Journal of Computational Physics 230 (2011) 8683–8697

Contents lists available at SciVerse ScienceDirect

Journal of Computational Physics

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

An efficient surrogate-based method for computing rare failure probability

Jing Li [a], Jinglai Li [b], Dongbin Xiu [a,*]

[a] *Department of Mathematics, Purdue University, West Lafayette, IN 47907, United States*
[b] *Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States*

ARTICLE INFO

Article history:
Received 31 December 2010
Received in revised form 13 July 2011
Accepted 11 August 2011
Available online 28 August 2011

Keywords:
Rare events
Failure probability
Importance sampling
Cross-entropy

ABSTRACT

In this paper, we present an efficient numerical method for evaluating rare failure probability. The method is based on a recently developed surrogate-based method from Li and Xiu [J. Li, D. Xiu, Evaluation of failure probability via surrogate models, J. Comput. Phys. 229 (2010) 8966–8980] for failure probability computation. The method by Li and Xiu is of hybrid nature, in the sense that samples of both the surrogate model and the true physical model are used, and its efficiency gain relies on using only very few samples of the true model. Here we extend the capability of the method to rare probability computation by using the idea of importance sampling (IS). In particular, we employ cross-entropy (CE) method, which is an effective method to determine the biasing distribution in IS. We demonstrate that, by combining with the CE method, a surrogate-based IS algorithm can be constructed and is highly efficient for rare failure probability computation—it incurs much reduced simulation efforts compared to the traditional CE-IS method. In many cases, the new method is capable of capturing failure probability as small as $10^{-12} \sim 10^{-6}$ with only several hundreds samples.

1. Introduction

Uncertainties in specifying material properties, geometric parameters, boundary conditions and applied loadings are unavoidable in describing real-life engineering systems. Traditionally, this has been catered for in an *ad hoc* way through the use of safety factors at the design stage. Such an approach is becoming less satisfactory in today's competitive design environment, for example, in minimum weight design of aircraft structures. Therefore, accurate evaluation of failure probability of a given system is becoming increasingly important and a fundamental problem in many fields such as risk management, structural design, reliability based optimization, etc.

Since in most cases the function separating the safe and failure domains is not known explicitly and can be highly irregular in high dimensional spaces, the standard integration rules such as Gauss quadrature, cubature, sparse grids, etc., are not directly applicable. The most commonly used method is Monte Carlo simulation (MCS), which requires one to simulate, also known as to sample, the underlying system repetitively. Though straightforward to implement, MCS can be highly time-consuming when the underlying system is complex, for each sample requires a full-scale numerical simulation of the system. To reduce the computational effort, many alternative non-sampling based methods have been developed, such as FORM/SORM (first-order/second-order reliability method) [13,9,6,12,20,27,26], RSM (response surface method) [29,8,3,21,10,11,19], etc. These methods usually incur much less simulation cost, compared to MCS, at the expense of reduced accuracy.

For systems with rare failure probability, which in this paper is defined as failure probability less than 10^{-5}, the problem becomes drastically more difficult. To this end, one needs to employ methods with sufficiently high accuracy so that the rare

* Corresponding author.
E-mail addresses: li176@math.purdue.edu (J. Li), jinglai@mit.edu (J. Li), dxiu@purdue.edu (D. Xiu).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2011.08.008

failure probability in the small tails of the distribution can be captured. And to this date none of the non-sampling methods can be highly effective in a general setting, due to their lack of high accuracy. Probably the most reliable method is based on MCS. To this end, the computational cost becomes much higher. As a rule-of-thumb, one usually needs about 10 samples in the failure domain to reliably estimate the failure probability. For rare probability this translates to prohibitively large number of samples. For example, for probability of $O(10^{-6})$, it is not uncommon to use $O(10^7)$, or even $O(10^8)$, number of samples for good accuracy.

To reduce the number of samples, one of the most widely used methods is importance sampling (IS). (See, for example, [22,28,4,25]). In IS, one seeks to sample the random variables from a different distribution, called *biasing distribution*, rather than the original one. The biasing distribution is constructed in such a way that more samples will land in the failure domain and thus results in (much) less total number of samples. The use of the biasing distribution, essentially a change of measure approach, is incorporated in the sampling estimate by adjusting the weight of each sample to ensure that the final estimate is unbiased. A successfully constructed IS method can significantly increase the efficiency of MCS. For the aforementioned example, for probability of $O(10^{-6})$, a carefully constructed IS method can reduce the total number of samples by several orders, e.g., $O(10^4)$, a drastic reduction from the $O(10^7)$ samples by the brute force MCS.

For practical engineering systems, computer simulations can be extremely time consuming. In many cases, one can only afford very limited number of simulations—nothing beyond a few hundreds. In this case, even the most effective IS method is not applicable. And one often has to resort to some highly problem dependent and/or *ad hoc* approaches to estimate the rare probability and then use safety factor to mitigate the impacts caused by the inaccuracy. It should be noted that the idea of combining IS with techniques such as FORM/SORM has been pursued and resulted in several interesting strategies. Also there are other types of approaches that do not employ IS, e.g., sequential Monte Carlo method ([15]), etc. These methods have their own specific strength and limitations, and will not be discussed in this paper. For detailed discussions, see, for example, [7,20,27,26,15].

The purpose of this paper is to present a new algorithm for computing rare probability, with significantly enhanced efficiency so that it can capture rare probability of less than 10^{-5} with (in many cases) a few hundreds of samples. A key feature of the method is that it is in a general setting and does not require approximations, transformations, or manipulations of the underlying systems. The method relies on a recently developed method by Li and Xiu [17], where a fundamental flaw in the traditional RSM was identified and an improvement – a surrogate-based hybrid method – was proposed. The method utilizes samples of the response surface, which hereafter will be referred to as *surrogate*, in the majority of the probability space and samples of the real system only in the region surrounding the failure mode. It achieves high accuracy in the failure probability estimation while incurring much reduced number of samples of the real system. In this paper the hybrid method from [17] is extended to the case of rare failure probability by incorporating the idea of importance sampling. In particular, the cross-entropy (CE) method for IS is adopted. The CE method ([24,5]) is a relatively new Monte Carlo technique for both estimation and optimization. In the estimation setting, the CE method provides an adaptive way to find a good biasing distribution for quite general problems. The biasing distribution is obtained by an optimization procedure that minimizes its distance, measured by cross-entropy between two distributions, from the *optimal* biasing distribution, which exists in theory but is unavailable in practice. Furthermore, an adaptive multilevel iterative algorithm is available to effectively compute the biasing distribution. In this paper we present a new algorithm based by combining the ideas of both the CE method and the hybrid method. The surrogate model is utilized in both the CE optimization step and the final IS integral evaluation. And the result is a highly efficient algorithm where the simulation cost is significantly reduced further, compared to the direct CE-IS method. In many test cases, the new algorithm is able to resolve failure probability as low as 10^{-6} with only several hundreds samples.

It should be emphasized that the purpose of this paper is *not* to compare different IS methods or other types of methods for rare probability. This is impossible to do owing to the large amount of existing methods. Rather, the paper presents an approach of enhancing certain IS strategies, by using the surrogate-based hybrid idea from [17], and demonstrates that the new method is rigorous, efficient, and more importantly, easy to implement. We restrict our discussion to the CE method, and leave the potential extension to other IS strategies to future research.

Throughout this paper, we assume that a surrogate model is available to use. The surrogate model can be constructed by simulations or given by physics laws or literature. Therefore we do not count the construction of the surrogate as part of the computational cost. Also, the proposed method does not seek to improve the accuracy of the surrogate, which is impossible to do if the surrogate is obtained from literature or physics laws. It should also be noted that improving the accuracy of the surrogate may not be useful at all [17]. In fact, the method does not require the surrogate model to be of high accuracy.

The rest of this paper is organized as the following. After presenting the formulation of rare failure probability computation in Section 2, we briefly review the key ingredients of the present method. These include importance sampling, cross-entropy method, and the hybrid method of [17]. The details of the new method are presented in Section 4, where two numerical algorithms are presented. Numerical examples are presented in Section 5 to demonstrate the effectiveness of the new algorithms.