Structural Safety 76 (2019) 15–27

Contents lists available at ScienceDirect

Structural Safety

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

Cross entropy-based importance sampling using Gaussian densities revisited

Sebastian Geyer*, Iason Papaioannou, Daniel Straub

Engineering Risk Analysis Group, Technische Universität München, Arcisstraße 21, 80333 München, Germany

ARTICLE INFO

Keywords:
Reliability analysis
Simulation
Importance sampling
Cross entropy method
Gaussian mixture
Expectation maximization
Sequential importance sampling

ABSTRACT

The computation of the probability of a rare (failure) event is a common task in structural reliability analysis. In most applications, the numerical model defining the rare event is nonlinear and the resulting failure domain often multimodal. One strategy for estimating the probability of failure in this context is the importance sampling method. The efficiency of importance sampling depends on the choice of the importance sampling density. A near-optimal sampling density can be found through application of the cross entropy method. The cross entropy method is an adaptive sampling approach that determines the sampling density through minimizing the Kullback-Leibler divergence between the theoretically optimal importance sampling density and a chosen parametric family of distributions. In this paper, we investigate the suitability of the multivariate normal distribution and the Gaussian mixture model as importance sampling densities within the cross entropy method. Moreover, we compare the performance of the cross entropy method to sequential importance sampling, another recently proposed adaptive sampling approach, which uses the Gaussian mixture distribution as a proposal distribution within a Markov Chain Monte Carlo algorithm. For the parameter updating of the Gaussian mixture within the cross entropy method, we propose a modified version of the expectation-maximization algorithm that works with weighted samples. To estimate the number of distributions in the mixture, the density-based spatial clustering of applications with noise (DBSCAN) algorithm is adapted to the use of weighted samples. We compare the performance of the different methods in several examples, including component reliability problems, system reliability problems and reliability in varying dimensions. The results show that the cross entropy method using a single Gaussian outperforms the cross entropy method using Gaussian mixture and that both distribution types are not suitable for high dimensional reliability problems.

1. Introduction

In structural reliability, the goal is to assess the effects of uncertain input variables on the performance of an engineering system. The main objective is to evaluate the probability of unsatisfactory performance of the system, the probability of failure P_F. Let $\boldsymbol{X}$ be a random vector with outcome space $\boldsymbol{x} \in \mathbb{R}^n$ that collects all uncertain input variables. The performance of a system can be assessed by means of the limit state function, denoted by $g(\boldsymbol{x})$. If the limit state function gives a value smaller or equal to zero the system fails, while otherwise the system has satisfactory performance. Hence, the event defining failure of the system is given by

$$F = \{\boldsymbol{x} \in \mathbb{R}^n : g(\boldsymbol{x}) \leqslant 0\} \tag{1}$$

The probability of failure is defined through the following integral:

$$P_F = \int_{\mathbb{R}^n} I(\boldsymbol{x}) \cdot f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{2}$$

In this definition, $f(\boldsymbol{x})$ is the joint probability density function (PDF) of $\boldsymbol{X}$ and $I(\boldsymbol{x})$ is an indicator function defined as

$$I(\boldsymbol{x}) = \begin{cases} 1 \text{ if } g(\boldsymbol{x}) \leqslant 0 \\ 0 \text{ else} \end{cases} \tag{3}$$

The function $g(\boldsymbol{x})$ depends on the outcome of an engineering model, and hence the probability of failure cannot be evaluated analytically. Therefore, the failure probability needs to be approximated by means of an efficient method that minimizes the number of model evaluations [1]. Several methods have been developed, including approximation methods such as the first/second order reliability method (FORM/SORM) [2,3] and sampling-based methods such as the Monte Carlo simulation (MCS) and its adaptive variants (e.g. [4–10]). The main advantage of sampling-based strategies is their robustness with respect to the complexity of the limit state function. Furthermore, they are asymptotically exact and, in contrast to approximation methods, information about the quality of the obtained result is more easily available.

* Corresponding author.
E-mail address: s.geyer@tum.de (S. Geyer).

https://doi.org/10.1016/j.strusafe.2018.07.001
Received 19 September 2017; Received in revised form 7 April 2018; Accepted 1 July 2018

Sampling methods are based on MCS; however, MCS is especially inefficient in estimating small failure probabilities, which is typically the goal in reliability assessment of engineering systems. The variance of the MCS estimate is inversely proportional to the target failure probability, which leads to prohibitively large sample sizes for obtaining sufficiently accurate estimates. The aim of advanced sampling-based methods is to enhance the efficiency of MCS through reducing the required sample size while keeping the variance of the estimate low.

A standard variant of MCS is the importance sampling (IS) method. IS aims at decreasing the variance of the MCS probability estimate by sampling from an alternative sampling density, the so-called IS density $h(\boldsymbol{x})$. Eq. (2) can be modified as follows:

$$P_F = \int_{\mathbb{R}^n} \frac{I(\boldsymbol{x})\cdot f(\boldsymbol{x})}{h(\boldsymbol{x})}\cdot h(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{4}$$

Provided that the support of $h(\boldsymbol{x})$ contains the failure domain, this modification does not alter the value of the integral. The IS estimate of P_F is given as follows:

$$\hat{P}_F = \frac{1}{n_s}\sum_{i=1}^{n_s} I(\boldsymbol{x}_i)\cdot\frac{f(\boldsymbol{x}_i)}{h(\boldsymbol{x}_i)}, \tag{5}$$

wherein the samples $\{\boldsymbol{x}_i,\ i = 1,\ ...,n_s\}$ are distributed according to $h(\boldsymbol{x})$. The efficiency of IS depends on the choice of the IS density. A near-optimal IS density can be found through application of the cross entropy (CE) method [11]. The CE method is an adaptive sampling approach that determines the sampling density through minimizing the Kullback-Leibler (KL) divergence between the theoretically optimal IS density and a chosen parametric family of probability distributions. Typically, the multivariate normal distribution is chosen as parametric distribution family, while recently the Gaussian mixture (GM) has been proposed for handling multimodal failure domains [7,12]. In this study we investigate the performance of these two distribution types within the CE method and propose a new updating scheme for estimating the parameters of the GM model through a modified expectation-maximization (EM) algorithm. Moreover, another recently proposed approach, called sequential importance sampling (SIS) [6], which also makes use of the Gaussian mixture, is discussed and its performance is compared to the one of the CE method.

The structure of the paper is as follows. First, the CE method is described and a general algorithm for its implementation is presented. Afterwards, the multivariate Gaussian and Gaussian mixture distribution types are introduced and their integration in the CE method is described; for integrating the Gaussian mixture model, a novel EM algorithm is proposed. Subsequently, a brief overview of SIS is given. The performance of the different methods is demonstrated by means of numerical examples. Finally, the results are summarized and the conclusions are presented.

2. Cross entropy method

For an efficient and reliable estimation of the failure probability with IS, the sampling density needs to be chosen carefully. In fact, there exists an optimal IS density, whose PDF is [13]:

$$p^*(\boldsymbol{x}) = \frac{I(\boldsymbol{x})\cdot f(\boldsymbol{x})}{\int_{\mathbb{R}^n} I(\boldsymbol{x})\cdot f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}} \tag{6}$$

Eq. (6) is the density of the random variables censored at the failure domain; its normalizing constant is the target probability of failure. The IS density of Eq. (6) leads to a variance of the probability estimate of zero. That is, a single sample of this density would lead to the exact value of the probability of failure. However, this density is not applicable in practice, as it requires a priori knowledge of the failure domain and the target failure probability.

Even though Eq. (6) cannot be used directly, one can use samples at the failure domain to identify a near-optimal IS density through fitting a distribution model. The CE method identifies the parameters of a chosen distribution model through minimizing the KL divergence between the sought sampling density and the optimal IS density of Eq. (6). The KL divergence is a measure of the difference between two PDFs, and is defined as [14]

$$D(p_1(\boldsymbol{x}),\, p_2(\boldsymbol{x})) = \mathrm{E}_{p_1}\left[\ln\left(\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\right)\right] = \int_{\mathbb{R}^n} p_1(\boldsymbol{x})\cdot\ln\left(\frac{p_1(\boldsymbol{x})}{p_2(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x}, \tag{7}$$

where $\mathrm{E}_{p_1}[\cdot]$ denotes the expectation with respect to $p_1(\boldsymbol{x})$. Eq. (7) is also known as the relative cross entropy [15]. In the following the terms cross entropy and KL divergence are used interchangeably.

The expression in Eq. (7) can be divided into two parts:

$$D(p_1(\boldsymbol{x}),\, p_2(\boldsymbol{x})) = \int_{\mathbb{R}^n} p_1(\boldsymbol{x})\cdot\ln(p_1(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} - \int_{\mathbb{R}^n} p_1(\boldsymbol{x})\cdot\ln(p_2(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} \tag{8}$$

Replacing $p_1(\boldsymbol{x})$ with the optimal (but unknown) IS density $p^*(\boldsymbol{x})$ and $p_2(\boldsymbol{x})$ with a parametric IS density $h(\boldsymbol{x};\boldsymbol{v})$, where $\boldsymbol{v}$ is the parameter vector, Eq. (8) can be rewritten to describe the cross entropy between these two PDFs:

$$D(p^*(\boldsymbol{x}),\, h(\boldsymbol{x};\boldsymbol{v})) = \int_{\mathbb{R}^n} p^*(\boldsymbol{x})\cdot\ln(p^*(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} - \int_{\mathbb{R}^n} p^*(\boldsymbol{x})\cdot\ln(h(\boldsymbol{x};\boldsymbol{v}))\,\mathrm{d}\boldsymbol{x} \tag{9}$$

As the parametric IS density appears in the second term only and $p^*(\boldsymbol{x})$ is invariant, a minimum of the cross entropy can be found by minimizing only the second part. Substituting Eq. (6) for $p^*(\boldsymbol{x})$ in Eq. (9) results in the following optimization problem [7]:

$$\arg\min_{\boldsymbol{v}} D(p^*(\boldsymbol{x}),\, h(\boldsymbol{x};\boldsymbol{v})) = \arg\max_{\boldsymbol{v}} \int_{\mathbb{R}^n} I(\boldsymbol{x})\cdot f(\boldsymbol{x})\cdot\ln(h(\boldsymbol{x};\boldsymbol{v}))\,\mathrm{d}\boldsymbol{x} \tag{10}$$

The IS density $h(\boldsymbol{x};\boldsymbol{v})$ found by the minimization of $D(p^*(\boldsymbol{x}),\, h(\boldsymbol{x};\boldsymbol{v}))$ is termed *near-optimal IS density*. For the efficient evaluation of Eq. (10), an alternative sampling density $h(\boldsymbol{x};\boldsymbol{w})$ is defined, which is of the same type as the near-optimal sampling density $h(\boldsymbol{x};\boldsymbol{v})$. The optimization program is modified accordingly:

$$\begin{aligned}\arg\min_{\boldsymbol{v}} D(p^*(\boldsymbol{x}),\, h(\boldsymbol{x};\boldsymbol{v})) &= \arg\max_{\boldsymbol{v}} \int_{\mathbb{R}^n} I(\boldsymbol{x})\cdot\ln(h(\boldsymbol{x};\boldsymbol{v}))\cdot W(\boldsymbol{x};\boldsymbol{w})\cdot h(\boldsymbol{x};\boldsymbol{w})\,\mathrm{d}\boldsymbol{x} \\ &= \arg\max_{\boldsymbol{v}} \mathrm{E}_{\boldsymbol{w}}[I(\boldsymbol{x})\cdot\ln(h(\boldsymbol{x};\boldsymbol{v}))\cdot W(\boldsymbol{x};\boldsymbol{w})]\end{aligned} \tag{11}$$

In this expression, $\mathrm{E}_{\boldsymbol{w}}[\cdot]$ denotes the mathematical expectation with respect to the alternative sampling density $h(\boldsymbol{x};\boldsymbol{w})$, while $W(\boldsymbol{x};\boldsymbol{w})$ is the likelihood ratio of the original sampling PDF to the alternative sampling PDF:

$$W(\boldsymbol{x};\boldsymbol{w}) = \frac{f(\boldsymbol{x})}{h(\boldsymbol{x};\boldsymbol{w})} \tag{12}$$

The expectation in Eq. (11) is approximated via IS. With referring to the i-th sample drawn from the density $h(\boldsymbol{x};\boldsymbol{w})$, $i = 1,\ ...,n_s$, the solution to Eq. (11) is approximated by

$$\arg\min_{\boldsymbol{v}} D(p^*(\boldsymbol{x}),\, h(\boldsymbol{x};\boldsymbol{v})) \approx \arg\max_{\boldsymbol{v}} \frac{1}{n_s}\cdot\sum_{i=1}^{n_s} I(\boldsymbol{x}_i)\cdot\ln(h(\boldsymbol{x}_i;\boldsymbol{v}))\cdot W(\boldsymbol{x}_i;\boldsymbol{w}) \tag{13}$$

If Eq. (13) has a global maximum, which is typically the case in structural reliability [4], this maximum can be found by taking the gradient with respect to $\boldsymbol{v}$ and setting the result to zero.

According to Eq. (13), obtaining a reliable estimate of the KL divergence requires that a substantial number of samples from $h(\boldsymbol{x};\boldsymbol{w})$ fall in the failure domain. The CE method solves this problem through introducing a series of intermediate failure domains that gradually approach the target failure domain. In this way, the CE optimization problem is solved for the optimal IS density of each intermediate failure domain using samples from the fitted parametric density obtained at the previous step. The l-th intermediate failure domain and corresponding indicator function are defined as follows: