Technical Note

Latin hypercube sampling used in the calculation of the fracture probability

Keqin Ding^a, Zegong Zhou^b & Chuntu Liu^a

^a*Institute of Mechanics, Chinese Academy of Sciences, Beijing, 100080, People's Republic of China*

^b*Taiyuan Heavy Machinery Institute, Taiyuan, 030024, People's Republic of China*

(Received 22 January 1996; revised 22 January 1997; accepted 12 May 1997)

The paper proposes Latin hypercube sampling combined with the stratified sampling of variance reduction technique to calculate accurate fracture probability. In the compound sampling, the number of simulations is relatively small and the calculation error is satisfactory. © 1998 Published by Elsevier Science Limited.

1 INTRODUCTION

In recent years, Monte Carlo simulation has been widely used in modern engineering structures. Owing to the complexity of the calculation models, the fact that each computer code usually has many input variables numbering in the hundreds and the fact that it often requires several hours to make a simulation run, only a limited number of simulation runs are feasible. Latin hypercube sampling (LHS) makes simulation runs of the Monte Carlo simulation less attractive to many researchers. But the LHS does not provide any reduction in the variance of the expected value of fracture probability (pf), and its calculative accuracy cannot be satisfied. In this paper a compound sampling in which the LHS is combined with a variance reduction technique named stratified sampling is proposed. As shown in the calculative example, compared with the results obtained from direct sampling 50,000 times, the maximum relative error of the compound sampling with 560 times is only 5.6% and the computer time of the compound sampling is about 1/1000 that of the direct sampling.

2 SAMPLING METHOD

2.1 LHS method^{1,2}

The Latin hypercube technique provides a constrained sampling scheme instead of random sampling according to the direct Monte Carlo method. Traditionally, random numbers are randomly generated between 0 and 1. They

are used to generate random variables according to the prescribed distribution function of each variable. In Latin hypercube sampling, the region between 0 and 1 is uniformly divided into N non-overlapping intervals for each random variable; where N is the number of random numbers which need to be generated for each random variable, i.e., number of simulation cycles. The N non-overlapping intervals are selected to be of the same probability of occurrence. Then, N different values in the N non-overlapping intervals are randomly selected for each random variable, i.e., one value per interval is generated. This can be accomplished by initially generating N random numbers within the range [0,1]. These values represent the percentage position of each generated value of a variable within an interval. Therefore, these values are linearly transformed to the random numbers in the non-overlapping intervals for each random variable using the following equation:

$$u_i = \frac{u}{N} + \frac{(i-1)}{N} \quad (1)$$

where $i = 1, 2, \dots, n$; u = a random number in the range [0,1]; and u_i = random number in the i th interval. From the above equation, it is quite obvious that there is only one generated value that is randomly selected within each of the N intervals for each random variable. This is due to the following relationship:

$$\frac{(i-1)}{N} < u_i < \frac{i}{N} \quad (2)$$

where $(i-1)/N$ and i/N are lower and upper bounds for the i th interval. Then n values obtained for x_1 (the first random

variable) are paired in a random manner (equally likely combination) with the n values of x_2 (the second random variable). These n pairs of (x_1, x_2) are combined in a random manner with the n values of x_3 (the third random variable) to form the first n -triplets (x_1, x_2, x_3) , and so on, until the $(k-2)$ th n -triplets $(x_1, x_2, x_{n-2}, x_{n-1}, x_n)$ (k = number of random variables) are formed. Thus, a $N \times k$ matrix is formed. It can be seen from Ref. 3 that a certain statistical correlation among columns of the matrix (e.g., the spearman rank correlation coefficient) may have a significant influence on the results of simulation. It naturally affects the bias and variance of the estimator.

The estimated values obtained from LHS is stable with respect to the number of simulations, but the LHS cannot provide the reduction in the variance of the estimated value of fracture probability. The conditional expectation (CE) variance reduction technique (VRT) and antithetic variates (AV) VRT can be used in combination with LHS to increase the accuracy of estimated value of fracture probability⁴.

2.2 The stratified sampling⁵

The basic idea of the stratified sampling is to partition the sample space of a dominant random variable, e.g., initial surface crack depth a_0 , into m mutually exclusive subregions. The pre-determined number of simulation in each subregion is proportional to its contribution to total fracture probability. Within each subregion the individual crack sizes are still selected according to the probability density function of a_0 . The probability of a crack of depth lying in the i -th subregion existing in the structure (e.g., the nozzle of a pressure vessel), q_i , is determined by the numerical integration when the probability density function of a_0 and the range of the i -th subregion are known.

The Monte Carlo simulation is used to calculate the p_f for the i -th crack depth subregion as

$$p_f^i = m_i/N_i \times q_i \quad (3)$$

where q_i is the probability of a initial crack having a depth within the boundaries of the i -th subregion; N_i is the number of sampling from the i -th subregion; M_i is the number of samples from the i -th subregion which failed.

The total fracture probability is then

$$p_f^{total} = \sum_{i=1}^m M_i/N_i \times q_i. \quad (4)$$

The total number of biased simulations N is $N = \sum_{i=1}^m N_i$, m is the total number of subregions.

2.3 The stratified sampling with LHS

Although LHS has the advantages in estimating the expected value of the probability of failure with a relatively small number of simulation cycles, it does not provide any reduction in the variance of the estimated probability of failure. In Ref. 4, the conditional expectation and antithetic

variates variance reduction technique are used in combination with LHS to estimate the probability of failure. But when we use these methods to calculate yearly fracture probabilities of a nozzle, the relative errors are very big. It is considered that the stratified sampling can reduce the variance of the estimated fracture probability than that of the conditional expectation method or the antithetic variates method. Therefore the Monte Carlo simulation with the stratified sampling combined with LHS to estimate the fracture probability is proposed in the paper. The basic idea of the compound sampling is as follows: according to the stratified sampling, the sample space of a dominant random variable (e.g., initial surface crack depth a_0 or fracture toughness δ_c) is partitioned into m mutually exclusive subregions and the number of simulation in each subregion is proportional to its contribution to the total fracture probability. Then, the estimated p_f value of each crack size subregion (i.e., each strata of the initial surface crack depth) is obtained by the Monte Carlo simulation with LHS.

3 EXAMPLE

We selected a nozzle of a pressure vessel in Ref. 6 as a calculative example to calculate yearly fracture probability (p_f) of the nozzle.

3.1 Formulae

We adopted $\delta_c \leq \delta(\text{COD})$ as fracture criterion and COD (crack opening displacement) design curve formulas of British Standard PD6493 to calculate applied COD:

$$\delta_c \leq 2\pi a e_y (e/e_y)^2 \quad e/e_y \leq 0.5 \quad (5a)$$

$$\delta_c \leq 2\pi a e_y (e/e_y - 0.25) \quad e/e_y \leq 0.5 \quad (5b)$$

where a = surface crack depth; δ_c = fracture toughness (critical COD); e = applied strain ($e = \sigma_m =$ membrane stress; $E =$ young's modulus); e_y = yield strain.

The fatigue crack depth (a) after applied cyclic stress N times can be calculated by Paris' formula:

$$da/dN = c(\Delta k)^m \quad (6)$$

$$\Delta k = k_t \Delta \sigma \sqrt{\pi a}$$

where c, m material constants of the Paris' formula, N = cycles of cyclic stress; $\Delta \sigma$ amplitude of cyclic stress; k_t the maximum stress concentration factor of the nozzle corner.

The maximum stress concentration factor of the nozzle corner can be calculated by the Decock formula:

$$k_t = \frac{2 + 2d(dt/DT)/D + 1.25d(DT)^{1/2}/D}{1 + t(dT/dt)^{1/2}t} \quad (7)$$

where d = internal diameter of the nozzle; t = wall thickness of the nozzle corner; D = internal diameter of the vessel; T = wall thickness of the vessel.

The vessel and nozzle were made of 16 MnR and 20 g

Table 1.

Random variables	p.d.f	Parameters
a_0	$\lambda e^{\lambda x} / (e^{-\lambda x_l} - e^{-\lambda x_u})$	$\lambda = 0.65, x_l = 0, x_u = 20 \text{ mm}$
σ_m	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$	$\mu = 165.2 \text{ MPa}, \sigma = 24.8 \text{ MPa}$
$\Delta\sigma$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$	$\mu = 82.6 \text{ MPa}, \sigma = 24.8 \text{ MPa}$
c	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right]$	$M = 2.33 \times 10^{-14}, = 2.3 \times 10^{-15}$
σ_y	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$	$\mu = 380.0 \text{ MPa}, \sigma = 22.84 \text{ MPa}$
δ_c	$\frac{\alpha}{\beta} \left(\frac{x}{\beta} \right)^{\alpha-1} \exp \left[-\left(\frac{x}{\beta} \right)^\alpha \right]$	$\alpha = 1.91, \beta = 0.138 \text{ mm}$

and its strengths were all lower than 540 mpa. According to experiment results, the fatigue crack growth rates of the chinese pressure vessel steels (i.e., 16 MnR and 20 g) can be expressed as

$$da/dN = 2.33 \times 10^{-14} (\Delta k)^{3.26} \quad (8)$$

where c is treated as a random variable and its mean value is assumed to be 2.33×10^{-14} , $da/dN = c(\Delta k)^{3.26}$, then the fatigue crack depth(a) after the applied cyclic stress N times is as follows:

$$a = \left[\frac{a_0^{0.63}}{1 - 756c(\Delta\sigma)^{3.26}Na_0^{0.63}} \right]^{1.59} \quad (9)$$

Substituting eqn (9) into eqn (5), $a_0, \sigma_y, \Delta\sigma, \sigma_m, \delta_c$ and c are treated as the random variables with known distributions and formulae, eqns (5), and (5) are considered as the probabilistic fracture model of the nozzle, then the yearly fracture probabilities are determined by a n -dimension integral ($n = 6$). In order to overcome the difficulties in the integration, the paper adopts the Monte Carlo simulation with compound sampling (stratified sampling with LHS) to estimate P_f of the nozzle.

3.2 Input data

3.2.1 Determined data

$D = 500 \text{ mm}$; $d = 147 \text{ mm}$; $t = T = 14 \text{ mm}$; $K_t (= 2.14)$ is obtained from the Decock formula; wall thickness in the direction of nozzle corner is equal to 20 mm; the upper and lower limits of initial crack depth x_u, x_l are 20 mm and zero, respectively; applied cyclic pressure is within 0~9 MPa; the cyclic stress within a year should be $252 \times m = 3.26$.

3.2.2 Random variables

The probability density functions (p.d.f.) and their distributional parameters of six random variables are listed in Table 1.

4 RESULTS

(1) By carrying out the Monte Carlo simulation 50,000 times with direct sampling, the yearly fracture probabilities(P_f) of the nozzle are obtained and shown in Fig. 1. These P_f values are considered as accurate values to estimate calculative errors of the stratified sampling and the stratified sampling with LHS.

(2) In the stratified sampling, according to the stratifying method in Ref. 6, the initial crack depth size region between zero and 20mm is divided into 10 over-overlapping sub-regions, i.e., (0,0-67), (0-67,1), (1,2), (2,3), (3,6), (6,8), (8,11), (11,14), (14,17), (17,20). The probability of a fracture event which arises from an initial surface crack with depth $a_0 < 1 \text{ mm}$ is very small; The probability of a surface crack with depth $a_0 > 14 \text{ mm}$ which leads to nozzle fracture is large, but its survival probability is very small ($< 1E - 4$). Therefore, the subregions of (0,0-67), (0-67,1), (14,17), (17,20) can be neglected in the sampling. The number of simulation of other subregions are 1000, 1000, 2000, 2000, 1500, 800, respectively. Using the stratified sampling to simulate 8,300 times, the yearly P_f are obtained and shown in Fig. 1. In comparison with results obtained from the direct sampling, the maximum relative

Fig. 1. Estimated fracture probability.

error is 21.7%. The computer time of the stratified method is about 6.6% that of the direct sampling.

(3) The conditional expectation and antithetic variates VRT with LHS are used to calculate the fracture probabilities of the nozzle. The simulation time is 1000. The results are shown in Fig. 1. Compared with the calculative values obtained from the direct sampling simulation 50,000 times, the maximum relative errors are 103.7% and 34.2%, respectively.

(4) The stratified sampling with LHS (the compound sampling). We carried out the stratified sampling as mentioned above, but the P_f arose from an initial crack lying in the 3rd–8th subregions are obtained by using the Monte Carlo simulation with LHS. The number of simulation of the 3rd–8th subregions are 80, 80, 150, 150, 50, 50, respectively. The results are shown in Fig. 1. In comparison with the results obtained from the direct sampling, the maximum relative error of the compound sampling is about 5.61%, and its computer time is about 0.1% that of the directing.

5 CONCLUSIONS

Many methods are currently being used for structural reliability assessment purposes. They include the First-Order Second-Moment method, the Advanced Second-Moment method, and Monte Carlo simulation methods, with and without variance reduction techniques. The main objective

of this paper is to introduce a compound sampling, i.e., the stratified sampling with LHS, to assess structural reliability. The results of the compound sampling are compared to the results of direct sampling, stratified sampling and the conditional expectation and antithetic variates variance reduction technique with LHS. The proposed stratified sampling with LHS is very accurate and efficient.

REFERENCES

1. Mackay, M. D., Beckman, R. J. and Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 1979, **2**, 239–245.
2. Ayyub, B.M. and Lai, Kwan-Ling, Selective sampling in simulation-based reliability assessment. *International Journal of Pressure Vessels and Piping*, 1991, **46**, 229–249.
3. Iman, R. I. and Conover, W. J. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics*, 1982, **B11**, 311–334.
4. Ayyub, B. M. and Lai, Kwan-Ling, Structural reliability assessment using Latin hypercube sampling. In *Proceedings of the International Conference on Structural Safety and Reliability, ICOSSAR89*, San Francisco, USA, 1989, pp. 174–184.
5. Rubinstein, R. Y., *Simulation and the Monte Carlo method*. John Wiley & Sons, New York, 1981.
6. Gao, Z., Fatigue crack growth in the nozzle corner of a pressure vessel. *International Journal of Pressure Vessels and Piping*, 1990, **42**, 1–13.