CALCULATION OF FIRST EXCURSION PROBABILITIES BY SUBSET SIMULATION

S.K. Au and J.L. Beck, M. ASCE
California Institute of Technology, Pasadena, CA 91125
siu@its.caltech.edu, *jimbeck@its.caltech.edu*

Abstract

A new simulation approach, called 'subset simulation', is applied to computing small first excursion probabilities for dynamical systems with stochastic excitations. The basic idea is to express the first excursion probability as a product of larger conditional failure probabilities by introducing intermediate failure boundaries. With a proper choice of the intermediate boundaries, the original problem of calculating a small failure probability, which is computationally demanding, is reduced to calculating a sequence of conditional probabilities, which are efficiently estimated by simulation using a special Markov chain. The proposed method is robust to the type of structural model (e.g., linear or nonlinear) and stochastic excitation model (e.g., stationary or nonstationary). Numerical studies are presented to demonstrate the efficiency of the method.

Introduction

For a system subjected to stochastic excitation, the failure probability that the response $x(t)$ will out-cross a threshold level b within the time interval $[0, T]$ is given by

$$P_F = P\{\exists t \in [0,T] : |x(t)| > b\} = \int I_F(\mathbf{z}) q(\mathbf{z}) d\mathbf{z} \tag{1}$$

where I_F is an indicator function: $I_F(\mathbf{z}) = 1$ if $\mathbf{z} \in F$ and $I_F(\mathbf{z}) = 0$ otherwise. Here, F denotes the failure region within the space of $\mathbf{z} = [z_1, \ldots, z_n]$, a vector consisting of values of the random variables which generate the stochastic excitation. For example, if the excitation is a linearly filtered white-noise process discretized at a number of time points, then $\mathbf{z}$ will be the discrete white-noise sequence. In (1), q is an n-dimensional probability density function (PDF) for $\mathbf{z}$. In practical applications, stochastic processes are often generated by some transformation of independent and identically distributed (i.i.d.) random variables. Thus, without much loss of generality, we will assume that $\mathbf{z}$ has i.i.d. components, that is, $q(\mathbf{z}) = \prod_{j=1}^{n} q_j(z_j)$.

Although P_F is written as an n-fold integral in (1), it is seldom evaluated by numerical integration, since the dimension n is too large and the failure region often has complicated geometry. Simulation methods offer a feasible means to compute P_F (Rubinstein 1981). Monte Carlo simulation (MCS) is well known to be robust to the type and dimension of the problem. Its main drawback, however, is that it is not suitable for finding small probabilities

(e.g., $P_F \leq 10^{-3}$), because the number of samples, and hence the number of structural analyses, required to achieve a given accuracy is proportional to $1/P_F$. Essentially, finding small probabilities requires information from rare samples corresponding to failure, and on average it would require many samples before one such sample occurs.

A new simulation approach, called subset simulation, is presented in this paper which bypasses the need to simulate rare samples for estimating small probabilities. By introducing intermediate failure boundaries, the failure probability is expressed as a product of conditional failure probabilities, the evaluation of which only requires simulation of more frequent failure events. Although the method is explained using the crossing problem of a scalar response process, it can be applied in general to vector crossing problems, as demonstrated in the numerical applications.

Basic Idea of Subset Simulation

Let $0 < b_1 < b_2 < \ldots b_m = b$ be an increasing sequence of threshold levels. For each b_i, define the failure event $F_i = \{\exists t \in [0,T] : |x(t)| > b_i\}$. Observe that, if the process $x(t)$ has crossed the level b_i ($i \geq 2$), it must have crossed the lower levels $b_{i-1}, \ldots, b_1$. Thus, $F_k = \cap_{i=1}^k F_i$, $k = 1, \ldots, m$. By definition of conditional probability, we have

$$\begin{aligned} P_F &= P(F_m) = P(\cap_{i=1}^m F_i) = P(F_m | \cap_{i=1}^{m-1} F_i) P(\cap_{i=1}^{m-1} F_i) = P(F_m|F_{m-1}) P(\cap_{i=1}^{m-1} F_i) \\ &= \ldots = P(F_1) \textstyle\prod_{i=1}^{m-1} P(F_{i+1}|F_i) \end{aligned} \tag{2}$$

Equation (2) expresses the failure probability as a product of a sequence of conditional probabilities $\{P(F_{i+1}|F_i)\}$ and $P(F_1)$. The idea of subset simulation is to estimate the failure probability P_F by estimating these quantities. Observe that, although P_F is small, by choosing the intermediate threshold levels appropriately, the conditional probabilities involved in (2) can be made sufficiently large so that they can be evaluated efficiently by simulation procedures. For example, suppose $P(F_1), P(F_{i+1}|F_i) \sim 0.1$, $i = 1, \ldots, 5$, then $P_F \sim 10^{-6}$ which is too small for efficient estimation by MCS. However, the conditional probabilities, which are of the order of 0.1, may be evaluated efficiently by simulation because the failure events are more frequent. The problem of simulating rare events in the original probability space is thus replaced by a sequence of simulations of more frequent events in the conditional probability spaces.

To compute P_F based on (2), one needs to compute the probabilities $P(F_1)$, $\{P(F_{i+1}|F_i)\}$. $P(F_1)$ can be readily estimated by MCS:

$$P(F_1) \approx \tilde{P}_1 = \frac{1}{N} \sum_{k=1}^{N} I_{F_1}(\mathbf{z}_k) \tag{3}$$

where $\{\mathbf{z}_k\}$ are i.i.d. samples simulated according to q. It is natural to compute the conditional failure probabilities based on an estimator similar to (3), which necessitates the simulation of samples according to the conditional distribution of the load vector $\mathbf{z}$ given that the corresponding response has exceeded the level b_i, i.e., $q(\mathbf{z}|F_i) = q(\mathbf{z}) I_{F_i}(\mathbf{z}) / P(F_i)$. Although one can obtain such samples as those simulated from q which lie in the failure

region F_i, it is not efficient to do so since on average it takes $1/P(F_i)$ samples before one such sample occurs. In general, the task of efficiently simulating conditional samples is not trivial. Fortunately, this can be achieved by an advanced Markov chain simulation technique, called the Metropolis method, which will be presented in the next section.

Markov Chain Simulation

The Metropolis method (Metropolis et al. 1953) is a powerful technique for simulating samples according to an arbitrary probability distribution. In this method, samples are simulated as the states of a Markov chain which has limiting (stationary) distribution equal to the target distribution. It has been recently applied to adaptive importance sampling for reliability analysis in Au and Beck (1999). The significance of the method to the current application is that we can efficiently simulate samples having the conditional distribution $q(\cdot|F_i)$. These Markov chain samples, which are dependent in general, can be used for statistical averaging as if they were i.i.d., although with some reduction in efficiency.

We first describe the method to simulate from the current sample the next Markov chain sample so that the limiting distribution of the samples is equal to $q(\mathbf{z}|F_i)$. The algorithm presented here is a modified version of the original Metropolis algorithm since it has been found that the latter cannot be applied to simulating random vectors with independent components when the dimension n is large. Let the current sample be $\mathbf{z}_k = [\mathbf{z}_k(1), \ldots, \mathbf{z}_k(n)]$. For every component $j = 1, \ldots, n$, simulate ξ uniformly on the interval $[\mathbf{z}_k(j) - l_j, \mathbf{z}_k(j) + l_j]$ for some chosen $l_j > 0$ which controls the maximum distance that the next sample can deviate from the current one. Compute the ratio $r = q_j(\xi)/q_j(\mathbf{z}_k(j))$. Set $\tilde{\mathbf{z}}(j) = \xi$ with probability $\min\{1, r\}$ and set $\tilde{\mathbf{z}}(j) = \mathbf{z}_k(j)$ with the remaining probability $1 - \min\{1, r\}$. After this process has been done for all j, check the location of the vector $\tilde{\mathbf{z}}$. If $\tilde{\mathbf{z}} \in F_i$, accept it as the next sample of the Markov chain, i.e., $\mathbf{z}_{k+1} = \tilde{\mathbf{z}}$; otherwise reject it and take the current sample as the next sample, i.e., $\mathbf{z}_{k+1} = \mathbf{z}_k$.

It is not difficult to verify that if the current sample is distributed as $q(\cdot|F_i)$, so is the next sample. In fact, writing out the transition PDF $p(\mathbf{z}_{k+1}|\mathbf{z}_k)$ of the Markov chain, one can prove the 'reversibility condition': $p(\mathbf{z}_{k+1}|\mathbf{z}_k)q(\mathbf{z}_k|F_i) = p(\mathbf{z}_k|\mathbf{z}_{k+1})q(\mathbf{z}_{k+1}|F_i)$. Thus, if the current sample $\mathbf{z}_k$ is distributed as $q(\cdot|F_i)$, then $p(\mathbf{z}_{k+1}) = \int p(\mathbf{z}_{k+1}|\mathbf{z}_k)q(\mathbf{z}_k|F_i)d\mathbf{z}_k = \int p(\mathbf{z}_k|\mathbf{z}_{k+1})q(\mathbf{z}_{k+1}|F_i)d\mathbf{z}_k = q(\mathbf{z}_{k+1}|F_i)\int p(\mathbf{z}_k|\mathbf{z}_{k+1})d\mathbf{z}_k = q(\mathbf{z}_{k+1}|F_i)$ since $\int p(\mathbf{z}_k|\mathbf{z}_{k+1})d\mathbf{z}_k = 1$. That is, the next Markov chain sample $\mathbf{z}_{k+1}$ will also be distributed as $q(\cdot|F_i)$, and the latter is indeed the stationary distribution for the generated Markov chain.

Utilizing this modified Metropolis method, the conditional probabilities can be computed as follows. Assume that we have already simulated samples according to q which are used for estimating $P(F_1)$ based on (3). From these samples, we can readily obtain some samples distributed as $q(\cdot|F_1)$, simply as those which have the corresponding response exceeding the first level b_1, i.e., those with $I_{F_1} = 1$. Starting from each of these samples, we can simulate Markov chain samples using the modified Metropolis method. These samples will also be distributed as $q(\cdot|F_1)$. They can be used to estimate $P(F_2|F_1)$ using an estimator similar to (3). Observe that the Markov chain samples which have the response exceeding the next level b_2 are distributed as $q(\cdot|F_2)$ and thus they provide 'seeds' for simulating

more samples according to $q(\cdot|F_2)$ to estimate $P(F_3|F_2)$. Repeating this process, we can compute the conditional probabilities of the higher levels until the level of interest b has been reached. In general, let $\{\mathbf{z}_k^{(i)} : k = 1, \ldots, N\}$ be the Markov chain samples with distribution $q(\cdot|F_i)$, $1 \leq i \leq m-1$, possibly coming from different chains. Then

$$P(F_{i+1}|F_i) \approx \tilde{P}_{i+1} = \frac{1}{N}\sum_{k=1}^{N} I_{F_{i+1}}(\mathbf{z}_k^{(i)}) \tag{4}$$

Note that the samples also yield conditional failure probability estimates for all threshold levels between b_i and b_{i+1}. Finally, combining (2), (3) and (4), $P_F \approx \tilde{P}_F = \prod_{i=1}^{m} \tilde{P}_i$.

Statistical Properties of Estimators

Conditional Estimator $\tilde{P}_i$ ($i \geq 2$): Although the Markov chain samples used for estimating the conditional probabilities based on (4) are in general dependent, all the estimators $\tilde{P}_i$, $i \geq 2$, still have the usual convergence properties of estimators using independent samples (Doob 1953). For example, $\tilde{P}_i$ converges almost surely (a.s.) to $P(F_i|F_{i-1})$ (strong law of large numbers), is unbiased, consistent, and Normally distributed as $N \to \infty$ (Central Limit Theorem). The coefficient of variation (c.o.v.) of $\tilde{P}_i$, δ_i, is given by:

$$\delta_i^2 = var[\tilde{P}_i]/P(F_i|F_{i-1})^2 = (1+\gamma_i)[1 - P(F_i|F_{i-1})]/P(F_i|F_{i-1})N \tag{5}$$

where γ_i is a correlation factor which can be estimated from simulation results. The term $[1 - P(F_i|F_{i-1})]/P(F_i|F_{i-1})N$ in (5) is the familiar term for the square of c.o.v. in MCS with N independent samples. The variance of $\tilde{P}_i$ can thus be considered as the variance in MCS with an effective number of independent samples $N/(1+\gamma_i)$. The value of γ_i depends on the choice of l_i. The efficiency of the estimator using dependent samples of a Markov chain (with $\gamma_i > 0$) is reduced compared to the case when the samples are independent ($\gamma_i = 0$), and smaller values of γ_i imply higher efficiency.

Failure Probability Estimator $\tilde{P}_F$: Since $\tilde{P}_1 \to P(F_1)$ and $\tilde{P}_i \to P(F_i|F_{i-1})$ ($i \geq 2$) a.s. as $N \to \infty$, $\tilde{P}_F \to P(F_1)\prod_{i=1}^{m-1} P(F_{i+1}|F_i) = P_F$ a.s. also. Due to the correlation between the conditional estimators $\{\tilde{P}_i\}$, it is biased for every N, but is asymptotically unbiased. The correlation is due to the fact that the samples used for computing $\tilde{P}_i$ which lie in F_i are used to start the Markov chains to compute $\tilde{P}_{i+1}$. It can be shown that the fractional bias, defined as $E|\tilde{P}_F - P_F|/P_F$, is $O(1/N)$. On the other hand, to the leading order, the c.o.v. of $\tilde{P}_F$, δ, is bounded above as follows:

$$\delta^2 = \frac{E|\tilde{P}_F - P_F|^2}{P_F^2} \leq \sum_{i,j=1}^{m} \delta_i \delta_j \tag{6}$$

where the upper bound corresponds to the case where the $\tilde{P}_i$s are fully correlated. From (5), $\delta_i^2 \sim O(1/N)$, and so (6) implies $\delta^2 \sim O(1/N)$, i.e., $\tilde{P}_F$ is a consistent estimator. The actual c.o.v. depends on the correlation between the $\tilde{P}_i$s at different intermediate levels. If all the $\tilde{P}_i$s were uncorrelated, then $\delta^2 = \sum_{i=1}^{m} \delta_i^2$. Although the $\tilde{P}_i$s are generally correlated, simulations show that δ^2 may be well approximated by $\sum_{i=1}^{m} \delta_i^2$. By estimating γ_i from

simulated samples and replacing $P(F_i|F_{i-1})$ with $\tilde{P}_i$ in (5), δ_i and hence an upper bound for δ based on (6) can be estimated in a simulation run.

To get an idea of the number of samples required to achieve a given accuracy in $\tilde{P}_F$, suppose $P(F_1) = P(F_{i+1}|F_i) = p_0$ and $\delta_i^2 = (1+\gamma)(1-p_0)/p_0 N$. Using (5) and (6), and noting that $m = \log P_F / \log p_0$, we conclude that to achieve a given c.o.v. of δ in the estimate $\tilde{P}_F$, the total number of samples required is roughly $N_T = mN = |\log P_F|^r \times (1+\gamma)(1-p_0)/p_0 |\log p_0|^r \delta^2$, where $r \leq 3$, depending on the actual correlation of the $\tilde{P}_i$s. Thus, for fixed p_0 and δ, $N_T \sim |\log P_F|^r$. Compared with MCS, where $N_T \sim 1/P_F$, this implies a substantial improvement in efficiency when estimating small probabilities.

Numerical Applications

Consider a five-story shear building with hysteretic behavior under earthquake motion modeled by a nonstationary stochastic process. The floor masses are 45.4×10^3 kg for all stories. The linear interstory stiffness for the first to fifth stories are 41.1×10^6, 38.5×10^6, 33.4×10^6, 25.6×10^6 and 15.2×10^6 N/m, respectively. Each story, of height 2.8 m, is modeled by a Bouc-Wen hysteretic element (Wen, 1976) with parameters $\alpha = \beta = 0.5$, $\eta = 1$ and has strength equal to 12×10^{-3} of the corresponding interstory stiffness. The small-amplitude natural frequency of the structure is 1.25 Hz. Rayleigh viscous damping is also assumed so that the first two modes have 5% of critical damping at small response amplitudes. The structure is subjected to base excitation modeled by Clough-Penzien filtered white noise modulated by an envelope function. The dominant and corner frequency of the Clough-Penzien spectrum are 2.5 Hz and 0.25 Hz, respectively, with corresponding damping parameters equal to 60% and 80%. The strength of the white noise is 2.5×10^{-3} m^2/s^3. The envelope function varies as $(t/4)^2$ for the first 4 sec., is equal to unity from 4 sec. to 24 sec., and finally decays as $\exp[-0.5(t-24)]$. The nonlinear response of the structure is computed using the Newmark constant acceleration scheme with a time spacing of 0.02 sec. and for a duration of 30 sec., which is sufficient to capture the whole earthquake response history. P-Δ effects have been taken into account in computing the nonlinear response. The total number of random variables involved in the problem is n=$30/0.02$+1=$1,501$. Failure is defined as the exceedence of the interstory drift of any one of the stories above a given threshold level b within the first 30 sec. $N = 500$ samples have been used to estimate each of the conditional probabilities. The parameter l_j is chosen to be 1 for all j.

Figure 1 shows 5 sets of simulation results for the failure probability. For comparison, the results using MCS with $100,000$ samples are also shown. The intermediate threshold levels b_i are chosen to be $1.5, 1.8, 2.1, 2.4, 2.7$ and 3.0 cm, whose failure probability estimates are shown with circles in the figure using a total number of samples equal to $500, 1000, 1500, 2000, 2500$ and 3000, respectively. The results by subset simulation and MCS agree well. To investigate the actual variability of the estimates and the validity of the upper bound on the c.o.v. in (6), the c.o.v. of the estimate over 100 sets of subset simulation results is plotted in Figure 2. The dash line shows an estimate of the upper bound computed based on (5) and (6) averaged over the 100 simulation runs. The actual c.o.v. lies close to the estimate assuming the $\tilde{P}_i$s are uncorrelated, showing that there is only a small reduction of efficiency due to the correlation between conditional probability estimates for different

levels. To compare the efficiency of subset simulation with MCS, the c.o.v. of the estimate at a particular probability level using the same number of samples as in subset simulation is plotted with squares with Figure 2. Note that the c.o.v. for MCS grows much faster with decreasing failure probability than for the subset simulation. Thus, subset simulation becomes more efficient compared with MCS as the failure probability gets smaller.

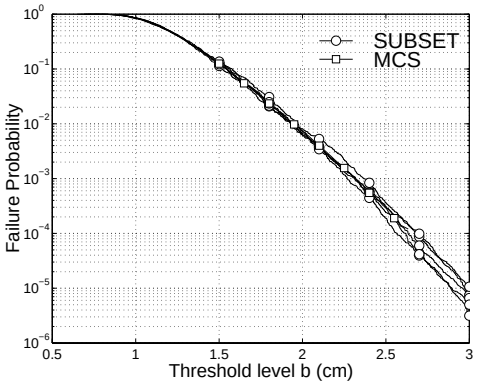

Figure 1. Failure probability estimate.

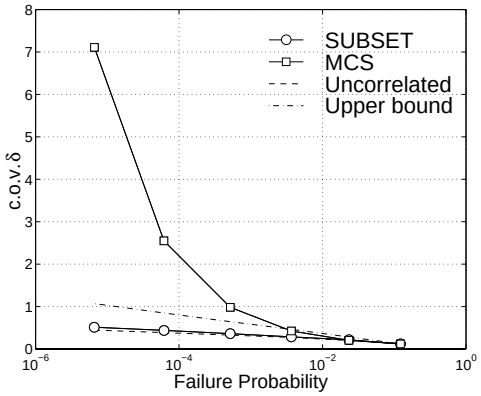

Figure 2. Coefficient of variation.

Conclusions

The need to simulate rare events is one of the major obstacles in applying simulation methods to estimating small failure probabilities. Subset simulation resolves this by breaking the problem into a sequence of problems of estimating larger conditional probabilities. The Metropolis method has been modified to efficiently compute the conditional probabilities. Theoretical estimates for the c.o.v. and results from numerical simulation demonstrate a substantial improvement in efficiency over Monte Carlo simulation.

Acknowledgements

This paper is based upon work partly supported by the Pacific Earthquake Engineering Research Center under National Science Foundation Cooperative Agreement No. CMS-9701568. This support is gratefully acknowledged.

References

Au, S.K. and Beck, J.L. (1999), "A new adaptive importance sampling scheme," *Structural Safety*, **21**, 135–158.

Doob, J.L. (1953), *Stochastic Processes*, John Wiley and Sons, New York.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, and A.H. Teller (1953), "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, **21**(6), 1087–1092.

Rubinstein, R.Y. (1981), *Simulation and the Monte-Carlo Method*, John Wiley and Sons, New York.

Wen, Y.K. (1976), "Method for random vibration of hysteretic systems," *Journal of Engineering Mechanics*, ASCE, **102**, 249–263.