Structural Safety 63 (2016) 71–80

Contents lists available at ScienceDirect

Structural Safety

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

A new maximum entropy-based importance sampling for reliability analysis

Hongzhe Dai [a,*], Hao Zhang [b], Wei Wang [a]

[a] School of Civil Engineering, Harbin Institute of Technology, Harbin 150090, China
[b] School of Civil Engineering, The University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 12 May 2015
Received in revised form 25 January 2016
Accepted 7 August 2016
Available online 24 August 2016

Keywords:
Density estimation
Importance sampling
Markov chain simulation
Reliability
Maximum-entropy

ABSTRACT

Importance sampling can be highly efficient if a good importance sampling density is constructed. Although the parametric sampling densities centered on the design points are often good choices, the determination of the design points can be a difficult and inefficient task itself, especially for problems with multiple design points, or highly nonlinear limit state functions. This paper introduces a nonparametric importance sampling method based on the Markov chain simulation and maximum-entropy density estimation (MEDE). In the proposed method, Markov chain simulation is utilized to generate samples that distribute asymptotically to the optimal importance sampling density. A nonparametric estimation of the optimal importance sampling density is then obtained using the MEDE technique. The conventional MEDE method is difficult for multi-dimensional problems as it needs to solve a set of simultaneous nonlinear integral equations. This paper developed a new MEDE technique for multivariate dataset. The method starts with using histogram to approximate a density. The multi-dimensional histogram is converted into a series of one-dimensional conditional PDFs in each dimension and the density is reconstructed by means of orthogonal expansion. Thus, the solution of MEDE is converted to a set of coefficients of the Legendre polynomials. The new importance sampling method is illustrated and compared with the classical kernel-based importance sampling using a number of numerical and structural examples.

1. Introduction

Importance sampling is a popular variance reduction technique for reliability analysis. The idea is to carry out simulation with sample points that concentrate in the failure region, as only these samples contribute to the failure probability. Consider a $d-$dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with a joint probability density function (PDF) $f(\mathbf{x})$. Let $g(\mathbf{x})$ represent the limit state function defined such that failure occurs when $g(\mathbf{x}) \leqslant 0$. With importance sampling, the failure probability, P_f, is calculated as

$$P_f = \int_{\mathbb{R}^d} \mathbf{1}[g(\mathbf{x}) \leqslant 0] \frac{f(\mathbf{x})}{h(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}, \tag{1}$$

in which $\mathbf{1}[A]$ is the indicator function for event A, having the value 1 if event A occurs and the value 0 otherwise; $h(\,)$ is the importance sampling density function, which is used to draw samples instead of from the actual probability density $f(\mathbf{x})$.

The efficiency and accuracy of importance sampling is critically dependent on the choice of $h(\mathbf{x})$. It has been shown that an optimal importance sampling density function $h_{\text{opt}}(\mathbf{x})$ exists [1]

$$h_{\text{opt}}(\mathbf{x}) = \frac{\mathbf{1}[g(\mathbf{x}) \leq 0] f(\mathbf{x})}{P_f}. \tag{2}$$

If $h_{\text{opt}}(\mathbf{x})$ is used the variance of the estimated P_f becomes zero regardless of the number of simulations. However, the optimal importance sampling function as defined in Eq. (2) generally is unknown as it involves P_f, which is to be determined. A number of techniques have been proposed to construct good importance sampling functions. A common approach is to choose the importance sampling density as a joint Gaussian distribution centered around the design point [2]. Although design points can often characterize the region of most interest, the search for the design points typically involves a constrained optimization problem, and can be a difficult and inefficient task itself, especially for problems with multiple design points, or highly nonlinear limit state functions [3].

Another approach is to use the kernel density estimation technique to construct nonparametric importance sampling densities. In the kernel density estimator, the density function is

* Corresponding author.
E-mail addresses: hzdai@hit.edu.cn (H. Dai), hao.zhang@sydney.edu.au (H. Zhang), wwang@hit.edu.cn (W. Wang).

http://dx.doi.org/10.1016/j.strusafe.2016.08.001
0167-4730/

approximated by a sum of kernel of functions centered at the data points, and a bandwidth associated with the kernel function is used to control the smoothness of the estimated densities. Ang et al. [4] proposed a kernel method to approximate the optimal importance sampling density. The method requires an initial Monte Carlo run to generate the samples in the failure region. Thus, it is not efficient for problems with small failure probabilities. Au and Beck [5] proposed an importance sampling scheme to improve the efficiency of the method in [4] by using the Markov chain simulation to generate the samples that populate in the regions of most interest. In Kurtz and Song [6], the importance sampling density is constructed using Gaussian mixture kernel. The difference between the constructed density and the optimal sampling density was measured by the Kullback–Leibler cross entropy. Various hybrid methods have also been proposed by combining the surrogate model (e.g., Kriging) and the importance sampling to further improve the sampling efficiency [7].

In summary, most of the existing methods for constructing non-parametric importance sampling density relies on the kernel-based density estimation. Since the kernel density has a shape similar to the optimal sampling density, a failure region of almost any shape can be handled. The kernel density method can also be extended to higher dimensions. However, the choice of the bandwidth of the kernel may heavily affect the accuracy of the kernel density estimator, particularly when the number of samples is not very large [8]. The resulting estimate may contain spurious data artifacts if the bandwidth is not chosen appropriately.

Maximum entropy (ME) approach is a widely used density-estimation method [9]. The key idea of the maximum-entropy density estimation (MEDE) is to find the probability density that maximizes its entropy under specified moment constraints. The moment constrained maximum entropy problem yields least biased probability density among all candidate densities that are consistent with available sample data [10–13]. However, the conventional MEDE becomes inefficient when handling multivariate data as the method requires to solve a set of simultaneous nonlinear integral equations, one for each moment constraint. In such cases, the dimensional reduction method can be used in conjunction with ME to estimate the probability of failure [14]. With the dimensional reduction method, a multi-dimensional function is decomposed in terms of the orthogonal component functions involving low-dimensional vector only, thus the maximum entropy works effectively and the curse of dimensionality can be mitigated. More recently, Zhang and Pandey [15] presented a fractional moments-based maximum entropy method for deriving the PDF of structural response, in which the fractional moment of a multi-variate function is obtained by a multiplicative form of dimensional reduction method. By using fractional moments as the constraints of maximum entropy, more statistical information can be obtained from the samples and thereby many limitations of traditional maximum entropy method can be overcome. Ref. [16,17] developed improved algorithms for computing the maximum entropy problem in higher-dimensional domains. The algorithms use multi-dimensional orthogonal polynomial basis in the dual space of Lagrange multipliers in the iterative optimization process.

This paper introduces a new maximum entropy-based importance sampling scheme. The proposed methodology involves the generation of samples that populate the important region by Markov chain simulation, and the construction of importance sampling density by the maximum entropy density estimation method. We developed a novel solution to the multi-dimensional entropy moment problem by converting a multi-dimensional histogram into a set of coefficients of the Legendre polynomials, thus obviating the need for dimension reduction techniques.

The paper first briefly introduces the Markov chain simulation in the context of importance sampling, followed by the description of the classical moment-constrained maximum-entropy density estimation. A new MEDE method based on the expansion of conditional PDF is then presented in Section 4. The procedure of the proposed importance sampling method is summarized in Section 5. Four examples are given to demonstrate the application and efficiency of the proposed method. Comparisons of the new method and the classical kernel-based importance sampling are made.

2. Markov chain simulation

The first step of the proposed importance sampling method is the generation of samples that cover the region of most interest. As seen in Eq. (2), the closed form expression for the optimal importance sampling density function is unknown, thus the direct Monte Carlo sampling procedure cannot be used. In this case, the Markov chain simulation can be employed to construct a Markov chain samples having a PDF which tends asymptotically to the optimal importance sampling function [5]. It should be noted that the adaptive Markov chain simulation (as introduced in [18,19]) may be more efficient than the classical Markov chain simulation. This study employs the classical Markov chain simulation because the main innovation of this paper is the multi-dimensional MEDE technique developed in Section 4. Markov chain simulation is introduced and discussed in many texts. For a complete discussion of the proposed importance sampling method, we briefly introduce the Markov chain simulation in the context of importance sampling.

The classical Markov chain simulation method is based on the random-walk Metropolis algorithm [20]. Let $\pi(\,)$ denote the density to be simulated, i.e., the target distribution density function. Suppose that a sample point $\mathbf{x}_i$ has been sampled. Then a point $\mathbf{y}$ is generated according to a proposal distribution $q(\cdot|\mathbf{x}_i)$, which is used to generate the next point of the Markov chain. Generally, $q(\cdot|\mathbf{x}_i)$ is symmetric with respect to its arguments. The proposal distribution governs the choice of the generated samples and consequently the efficiency of the Metropolis algorithm. The point $\mathbf{y}$ is accepted (i.e., $\mathbf{x}_{i+1} = \mathbf{y}$) with probability

$$\alpha = \min\left(1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x}_i)}\right). \tag{3}$$

The point $\mathbf{y}$ is rejected with the remaining probability $(1-\alpha)$, in this case the state is not updated, i.e., $\mathbf{x}_{i+1} = \mathbf{x}_i$.

Let the optimal importance sampling function in Eq. (2) be the target distribution of the Markov chain. The acceptance probability α is then given by

$$\alpha = \frac{\mathbf{1}[g(\mathbf{y}) \le 0] f(\mathbf{y})}{\mathbf{1}[g(\mathbf{x}_i) \le 0] f(\mathbf{x}_i)}. \tag{4}$$

It can be seen that the evaluation of Eq. (4) only requires the ratio of the target distribution between consecutive states, and P_f is not needed. This implies that the sequence of points $\{\mathbf{x}_i\}$ generated by the above algorithm can populate the important region according to the optimal sampling density function. This simulation procedure can work equally well with problems with very small probability of failure [5]. In this study, the multivariate Gaussian distribution is chosen as the proposal distribution. The first 10% of the samples in the Markov chain are discarded to eliminate the bias due to the initially chosen starting points.

3. Solution to the problem of multi-dimensional moments

The moment-constrained maximum entropy problem yields an estimate of a probability density with the largest entropy (i.e.,