Comparative Study of Latin Hypercube Sampling and Monte Carlo Method in Structural Reliability Analysis

Zichuan Wang *

Department of Civil Engineering, Tongji University, Shanghai 200092, China

* Corresponding Author Email: zichuan0304@tongji.edu.cn

Abstract. The Monte Carlo (MC) Method is a crucial approach to approximately calculate the reliability of structures, although it requires a large amount of calculation comprising millions of sampling realizations. The Latin Hypercube Sampling (LHS) method has been raised to improve efficiency, and in which situation this method can significantly outperform the MC method is a pivotal problem that needs to be resolved. In this thesis, comparative sampling is conducted with two methods to calculate the failure probability of a steel beam, and merits, as well as demerits, are concluded with a comparative analysis of the sampling results. The thesis compares the fundamental principles and methodology of the two methods, and the variables involved in reliability analysis are expounded. It is emphasized that the LHS method has the advantage of stratified sampling, and variables consist of loads, geometrical dimensions, and material properties. The details of the example beam are given, followed by the sampling done by two methods. The analysis extracts the noteworthy statistics in sampling, propounding the differences: the LHS method can markedly reduce the sampling number needed to obtain reliable figures with a higher coefficient of variation. Further research should be carried out to study the influence of various distributions on the difference between the two methods. Overall, this thesis provides insights for researchers in the reliability analysis field to utilize the LHS method more sensibly.

Keywords: Structural reliability analysis; Latin Hypercube Sampling; Monte Carlo Method; sampling; statistical variability.

1. Introduction

The probability of failure is pivotal while designing and analyzing a structure [1]. Since the commencement of research on reliability in the field of structural engineering in the mid-20th century, this field's outcomes have profoundly impacted practical engineering. Numerous methods have been developed to calculate or denote the reliability of structures, such as the reliability index. Currently, the MC method has been accepted as a powerful tool to deal with many engineering problems regarding reliability [2]. It is actually one of the simplest and most widely utilized methods employed to predict failures like collapse behavior [3]; however, the process of analysis will occupy a large number of computational resources, rendering it time-consuming. As a result, several alternate methods have been invented, including the LHS method and the response surface methodology [4]. In general, the LHS method is regarded as a branch of the MC method simulation, as it is often the preferred sampling method based on the Monte Carlo principle while having a more efficient manner of stratifying across the range of each sampled variable [5]. To fully comprehend the practicality of the LHS method, thus improving the working efficiency of reliability analysis, comparative research should be carried out on these two methods in a scrupulously designed example and feasible and trustable comparing method, which are the major obstacles in research.

A wide range of studies have been done on the MC method and LHS method. Michael D.Shields reviewed several sampling methods, including the LHS method and proposed Latinized stratified sampling [6]. CJ Sallaberry described a way to enlarge the size of a Latin hypercube sample [7]. the Local Domain Monte Carlo Simulation was studied by H.J.Pradlwarter to improve the correctness of variance of the estimated failure probability [8]. Such research undoubtedly modified and enhanced the two methods, respectively. Nevertheless, in most of the common engineering analysis scenarios, like ANSYS, the standard MC method is applied to conduct numerical simulation, explaining the importance of figuring out the most suitable situations for these two methods. Similar research has been done in the field of fluid flow, whereas the novel outcomes in the structural reliability realm need to be explored [9]. In this thesis, an example based on the structural component is chosen to testify and compare the crucial parameters of two sampling results since this way is more accessible and closer to reality in engineering. Stark differences and concise conclusions are expected to illustrate the results of sampling.

Specifically, the goal of this thesis is to illustrate and compare the two methods. Firstly, they are illustrated and several variables accounting for the variability of the structure are listed. Secondly, consists of the explanation of steel-beam example's calculation formulas and statistical parameters. Finally, the sampling results by two methods are given in the form of figures and tables, as well as a detailed analysis.

2. Analysis of reliability calculation methods

2.1. The basic principle of structural reliability analysis of MC method

The reliability of structure is defined as the probability that the building can complete its expected targets in designed expectancy, and the random variable $X = (x_1, x_2, \dots, x_n)$ indicates the uncertainty of the building, such as geometric and material properties. Thus, the extreme status of structure can be formulated using the function (1):

$$Z = g(X) = r(X) - s(X) \tag{1}$$

When $g(X) > 0$, the structure remains safe, and the corresponding probability is regarded as reliable, whereas when $g(X) < 0$, the structure fails with the corresponding failure probability. Let r and s denote the extreme strength of material and structure stress, respectively, and their probabilities are independent; the reliability of structure can be expressed as formula (2):

$$R = P(r > s) = \int_{-\infty}^{\infty} f_s(s)[\int_{s}^{\infty} f_r(r)dr]\, ds \tag{2}$$

If function Z has explicit expression, numerical integration can be utilized to solve it. However, due to the complexity and variability of structure and loads, explicit expression cannot be obtained in most cases. Therefore, when dealing with structure systems with random variables, especially non-linear problems, the most efficient way is the MC method based on the statistical sampling theory. The calculation methods of structural system reliability generally include the failure mode method, MC method, response surface method and stochastic finite element method. The MC method is the most direct calculation way, which obtains the failure probabilities by sampling a large number of random variables. The failure probability can be obtained with formula (3):

$$P_f = \frac{1}{N}\sum_{i=1}^{N} I[g(X)] \tag{3}$$

Where, $I[g(X)]$—index function, $I[g(X)] = I[g(x_1, x_2, \dots, x_n)] = \begin{cases} 0, g(X) \geq 0 \\ 1, g(X) < 0 \end{cases}$; P_f—failure probability of the structure, and $P_f = 1 - R$; N—the sampling number.

The scope of application of the MC method is comparatively broad. As long as the model is accurate and the sampling number is abundant, the result will be highly trustworthy. The advanced performance of modern computers lays a solid hardware foundation for the application of this method.

In conclusion, the advantage of the MC method is its conciseness and accessibility, although, in order to obtain precise failure probability, it requires an exorbitant calculation number. Especially when encountering a small failure probability, the number of times of calculation by the MC method is often as many as tens of thousands or even hundreds of thousands, and the calculation time is too long.

2.2. The basic principle of structural reliability analysis of LHS method

LHS method was proposed by McKay et al. in 1979, which is a multidimensional stratified sampling method. It first determines the sampling number N, then divides the variables' probability distribution functions into N non-overlapping subintervals, in which the independent equal probability sampling is conducted respectively. This way, the large amount of repetitive sampling work in the MC method can be averted. By comparison, the LHS method has a better performance in the estimation of mean value and variance.

In order to assure that the random numbers sampled come from each subinterval, the random number V_i in the subinterval i should comply with the equation (4):

$$V_i = \frac{V}{N} + \frac{i-1}{N} \tag{4}$$

Where, $i = 1,2, \dots N$; N denotes the desired sampling number; V is a random number from the sampling of the uniform distribution with a range of (0,1); V_i is the random number in the subinterval i. Only one random number should be generated from every subinterval. Each random element V_i obtained from the above method needs to be mapped according to the function (5):

$$\hat{x}_{ij} = {F_j}^{-1}(V_{ij}) \tag{5}$$

Where, ${F_j}^{-1}$ represents the inverse function of the variable j's cumulative distribution function. Therefore, the random numbers deriving from N subintervals will convert into N sampling values following the cumulative distribution function (CDF). Finally, all these values will be rearranged, as the random permutations will be conducted on the order number of every subinterval.

Suppose the LHS method is utilized on a sampling project consisting of two variables and requiring five realizations. In that case, one possible solution can be shown in Fig. 1 in which the values of either axis stand for V_i, rather than the actual values of variables that can be obtained with the formula (5). It is clearly displayed from this figure that all points are well separated without the tendency of congregation.

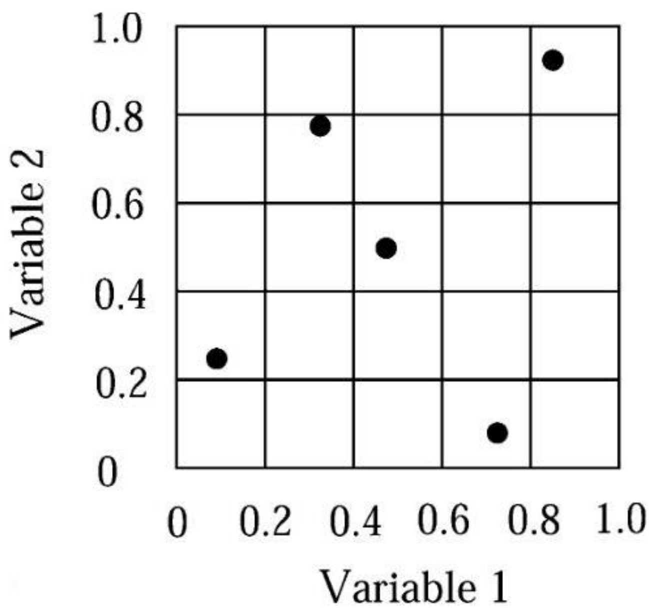

Fig 1. LHS method (two variables and five realizations) [10].

The LHS method has the advantage of providing stable projected parameters since it can acquire the sampling displaying the characteristics of distributions from a relatively smaller amount of sampling. Meanwhile, the same accuracy requires a far larger amount of sampling if it is done with the MC method.

2.3. The random variables in structural reliability analysis

2.3.1. Basic information about variables

In realistic engineering structures, loads, ultimate stress, material performance, and geometric properties are random variables that follow certain kinds of probability distributions. When dealing with the analysis of structural reliability, the random variables can be classified into three types: sizes of structures, physical properties of materials, and external effects on structures like loads and temperature change. Before conducting the analysis, the observation values or experimental data of random variables should be gathered, and statistical analysis should be carried out, which will provide statistical properties such as mean, variance, and coefficient of variation. Most of the random variables obey normal distribution, lognormal distribution, or extreme value distribution type-I (also known as Gumbel Distribution).

2.3.2. The statistical analysis of material properties

There are many indexes that represent the mechanical properties of materials, and the commonly-used indicators are ultimate strength, yield strength, fatigue strength, elongation, and elastic modulus. These variables generally comply with normal distribution, or approximately. Currently, the manufacturers provide engineers with definite value or range, as the dimensional data are generally given in the form of nominal dimensions or tolerances. Consequently, when applying these data in a probabilistic design, it is necessary to derive the mean and standard deviation from them.

2.3.3. The statistical analysis of working loads

When a definite value *A* regarding the material property is given, this value can be used as the mean value of this variable, and its standard deviation is obtained using the coefficient of variation *V*, as displayed in equation (6):

$$\sigma_A = V \times \mu \tag{6}$$

The load applied to the structure can be force, moment, stress, power, temperature, etc. The load can be divided into static load, dynamic load, stable, unstable, and other types.

2.3.4. The distribution and statistical variance of geometrical dimensions

The statistical distribution and deviation also vary randomly due to machining errors. The machining dimensions are the result of the combined effect of several random factors, which generally also comply with normal distributions. The geometrical dimensions are generally given in the form of specified tolerances, which can be treated according to the 3σ rule. If the actual range of variation of dimension *D* is *D*±*T*, then $3\sigma=T$. Therefore, the formula for standard deviation is equation (7):

$$\sigma = T/3 \tag{7}$$

3. Comparative sampling based on the example

3.1. The basic information about the sampling example

In order to make a comparison between the MC method and the LHS Method, a structural example should be created on which these two methods will be tested. To simplify it, a steel beam is chosen as the example, which is made of stainless steel [11]. The specific type is austenitic, and the product type is cold-rolled coil/sheet (C). The mean value of its yield strength $f_{y,mean}$ is $312N/mm^2$ and the standard deviation is $15.2\ N/mm^2$. Therefore, the coefficient of variation is 0.049. The dimensional variations also need to be taken into consideration. According to the statistics of I-shaped beams, the variations can be listed in the Table 1.

Table 1. Dimensional variation of key dimensions of I-sections [11].

Dimension	Depth(h)	Breadth(b)	Web thickness(t_w)	Flange thickness(t_f)
Mean	1.0141	0.9977	0.9991	0.9994
Standard deviation	0.0369	0.0132	0.0151	0.0182
Coefficient of variation	0.0364	0.0132	0.0151	0.0182

It is noteworthy that all the variables concerning the properties of steel beams comply with normal distribution and the data in this table are ratios of measured to nominal values. The section of the sample beam is chosen as 300×150×6.5×9, which means the depth of the I-shaped beam is 300mm, the breadth is 150mm, the web thickness is 6.5mm, and the flange thickness is 9mm. The material of this beam is identical to the stainless steel in Table 1. Besides, the self-weight of this beam is $36.7kg/m$. After calculation, the crucial sectional properties are listed here: $W_x = 462.17\ cm^3, I_x = 6932.52cm^4$. With these data and the formula of steel structure (10), the ultimate bending moment can be calculated.

$$M_x = \gamma_x W_{nx} f_y \tag{8}$$

Where, γ_x denotes the plastic adaption coefficient of cross-section, and in this case, it can be chosen as 1.05; W_{nx} denotes the cross-sectional net moment of inertia. The result of M_x is $151.41kN \cdot m$. Since the target of the simulation is comparing two sampling methods, the shear stress calculation is omitted. However, these data cannot display the property of variables, which means that during the sampling using the two methods mentioned above, all parameters should be utilized in the form of variables.

Besides, assume that this beam bears the live load and dead load of a square area of 2×8m^2, and the values are $2.5kN/m^2$ and $5.5kN/m^2$ respectively. In the calculation of loads, variability is dismissed since the concrete situation of load variability varies in different regions and no specific data are provided, and the figures above are concluded from codes in different countries. The variability of the length of the beam is dismissed, too. The researched beam is simply supported with two main beams at its either end. With these statistics, the actual internal force in the middle of the beam can be expressed by formula (9):

$$M_{max} = \frac{1}{8} q l^2 \tag{9}$$

Were, $q = 2.5 \times 2 + 5.5 \times 2 + 0.367 = 16.367kN/m;\ l = 8m;\ M_{max} = 130.936kN \cdot m.$

3.2. The sampling process and result using MC and LHS method

With the MC method, the sampling numbers were set as 3,000, 5,000, 50,000, 200,000 and 500,000 on MATLAB. The failure probability and the coefficient of variability of the beam have been calculated, and the results can be seen in Table 2.

Table 2. Results of the MC method.

Sampling numbers	3,000	5,000	50,000	200,000	500,000
Failure probability (one example)	0.0167	0.0108	0.0122	0.0128	0.0130
Coefficient of variation	0.162	0.124	0.0383	0.0178	0.0131
Average probability of 100 tests	0.0134	0.0130	0.0129	0.0130	0.130

It should be noticed that the failure probability is derived from one random set of sampling without any further process. In contrast, the coefficient of variation comes from the statistical analysis of the failure probability. In detail, for every set of sampling with different aggregate numbers (like 3,000, 5,000), it was conducted 100 times in order to obtain a reliable coefficient of variation.

With the LHS Method, in order to compare the two methods, the sampling numbers were also set as 3,000, 5,000, 50,000, 200,000 and 500,000. The results are shown in Table 3.

Table 3. Results of the LHS method.

Sampling numbers	3,000	5,000	50,000	200,000	500,000
Failure probability (one example)	0.0117	0.0112	0.0132	0.0130	0.00130
Coefficient of variation	0.123	0.100	0.0346	0.0168	0.00953
Average probability of 100 tests	0.0130	0.0130	0.0130	0.0129	0.0130

3.3. The analysis and comparison of the MC method and LHS method

One example of the results of the two methods can be seen in Fig. 2, from which it is clearly shown that when the sampling number reaches 50,000, the failure probability calculated by the LHS method converges on a stable value, while the results of the MC method are rather volatile and still fluctuate on a small scale after reaching 50,000 times of sampling. Therefore, if the target of sampling is to obtain a comparatively reliable probability figure, the sampling numbers required for the LHS Method are approximately a quarter of these of the MC method, which indicates that the use of the LHS Method can exceedingly reduce the sampling numbers.

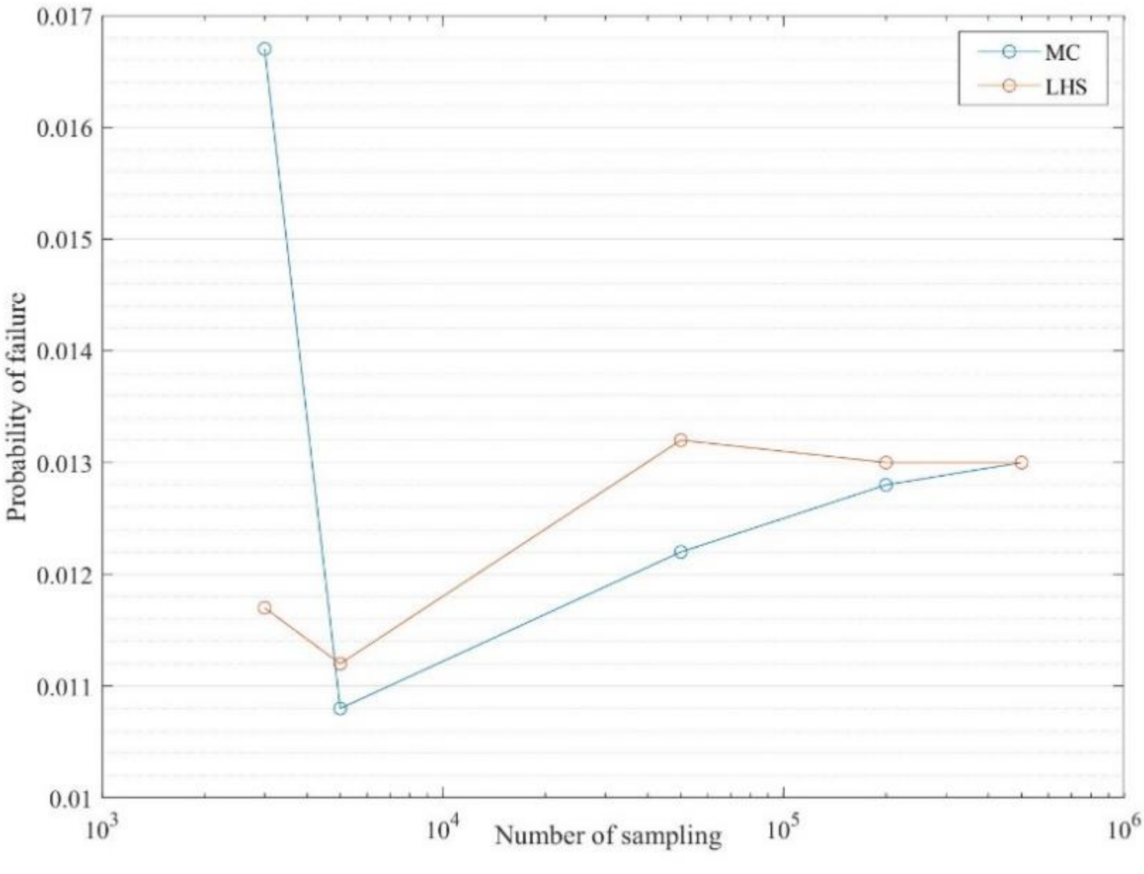

Fig 2. Tendency of the probability of failure with the sampling numbers increasing.

Although this example illustrates some properties of these two methods, due to the variability of examples, further investigation should be carried out. It is also analyzed that the results of coefficient of variation vary remarkably too, which is displayed in Fig. 3.

Obviously, as the sampling number increases, the coefficient of variation drops in both methods, although there is a significant difference in values. The figures for the LHS method are consistently higher than these of the MC method, but there is a narrowing trend in the gap between them. Therefore, the brief conclusion is that the sampling efficiency of the LHS method is higher than that of the MC method, and with the increase of times of sampling, both methods tend to be more accurate (the accuracy of results is inversely proportional to the coefficient of variation) while the upward accuracy rate of the MC method is more rapid. It should be complemented that after the sampling number reaches 50,000, there is no marked difference between the two methods.

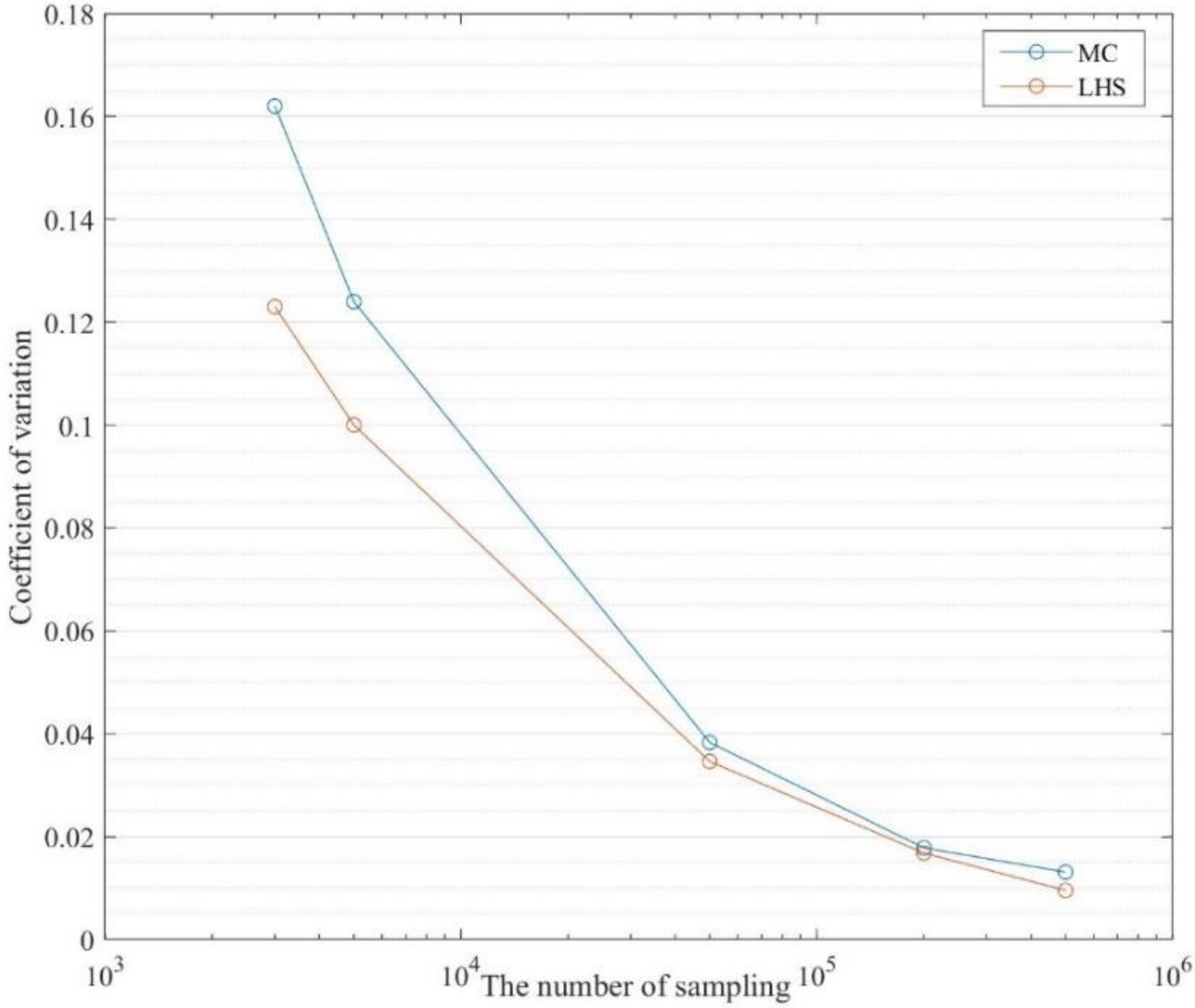

Fig 3. Difference of COV in two methods.

There is also a notable difference in the elapsed time of code in MATLAB. When the program is operated on a computer with the CPU of i7-6700k and the sampling number is set at 5,000, it takes 12.41s and 3.88s for the MC method and LHS method to run 100 times, respectively. This comparison indicates that the time needed for the LHS method is far less than the traditional MC method.

It is also necessary to study the difference between these two methods under other distributions and variables. Thus, in the next step of sampling, the geometric properties of the tested beam are deemed as constants, leading to the yield strength of the material becoming the only variable in the reliability analysis. The main body of the sampling is analogous to the part 3.2 and 3.3 while the W_x is $462.17\ cm^3$ in this case, and the extracted results are shown in the Table. 4 and Fig. 4. Since the goal of this group of sampling is to compare the trend of efficiency changes between two methods which can be appraised by the coefficient of variation when variables are varied, the other parameters were dismissed.

Table 4. Results of the comparative sampling process.

Sampling numbers	3,000	5,000	50,000	200,000	500,000
Coefficient of variation (MC)	0.343	0.300	0.0865	0.0474	0.0267
Coefficient of variation (LHS)	0.0540	0.0282	0.00213	0.000879	0.0000714

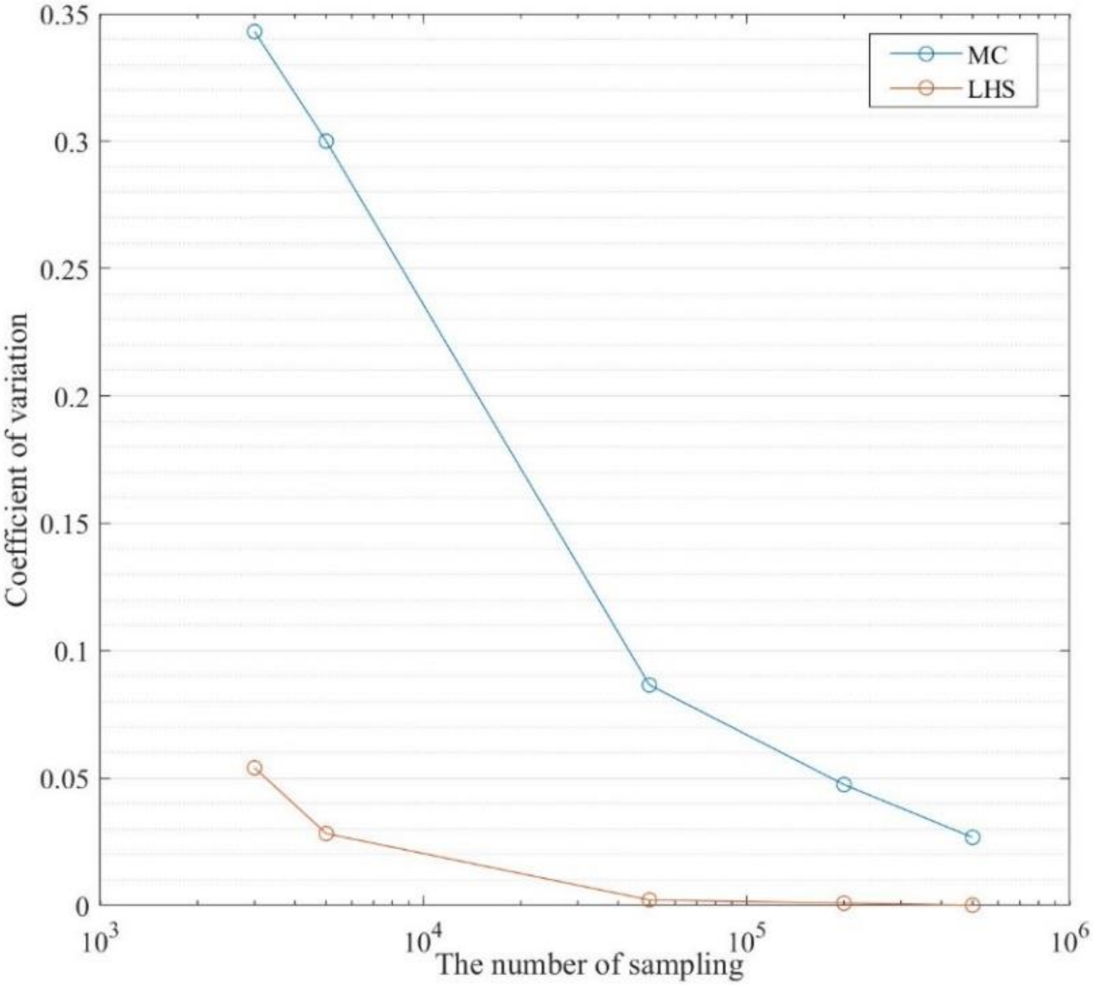

Fig 4. Results of the comparative sampling process.

It can be inferred from the figure that when the number of variables shrinks, the accuracy difference between these two methods enlarges. Meanwhile, unlike the results in Fig.2, the coefficient of variation calculated by the LHS Method at every time of sampling is extremely lower than that of the MC method. At the same time, its dropping rate is far higher, which is plausibly exemplified by the figure for the LHS Method in the last column of table 4. In addition, the absolute values of the MC method are higher than the former simulation while the absolute values of the LHS Method are lower. With these analyses, it can be concluded that the LHS Method is more accurate and suitable when dealing with sampling situations with fewer variables.

4. Conclusion

The MC method and LHS method are potent tools for tackling structural reliability problems. This essay mainly discusses the methodology and principles of these two methods and the vital elements that should be considered during the failure probability analysis. Sampling procedures are analyzed and elaborated, and several formulas which facilitate the calculation of parameters of variables are proposed. The highlight of this essay concentrates on the steel-beam sampling example, and practical conclusions have been extracted: the MC method lags behind the LHS method since it requires more sampling times to achieve the same statistical precision as the LHS method. The deeper insight is that the less complex the variables are, the more efficient the LHS method is, as the advantage of this method will be degraded by increasing the sampling number. The comparison in terms of time effectiveness is also conducted, demonstrating that the LHS method is significantly time-saving. Overall, the LHS method is an excellent substitute for the MC method in structural reliability analysis.

To further enhance the accessibility and efficiency of reliability analysis, future research should aim at optimization and modification of the LHS method and the development of feasible ways to obtain structural variables. The comparisons and suggestions of structural reliability analysis methods have been made in this thesis to provide advice to structural engineers.

References

[1] William L. Oberkampf, Sharon M. DeLand, et al. Error and uncertainty in modeling and simulation. Reliability Engineering & System Safety, 2002, 75(3): 333-357.

[2] Luigi Carassale, Giovanni Solari. Monte Carlo simulation of wind velocity fields on complex structures. Journal of Wind Engineering and Industrial Aerodynamics, 2006, 94(5): 323-339.

[3] M. Gordini, M.R. Habibi, M.H. Tavana, M. TahamouliRoudsari, M. Amiri. Reliability Analysis of Space Structures Using Monte-Carlo Simulation Method. Structures, 2018, 14: 209-219.

[4] André I. Khuri, Siuli Mukhopadhyay. Response Surface Methodology. WIREs Computational Statistics, 2010, 2(2):128-149.

[5] J.C. Helton, F.J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety, 2003, 81(1): 23-69.

[6] Michael D. Shields, Jiaxin Zhang. The generalization of Latin hypercube sampling. Reliability Engineering & System Safety, 2016, 148: 96-108.

[7] C.J. Sallaberry, J.C. Helton, S.C. Hora. Extension of Latin hypercube samples with correlated variables. Reliability Engineering & System Safety, 2008, 93(7): 1047-1059.

[8] H.J. Pradlwarter, G.I. Schuëller. Local Domain Monte Carlo Simulation. Structural Safety, 2010, 32(5): 275-280.

[9] J.C. Helton, F.J. Davis, J.D. Johnson. A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. Reliability Engineering & System Safety, 2005, 89(3): 305-330.

[10] A. Olsson, G. Sandberg, O. Dahlblom. On Latin hypercube sampling for structural reliability analysis. Structural Safety, 2003, 25(1): 47-68.

[11] S. Afshan, P. Francis, N.R. Baddoo, L. Gardner. Reliability analysis of structural stainless steel design provisions. Journal of Constructional Steel Research, 2015, 114: 293-304.