Estimation of small failure probabilities in high dimensions by subset simulation

Siu-Kui Au, James L. Beck*

Division of Engineering and Applied Science, California Institute of Technology, Mail Code 104-44, Pasadena, CA 91125, USA

Abstract

A new simulation approach, called 'subset simulation', is proposed to compute small failure probabilities encountered in reliability analysis of engineering systems. The basic idea is to express the failure probability as a product of larger conditional failure probabilities by introducing intermediate failure events. With a proper choice of the conditional events, the conditional failure probabilities can be made sufficiently large so that they can be estimated by means of simulation with a small number of samples. The original problem of calculating a small failure probability, which is computationally demanding, is reduced to calculating a sequence of conditional probabilities, which can be readily and efficiently estimated by means of simulation. The conditional probabilities cannot be estimated efficiently by a standard Monte Carlo procedure, however, and so a Markov chain Monte Carlo simulation (MCS) technique based on the Metropolis algorithm is presented for their estimation. The proposed method is robust to the number of uncertain parameters and efficient in computing small probabilities. The efficiency of the method is demonstrated by calculating the first-excursion probabilities for a linear oscillator subjected to white noise excitation and for a five-story nonlinear hysteretic shear building under uncertain seismic excitation. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Markov chain Monte Carlo method; Monte Carlo simulation; Reliability; First excursion probability; First passage problem; Metropolis algorithm

1. Introduction

The determination of the reliability of a system or component usually involves calculating its complement, the probability of failure:

$$P_F = P(\boldsymbol{\theta} \in F) = \int I_F(\boldsymbol{\theta}) q(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \qquad (1)$$

where $\boldsymbol{\theta} = [\theta_1, ..., \theta_n] \in \Theta \subset \mathbb{R}^n$ represents an uncertain state of the system with probability density function (PDF) q; F is the failure region within the parameter space Θ; I_F is an indicator function: $I_F(\boldsymbol{\theta}) = 1$ if $\boldsymbol{\theta} \in F$ and $I_F(\boldsymbol{\theta}) = 0$ otherwise. In practical applications, dependent random variables may often be generated by some transformation of independent random variables, and so it is assumed without much loss of generality that the components of $\boldsymbol{\theta}$ are independent, that is, $q(\boldsymbol{\theta}) = \prod_{j=1}^{n} q_j(\theta_j)$, where for every j, q_j is a one-dimensional PDF for θ_j.

Although P_F is written as an n-fold integral over the parameter space Θ in Eq. (1), it cannot be efficiently evaluated by means of direct numerical integration if the dimension n is not small or the failure region has complicated geometry.

Both of these conditions are likely to be encountered in real applications. Simulation methods then offer a feasible means to compute P_F. Monte Carlo simulation (MCS) [1,2] 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 system 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 a failure occurs.

Importance sampling techniques [1,3–5] have been developed over the past few decades to shift the underlying distribution towards the failure region so as to gain information from rare events more efficiently. The success of the method relies on a prudent choice of the importance sampling density (ISD), which undoubtedly requires knowledge of the system in the failure region. When the dimension n of the uncertain parameter space is not too large and the failure region F is relatively simple to describe, many schemes for constructing the ISD, such as those based on design point(s) (e.g. [4,6,7–10]) or adaptive pre-samples [11–14], are found to be useful. The design point(s) or pre-samples are often obtained numerically by optimization

* Corresponding author. Tel.: +1-626-395-4132; fax: +1-626-568-2719.
E-mail address: jimbeck@caltech.edu (J.L. Beck).

or simulation where the integrand function $I_F(\boldsymbol{\theta})q(\boldsymbol{\theta})$ is directly used.

When the dimension n is large and the complexity of the problem increases, however, it may be difficult to gain sufficient knowledge to construct a good ISD [15]. Therefore, the development of efficient simulation methods for computing small failure probabilities in high-dimensional parameter spaces remains a challenging problem. Our objective has been to develop successful simulation methods that can adaptively generate samples, which explore the failure region efficiently, while at the same time retain the robustness of MCS to the dimension of the uncertain parameters and the complexity of the failure region.

A new simulation approach, called subset simulation, is presented to compute failure probabilities. It is robust to dimension size and efficient in computing small probabilities. In this approach, the failure probability is expressed as a product of conditional probabilities of some chosen intermediate failure events, the evaluation of which only requires simulation of more frequent events. The problem of evaluating a small failure probability in the original probability space is thus replaced by a sequence of simulations of more frequent events in the conditional probability spaces. The conditional probabilities, however, cannot be evaluated efficiently by common techniques, and therefore a Markov chain MCS method based on the Metropolis algorithm [16] is used.

2. Basic idea of subset simulation

For convenience, we will use F to denote the failure event as well as its corresponding failure region in the uncertain parameter space. Given a failure event F, let $F_1 \supset F_2 \supset \cdots \supset F_m = F$ be a decreasing sequence of failure events so that $F_k = \bigcap_{i=1}^{k} F_i$, $k = 1, \dots, m$. For example, if failure of a system is defined as the exceedence of an uncertain demand D over a given capacity C, that is, $F = \{D > C\}$, then a sequence of decreasing failure events can simply be defined as $F_i = \{D > C_i\}$, where $C_1 < C_2 < \dots < C_m = C$. By definition of conditional probability, we have

$$P_F = P(F_m) = P\left(\bigcap_{i=1}^{m} F_i\right) \tag{2}$$

$$= P\left(F_m \Big| \bigcap_{i=1}^{m-1} F_i\right) P\left(\bigcap_{i=1}^{m-1} F_i\right)$$

$$= P(F_m|F_{m-1}) P\left(\bigcap_{i=1}^{m-1} F_i\right) = \dots$$

$$= P(F_i) \prod_{i=1}^{m-1} P(F_{i+1}|F_i)$$

Eq. (2) expresses the failure probability as a product of a sequence of conditional probabilities $\{P(F_{i+1}|F_i) : i = 1, \dots, m-1\}$ and $P(F_1)$. The idea of subset simulation is to estimate the failure probability P_F by estimating these quantities. When the probability of failure is estimated by means of simulation, the difficulty often increases with decreasing failure probability. Basically, the smaller the P_F, the more rare the failure event is, and the more the number of samples required to realize failure events for computing P_F. Observe that, although P_F is small, by choosing the intermediate failure events $\{F_i : i = 1, \dots, m-1\}$ appropriately, the conditional probabilities involved in Eq. (2) can be made sufficiently large so that they can be evaluated efficiently by simulation procedures. For example, suppose $P(F_1), P(F_{i+1}|F_i) \sim 0.1$, $i = 1, 2, 3$, then $P_F \sim 10^{-4}$ 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 Eq. (2), one needs to compute the probabilities $P(F_1), \{P(F_{i+1}|F_i) : i = 1, \dots, m-1\}$. $P(F_1)$ can be readily estimated by MCS:

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

where $\{\boldsymbol{\theta}_k : k = 1, \dots, N\}$ are independent and identically distributed (i.i.d.) samples simulated according to PDF q. It is natural to compute the conditional failure probabilities based on an estimator similar to Eq. (3), which necessitates the simulation of samples according to the conditional distribution of $\boldsymbol{\theta}$ given that it lies in F_i, that is, $q(\boldsymbol{\theta}|F_i) = q(\boldsymbol{\theta})I_{F_i}(\boldsymbol{\theta})/P(F_i)$. Although one can follow a 'direct' MCS approach and 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 a Markov chain MCS method based on the Metropolis algorithm that is presented in the Section 3.

3. Markov chain MCS

Markov chain MCS, in particular, the Metropolis method [2,16], 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, under the assumption of ergodicity, has the target distribution as its limiting stationary distribution. It has been recently applied to adaptive importance sampling for reliability analysis to construct an asymptotically optimal importance sampling density [14]. The significance of the Metropolis algorithm to the current application is that if

we can simulate a sample having the conditional distribution $q(\cdot|F_i)$, then we can use the method to simulate as the next state of the Markov chain a new sample which will also be distributed as $q(\cdot|F_i)$. (Even if the current sample is not distributed as $q(\cdot|F_i)$, the limiting distribution property guarantees that the distribution of simulated samples will tend to $q(\cdot|F_i)$ as the number of Markov steps increases.) 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 stationary distribution of the Markov chain is equal to $q(\boldsymbol{\theta}|F_i) = q(\boldsymbol{\theta})I_{F_i}(\boldsymbol{\theta})/P(F_i) = [\prod_{j=1}^{n} q_j(\boldsymbol{\theta}_j)]I_{F_i}(\boldsymbol{\theta})/P(F_i)$, where $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n]$. The algorithm presented here is a new modified version of the original Metropolis algorithm since it has been found that the latter has difficulties when simulating random vectors with many independent components. The reader is referred to the Appendix for the description of the original Metropolis algorithm and the problems encountered in high dimensional simulation.

3.1. Modified Metropolis algorithm

For every $j = 1, ..., n$, let $p_j^*(\xi|\theta)$, called the 'proposal PDF', be a one-dimensional PDF for ξ centered at θ with the symmetry property $p_j^*(\xi|\theta) = p_j^*(\theta|\xi)$. Generate a sequence of samples $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ...\}$ from a given sample $\boldsymbol{\theta}_1$ by computing $\boldsymbol{\theta}_{k+1}$ from $\boldsymbol{\theta}_k = [\boldsymbol{\theta}_k(1), ..., \boldsymbol{\theta}_k(n)]$, $k = 1, 2, ...$, as follows:

1. *Generate a 'candidate' state* $\tilde{\boldsymbol{\theta}}$: For each component $j = 1, ..., n$, simulate ξ_j from $p_j^*(\cdot|\boldsymbol{\theta}_k(j))$. Compute the ratio $r_j = q_j(\xi_j)/q_j(\boldsymbol{\theta}_k(j))$. Set $\tilde{\boldsymbol{\theta}}(j) = \xi_j$ with probability $\min\{1, r_j\}$ and set $\tilde{\boldsymbol{\theta}}(j) = \boldsymbol{\theta}_k(j)$ with the remaining probability $1 - \min\{1, r_j\}$.
2. *Accept/reject* $\tilde{\boldsymbol{\theta}}$: Check the location of $\tilde{\boldsymbol{\theta}}$. If $\tilde{\boldsymbol{\theta}} \in F_i$, accept it as the next sample, i.e. $\boldsymbol{\theta}_{k+1} = \tilde{\boldsymbol{\theta}}$; otherwise reject it and take the current sample as the next sample, i.e. $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$.

In brief, we first generate a candidate state $\tilde{\boldsymbol{\theta}}$ (dependent on the current state) and then take either $\tilde{\boldsymbol{\theta}}$ or $\boldsymbol{\theta}_k$ as the next sample $\boldsymbol{\theta}_{k+1}$ according to whether the candidate state lies in F_i or not.

We now show that the next sample will be distributed as $q(\cdot|F_i)$ if the current sample is, and hence $q(\cdot|F_i)$ is the stationary distribution of the Markov chain. Since all the Markov chain samples lie in F_i, it is sufficient to consider the transition between the states in F_i, which is governed by Step 1. According to Step 1, the transition of the individual components of $\boldsymbol{\theta}_k$ are independent, so the transition PDF of the Markov chain between any two states in F_i can be

expressed as a product of the component transition PDFs:

$$p(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k) = \prod_{j=1}^{n} p_j(\boldsymbol{\theta}_{k+1}(j)|\boldsymbol{\theta}_k(j)) \tag{4}$$

where p_j is the transition PDF for the jth component of $\boldsymbol{\theta}_k$. For $\boldsymbol{\theta}_{k+1}(j) \neq \boldsymbol{\theta}_k(j)$,

$$p_j(\boldsymbol{\theta}_{k+1}(j)|\boldsymbol{\theta}_k(j)) = p_j^*(\boldsymbol{\theta}_{k+1}(j)|\boldsymbol{\theta}_k(j))\min\left\{1, \frac{q_j(\boldsymbol{\theta}_{k+1}(j))}{q_j(\boldsymbol{\theta}_k(j))}\right\} \tag{5}$$

Using Eq. (5), together with the symmetry property of p_j^* and the identity $\min\{1, a/b\}b = \min\{1, b/a\}a$ for any positive numbers a,b, it is straightforward to show that p_j satisfies the 'reversibility' condition with respect to q_j:

$$p_j(\boldsymbol{\theta}_{k+1}(j)|\boldsymbol{\theta}_k(j))q_j(\boldsymbol{\theta}_k(j)) = p_j(\boldsymbol{\theta}_k(j)|\boldsymbol{\theta}_{k+1}(j))q_j(\boldsymbol{\theta}_{k+1}(j)) \tag{6}$$

Note that the equality in Eq. (6) is trivial when $\boldsymbol{\theta}_{k+1}(j) = \boldsymbol{\theta}_k(j)$. Combining Eqs. (4) and (6) and the fact that all the states lie in F_i, the transition PDF for the whole state $\boldsymbol{\theta}_k$ also satisfies the following reversibility condition with respect to $q(\cdot|F_i)$:

$$p(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k)q(\boldsymbol{\theta}_k|F_i) = p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})q(\boldsymbol{\theta}_{k+1}|F_i) \tag{7}$$

Thus, if the current sample $\boldsymbol{\theta}_k$ is distributed as $q(\cdot|F_i)$, then

$$p(\boldsymbol{\theta}_{k+1}) = \int p(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k)q(\boldsymbol{\theta}_k|F_i)d\boldsymbol{\theta}_k \tag{8}$$

$$= \int p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})q(\boldsymbol{\theta}_{k+1}|F_i)d\boldsymbol{\theta}_k|, \text{ by (7)}$$

$$= q(\boldsymbol{\theta}_{k+1}|F_i) \int p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})d\boldsymbol{\theta}_k$$

$$= q(\boldsymbol{\theta}_{k+1}|F_i)$$

since $\int p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})d\boldsymbol{\theta}_k = 1$. This shows that the next Markov chain sample $\boldsymbol{\theta}_{k+1}$ will also be distributed as $q(\cdot|F_i)$, and so the latter is indeed the stationary distribution for the generated Markov chain.

Step 1 can be viewed as a 'local' random walk in the neighborhood of the current state $\boldsymbol{\theta}_k$, while Step 2 ensures that the next sample always lie in F_i so as to produce the correct conditioning in the samples. Thus, Step 2 is in principle similar to the direct MCS approach in that both are based on accepting samples that lie in F_i. However, the 'acceptance' rate for the modified Metropolis algorithm should be considerably higher than for direct MCS because the candidate state $\tilde{\boldsymbol{\theta}}$ is simulated in the neighborhood of the current state $\boldsymbol{\theta}_k \in F_i$ if the proposal PDFs p_j^* are sufficiently local and so $\boldsymbol{\theta}_k$ should have a high probability of lying in F_i. Thus, Markov chain simulation accelerates the efficiency of exploring the failure region. The higher acceptance rate, however, is gained at the expense of introducing dependence between successive samples, which inevitably

reduces the efficiency of the conditional failure probability estimators, as shown later.

4. Subset simulation procedure

Utilizing the modified Metropolis method, subset simulation proceeds as follows. First, we simulate samples by direct MCS to compute $\tilde{P}_1$ from Eq. (3) as an estimate for $P(F_1)$. From these MCS samples, we can readily obtain some samples distributed as $q(\cdot|F_1)$, simply as those which lie in F_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 $\tilde{P}_2$ similar to Eq. (3). Observe that the Markov chain samples which lie in F_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 conditional levels until the failure event of interest, $F\ (=F_m)$, has been reached. At the ith conditional level, $1 \le i \le m-1$, let $\{\boldsymbol{\theta}_k^{(i)} : k = 1, \ldots, N\}$ be the Markov chain samples with distribution $q(\cdot|F_i)$, possibly coming from different chains generated by different seeds. Then

$$P(F_{i+1}|F_i) \approx \tilde{P}_{i+1} = \frac{1}{N} \sum_{k=1}^{N} I_{F_{i+1}}(\boldsymbol{\theta}_k^{(i)}) \qquad (9)$$

Finally, combining Eqs. (2), (3) and (9), the failure probability estimator is

$$\tilde{P}_{\mathrm{F}} = \prod_{i=1}^{m} \tilde{P}_i \qquad (10)$$

5. Implementation issues

5.1. Choice of proposal PDFs $\{p_j^*\}$

The proposal PDFs $\{p_j^*\}$ affect the deviation of the candidate state from the current state, and control the efficiency of the Markov chain samples in populating the failure region. Simulations show that the efficiency of the method is insensitive to the *type* of the proposal PDFs, and hence those which can be operated easily may be used. For example, the uniform PDF centered at the current sample with width $2l_j$ is a good candidate for p_j^*, and this is used in the examples in this work. The *spread* of the proposal PDFs affects the size of the region covered by the Markov chain samples, and consequently it controls the efficiency of the method. Small spread tends to increase the dependence between successive samples due to their proximity, slowing down convergence of the estimator and it may also cause ergodicity problems (see later). Excessively large spread, however, may reduce the acceptance rate, increasing the number of repeated Markov chain samples and so slowing down convergence. The optimal choice of the spread of p_j^* is therefore a trade-off between acceptance rate and correlation due to proximity. Roughly speaking, the spread of p_j^* may be chosen as some fraction of the standard deviation of the corresponding component θ_j as specified by the PDF $q(\boldsymbol{\theta})$, although the optimal choice depends on the particular type of problem.

5.2. Choice of intermediate failure events

The choice of the intermediate failure events $\{F_i\}$ plays a key role in the subset simulation procedure. Two issues are basic to the choice of the intermediate failure events. The first is the parameterization of the target failure event F which allows the generation of intermediate failure events by varying the value of the defined parameter. The second is the choice of the specific sequence of values of the defined parameter, which affects the values of the conditional probabilities $\{P(F_{i+1}|F_i)\}$ and hence the efficiency of the subset simulation procedure.

Many failure events encountered in engineering applications can be defined using a combination of union and intersection of some component failure events. In particular, consider a failure event F of the following form:

$$F = \bigcup_{j=1}^{L} \bigcap_{k=1}^{L_j} \{D_{jk}(\boldsymbol{\theta}) > C_{jk}(\boldsymbol{\theta})\} \qquad (11)$$

where $D_{jk}(\boldsymbol{\theta})$ and $C_{jk}(\boldsymbol{\theta})$ may be viewed as the demand and capacity variables of the (j,k) component of the system. The failure event F in Eq. (11) can be considered as the failure of a system with L sub-systems connected in series, where the jth sub-system consists of L_j components connected in parallel.

In order to apply subset simulation to compute the failure probability P_{F}, it is desirable to parameterize F with a single parameter so that the sequence of intermediate failure events $\{F_i : i = 1, \ldots, m-1\}$ can be generated by varying the parameter. This can be accomplished as follows. For the failure event in Eq. (11), define the 'critical demand-to-capacity ratio' (CDCR) Y as:

$$Y(\boldsymbol{\theta}) = \max_{j=1,\ldots,L} \min_{k=1,\ldots,L_j} \frac{D_{jk}(\boldsymbol{\theta})}{C_{jk}(\boldsymbol{\theta})} \qquad (12)$$

Then it can be easily verified that

$$F = \{Y(\boldsymbol{\theta}) > 1\} \qquad (13)$$

and so the sequence of intermediate failure events can be generated as:

$$F_i = \{Y(\boldsymbol{\theta}) > y_i\} \qquad (14)$$

where $0 < y_1 < \cdots < y_m = 1$ is a sequence of (normalized) intermediate threshold values.

Similarly, consider a failure F of the form:

$$F = \bigcap_{j=1}^{L} \bigcup_{k=1}^{L_j} \{D_{jk}(\boldsymbol{\theta}) > C_{jk}(\boldsymbol{\theta})\} \qquad (15)$$

which can be considered as the failure of a system with L sub-systems connected in parallel with the jth sub-system consisting of L_j components connected in series. One can easily verify that the definition

$$Y(\boldsymbol{\theta}) = \min_{j=1,\dots,L} \max_{k=1,\dots,L_j} \frac{D_{jk}(\boldsymbol{\theta})}{C_{jk}(\boldsymbol{\theta})} \qquad (16)$$

satisfies Eq. (13) and hence the sequence of failure events can again be generated based on Eq. (14).

The foregoing discussion can be generalized to failure events consisting of multiple stacks of union and intersection. Essentially, Y is defined using 'max' and 'min' in the same order corresponding to each occurrence of union ($\cup$) and intersection ($\cap$) in F, respectively. As another example, consider the first excursion failure of the interstory drift in any one story of a n_s-story building beyond a given threshold level b. Let the interstory drift response $\{X_j(t; \boldsymbol{\theta}) : j = 1, \dots, n_s\}$ be computed at the n_t time instants $t_1, \dots, t_{n_t}$ within the duration of interest. Then:

$$F = \bigcup_{j=1}^{n_s} \bigcup_{k=1}^{n_t} \{|X_j(t_k; \boldsymbol{\theta})| > b\} = \{Y(\boldsymbol{\theta}) > 1\} \qquad (17)$$

where

$$Y(\boldsymbol{\theta}) = \min_{j=1,\dots,n_s} \max_{k=1,\dots,n_t} \frac{|X_j(t_k)|}{b} \qquad (18)$$

The choice of the sequence of intermediate threshold values $\{y_1, \dots, y_m\}$ appearing in the parameterization of intermediate failure events affects the values of the conditional probabilities and hence the efficiency of the subset simulation procedure. If the sequence increases slowly, then the conditional probabilities will be large, and so their estimation requires less samples N. A slow sequence, however, requires more simulation levels m to reach the target failure event, increasing the total number of samples $N_T = mN$ in the whole procedure. Conversely, if the sequence increases too rapidly that the conditional failure events become rare, it will require more samples N to obtain an accurate estimate of the conditional failure probabilities in each simulation level, which again increases the total number of samples. It can thus be seen that the choice of the intermediate threshold values is a trade-off between the number of samples required in each simulation level and the number of simulation levels required to reach the target failure event.

The choice of the intermediate threshold values $\{y_i : i = 1, \dots, m-1\}$ deserves a detailed study which is left for future work. One strategy is to choose the y_i a priori, but then it is difficult to control values of the conditional probabilities $P(F_i|F_{i-1})$. Therefore, in this work, the y_i are chosen 'adaptively' so that the estimated conditional prob-

abilities are equal to a fixed value $p_0 \in (0, 1)$. This is accomplished by choosing the intermediate threshold level y_i ($i = 1, \dots, m-1$) as the $(1 - p_0)N$th largest value (i.e. an order statistic) among the CDCRs $\{Y(\boldsymbol{\theta}_k^{(i-1)}) : k = 1, \dots, N\}$ where the $\boldsymbol{\theta}_k^{(i-1)}$ are the Markov chain samples generated at the $(i-1)$th conditional level for $i = 2, \dots, m-1$, and the $\boldsymbol{\theta}_k^{(0)}$ are the samples from the initial MCS. Here, p_0 is assumed to be chosen so that $p_0 N$ and hence $(1 - p_0)N$ are positive integers, although this is not strictly necessary. This choice of the intermediate threshold levels implies that they are dependent on the conditional samples and will vary in different simulation runs. In the examples presented in this paper, the value p_0 is chosen to be 0.1, which is found to yield good efficiency.

6. Statistical properties of the estimators

In this section, we present results on the statistical properties of the estimators $\tilde{P}_i$ and $\tilde{P}_F$. They are derived assuming that the Markov chain generated according to the modified Metropolis method is (theoretically) ergodic, that is, its stationary distribution is unique and independent of the initial state of the chain. A discussion on ergodicity will follow after this section. It is assumed in this section that the intermediate failure events are chosen a priori. In the case where the intermediate threshold levels are chosen dependent on the conditional samples and hence vary in different simulation runs, as is the case in the examples presented in this paper, the derived results should hold approximately, provided such variation is not significant. Nevertheless, this approximate analysis is justified since the objective is to have an assessment of the quality of the probability estimate based on information available in one simulation run.

6.1. MCS estimator $\tilde{P}_1$

As well-known, the MCS estimator $\tilde{P}_1$ in Eq. (3) computed using the i.i.d. samples $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N\}$ converges almost surely to P_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 $\tilde{P}_1$, δ_1, defined as the ratio of the standard deviation to the mean of $\tilde{P}_1$, is given by:

$$\delta_1 = \sqrt{\frac{1 - P_1}{P_1 N}} \qquad (19)$$

6.2. Conditional probability estimator $\tilde{P}_i$ ($2 \le i \le m$)

Since the Markov chains generated at each conditional level are started with samples (selected from the previous simulation level) distributed as the corresponding target conditional PDF, the Markov chain samples used for computing the conditional probability estimators based on Eq. (9) are identically distributed as the target conditional

PDF. Using this result and taking expectation on both side of Eq. (9) shows that the conditional estimators $\tilde{P}_i$ ($i = 2, ..., m$) are unbiased. On the other hand, the Markov chain samples are dependent. In spite of this dependence, all the estimators $\tilde{P}_i$, still have the usual convergence properties of estimators using independent samples [17]. For example, $\tilde{P}_i$ converges almost surely to $P(F_i|F_{i-1})$ (Strong Law of Large Numbers), is consistent, and Normally distributed as $N \to \infty$ (Central Limit Theorem).

An expression for the c.o.v. for $\tilde{P}_i$ is next derived. At the $(i-1)$th conditional level, suppose that the number of Markov chains (each with a possibly different starting point) is N_c, and N/N_c samples have been simulated from each of these chains, so that the total number of Markov chain samples is N. Although the samples generated by different chains are in general dependent because the seeds for each chain may be dependent, it is assumed for simplicity in analysis that they are uncorrelated through the indicator function $I_{F_i}(\cdot)$, i.e. $E[I_{F_i}(\boldsymbol{\theta})I_{F_i}(\boldsymbol{\theta}')] - P(F_i|F_{i-1})^2 = 0$ if $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ are from different chains. Let $\boldsymbol{\theta}_{jk}^{(i)}$ be the kth sample in the jth Markov chain at simulation level i. For simplicity in notation, let $I_{jk}^{(i)} = I_{F_i}(\boldsymbol{\theta}_{jk}^{(i-1)})$ and $P_i = P(F_i|F_{i-1})$, $i = 2, ..., m$. Then:

$$E[\tilde{P}_i - P_i]^2 = E\left[\frac{1}{N}\sum_{j=1}^{N_c}\sum_{k=1}^{N/N_c}(I_{jk}^{(i)} - P_i)\right]^2$$

$$= \frac{1}{N^2}\sum_{j=1}^{N_c}E\left[\sum_{k=1}^{N/N_c}(I_{jk}^{(i)} - P_i)\right]^2 \quad (20)$$

For the jth chain,

$$E\left[\sum_{k=1}^{N/N_c}(I_{jk}^{(i)} - P_i)\right]^2 = \sum_{k,l=1}^{N/N_c}E(I_{jk}^{(i)} - P_i)(I_{jl}^{(i)} - P_i)$$

$$= \sum_{k,l=1}^{N/N_c}R_i(k - l) \quad (21)$$

where

$$R_i(k) = E(I_{jl}^{(i)} - P_i)(I_{j,l+k}^{(i)} - P_i) = E[I_{jl}^{(i)}I_{j,l+k}^{(i)}] - P_i^2 \quad (22)$$

is the covariance between $I_{jl}^{(i)}$ and $I_{j,l+k}^{(i)}$, for any $l = 1, ..., N/N_c$, and it is independent of l due to stationarity. It is also independent of the chain index j since all chains are probabilistically equivalent. Evaluating the sum in Eq. (21) with respect to $k - l$,

$$E\left[\sum_{k=1}^{N/N_c}(I_{jk}^{(i)} - P_i)\right]^2 = \frac{N}{N_c}R_i(0) + 2\sum_{k=1}^{N/N_c - 1}\left(\frac{N}{N_c} - k\right)R_i(k) \quad (23)$$

$$= \frac{N}{N_c}\left[R_i(0) + 2\sum_{k=1}^{N/N_c - 1}\left(1 - \frac{kN_c}{N}\right)R_i(k)\right]$$

Substituting Eq. (23) into (20) yields:

$$E[\tilde{P}_i - P_i]^2 = \frac{1}{N}\left[R_i(0) + 2\sum_{k=1}^{N/N_c - 1}\left(1 - \frac{kN_c}{N}\right)R_i(k)\right] \quad (24)$$

$$= \frac{R_i(0)}{N}\left[1 + 2\sum_{k=1}^{N/N_c - 1}\left(1 - \frac{kN_c}{N}\right)\rho_i(k)\right]$$

where

$$\rho_i(k) = R_i(k)/R_i(0) \quad (25)$$

is the correlation coefficient at lag k of the stationary sequence $\{I_{jk}^{(i)} : k = 1, ..., N/N_c\}$. Finally, since $I_{jk}^{(i)}$ is a Bernoulli random variable, $R_i(0) = \text{Var}[I_{jk}^{(i)}] = P_i(1 - P_i)$, and so the variance of $\tilde{P}_i$ is given by

$$\sigma_i^2 = E[\tilde{P}_i - P_i]^2 = \frac{P_i(1 - P_i)}{N}[1 + \gamma_i] \quad (26)$$

where

$$\gamma_i = 2\sum_{k=1}^{N/N_c - 1}\left(1 - \frac{kN_c}{N}\right)\rho_i(k) \quad (27)$$

The c.o.v. δ_i of $\tilde{P}_i$ is thus given by:

$$\delta_i = \sqrt{\frac{1 - P_i}{P_iN}(1 + \gamma_i)} \quad (28)$$

The covariance sequence $\{R_i(k) : i = 0, ..., N/N_c - 1\}$ can be estimated using the Markov chain samples $\{\boldsymbol{\theta}_{jk}^{(i-1)} : j = 1, ..., N_c; k = 1, ..., N/N_c\}$ at the $(i - 1)$th conditional level by:

$$R_i(k) \approx \tilde{R}_i(k) = \left(\frac{1}{N - kN_c}\sum_{j=1}^{N_c}\sum_{l=1}^{N/N_c - k}I_{jl}^{(i)}I_{j,l+k}^{(i)}\right) - \tilde{P}_i^2 \quad (29)$$

from which the correlation sequence $\{\rho_i(k) : k = 1, ..., N/N_c - 1\}$ and hence the correlation factor γ_i in Eq. (27) can also be estimated. Consequently, the c.o.v. δ_i for the conditional probability estimator $\tilde{P}_i$ can be estimated by Eq. (28), where P_i is approximated by $\tilde{P}_i$ using Eq. (9).

The factor $(1 - P_i)/P_iN$ in Eq. (28) is the familiar one for the square of the c.o.v. of the MCS estimators with N independent samples. The c.o.v. of $\tilde{P}_i$ can thus be considered as the one in MCS with an effective number of independent samples $N/(1 + \gamma_i)$. The efficiency of the estimator using dependent samples of a Markov chain ($\gamma_i > 0$) is therefore reduced compared to the case when the samples are independent ($\gamma_i = 0$), and smaller values of γ_i imply higher efficiency. The value of γ_i depends on the choice of the spread of the proposal PDFs (Section 5.1).

6.3. Failure probability estimator $\tilde{P}_F$

Since $\tilde{P}_1 \to P(F_1)$ and $\tilde{P}_i \to P(F_i|F_{i-1})$ ($2 \le i \le m$) almost surely as $N \to \infty$, $\tilde{P}_F \to P(F_1)\prod_{i=1}^{m-1}P(F_{i+1}|F_i) = P_F$ almost surely also. Due to the correlation between the conditional estimators $\{\tilde{P}_i\}$, $\tilde{P}_F$ is biased for every N, but it

is asymptotically unbiased. This 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 $\tilde{P}_{i+1}$. The results concerning the mean and c.o.v. of the failure probability estimator $\tilde{P}_F$ are summarized in the following two propositions.

Proposition 1. $\tilde{P}_F$ is biased for every N. The fractional bias is bounded above by:

$$\left| E\left[\frac{\tilde{P}_F - P_F}{P_F} \right] \right| \leq \sum_{i>j} \delta_i \delta_j + o(1/N) = O(1/N) \tag{30}$$

$\tilde{P}_F$ is thus asymptotically unbiased and the bias is $O(1/N)$.

Proof. Define $Z_i = (\tilde{P}_i - P_i)/\sigma_i$, then it is clear that $E[Z_i] = 0$ and $E[Z_i^2] = 1$, and $\tilde{P}_i = P_i + \sigma_i Z_i$.

$$\frac{\tilde{P}_F - P_F}{P_F} = \prod_{i=1}^m \tilde{P}_i/P_i - 1 \tag{31}$$

$$= \prod_{i=1}^m (1 + \delta_i Z_i) - 1$$

$$= \sum_{i=1}^m \delta_i Z_i + \sum_{i>j} \delta_i \delta_j Z_i Z_j + \sum_{i>j>k} \delta_i \delta_j \delta_k Z_i Z_j Z_k + \cdots + \prod_{i=1}^m \delta_i Z_i$$

Taking expectation and using $E[Z_i] = 0$,

$$E\left[\frac{\tilde{P}_F - P_F}{P_F} \right] = \sum_{i>j} \delta_i \delta_j E[Z_i Z_j] + \sum_{i>j>k} \delta_i \delta_j \delta_k E[Z_i Z_j Z_k] + \cdots$$
$$+ \left(\prod_{i=1}^m \delta_i \right) E\left[\prod_{i=1}^m Z_i \right] \tag{32}$$

If $\{Z_i\}$ are uncorrelated, $E[Z_i Z_j]$, $E[Z_i Z_j Z_k]$,..., are all zero, and hence $\tilde{P}_F$ will be unbiased. In general, however, $\{Z_i\}$ are correlated, so $\tilde{P}_F$ is biased for every N. On the other hand, since δ_i is $O(1/\sqrt{N})$ from Eq. (28), the first term in Eq. (32) is $O(1/N)$ while the remaining sums of higher products of the δ_i's are $o(1/N)$. Taking absolute value on both sides of Eq. (32), using the Cauchy–Schwarz inequality to obtain $|E[Z_i Z_j]| \leq \sqrt{E[Z_i^2]E[Z_j^2]} = 1$ and applying it to the R.H.S. of Eq. (32) proves the required proposition.

$\square$

Proposition 2. The c.o.v. δ of $\tilde{P}_F$ is bounded above by:

$$\delta^2 = E\left[\frac{\tilde{P}_F - P_F}{P_F} \right]^2 \leq \sum_{i,j=1}^m \delta_i \delta_j + o(1/N) = O(1/N) \tag{33}$$

$\tilde{P}_F$ is thus a consistent estimator and its c.o.v. δ is $O(1/\sqrt{N})$.

Proof. From Eq. (31)

$$E\left[\frac{\tilde{P}_F - P_F}{P_F} \right]^2 = E\left[\sum_{i=1}^m \delta_i Z_i + \sum_{i>j} \delta_i \delta_j Z_i Z_j + \cdots + \prod_{i=1}^m \delta_i Z_i \right]^2 \tag{34}$$

$$= \sum_{i,j=1}^m \delta_i \delta_j E[Z_i Z_j] + o(1/N)$$

$$\leq \sum_{i,j=1}^m \delta_i \delta_j + o(1/N) = O(1/N)$$

since $E[Z_i Z_j] \leq 1$ and $\delta_i = O(1/\sqrt{N})$.

Note that the c.o.v. δ in Eq. (33) is defined through the expected deviation about the target failure probability P_F instead of $E[\tilde{P}_F]$ so that the effects of bias are accounted for. The upper bound corresponds to the case when the conditional probability estimators $\{\tilde{P}_i\}$ are fully correlated. The actual c.o.v. depends on the correlation between the $\tilde{P}_i$s. If all the $\tilde{P}_i$s were uncorrelated, then

$$\delta^2 = \sum_{i=1}^m \delta_i^2 \tag{35}$$

Although the $\tilde{P}_i$s are generally correlated, simulations show that δ^2 may be well approximated by $\sum_{i=1}^m \delta_i^2$. This will be illustrated in Section 8.

To get an idea of the number of samples required to achieve a given accuracy in $\tilde{P}_F$, consider the case where $P(F_1) = P(F_{i+1}|F_i) = p_0$. Assume $\delta_i^2 = (1 + \gamma)(1 - p_0)/p_0 N$ to be the same for all levels. Using Eqs. (28) and (34), and noting that the number of simulation levels to achieve the target failure probability is $m = \log P_F/\log p_0$, 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

$$N_T \approx mN = |\log P_F|^r \frac{(1 + \gamma)(1 - p_0)}{p_0 |\log p_0|^r \delta^2} \tag{36}$$

where $r \leq 3$ depends on the actual correlation of the $\tilde{P}_i$s. Thus, for a fixed p_0 and δ, $N_T \propto |\log P_F|^r$. Compared with MCS, where $N_T \propto 1/P_F$, this implies a substantial improvement in efficiency when estimating small probabilities.

$\square$

7. Ergodicity of subset simulation procedure

The foregoing discussion assumes that the Markov chain generated according to the Metropolis method is ergodic, which guarantees that the conditional probability estimate based on the Markov chain samples from a single chain will tend to the corresponding theoretical conditional probability as $N \to \infty$. Theoretically, ergodicity can be always achieved by choosing a sufficiently large spread in the proposal PDFs $\{p_j^*\}$. Practically, with a finite number of Markov chain

samples, ergodicity often becomes an issue of whether the Markov chain samples used to estimate the conditional probabilities populate sufficiently well the important regions of the failure domain, where the main contributions to the failure probability come from. Intuitively, if there are some parts of the failure region of significant probability content that are not well visited by the Markov chain samples, the contribution of the probability mass from such regions will not be reflected in the estimator, and the conditional probability estimate will be significantly biased.

For a Markov chain started at a single point, ergodicity problems may arise due to the existence of disconnected failure regions that are separated by safe regions whose size is large compared to the spread of the proposal PDFs $\{p_j^*\}$. To see this, recall that to go from the current Markov state to the next, a candidate state is generated in the neighborhood of the current state according to the proposal PDFs. To transit from one failure region to another, the candidate state has to lie in the second failure region, but this is unlikely to happen if the spread of the proposal PDF is small compared to the safe region between the two failure regions. (Any candidate state that falls in the safe region is rejected by the Metropolis algorithm.) So, the safe region prohibits transition of the Markov chain between disconnected failure regions, resulting in ergodicity problems. The situation of disconnected failure regions may arise, for example, in failure problems for systems with components connected in series.

By using the samples from multiple Markov chains with different initial states obtained from previous conditional levels, as in the present methodology, ergodicity problems due to disconnected failure regions may be resolved. To see this, first note that the subset simulation procedure begins with MCS, which produces independent samples distributed in the whole parameter space. The MCS samples which lie in F_1 populate the different regions of F_1 according to the relative importance (probability content) of the regions. We can expect these samples populate F_1 sufficiently well, since otherwise the c.o.v. of $\tilde{P}_1$ will be large and subsequently the c.o.v. of $\tilde{P}_F$ will not be acceptable as the errors accumulate through subsequent conditional simulation levels. For the next conditional level, we simulate multiple Markov chains, starting at the conditional MCS samples distributed among different regions in F_1. This allows us to have seeds in the different regions of F_1, which could possibly be disconnected. The Markov chain initiated in each disconnected region will populate it at least locally, and so the combined Markov chains correctly account for the contributions from the regions to the conditional probability estimate.

In this manner, the contribution from disconnected failure regions in higher conditional levels can also be accounted

for as the Markov chain samples in these regions propagate through higher conditional levels. This argument holds, of course, as long as unimportant failure regions (i.e. those with negligible contribution) at lower conditional levels remain unimportant at higher conditional levels, since otherwise there may not be enough seeds (if any) at lower conditional levels to develop more Markov chain samples to account for their importance at higher levels. In conclusion, even though a Markov chain started from a single state may have ergodicity problems with disconnected failure regions, using Markov chain samples from multiple chains, as in the proposed method, can be expected to achieve practical ergodicity in spite of the existence of disconnected failure regions.

The foregoing argument suggests that the subset simulation procedure is likely to produce an ergodic estimator for failure probability, nevertheless it offers no guarantee for practical ergodicity. Whether ergodicity problems become an issue depends on the particular application and the choice of the proposal PDFs. It is worth-noting that similar issues related to ergodicity are expected to arise in any simulation method which tries to conclude 'global' information from some known local information, assuming implicitly that the known local information dominates the problem. For example, importance sampling using design point(s) implicitly assumes that the main contribution of failure probability comes from the neighborhood of the known design points and there are no other design points of significant contribution. Thus, in situations such as when the failure region is highly concave or there are other unknown design points, the importance-sampling estimator is biased and has an ergodicity problem. In view of this, one should appreciate the ergodic property of standard MCS, since it is a totally global procedure in the sense that it does not accumulate information about the failure region developed from local states only.

8. Examples

The subset simulation methodology is applied to solving first-excursion failure probabilities for two examples. In these examples, the input $W(t)$ is a Gaussian white noise process with spectral intensity S. The response of the system is computed at the discrete time instants $\{t_k = (k-1)\Delta t : k = 1, \dots n\}$, where the sampling interval is assumed to be $\Delta t = 0.02$ s and the duration of study is T = 30 s, so that the number of time instants is $n = T/\Delta t + 1 = 1501$. The uncertain state vector $\boldsymbol{\theta}$ then consists of the sequence of i.i.d. standard Gaussian random variables which generate the white noise input at the discrete time instants, $\{W(t_k) = \sqrt{2\pi S/\Delta t}\boldsymbol{\theta}_k : k = 1, \dots, n\}$, and so the number of uncertain parameters involved in the problem is $n = 1501$. The intermediate thresholds $\{b_i\}$ are chosen 'adaptively' (see Section 5.2) so that the estimates for the conditional probabilities
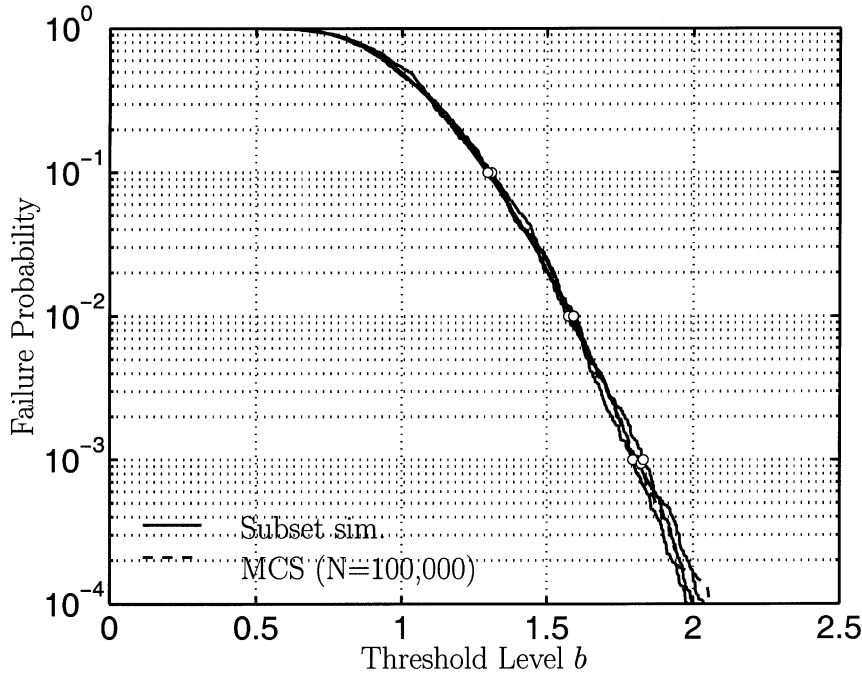
Fig. 1. Failure probability estimate for Example 1.

$\{P(F_i|F_{i-1})\}$ are equal to $p_0 = 0.1$. The proposal PDFs $\{p_j^*\}$ are chosen to be uniform PDFs with width 2, which are found to give good efficiency. The two examples are described first, followed by a discussion of the simulation results.

8.1. Example 1 SDOF linear oscillator

Consider a single degree-of-freedom (DOF) oscillator with natural frequency $\omega = 7.85$ rad/s (1.25 Hz) and damping ratio $\zeta = 2\%$ subjected to white noise excitation:

$$\ddot{X}(t) + 2\zeta\omega\dot{X}(t) + \omega^2 X(t) = W(t) \tag{37}$$

The system is assumed to start from rest, that is, $X(0) = 0$ and $\dot{X}(0) = 0$. The spectral intensity for the white noise is assumed to be $S = 1$.

Failure is defined as the displacement response exceeding a threshold level b within the first 30 s, that is,

$$F = \bigcup_{k=1}^{n} \{|X(t_k)| > b\} = \{\max_{k=1,\dots,n} |X(t_k)| > b\} \tag{38}$$

The intermediate failure events $\{F_i\}$ can then be chosen as:

$$F_i = \{\max_{k=1,\dots,n} |X(t_k)| > b_i\} \tag{39}$$

where $b_1 < \dots < b_m = b$ are the intermediate thresholds.

8.2. Example 2 Five-story nonlinear hysteretic shear building

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 2.7 m, is modeled by a Bouc–Wen hysteretic element [18] 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 $\ddot{a}(t)$ modeled by Clough–Penzien filtered white noise $W(t)$ modulated by an envelope function $e(t)$:

$$\ddot{a}(t) + 2\zeta_{s2}\omega_{s2}\dot{a}(t) + \omega_{s2}^2 a(t) = 2\zeta_{s1}\omega_{s1}\dot{a}_1(t) + \omega_{s1}^2 a_1(t) \tag{40}$$

$$\ddot{a}_1(t) + 2\zeta_{s1}\omega_{s1}\dot{a}_1(t) + \omega_{s1}^2 a_1(t) = e(t)W(t) \tag{41}$$

where $\omega_{s1} = 15.7$ rad/s (2.5 Hz) and $\omega_{s2} = 1.57$ rad/s (0.25 Hz) are the dominant and lower-cutoff frequency of the spectrum, respectively; $\zeta_{s1} = 0.6$ and $\zeta_{s2} = 0.8$ are the
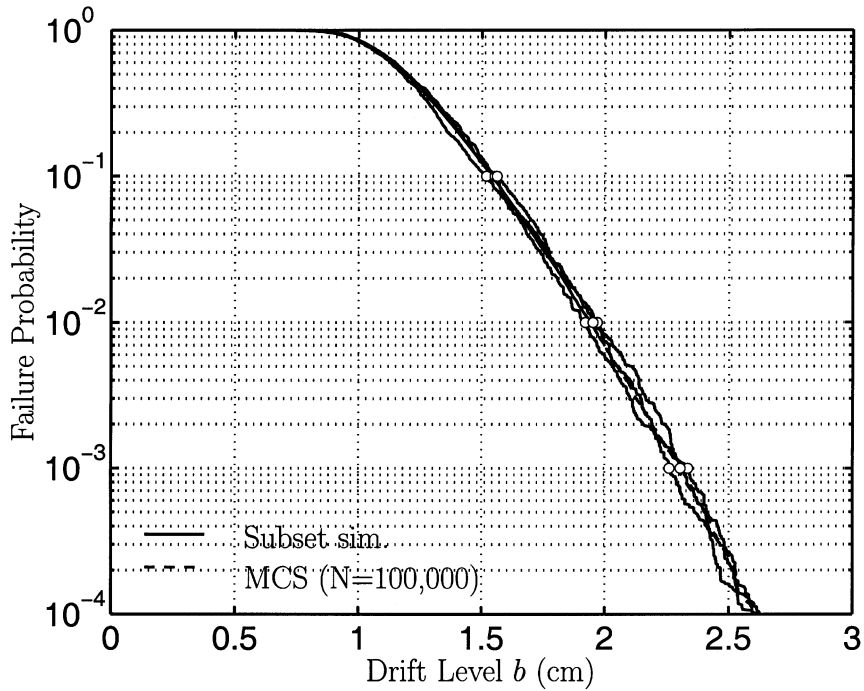
Fig. 2. Failure probability estimate for Example 2.

damping parameters associated with the dominant and lower-cutoff frequency, respectively. The envelope function $e(t)$ is assumed to vary quadratically as $(t/4)^2$ for the first 4 s, then settle at unity for 20 s, and finally decay as $\exp[-(t-24)^2/2]$ starting from $t = 24$ s. The spectral intensity for the Gaussian white noise $W(t)$ in Eq. (41) is assumed to be $S = 2.5 \times 10^{-3} \mathrm{m}^2/\mathrm{s}^3$. The nonlinear response of the structure is computed using the Newmark constant
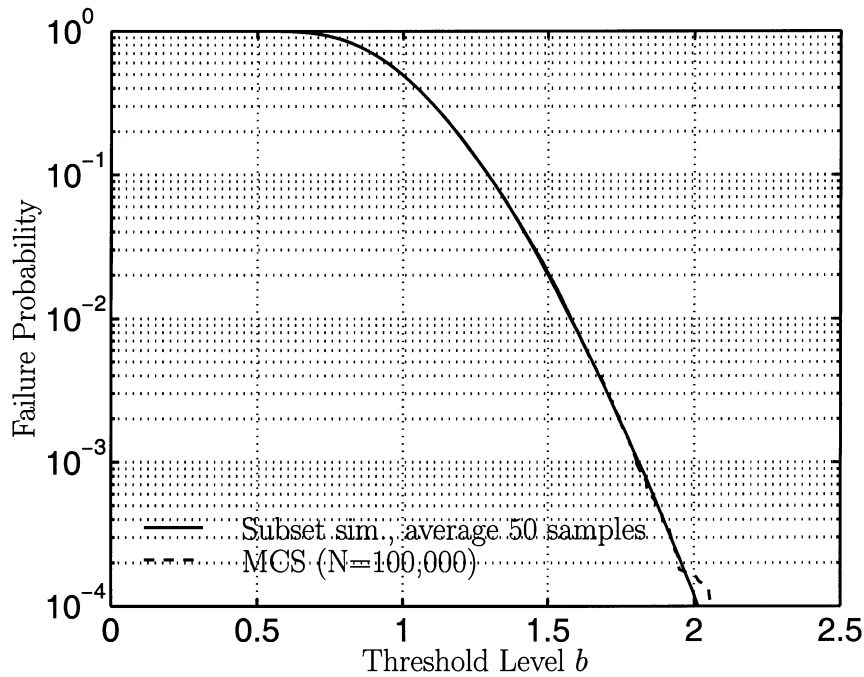
Fig. 3. Sample average of failure probability estimates from 50 simulation runs for Example 1.
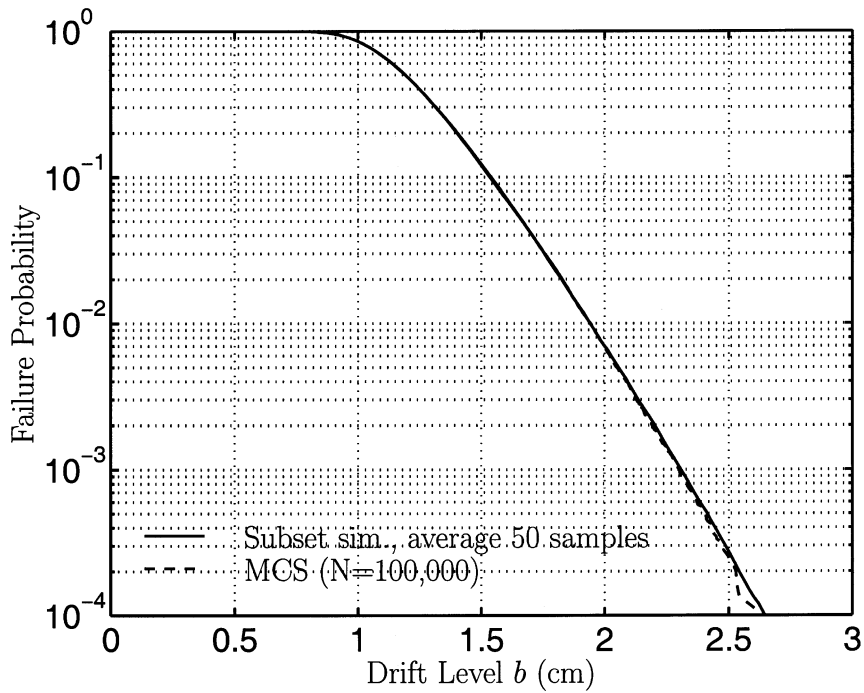
Fig. 4. Sample average of failure probability estimates from 50 simulation runs for Example 2.

acceleration scheme for a duration of 30 s, which is sufficient to capture the whole earthquake response history. *P–Δ* effects have been taken into account in computing the nonlinear response.

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 s. Let $X_j(t_k)$ denote the interstory drift response for the jth story at time t_k, $j = 1, ..., 5$,
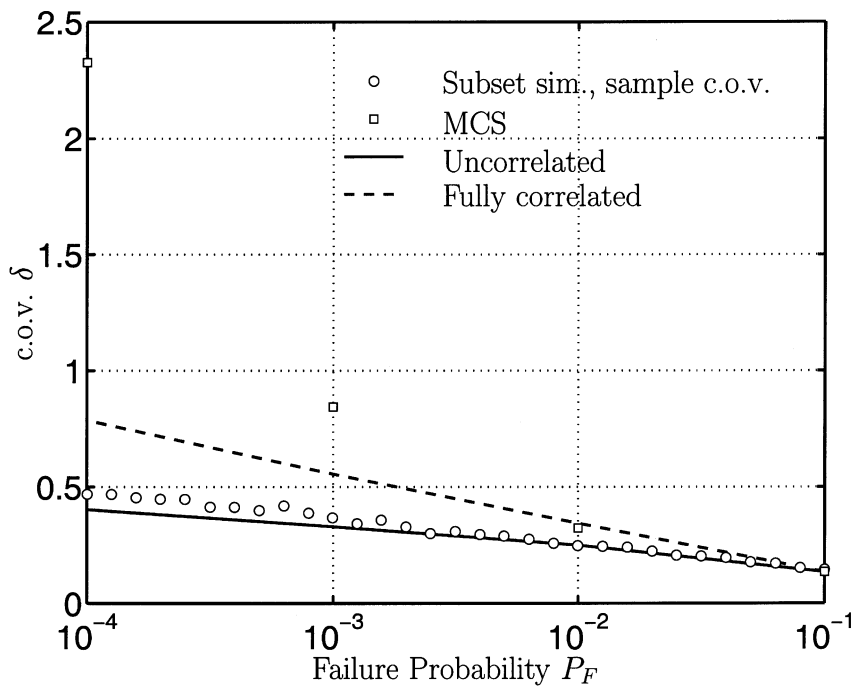
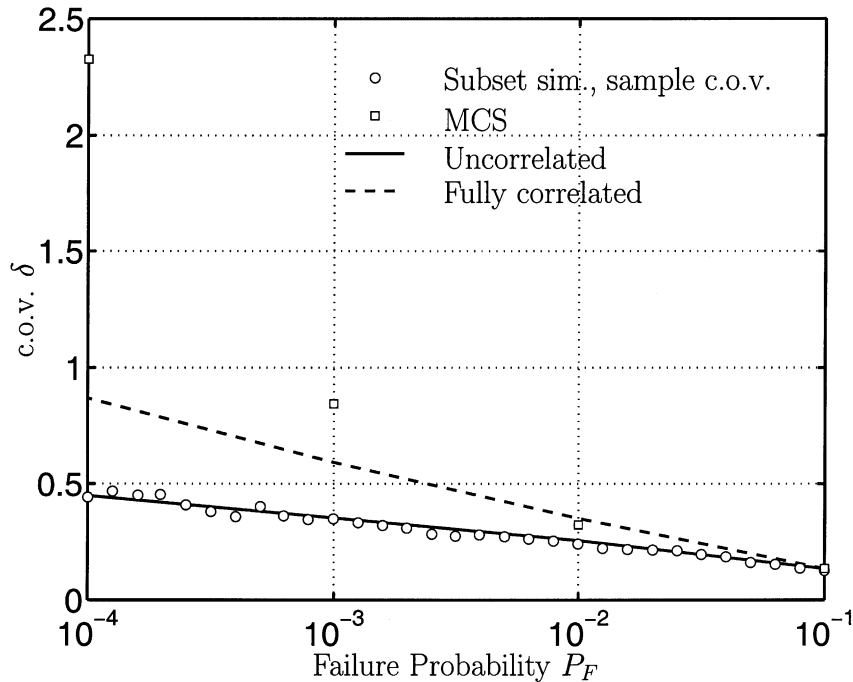Fig. 5. Coefficient of variation of failure probabilty estimate for Example 1.

Fig. 6. Coefficient of variation of failure probability estimate for Example 2.

$k = 1, ..., n$. Then the failure event of interest is

$$F = \bigcup_{j=1}^{5} \bigcup_{k=1}^{n} \{|X_j(t_k)| > b\} = \{\max_{j=1,...,5} \max_{k=1,...,n} |X_j(t_k)| > b\}$$

(42)

The intermediate failure events $\{F_i\}$ can then be chosen as:

$$F_i = \{\max_{j=1,...,5} \max_{k=1,...,n} |X_j(t_k)| > b_i\}$$

(43)

where $\{b_i\}$ are the intermediate interstory drift thresholds.

8.3. Discussion of results

Figs. 1 and 2 show the failure probability estimates from three independent simulation runs for Examples 1 and 2, respectively. In these figures, the failure probability estimates corresponding to the intermediate thresholds are shown with circles, which are computed using a total number of samples equal to 500, 1000, 1500 and 2000, respectively. That is, $N = 500$ samples have been used to estimate each of the conditional probabilities. Note that a single subset simulation run yields failure probability estimates for all threshold levels up to the largest one considered, not just for the b_i. For comparison, the results using MCS with 100,000 samples are also shown in the figures. It is seen in both examples that the results by subset simulation and MCS agree well.

To investigate the bias of the failure probability estimates, the average of failure probability estimates over 50

independent simulation runs is computed. The results are compared with those obtained by MCS in Figs. 3 and 4 for Examples 1 and 2, respectively. It is seen in both figures that the average of the failure probability estimates almost coincide with the MCS results, except in the low probability region where the error in the MCS estimates is significant. These results show that the bias due to the correlation between conditional probability estimates at different levels is negligible, and the failure probability estimate obtained by subset simulation is practically unbiased.

To investigate the actual variability of the estimates, as well as the validity of the upper bound on the c.o.v. in Eq. (33), the sample c.o.v. of the failure probability estimates over 50 sets of independent subset simulation runs are plotted with circles in Figs. 5 and 6 for Example 1 and 2, respectively. In these figures, the dashed line shows the upper bound computed using Eq. (33) averaged over the 50 simulation runs. The averaged estimates of the c.o.v. based on Eq. (35), which assumes the conditional probability estimators are uncorrelated, are also plotted with a solid line in the figure. In both figures, the sample c.o.v. (circles) lie close to the one assuming uncorrelated conditional probability estimates (solid line), showing that there is only a small reduction in efficiency due to the correlation. The smaller the failure probability and hence higher the simulation level, the more the sample c.o.v. is greater than the one assuming uncorrelated conditional probability estimates, since correlation effects accumulate with increasing simulation levels. To compare the efficiency of subset simulation with MCS, the c.o.v. of a MCS estimate at a particular

failure probability level using the same total number of samples as in subset simulation (that is, $N_T = 500, 1000, 1500, 2000$ for $P_F = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$, respectively) is plotted with squares in Figs. 5 and 6 for Examples 1 and 2, respectively. At a given failure probability level P_F, the c.o.v. of failure probability estimates computed by subset simulation and MCS using the same total number of samples (and hence approximately the same computational effort) can be compared. In Figs. 5 and 6, the c.o.v. achieved by subset simulation and MCS are approximately the same in the large probability region (e.g., $P_F \sim 0.1$). In particular, the values of c.o.v. for subset simulation (circle) and MCS (square) coincide at $P_F = 0.1$, since according to the subset simulation procedure with $p_0 = 0.1$ this probability estimate is computed based on an initial MCS. While the c.o.v. for subset simulation grows approximately in a linear fashion with the logarithm of decreasing failure probabilities, the c.o.v. for MCS grows exponentially. Thus, subset simulation becomes more and more efficient compared with MCS as the target probability of failure gets smaller.

Note that, from a theoretical point of view, Example 2 is much more complex than Example 1, since it involves the first excursion failure of a vector-valued response process, the system is nonlinear hysteretic, and the excitation is non-white and nonstationary. In spite of this difference in complexity, however, the trend of the c.o.v. with decreasing failure probability in both examples is similar, and so is the efficiency of subset simulation. Thus, it is expected that subset simulation can be applied efficiently to first excursion problems for a wide range of dynamical systems.

9. Conclusions

One of the major obstacles in applying simulation methods to estimating small failure probabilities is the need to simulate rare events. Subset simulation resolves this by breaking the problem into the estimation of a sequence of larger conditional probabilities. The Metropolis method has been modified to efficiently estimate the conditional probabilities by Markov chain simulation. Theoretical estimates for the coefficient of variation and results from numerical simulation demonstrate a substantial improvement in efficiency over standard MCS.

It should be noted that the choice of the proposal PDFs is crucial to the success of the subset simulation methodology. This choice deserves a detailed study when the method is applied to different types of problems. Also, better choices of intermediate threshold levels may be developed to further improve the efficiency of subset simulation. Future research may also focus on using the conditional samples generated in the subset simulation procedure for system analysis after failure occurs.

Appendix A

This appendix describes the original Metropolis algorithm and explains why it is not applicable to simulating random vectors with independent components in high dimensions.

A.1. Metropolis algorithm

Let the proposal PDF, $p^*(\xi|\theta)$, be a chosen n-dimensional joint PDF for ξ centered at θ with the symmetry property $p^*(\xi|\theta) = p^*(\theta|\xi)$. Generate a sequence of samples $\{\theta_1, \theta_2, ...\}$ from a given sample θ_1 by computing θ_{k+1} from θ_k, $k = 1, 2, ...$, as follows:

1. Generate a 'candidate' state $\tilde{\theta}$: Simulate ξ according to $p^*(\xi|\theta_k)$. Compute the ratio $r = q(\xi)/q(\theta_k)$. Set $\tilde{\theta} = \xi$ with probability $\min\{1, r\}$ and set $\tilde{\theta} = \theta_k$ with the remaining probability $1 - \min\{1, r\}$.
2. Accept/reject $\tilde{\theta}$: Check the location of $\tilde{\theta}$. If $\tilde{\theta} \in F_i$, accept it as the next sample, i.e. $\theta_{k+1} = \tilde{\theta}$; otherwise reject it and take the current sample as the next sample, i.e. $\theta_{k+1} = \theta_k$.

Essentially, the original Metropolis algorithm differs from the modified Metropolis algorithm in the way the candidate state is generated in Step 1. Note that the Markov chain samples generated according to the Metropolis algorithm are not all distinct. In particular, there is a non-zero probability, P_R, that the next state in the Markov chain will be equal to the current state. A large P_R implies that the current and the next state will be highly correlated. This generally increases the statistical variability of the estimator $\tilde{P}_i$ and reduces the efficiency of the simulation process. It is thus essential that P_R be generally small among the samples in order to achieve satisfactory efficiency.

It has been found that when the uncertain parameters θ are independent and the dimension n is large, P_R is close to 1, rendering the original Metropolis algorithm almost inapplicable. To demonstrate this, consider the case when $q(\theta) = \prod_{j=1}^n q(\theta_j)$, where $q(\theta_j)$ is the one-dimensional PDF for the component θ_j, being equal for all components j. As a common choice, assume that $p^*(\xi|\theta) = \prod_{j=1}^n p^*(\xi_j|\theta_j)$, where $p^*(\xi_j|\theta_j)$ denotes the one-dimensional PDF for ξ_j centered at θ_j with the symmetric property $p^*(\xi_j|\theta_j) = p^*(\theta_j|\xi_j)$. Let $r = \prod_{j=1}^n q(\xi_j)/q(\theta_j)$ and let u be uniformly distributed on $[0,1]$, independent of everything else. According to the Metropolis algorithm, the event $\{r < 1\}$ and an independent failed Bernoulli trial (with probability

$1 - r$) imply that the next state will be equal to the current state. Thus,

$$P_R \geq P(r < 1, u > r) \tag{A1}$$

$$= P(u > r | r < 1)P(r < 1)$$

$$= E[1 - r | r < 1]P(r < 1)$$

$$= \{1 - E[r | r < 1]\}P(r < 1)$$

To assess the quantities appearing in Eq. (A1), first note that, since $\{(\xi_j, \theta_j) : j = 1, ..., n\}$ are independently and identically distributed as $p^*(\xi_j | \theta_j)q(\theta_j)$, we have

$$\frac{\log(r)}{n} = \frac{1}{n} \sum_{j=1}^{n} \log \frac{q(\xi_j)}{q(\theta_j)} \to E\left[\log \frac{q(\xi)}{q(\theta)} \right] \tag{A2}$$

where the convergence is in an almost sure sense due to the Strong Law of Large Numbers; ξ and θ in Eq. (A2) are jointly distributed as $p^*(\xi | \theta)q(\theta)$. Since $\exp(\cdot)$ is a convex function, Jensen's inequality gives:

$$\exp\left(E\left[\log \frac{q(\xi)}{q(\theta)} \right] \right) \leq E\left[\exp\left(\log \frac{q(\xi)}{q(\theta)} \right) \right] = E\left[\frac{q(\xi)}{q(\theta)} \right] \tag{A3}$$

Further,

$$E\left[\frac{q(\xi)}{q(\theta)} \right] = \int \int \frac{q(\xi)}{q(\theta)} p^*(\xi | \theta)q(\theta)\mathrm{d}\theta\mathrm{d}\xi \tag{A4}$$

$$= \int q(\xi) \int p^*(\theta | \xi)\mathrm{d}\theta\mathrm{d}\xi \qquad \text{since} \qquad P^*(\xi | \theta) = p^*(\theta | \xi)$$

$$= \int q(\xi)\mathrm{d}\xi \qquad \text{since} \qquad \int p^*(\theta | \xi)\mathrm{d}\theta = 1$$

$$= 1$$

Combining Eqs. (A3) and (A4) gives:

$$E\left[\log \frac{q(\xi)}{q(\theta)} \right] = -C < 0 \tag{A5}$$

for some fixed positive constant C. Thus, Eqs. (A2) and (A5) imply, almost surely as $n \to \infty$,

$$r \to \exp(-nC) \tag{A6}$$

and hence r is exponentially small when the dimension n of the uncertain parameter space is large. It follows that $E[r | r < 1] \to 0$, $P(r < 1) \to 1$ and so $P_R \to 1$ as $n \to \infty$. Thus, it is unlikely that the Metropolis chain will transit to any other new distinct state as the dimension n is large, in which case nearly all the Markov chain samples will be equal.

Although it has not been demonstrated analytically, numerical simulations show that the above phenomenon occurs in more general situations; for example, when the components θ_j are not all identically distributed. In contrast, the modified Metropolis algorithm presented in Section 3 is applicable in high dimensions. To see this, let $r_j =$

$q_j(\xi_j)/q_j(\theta_j)$, $j = 1, ..., n$ and let $\{u_j : j = 1, ..., n\}$ be independent and uniformly distributed on [0,1], independent of everything else. For the modified Metropolis algorithm, note that the next state is equal to the current state either when the candidate state generated in Step 1 of the algorithm is equal to the current state or when the candidate state does not lie in F_i and hence is rejected in Step 2. So,

$$P_R = P\left(\left\{ \bigcap_{j=1}^{n} r_j < 1, u_j > r_j \right\} \cup \tilde{\boldsymbol{\theta}} \notin F_i \right) \tag{A7}$$

$$\leq P\left(\bigcap_{j=1}^{n} r_j < 1, u_j > r_j \right) + P(\tilde{\boldsymbol{\theta}} \notin F_i)$$

$$= \prod_{j=1}^{n} P(r_j < 1, u_j > r_j) + P(\tilde{\boldsymbol{\theta}} \notin F_i)$$

$$= \prod_{j=1}^{n} P(u_j > r_j | r_j < 1)P(r_j < 1) + P(\tilde{\boldsymbol{\theta}} \notin F_i)$$

$$= \prod_{j=1}^{n} \{1 - E[r_j | r_j < 1]\}P(r_j < 1) + P(\tilde{\boldsymbol{\theta}} \notin F_i)$$

Since the factors in the first product are always less than one, the first term will often tend to zero as $n \to \infty$. In this case, combining Eq. (A7) with the fact that $P_R \geq P(\tilde{\boldsymbol{\theta}} \notin F_i)$, it can be concluded that $P_R \to P(\tilde{\boldsymbol{\theta}} \notin F_i)$ as $n \to \infty$. This result can be expected intuitively, since when n is large, it is unlikely that the candidate state is equal to the current state, as this requires all the n 'component candidate states' $\{\xi_j : j = 1, ..., n\}$ be rejected in Step 1 of the algorithm. The event that the next state is equal to the current state then nearly corresponds to the event where a candidate state is rejected for not lying in F_i. Consequently, when the dimension n is large, P_R can be expected not to increase systematically with n, and hence the modified Metropolis algorithm is applicable even when the dimension is large.

References

[1] Rubinstein RY. Simulation and the Monte-Carlo method. New York: Wiley, 1981.

[2] Fishman GS. Monte Carlo: concepts, algorithms, and applications. New York: Springer, 1996.

[3] Hammersley JM, Handscomb DC. Monte-Carlo methods. London: Methuen, 1964.

[4] Schuëller GI, Stix R. A critical appraisal of methods to determine failure probabilities. Struct Safety 1987;4:293–309.

[5] Shinozuka M. Basic analysis of structural safety. J Struct Engng, ASCE 1983;109:721–40.

[6] Melchers RE. Importance sampling in structural systems. Struct Safety 1989;6:3–10.

[7] Hohenbichler M, Rackwitz R. Improvement of second-order reliability estimates by importance sampling. J Engng Mech, ASCE 1988;114(12):2195–8.

[8] Papadimitriou C, Beck JL, Katafygiotis LS. Asymptotic expansions

for reliabilities and moments of uncertain dynamic systems. J Engng Mech, ASCE 1997;123(12):1219–29.

[9] Der Kiureghian A, Dakessian T. Multiple design points in first and second-order reliability. Struct Safety 1998;20:37–49.

[10] Au SK, Papadimitriou C, Beck JL. Reliability of uncertain dynamical systems with multiple design points. Struct Safety 1999;21:113–33.

[11] Bucher CG. Adaptive sampling — an iterative fast Monte Carlo procedure. Struct Safety 1988;5:119–26.

[12] Karamchandani A, Bjerager P, Cornell CA, Adaptive importance sampling. In: Proceedings of the Fifth ICOSSAR, p. 855–62, San Francisco, 1989.

[13] Ang GL, Ang AH-S, Tang WH. Optimal importance sampling density estimator. J Engng Mech, ASCE 1992;118(6):1146–63.

[14] Au SK, Beck JL. A new adaptive importance sampling scheme. Struct Safety 1999;21:135–58.

[15] Schuëller GI, Pradlwarter HJ, Pandey, MD. Methods for reliability assessment of nonlinear systems under stochastic dynamic loading - a review. In Proceedings of EURODYN'93, p. 751–9. Balkema, 1993.

[16] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH. Equations of state calculations by fast computing machines. J Chem Phys 1953;21(6):1087–92.

[17] Doob JL. Stochastic processes. New York: Wiley, 1953.

[18] Wen YK. Method for random vibration of hysteric systems. J Eng Mech, ASCE 1976;102:249-63