

# Weighted Latin Hypercube Sampling to Estimate Clearance-to-stop for Probabilistic Design of Seismically Isolated Structures in Nuclear Power Plants

Han, Minsoo<sup>1)</sup> · Hong, Kee-Jeung<sup>2)\*</sup> · Cho, Sung-Gook<sup>1)</sup>

<sup>1)</sup>R&D Center, Innose Tech Co. Ltd, <sup>2)</sup>Department of Civil and Environmental Engineering, Kookmin University

## / A B S T R A C T /

This paper proposes extension of Latin Hypercube Sampling (LHS) to avoid the necessity of using intervals with the same probability area where intervals with different probability areas are used. This method is called Weighted Latin Hypercube Sampling (WLHS). This paper describes equations and detail procedure necessary to apply weight function to WLHS. WLHS is verified through numerical examples by comparing the estimated distribution parameters with those from other methods such as Random Sampling and Latin Hypercube Sampling. WLHS provides more flexible way on selecting samples than LHS. Accuracy of WLHS estimation on distribution parameters is depending on the selection of weight function. The proposed WLHS is applied to seismically isolated structures in nuclear power plants. In this application, clearance-to-stops (CSs) calculated using LHS proposed by Huang et al. [1] and WLHS proposed in this paper, respectively, are compared to investigate the effect of choosing different sampling techniques.

**Key words:** Probability, Distribution parameter, Latin hypercube sample, Weight function, Design

## 1. Introduction

Sampling technique is used to estimate uncertainty inherent in population. As the type of probability distribution of population is unknown in most sociology and engineering problems, the expectation of random variable or the occurrence probability of subset can be directly calculated through sampling. Also, in a case where the type of probability distribution of population is known or properly assumable, it is possible to calculate the occurrence probability of subset based on the distribution parameters of population estimated through sampling.

It is often difficult to define the probability distribution of inputs entered in a system to define problems in sociology and engineering. Even if the inputs are fortunately known as normal distribution, the outputs produced by a system may have other types of distribution than normal distribution if the correlation between random variables is nonlinear or the system itself has a nonlinear feature. Nevertheless, in pro-

babilistic design for nuclear power plants associated to ASCE 43 [2], the probability distribution of outputs produced by the system is properly assumed in order to increase the efficiency in calculation through an allowance for some degree of error [1].

To estimate the mentioned uncertainties, one can use well known techniques such as 1) Random Sampling [3, 4], 2) Importance Sampling [5], and 3) Latin Hypercube Sampling [6]. As Random Sampling (RS) uses samples selected randomly among population, it has a disadvantage of a lengthy calculation time because a large number of samples are required especially in a case where a probability to be calculated is located in the tail of probability distribution. When there is a need to calculate the probability in the tail, Importance Sampling is a method for selecting, randomly in the tail area, a far smaller sample than RS. Latin Hypercube Sampling (LHS) is a method for dividing the whole interval of random variables into several intervals with the same probability area and for selecting samples from each interval. LHS can estimate the probability distribution approximately using a far smaller sample than RS, and is effective in estimating the distribution parameters in a case where the type of the probability distribution of random variable is known or properly assumable. For this reason, LHS is preferred in probabilistic structural design for an efficient calculation [1].

\*Corresponding author: Hong, Kee-Jeung

E-mail: [kjhong@kookmin.ac.kr](mailto:kjhong@kookmin.ac.kr)

(Received December 30, 2017; Revised January 20, 2018; Accepted February 7, 2018)

This study proposes extension of LHS to avoid the necessity of using intervals with the same probability area. In a case where intervals with different probability areas are used, it is necessary to make a correction using proper weight function. This method is called Weighted Latin Hypercube Sampling (WLHS). This study establishes equations necessary to apply weight function to WLHS, and verifies WLHS through numerical examples.

The proposed WLHS is also applied to seismically isolated structures in nuclear power plants. Huang et al. [1] considered the uncertainties using LHS for a study on seismic performance evaluation of seismically isolated structures in NPPs. The same numerical example as Huang et al. [1] is used in this paper with the site-specific ground motions of the nuclear power plant in Diablo Canyon, which is western region in the USA. In this application, clearance-to-stops (CSs) calculated using LHS proposed by Huang et al. [1] and WLHS proposed in this paper, respectively, are compared to investigate the effect of choosing different sampling techniques.

## 2. Weighted Latin Hypercube Sampling (WLHS)

In order to see how to extend LHS to propose WLHS, it is helpful for us to review on LHS. Based on this review, weight function and distribution parameters can be defined in a reasonable way.

### 2.1 Latin Hypercube Sampling (LHS)

If the number of samples is the same as that of population, an exact probability distribution can be found. However, in real situation, probability distribution can be estimated using samples far smaller than population in quantity. In Latin Hypercube Sampling, the total probability area created between probability density function and horizontal axis (random variable axis) is divided into n intervals with the same areas, and n samples are created through a random selection within each interval or a selection of the probability-area center in each interval. The probability-area center refers to the value of random variables at the center of probability area in each interval.

If the population of random variable  $X$  follows normal distribution, sample mean  $\bar{X}$  and unbiased sample variance  $\bar{s}^2$  can be estimated as follows, based on samples selected using LHS. The well-known derivation of these equation is shown in Appendix as a reference.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \tag{1}$$

$$\bar{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \tag{2}$$

Using sample mean and sample variance calculated by these equations, the mean and variance of population can be approximately estimated. Furthermore, using these calculated distribution parameters, the normal probability distribution of population can be approximated.

If  $Y$  has log-normal distribution, ' $X = \ln Y$ ' follows normal distribution, and  $\bar{X}$ , which is the sample mean of  $X$ , is defined as follows:

$$\bar{X} = \ln Y_1 + \ln Y_2 + \dots + \ln Y_n = \frac{1}{n} \sum_{i=1}^n \ln Y_i \tag{3}$$

where  $Y_i$  refers to a sample selected from the  $i^{\text{th}}$  interval of log-normal distribution, and  $\bar{\theta}$  is the sample median of the log-normal distribution. The Equations above can be rewritten for  $\bar{\theta}$  as follows:

$$\bar{\theta} = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln Y_i\right) \tag{4}$$

The sample lognormal variance of  $Y$ ,  $\bar{\beta}^2$ , is the same as the sample variance of  $\ln Y$ ,  $\bar{S}_{\ln Y}^2$  so that it can be defined as follows:

$$\bar{\beta}^2 = \bar{S}_{\ln Y}^2 = \frac{1}{n-1} \sum_{i=1}^n (\ln Y_i - \ln \bar{\theta})^2 \tag{5}$$

### 2.2 Weighted Latin Hypercube Sampling (WLHS)

In the afore-mentioned Latin Hypercube Sampling, each interval is set so that probability area of each interval is the same, and one sample is selected from each interval. However, this method of setting the sample intervals need to have more flexible way because of the following two reasons. First, sample data may not exist in an interval set by LHS in a case where the samples are collected from the uncontrolled observations. Second, there is possibility that a sample interval can be optimized to accurately estimate the probability distribution of population using a smaller number of samples than LHS.

#### 2.2.1 Weight function for a single random variable

In WHLS, the total probability area created by probability density function and horizontal axis (random variable axis) is divided into n intervals, and samples are selected from each interval as similarly as in LHS. These intervals need not have the same probability area, unlike LHS. In LHS method, the total probability area is divided into n intervals with the same areas, and n samples are selected from these intervals. Therefore, the probability of selecting each sample is the same as  $\frac{1}{n}$ . On the other hand, in WLHS method, the probability area in each interval is different from each other. Therefore, probabilities of selecting samples

in these intervals are different from each other. As mentioned above, the probability area ( $A_i$ ) of each interval is the same as probability of selecting a sample from that interval, so that it is used as a weight function. This sampling method is referred to as Weighted Latin Hypercube Sampling (WLHS). The weight defined this way is represented by Eq. (1).

$$w_i = A_i \quad (6)$$

where  $i$  refers to a natural number ranging from 1 to  $n$ , the number of sample.  $\sum_{i=1}^n w_i = \sum_{i=1}^n A_i = 1$  should be satisfied.

### 2.2.2 Weight function for two or more random variables

If the probability distribution of population is defined by two random variables  $X$  and  $Y$ , weight  $w_{X,j}$  for a sample selected in the  $j^{\text{th}}$  sample interval ( $a, b$ ) of random variable  $X$  is defined as the probability area of the  $j^{\text{th}}$  interval, and weight  $w_{Y,k}$  for a sample selected in the  $k^{\text{th}}$  sample interval ( $c, d$ ) of random variable  $Y$  is defined in the same way. The probability (or weight  $w_i$ ) that the  $j^{\text{th}}$  sample of  $X$  and the  $k^{\text{th}}$  sample of  $Y$  can be selected at the same time is the same as probability volume created by joint probability distribution in joint sample region of  $X$  and  $Y$ . This is represented by Eq. (7).

$$w_i = P(a < X \leq b, c < Y \leq d) = \int_c^d \int_a^b f_{XY}(x, y) dx dy \quad (7)$$

where if two variables  $X$  and  $Y$  are mutually independent, the equation above is simplified as follows. Furthermore, in a case where random variables are more than two, weight can be defined in the same way.

$$\begin{aligned} w_i = P(a < X \leq b, c < Y \leq d) &= \int_c^d \int_a^b f_{XY}(x, y) dx dy \\ &= \int_c^d f_Y(y) dy \int_a^b f_X(x) dx = w_{X,j} \times w_{Y,k} \end{aligned} \quad (8)$$

### 2.2.3 Sample mean and sample variance

In the case of using WLHS, the probability of selecting each sample  $X_i$  is the same as weight  $w_i$  as mentioned above. Considering this, weighted sample mean  $\bar{X}$  and unbiased weighted sample variance  $\bar{S}^2$  can be defined as follows. The derivation of these equations is described in Appendix. It is assumed that random variable  $X$  has normal distribution in derivation for unbiased weighted sample variance.

$$\bar{X} = w_1 X_1 + w_2 X_2 + \dots + w_n X_n = \sum_{i=1}^n w_i X_i \quad (9)$$

$$\bar{S}^2 = \frac{1}{1 - \sum_{i=1}^n w_i^2} \sum_{i=1}^n w_i (X_i - \bar{X})^2 \quad (10)$$

If weight  $w_i$  in the two equations above is  $\frac{1}{n}$ , they are simplified to those equations defined in LHS. Eq. (10) is the same as the equation defined by Wiki [7], GNU Scientific Library [8] and Madansky [9] for unbiased weighted sample variance. On the other hand, the derived equation for weighted sample variance by Copper et al. [10] fails to meet the unbiased condition of sample variance. Equations failing to meet unbiased condition have been used for some programs and research results without recognition of such confusion [11].

Equations derived above for WLHS can be easily extended to log-normal distribution. Let  $Y$  has log-normal distribution. Then, ' $X = \ln Y$ ' has normal distribution, so that sample mean  $\bar{X}$  of  $X$  is defined as follows:

$$\begin{aligned} \bar{X} = \ln(\bar{\theta}) &= w_i \\ &= \frac{1}{n} \sum_{i=1}^n \ln Y_i \end{aligned} \quad (11)$$

where  $Y_i$  is a sample selected from the  $i^{\text{th}}$  sample interval of log-normal distribution, and  $\bar{\theta}$  is sample median of population with log-normal distribution and defined as follows:

$$\bar{\theta} = \exp\left(\sum_{i=1}^n w_i \ln Y_i\right) \quad (12)$$

The sample lognormal variance of population,  $\bar{\beta}^2$ , is the same as the sample variance of  $\ln Y$ ,  $\bar{S}_{\ln Y}^2$ . Therefore,  $\bar{\beta}^2$  is defined as follows:

$$\bar{\beta}^2 = \bar{S}_{\ln Y}^2 = \frac{1}{1 - \sum_{i=1}^n w_i^2} \sum_{i=1}^n w_i (\ln Y_i - \ln \bar{\theta})^2 \quad (13)$$

Sampling intervals in WLHS can be selected more flexibly than in LHS, in consideration of users' conveniences or the limits of given data. In other words, the mean and variance of population with normal distribution (or lognormal distribution) can be approximated more flexibly using sample mean (or median) and sample variance (sample lognormal variance) calculated according to the equations presented in this section. Based on this calculation, tail values of the normal distribution of population can also be approximated using the approximated distribution parameters.

## 3. Numerical examples

How accurately WLHS can estimate the probability distribution of

population is reviewed according to procedures such as 1) to assume population whose probability distribution is known, 2) to select samples by different sampling methods such as Random Sampling, Latin Hypercube Sampling and Weighted Latin Hypercube Sampling, and 3) to compare how accurately each sampling method can estimate the distribution parameters. Sampling intervals and weights are properly assumed in WLHS as described in the previous section.

### 3.1 Assumed probability distribution of population

It is assumed that population has standard normal distribution. In other words, the mean and standard deviation of population are assumed as 0 and 1, respectively. Figs. 1 and 2 show probability density function (PDF) and cumulative distribution function (CDF) of that probability distribution. Samples are selected from the population using Random Sampling, LHS and WLHS. Sample mean and standard deviation are calculated, and their accuracy is verified through their comparison with the known mean and standard deviation of population. Note

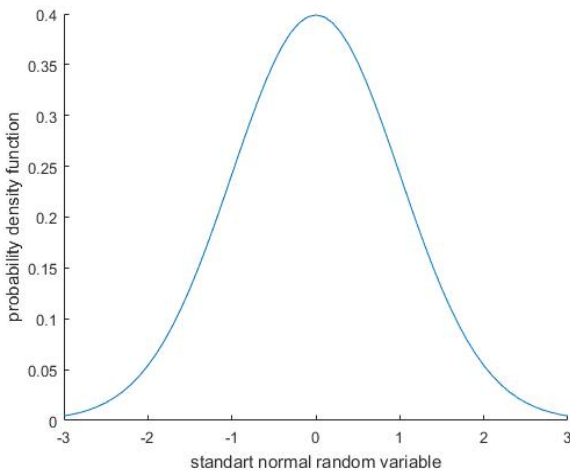


Fig. 1. PDF of the standard normal distribution

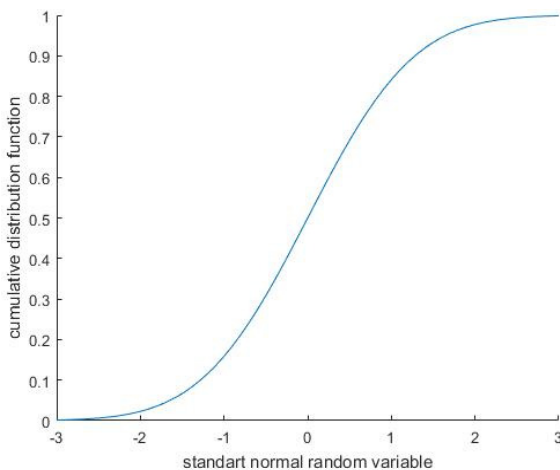


Fig. 2. CDF of the standard normal distribution

that since it is shown in Section 2 how to estimate distribution parameters for an example having log-normal distribution, one can easily conduct analysis for it. However, the example having standard normal distribution is considered in this paper.

### 3.2 Weighted Latin Hypercube Sampling

Samples are selected from population using Random Sampling, Latin Hypercube Sampling and Weighted Latin Hypercube Sampling. In Random Sampling, samples are selected randomly from standard normal distribution. In Latin Hypercube Sampling and Weighted Latin Hypercube Sampling, sample intervals are set according to the procedures mentioned in Section 2, and samples are selected at the probability-area center in each interval. To select samples in WLHS, sample weight  $w_i$  should be defined according to Section 2.2. As described in Eq. (6), the weight can be considered as the probability area of each sampling interval,  $A_i$ . Sampling interval can be easily defined by making the probability area of each interval of random variable equal to weight,  $w_i$ . The weight  $w_i$  can be defined depending on distribution of collected data, or empirical intuition of researchers.

In this example, weight  $w_i$  is chosen to have the following definition.

$$w_i = A_i = \frac{c}{i^\alpha} \tag{14}$$

where  $\alpha$  is a positive real number, and normalization coefficient  $c = [\sum_{i=1}^n \frac{1}{i^\alpha}]^{-1}$  is defined to meet  $\sum_{i=1}^n w_i = 1$ . When  $\alpha$  is 0,  $w_i = \frac{1}{n}$  is established, so that the same samples as in LHS are selected. In this example, the results of WLHS produced with the change of  $\alpha$  are compared with the results from RS and LHS. As each sampling interval can be defined using that weight, the center of probability area of sampling interval is selected as sample point according to the following detail procedures.

- 1) Weight  $w_i$  is arranged in the order to be applied to desired locations from the left side to the right side of probability distribution. In this example, weight is selected to be symmetric with respect to the mean. Fig. 3 shows weight  $w_i$  under the conditions of  $\alpha = 0.5$  and the number of samples,  $n = 10$ . It can be found that the weights of WLHS are larger in the middle and smaller in both tails than those of LHS.
- 2) The center of probability area of each sampling interval is located using cumulative distribution function. The center of probability area of the 1<sup>st</sup> sample interval is under the condition of cumulative distribution function  $F_1 = \frac{A_1}{2} = \frac{w_1}{2}$ , and that of the  $i^{\text{th}}$  sample

interval is under the condition of  $F_i = \frac{F_{i-1} + (w_{i-1} - w_i)}{2}$ . As this example assumes that the population follows standard normal distribution,  $F_i$  is equivalent to standard normal cumulative distribution

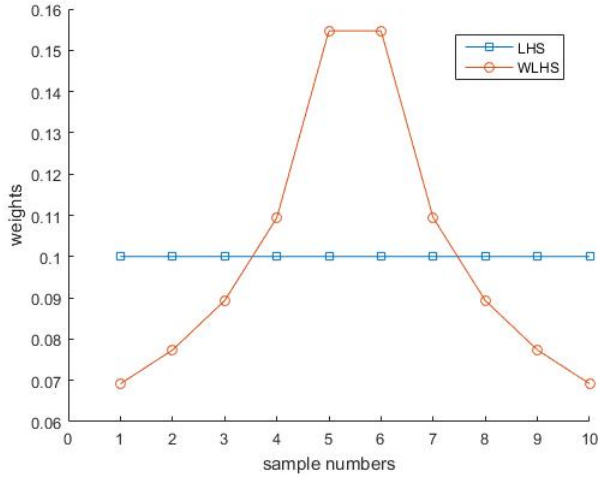


Fig. 3. Symmetric weights of samples in WLHS and LHS

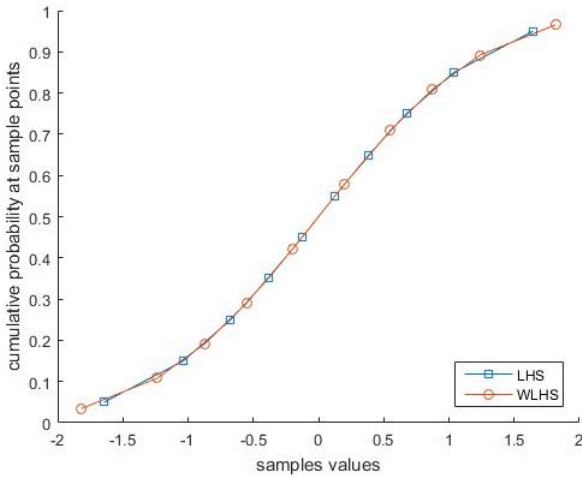


Fig. 4. Cumulative probability vs. sample values

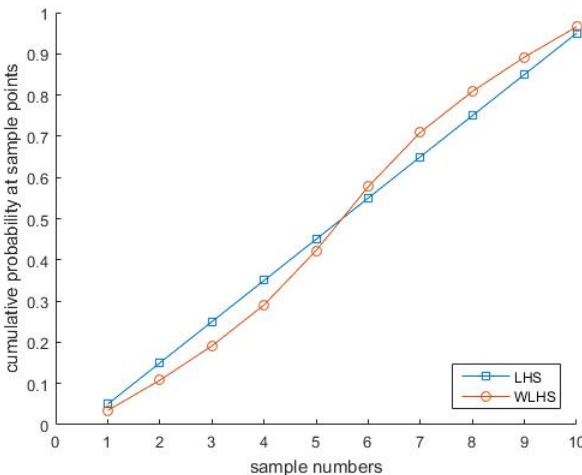


Fig. 5. Cumulative probability vs. sample numbers

function  $\Phi_i$ .

3) Sample value,  $X_i$ , at the center of probability area of sample interval is found using inverse cumulative distribution function. In other words,  $X_i = F^{-1}(F_i)$  where  $F^{-1}$  refers to inverse cumulative distribution function. Since the population is assumed to follow standard normal distribution,  $F^{-1}$  is equal to  $\Phi^{-1}$ . Fig. 4 shows sample value and cumulative probability at the center of probability area of each sampling interval under the conditions of  $\alpha = 0.5$  and the number of samples,  $n = 10$ . Fig. 5 shows the cumulative probability of samples over their sample numbers. This figure shows that, although samples in LHS are separated by the same intervals in probability, the intervals in WLHS are not same each other and relatively longer in the center and shorter in both tails.

### 3.3 Estimation of distribution parameters

For the sake of convenient calculation in the case of many actual probability and statistics problems, probability distributions are frequently estimated on the assumption that population has normal distribution. This example assumes that population follows standard normal distribution. If the mean and variance of population can be approximately estimated using sample mean and sample variance, the probability distribution of population can be easily approximated by these sample distribution parameters.

Since all samples in LHS have the same weights, the sample mean and sample variance are estimated according to Eqs. (1) and (2) in Section 2.1, respectively. Likewise, since all samples in RS have the same weights, the same equations as in LHS are used. On the other hand, since samples in WLHS have different weights, the sample mean and sample variance are estimated according to Eqs. (9) and (10) in Section 2.2, respectively.

Figs. 6 and 7 show the sample mean and sample standard deviation that vary according to the number of samples under the condition of  $\alpha = 0.5$ . The sample mean under LHS and WLHS is exactly the same as population mean. The reason is that LHS and WLHS selected samples symmetrically in this example so that the sample mean could be maintained to be exact to the population mean. In the cases of LHS and WLHS, mean and standard deviation of population are estimated quite efficiently and accurately comparing to RS. As shown in these figures, large enough number of samples in RS are necessary to approximate the mean and standard deviation of population.

Figs. 8 and 9 show sample mean and sample standard deviation that vary according to the number of samples under the condition of  $\alpha = 1.0$  respectively. Fig. 8 shows that the sample means under LHS and WLHS are exactly the same as population mean due to the above-mentioned reason. Fig. 9 shows that sample standard deviation under WLHS esti-

mates population standard deviation less accurately compared to LHS. The reason is that, as samples are concentrated on both tails with the increase of  $\alpha$ , sample standard deviation is overestimated compared with population standard deviation. This trend become more extreme

with the increase of  $\alpha$ . In the case of using WLHS, samples are relatively concentrated on both tails, and thus sample standard deviation can be estimated at larger than population standard deviation. It is necessary to conduct a study in future on how samples can be properly distributed considering this effect.

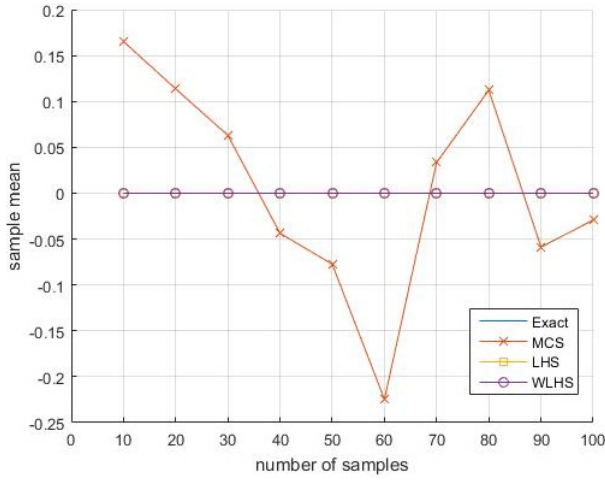


Fig. 6. Sample mean over number of samples when  $\alpha = 0.5$

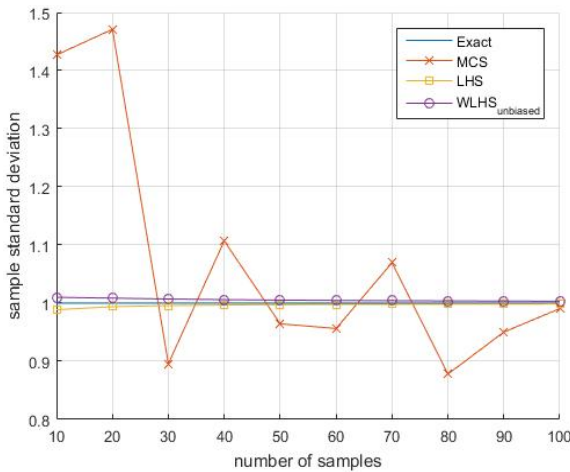


Fig. 7. Sample standard deviation over number of samples when  $\alpha = 0.5$

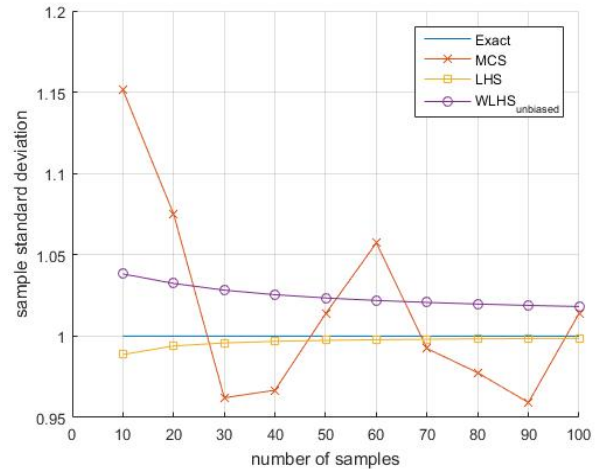


Fig. 9. Sample standard deviation over number of samples when  $\alpha = 1.0$

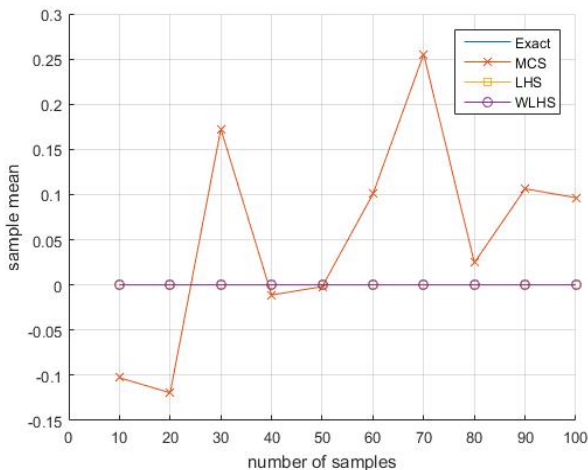


Fig. 8. Sample mean over number of samples when  $\alpha = 1.0$

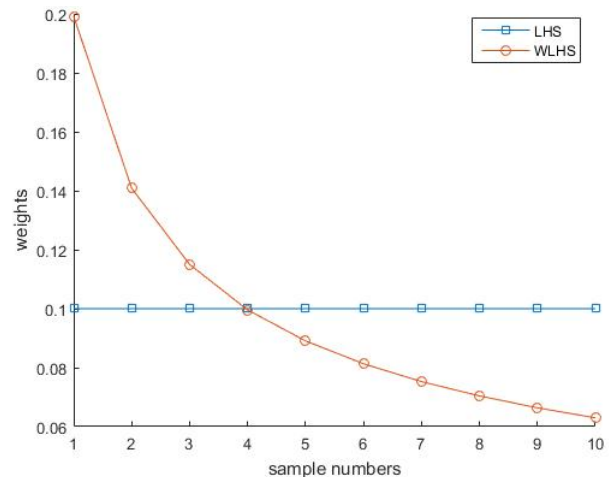


Fig. 10. Asymmetric weights of samples in WLHS and LHS

from the left to the right on the graph, weight becomes lower. This means that samples are selected more densely in the right tail.

Figs. 11 and 12 respectively show sample mean and sample standard deviation that vary according to the number of samples under the condition of  $\alpha = 0.5$ . Fig. 11 shows that sample mean of LHS is exactly the same as population mean, but that sample mean of WLHS estimates population mean with a little error. As WLHS selects sample asymmetrically, sample mean is biased toward where samples are more concentrated. As shown in Fig. 12, LHS estimates population standard deviation more accurately than WLHS.

Figs. 13 and 14 respectively show sample mean and sample standard deviation that vary according to the number of samples under the condition of  $\alpha = 1.0$ . Fig. 13 shows that sample mean of LHS is exactly the same as population mean, but that sample mean of WLHS estimates population mean with a little error. This is attributed to the same reason as that of Fig. 11. In Fig. 14, LHS estimates population standard deviation more accurately than WLHS. Therefore, it is necessary to

study in future on how to distribute samples effectively in WLHS in order to minimize the difference between the sample standard deviation and the population standard deviation.

#### 4. Application to seismically isolated structures in nuclear power plants

Huang et al. [1] considered the uncertainties using LHS for a study on seismic performance evaluation of seismically isolated structures in NPPs. The same numerical example as Huang et al. [1] is used in this paper with the site-specific ground motions of the nuclear power plant in Diablo Canyon, which is western region in the USA. In this numerical example, clearance-to-stops (CSSs) calculated using LHS proposed by Huang et al. [1] and WLHS, respectively, are compared to investigate the effect of using different sampling techniques.

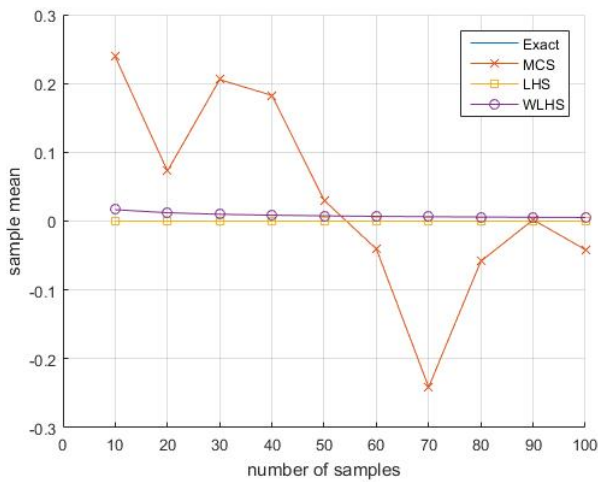


Fig. 11. Sample mean over number of samples when  $\alpha = 0.5$  and asymmetrically sampled

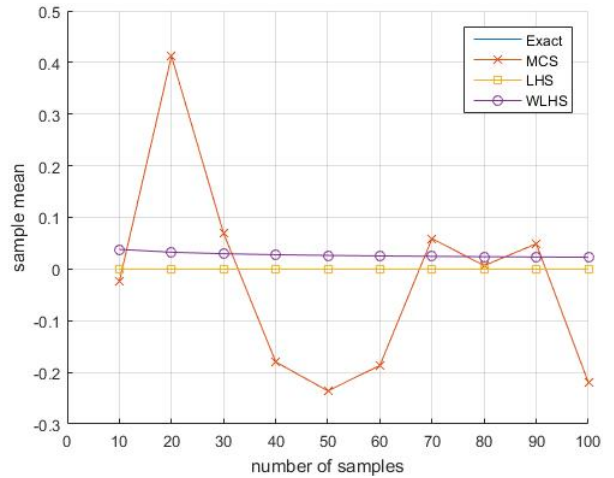


Fig. 13. Sample mean over number of samples when  $\alpha = 1.0$  and asymmetrically sampled

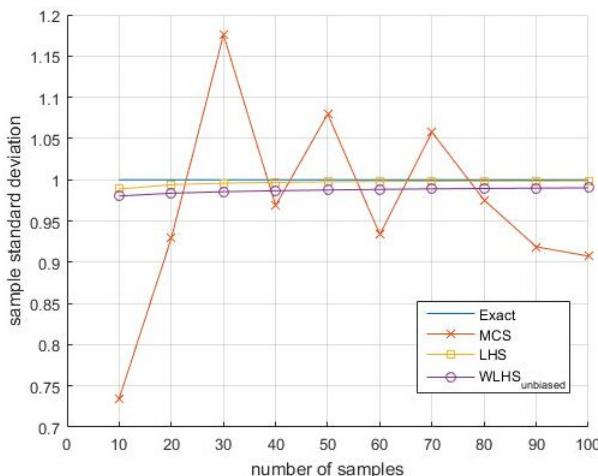


Fig. 12. Sample standard deviation over number of samples when  $\alpha = 0.5$  and asymmetrically sampled

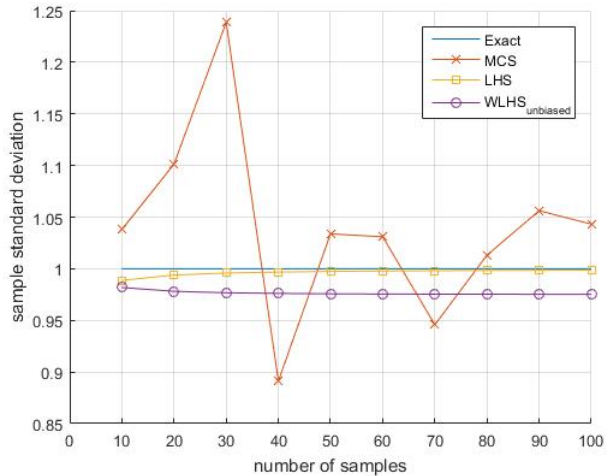


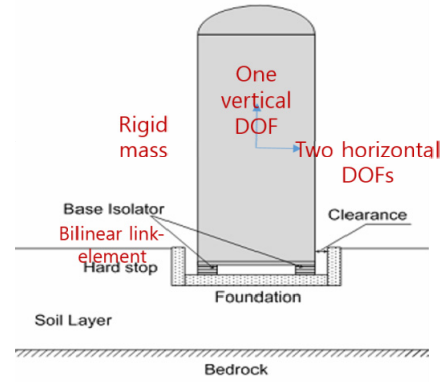
Fig. 14. Sample standard deviation over number of samples when  $\alpha = 1.0$  and asymmetrically sampled

### 4.1 Lumped-mass model for a seismically isolated structure

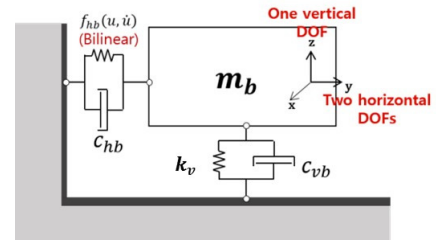
Huang et al. [1] represented behaviours of seismically isolated structures by a lumped-mass model as shown in Fig. 15, to which seismic isolation devices such as friction-pendulum bearing, low-damping rubber bearing, and lead-rubber bearing were applied. In this paper, among these three isolation devices, a lead-rubber bearing is selected to implement for the lumped-mass model. The lumped-mass model represents an upper structure as a lumped mass  $m_b$ , which has two horizontal degrees of freedom and one vertical degree of freedom as shown in Fig. 15. The spring and damper are introduced to express stiffness and damping effect in each degree of freedom. The restoring force of nonlinear spring in two horizontal directions is expressed as a bilinear function  $f_h(u, \dot{u})$ , and the linear spring stiffness in vertical direction is expressed as constant  $k_v$ . The restoring force of nonlinear spring in horizontal direction can be represented by a bilinear function, which is defined with secondary horizontal stiffness  $k_2$ , yield strength  $f_y$ , and yield displacement  $u_y$  as shown in Fig. 16. Yield strength can be alternatively expressed by the y-intercept  $Q_d$ , and primary horizontal stiffness  $k_1$  can be calculated as dividing yield strength  $f_y$  by yield displacement  $u_y$ . The present paper employs OpenSEES (2016) to perform nonlinear response history analyses (RHA) using this lumped-mass model.

For known horizontal yield displacement  $u_y$ , y-intercept  $Q_d$  and horizontal natural period  $T_d$ , one can easily determine the first and secondary horizontal stiffness  $k_1$  and  $k_2$ , respectively, and the yield strength  $f_y$ . For known  $T_v$ , one can easily determine the vertical stiffness  $k_v$ . In the seismic isolation system, we assume that the horizontal yield displacement is 25 mm and the vertical natural period  $T_y$  is 0.05 sec. Therefore,  $Q_d$  and  $T_d$  are considered as random variables in this numerical example. Nine isolation systems are considered by setting three cases of  $T_d$  such as 2, 3, and 4 sec, and three cases of  $Q_d$  such as 0.03  $W$ , 0.06  $W$ , and 0.09  $W$ . Among these, only two examples, representing the smallest and the largest differences between CSs calculated by LHS and WLHS, are selected to be shown.

A probability distribution of properties of seismic isolation systems occur during manufacturing process and quality control for seismic isolation units. Scale factors are defined to represent this distribution as these are multiplied with the mean values of properties to generate samples. The mean value of the scale factor is one, and the standard deviation are defined according to cases where quality control is excellent or good [1]. If the properties under excellent quality control have 95% confidence within  $\pm 10\%$  interval from the best-estimated value, it has a normal distribution with 0.05 of standard deviation. If the properties under good quality control have 95% confidence within  $\pm 20\%$  interval from the best-estimated value, it has a normal distribution with 0.1 of



(a) Simplified shape of isolated structure



(b) Analysis model with a lumped mass

Fig. 15. Isolated structure in nuclear power plants

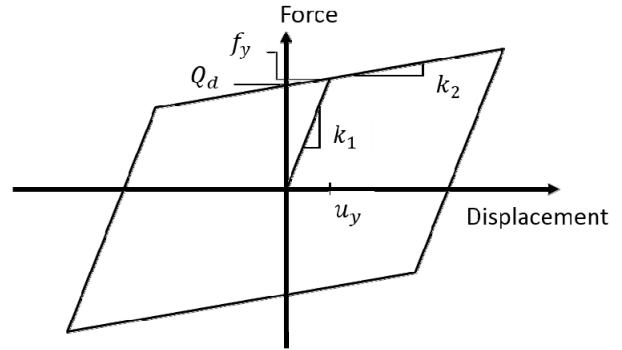


Fig. 16. Hysteretic bilinear response of a LR bearing in horizontal direction

standard deviation. Table 1 show 30 samples of the scale factor extracted using LHS and WLHS for cases whose quality control is excellent and good. In these numerical examples, we consider three properties,  $k_2$ ,  $Q_d$  and  $k_v$ , as random variables under excellent quality control, and  $\alpha$  in Eq. (14) is assumed to be 0.5 for determining sample weights in WLHS.

### 4.2 Input ground motion

In this example, 30 sets of input ground motions are selected as same as Huang et al. [1], and matched with the site-specific design spectrum of nuclear power plants in Diablo Canyon, western region in the USA. The matched input ground motions are considered as Design



**Table 1.** Scale factors of LHS and WLHS for isolation system

No.	Latin hypercube sampling		Weighted latin hypercube sampling	
	Excellent control	Good control	Excellent control	Good control
1	0.894	0.787	0.884	0.768
2	0.918	0.836	0.906	0.813
3	0.931	0.862	0.919	0.837
4	0.940	0.881	0.928	0.855
5	0.948	0.896	0.935	0.870
6	0.955	0.910	0.941	0.883
7	0.961	0.922	0.947	0.895
8	0.966	0.933	0.953	0.906
9	0.971	0.943	0.958	0.916
10	0.976	0.952	0.963	0.927
11	0.981	0.961	0.969	0.937
12	0.985	0.970	0.974	0.948
13	0.989	0.979	0.980	0.960
14	0.994	0.987	0.987	0.973
15	0.998	0.996	0.995	0.990
16	1.002	1.004	1.005	1.010
17	1.006	1.013	1.013	1.027
18	1.011	1.021	1.020	1.040
19	1.015	1.030	1.026	1.052
20	1.019	1.039	1.031	1.063
21	1.024	1.048	1.037	1.073
22	1.029	1.057	1.042	1.084
23	1.034	1.067	1.047	1.094
24	1.039	1.078	1.053	1.105
25	1.045	1.090	1.059	1.117
26	1.052	1.104	1.065	1.130
27	1.060	1.119	1.072	1.145
28	1.069	1.138	1.081	1.163
29	1.082	1.164	1.094	1.187
30	1.106	1.213	1.116	1.232

Basis Earthquakes (DBE). A single set of the input ground motions consists of two horizontal ground motions and one vertical ground motion.

It is reasonable to make intensity of input ground motions different in each direction if input ground motions in both directions are produced from one Uniform Hazard Response Spectrum (UHRS). Huang et al. [1] introduced the ratio,  $F_h$ , of input ground motion in horizontal direction and the ratio,  $F_v$ , in vertical direction to the geometric means. These ratios are considered as random variables, and their distributions are defined based on statistics of measured earthquakes. For  $F_h$ , the median value,  $\theta_h$ , is 1.3 and the logarithmic standard deviation,  $\beta_h$ , is 0.13. For  $F_v$ , the median value,  $\theta_v$ , is 1.0 and the logarithmic standard deviation,  $\beta_v$ , is 0.18. To maintain the same geometric mean in horizontal directions, the strong-axis ground motion is defined as  $F_h$  times of the geometric mean while the weak-axis ground motion in the

**Table 2.** Scale factors of LHS and WLHS for input ground motions

No.	Latin hypercube sampling		Weighted latin hypercube sampling	
	Horizontal direction	Vertical direction	Horizontal	Vertical
1	0.986	0.682	0.961	0.658
2	1.050	0.744	1.019	0.714
3	1.086	0.780	1.052	0.746
4	1.113	0.807	1.077	0.771
5	1.136	0.830	1.098	0.791
6	1.156	0.850	1.117	0.810
7	1.174	0.868	1.134	0.827
8	1.191	0.886	1.150	0.844
9	1.207	0.902	1.166	0.860
10	1.222	0.918	1.182	0.877
11	1.236	0.933	1.198	0.893
12	1.251	0.948	1.215	0.911
13	1.265	0.963	1.234	0.930
14	1.279	0.978	1.256	0.953
15	1.293	0.993	1.284	0.983
16	1.307	1.008	1.317	1.018
17	1.321	1.023	1.346	1.049
18	1.336	1.039	1.369	1.075
19	1.351	1.055	1.390	1.098
20	1.367	1.072	1.410	1.119
21	1.383	1.090	1.430	1.141
22	1.401	1.109	1.449	1.162
23	1.419	1.129	1.469	1.185
24	1.439	1.151	1.491	1.209
25	1.462	1.176	1.514	1.235
26	1.488	1.205	1.539	1.264
27	1.518	1.239	1.569	1.298
28	1.556	1.283	1.606	1.340
29	1.610	1.345	1.658	1.401
30	1.714	1.467	1.759	1.519

orthogonal direction  $\frac{1}{F_h}$  times. Thirty samples of  $F_h$  (Table 2) are extracted using LHS and WLHS ( $\alpha$  in Eq. (14) is assumed to be 0.5 for determining sample weights), respectively, and then the maximum-minimum spectra compatible ground motions are defined by multiplying the samples of  $F_h$  or  $\frac{1}{F_h}$  with 30 sets of the selected ground motions. Fig. 17(a), (b) show the distributed maximum-minimum spectra compatible ground motions of the design basis earthquake in horizontal directions determined by WLHS. The similar figures determined by LHS can be found in Huang et al. [1].

In the same way, thirty samples of  $F_v$  (Table 2) are extracted using LHS and WLHS, respectively, and then the samples of ground motions are determined by multiplying the samples of  $F_v$  with 30 sets of the selected ground motions. Fig. 17(c) show the distributed ground

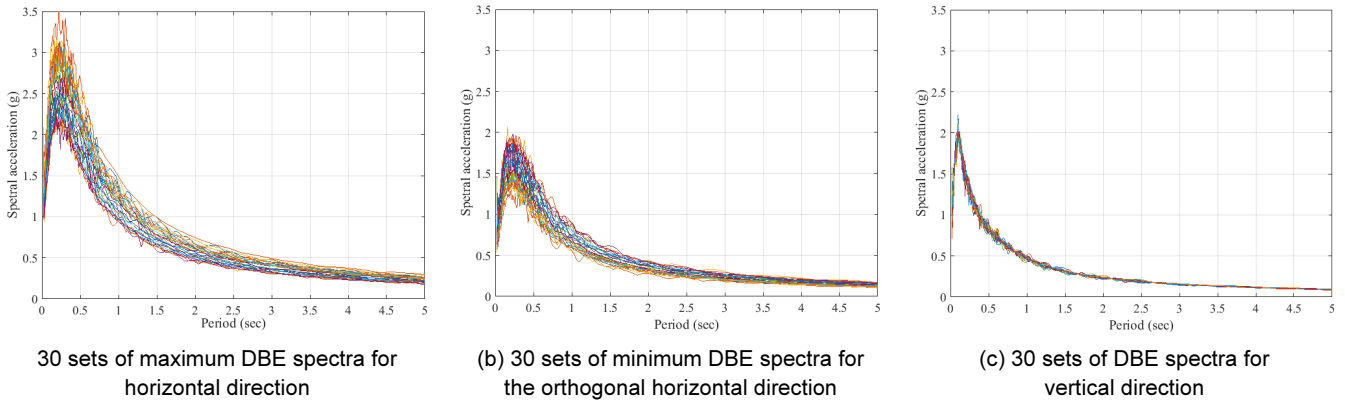


Fig. 17. 5% damped response spectra for the Diablo Canyon site by WLHS

motions of the design basis earthquake in vertical direction determined by WLHS. The similar figures determined by LHS also can be found in Huang et al. [1].

### 4.3 Effect of WLHS on CS

The effect of sampling techniques on CS is discussed in this section. Since CS is defined as 90%-ile displacement under 150% DBE according to ASCE 4 draft [12], the DBE shown in Fig. 17 are amplified by 150% and then applied as input ground motions to the lumped-mass model in Fig. 15(b). Among the 9 cases of isolation systems mentioned in Section 4.1, only two examples, representing the smallest and the largest differences between CSs calculated by LHS and WLHS, are selected to be shown in this section. The CDF of displacements based on the isolation system with  $T_d$  2 sec and  $Q_d$  0.03W and isolation system with  $T_d$  3 sec,  $Q_d$  0.03W are shown in Figs. 18 and 19, respectively. Figs. 18(a) and 19(a) compare the CDF of sample displacements and the corresponding estimated CDF calculated by LHS and WLHS. Figs. 18(b) and 19(b) show the enlarged view of the right tails in the CDFs. Here, the blue points indicate samples acquired through LHS and the red points indicates samples through WLHS. The estimated CDF is shown with dotted lines by calculating sample medians and sample logarithmic standard deviations.

According to Fig. 18(b), the calculated CSs (90%-ile displacements based on the estimated CDFs) by LHS and WLHS are 1080 mm and 1076 mm, respectively, so that the choice of sampling techniques essentially does not make any difference. However, the choice of the sampling techniques makes 16% difference in CS since the calculated CSs by LHS and WLHS are 1589 mm and 1549 mm, respectively, as shown in Fig. 19(b). These results indicate that a choice of sampling techniques can cause significant difference in CS. Therefore, it is necessary to standardize the sampling technique for calculating CS.

Figs. 18(a) and 19(a) show that there are significant differences between the samples and the estimated CDFs. This occurs because the

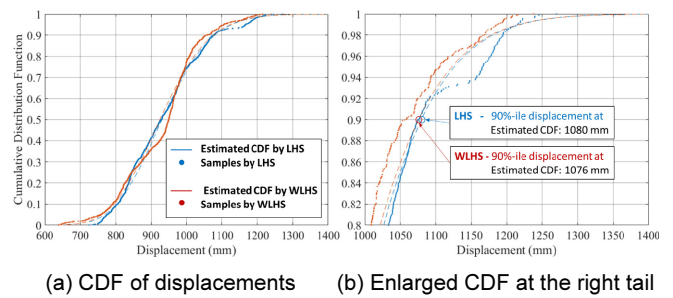


Fig. 18. Compare displacements under 150% DBE by LHS and WLHS (with isolation system: = 2 sec, = 0.03 W)

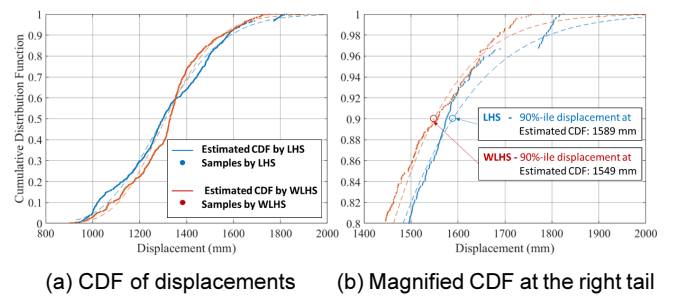


Fig. 19. Compare displacements under 150% DBE by LHS and WLHS (with isolation system: =3 sec, =0.03 W)

assumed lognormal distribution is different from the actual distribution of population, and/or because there is randomness in sampling procedure. This difference can be reduced by extended case studies and in-depth reviews on the sampling procedure.

## 5. Conclusion

This study presented Weighted Latin Hypercube Sampling which is expansion of Latin Hypercube Sampling by applying weights to samples. WLHS allows more flexible way of selecting samples. The results are summarized as follows:

- 1) The WLHS, which is an extension of Latin Hypercube Sampling (LHS)

to avoid the necessity of using sampling intervals having the same probability area, has been proposed. WLHS provides more flexible way on selecting samples by using sampling intervals having different probability areas.

- 2) To approximate distribution parameters of population, the equations and procedures for calculating sample mean and sample variance of normal distribution (or sample median and sample lognormal variance of log-normal distribution) were clearly described with consideration of applying sample weights.
- 3) Accuracy of WLHS estimation on distribution parameters is depending on the selection of weight function. By comparing distribution parameters estimated by WLHS with the results of RS and LHS through numerical examples, it is found that distribution parameters can be estimated quite accurately using proper weights.
- 4) In WHLS where samples are distributed symmetrically with respect to mean, population mean is found to be accurately estimated by sample mean. On the other hand, in a case where they are asymmetrically distributed, the accuracy of an estimation of population mean is found to be relatively low.
- 5) In WHLS where samples are distributed symmetrically with respect to mean, samples are relatively concentrated on both tails, so that sample standard deviation is overestimated compared to population standard deviation.
- 6) In WLHS where samples are distributed asymmetrically, LHS is found to estimate population mean and population standard deviation more accurately than WLHS.
- 7) There is a need for a further study on how to distribute samples properly in WLHS to estimate distribution parameters with sufficient accuracy.
- 8) A choice of sampling techniques between LHS and WLHS can cause significant difference in determining CS. Since it would be better to minimize this randomness caused by choice of sampling techniques, reviews on the sources of this differences should be performed in future study.
- 9) There are significant differences between the samples and the estimated CDFs. This occurs because the assumed lognormal distribution is different from the actual distribution of population, and/or because there is randomness in sampling procedure. This difference can be reduced by extended case studies and in-depth reviews on the sampling procedure.
- 10) Although ASCE 43 [2] requires the probabilistic design technique in NPP, the sampling method in the probabilistic design technique is not clearly mentioned. The sampling method proposed in the present study can be one alternative. In order to guarantee consistent design regardless of choice of sampling techniques, it is necessary to establish a consistent sampling technique based on numerous examples in future.

## / ACKNOWLEDGEMENT /

This work was supported by the Nuclear Research and Development Program in the form of a National Research Foundation Grant funded by the Ministry of Science, ICT and Future Planning (MSIP) in the Republic of Korea. This research (ID: 2014151010170B) was also supported by Korea Institute of Energy Technology Evaluation and Planning (KETEP).

## / REFERENCES /

1. Nuclear Standards Committee, Seismic Design Criteria for Structures, Systems, and Components in Nuclear Facilities, American Society of Civil Engineers, Reston, VA, c2005.
2. Huang YN, Whittaker AS, Kennedy RP, Mayes RL, Assessment of base-isolated nuclear structures for design and beyond-design basis earthquake shaking. Tech. Rep. MCEER-09-0008, Buffalo Univ., New York, c2009 Aug.
3. Hammersley JM, Handscomb DC, Monte carlo methods, Methuen, London, c1964.
4. Rubinstein RY, Kroese DP, Simulation and the monte carlo method, John Wiley & Sons; c2016 Oct 20.
5. Siegmund D, Importance sampling in the Monte Carlo study of sequential tests, The Annals of Statistics, 1976 Jul 1:673-684.
6. McKay MD, Beckman RJ, Conover WJ, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 1979 May 1;21(2):239-245.
7. The Wiki [Internet]. Weighted arithmetic mean, 2016. Available from: [https://en.wikipedia.org/wiki/Weighted\\_arithmetic\\_mean#Weighted\\_sample\\_covariance](https://en.wikipedia.org/wiki/Weighted_arithmetic_mean#Weighted_sample_covariance).
8. GNU Scientific Library. Reference Manual: Weighted Samples [Internet], Gnu.org. Retrieved 2017 Dec 22. Available from: [https://www.gnu.org/software/gsl/manual/html\\_node/Weighted-Samples.html](https://www.gnu.org/software/gsl/manual/html_node/Weighted-Samples.html)
9. Madansky A, Alternative Approaches to Significance Testing with Weighted Means: The Analytical Group Inc, Page [Internet]. Available from: <http://www.analyticalgroup.com/download/Alternative%20Approaches.pdf>
10. Cooper H, Hedges LV, Valentine JC, editors, The handbook of research synthesis and meta-analysis, Russell Sage Foundation; c2009 Feb 5.
11. Maletta H, Weighting: SPSS Tutorials Page [Internet], 2007 Mar 12. Available from: <http://spsstools.net/static/tutorials/WEIGHTING.pdf>
12. Nuclear Standards Committee, Seismic analysis of safety-related nuclear structures and commentary, American Society of Civil Engineers, Reston, VA, c2013.
13. Stone CJ, A course in probability and statistics, Belmont: Duxbury Press; c1996.

**/ Appendix /**

In this appendix, equations for defining sample mean and unbiased sample variance is derived using Weighted Latin Hypercube Sampling. Before deriving this, it is helpful to review the process of deriving equations for sample mean and unbiased sample variance using Latin Hypercube Sampling. In LHS, each sample has the same selection probability (or weights)  $\frac{1}{n}$ , so that the sample mean  $\bar{X}$  can be written as follows:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \tag{A-1}$$

where  $X_i$  refers to the  $i^{\text{th}}$  sample, and  $n$  refers to the number of samples. Sample variance  $\bar{S}^2$  can be defined as follows where sample mean  $\bar{X}$  is used.

$$\bar{s}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \tag{A-2a}$$

$$= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2 \tag{A-2b}$$

$$= \frac{1}{n} \sum_{i=1}^n \{ (X_i - \bar{X})^2 + (\bar{X} - \mu)^2 + 2(X_i - \bar{X})(\bar{X} - \mu) \} \tag{A-2c}$$

$$= \frac{1}{n} \sum_{i=1}^n \{ (X_i - \bar{X})^2 + (\bar{X} - \mu)^2 \} + \frac{2(\bar{X} - \mu)}{n} \sum_{i=1}^n (X_i - \bar{X}) \tag{A-2d}$$

Since sample mean  $\bar{X}$  and population mean  $\mu$  do not vary according to  $i$ , they can move out from the summation in the last term of Eq. (A-2c). Note that Eq. (A-3) is easily derived based on Eq. (A-1).

$$\sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n X_i - n\bar{X} = 0 \tag{A-3}$$

Because of Eq. (A-3), the last term of Eq. (A-2d) is removed. Applying unbiased variance condition, the expectation of sample variance (Eq. (A-2)) should be the same as population variance, Eq. (A-4) is derived.

$$\sigma^2 = E[\bar{S}^2] = E\left[ \frac{1}{n} \sum_{i=1}^n \{ (X_i - \bar{X})^2 + (\bar{X} - \mu)^2 \} \right] \tag{A-4}$$

By multiplying the both sides by number of samples,  $n$ , the above equation becomes

$$n\sigma^2 = E\left[ \sum_{i=1}^n (X_i - \bar{X})^2 \right] + nE[(\bar{X} - \mu)^2] \tag{A-5}$$

where if  $X_i$  follows normal distribution with population variance  $\sigma^2$ ,  $\bar{X}$  follows normal distribution with variance  $\frac{\sigma^2}{n}$ . Therefore, the 2<sup>nd</sup> term on the right hand side in Eq. (A-5) is as follows:

$$E[(\bar{X} - \mu)^2] = \text{var}(\bar{X}) = \frac{\sigma^2}{n} \tag{A-6}$$

Applying Eq. (A-6) to Eq. (A-5) gives

$$(n-1)\sigma^2 = E\left[ \sum_{i=1}^n (X_i - \bar{X})^2 \right] \tag{A-7}$$

Eq. (A-7) can be rewritten for  $\sigma^2$  using unbiased variance condition as follows:

$$\sigma^2 = E[\bar{S}^2] = E\left[ \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \right] \tag{A-8}$$

Let expectations in both sides of Eq. (A-8) are removed. Then, unbiased sample variance is derived as follows:

$$\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \tag{A-9}$$

If square-root is taken in Eq. (A-9), unbiased sample standard deviation  $\bar{S}$  is represented by Eq. (A-10) [13].

$$\bar{S} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \tag{A-10}$$

Sample mean and sample variance of WLHS can be derived according to the same process and procedure as those in LHS. In WLHS, each sample has different selection probability (or weights)  $w_i$ , and so sample mean  $\bar{X}$  can be defined as follows:

$$\bar{X} = w_1 X_1 + w_2 X_2 + \dots + w_n X_n = \sum_{i=1}^n w_i X_i \tag{A-11}$$

where  $X_i$  refers to the  $i^{\text{th}}$  sample;  $n$  refers to the number of samples. The sum of weight  $w_i$  should be 1,  $1 = \sum_{i=1}^n w_i$ . Sample variance  $\bar{S}^2$  can be defined and developed as follows using  $\bar{X}$ .

$$s^{-2} = \sum_{i=1}^n w_i (X_i - \mu)^2 \quad (\text{A-12a})$$

$$= \sum_{i=1}^n w_i (X_i - \bar{X} + \bar{X} - \mu)^2 \quad (\text{A-12b})$$

$$= \sum_{i=1}^n w_i \{ (X_i - \bar{X})^2 + (\bar{X} - \mu)^2 + 2(X_i - \bar{X}) + (\bar{X} - \mu) \} \quad (\text{A-12c})$$

$$= \sum_{i=1}^n w_i \{ (X_i - \bar{X})^2 + (\bar{X} - \mu)^2 \} + 2(\bar{X} - \mu) \sum_{i=1}^n w_i (X_i - \bar{X}) \quad (\text{A-12d})$$

Since sample mean  $\bar{X}$  and population mean  $\mu$  do not vary according to  $i$ , they can move out from the summation in the last term of Eq. (A-12d). The following equation is easily derived based on  $1 = \sum_{i=1}^n w_i$  and Eq. (A-11).

$$\sum_{i=1}^n w_i (X_i - \bar{X}) = \sum_{i=1}^n w_i X_i - \bar{X} \sum_{i=1}^n w_i = 0 \quad (\text{A-13})$$

Because of Eq. (A-13), the last term of Eq. (A-12d) is removed. Applying unbiased variance condition, the expectation of sample variance (Eq. (A-12)) should be the same as population variance, Eq. (A-14) is derived.

$$\sigma^2 = E[\bar{S}^2] = E\left[ \sum_{i=1}^n w_i (X_i - \bar{X})^2 + \sum_{i=1}^n w_i (\bar{X} - \mu)^2 \right] \quad (\text{A-14})$$

As sample mean  $\bar{X}$  and population mean  $\mu$  do not vary according to  $i$ , they can move out from the summation of the last term in Eq. (A-14).

$$\sigma^2 = E\left[ \sum_{i=1}^n w_i (X_i - \bar{X})^2 \right] + E\left[ (\bar{X} - \mu)^2 \sum_{i=1}^n w_i \right] \quad (\text{A-15})$$

where  $1 = \sum_{i=1}^n w_i$ . Let  $X_i$  have normal distribution with population variance  $\sigma^2$ . Then,  $\bar{X}$ , which is defined by linear combination (Eq.

(A-11) of  $X_i$ , also has normal distribution. Therefore, the last term of the equation above is

$$E[(\bar{X} - \mu)^2] = \text{var}(\bar{X}) \quad (\text{A-16a})$$

$$= \text{var}\left( \sum_{i=1}^n w_i X_i \right) \quad (\text{A-16b})$$

$$= \sum_{i=1}^n w_i^2 \times \text{var}(X_i) \quad (\text{A-16c})$$

$$= \sigma^2 \sum_{i=1}^n w_i^2 \quad (\text{A-16d})$$

Applying this equation to Eq. (A-15), Eq. (A-17) is derived.

$$\left( 1 - \sum_{i=1}^n w_i^2 \right) \sigma^2 = E\left[ \sum_{i=1}^n w_i (X_i - \bar{X})^2 \right] \quad (\text{A-17})$$

Eq. (A-17) is rewritten for  $\sigma^2$  using unbiased variance condition, Eq. (A-18) is derived.

$$\sigma^2 = E[\bar{S}^2] = E\left[ \frac{1}{1 - \sum_{i=1}^n w_i^2} \sum_{i=1}^n w_i (X_i - \bar{X})^2 \right] \quad (\text{A-18})$$

Let expectations in both sides of Eq. (A-18) removed. Then, unbiased sample variance is derived as follows:

$$\bar{S}^2 = \frac{1}{1 - \sum_{i=1}^n w_i^2} \sum_{i=1}^n w_i (X_i - \bar{X})^2 \quad (\text{A-19})$$

If square-root is taken in Eq. (A-19), unbiased sample standard deviation  $\bar{S}$  is represented by Eq. (A-20).

$$\bar{S} = \sqrt{\frac{1}{1 - \sum_{i=1}^n w_i^2} \sum_{i=1}^n w_i (X_i - \bar{X})^2} \quad (\text{A-20})$$