  $l$
- $d_c$  Demand for products at retailer  $c$
- $msm$  Minimum transportation quantity from supplier  $s$  to manufacturer  $m$
- $f_w^l$  Fixed cost of opening a warehouse  $w$  of capacity  $l$
- $pm_{sm}$  Purchasing cost of material from supplier  $s$  by plant  $m$
- $tr_{sm}$  Transportation cost per unit from plant  $m$  to warehouse  $w$
- $tr_{wc}$  Transportation cost per unit from warehouse  $w$  to retailer  $c$
- $pc_m$  Production cost for a product at plant  $m$
- $np$  Price of a product
- $ls_c$  Lost sales cost at retailer  $c$

##### Random Parameters

- $\alpha_{sk} = \begin{cases} 1 & \text{if supplier } s \text{ is operated in scenario } k \\ 0 & \text{if a disruption occurs at supplier } s \text{ in scenario } k \end{cases}$
- $\delta_{wk} = \begin{cases} 1 & \text{if warehouse } w \text{ is operated in scenario } k \\ 0 & \text{if a disruption occurs at warehouse } w \text{ in scenario } k \end{cases}$
- $\beta_{mk} = \begin{cases} 1 & \text{if plant } m \text{ is operated in scenario } k \\ 0 & \text{if a disruption occurs at plant } m \text{ in scenario } k \end{cases}$
- $p_k$  The probability of disruption occurs in scenario  $k$

##### First-stage Decision Variables

- $x_w^l = \begin{cases} 1 & \text{if warehouse } w \text{ is operated with size } l \\ 0 & \text{otherwise} \end{cases}$
- $y_s = \begin{cases} 1 & \text{if supplier } s \text{ is selected} \\ 0 & \text{otherwise} \end{cases}$

##### Second-stage Decision Variables

- $QSM_{smk}$  Quantity of raw material purchased from supplier  $s$  by plant  $m$  in scenario  $k$
- $QMW_{mwk}$  Quantity of products shipped from plant  $m$  to warehouse  $w$  in scenario  $k$
- $QWC_{wck}$  Quantity of products shipped from warehouse  $w$  to retailer  $c$  in scenario  $k$
- $LD_{ck}$  Quantity of sales lost at retailer  $c$  in scenario  $k$

##### 3.2.1 Objective function

The objective is to maximize the expected supply chain profit ( $Z$ ) which is the difference between the total cost and the expected revenue.

$$\begin{aligned} \text{Maximize } Z = & - \sum_{w \in W} \sum_{l \in L} f_w^l x_w^l + \sum_{k \in K} p_k \left[ np \left( \sum_{w \in W} \sum_{c \in C} QWC_{wck} \right) \right. \\ & - \left( \sum_{s \in S} \sum_{m \in M} pm_{sm} QSM_{smk} \right) \\ & - \left( \sum_{m \in M} \sum_{w \in W} tr_{mw} QMW_{mwk} + \sum_{w \in W} \sum_{c \in C} tr_{wc} QWC_{wck} \right) \\ & \left. - \sum_{m \in M} pc_m \left( \sum_{w \in W} QMW_{mwk} \right) - \sum_{c \in C} ls_c LD_{ck} \right] \end{aligned}$$

##### 3.2.2 Constraints

###### Supplier capacity

$$\sum_{m \in M} QSM_{smk} \leq cap_s \alpha_{sk} y_s, \quad \forall s \in S, k \in K$$

###### Inter-stage flow

$$msm \cdot \alpha_{sk} y_s \leq QSM_{smk} \leq cap_s \alpha_{sk} y_s, \quad \forall s \in S, k \in K$$

###### Production capacity

$$\sum_{w \in W} QMW_{mwk} \leq cap_m \beta_{mk}, \quad \forall m \in M, k \in K$$

###### Material flow between suppliers and plants

$$\sum_{s \in S} QSM_{smk} = \sum_{w \in W} QMW_{mwk}, \quad \forall m \in M, k \in K$$

###### Warehouse capacity

$$\begin{aligned} \sum_{m \in M} QMW_{mwk} & \leq \sum_{l \in L} cap_w \delta_{wk} x_w^l, \quad \forall w \in W, k \in K \\ \sum_{l \in L} x_w^l & \leq 1, \quad \forall w \in W \end{aligned}$$

Product flow between warehouse and retailers

$$\sum_{m \in M} QMW_{mwk} = \sum_{C \in C} QWC_{wck}, \quad \forall w \in W, k \in K$$

Demand requirement

$$\sum_{w \in W} QWC_{wck} + LD_{ck} = d_c, \quad \forall c \in C, k \in K$$

Non-negativity

$$QSM_{smk}, QMW_{mwk}, QWC_{wck}, LD_{ck} \geq 0$$

### 3.3 Solution Methodology

We use two-stage stochastic programming which is composed of 2 Stages: 1) Simulated Annealing algorithm (SA) which is used to find the initial solution. Then, 2) disruptions are generated to evaluate the solution's objective function. In the first stage, we used an SA to find the solution which includes 2 steps. The first step is to set the model's initial parameters. The second step is to find the initial solution, which is the fixed cost of suppliers' location, warehouse's location and the size of the warehouse. In the second stage is to generate disruptive scenarios by using two types of LHS techniques. We generate a sample average problem instead of finding the answer to all possible events. This technique is called sample average approximation (SAA)[3]. Then we find the value of the objective function, which is the mean of the maximum profit that we obtained under disruptive scenarios by using the CPLEX program. The structure of finding the solution as shown in Figure 2.

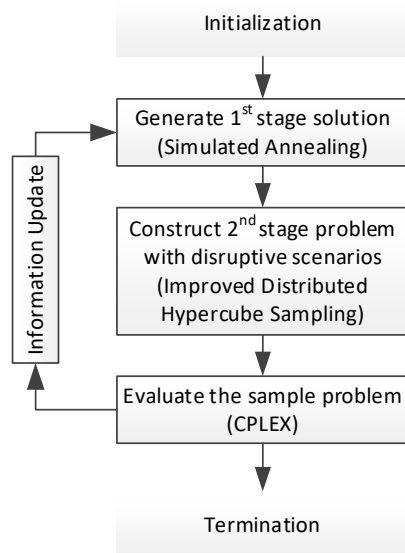


Figure 2. The general framework of the proposed methodology

#### 3.3.1 Simulated Annealing

The Simulated Annealing method [7] is used to find the near optimal solution.

#### 3.3.2 Disruptive Scenarios

In this study, we use IHS to generate disruptive scenarios to ensure that all disruptive scenarios will occur in each input variables. We expect to obtain solutions that have a smaller standard deviation than the LHS techniques.

Euclidean distances that use in IHS can be calculated from the equation 2, should be close to or equal to  $d_{opt}$ , which can be calculated from equation 3. This techniques has a distribution of sample points covering the sample space and have low coefficients of variation. It is one of the most popular sampling techniques used in improving the performance of a solution. This research utilizes the sampling techniques by the hypercube distribution method to randomly sample the disruptions to ensure that random events can represent all events.

$$d(x_i, x_j) = d_{ij} = \left[ \sum_{k=1}^S |x_{ik} - x_{jk}|^2 \right]^{1/2}, \quad (2)$$

$$i \in \{1, \dots, (N-1)\}, j \in \{(i+1), \dots, N\}$$

$$d_{opt} = \frac{N}{\sqrt{N}} \quad (3)$$

Where

$x_{ik}$  Integer at scenarios  $i$  of variable  $k$

$x_{jk}$  Integer at scenarios  $j$  of variable  $k$

$d(x_i, x_j)$  Distance between  $x_{ik}$  and  $x_{jk}$ ,  $i \neq j$

$d_{opt}$  Optimal distance

IHS is a matrix  $(I)$  with dimensions  $N \times S$  where  $N$  is the number of samples and  $S$  is the number of independent variables. The design steps are as follows.

**Step 1:** Define initial parameters  $(N, S, D)$  where  $D$  is the number of sets of numbers generated.

**Step 2:** Create the final  $P(N, S)$  matrix from the design approach.

**Step 3:** Define the initial parameter  $r := N-1$

**Step 4:** When the required number of cycles  $(r \leq 2)$  is not reached

For the answer matrix. Complete the following steps.

1. Create a matrix  $A(c, S)$  by  $c = r, (D-1) + 1, \dots, rd$  and  $d = 1, \dots, D$ . The random number of integers is  $l$  to  $N$  and is not chosen from the matrix  $P(N, S)$ .

2. Find the distance between the points of  $P(r+1, S)$  and  $A(l, S)$  from equation 2, where  $l = 1, \dots, r, d$ .

3. Put an integer in matrix  $A$  where  $l$  is the distance between the nearest point and  $d_{opt}$  from equation 3. Put in the matrix  $P(r, S)$

**Step 5** Take the integer value from  $l$  to  $N$  and not be chosen from the final matrix  $P(N, S)$ . Put in the matrix  $P(l, S)$

**Step 6** Create the Matrix  $I(N, S)$  from Equation 1.

Then, Compare the distribution of random numbers obtained from the LHS techniques and IHS techniques. As shown in Figure 3.

Figure 3. shows a scatter plot of LHS techniques. The number of samples size  $(N)$  is 40 with two input variables  $(S)$ . When dividing the sample space into  $5 \times 5$  in each row and each column it contains eight points, which shows that the sample points are uniformly

distributed in the corresponding row and column. It shows that each input variables has all portions of its range. But the distribution of sample points is also adjacent, not spreading the sample space. Because of the design, there are no conditions to find the distance between sample points. While Figure 4. shows that the dispersion of the sample points has spread over the sample area.

In this paper, the sampling techniques by LHS and IHS method is used as a set of disruptive scenarios in the evaluation of the objective function.

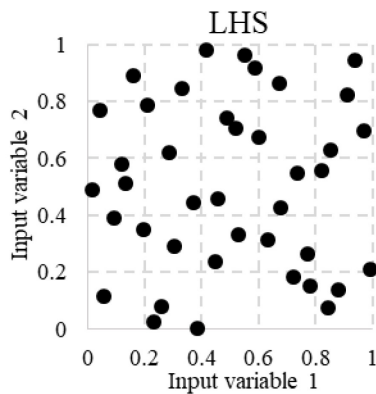


Figure 3. Scatter plot of samples generated by LHS

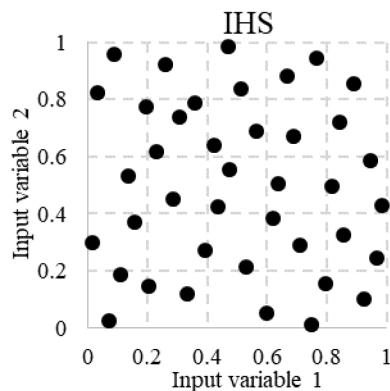


Figure 4. Scatter plot of samples generated by IHS

#### 4 Result and discussions

We compared the two sampling techniques of LHS and IHS, which are used to randomize disruption events in designing supply chain network problem. The sample

size ( $N$ ) was varied from 20, 50, and 100 respectively. Each sample size run was repeated 10 times and calculate their average and standard deviation, as shown in Table 1.

Based on the comparison of the standard deviation, IHS outperforms LHS. It reduces the standard deviation by 22.45%, 15.54%, and 21.99%, for all three sample sizes. Furthermore, smaller sample size with IHS provides the similar or lower standard deviation of those of larger sample size with LHS.

#### 5 Conclusion

This study presents a comparison of two sampling techniques Latin Hypercube Sampling and Improve Distributed Hypercube Sampling. The IHS is used with two-stage stochastic programming to solve a supply chain network design. IHS provides a lower standard deviation, smaller sample size, and better computation efficiency. In addition, a modification of LHS or Optimal Latin Hypercube Sampling will be further studied to improve computational efficiency.

#### References

1. Haraguchi M, Lall U (2015) Flood risks and impacts: A case study of Thailand’s floods in 2011 and research questions for supply chain decision making. *Int J Disaster Risk Reduct* 14:256–272
2. Rienkhemaniyom K, Yuangyai C, Janjarassuk U A two-stage stochastic program for supply chain design under disruption
3. Shapiro A, Ruszczyński A (2003) Monte Carlo sampling methods. *Handbooks Oper Res Manag* 10:353–425
4. Beachkofski BK, Grandhi R V (2002) Improved Distributed Hypercube Sampling. *AIAA Pap* 1–8
5. Randall D. Manteufel (2001) Distributed Hypercube Sampling Algorithm. *Am Inst Aeronaut Astronaut*
6. Rienkhemaniyom K (2015) Resilience improvement in supply chain network design. In: *Asia Pacific Industrial Engineering and Management Systems Conference*. pp 475–482
7. Sastry K, Goldberg D, Kendall G (2005) *Search Methodologies*

Table 1: Comparison of objective values between two sampling techniques with different samples sizes

Sampling techniques	N = 20		N = 50		N = 100	
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
Latin Hypercube	12,486,781.44	244,354.22	11,876,213.93	197,586.00	11,680,288.82	194,683.05
Improve Distributed Hypercube	12,228,495.99	199,558.44	11,943,134.72	171,003.85	11,676,678.56	159,589.15
Difference	-	-22.45%	-	-15.54%	-	-21.99%