Computers and Structures 153 (2015) 239–251

Contents lists available at ScienceDirect

Computers and Structures

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

A generalized Subset Simulation approach for estimating small failure probabilities of multiple stochastic responses

Hong-Shuang Li [a,*], Yuan-Zhuo Ma [a], Zijun Cao [b]

[a] Key Laboratory of Fundamental Science for National Defense-Advanced Design Technology of Flight Vehicles, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[b] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 2 October 2013
Accepted 13 October 2014
Available online 12 November 2014

Keywords:
Subset Simulation
Multiple stochastic responses
Reliability analysis
High dimension
Small failure probability

ABSTRACT

Estimating failure probabilities of multiple stochastic responses using a single run of Subset Simulation remains a challenging task in structural reliability analysis. To address this issue, this study develops a generalized Subset Simulation (GSS) approach, in which a unified intermediate event is defined to drive the simulation procedure progressively approaching multiple failure regions simultaneously. It bypasses the sorting difficulty arising from the multiple stochastic responses. The failure probabilities of multiple stochastic responses are obtained simultaneously by a single run of GSS. Finally, four representative examples are used to demonstrate the efficiency, accuracy and robustness of the proposed GSS approach.

1. Introduction

Estimating failure probability P_F of a failure mode concerned is one of the most challenging and fundamental problem in structural reliability analysis. For a given failure mode, its corresponding P_F can be expressed as the evaluation of the following multi-dimensional integral [1,2]

$$P_F = \mathrm{P}(\mathbf{X} \in F) = \int I_F(\mathbf{X}) q(\mathbf{X}) d\mathbf{X} \tag{1}$$

where $\mathbf{X} = [x_1, \ldots, x_n] \in \Omega \subset R^n$ represents an n-dimensional random vector, in which $x_1, \ldots, x_n$ are n random parameters of the structural system concerned; $q(\mathbf{X})$ is the joint probability density function (PDF) of $\mathbf{X}$; and $F = \{g(\mathbf{X}) < 0\}$ is the target failure region in the parameter space. Note that $g(\mathbf{X})$ is the limit state function (LSF) that divides the parameter space Ω into a safety region with $g(\mathbf{X}) \geqslant 0$ and a failure region with $g(\mathbf{X}) < 0$. Without much loss of generality, it is assumed that the components (i.e., $x_1, \ldots, x_n$) of $\mathbf{X}$ are independent so that $q(\mathbf{X}) = \prod_{i=1}^{n} q_i(x_i)$, where q_i, $i = 1, 2, \ldots, n$, are the marginal PDFs for x_i. Usually, the dependent samples can be generated by transformation of independent ones in applications [3–5].

In general, P_F cannot be efficiently evaluated by direct numerical integration because multi-dimensional integral is involved (see Eq. (1)) and the failure region can be very complicated. Many numerical techniques have been developed for estimating P_F. They can be roughly classified into two categories by their features. In the first category, the LSF is approximated by the first order or second order Taylor series expansion around a reference point (i.e., the so-called most probable point or design point) [1,2,6–9]. Then, the reliability index is calculated through the simple approximated LSF and the probabilistic information of the input random vector. The concept and calculation procedure of these methods are simple and they also were observed efficiently in many previous studies [1,2,6–9]. However, this category of methods cannot be applied to high dimensional problems ($n > 20$) [10] or the problems with highly nonlinear LSFs. They may also trigger convergence issues when the required gradient and/or Hessian matrix of the LSF are calculated by finite difference methods.

The second category of approaches for estimating the failure probability are collectively referred to as simulation based methods, including the crude Monte Carlo Simulation (MCS) [11,12], important sampling (IS) methods [12,13], directional simulation (DS) [14–16], line sampling (LS) method [17,18], and Subset Simulation (SS) [19]. The crude MCS [11,12] is one of the most well-known simulation methods for reliability analysis because it is independent of the complexity and dimension of the problem. However, the crude MCS suffers from its inefficiency at small probability levels due to the demand of a huge amount of samples [13–22]. To address this drawback, many variance reduction techniques have been proposed. Based on a prudent choice of the

* Corresponding author. Tel.: +86 25 84890119; fax: +86 25 84891422.
E-mail address: hongshuangli@nuaa.edu.cn (H.-S. Li).

http://dx.doi.org/10.1016/j.compstruc.2014.10.014
0045-7949/

importance sampling density (ISD), IS methods shift the underlying distributions toward the target failure region so that more samples are generated in the failure region [12,13]. Many schemes employ design points to construct the ISD, which serve as the new sampling center. This strategy is usually appropriate when the dimension n of the problem is relatively low and the failure region is relatively simple [20,22,23], while its efficiency decreases as the number of dimensions and/or the complexity of the problem increase [21,24]. DS [14–16] generates samples in an independent standard normal space with polar coordinates. It reduces the number of dimensions of the problem from n to $n-1$ by interpolation or solving the nonlinear equation. DS presents much better efficient and accurate than the crude MCS implemented in Cartesian coordinates when the LSF is close to a spherical surface, while no advantage for problems with linear LSFs. LS method [17,18] aims to efficiently resolve the high dimensional difficulty in reliability analysis through interpolation in the important direction and random sampling of the rest $n-1$ dimensions. It works efficiently when the important direction is close to the steepest descent direction (i.e., optimal important direction) of the LSF. It can be seen that LS requires prior information on the important direction in the standard normal space. However, such information might not be available in many cases. SS, which was first pioneered by Au and Beck [19], is efficient for estimating small failure probabilities and robust to dimensions. It converts a small failure probability into a product of relatively large conditional probabilities by introducing intermediate events. SS has been recognized as one of the most successful reliability methods in the past decade [25]. It has been widely applied to estimate the small failure probabilities of general dynamical systems encountered in engineering reliability analysis [19,26]. Applications of SS for some reliability benchmark problems [25,27] and several variants of SS are also available in previous studies, e.g., Refs. [28–33].

Most of reliability methods described above are developed for estimating the failure probability of a single stochastic response (i.e., a single LSF), but relatively few efforts were devoted to the capacity of estimating, simultaneously, the failure probabilities of multiple stochastic responses (i.e., multiple LSFs), which usually arises in reliability-based design when multi failure criteria are considered. The crude MCS is able to estimate the failure probabilities of multiple stochastic responses simultaneously, but it suffers from lack of efficiency at small failure probability levels. Compared with the crude MCS, SS improves significantly the efficiency and resolution of estimating the failure probability at small probability levels, but it remains a difficulty to estimate failure probabilities of multiple stochastic responses simultaneously. A parallel Subset Simulation (PSS) approach has been proposed for multiple stochastic responses [33]. In the PSS, a principle variable that is correlated with all the LSFs of interest is defined to drive the simulation to gradually approach the multiple failure regions. However, determination of a proper principle variable in the PSS is a not trivial task. In addition, it is difficult to theoretically verify the correlation between the principle variable and all the LSFs concerned in the subsequent SS levels after the level of MCS although the empirical verification on the correlation is possible using numerical examples [33].

The objective of this study is to develop a generalized Subset Simulation (GSS) approach to simultaneously estimate the failure probabilities of multiple stochastic responses. The proposed GSS inherits the excellent properties of the original SS, e.g., robustness to dimension, high efficiency at small probability levels, and independence to model complexity, etc., and is applicable for estimating failure probabilities of multiple stochastic responses. In the GSS approach, a unified intermediate event is constructed to resolve the sorting difficulty in the original SS. The determination of intermediate events for each stochastic response in GSS is identical with that in the original SS, while the unified intermediate events drive the simulation to, progressively and simultaneously, approach the multiple failure regions defined by multiple stochastic responses in a single run of simulation.

The paper is organized as follows. Section 2 presents the fundamental principle and implementation procedure of the original SS for a single stochastic response, followed by the development of the proposed GSS for multiple stochastic responses. Then, the proposed GSS approach is illustrated through two numerical examples and two high dimensional reliability benchmark problems. Results obtained from the proposed GSS are compared with those obtained from the original SS and the crude MCS. Finally, conclusions are given in the last section.

2. Subset Simulation for a single stochastic response

The proposed GSS is developed from the original SS. To facilitate the understanding of the proposed GSS, let us first briefly review its fundamental principle and implementation procedure of the original SS.

2.1. SS principle

The basic idea of SS is to convert a small failure probability into a product of a sequence of relatively large conditional probabilities by introducing the intermediate events adaptively [19]. Let F denote the target failure event defined as $F=\{g(\mathbf{X})\leqslant b\}$, where b is the desired response threshold for a performance index in a structure of interest. The failure occurs when $g(\mathbf{X})\leqslant b$. Let $F_1\supset F_2\supset\cdots\supset F_m=F$ denote a sequence of nested intermediate events. Then, the target failure probability can be expressed as [19]

$$P_F=\mathrm{P}(F)=\mathrm{P}(F_m)=\mathrm{P}(F_m|F_{m-1})\mathrm{P}(F_{m-1})=\cdots$$
$$=\mathrm{P}(F_1)\prod_{i=2}^{m}\mathrm{P}(F_i|F_{i-1}) \tag{2}$$

The intermediate events have similar expressions to the target failure event F, i.e., $F_i=\{g(\mathbf{X})\leqslant b_i,\ i=1,\ldots,m\}$ $(b=b_m<\cdots<b_2<b_1)$, where m is the total number of intermediate events. They can be determined in an adaptive way if a constant p_0 is given to the conditional probability [19]. Note that generating conditional samples is pivotal to implementing the original SS. A modified Metropolis–Hasting algorithm is developed and applied to generate the conditional samples in SS [19]. Details of the algorithm of the original SS are referred to Ref. [19].

2.2. Implementation procedure

Without loss of generality, N samples and a constant value p_0 are employed in each simulation level of SS.

(1) SS starts with the crude MCS to generate N samples $\mathbf{X}_i$ $(i=1,\ldots,N)$ in the parameter space Ω. Then, the first response threshold b_1 is determined by setting the first conditional probability $\mathrm{P}(F_1)$ be equal to p_0 from the point view of statistical inference. This is simply achieved by selecting the first Np_0 samples from the ascending sequence of response values $\{g(\mathbf{X}_i)\}$. If Np_0 is not an integer, one can use a new constant $[Np_0]/N$ to replace p_0 to determine b_1. For simplification, the symbol "p_0" is still used to denote the conditional probability in the remaining part of the paper. The selected Np_0 samples are used as the seed samples for generating conditional samples in the next simulation level. It is obvious that the first intermediate event F_1 is conditional on the whole parameter space Ω. Therefore, these Np_0 samples belong to the first intermediate event