Aerospace Science and Technology 29 (2013) 253–261

Contents lists available at SciVerse ScienceDirect

Aerospace Science and Technology

www.elsevier.com/locate/aescte

A novel adaptive importance sampling algorithm based on Markov chain and low-discrepancy sequence

Xiukai Yuan [a,*], Zhenzhou Lu [b], Changcong Zhou [b], Zhufeng Yue [c]

[a] Department of Aeronautics, Xiamen University, Xiamen 361005, PR China
[b] School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, PR China
[c] School of Mechanics, Civil Engineering & Architecture, Northwestern Polytechnical University, Xi'an 710072, PR China

ARTICLE INFO

Article history:
Received 11 January 2012
Received in revised form 18 March 2013
Accepted 22 March 2013
Available online 29 March 2013

Keywords:
Reliability
Importance sampling
Markov chain
Low-discrepancy sequence
Monte Carlo simulation

ABSTRACT

A novel adaptive importance sampling method is proposed to estimate the structural failure probability. It properly utilizes Markov chain algorithm to form an adaptive importance sampling procedure. The main concept is suggesting the proposal distributions of Markov chain as the importance sampling density. Markov chain states can adaptively populate the important failure regions thus the importance sampling based on them will yield an efficient and accurate estimate of the failure probability. Compared with existent methods, it does not need the solution of the design point(s) or the pre-sampling in the failure region. Various examples are given to demonstrate the advantages of the proposed method.

1. Introduction

In structural reliability analysis, the failure probability can be defined as the following multifold probability integral:

$$P_F = \int \cdots \int_F f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int \cdots \int_{R^n} I_F(\boldsymbol{x}) f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{1}$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ is the n-dimensional vector of random variables; $f(\boldsymbol{x})$ is the joint probability density function (PDF) of $\boldsymbol{x}$ and $I_F(\boldsymbol{x})$ is the indicator function of the failure region $F = \{\boldsymbol{x}\colon g(\boldsymbol{x}) \leqslant 0\}$ (where $g(\boldsymbol{x})$ is the limit state function): $I_F(\boldsymbol{x}) = 1$ if $g(\boldsymbol{x}) \leqslant 0$ and $I_F(\boldsymbol{x}) = 0$ otherwise.

In order to evaluate the failure probability P_F, it is necessary to compute an integral over n-dimension space. However, in practice, it is difficult to evaluate P_F when the number of the dimensions n is large or the failure region F has a complicated shape. In addition, the efficiency of the computing method is the main concern when dealing with a practical problem, as reliability analysis with finite element model is usually quite time-consuming.

Several methods have been developed in order to solve the integral. Among which Monte Carlo simulation (MCS) [25] method is one of the most widely used methods for handling complex problems. However, it shows a poor computational efficiency in dealing with problems of small failure probability or problems with costly finite element analyses. Hence, many variance reduction techniques have been developed to address this disadvantage, such as importance sampling [1,4,5,8,10–13,15,18,20,21], subset simulation [6,7], and line sampling technique [23,24], etc.

Importance sampling techniques have been developed over the past few decades, and have been one of the most prevalent approaches in the structural reliability analysis. Its basic idea is to shift underlying distribution towards the failure region to gain information from rare events more efficiently. It draws samples of the random variables from a pre-selected distribution which is concentrated in the important failure region. The success of the method relies on the proper choice of the importance sampling density (ISD). The optimal choice for the ISD is practically infeasible since it requires knowledge of failure probability a priori. Several methods have been developed to construct the ISD, which is the approximate optimal ISD or a different one. At present, the most prevalent choices are those based on design point(s) [5,8,18] or kernel density estimators [1,4,10–13,15,20,21].

A design point is usually obtained by solving a constrained optimization problem. When multiple design points exist, the search for them may be difficult as it requires more sophisticated algorithms for the optimization problem. Thus in an importance sampling process, the search for design point(s) may occupy a great portion of the total computational effort.

* Corresponding author.
E-mail addresses: xiukaiyuan@xmu.edu.cn (X. Yuan), zhenzhoulu@nwpu.edu.cn (Z. Lu).

1270-9638/$ – see front matter
http://dx.doi.org/10.1016/j.ast.2013.03.008

The kernel sampling density (KSD) is constructed based on the samples in the failure region which are simulated according to $f(\boldsymbol{x})$. It tends to the optimal ISD as the number of samples increases, thus it can be served as a good choice for the ISD. However, it needs pre-sampling in the failure region F. The points in F can be generated using standard rejection sampling from the original PDF [1], however this is extremely inefficient in cases where the failure probability is small. Another method which adopts Markov chain Metropolis algorithm [19] can overcome this drawback [4]. In this method, a number of points in F are simulated as the states of a Markov chain and their PDF asymptotically tends to the optimal ISD as the number of Markov chain steps increases. Subsequently, a KSD estimator can be constructed based on these points. The total computational cost of this approach mainly consists of the simulation of Markov chain samples in order to construct the KSD and the subsequent simulation of importance sampling in the estimation of the failure probability. Some other adaptive importance sampling strategies have also been developed, i.e., the non-parametric adaptive ISD [20] and the cross-entropy (CE) method [10,11,13]. In [20], a non-parametric adaptive importance sampling procedure was proposed which adapts the auxiliary PDF (weight and kernels) by using an algorithm at each iteration to approach the optimal ISD. The cross-entropy method [10,11,13] can be seen as an adaptive importance sampling procedure. The cross-entropy is a measure of closeness between two sampling distributions. The basic idea of the cross-entropy method is to choose the ISD in a specified class of densities such that the cross-entropy between the optimal ISD and implementing ISD is minimal. These methods provide alternative and adaptive ways of approximating the optimal ISD. Other strategies can also be found in literatures to further improve the performance of importance sampling, i.e., adopting the splitting strategy as given in [21] or the surrogate model (Kriging model) as in [15].

Recently, quasi-Monte Carlo (QMC) has gained more and more popularity in studying multidimensional integration [9,14,22,26]. Compared with the traditional MCS method, quasi-Monte Carlo method replaces the pseudo-random numbers used in MCS by low-discrepancy sequences. Sobol [26] drew the conclusion that applying quasi-Monte Carlo is more promising than Monte Carlo algorithms and Bosserta et al. [9] also pointed out that the numerical integration converges faster with a higher accuracy by using quasi-random numbers compared with using pseudo-random numbers. The quasi-Monte Carlo method which is also called the low-discrepancy sampling method has been applied in many fields, i.e. computational physics, statistics and mathematical finance. Recently, Nie and Ellingwood [22] and Dai and Wang [14] applied it to structural reliability analysis. Various low-discrepancy sequences were investigated in both Refs. [22] and [14]. Dai and Wang successfully combined the low-discrepancy sequences with traditional importance sampling technique to further improve the computational efficiency.

In this paper, a novel importance sampling approach, called Markov chain importance sampling method (MCIS), is proposed to efficiently handle structural reliability analysis. The innovative part of this contribution is that the proposed method modify the Markov chain Metropolis algorithm into an adaptive importance sampling procedure, thus the proposed method combines the advantages of both the Markov chain Metropolis algorithm and the importance sampling method. The basic idea of the proposed method is to take the proposal distribution of Markov chain as the ISD, and the Markov chain time steps are selected as the importance sampling centers. According to the property of the Metropolis algorithm, Markov chain states can adaptively populate the important region, thus the proposed importance sampling based on these states will yield an efficient and accurate estimate of the failure probability. The most desirable feature is that it can complete the importance sampling with only an initial point, no more information is needed, e.g. design point(s) or pre-samples. Furthermore, the low-discrepancy sequence is adopted in the proposed method to further improve the accuracy.

2. Original Markov chain Metropolis algorithm

Markov chain, also called Metropolis–Hastings algorithm, is a very useful and powerful tool for simulating samples according to an arbitrary PDF and it has become more and more widely used in structural reliability analysis [4,6]. In the Metropolis–Hastings algorithm, the samples are simulated as the states of a special Markov chain whose limiting stationary distribution is equal to the target PDF. This means that the PDF of the samples, which are Markov chain states, converges to a unique limit, the target PDF, as the number of Markov steps increases. In structural reliability analysis, this algorithm can be used to efficiently simulate the samples in the failure region and the samples can adaptively populate the important failure region. Taking the optimal sampling density $f(\boldsymbol{x}|F) = I_F(\boldsymbol{x})f(\boldsymbol{x})/P_F$ as the stationary distribution, the Markov chain states $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ (M is the total number of Markov chain time steps) are asymptotically distributed as $f(\boldsymbol{x}|F)$. See [4,6,19] and [3] for more details of applying the Metropolis–Hastings algorithm to reliability analysis.

The original Markov chain procedure for simulating samples $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ with limit stationary distribution equal to target PDF $f(\boldsymbol{x}|F)$ is briefly described below.

(1) Select an initial state of Markov chain.

A 'seed' point $\boldsymbol{x}_1$ should be selected in the failure region F either randomly or deterministically to start the sequence. It may be obtained as a point simulated according to the PDF $f(\boldsymbol{x})$, or be assigned based on engineering judgment as a point that leads to system failure. In most structural systems of practical interest the determination of the point that lead to failure does not present serious difficulties. Even if the current sample is not distributed as $f(\boldsymbol{x}|F)$, the limit distribution property of Markov chain guarantees that the distribution of simulated samples will tend to $f(\boldsymbol{x}|F)$ as the number of Markov chain steps increases [3,4].

(2) Determine the $(j+1)$-th time step state $\boldsymbol{x}_{j+1}$ of Markov chain.

Suppose the Markov chain is in j-th time step $\boldsymbol{x}_j$, then generate a new point $\boldsymbol{\varepsilon}$ from a chosen PDF, $f^*(\boldsymbol{\varepsilon}|\boldsymbol{x}_j)$, called the 'proposal distribution', which is a PDF for $\boldsymbol{\varepsilon}$ depending on $\boldsymbol{x}_j$. Calculate the ratio $r = f(\boldsymbol{\varepsilon}|F)/f(\boldsymbol{x}_j|F)$. If $r > 1$, then $\boldsymbol{\varepsilon}$ is accepted as the $(j+1)$-th state, that is $\boldsymbol{x}_{j+1} = \boldsymbol{\varepsilon}$, otherwise $\boldsymbol{\varepsilon}$ is accepted as the $(j+1)$-th state with probability r and $\boldsymbol{x}_j$ is accepted as the $(j+1)$-th state with the remaining probability $1-r$, i.e., $\boldsymbol{x}_{j+1} = \boldsymbol{x}_j$, where a repeated state is obtained.

(3) Repeat step (2) until M Markov chain states, $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$, are obtained.

3. Markov chain importance sampling method

The proposed method is to modify the original Markov chain Metropolis algorithm to establish an adaptive importance sampling procedure. It takes the proposal distribution of Markov chain as the ISD, and the Markov chain time step states are selected as the sampling centers, thus it can properly combine the advantages of both the Markov chain Metropolis algorithm and the importance sampling method. The Markov chain states can adaptively populate the importance failure region, thus the importance sampling centered on these states owns high efficiency in exploring the important failure region.

The procedure of the proposed method is illustrated firstly. It stems form the original Markov chain algorithm, but several mod-