

## Methods for Uncertainty Analysis for Structural Integrity Problems

R.E. Kurth, P. Baybutt

*Battelle's Columbus Laboratories, 505 King Avenue, Columbus, Ohio 43201, U.S.A.*

### Abstract

The determination of the expected life of a structure requires the treatment of many random variables. Bounds for the expected time to failure due to fatigue crack growth are assessed using three uncertainty analyses techniques: Monte Carlo; Latin Hypercube Sampling; and a new method developed at Battelle called RASCAL. The comparisons made during this study indicate that the RASCAL method can reproduce the 0.1 to 99.9 percentile values of the Monte Carlo analysis within 4% for one-half to one-fifth of the cost. The LHS method does not perform as well as the RASCAL method.

### 1. Introduction

Providing credible bounds for the remaining life of a structure is important in order to answer two important questions: firstly, what is the probability of the structure failing as it ages?; secondly, what are the important variables contributing to the uncertainty in the structural response? There are several methods available for performing these types of analysis, however the most widely accepted ones in the nuclear industry are Monte Carlo simulation, Latin Hypercube Sampling (LHS), and the Discrete Probability Distribution (DPD) method. Each of these methods have been investigated for study in structural reliability uncertainty analysis. The DPD method will not work for many structural analysis problems because of the large amounts of computer storage which is required [1]. To address this problem a new probabilistic technique, denoted RASCAL, has been developed at Battelle which can treat discrete distributions in the same manner as the DPD method but does not have the computer storage limitations of the DPD algorithm. Additionally, the RASCAL method can perform calculations similar to importance sampling with the Monte Carlo method regardless of the distributional family being studied.

This is an important advantage of the RASCAL method over the Monte Carlo technique since the importance sampling in the RASCAL method is defined by input and does not require any coding changes. The Monte Carlo technique does require changes to be made to the code in many instances in order to change the weighting distributions for the importance sampling methods. The LHS method has been used extensively for the uncertainty analysis of large computer programs.

Each of these techniques have been applied to fatigue crack growth analysis which is an important problem in many structural analysis calculations. While there are other important calculations needed for structural integrity problems the crack growth problem has been chosen to illustrate the performance of each of the uncertainty analysis methods. It is important to note that the results of these studies are only applicable to structural calculations whose characteristics are similar to the fatigue crack problem. Because of space limitations the results of other studies will not be discussed in detail, but rather the conclusions drawn from this work will be presented.

## 2. Uncertainty Analysis Methods Description

Monte Carlo. The Monte Carlo technique is a simple method for adding a probabilistic structure to a deterministic model. If each of the individual variables has been characterized by a PDF during data analysis, then the following procedure is used during a Monte Carlo simulation. The CDF of each individual variable is generated by integrating the PDF. A random number between 0 and 1 is generated, call it  $r_1$ . The CDF is inverted and the value for the variable  $L_j(1)$ , is determined. This method of choosing the value is repeated for each individual variable. A value of a variable representing the composite variable is then calculated as

$$L_c(1) = f(L_1(1), L_2(1), \dots, L_n(1))$$

The entire process is repeated a large number of times, designed  $M$ . What results is an  $M$ -dimensional vector of a composite variable  $(L_c(1), L_c(2), \dots, L_c(M))$ . This vector is used to construct a histogram of the composite variable. This histogram can now be analyzed statistically to obtain estimates of the mean, kurtosis, probability of a variable being exceeded and so on. Obviously, in the limit as  $M$  tends to infinity, the continuous distribution will be asymptotically approached. Equally obvious, the computer time will also increase. The value of  $M$  increases dramatically as the probability of the event being simulated becomes smaller. In these low probability calculations importance sampling schemes<sup>[2]</sup> are used. For this study only the basic Monte Carlo algorithm

is used so no further discussion of importance sampling methods is provided.

Latin Hypercube Sampling (LHS) Method. The basic premise of LHS is to insure that all portions of the sample space of the input variables  $X_j$  are sampled. Each  $X_j$  is divided into  $K$  strata of equal marginal probability  $1/K$ , denoted by  $X_{jk}$  where  $k = 1, 2, \dots, K$ . For each stratum, one value of  $X_{jk}$  is obtained by sampling in this interval. LHS then matches the  $X_{jk}$  randomly to use as input to the function being evaluated to obtain values for the response. This method of sampling is an extension of quota sampling. Additionally, it can be viewed as a  $K$ -dimensional extension of Latin square sampling, from which it obtains its name.

Discrete Probability Distributions (DPD's). The description of DPD's follows the conventions set forth by Kaplan [3] and Kurth and Cox [4]. In this method, the initial values of the variables are discretized into  $m$  values. Each value of each variable is then assigned a probability of occurrence. Additionally, the various forms of any probabilistic function are assigned a probability of being correct. If these discrete values are paired with their probabilities, the following vectors of ordered pairs result for two variables  $X$  and  $Y$ :

$$X = [(X_1, p_1), (X_2, p_2), \dots, (X_m, p_m)]$$

$$Y = [(Y_1, q_1), (Y_2, q_2), \dots, (Y_m, q_m)]$$

The number of discrete points in each of these vectors has been chosen to be the same although it is not necessary to do so. The addition of two discrete vectors is defined by

$$Z = Y + X.$$

$$Z = (Y_i, p_i) + (X_j, q_j), \text{ and}$$

$$Z = (X_j + Y_i, p_i * q_j) \text{ for all } i \text{ and } j$$

Therefore, the addition of two vectors containing  $m$  ordered pairs each results in a vector which has  $m^2$  ordered pairs. The multiplication of DPD's is similarly defined:

$$Z = X * Y$$

$$= [(X_j * Y_i, p_i * q_j)] \text{ for all } i \text{ and } j$$

For the combination of a large number of variables, the amount of computer storage increases very quickly. If there are  $k$  variables, each described by  $M$  discrete points, then the vector will contain  $M^k$  ordered pairs. Since, even for relatively small values of  $M$  and  $K$  (on the order of 20), the computer storage capability will quickly be exceeded, it is necessary to have some procedure for reducing this vector's size. This leads to the use of a condensation procedure described in references

[3] and [4]. However not all problems can be handled by a condensation technique and, therefore, the development of a modified DPD scheme resulted which has fixed storage requirements.

Random Sampling Condensation Algorithm. The use of the DPD method for determining the statistical characteristics of a response of interest is appealing because it can easily be put in a data base format and the logic is independent of the distributional form. However, the computer storage requirements can quickly exceed any machine's limits. For example, if a response,  $R$ , is equal to the sum of 10 variables, denoted  $x_i$  for  $i = 1, 2, \dots, 10$ , then the storage requirements are equal to  $2 \times N_D^{10}$ , where  $N_D$  is the number of discrete points in each interval. (There is a factor of 2 because both the value and the probability of the combination of inputs occurring must be stored.) To make the example more concrete, assume that  $R$ , the response, is simply equal to the sum of all the  $x_i$ 's which have been discretized into intervals of 20 each. In this case, the number of discrete values which will be generated to define  $R$  is equal to  $20^{10}$ . From a statistical standpoint, most of these  $20^{10}$  points are not meaningful if the PDF for  $R$  is being approximated. The situation is analogous to the problem of determining how many black balls are in a jar of black and white balls. If a jar could be found which is large enough to hold  $20^{10}$  balls which contains an equal number of black and white balls, one solution is to tip over the jar and count all of the white balls. This is what the DPD method does in an analogous fashion. If all of the  $20^{10}$  values for  $R$  are available, is there not some way to sample from the space of  $R$  to estimate the PDF? The answer is, of course, yes; and the RASCAL method has been developed to perform this sampling.

### 3. Uncertainty Analysis Comparison For Fatigue Crack Growth

The calculation of the mean crack size and its associated standard deviation, defined in the classical sense, was performed for a fatigue crack growth problem where the crack size at time  $t$ , denoted  $a_t$ , is related to the initial crack size,  $a_0$ , the applied stress,  $S$ , and a random parametric value,  $c$ , by the Paris' law equation. Table 1 presents the distributional families and their parameters used for each of the input variables. Since an analytic solution of this problem is not available a 2000 trial Monte Carlo simulation is taken to represent "truth", i.e. the correct final value of the crack size. The results of the various analyses are given in Table 2. For these analyses the uncertainty measure is the standard deviation. In Table 2 the runs identified as LHS followed by a number indicate that

the LHS method was used with the given number of strata, e.g. LHS-20 represents the LHS analysis with twenty strata. A similar notation is used for the RASCAL analyses. Also the number in parentheses below the identification of the method type provides the computer time used for each of the analyses<sup>1</sup>.

As Table 2 indicates the RASCAL method provides excellent agreement in the estimate of the mean values of the crack size and good agreement with the standard deviation. Additionally, the computational time of the RASCAL method is a factor of 2 to 6 faster than the Monte Carlo. The LHS analysis provides reasonable agreement in the mean values but poor agreement in the estimate of the standard deviation.

#### 4.0 Summary and Conclusions

The results of one study, important to structural analysis has been presented. In this work the effect of fatigue crack growth on the structure's reliability is considered. The random nature of the initial crack size, applied load, and functional relationship were all included in the analysis. It was found that the RASCAL method of analysis was the most appropriate method for simulating the growth of cracks when the simple Paris law type of empirical correlation is used. The Monte Carlo analysis is much more expensive to use than RASCAL without a cotresponding increase in the accuracy of the results. The LHS sampling method performed very poorly.

Other studies, not presented here because of space limitations, have indicated that, if significantly longer running computer codes are being used, e.g. finite element stress analysis codes, are being used in the analysis then the LHS method becomes more attractive if only basic uncertainty information is desired. However, in this case, replicate LHS design methods should be used. If the code is small enough so that many thousands of Monte Carlo trials can be used then the Monte Carlo analysis should be adopted since classic estimates of the confidence limits on the various parameters can be estimated. It is not clear that either the RASCAL or LHS method can provide such estimates without further assumptions about the output being made which may not be able to be justified by the physics of the problem.

#### ACKNOWLEDGMENT

This work was supported by Battelle Corporate Technical Development, contract B1333-1430.

<sup>1</sup>The Monte Carlo analysis used a bubble sorting method which dramatically reduces the computational time.

**Table 1. Random Parameter Definition For Crack Growth Analysis**

<u>Variable Definition</u>	<u>Parameter 1</u>	<u>Parameter 2</u>
Initial crack size	0.005	0.01
Paris' parameter	$1.0 \times 10^{-14}$	$9.11 \times 10^{-12}$
Applied load	10.0	35.0

All variables are Rayleigh distributed:

$$f(x) = [(x-p_1)/p_2] \cdot \exp[-(x-p_1)^2/(2 \cdot p_2)]$$

**Table 2. Results of Uncertainty Analysis of Crack Growth Model**

<u>Method</u>	<u>Time (cycles)</u>					
	<u>2000</u>		<u>6000</u>		<u>10000</u>	
	<u>mean</u>	<u>std dev<sup>2</sup></u>	<u>mean</u>	<u>std dev</u>	<u>mean</u>	<u>std dev</u>
LHS-20 ( 2.7)	.01215	.00427	.01386	.00856	.01788	.02464
LHS-50 (13.0)	.01168	.00344	.02176	.00684	.03162	.00957
RASCAL-20 (13.8)	.01149	.00345	.01254	.01632	.01545	.03592
RASCAL-50 (36.9)	.01149	.00337	.01224	.01317	.01522	.03395
MONTE CARLO (52.9)	.01161	.00777	.01276	.01891	.01584	.03973

**References**

- [1] Martz, H.F., et. al., "A Comparison of Methods for Uncertainty Analysis of Nuclear Power Plant Safety System Fault Tree Models", NUREG/CR-3263, April 1983
- [2] Halton, J.H., "A Retrospective and Prospective Survey Of The Monte Carlo Method", SIAM Review, 12, 1, 1961
- [3] Kaplan, S., "On The Method of Discrete Probability Distributions In Risk And Reliability Calculations - Application To Seismic Risk Assessment", Risk Analysis, 1, No.3, 1981
- [4] Kurth, R.E. and Cox, D.C., "Discrete Probability Distributions For Probabilistic Fracture Mechanics", to be published in Risk Analysis, 1985

<sup>2</sup>Standard deviation