Structural Safety 59 (2016) 42–52

Contents lists available at ScienceDirect

Structural Safety

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

Cross-entropy-based adaptive importance sampling using von Mises-Fisher mixture for high dimensional reliability analysis

Ziqi Wang [a], Junho Song [b,*]

[a] Earthquake Engineering Research & Test Center, Guangzhou University, Guangzhou, China[1]
[b] Department of Civil and Environmental Engineering, Seoul National University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 28 December 2014
Received in revised form 6 November 2015
Accepted 13 November 2015
Available online 23 December 2015

Keywords:
Adaptive sampling
Cross entropy
High dimensional space
Importance sampling
Monte Carlo simulations
Structural reliability
Von Mises-Fisher distribution

ABSTRACT

In order to address challenges in performing importance sampling in a high dimensional space of random variables, the paper develops a cross-entropy-based adaptive importance sampling technique that employs a von Mises-Fisher mixture as the sampling density model. By small-size pre-samplings, the proposed approach first finds a near-optimal sampling density by minimizing the Kullback–Leibler cross entropy between a von Mises-Fisher mixture model and the absolute best importance sampling density. To facilitate the minimization process, updating rules for parameters of the von Mises-Fisher mixture model are derived. Various practical issues associated with the updating rules are discussed and heuristic rules to improve the performance of the importance sampling are introduced. At the stage of final sampling, two slightly different sampling strategies are proposed to provide analysis options. Three numerical examples are investigated to test and demonstrate the proposed importance sampling method. The numerical examples show that the proposed approach, applicable to both component and system reliability problems, has superior performance for high dimensional reliability analysis problems with low failure probabilities.

1. Introduction

An important goal of reliability analysis is to determine the probability that a prescribed performance requirement of a system is satisfied, given that certain parameters associated with the system capacity and/or demand are stochastic. In engineering practice the performance of a system is often estimated by computationally demanding numerical algorithms, thus the limit-state surface in the random variable space may be not only complex but also implicitly defined. Due to the complex topology of limit-state surfaces, analytical solutions of reliability problems either at component or system level are usually unavailable. As a result, approximate reliability methods such as first- and second-order reliability methods (FORM and SORM) [1,2], response surface methods (RSM) [3–11], and various sampling techniques [12–15] have gained wide popularity.

Compared with other reliability methods, sampling-based methods have the benefits of avoiding errors from approximations of limit state surfaces, being insensitive to the complexity of limit-state functions, and being straightforward to implement. However, since many practical structural reliability problems are characterized by low failure probabilities, the brute force Monte Carlo sampling (MCS) scheme based on the original or transformed joint probability density function may become computationally inefficient. As a result the importance sampling (IS) has been developed as an efficient alternative for structural reliability analysis. One widely used IS scheme is to employ a Gaussian joint density whose mean is located at the design point as the IS density. However, this approach becomes inefficient or inaccurate when there are multiple design points or critical regions in the random variable space. Moreover, this approach in general does not work well for high dimensional reliability problems [15,16].

Another attractive IS scheme is the cross-entropy-based adaptive importance sampling (CE-AIS) [17,18]. The CE-AIS obtains a near-optimal IS density by minimizing the Kullback–Leibler cross entropy using pre-samples. The cross entropy is used as a measure of the difference between the absolute best IS density and the current IS density model. Recently, a notable CE-AIS approach [18] was proposed, in which a Gaussian mixture (GM) model was employed as the IS density model to enhance the adaptability to the complex shape of the absolute best IS density of component and system reliability problems. It has been shown that this CE-AIS-GM performs well for reliability problems with dimensions

* Corresponding author. Tel.: +82 2 880 7369; fax: +82 2 873 2684.
E-mail address: junhosong@snu.ac.kr (J. Song).
[1] Formerly, Department of Bridge Engineering, Southwest Jiaotong University, Chengdu, China.

http://dx.doi.org/10.1016/j.strusafe.2015.11.002
0167-4730/

up to 50. However, it was found that CE-AIS-GM practically does not work well for high dimensional problems with dimensions higher than 100. In fact, there are reasons to expect that Gaussian based IS density models would not work properly for high dimensional reliability analysis. This perspective will be discussed in detail in the subsequent sections.

To address challenges in performing importance sampling in high dimensional space of random variables, this paper proposes a CE-AIS technique employing a von Mises-Fisher mixture as the IS density model. After providing a brief review of properties of high dimensional standard normal space, the sampling scheme in polar coordinates is discussed and formulated. Based on the formulation, a CE-AIS method using a von Mises-Fisher mixture (vMFM) is developed and rules for updating parameters of vMFM based on pre-samples are derived. Various practical issues associated with the updating are discussed and heuristic rules are introduced to improve the performance of the proposed importance sampling method. At the stage of final sampling, two slightly different sampling strategies of CE-AIS-vMFM are proposed to offer analysis options. Three numerical examples illustrate the accuracy and efficiency of the proposed CE-AIS-vMFM approaches.

2. Properties of high dimensional standard normal space

Consider an n dimensional standard normal space, which one can achieve for a given structural reliability problem through a proper transformation [2], the distance from the origin in the space, r follows the χ-distribution with n degrees of freedom. If n is relatively large, the probability density function (PDF) of the χ-distribution will be concentrated around $r = \sqrt{n}$. Fig. 1 shows the probability density function (PDF) of $r - \sqrt{n}$ for several selected values of n. It is seen that as n increases, the distribution of r approaches a normal distribution concentrated around $\sqrt{n}$. In fact, if $n \to \infty$, $\chi_n \approx \mathcal{N}(\sqrt{n}, 1/2)$ [16,19], that is, if the dimension n is sufficiently large, the distance from the origin can be approximated by a normal distribution with mean $\sqrt{n}$ and variance 1/2. The aforementioned observation indicates that almost all probability information of a high dimensional standard normal space is captured by a relatively narrow "important ring" region with radius $\sqrt{n} \pm \varepsilon$ where ε is a small perturbation. For example, for $n = 400$, more than 95% of the probability is contained within the ring of width 20 ± 1 and more than 99.99% is contained within the ring of width 20 ± 2. Since $\sqrt{n} \pm \varepsilon$ is relatively large and most realistic structural reliability problems have a reliability index of the order 2–6, the failure domain of most practical structural reliability problems would intersect the important ring. Consequently, for successful application of a reliability method (e.g. first-order or second-order reliability method (FORM/SORM), response surface or sampling) to high dimensional problems, it is critical to capture the intersection between the important ring and the failure domain. In other words, it is important to characterize a subset of the important ring for effective reliability analysis in high dimensional problems.

There is no absolute definition of *high* dimension. However, for this study it seems reasonable to define high dimension in terms of the coefficient of variation (c.o.v.) of the χ-distribution, which is expressed as

$$c_v(n) = \frac{\Gamma\left(\frac{n}{2}\right)\sqrt{\frac{n}{2} - \frac{\Gamma\left(\frac{1+n}{2}\right)^2}{\Gamma\left(\frac{n}{2}\right)^2}}}{\Gamma\left(\frac{1+n}{2}\right)} \tag{1}$$

where $\Gamma(\cdot)$ is the gamma function. The variation of c_v with dimensions is illustrated in Fig. 2. Now high dimension can be defined as a dimension n which satisfies

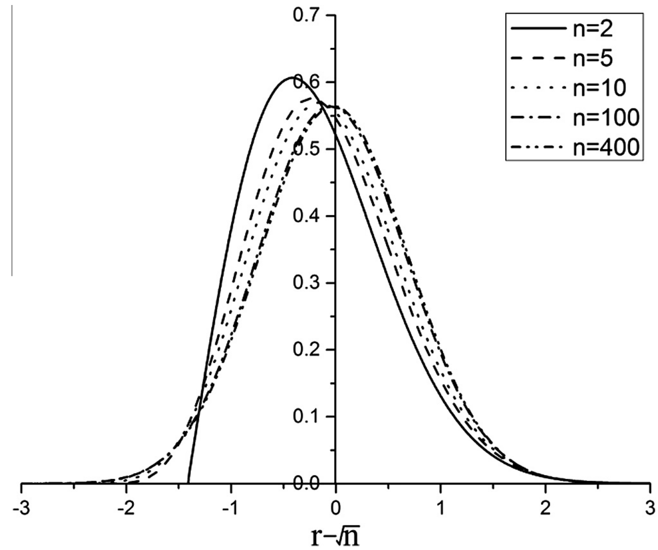

Fig. 1. PDF of χ-distribution for various dimensions.

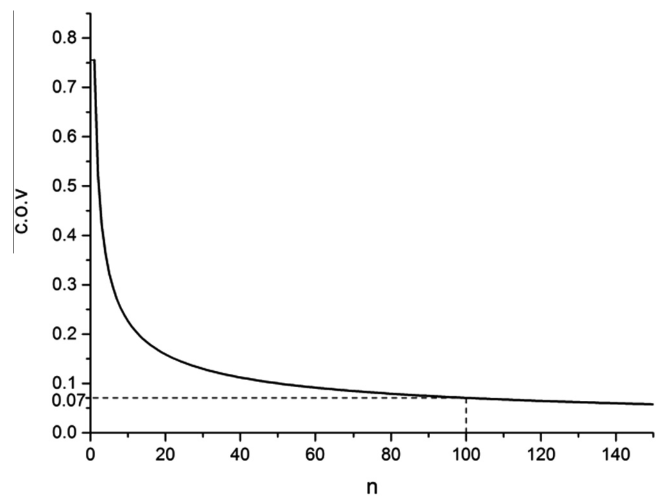

Fig. 2. Variation of c_v with dimension.

$$c_v(n) \leqslant Tol \tag{2}$$

where Tol is a prescribed tolerance on the c.o.v. For example, if $Tol = 0.1$, high dimension is defined as $n \geqslant 51$; if $Tol = 0.05$, the definition is $n \geqslant 201$. In this paper, we regard high dimension as $n \geqslant 100$, which corresponds to the tolerance of c.o.v. = 0.07.

3. Importance sampling for high dimensional reliability analysis

The failure probability of a reliability problem defined in the n-dimensional standard normal space can be described in terms of the distance from the origin, i.e.

$$P_f = \int_0^\infty \theta(r) f_\chi(r) dr \tag{3}$$

in which $f_\chi(r)$ is the PDF of the χ-distribution with n degrees of freedom representing the random distance from the origin, and $\theta(r)$ is the failure ratio of the hypersphere surface area with radius r. The failure ratio $\theta(r)$ represents the percentage of the surface area of the hypersphere with radius r that belongs to the failure domain. Note that in a standard normal space points on a hypersphere is