Simulation Modelling Practice and Theory 49 (2014) 287–304

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

A survey of rare event simulation methods for static input–output models

Jérôme Morio [a,*], Mathieu Balesdent [b], Damien Jacquemart [b,c], Christelle Vergé [b,d,e]

[a] Onera – The French Aerospace Lab, BP 74025, 31055 Toulouse Cedex, France
[b] Onera – The French Aerospace Lab, BP 80100, 91123 Palaiseau Cedex, France
[c] INRIA Rennes, ASPI Applications of Interacting Particle Systems to Statistics, Campus de Beaulieu, 35042 Rennes, France
[d] INRIA Bordeaux, 351 cours de la Libération, 33405 Talence Cedex, France
[e] CNES, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France

ARTICLE INFO

Article history:
Received 28 August 2013
Received in revised form 12 September 2014
Accepted 13 October 2014

Keywords:
Monte-Carlo methods
Rare event
Input–output model
Simulation

ABSTRACT

Crude Monte-Carlo or quasi Monte-Carlo methods are well suited to characterize events of which associated probabilities are not too low with respect to the simulation budget. For very seldom observed events, such as the collision probability between two aircraft in airspace, these approaches do not lead to accurate results. Indeed, the number of available samples is often insufficient to estimate such low probabilities (at least 10^6 samples are needed to estimate a probability of order 10^{-4} with 10% relative error with Monte-Carlo simulations). In this article, one reviewed different appropriate techniques to estimate rare event probabilities that require a fewer number of samples. These methods can be divided into four main categories: parameterization techniques of probability density function tails, simulation techniques such as importance sampling or importance splitting, geometric methods to approximate input failure space and finally, surrogate modeling. Each technique is detailed, its advantages and drawbacks are described and a synthesis that aims at giving some clues to the following question is given: "which technique to use for which problem?".

1. Introduction

Rare event estimation has become a large area of research in the reliability engineering and system safety domains. A significant number of methods has been proposed to reduce the computation burden for the estimation of rare events from sampling to extreme value theory. However it is often difficult to determine which algorithm is the most adapted to a given problem. Moreover, the existing survey articles on rare events are often focused on specific algorithms [1–3]. The novelties of this article are thus to provide a broad view of the current available techniques to estimate rare event probabilities described with a unified notation and to provide some clues to answer this question: which rare event technique is the most adapted to a given situation?

The general problem considered in this article is analysed in a first section and then all the different methods are described separately. Their advantages and drawbacks are also given. Finally, a synthesis helps the reader to determine the most appropriate method to a given rare event estimation problem.

* Corresponding author. Tel.: +33 5 62 25 26 63.
E-mail addresses: jerome.morio@onera.fr (J. Morio), mathieu.balesdent@onera.fr (M. Balesdent), damien.jacquemart@onera.fr (D. Jacquemart), christelle.verge@onera.fr (C. Vergé).

http://dx.doi.org/10.1016/j.simpat.2014.10.007
1569-190X/

Let us consider a d-dimensional random vector $\boldsymbol{X}$ with a probability density function (PDF) h_0, ϕ a continuous positive scalar function $\phi : \mathbb{R}^d \to \mathbb{R}$ and S a threshold. The different components of $\boldsymbol{X}$ will be denoted $\mathbf{X} = (X^1, X^2, \ldots, X^d)$ in the following. The function ϕ is static, *i.e.*, does not depend on time, and represents for instance an input–output model. This kind of model is notably used in numerous engineering applications [4–9]. We assume that the output $Y = \phi(\mathbf{X})$ is a scalar random variable. In this article, we propose to review different algorithms that can be efficient to estimate the probability $P = P(\phi(\mathbf{X}) > S)$ when this quantity is rare relatively to the available simulation budget N, that is when $P < \frac{1}{N}$. For the sake of conciseness, the issue of extreme quantile estimation is not addressed even if the vast majority of the methods that are presented in the paper can be adapted to this specific case. The case of dynamic systems modeled with Markov chains is also not considered in this paper. Specific algorithm extensions for large complex systems modeled by a network or a coherent fault tree are completely detailed in [10] and will not be much developed here. It corresponds to the case where the inputs $X^i, i = 1, \ldots, d$ follow a Bernoulli distribution and the output is equivalent to an indicator function.

2. Monte-Carlo methods

A simple way to estimate a probability is to consider Crude Monte-Carlo (CMC) [11–16]. For that purpose, one generates N independent and identically distributed (i.i.d.) samples $\mathbf{X}_1, \ldots, \mathbf{X}_N$ from the PDF h_0 and computes their outputs with the function $\phi : \phi(\mathbf{X}_1), \ldots, \phi(\mathbf{X}_N)$. The probability $P(\phi(\mathbf{X}) > S)$, also called failure probability, is then estimated with

$$\widehat{P}^{CMC} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\phi(\mathbf{X}_i)>S}, \tag{1}$$

where $\mathbf{1}_{\phi(\mathbf{X}_i)>S}$ is equal to 1 if $\phi(\mathbf{X}_i) > S$ and 0 otherwise. This estimation converges to the real probability as shows the law of large numbers [13]. The positive and negative aspects of CMC are described in Table 1. A possible indicator of the estimation efficiency is notably its relative deviation. The relative deviation or relative error RE of an estimator $\widehat{P}$ of P is given by the following ratio:

$$RE(\widehat{P}) = \frac{\sigma_{\widehat{P}}}{\mathbb{E}(\widehat{P})}, \tag{2}$$

with $\sigma_{\widehat{P}}$ the standard deviation of $\widehat{P}$ and $\mathbb{E}$ the mathematical expectation. The relative error is said bounded when $RE(\widehat{P})$ remains bounded when $P \longrightarrow 0$ [17,18]. In that case, the number of samples needed to get a specified relative error is bounded whatever the rarity of $\phi(\mathbf{X}) > S$. The logarithmic efficiency LE can also be defined for an unbiased estimator $\widehat{P}$ with [17,18],

$$LE(\widehat{P}) = \lim_{P\to 0}\frac{\log(\mathbb{E}(\widehat{P}^2))}{\log(P)} = 2. \tag{3}$$

Logarithmic efficiency is a necessary but not sufficient condition for bounded relative error. Characterizing the rare event probability estimate with these concepts is very important even if they are often difficult to verify in practice.

Since $\widehat{P}^{CMC}$ is unbiased, the relative error of the estimator $\widehat{P}^{CMC}$ is given by the ratio $\frac{\sigma_{\widehat{P}^{CMC}}}{P}$ with $\sigma_{\widehat{P}^{CMC}}$, the standard deviation of $\widehat{P}^{CMC}$. Knowing the true probability P of the event $(\phi(\mathbf{X}) > S)$, one has [11,19]

$$\frac{\sigma_{\widehat{P}^{CMC}}}{P} = \frac{1}{\sqrt{N}}\frac{\sqrt{P-P^2}}{P}. \tag{4}$$

Considering rare event probability estimation, that is when P takes low values, one obtains

$$\lim_{P\to 0}\frac{\sigma_{\widehat{P}^{CMC}}}{P} = \lim_{P\to 0}\frac{1}{\sqrt{NP}} = +\infty. \tag{5}$$

The relative deviation is consequently unbounded. For instance, to estimate a probability P of order 10^{-4} with a 10% relative deviation, at least 10^6 samples are required. The simulation budget is thus an issue when the computation time required to obtain a sample $\phi(\mathbf{X}_i)$ is not negligible. CMC is thus not adapted to rare event estimation and a wide collection of statistic and simulation methods has been developed. The following sections describe the different available alternatives to CMC to

Table 1
Advantages and drawbacks of CMC methods.

Advantages of CMC	Drawbacks of CMC
Simple implementation	Slow convergence
Information on ϕ not needed	Significant simulation budget for rare events
No bias	