Simulation Modelling Practice and Theory 19 (2011) 2244–2255

Contents lists available at SciVerse ScienceDirect

Simulation Modelling Practice and Theory

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

Influence of input PDF parameters of a model on a failure probability estimation

Jérôme Morio *

Onera - The French Aerospace Lab, BP 80100, FR-91123 Palaiseau Cedex, France

ARTICLE INFO

Article history:
Received 14 February 2011
Received in revised form 5 August 2011
Accepted 6 August 2011
Available online 9 September 2011

Keywords:
Sensitivity analysis
Rare event simulation
Monte Carlo methods
Sobol indices
Importance sampling
Importance splitting

ABSTRACT

Critical probability estimation is of major interest in safety and reliability applications. In this article, we focus on a black-box model with multidimensional random input X and one random output Y. We consider the estimation of probability P that Y exceeds a threshold S. We assume that the random input X follows a multidimensional parametric density with parameters δ and thus the probability P will depend on the values of δ. In this paper, we analyze the sensitivity of the critical probability P to the model parameters δ. We propose a methodology that estimates Sobol indices with low computation cost. This strategy enables us to determine which statistical parameters have a great influence on the value of the probability and require a valuable determination. The last part of this article applies the proposed technique on a realistic case of missile collateral damage estimation.

1. Introduction

Simulation and computing take a larger part in the dimensioning of real systems. In this article, we focus on rare failure probability estimation of a black-box model. This estimation is performed with Monte Carlo simulations or with more adapted techniques like importance sampling or importance splitting. These algorithms assume that the input PDF (probability density function) parameters (such as the mean and the variance in the case of Gaussian input density) are perfectly known and well-determined. It is of course not always the case in realistic situations. In this article, we propose to analyze the influence of the uncertainty of input density parameters on a failure probability estimation. It is an original and important issue in reliability and safety since it is often of major interest to determine which input density parameters have to be well estimated in order to not bias the probability estimation. Similar approaches have been proposed in the case of local sensitivity method with importance splitting [1–3], extreme value theory [4,5], fault tree [6,7] or probability bounding [8,9].

For that purpose, we propose a method to estimate Sobol indices of the input PDF parameters that characterizes their influence on the rare event probability. Even if rare event estimation and global sensitivity method are generally well-known, their combined use in this article is not very developed in the scientific literature [10,11]. The main difficulty of sensitivity analysis methods is the high required number of samples generated with the simulation code in order to obtain an accurate estimation. The approach proposed in this article requires some samples to estimate the failure probability for a set of input density parameters. Then, the sensitivity analysis does not need the generation of new samples which is another innovative aspect of this article.

In this article, we describe the problem statement and propose a three stage methodology to analyze the sensitivity of a rare probability estimation to the uncertainty of input density parameters. Sobol indices of each input density parameter are

* Tel.: +33 1 80 38 66 54.
E-mail address: jerome.morio@onera.fr

1569-190X/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.simpat.2011.08.003

then determined to rank their different influences. The last part of the article concerns the application of this approach on missile collateral damage probability estimation.

2. Problem statement

In this section, we propose to analyze the general context of the proposed study. We consider a black-box system ϕ modeled by a continuous scalar function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$. The input of this function is a d-dimensional random variable $X = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ that follows a probability density function f and the output is defined by the variable Y such as $Y = \phi(X)$. Each random variable $X^{(i)}$ follows a PDF $f_i(\delta_i)$ with δ_i, a set of PDF parameters. The PDF parameter δ_i is for instance the variance of a Gaussian PDF. Without any lack of generality, we will suppose to simplify the notation in the following that each input PDF only depends on one parameter.

In this paper, we focus on the estimation of a failure probability $P(\phi(X) > S) = P$ with S a threshold. A simple way to estimate this probability is to consider Monte Carlo methods [12–16]. For that purpose, one generates independent and identically distributed samples $X_1, \ldots, X_N$ from the PDF f and then estimates the probability with

$$P^{MC} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(X_i) > S} \tag{1}$$

where $\mathbf{1}_{\phi(X_i)>S}$ is equal to 1 if $\phi(X_i) > S$ and 0 otherwise. When S has a high value, the probability is low and cannot be accurately estimated with Monte Carlo. The estimate relative deviation is indeed too important in this case. Alternative methods such as importance sampling (IS) method [17–25] or importance splitting (ISp) [26–29] can be used.

In many situations, the PDF model and its parameters are uncertain but are assumed constant to simplify the problem. This situation is of course not realistic. In this paper, we assume that the PDF of parameter δ_i is defined by g_i. The PDF g_i can be for instance a uniform random variable on the interval $[\langle \delta_i \rangle - c_i, \langle \delta_i \rangle + c_i]$ where $\langle \delta_i \rangle$ is the mean parameter of δ_i and where c_i is experimentally determined or also a Gaussian PDF with mean parameter $\langle \delta_i \rangle$ and a variance parameter v_i. This uncertainty has an influence on the probability estimation and consequently the probability P depends on the value of $\delta_1, \delta_2, \ldots, \delta_d$.

Nevertheless, in order to estimate Sobol indices of $\delta_1, \delta_2, \ldots, \delta_d$, all these parameters have to be independent. It is indeed a sine qua none condition for Sobol indice estimation. This is of course a limit of the proposed approach since it is not always true for every kind of densities. As in every application of Sobol indices, it is thus necessary to be careful that this independence assumption is valid, otherwise Sobol indice estimations are not valuable. Contrary to the usual application of Sobol indices, the $X^{(i)}$ components are not required to be independent but have to be described by PDF with independent parameters.

In this paper, we propose a methodology to study the influence of $(\delta_1, \delta_2, \ldots, \delta_d)$ on the probability P with a low computation cost.

3. Sensitivity analysis of model parameters on a failure probability

It consists in analyzing the influence of the input PDF parameters δ_i on the value of the probability P. The proposed strategy is the following:

- Determine an efficient sampling density h to estimate P for a fixed value of $(\delta_1, \delta_2, \ldots, \delta_d)$. One can notably choose that $(\delta_1, \delta_2, \ldots, \delta_d) = (\langle \delta_1 \rangle, \langle \delta_2 \rangle, \ldots, \langle \delta_d \rangle)$. h_{opt} is theoretically defined with:

$$h_{opt} = \frac{\mathbf{1}_{\phi(X)>S} f(X)}{P(\langle \delta_1 \rangle, \langle \delta_2 \rangle, \ldots, \langle \delta_d \rangle)} \tag{2}$$

  The PDF h_{opt} cannot be derived since it depends on the unknown probability P. Nevertheless rare event methods enable to generate samples according to efficient sampling PDF h that approximates the PDF h_{opt}. The more h is close to h_{opt}, the more efficient is the rare event algorithm. At the end of this stage, we can obtain a set of N samples $X_1, \ldots, X_N$ generated from h and estimate the probability $\widehat{P}(\langle \delta_1 \rangle, \langle \delta_2 \rangle, \ldots, \langle \delta_d \rangle)$.
- Estimate the probability depending on the value of the model parameters δ_i with importance sampling thanks to the following equation:

$$\widehat{P}(\delta_1, \ldots, \delta_d) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(X_i)>S} \frac{f_{\delta_1, \ldots, \delta_d}(X_i)}{h(X_i)} \tag{3}$$

  where the term $f_{\delta_1, \ldots, \delta_d}$ means that f depends on $\delta_1, \ldots, \delta_d$. One has thus determined a direct link between probability estimate $\widehat{P}$ and model parameters $\delta_1, \ldots, \delta_d$.
- Make a sensitivity analysis with, for instance, Sobol indices to determine which density parameters are the most influential on the value of $\widehat{P}$. If the variation of a model parameter δ_i gives a large variation of the value $\widehat{P}$, then $\widehat{P}$ is very sensitive to δ_i. A ranking of the different input density parameters can then be obtained and one can determine which input density parameters have to be accurately estimated to have confidence in the model.

These stages are analyzed in detail in the following of the article.