Structural Safety 66 (2017) 62–73

Contents lists available at ScienceDirect

Structural Safety

journal homepage: www.elsevier.com/locate/strusafe

Subdomain sampling methods – Efficient algorithm for estimating failure probability

C. Hsein Juang [a,b], Wenping Gong [b,c,*], James R. Martin [b]

[a] MOE Key Laboratory of Road and Traffic Engineering, College of Transportation Engineering, Tongji University, Shanghai 201804, China
[b] Glenn Department of Civil Engineering, Clemson University, SC 29634-0911, USA
[c] Faculty of Engineering, China University of Geosciences, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 20 October 2015
Received in revised form 31 January 2017
Accepted 5 February 2017

Keywords:
Monte Carlo simulation
Hasofer-Lind reliability index
Total probability theorem
Importance sampling
Subdomain sampling method
Supported excavations

ABSTRACT

Uncertainties in the solution model and its input parameters make it difficult to ascertain the performance of an engineering system. While Monte Carlo simulation methods may be used to model the uncertain performance of such system, computational efficiency is a great challenge. To this end, subdomain sampling method (SSM), an efficient algorithm for estimating the failure probability of a system, is proposed in this study. The SSM involves a few steps. First, the possible domain of uncertain input variables of the system of concern is partitioned into a set of subdomains. Then, samples of uncertain variables are generated in each and every domain separately. Among these generated samples, those that lead to failure of the system are identified through a deterministic analysis. Finally, the failure probability is estimated using the total probability theorem. This SSM approach is referred to as the *coarse* subdomain sampling method, which is a fast algorithm with a generally acceptable accuracy. To reduce the variation of the failure probability estimate, a *refined* SSM is further developed by combining the coarse SSM with the importance sampling method. The accuracy and the efficiency of the proposed subdomain sampling methods, the coarse and refined SSMs, are demonstrated with two supported excavation problems.

1. Introduction

Uncertainties in the solution models and input parameters are often unavoidable in the analysis of a civil engineering system. In an uncertain environment, it is difficult to evaluate the performance of an engineering system with certainty. To cope with the uncertainty in the evaluated performance, a conservative factor of safety (FS) is often adopted in the design. Though the FS-based approach is simple, the "true" safety level of the resulting design is generally not known. While a design could be made conservative by adopting a sufficiently large FS, under-design (unsafe design) may still happen in the face of model and parameters uncertainties. To account for the uncertainties explicitly in the analysis and design, use of reliability-based design (RBD) method has long been advocated [20,6,2]. Within the context of RBD, the performance of the engineering system (called "system performance" herein) is usually studied with probabilistic methods that could explicitly account for the uncertainties in the solution model and input parameters. The outcome of the probabilistic analysis is a failure probability (P_f), expressed as:

$$P_f = \Pr[g(\boldsymbol{x}) \leqslant 0] = \int_{g(\boldsymbol{x})\leqslant 0} f(\boldsymbol{x})d\boldsymbol{x} \tag{1}$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_{n_x}]^T$ is a vector of uncertain variables x_i ($i = 1, 2, \cdots, n_x$), in which the subscript n_x is the dimension of uncertain variables; $g(\boldsymbol{x})$ is the performance function or limit state function, which is formulated such that $g(\boldsymbol{x}) \leq 0$ denotes the failure of the engineering system; and, $f(\boldsymbol{x})$ is the joint probability density function (PDF) of uncertain variables $\boldsymbol{x}$.

Difficulty in deriving the closed-form solution of the multifold probability integral in Eq. (1) has led to various approximation methods, such as first order second moment method (FOSM) (e.g., [26], first order reliability method (FORM) (e.g., [30,27], and point estimate-based moment method (PEM) (e.g., [38,39]. The outcome of these approximation methods is a reliability index (β), which may be related to the failure probability (P_f) as follows:

$$P_f = \Phi(-\beta) \tag{2}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal variable.

* Corresponding author at: Glenn Department of Civil Engineering, Clemson University, SC 29634-0911, USA.
E-mail address: wenping@clemson.edu (W. Gong).

http://dx.doi.org/10.1016/j.strusafe.2017.02.002
0167-4730/

Though these approximation methods have been widely used in engineering practices, some concerns remain: (1) the accuracy of the approximation methods might become an issue if the performance function is highly nonlinear and/or high-dimensional; (2) the evaluation of the partial derivative of the performance function might be an obstacle when using FOSM or FORM, especially in situations where the system performance could only be analyzed using numerical methods; and (3) the distribution of the performance function is approximated with its moments of finite order, both the evaluation of the moments and the approximation of the distribution might introduce errors. To avoid the drawbacks of these approximation methods, sampling-based methods such as Monte Carlo simulation (MCS) (e.g., [34,7] are usually employed. Within the context of MCS, samples of uncertain variables are drawn from the joint PDF of uncertain variables. Once the samples are generated, the system performance for each of these samples can be evaluated using a deterministic model, and then the failure probability is readily estimated.

$$P_f = \frac{1}{N}\sum_{i=1}^{i=N} I(\boldsymbol{x_i}) \tag{3}$$

where N is the number of samples of uncertain variables; $I(\boldsymbol{x_i})$ is an indicator function of the ith sample of uncertain variables $\boldsymbol{x_i}$, which is defined as follows: if $g(\boldsymbol{x}_i) \leq 0$ then $I(\boldsymbol{x}_i) = 1.0$, otherwise $I(\boldsymbol{x}_i) = 0$.

MCS-based approach generally yields an unbiased estimate of the failure probability; however, the required number of samples of uncertain variables, and thusly the number of deterministic evaluations of the system performance, may be too large to be computationally efficient, especially for problems of low failure probabilities. The issue of low computational efficiency becomes more profound when the system performance can only be evaluated using numerical methods. In such circumstances, various sophisticated sampling methods, such as importance sampling (e.g., [1,3,19] and subset simulation (e.g., [4,8,22], have been proposed to improve the computational efficiency. Nevertheless, the concern regarding the computational inefficiency of the sampling-based methods has not been settled fundamentally. For example, the construction of the importance sampling density function, which is an essential step in the importance sampling method, is often a challenging task as it requires the identification of the failure domain. In the case of the subset simulation, the required number of samples is largely dependent upon the magnitude of the failure probability. That is to say, effort to improve the computational efficiency through a new sampling method is deemed a worthwhile pursuit.

In this study, subdomain sampling methods (SSMs) are proposed for estimating the failure probability. Within the framework of the proposed SSMs, the possible domain of uncertain variables is partitioned into a set of subdomains with the aid of a distance index that is formulated in the standard normal space. Then, samples of uncertain variables are generated in each and every subdomain separately using a sampling algorithm suggested by Gong et al. [15]. Among these generated samples, deterministic analysis of the system performance can readily be conducted; and, the samples that lead to the failure of the system are identified. Next, the failure probability of the engineering system is estimated using the total probability theorem. This approach is referred to herein as the *coarse* SSM. Note that the failure samples identified with the coarse SSM can be used for constructing the importance sampling density function, which is a required element in the importance sampling method. Thus, a refined SSM, which takes the advantages of both the coarse SSM and the importance sampling method, is further developed for the analysis of critical systems.

This paper is organized as follows. First, the coarse SSM is introduced. Second, the refined SSM is presented. Third, two supported excavation problems are studied and used as demonstrative examples to investigate the accuracy and the efficiency of the proposed SSMs. A series of comparisons are made between the proposed SSMs and the existing methods such as subset simulation, FORM, and crude MCS. Finally, the conclusions are drawn on the basis of the results presented.

2. Coarse subdomain sampling method

In reference to Eq. (3), the samples that are located in the failure domain contribute the most to the failure probability estimate; however, the density of the samples, generated with the crude MCS, is proportional to the joint PDF and most samples are located in the region of high joint density values. As such, the crude MCS may not be computationally efficient. To overcome the limitations of the crude MCS, subdomain sampling methods (SSMs) are proposed in this study. Within the framework of SSMs, the possible domain of uncertain variables is partitioned into a set of subdomains; then, samples of uncertain variables are generated in each and every subdomain separately. These generated samples are "uniformly" distributed in the domain of uncertain variables and a larger number of samples are located in the failure domain; thus, the failure probability could be estimated with higher accuracy. Here, a distance index (d), which is employed to partition the domain of uncertain variables, is formulated based upon the Hasofer-Lind reliability index [21,27]:

$$d = \sqrt{[\boldsymbol{n}]^{\mathrm{T}}[\boldsymbol{R}]^{-1}[\boldsymbol{n}]} \tag{4}$$

where $\boldsymbol{R}$ is the correlation matrix among the equivalent standard normal variables $\boldsymbol{n} = [n_1, n_2, \ldots, n_{n_x}]^{\mathrm{T}}$, which may be estimated from the correlation matrix among the original uncertain variables $\boldsymbol{x} = [x_1, x_2, \ldots, x_{n_x}]^{\mathrm{T}}$ using the transformation suggested by Der Kiureghian and Liu [11]; and, the component n_i in $\boldsymbol{n}$ is related to the uncertain variable x_i in $\boldsymbol{x}$ as follows.

$$n_i = \Phi^{-1}[F(x_i)] \tag{5}$$

where $F(x_i)$ is the CDF of the uncertain variable x_i ($i = 1, 2, \cdots, n_x$). It is worth noting that d^2 is distributed as chi-square distribution with n_x degrees of freedom [15].

With the distance index formulated in Eq. (4), a possible domain of uncertain variables, denoted as $[0, d_{\max})$, is readily located.

$$\chi^2_{n_x}(d^2_{\max}) = \varepsilon \tag{6}$$

where $\chi^2_{n_x}(\cdot)$ is the chi-square CDF with n_x degrees of freedom, and ε is a probability that is relatively low and negligible. As can be seen from Eq. (6), the probability of the samples being located in and outside this possible domain, $[0, d_{\max})$, are $(1-\varepsilon)$ and ε, respectively.

2.1. Subdomain partition

In reference to Gong et al. [15], the likelihoods of the samples of uncertain variables being located in the subdomains, denoted as $(p_{d1}, p_{d2}, p_{d3}, \ldots)$, should be taken as a decreasing sequence, in the subdomain partition for the purpose of being computationally efficient.

$$\begin{aligned} p_{di} &= \Pr\left[d_{i-1} \leqslant \sqrt{[\boldsymbol{n}]^{\mathrm{T}}[\boldsymbol{R}]^{-1}[\boldsymbol{n}]} < d_i\right] = \Pr\left[d^2_{i-1} \leqslant d^2 < d^2_i\right] \\ &= \chi^2_{n_x}\left(d^2_i\right) - \chi^2_{n_x}\left(d^2_{i-1}\right) \end{aligned} \tag{7}$$

where p_{di} is the likelihood of the uncertain variables being located in the ith subdomain $[d_{i-1}, d_i)$. Similar to those in Gong et al. [15], $p_{d1} = 1/2$, $p_{d2} = 1/2^2$, $p_{d3} = 1/2^3$, ... are taken in this study.