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

Jérôme Morio, Mathieu Balesdent, Damien Jacquemart, Christelle Vergé

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

Jérôme Morio [1,*]
Mathieu Balesdent [2] Damien Jacquemart [2,3] Christelle Vergé [2,4,5]

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 modelling. 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?".

Key words: Monte-Carlo methods, Rare event, Input-output model, Simulation

* corresponding author
Email addresses: jerome.morio@onera.fr (Jérôme Morio), mathieu.balesdent@onera.fr (Mathieu Balesdent), damien.jacquemart@onera.fr (Damien Jacquemart), christelle.verge@onera.fr (Christelle Vergé).
[1] Onera - The French Aerospace Lab, BP 74025, 31055 Toulouse Cedex, France Tel.: +33 5 62 25 26 63
[2] Onera - The French Aerospace Lab, BP 80100, 91123 Palaiseau Cedex, France
[3] INRIA Rennes, ASPI Applications of interacting particle systems to statistics, campus de Beaulieu, 35042 Rennes, France
[4] INRIA Bordeaux, 351 cours de la Libération, 33405 Talence Cedex, France
[5] CNES, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France

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.
Let us consider a d-dimensional random vector $\mathbf{X}$ 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 $\mathbf{X}$ will be denoted $\mathbf{X} = (X^1, X^2, ..., X^d)$ 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 modelled 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, $i = 1, ..., d$ 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 $\mathbf{X}_1, ..., \mathbf{X}_N$ from the PDF h_0 and computes their outputs with the function ϕ: $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$. The probability $P(\phi(\mathbf{X}) > S)$, also called failure probability, is then estimated with

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

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

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

Table 1
Advantages and drawbacks of CMC methods.

$\hat{P}$ of P is given by the following ratio:

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

with $\sigma_{\hat{P}}$ the standard deviation of $\hat{P}$ and $\mathbb{E}$ the mathematical expectation. The relative error is said bounded when $RE(\hat{P})$ 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 $\hat{P}$ with [17,18],

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

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

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

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

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

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 improve probability estimations, *i.e.*, to reduce the number of required samples, increase the estimation accuracy, and thus decrease $RE(\hat{P})$.

3. Statistical techniques

Statistical techniques enable to derive a probability estimate and associated confidence intervals with a fixed set of samples $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$. The main statistical approaches, extreme value theory and large deviation theory, model the behaviour of the PDF tails. Let us review their theoretical founding.

3.1. *Extreme value theory*

Extreme value theory (EVT) [20,21] characterizes the distribution tails of a random variable, based on a reasonable number of observations. Thanks to its general applicative conditions, this theory has been widely used for describing extreme meteorological phenomena with applications such as hydrology [22], snowfall [23], but also in finance and insurance [20,24], and engineering [25].

3.1.1. *Law of sample maxima*

EVT is notably very useful when one has to work with only a fixed set of data. One consequently assumes in the following that a finite set of i.i.d. samples $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$ of the output is available, but also that one cannot generate new samples of $\phi(\mathbf{X})$. The associated ordered sample set is defined with $\phi(\mathbf{X}_{(1)}) \leq \phi(\mathbf{X}_{(2)}) \leq ... \leq \phi(\mathbf{X}_{(N)})$. EVT enables to estimate for some threshold S the probability $P(\phi(\mathbf{X}) > S)$.
The founder theorem of EVT [20,26,27] is that, under some conditions, the maxima of an i.i.d. sequence converge to a generalized extreme value (GEV) distribution G_ξ, which admits the following cumulative distribution function (CDF)

$$G_\xi(x) = \begin{cases} \exp(-\exp(-x)), & \text{for } \xi = 0, \\ \exp\left(-(1+\xi x)^{-\frac{1}{\xi}}\right), & \text{for } \xi \neq 0. \end{cases} \tag{6}$$

The set of GEV distributions is composed of three distinct types, characterized by $\xi = 0$, $\xi > 0$ and $\xi < 0$ that correspond to the Gumbel, Fréchet and Weibull distributions respectively. Let us define G, the CDF of the i.i.d. samples $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$.
Theorem 3.1 *Suppose there exist a_N and b_N, with $a_N > 0$ such that, for all $y \in \mathbb{R}$*

$$P\left(\frac{\phi(\mathbf{X}_{(N)}) - b_N}{a_N} \leq y\right) = G^N(a_N y + b_N) \overset{N\to\infty}{\longrightarrow} G(y),$$

where G is a non degenerate CDF, then G is a GEV distribution G_ξ. In this case, one denotes $G \in MDA(\xi)$ (MDA=maximum domain of attraction).
The sequences a_N and b_N are computed in [20] for most well-known PDF. An approximation of $P(\phi(\mathbf{X}) > S)$ [20] for large values of S and N can also be obtained:

$$\hat{P}^{EVT}(\phi(\mathbf{X}) > S) \approx \frac{1}{N}\left(1 + \xi\left(\frac{S - b_N}{a_N}\right)\right)^{-\frac{1}{\xi}}. \tag{7}$$

The GEV approach is notably used when only samples of maxima are available. In that case, the different parameters of the GEV distribution are obtained by determining maximum likelihood or probability weighted moment estimators. When samples of maxima are not available, it is required to group the samples $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$ into blocks and fit the GEV using the maximum of each block (block maxima method). The main difficulty is to determine an efficient sample size for the different blocks.

3.1.2. *Peak over threshold approach*

Instead of grouping the samples into block maxima, POT considers the largest samples $\phi(\mathbf{X}_i)$ to estimate the probability $P(\phi(\mathbf{X}) > S)$.

There are two equivalent ways of analyzing extremes with POT. The most common is to characterize the distribution of samples above a threshold u, which is given by the generalized Pareto CDF. An alternative is to use a Poisson point process which counts the number of threshold exceedances. This approach is not developed in this article, but one can refer to [27] for more details. The first paper linking the EVT with the distribution of a threshold exceedance is [28]. Later, De Haan obtains a result of the same type, with a slightly simplified conclusion, using slow varying functions [29]. The following theorem [20] can be then obtained:

Theorem 3.2 *Let us assume that the distribution function G of i.i.d. samples $\phi(\mathbf{X}_1)$,..., $\phi(\mathbf{X}_N)$ is continuous. Set $y^* = \sup\{y,\, G(y) < 1\} = \inf\{y,\, G(y) = 1\}$. Then, the two following assertions are equivalent*

(i) $G \in MDA(\xi)$,

(ii) there exists a positive and measurable function $u \mapsto \beta(u)$ such that

$$\lim_{u \mapsto y^*} \sup_{0<y<y^*-u} |G^u(y) - H_{\xi,\beta(u)}(y)| = 0,$$

where $G^u(y) = P(\phi(\mathbf{X}) - u \leq y | \phi(\mathbf{X}) > u)$, and $H_{\xi,\beta(u)}$ is the CDF of a generalized Pareto distribution (GPD) with shape parameter ξ and scale parameter $\beta(u)$.

The expression of the GPD distribution function is the following

$$H_{\xi,\beta}(x) = \begin{cases} 1 - \exp\left(-\frac{x}{\beta}\right), \text{ for } \xi = 0, \\ 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}, \text{ for } \xi \neq 0. \end{cases} \tag{8}$$

This theorem is in fact useful to estimate a probability of exceedance. Indeed, the probability $P(\phi(\mathbf{X}) > S)$ can be rewritten as

$$P(\phi(\mathbf{X}) > S) = P(\phi(\mathbf{X}) > S | \phi(\mathbf{X}) > u) P(\phi(\mathbf{X}) > u). \tag{9}$$

for $S > u$. A natural estimate of $P(\phi(\mathbf{X}) > u)$ is given by

$$\hat{P}^{CMC}(\phi(\mathbf{X}) > u) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i) > u}. \tag{10}$$

With the Theorem 3.2 and for significant value of u, one obtains

$$\hat{P}(\phi(\mathbf{X}) > S | \phi(\mathbf{X}) > u) = 1 - H_{\xi,\beta(u)}(S - u). \tag{11}$$

The estimate of $P(\phi(\mathbf{X}) > S)$ is then built with

$$\hat{P}^{POT}(\phi(\mathbf{X}) > S) = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i) > u} \right) \times \left(1 - H_{\xi,\beta(u)}(S - u)\right). \tag{12}$$

The mathematical justification of Eq. 11 and Eq. 12 is notably discussed in [21], [30], [31], or [32] for a given set of samples to determine if this set is suitable for the application of POT. Three parameters have to be determined in the POT probability estimate of Eq. 12: the threshold u and the couple $(\xi, \beta(u))$. The choice of u is very influent since it determines the samples that are used in the estimation of $(\xi, \beta(u))$. Indeed, a high threshold leads to consider only a small number of samples in the estimation of $(\xi, \beta(u))$ and thus their estimate can be then spoiled by a large variance whereas a low threshold

Advantages of EVT	Drawbacks of EVT
No need to resample	Complex estimation of the adequate parameters $(u, \xi, \beta(u))$ or of the block maxima size.
Can be applied with a relatively low value of N	Less efficient than simulation methods when resampling is possible

Table 2
Advantages and drawbacks of EVT.

introduces a bias in the probability estimate [33]. There are several methods to determine a valuable threshold u knowing the samples. The most well-known ones are the Hill plot and the mean excess plot [20]. These methods are nevertheless very empirical since they are based on graphical interpretation. It is often necessary in practice to compare the estimates of u given by the different methods. Once the value of u is set, the parameters $(\xi, \beta(u))$ are often estimated by maximum likelihood [34] or more occasionally by the method of moments [35]. The estimate $\hat{P}^{POT}(\phi(\mathbf{X}) > S)$ given in Eq. 12 for $S > u$ is then completely defined. A review of these different methods can be found in [36]. It is not possible, to our knowledge, to control the probability error estimate in EVT. Nevertheless, the use of boostrap on samples $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$ [37] can give some information on the efficiency of EVT.

3.1.3. *Block maxima versus POT*

The POT method takes into account all relevant high samples $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$ whereas the block maxima method can miss some of these high samples and, on the same time, consider some lower samples in its probability estimation. Thus, POT seems to be more appropriate for the design of sample PDF tail. Nevertheless, the block maxima method is preferable when the available samples are not exactly i.i.d. or when only samples of maxima are available. For instance, the samples of a monthly river maximum height correspond to this situation. Finally, the tuning of block maxima size turns out to be easier than the tuning of POT threshold u in many situations [38]. The advantages and drawbacks of EVT are presented in Table 2.

3.2. *Large deviation theory*

The large deviation theory (LDT) characterizes the asymptotic behaviour of PDF sequence tails [39–41] and more precisely, it analyses how a PDF sequence tail deviates from its typical behaviour described by the law of large numbers. LDT can be used to evaluate the convergence of rare event algorithms [42–46]. Let us define $H_N = J(\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N))$ a random variable indexed by N with J a continuous scalar function, H its mathematical expectation and $V_N = H_N - H$. One says that V_N satisfies the principle of large deviations with a continuous rate function I if the following limit exists:

$$\lim_{N \to \infty} \frac{1}{N} \ln[P(\mid V_N \mid > \gamma)] = -I(\gamma). \tag{13}$$

The existence of this limit implies for a large value of N that

$$P(\mid V_N \mid > \gamma) \approx \exp\left(-NI(\gamma)\right). \tag{14}$$

The probability decays exponentially as N grows to infinity, at a rate depending on γ. This approximation is a well-known result of LDT. If the limit does not exist, then $P(\mid V_N \mid > \gamma)$ has a too singular behaviour or decreases faster than exponential decay. If the limit is equal to 0, then the tail $P(\mid V_N \mid > \gamma)$ decreases with N slower than $\exp(-Na)$ with $a > 0$. The computation of the rate function I is not obvious but can be obtained through the Gärtner-Ellis theorem [47]. Let us define the function $\lambda(\theta)$ of V_N with

$$\lambda(\theta) = \lim_{N\to\infty} \frac{1}{N} \ln\left[\mathbb{E}\left(\exp\left(N\theta V_N\right)\right)\right], \tag{15}$$

with $\theta \in \mathbb{R}$.

Theorem 3.3 Gärtner-Ellis theorem *If the function $\lambda(\theta)$ of the variable V_N exists and is differentiable for all $\theta \in \mathbb{R}$, then V_N satisfies the principle of large deviations and $I(\gamma)$ is given by*

$$I(\gamma) = \sup_{\theta\in\mathbb{R}} \left[\theta\gamma - \lambda(\theta)\right].$$

In the specific case of a scalar function J, one can derive the Cramér theorem from Gärtner-Ellis theorem [47].

Theorem 3.4 Cramér theorem *If $V_N = \frac{1}{N}\sum_{i=1}^{N} J(\phi(\mathbf{X}_i))$ where the random variables $J(\phi(\mathbf{X}_i))$ are i.i.d, the rate function is given by*

$$I(\gamma) = \sup_{\theta\in\mathbb{R}} \left[\theta\gamma - \lambda(\theta)\right],$$

with

$$\lambda(\theta) = \ln\left[\mathbb{E}\left(\exp\left(\theta J(\phi(\mathbf{X}))\right)\right)\right].$$

This theorem only holds for light tail distributions.

Let us consider the Monte-Carlo probability estimate given in Eq. 1. In that case, one has $J(\phi(.)) = \mathbf{1}_{\phi(.)}$. The random variable $J(\phi(\mathbf{X}_i))$ follows a Bernoulli distribution of mean P. The sequence V_N is defined with

$$V_N = \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i)>S}\right) - P. \tag{16}$$

The functions $\lambda(\theta)$ and $I(\gamma)$ can be derived for some well-known PDF. In the case of Bernoulli distributions of mean P, one has

$$\lambda(\theta) = P\exp(\theta) + 1 - P, \tag{17}$$

and

$$I(\gamma) = \gamma\ln\left(\frac{\gamma}{P}\right) + (1-\gamma)\ln\left(\frac{1-\gamma}{1-P}\right). \tag{18}$$

One can then obtain the convergence speed of the Monte-Carlo probability estimate in function of the number of samples with the following equation

$$\lim_{N\to\infty} \frac{1}{N}\ln[P(\mid V_N \mid > \gamma)] = -I(\gamma) = -\gamma\ln\left(\frac{\gamma}{P}\right) - (1-\gamma)\ln\left(\frac{1-\gamma}{1-P}\right). \tag{19}$$

The quantity $I(\gamma)$ corresponds to the relative entropy (Kullback-Leibler divergence) of a coin toss with bias γ with respect to true value P. In a lot of situations, the large deviation rate function is the Kullback-Leibler divergence [47].

LDT cannot in fact be applied directly to determine a rare event probability in a realistic practical case where the density of Y is not known *a priori*. LDT can be useful to analyze the deviation of a probability estimate, notably if the probability estimate is a sum of random variables as shows Eq. 19. for the CMC estimate. Specific surveys on LDT can be found in [3,48].

4. Importance sampling

4.1. *Principle of importance sampling*

The objective of importance sampling (IS) is to reduce the variance of the Monte-Carlo estimator $\widehat{P}^{CMC}$ [17,19,49–53]. The main idea is to generate the samples $\mathbf{X}_1, ..., \mathbf{X}_N$ with an auxiliary PDF h that is able to generate more samples such that $\phi(\mathbf{X}) > S$ than PDF h_0 and then to introduce a weight in the probability estimate to take into account the change in the PDF generating the samples. The IS probability estimate $\widehat{P}^{IS}$ is then given with

$$\widehat{P}^{IS} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i)>S} \frac{h_0(\mathbf{X}_i)}{h(\mathbf{X}_i)}. \tag{20}$$

The term $\widehat{P}^{IS}$ is an unbiased estimate of the probability P. Its variance is given by the following equation:

$$Var\left(\widehat{P}^{IS}\right) = \frac{Var\left(\mathbf{1}_{\phi(\mathbf{X})>S} w(\mathbf{X})\right)}{N}, \tag{21}$$

with $w(\mathbf{X}) = \frac{h_0(\mathbf{X})}{h(\mathbf{X})}$. The term $w(\mathbf{X})$ is often called the likelihood function in the importance sampling literature. The variance of $\widehat{P}^{IS}$ strongly depends on the choice of h. If h is well-chosen, the IS estimate has then a much smaller variance than Monte-Carlo estimate and conversely. The objective of IS is to decrease the estimation variance and one can thus define an optimal IS auxiliary density that minimizes the variance $Var\left(\widehat{P}^{IS}\right)$. Since variances are non negative quantities, the optimal auxiliary density h_{opt} is determined by cancelling the variance in Eq. 21. It is well-known that h_{opt} is then defined with [54]

$$h_{opt}(\mathbf{X}) = \frac{\mathbf{1}_{\phi(\mathbf{X})>S} h_0(\mathbf{X})}{P}. \tag{22}$$

The optimal auxiliary density h_{opt} depends unfortunately on the probability P that one tries to estimate and is unusable in practice. Nevertheless, h_{opt} can be useful to determine an efficient sampling PDF. Indeed, a valuable sampling auxiliary PDF h will be close to the PDF h_{opt} relative to a given criterion. An optimization of the auxiliary sampling PDF is then necessary. In some specific cases or specific functions ϕ, importance sampling probability estimate can have a bounded relative error as demonstrated in [55,56] or logarithmic efficiency in [57,58].
Specific surveys on IS have been proposed such as in [1,59], and thus, the complete list of possible importance algorithms will not be described for the sake of conciseness. We only review the main algorithms in the next sections.

4.2. *Cross entropy optimization of importance sampling auxiliary density*

Let us define $h_{\boldsymbol{\lambda}}$, a family of PDF indexed by a parameter $\boldsymbol{\lambda} \in \Delta$ where Δ is the multidimensional space of PDF parameters. The parameter $\boldsymbol{\lambda}$ is, for instance, the mean and the covariance matrix in the case of Gaussian densities. The objective of IS with cross entropy (CE) is to determine the parameter $\boldsymbol{\lambda}_{opt}$ that minimizes the Kullback-Leibler divergence between $h_{\boldsymbol{\lambda}_{opt}}$ and h_{opt} [60,61]. The value of $\boldsymbol{\lambda}_{opt}$ is thus obtained with

$$\boldsymbol{\lambda}_{opt} = \underset{\boldsymbol{\lambda} \in \Delta}{\operatorname{argmin}} \left\{ D(h_{opt}, h_{\boldsymbol{\lambda}}) \right\}, \tag{23}$$

where D is the Kullback-Leibler divergence defined between PDF p and PDF q by

$$D(q,p) = \int_{\mathbb{R}^d} q(\mathbf{x}) \ln(q(\mathbf{x})) d\mathbf{x} - \int_{\mathbb{R}^d} q(\mathbf{x}) \ln(p(\mathbf{x})) d\mathbf{x}. \tag{24}$$

Determining the parameter $\boldsymbol{\lambda}_{opt}$ with Eq. 23 is not obvious since it depends on the unknown PDF h_{opt}. In fact, it can be shown [60] that Eq. 23 is equivalent to the following one

$$\boldsymbol{\lambda}_{opt} = \underset{\boldsymbol{\lambda} \in \Delta}{\operatorname{argmax}} \left\{ \mathbb{E}\left[\mathbf{1}_{\phi(\mathbf{X}) > S} \ln\left(h_{\boldsymbol{\lambda}}(\mathbf{X}) \right) \right] \right\}. \tag{25}$$

In practice, one does not focus directly on Eq. 25 since it requires the knowledge of some samples of $\mathbf{X}$ so that $\phi(\mathbf{X}) > S$. In most realistic applications, it is not the case. Thus, one proceeds iteratively to estimate $\boldsymbol{\lambda}_{opt}$ with an increasing sequence of thresholds

$$\gamma_0 < \gamma_1 < \gamma_2 < ... < \gamma_k < ... \leq S, \tag{26}$$

chosen adaptively using quantile definition. At the iteration k, the value $\boldsymbol{\lambda}_{k-1}$ is available and one determines in practice

$$\boldsymbol{\lambda}_k = \underset{\boldsymbol{\lambda} \in \Delta}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i) > \gamma_k} \frac{h_0(\mathbf{X}_i)}{h_{\boldsymbol{\lambda}_{k-1}}(\mathbf{X}_i)} \ln(h_{\boldsymbol{\lambda}}(\mathbf{X}_i)), \tag{27}$$

where the samples $\mathbf{X}_1, ..., \mathbf{X}_N$ are generated with $h_{\boldsymbol{\lambda}_{k-1}}$. The probability $\widehat{P}^{CE}$ is then estimated with IS at the last iteration. The cross entropy optimization algorithm for the IS density is described more precisely by the following scheme

(i) $k = 1$, define $h_{\boldsymbol{\lambda}_0} = h_0$ and set $\rho \in]0,1[$.

(ii) Generate the population $\mathbf{X}_1, ..., \mathbf{X}_N$ according to the PDF $h_{\boldsymbol{\lambda}_{k-1}}$ and apply the function ϕ in order to have $Y_1 = \phi(\mathbf{X}_1), ..., Y_N = \phi(\mathbf{X}_N)$.

(iii) Compute $\gamma_k = min(S, Y_\rho)$ where Y_ρ denotes the empirical ρ-quantile of $Y_1, ..., Y_N$.

(iv) Optimize the parameters of the auxiliary PDF family with

$$\boldsymbol{\lambda}_k = \underset{\boldsymbol{\lambda} \in \Delta}{\operatorname{argmax}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[\mathbf{1}_{\phi(\mathbf{X}_i) > \gamma_k} \frac{h_0(\mathbf{X}_i)}{h_{\boldsymbol{\lambda}_{k-1}}(\mathbf{X}_i)} \ln\left[h_{\boldsymbol{\lambda}}(\mathbf{X}_i) \right] \right] \right\}.$$

(v) If $\gamma_k < S$, $k \leftarrow k+1$, back to the step (ii).

(vi) Estimate the probability $\widehat{P}^{CE}(\phi(\mathbf{X} > S)) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i) > S} \frac{h_0(\mathbf{X}_i)}{h_{\boldsymbol{\lambda}_k - 1}(\mathbf{X}_i)}$.

The advantages of and drawbacks of CE are presented in Table 3. CE is a very practical algorithm to approximate the optimal sampling density. Nevertheless, the choice of the parametric family density $h_{\boldsymbol{\lambda}}$ has to be done carefully to obtain valuable results. Due to

Advantages of CE	Drawbacks of CE
Simple optimization for exponential PDF family	Strong influence of the initial parametric density choice
Fast computation	Difficult to apply in cases where the optimal auxiliary density is multimodal

Table 3
Advantages and drawbacks of CE.

the adaptiveness of the algorithm, it is difficult to ensure the robustness (logarithmic efficiency) of the CE estimate in the general case [62]. The concept of probabilistic bounded relative error is then proposed.

4.3. *Non parametric adaptive importance sampling*

The objective of non parametric adaptive importance sampling (NAIS) technique [63–66] is to approximate the IS optimal auxiliary density given in Eq. 22 with kernel density function [67]. NAIS does not require the choice of a PDF family and is thus more flexible than a parametric model. The iterative principle is relatively similar to the CE optimization and is described by the following steps. For the sake of simplicity, the algorithm is presented with a Gaussian kernel but other kinds of kernel can be used.

(i) $k = 1$ and set $\rho \in]0, 1[$.

(ii) Generate the population $\mathbf{X}_1^{(k)}, ..., \mathbf{X}_N^{(k)}$ according to the PDF h_{k-1}, apply the function ϕ in order to have $Y_1^{(k)} = \phi(\mathbf{X}_1^{(k)}), ..., Y_N^{(k)} = \phi(\mathbf{X}_N^{(k)})$.

(iii) Compute $\gamma_k = min(S, Y_\rho^{(k)})$ where $Y_\rho^{(k)}$ denotes the empirical ρ-quantile of $Y_1^{(k)}, ..., Y_N^{(k)}$.

(iv) Estimate $I_k = \frac{1}{kN} \sum_{j=1}^{k} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i^{(j)}) \geq \gamma_k} \frac{h_0(\mathbf{X}_i^{(j)})}{h_{j-1}(\mathbf{X}_i^{(j)})}$.

(v) Update the Gaussian kernel sampling PDF with

$$h_k(\mathbf{X}) = \frac{1}{kNI_k \det(B_k)} \sum_{j=1}^{k} \sum_{i=1}^{N} w_j(\mathbf{X}_i^{(j)}) K_d \left(B_k^{-1} \left(\mathbf{X} - \mathbf{X}_i^{(j)} \right) \right). \tag{28}$$

where K_d is standard d-dimensional Gaussian function with zero mean and a diagonal covariance matrix $B_k = diag(b_k^1, ..., b_k^d)$ and $w_j(.) = \mathbf{1}_{\phi(.) \geq \gamma_k} \frac{h_0(.)}{h_{j-1}(.)}$. The adapted coefficient in the matrix B_{k+1} can be optimized according to the AMISE (asymptotic mean integrated square error) criterion [11] and [68].

(vi) If $\gamma_k < S$, $k \leftarrow k + 1$, back to the step (ii).

(vii) Estimate the probability $\widehat{P}^{NAIS}(\phi(\mathbf{X}) > S) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i^{(k)}) > S} \frac{h_0(\mathbf{X}_i^{(k)})}{h_{k-1}(\mathbf{X}_i^{(k)})}$.

The advantages of and drawbacks of NAIS are presented in Table 4. The use of kernel density function enables a more flexible and general model than CE. It become very difficult to apply NAIS in cases where the input dimension d is greater than 10 due to the numerical cost induced by the use of kernel density [66].

Advantages of NAIS	Drawbacks of NAIS
No choice of a parametric density	Computation time
Efficient in cases where the optimal auxiliary density is multimodal	Inapplicable when d is greater than 10

Table 4
Advantages and drawbacks of NAIS.

4.4. *Simple changes of measure*

The use of CE or NAIS is not always necessary, notably in simple cases of function $\phi(.)$. Conventional changes of density h_0 can then be efficient to decrease the probability estimate variance. Scaling and translation can be applied on the initial PDF h_0. Scaling consists in defining the auxiliary PDF h so that

$$h(\mathbf{X}) = \frac{1}{a} h_0 \left(\frac{\mathbf{X}}{a} \right), \tag{29}$$

with $a \in \mathbb{R}^*$. Translation is another simple change of density that can be applied in IS. The new auxiliary density is defined with translation by

$$h(\mathbf{X}) = h_0(\mathbf{X} - \mathbf{c}), \tag{30}$$

with $\mathbf{c} \in \mathbb{R}^d$. The choices of a and c for each method strongly influence the importance sampling efficiency. Valuable values of a and c are not obvious to find without some knowledge of the function ϕ.

4.5. *Exponential twisting*

The principle of exponential twisting is very similar to LDT and saddle point approximation [69–72]. The main idea of exponential twisting is to define the auxiliary density on the output $Y = \phi(\mathbf{X})$ with

$$h(y) = \exp(\theta y - \lambda(\theta)) g(y), \tag{31}$$

where g is the density of random variable Y and $\lambda(\theta) = \ln\left(\mathbb{E}\left(\exp\left(\theta Y\right)\right)\right)$. The probability is then determined with

$$\mathbb{P}^{TW} = \mathbb{E}\left(\mathbf{1}_{Y>S} \frac{g(Y)}{h(Y)} \right).$$

The variable Y has to get exponential moments so that $\lambda(\theta)$ to be finite for at least some values of $\theta \in \mathbb{R}$. The PDF $h(y)$ depends on the parameter θ. An optimal value θ_{opt} can be obtained with saddle point approximation with

$$\left. \frac{d\lambda(\theta)}{d\theta} \right|_{\theta=\theta_{opt}} = S. \tag{32}$$

The parameter θ_{opt} is estimated numerically. Exponential twisting can thus only be applied in some specific cases, notably if $Y = \sum_{i=1}^{d} X^i$ (function used in some queueing models) or if the density g is analytically known. In the case of a sum of random variables, this estimator has a bounded relative error if the input has a light tail [73,74]. In case of large deviation probabilities and under some general conditions, logarithmic efficiency is guaranteed with exponential twisting importance sampling [75].

5. FORM/SORM

First/second-order reliability methods (FORM/SORM) [76–79] are considered as reliable computational methods for structural reliability. FORM is an analytical approximation in which the reliability index is interpreted as the minimum distance from the origin to the limit state surface in standardized normal input space. This limit state surface characterizes the input region where $\phi(\mathbf{X}) > S$. The most probable failure point (design point) is searched using mathematical programming methods. Since the performance function is approximated by a linear function at the design point, accuracy problems occur when the performance function is strongly nonlinear or if the most probable failure point is not unique [80]. The second-order reliability method (SORM) has been established as an attempt to improve the accuracy of FORM. SORM approximates the limit state surface at the design point by a second-order surface.

FORM/SORM method are applied in four stages to estimate $P(\phi(\mathbf{X}) > S)$:

(i) Apply a transformation T on the input $\mathbf{X}$ such that $\mathbf{R} = T(\mathbf{X})$ with $\mathbf{R}$ a normal reduced centered PDF. Depending on the available information on the PDF of $\mathbf{X}$, several transformations can be proposed [81–86]. See Table 5 for details on the correspondence between assumptions and transformations.

(ii) Evaluate the most probable failure point β such that

$$\beta = \underset{\mathbf{R}}{\operatorname{argmin}} \mid\mid \mathbf{R} \mid\mid, \tag{33}$$

subject to the constraint $S - \phi(T^{-1}(\mathbf{R})) = 0$ and where $\mid\mid . \mid\mid$ is the Euclidian norm. The constraint $S - \phi(T^{-1}(\mathbf{R})) = 0$ defines the limit of failure space for variable $\mathbf{R}$. The parameter β is the design point and $\mid\mid \beta \mid\mid$ is the reliability index. Several algorithms have been proposed to solve this optimization problem as proposed in [82,83,87,88].

(iii) Approximate the surface $S - \phi(T^{-1}(\mathbf{R})) = 0$ at the solution β. In the case of FORM, this surface is a hyperplane and it is a paraboloid in the case of SORM [89].

(iv) Estimate the failure probability with, in the case of FORM :

$$\hat{P}^{FORM}(\phi(\mathbf{X}) > S) = \Omega(- \mid\mid \beta \mid\mid), \tag{34}$$

where Ω is the CDF of a normal reduced and centered PDF. In the case of SORM, the failure probability is given by [90]

$$\hat{P}^{SORM}(\phi(\mathbf{X}) > S) = \Omega(- \mid\mid \beta \mid\mid) \prod_{i=1}^{d-1} (1 - \beta\kappa_i)^{-\frac{1}{2}}, \tag{35}$$

where κ_i denotes the principal curvature of $S - \phi(T^{-1}(\mathbf{R}))$ at the design point β. The term κ_i is defined with

$$\kappa_i = \left. \frac{\partial^2 (S - \phi(T^{-1}(\mathbf{R})))}{\partial^2 R^i} \right|_{\mathbf{R}=\beta}, \tag{36}$$

with R^i, $i = 1, ..., d$, a component of the vector $\mathbf{R}$. A first order saddle point approximation (FOSPA) [91,92] method has also been proposed as an improvement to FORM/SORM. It consists in using LDT and the saddle point approximation [69–72] which considers the function

Assumptions on the PDF of $\mathbf{X}$	Corresponding transformations T
X is Gaussian with uncorrelated components	Hasofer-Lind transformation
X has independent components (not assumed to be Gaussian)	Diagonal transformation
Only the marginal laws of $\mathbf{X}$ and their covariance are known	Nataf tranformation
The complete law of $\mathbf{X}$ is known	Rosenblatt transformation

Table 5
Possible transformations T depending on the assumptions on the PDF of $\mathbf{X}$.

$$\lambda(\theta) = \ln\left[\mathbb{E}\left(\theta\phi(\mathbf{X})\right)\right], \tag{37}$$

to estimate the repartition function of $\phi(\mathbf{X})$. Indeed, it is possible to show that

$$P(\phi(\mathbf{X}) > S) \approx 1 - \Omega\left(w + \frac{1}{w}\ln\left(\frac{v}{w}\right)\right), \tag{38}$$

with

$$w = \text{sign}(\theta_s)(2(\theta_s S - \lambda(\theta_s)))^{\frac{1}{2}}, \tag{39}$$

and

$$v = \theta_s\left(2\frac{d^2\lambda(\theta)}{d\theta^2}|_{\theta=\theta_s}\right)^{\frac{1}{2}}. \tag{40}$$

The parameter θ_s is the saddle point and is the solution of the equation

$$\frac{d^2\lambda(\theta)}{d\theta^2}|_{\theta=\theta_s} = S. \tag{41}$$

The approximation proposed in Eq. 38 is not easily computable in the general case. It is thus often necessary to linearize the function ϕ near the most probable failure point with the constraint $S - \phi(\mathbf{X}) = 0$ and also to linearize the function λ. These linearizations simplify the estimation of $\lambda(\theta)$ in Eq. 37 and of θ_s. The moment method is also used to approximate the function λ in [91,93,94].
The advantage of and drawbacks of geometric methods such as FORM/SORM/FOSPA are given in Table 6. These methods do not require a large simulation budget to obtain a valuable result. Nevertheless, the different assumptions require that one has to be careful when one applies FORM/SORM/FOSPA to a realistic case of function ϕ. There is also no control of the error in FORM/ SORM. However, it is possible from FORM/SORM to determine an importance sampling auxiliary density and then to sample with it to estimate the rare event probability.

6. Line sampling

6.1. *Principle*

The underlying idea of Line Sampling (LS) [95–97] is to employ lines instead of random points in order to probe the failure domain of the system, *i.e.* $\mathbf{X}$ so that $\phi(\mathbf{X}) > S$. It has to be applied on input random variables that have zero-mean standard normal density. Let us first assume that $\mathbf{X}$ follows a multidimensional zero-mean standard normal

Advantages of FORM/SORM/FOSPA	Drawbacks of FORM/SORM/FOSPA
Necessary simulation budget very restricted	Difficult to apply when the optimal auxiliary density is multimodal
	Necessary transformation on input variables if they are not Gaussian
	Not adapted to non linear and to high dimensional function ϕ
	No possible control of the error

Table 6
Advantages and drawbacks of FORM/SORM.

distribution and also define the set $\mathbf{A} = \{\mathbf{X} \in \mathbb{R}^d | \phi(\mathbf{X}) > S\}$. The set $\mathbf{A}$ can be also expressed in the following way

$$\mathbf{A} = \{\mathbf{X} \in \mathbb{R}^d | X^1 \in \mathbf{A}_1(\mathbf{X}^{-1})\}. \tag{42}$$

where the set $\mathbf{A}_1(\mathbf{X}^{-1})$ is defined on $\mathbb{R}$ and depends on $\mathbf{X}^{-1} = (X^2, X^3, ..., X^d)$. Similar sets $\mathbf{A}_1$ can be defined with respect to any direction in the random parameter space and for all measurable $\mathbf{A}$. The failure probability $P(\phi(\mathbf{X}) > S)$ can be written with integrals in the following way :

$$\begin{aligned} P &= \int_{\mathbb{R}^d} \mathbf{1}_{\phi(\mathbf{X})>S} h_0(\mathbf{X}) d\mathbf{X}, \\ &= \int_{\mathbb{R}^d} \mathbf{1}_{\mathbf{X}\in\mathbf{A}} h_0(\mathbf{X}) d\mathbf{X}, \\ &= \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} \mathbf{1}_{X^1\in\mathbf{A}_1} h_0(\mathbf{X}) dX^1 d\mathbf{X}^{-1}. \end{aligned}$$

It can then be rewritten with mathematical expectation over the variable $\mathbf{X}^{-1}$ thanks to the Gaussian assumptions with

$$P = \mathbb{E}\left(P(X^1 \in \mathbf{A}_1 | \mathbf{X}^{-1})\right). \tag{43}$$

The failure probability is described as the expectation of the continuous random variable $P(X^1 \in \mathbf{A}_1)$ relatively to the variable $\mathbf{X}^{-1}$. This expectation is replaced in practice in LS by its Monte-Carlo estimate

$$\hat{P}^{LS} = \frac{1}{N_C} \sum_{i=1}^{N_C} (P(X^1 \in \mathbf{A}_1(\mathbf{X}_i^{-1}))), \tag{44}$$

where $(\mathbf{X}_1^{-1}), ..., (\mathbf{X}_{N_C}^{-1})$ are samples of the random variable $\mathbf{X}^{-1}$. It is still necessary to estimate the probability $P(X^1 \in \mathbf{A}_1(\mathbf{X}_i^{-1}))$, that is

$$P(X^1 \in \mathbf{A}_1(\mathbf{X}_i^{-1})) = \int_{\mathbb{R}} \mathbf{1}_{X^1\in\mathbf{A}_1(\mathbf{X}_i^{-1})} \omega(X^1) dX^1, \tag{45}$$

where ω is a zero-mean standard normal variable. It is possible to show that this integral can be approximated with

Advantages of LS	Drawbacks of LS
Necessary simulation budget restricted	Difficult to apply when the optimal auxiliary density is multimodal
Simple implementation	Necessary transformation on input variables if they are not Gaussian
	Need a priori information on ϕ

Table 7
Advantages and drawbacks of LS.

$$P(X^1 \in \mathbf{A}_1(\mathbf{X}_i^{-1})) \approx \int_{c_i}^{\infty} \omega(\mathbf{X^1}) d\mathbf{X^1}, \tag{46}$$

where c_i is the value of X^1 such that $\phi(c_i, \mathbf{X}_i^{-1}) = S$. This approximation is only valuable if there is only one intersection point between the input failure region and the chosen sampling direction. The variance of LS estimate is always lower or equal to the CMC estimation [95]. Nevertheless, to our knowledge, the logarithmic efficiency of this algorithm has never been provided.

6.2. *Algorithm*

The computational steps of the algorithm are:

(i) Assume $\mathbf{X}$ follows a centered Gaussian PDF. If it is not the case, apply a transformation on $\mathbf{X}$ described in Table 5.

(ii) In the standard normal space, determine the unit important direction vector $\alpha \in \mathbb{R}^d$. It is the direction that enables to reach the curve $S - \phi(\mathbf{X}) = 0$ with the shortest path to the origin. This direction can be found with Monte-Carlo Markov chain methods [98]. To simplify the notations, one assumes that the important direction vector is $\alpha = (1, 0, ..., 0)$. If it is not the case, a rotation has to be applied to the variable $\mathbf{X}$.

(iii) Generate N_C samples $\mathbf{X}_1^{-1}, ..., \mathbf{X}_{N_C}^{-1}$ of the variable $\mathbf{X}^{-1}$ and estimate for each of these samples the probability $P(X^1 \in \mathbf{A}_1(\mathbf{X}_i^{-1}))$ using Eq. 46.

(iv) Estimate the LS probability estimate with

$$\hat{P}^{LS} = \frac{1}{N_C} \sum_{i=1}^{N_C} (P(X^1 \in \mathbf{A}_1(\mathbf{X}_i^{-1}))). \tag{47}$$

A joint use of Monte-Carlo simulations and line sampling, that does not need the knowledge of the direction α has been proposed in [99,100]. It requires nevertheless some *a priori* information on $\phi(.)$ in order to be efficient. The advantages and drawbacks of LS are presented in Table 7.

7. Adaptive splitting technique

7.1. *Principle*

The idea of importance splitting, also called subset sampling, subset simulation or sequential Monte-Carlo, is to decompose the sought probability in a product of conditional probabilities that can be estimated with a reasonable simulation budget. It has firstly been proposed in a physical context in 1951 [101], and numerous variants have been then worked out. Considering the set $\mathbf{A} = \{\mathbf{X} \in \mathbb{R}^d | \phi(\mathbf{X}) > S\}$, the objective of adaptive splitting technique (AST) [102–106] is to determine the probability $P(\mathbf{X} \in \mathbf{A}) = P(\phi(\mathbf{X}) > S)$. For that purpose, the principle of AST [107–113] is to iteratively estimate supersets of $\mathbf{A}$ and then to estimate $P(\mathbf{X} \in \mathbf{A})$ with conditional probabilities. Let us define $\mathbf{A}_0 = \mathbb{R}^d \supset \mathbf{A}_1 \supset ... \supset \mathbf{A}_{n-1} \supset \mathbf{A}_n = \mathbf{A}$, a decreasing sequence of $\mathbb{R}^d$ subsets with smallest element $\mathbf{A} = \mathbf{A}_n$. The probability $P(\mathbf{X} \in \mathbf{A})$ can be then rewritten in the following way:

$$P(\mathbf{X} \in \mathbf{A}) = \prod_{k=1}^{n} P(\mathbf{X} \in \mathbf{A}_k | \mathbf{X} \in \mathbf{A}_{k-1}), \tag{48}$$

where $P(\mathbf{X} \in \mathbf{A}_k | \mathbf{X} \in \mathbf{A}_{k-1})$ is the probability that $\mathbf{X} \in \mathbf{A}_k$ knowing that $\mathbf{X} \in \mathbf{A}_{k-1}$. An optimal choice of the sequence $\mathbf{A}_k$, $k = 0, ..., n$ is given when $P(\mathbf{X} \in \mathbf{A}_k | X \in \mathbf{A}_{k-1}) = \rho$, where ρ is a constant, that is when all the conditional probabilities are equal. The variance of $P(\mathbf{X} \in \mathbf{A})$ is indeed minimized in this configuration as shown in [114,115]. Consequently, if each $P(\mathbf{X} \in \mathbf{A}_k | \mathbf{X} \in \mathbf{A}_{k-1})$ is well estimated, then the probability $P(\mathbf{X} \in \mathbf{A})$ is estimated more accurately with AST than with a direct estimation by Monte-Carlo [116].
Let us define h^k the density of $\mathbf{X}$ restricted to the set $\mathbf{A}_k$. The subset $\mathbf{A}_k$ can be defined with $\mathbf{A}_k = \{\mathbf{X} \in \mathbb{R}^d | \phi(\mathbf{X}) > S_k\}$ for $k = 0, ..., n$ with $S = S_n > S_{n-1} > ... > S_k > ... > S_0$. Determining the sequence $\mathbf{A}_k$ is equivalent to choose some values for S_k, with $k = 0, ..., n$. The values of S_k for $k = 0, ..., n$ can be determined in an adaptive manner to perform valuable results [116] using ρ-quantile of samples generated with the PDF h^k.

7.2. *Algorithm*

The different stages of AST to estimate $P(\phi(\mathbf{X}) > S)$ are the following ones:

(i) Set $k = 0$, $\rho \in]0, 1[$ and $h^0 = h_0$

(ii) Generate N samples $\mathbf{X}_1^{(k)}, ..., \mathbf{X}_N^{(k)}$ from h^k and apply the function ϕ in order to have $Y_1^{(k)} = \phi(\mathbf{X}_1^{(k)}), ..., Y_N^{(k)} = \phi(\mathbf{X}_N^{(k)})$

(iii) Estimate the ρ-quantile $\gamma_\rho^{(k)}$ of the samples $Y_1^{(k)}, ..., Y_N^{(k)}$.

(iv) Determine the subset $\mathbf{A}_{k+1}$ with $\mathbf{A}_{k+1} = \{\mathbf{X} \in \mathbb{R}^d | \phi(\mathbf{X}) > \gamma_\rho^{(k)}\}$ and the conditional density h^{k+1}.

(v) If $\gamma_\rho^{(k)} < S$, set $k \leftarrow k + 1$ and go back to stage (ii). Otherwise, estimate the probability with

$$\hat{P}^{AST} = (1 - \rho)^k \times \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i^{(k)}) > S}.$$

Advantages of AST	Drawbacks of AST
Applicable in high dimensions and non linear systems	Important simulation budget
Efficient on very rare events ($P < 10^{-6}$)	Difficult to apply on non Gaussian inputs

Table 8
Advantages and drawbacks of AST.

Generating directly independent samples from the h^k conditional densities is in most cases impossible as they are usually unknown [102,117]. Nevertheless, AST provides an iterative way to do it, yet in a dependent fashion using a h_0-reversible Markovian kernel $K(\mathbf{X}, \cdot)$. With such a kernel and $\mathbf{X}_k$ following the density h^k, one can distribute random variable Ξ_k according to h^k with the following proposal/refusal method [116]:

$$\Xi_k = \Xi_k(\mathbf{X}_k) = \begin{cases} K(\mathbf{X}^k, \cdot), \text{ if } K(\mathbf{X}^k, \cdot) \in \mathbf{A_k}, \\ \mathbf{X}^k, \qquad \text{otherwise.} \end{cases}$$

This proposal/refusal algorithm enables to generate any number of samples according to h^k in a relative simple manner. It also enables us to keep constant the number of samples to estimate each $P(\mathbf{X} \in \mathbf{A}_{k+1} | \mathbf{X} \in \mathbf{A}_k)$. This operation has to be applied for each density h^k. The generated samples are unfortunately dependent and identically distributed according to h^k. Up to now, there is no way to do this in an independent fashion. However, under mild conditions, it can be shown [117] that applying the proposal/refusal method several times may decrease variance.
The advantages and drawbacks of AST are described in Table 8. AST is often applied to estimate very rare events ($P < 10^{-6}$). For higher probabilities, other simulation methods as IS are more efficient than AST [116]. The logarithmic efficiency has been proved for splitting with fixed levels in [118].

8. CMC inspired methods

Even if CMC is not adapted to rare event estimations, CMC can nevertheless be slightly improved with the use of stratified sampling of Latin hypercube sampling as described the following subsections.

8.1. *Stratified Sampling*

The principle of stratified sampling (SS) is very similar to CMC [119]. The idea is to propose more samples in the input space so that $\mathbf{1}_{\phi(\mathbf{X})>S} = 1$. SS consists thus in partitioning the support of $\mathbf{X}$, defined by $\mathbb{R}^d$ in the general case as proposed in Section 1, in several subsets $\mathbb{Q}_i$, $i = 1, ..., m$ such that $\mathbb{Q}_i \bigcap \mathbb{Q}_j = \emptyset$ for $i \neq j$, and $\bigcup_i \mathbb{Q}_i = \mathbb{R}^d$. One then generates n_i i.i.d. samples $\mathbf{X}_1^i, ..., \mathbf{X}_{n_i}^i$ from the PDF $h_{\mathbb{Q}_i}$ defined with

$$h_{\mathbb{Q}_i}(\mathbf{X}) = \mathbf{1}_{\mathbf{X} \in \mathbb{Q}_i} \frac{h_0(\mathbf{X})}{d_i}, \tag{49}$$

where d_i is defined by

Advantages of SS	Drawbacks of SS
Simple implementation	Necessary information on function ϕ
Potential decrease of CMC relative deviation	Subset definition strongly influences probability estimate accuracy

Table 9
Advantages and drawbacks of stratified sampling.

$$d_i = \int_{\mathbb{Q}_i} h_0(\mathbf{x}) d\mathbf{x}. \tag{50}$$

The required number of samples N in SS is equal to

$$N = \sum_{i=1}^{m} n_i.$$

The SS probability estimate $\hat{P}^{SS}$ is then obtained with

$$\hat{P}^{SS} = \sum_{i=1}^{m} d_i \hat{P}_{h_{\mathbb{Q}_i}}, \tag{51}$$

where $\hat{P}_{h_{\mathbb{Q}_i}}$ is defined as

$$\hat{P}_{h_{\mathbb{Q}_i}} = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{1}_{\phi(\mathbf{X}_j^i)>S}. \tag{52}$$

The relative deviation of $\hat{P}^{SS}$ depends notably on n_i and $h_{\mathbb{Q}_i}$, and is given by the following equation [120]

$$\frac{\sigma_{\hat{P}^{SS}}}{P} = \frac{1}{P} \sqrt{\sum_{i=1}^{m} d_i \frac{P_{h_{\mathbb{Q}_i}}(1 - P_{h_{\mathbb{Q}_i}})}{n_i}}, \tag{53}$$

where $P_{h_{\mathbb{Q}_i}}$ is the true value of $\hat{P}_{h_{\mathbb{Q}_i}}$. If $m = 1$, the previous equation corresponds to the CMC relative deviation given in Eq. 4. The choice of the subsets $\mathbb{Q}_i$ and of n_i is thus very important in order to reduce the Monte-Carlo estimator variance, but requires some information on the input-output function ϕ. If one has no clue on where $\mathbf{1}_{\phi(\mathbf{X})>S} = 1$ in the input space, the method of stratified sampling is not applicable and can increase the Monte-Carlo relative deviation if $\mathbb{Q}_i$ and n_i are not adapted to ϕ. An adaptive version of SS has been proposed in [121]. Table 9 sums up the characteristics of stratified sampling estimator. An extended version of SS called coverage Monte-Carlo method in [122,123] has been proposed for specific systems represented by a fault tree or a network using its minimal cuts to improve the probability estimation. For the same kind of systems, recursive variance reduction methods described in [124,125], have also been proposed and have some links with SS. They are one of the most efficient methods for this application [126].

8.2. *Monte-Carlo method with Latin Hypercube Sampling*

Latin hypercube sampling (LHS) [127–132] can be used instead of stratified sampling when the subsets $\mathbb{Q}_i$ are difficult to estimate. The principle is to stratify in an independent fashion each of the d input dimensions $\mathbf{X} = (X^1, X^2, ..., X^d)$ into N equipossible intervals of probability $\frac{1}{N}$. For a given dimension k, one generates one sample in each interval

Advantages of LHS	Drawbacks of LHS
Simple implementation	Weak potential decrease of CMC relative deviation

Table 10
Advantages and drawbacks of LHS.

according to the conditional joint law of h_0 for the dimension k and thus obtains N scalar samples. The random matching between the scalar samples in the different dimensions enables to obtain a N d-tuple $\mathbf{X}_1, ..., \mathbf{X}_N$ that describes a LHS. The probability with LHS is estimated in the same way as Monte-Carlo with

$$\hat{P}^{LHS} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\phi(\mathbf{X}_i)>S}. \tag{54}$$

This estimate is unbiased and its relative deviation is always lower than CMC [133,134]. The advantages and drawbacks of LHS are described in Table 10. In [135], the use of LHS allows to decrease by $\sqrt{2}$ the relative deviation of the Monte-Carlo method. This reduction is interesting and divides by 2 the computational effort. It is nevertheless possible to obtain a better decrease of the estimate variance with statistic or simulation techniques dedicated to rare event estimation. Some information about the relative error bound of LHS sampling can be found in [15]. The logarithmic efficiency of this algorithm has not been proved.

9. Other simulation algorithms

9.1. *Control Variates*

The control variate method [136,137] is a variance reduction technique used in Monte-Carlo methods. The principle is the following. Let us define the random variable $H = \mathbf{1}_{\phi(\mathbf{X})>S}$. One has $\mathbb{E}(H) = P$ and can define a random variable m such that $\mathbb{E}(m) = \tau$. One can also define the variable H^* so that, given a coefficient c,

$$H^* = H + c(m - \tau). \tag{55}$$

The variable H^* is also an unbiased estimator of P for any choice of the coefficient c. The variance of H^* is given by

$$Var(H^*) = Var(H) + c^2 Var(m) + 2c\ Cov(H, m), \tag{56}$$

where $Cov(H, m)$ is the covariance between H and m. It can be shown that choosing the optimal coefficient c^* defined by

$$c^* = \frac{-Cov(H, m)}{Var(m)}, \tag{57}$$

minimizes the variance of H^*. In that case, the variance H^* is equal to

$$Var(H^*) = (1 - \rho^2) Var(H), \tag{58}$$

where ρ is the correlation coefficient between H and m. Unfortunately, the optimal coefficient c^* is not available and thus, different techniques allow to choose efficient values of c. When the system can be bounded, that is, if one can determine ϕ_L and ϕ_R such

that $\phi_L(\mathbf{X}) < \phi(\mathbf{X}) < \phi_R(\mathbf{X})\ \forall \mathbf{X}$, the use of control variates can decrease the variance of the probability estimate. Such developments have notably been proposed in [138] for fault trees.

9.2. *Antithetic variates*

The antithetic variate (AV) algorithm [52,139] is a variance reduction technique. Let us assume that one has two random variables H_1 and H_2 with the same probability law of $H = \mathbf{1}_{\phi(\mathbf{X})>S}$. One has then

$$\mathbb{E}(H) = \frac{1}{2}(\mathbb{E}(H_1) + \mathbb{E}(H_2)) = \mathbb{E}\left(\frac{H_1 + H_2}{2}\right), \tag{59}$$

and also

$$Var\left(\frac{H_1 + H_2}{2}\right) = \frac{Var(H_1) + Var(H_2) + 2Cov(H_1, H_2)}{4}. \tag{60}$$

If H_1 and H_2 are i.i.d, then $Cov(H_1, H_2) = 0$ and one obtains the same variance as Monte-Carlo estimate. The principle of AV is to obtain samples so that $Cov(H_1, H_2) < 0$. For instance, if X follows a multidimensional normal PDF with mean μ and covariance matrix Σ, then $X' = 2\mu - X$ follows the same law as X. In that case, one can generate $H_1 = \mathbf{1}_{\phi(X)>S}$ and $H_2 = \mathbf{1}_{\phi(X')>S}$ and reduce the variance of the Monte-Carlo estimate on P.
Control and antithetic variates cannot be easily applied in cases where the function ϕ is not known analytically which reduces the potential applicability of these methods. Recent results have thrown an important doubt about their interest [140]. Dagger sampling, described in [141] and more recently in [142], is an extension of antithetic variable method. It improves CMC estimate for specific systems such as networks or fault trees.

10. Use of metamodels in rare event probability estimation

Being able to build an efficient surrogate model which allows to reduce the number of calls to the expensive input-output function ϕ while keeping a good accuracy is a key point in rare event probability estimation. A great number of methods have been proposed and compared in recent years. For the sake of conciseness, in this paper, we do not review all the methods present in the literature which is very profuse on this subject. A survey of the different metamodel methods can be found in [80]. In this section, we present the main surrogate models which have been got underway with importance sampling and Monte-Carlo estimators. Classical deterministic surrogate models such as polynomials, splines have been tested and compared to neural networks and first order reliability method (FORM) [143–145]. Chaos Polynomials have been associated with Monte-Carlo sampling to estimate failure probabilities [146]. Support vector machines have also been employed to estimate the domains of failure [147] and been coupled to rare event estimator such as subset sampling [148].
Kriging method [149–151] presents some advantages in rare event probability estimation. Indeed, this surrogate model is based on a Gaussian process, that allows to estimate the variance of the prediction error and consequently to define a confidence domain of the surrogate model. This indicator can be directly used to refine the model, *i.e.*,

Advantages	Drawbacks
Allow to greatly reduce computation time	Induce approximation errors due to the surrogate model
Allow to use greater simulation budget	Require knowledge on ϕ to build a consistent model especially when $\phi(\mathbf{X}) > S$

Table 11
Advantages and drawbacks of metamodel probability estimate.

to choose new points to evaluate the real function that allow to improve the accuracy of the model. Kriging has been extensively used with classical Monte Carlo estimator [152], Importance sampling method [145,153–155], importance sampling with control variates [156] or subset simulation [157–159]. The way to refine the Kriging model is a key point and different strategies have been proposed [155,160,161] to exploit the complete probabilistic description given by the Kriging to evaluate the minimal number of points on the real expensive input-output function. A numerical comparison of different Kriging based methods to estimate a probability of failure can be found in [162].

The advantages and drawbacks of metamodel-based rare event probability estimators are given in Table 11.

11. Synthesis

The proposed synthesis of this article consists of a series of questions than can help the reader to choose the appropriate methods for his estimation problem.

(i) Is it possible to use the function ϕ to resample? If resampling is not possible, that is if one considers only a fixed set of samples $\phi(\mathbf{X}_1), ..., \phi(\mathbf{X}_N)$, the only available methods are EVT and metamodel probability estimate. If resampling is possible, the other simulation methods presented in this article are more efficient than EVT.

(ii) Is the density of Y or the function ϕ analytically known? If it is the case, then it can be interesting to focus on LDT, exponential twisting, simple changes of importance sampling, control variates and antithetic variates. If these methods are not efficient, then more general algorithms are more complex to implement but should be efficient.

(iii) Is the input region which gives $\phi(\mathbf{X}) > S$ approximately known? If yes, then SS and FORM/SORM/FOSPA are adapted.

(iv) Is the input region which gives $\phi(\mathbf{X}) > S$ multimodal? If yes or if the answer to this question is not known, the use of CE, FORM/SORM/FOSPA is not advised.

(v) What is the dimension d of the problem? If $d < 10$ (value given as an order of magnitude), NAIS, FORM/SORM/FOSPA and LS can be considered. If $d > 10$, AST and CE are the most efficient algorithms.

(vi) What is the available simulation budget N? If $N > 1000$ (value given as an order of magnitude), then CE, NAIS and AST are adapted. If $N < 1000$, FORM/SORM/FOSPA and LS have to be used. CE, NAIS and AST can also be applied when $N < 1000$ but jointly used with a surrogate model.

(vii) Is the function ϕ highly non linear? If it is the case, then FORM/SORM/FOSPA, LS and surrogate model can imply a bias in the estimation and has to be applied carefully whereas AST is adapted.

(viii) Is it possible to prove that the probability estimate has a bounded relative error or is logarithmic efficient ? IS with exponential twisting or with CE optimisation (in a specific context) and AST have been proved to have good robustness properties in certain applications.

Table 12 sums up these different answers. It is often difficult to practically choose the most efficient rare event method for a given problem. Indeed, as described in this article, a large collection of methods is available to estimate rare event probability with more or less accuracy depending on the problem characteristics. The answers to all the previous questions can guide the reader to an appropriate algorithm. An open topic on rare event estimation is the analysis of the robustness properties of the different probability estimates in very general cases. It would ease the comparison of the different algorithms to determine which method could potentially lead to the required simulation budget for a fixed relative error.

	Impossibility of resampling	Density of Y known	ϕ known analytically	ϕ and Y unknown	Region $Y > S$ partially known	Region $Y > S$ disjoint - info not available	$d < 10$	d> 10	$N > 1000$	$N < 1000$	ϕ non linear
AST				√				√	√		√
Anti.Var.		√	√								
CE				√	√	×		√	√		
Cont.Var.		√	√								
EVT	√										
Exp.Tw		√	√								
FORM /SORM				√	√	×				√	×
LDT		√	√								
LS				√		×	√			√	
Surrogates	√				√			×		√	×
NAIS				√	√		√	×	√		
SS					√		√				

Table 12
Synthesis table - √(resp. ×): the method presents some advantages (resp. drawbacks) for the considerate characteristic

12. Acknowledgements

The work of Jérôme Morio has received funding from the European Community's Seventh Framework Program (FP7/2012-2015) under grant agreement number 284802 HIPOW project. The work of Damien Jacquemart is financially supported by DGA (Direction Générale de l'Armement) and Onera. The work of Christelle Vergé is financially supported by CNES (Centre National d'Etudes Spatiales) and Onera. The authors thank the anonymous reviewer for his valuable remarks.

References

[1] P. J. Smith, M. Shafi, H. Gao, Quick simulation: a review of importance sampling techniques in communications systems, Selected Areas in Communications, IEEE Journal on 15 (4) (1997) 597–613.

[2] M. Rocco, Extreme value theory for finance: A survey, Bank of Italy Occasional 99 (2012) 1–72.
[3] S. Juneja, P. Shahabuddin, Chapter 11 rare-event simulation techniques: An introduction and recent advances, in: S. G. Henderson, B. L. Nelson (Eds.), Simulation, Vol. 13 of Handbooks in Operations Research and Management Science, Elsevier, 2006, pp. 291–350.
[4] A. J. Keane, P. B. Nair, Computational Approaches for Aerospace Design, John Wiley & Sons, Ltd, 2005.
[5] J. Morio, R. Pastel, F. Le Gland, Missile target accuracy estimation with importance splitting, Aerospace Science and Technology 25 (1) (2013) 40–44.
[6] I. Banerjee, S. Pal, S. Maiti, Computationally efficient black-box modeling for feasibility analysis, Computers & Chemical Engineering 34 (9) (2010) 1515 – 1521.
[7] A. Grancharova, J. Kocijan, T. A. Johansen, Explicit output-feedback nonlinear predictive control based on black-box models, Engineering Applications of Artificial Intelligence 24 (2) (2011) 388–397.
[8] V. Kreinovich, S. Ferson, A new Cauchy-based black-box technique for uncertainty in risk analysis, Reliability Engineering & System Safety 85 (1-3) (2004) 267–279.
[9] K. Worden, C. Wong, U. Parlitz, A. Hornstein, D. Engster, T. Tjahjowidodo, F. Al-Bender, D. Rizos, S. Fassois, Identification of pre-sliding and sliding friction dynamics: Grey box and black-box models, Mechanical Systems and Signal Processing 21 (1) (2007) 514–534.
[10] H. Cancela, M. El Khadiri, G. Rubino, Rare event analysis by Monte Carlo techniques in static models, John Wiley & Sons, Ltd, 2009, pp. 145–170.
[11] B. W. Silverman, Density estimation for statistics and data analysis, in: Monographs on Statistics and Applied Applied Probability, London: Chapman and Hall, 1986.
[12] G. A. Mikhailov, Parametric Estimates by the Monte Carlo Method, VSP, Utrecht (NED), 1999.
[13] I. M. Sobol, A Primer for the Monte Carlo Method, CRC Press, Boca Raton, Fl., 1994.
[14] C. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, 2005.
[15] H. Niederreiter, J. Spanier, Monte Carlo and Quasi-Monte Carlo Methods, Springer, 2000.
[16] G. S. Fishman, Monte-Carlo: Concepts, Algorithms, and Applications, Springer, New York, 1996.
[17] P. L'Ecuyer, M. Mandjes, B. Tuffin, Importance Sampling in Rare Event Simulation, John Wiley & Sons, Ltd, 2009, pp. 17–38.
[18] P. L'Ecuyer, J. H. Blanchet, B. Tuffin, P. W. Glynn, Asymptotic robustness of estimators in rare-event simulation, ACM Trans. Model. Comput. Simul. 20 (1) (2010) 1–37.
[19] D. P. Kroese, R. Y. Rubinstein, Monte-Carlo methods, Wiley Interdisciplinary Reviews: Computational Statistics 4 (1) (2012) 48–58.
[20] P. Embrechts, C. Kluppelberg, T. Mikosch, Modelling extremal events for insurance and finance, ZOR Zeitschrift for Operations Research Mathematical Methods of Operations Research 97 (1) (1994) 1–34.
[21] S. Kotz, S. Nadarajah, Extreme Value Distributions. Theory and Applications, Imperial College Press, London, 2000.
[22] E. Towler, B. Rajagopalan, E. Gilleland, R. S. Summers, D. Yates, R. W. Katz, Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory, Water Resources Research 46 (11) (2010) 1–11.
[23] J. Blanchet, C. Marty, M. Lehning, Extreme value statistics of snowfall in the Swiss alpine region, Water Resources Research 45 (5) (2009) 1–12.
[24] R. D. Reiss, M. Thomas, Statistical Analysis of Extreme Values, Birkhauser, 1997.
[25] E. Castillo, A. Hadi, J. Sarabia, Extreme Value and Related Models with Applications in Engineering and Science, John Wiley & Sons, New Jersey, 2005.
[26] B. Gnedenko, Sur la distribution limite du terme maximum d'une série aléatoire, Annals of Mathematics 44 (3) (1943) 423–453.
[27] S. Resnick, Extreme values, regular variation and point process, New York, 1987.
[28] J. Pickands, Statistical inference using extreme order statistics, Annals of Statistics 3 (1) (1975) 119–131.
[29] L. de Haan, Slow variation and characterization of domains of attraction, in: J. de Oliveira (Ed.), Statistical Extremes and Applications, Vol. 131 of NATO ASI Series, Springer Netherlands, 1984, pp. 31–48.
[30] M. I. Fraga Alves, M. Ivette Gomes, Statistical choice of extreme value domains of attraction - a comparative analysis, Communications in Statistics - Theory and Methods 25 (4) (1996) 789–811.

[31] H. Drees, L. de Haan, D. Li, Approximations to the tail empirical distribution function with application to testing extreme value conditions, Journal of Statistical Planning and Inference 136 (10) (2006) 3498–3538.
[32] R. B. D'Agostino, M. A. Stephens, Goodness-of-fit Techniques, Vol. 68 of Statistics: a Series of Textbooks and Monographs, Dekker, New York, 1996.
[33] A. L. M. Dekkers, L. De Haan, On the estimation of the extreme-value index and large quantile estimation, The Annals of Statistics 17 (4) (1999) 1795–1832.
[34] S. G. Coles, An introduction to statistical modeling of extreme values, Springer, New York, 2001.
[35] J. Hosking, J. Wallis, Parameter and quantile estimation for the generalized Pareto distribution, Technometrics 29 (3) (1987) 339–349.
[36] C. Neves, M. Fraga Alves, Reiss and Thomas' automatic selection of the number of extremes, Computational Statistics and Data Analysis 47 (4) (2004) 689–704.
[37] J. Geluk, L. de Haan, On bootstrap sample size in extreme value theory, Econometric Institute Research Papers EI 2002-40, Erasmus University Rotterdam, Erasmus School of Economics (ESE), Econometric Institute (2002).
[38] A. Ferreira, L. De Haan, On the block maxima method in extreme value theory submitted.
URL http://arxiv.org/pdf/1310.3222.pdf
[39] S. R. S. Varadhan, Special invited paper: Large deviations, The Annals of Probability 36 (2) (2008) 397–419.
[40] A. Dembo, O. Zeitouni, Large deviations techniques and applications, Springer-Verlag, New York, 1998.
[41] F. den Hollander, Large Deviations, American Mathematical Society, New York, 2008.
[42] P. Dupuis, H. Wang, Importance sampling, large deviations, and differential games, Stochastics and Stochastics Reports 76 (2004) 481–508.
[43] J. S. Sadowsky, Evaluation of large deviation probabilities via importance sampling, in: Signals, Systems and Computers, 1994 Conference Record of the Twenty-Eighth Asilomar Conference on, Vol. 1, 1994, pp. 30–34.
[44] P. Glasserman, Y. Wang, Counter examples in importance sampling for large deviations probabilities, Ann. Appl. Prob. 7 (1997) 731–746.
[45] T. Dean, P. Dupuis, Splitting for rare event simulation: a large deviation approach to design and analysis, Stochastic processes and their applications 119 (2) (2009) 562–587.
[46] P. Del Moral, J. Garnier, Genealogical particle analysis of rare events, The Annals of Applied Probability 15 (2005) 2496–2534.
[47] H. Touchette, The large deviation approach to statistical mechanics, Physics Reports 478 (1-3) (2009) 1–69.
[48] J. Blanchet, H. Lam, State-dependent importance sampling for rare-event simulation: An overview and recent advances, Surveys in Operations Research and Management Science 17 (1) (2012) 38–59.
[49] G. C. Orsak, B. Aazhang, A class of optimum importance sampling strategies, Information Sciences 84 (1-2) (1995) 139–160.
[50] J.-F. Richard, W. Zhang, Efficient high-dimensional importance sampling, Journal of Econometrics 141 (2) (2007) 1385–1411.
[51] S. Engelund, R. Rackwitz, A benchmark study on importance sampling techniques in structural reliability, Structural Safety 12 (4) (1993) 255–276.
[52] J. Hammersley, D. Handscomb, Monte-Carlo methods, Methuen, London, 1964.
[53] R. E. Melchers, Importance sampling in structural systems, Structural Safety 6 (1) (1989) 3–10.
[54] J. A. Bucklew, Introduction to Rare Event Simulation, Springer, 2004.
[55] D. L. McLeish, Bounded relative error importance sampling and rare event simulation, ASTIN Bulletin 40 (2010) 377–398.
[56] S. Juneja, Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error, Queueing Systems 57 (2-3) (2007) 115–127.
[57] P. Dupuis, H. Wang, Dynamic importance sampling for uniformly recurrent Markov chains, The Annals of Applied Probability 15 (1A) (2005) 1–38.
[58] P. Dupuis, A. D. Sezer, H. Wang, Dynamic importance sampling for queueing networks, The Annals of Applied Probability 17 (4) (2007) 1306–1346.
[59] S. T. Tokdar, R. E. Kass, Importance sampling: a review, Wiley Interdisciplinary Reviews: Computational Statistics 2 (1) (2010) 54–60.

[60] R. Rubinstein, D. Kroese, The Cross-Entropy Method : A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning (Information Science and Statistics), Springer, 2004.
[61] T. H. de Mello, R. Y. Rubinstein, Rare event estimation for static models via cross-entropy and importance sampling, John Wiley, New York, 2002.
[62] B. Tuffin, A. Ridder, Probabilistic bounded relative error for rare event simulation learning techniques, in: Simulation Conference (WSC), Proceedings of the 2012 Winter Simulation Conference, 2012, pp. 1–12.
[63] P. Zhang, Nonparametric importance sampling, Journal of the American Statistical Association 91 (434) (1996) 1245–1253.
[64] J. C. Neddermeyer, Non-parametric partial importance sampling for financial derivative pricing, Quantitative Finance 11 (2010) 1193–1206.
[65] J. C. Neddermeyer, Computationally efficient nonparametric importance sampling, Journal of the American Statistical Association 104 (2009) 788–802.
[66] J. Morio, Extreme quantile estimation with nonparametric adaptive importance sampling, Simulation Modelling Practice and Theory 27 (0) (2012) 76–89.
[67] M. P. Wand, M. C. Jones, Kernel Smoothing, Chapman and Hall, New York, 1994.
[68] I. K. Glad, N. L. Hjort, N. G. Ushakov, Mean-squared error of kernel estimators for finite values of the sample size, Journal of Mathematical Sciences 146 (2007) 5977–5983.
[69] H. E. Daniels, Saddlepoint approximations in statistics, Ann. Math. Statist 25 (4) (1954) 631–650.
[70] J. L. Jensen, Saddlepoint Approximations, Oxford University Press, USA, 1995.
[71] S. Huzurbazar, Practical saddlepoint approximations, The American Statistician 53 (3) (1999) 225–232.
[72] C. Goutis, G. Casella, Explaining the saddlepoint approximation, The American Statistician 53 (3) (1999) 216–224.
[73] S. Asmussen, Applied probability and queues, Springer, New York, 2003.
[74] D. Siegmund, Importance sampling in the Monte Carlo study of sequential tests, The Annals of Statistics 4 (4) (1976) 673–684.
[75] A. B. Dieker, M. Mandjes, On asymptotically efficient simulation of large deviation probabilities, Advances in Applied Probability 37 (2) (2005) 539–552.
[76] Z. Yan-Gang, O. Tetsuro, A general procedure for first/second-order reliability method (FORM/SORM), Structural Safety 21 (2) (1999) 95–112.
[77] R. Bjerager, Methods for structural reliability computation, Springer Verlag, New York, 1991, pp. 89–136.
[78] H. . Madsen, S. Krenk, N. C. Lind, Methods of structural safety, Springer-Verlag, Englewood Cliffs, 1986.
[79] T. Lassen, N. Recho, Fatigue Life Analyses of Welded Structures, ISTE Wiley, New York, 2006.
[80] B. Sudret, Meta-models for structural reliability and uncertainty quantification, in: Asian-Pacific Symposium on Structural Reliability and its Applications, Singapore, Singapore, 2012.
[81] A. Nataf, Distribution des distributions dont les marges sont données (in French), Comptes rendus de l'Académie des Sciences 225 (1962) 42–43.
[82] A. Hasofer, N. Lind, An exact and invariant first-order reliability format, Journal of Engineering Mechanics 100 (1974) 111–121.
[83] L. Pei-Ling, A. D. Kiureghian, Optimization algorithms for structural reliability, Structural Safety 9 (3) (1991) 161–177.
[84] M. Rosenblatt, Remarks on a multivariate transformation, Annals of Mathematical Statistics 23 (1952) 470–472.
[85] R. Lebrun, A. Dutfoy, An innovating analysis of the Nataf transformation from the copula viewpoint, Probabilistic Engineering Mechanics 24 (3) (2009) 312–320.
[86] R. Lebrun, A. Dutfoy, A generalization of the Nataf transformation to distributions with elliptical copula, Probabilistic Engineering Mechanics 24 (2) (2009) 172–178.
[87] R. Rackwitz, B. Flessler, Structural reliability under combined random load sequences, Computers and Structures 9 (5) (1978) 489–494.
[88] O. D. et H.O Madsen, Structural Reliability Methods, John Wiley and Sons, New York, 1996.
[89] P. Bjerager, On computation methods for structural reliability analysis, Structural Safety 9 (2) (1990) 79–96.

[90] K. Breitung, Asymptotic approximation for multinormal integrals, Journal of Engineering Mechanics 110 (3) (1984) 357–366.
[91] B. Huanh, X. Du, Probabilistic uncertainty analysis by mean-value first order saddlepoint approximation, Reliability Engineering and System Safety 93 (2) (2008) 325–336.
[92] X. Du, A. Sudjianto, First order saddlepoint approximation for reliability analysis, AIAA Journal 42 (6) (2004) 1199–1207.
[93] Z. Lu, S. Song, Z. Yue, J. Wang, Reliability sensitivity method by line sampling, Structural Safety 30 (6) (2008) 517–532.
[94] Y.-G. Zhao, T. Ono, Moment methods for structural reliability, Structural Safety 23 (1) (2001) 47–75.
[95] P. S. Koutsourelakis, H. J. Pradlwarter, G. I. Schueller, Reliability of sructures in high dimensions, part I algorithms and application, Probabilistic Engineering Mechanics 19 (2004) 409–417.
[96] P. Koutsourelakis, Reliability of structures in high dimensions. part II. theoretical validation, Probabilistic Engineering Mechanics 19 (4) (2004) 419–423.
[97] G. I. Schueller, H. J. Pradlwarter, P. S. Koutsourelakis, A critical appraisal of reliability estimation procedures for high dimensions, Probabilistic Engineering Mechanics 19 (2004) 463–474.
[98] C. A. Schenk, H. J. Pradlwarter, G. I. Schuëller, Realistic and efficient reliability estimation in space engineering, in: Proceedings of the 16th ASCE Engineering Mechanics Conference, Seattle, 2003.
[99] P. Bjerager, Probability integration by directional simulation, Journal of Engineering Mechanics 114 (8) (1988) 1288–1302.
[100] J. Nie, B. R. Ellingwood, A new directional simulation method for system reliability. part II: application of neural networks, Probabilistic Engineering Mechanics 19 (4) (2004) 437–447.
[101] H. Kahn, T. Harris, Estimation of particle transmission by random sampling, Appl. Math. Ser. 12 (1951) 27–30.
[102] F. Cérou, P. Del Moral, T. Furon, A. Guyader, Rare event simulation for a static distribution, Vol. 7, RESIM, 2008, pp. 107–115.
[103] F. Cérou, A. Guyader, Adaptive particle techniques and rare event estimation, in: Proceedings of ESAIM conference, Vol. 19, 2007, pp. 65–72.
[104] P. L'Ecuyer, V. Demers, B. Tuffin, Splitting for rare event simulation, in: Proceedings of the 2006 Winter Simulation Conference, 2006, pp. 137–148.
[105] M. Villén-Altamirano, J. Villén-Altamirano, Restart: a straightforward method for fast simulation of rare events, 1994, pp. 282–289.
[106] P. Glasserman, P. Heidelberger, P. Shahabuddin, T. Zajic, Splitting for rare event simulation: analysis of simple cases, in: Proceeding of the 1996 Winter Simulation Conference, 1996, pp. 302–308.
[107] M. N. Rosenbluth, A. W. Rosenbluth, Monte-Carlo calculation of the average extension of molecular chains, J. Chem. Phys 356 (1955) 356–359.
[108] P. L'Ecuyer, F. Le Gland, P. Lezaud, B. Tuffin, Splitting Techniques, John Wiley & Sons, Ltd, 2009, pp. 39–61.
[109] P. Del Moral, Feynman-Kac Formulae, Genealogical and Interacting Particle Systems with Applications. Probability and its Applications., Springer, New York, 2004.
[110] S. K. Au, Reliability-based design sensitivity by efficient simulation, Computers and Structures 83 (2005) 1048–1061.
[111] S. K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulations, Probabilistic Engineering Mechanics 16 (4) (2001) 263–277.
[112] Z. I. Botev, D. P. Kroese, Efficient Monte-Carlo simulation via the generalized splitting method., Statistics and Computing 22 (1) (2012) 1–16.
[113] Z. I. Botev, D. P. Kroese, An efficient algorithm for rare-event probability estimation, combinatorial optimization, and counting, Methodol. Comput. Appl. Probab. 10 (4) (2012) 471–505.
[114] A. Lagnoux, Rare event simulation, Probability in the Engineering and Informational science 20 (2006) 45–66.
[115] F. Cérou, P. Del Moral, F. Le Gland, P. Lezaud, Genetic genealogical models in rare event analysis, INRIA report 1797 (2006) 1–30.
[116] F. Cérou, P. Del Moral, T. Furon, A. Guyader, Sequential Monte-Carlo for rare event estimation, Statistics and Computing 22 (2012) 795–808.

[117] L. Tierney, Markov chains for exploring posterior distributions, Annals of Statistics 22 (1994) 1701–1762.
[118] F. Cérou, P. Del Moral, A. Guyader, A nonasymptotic theorem for unnormalized Feynman-Kac particle models, Ann. Inst. H. Poincaré Probab. Statist. 47 (3) (2011) 629–649.
[119] W. G. Cochran, Sampling Techniques, Wiley, New York, 1977.
[120] M. Keramat, R. Kielbasa, A study of stratified sampling in variance reduction techniques for parametric yield estimation, Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on 45 (5) (1998) 575–583.
[121] M. M. Zuniga, J. Garnier, E. Remy, E. de Rocquigny, Adaptive directional stratification for controlled estimation of the probability of a rare event, Reliability Engineering & System Safety 96 (12) (2011) 1691 – 1712.
[122] R. M. Karp, M. G. Luby, A new Monte-Carlo method for estimating the failure probability of an n-component system, Tech. Rep. UCB/CSD-83-117, EECS Department, University of California, Berkeley (1983).
[123] H. Kumamoto, T. Tanaka, K. Inoue, A new Monte Carlo method for evaluating system-failure probability, Reliability, IEEE Transactions on R-36 (1) (1987) 63–69.
[124] H. Cancela, M. El Khadiri, A recursive variance-reduction algorithm for estimating communication-network reliability, Reliability, IEEE Transactions on 44 (4) (1995) 595–602.
[125] H. Cancela, M. El Khadiri, The recursive variance-reduction simulation algorithm for network reliability evaluation, Reliability, IEEE Transactions on 52 (2) (2003) 207–212.
[126] D. P. Kroese, T. Taimre, Z. I. Botev, Handbook of Monte Carlo Methods, John Wiley & Sons, Hobokem, USA, 2011.
[127] M. D. McKay, R. J. Beckman, W. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (1979) 239 – 245.
[128] M. D. MacKay, Latin hypercube sampling as a tool in uncertainty analysis of computer models, in: WSC '92 Proceedings of the 24th conference on Winter simulation, ACM, New York, 1992, pp. 557–564.
[129] R. L. Inman, J. C. Helson, J. E. Campbell, An approach to sensitivity analysis of computer models: Part I - introduction, input variable selection and preliminary variable assessment, Journal of Quality Technology 13 (3) (1981) 174–183.
[130] Z. Keqin D., Zegong, L. Chuntu, Latin hypercube sampling used in the calculation of the fracture probability, Reliability Engineering & System Safety 59 (2) (1998) 239–242.
[131] B. Ayyub, L. Kwan-Ling, Structural reliability assessment using Latin hypercube sampling, in: Proceedings of the International Conference on Structural Safety and Reliability, ICOSSAR'89, San Francisco, USA, 1989, pp. 174–184.
[132] P. Zhang, P. Breitkopf, C. Knopf-Lenoir, W. Zhang, Diffuse response surface model based on moving latin hypercube patterns for reliability-based design optimization of ultrahigh strength steel nc milling parameters, Struct. Multidiscip. Optim. 44 (5) (2011) 613–628.
[133] M. Keramat, R. Kielbasa, Modified latin hypercube sampling Monte-Carlo (MLHSMC) estimation for average quality index, Analog Integr. Circuits Signal Process. 19 (1) (1999) 87–98.
[134] M. Keramat, R. Kielbasa, Worst case efficiency of Latin hypercube sampling Monte-Carlo (LHSMC) yield estimator of electrical circuits, in: Proc. IEEE Int. Symp. Circuits Syst., Hong Kong, 1997, pp. 1660–1663.
[135] G. Olsson, A. Sandberg, O. Dahlblom, On latin hypercube sampling for structural reliability analysis, Structural Safety 25 (1) (2003) 47–68.
[136] S. P. Meyn, Control Techniques for Complex Networks, Cambridge University Press, 2007.
[137] P. Glasserman, Monte-Carlo Methods in Financial Engineering, Springer, New York, 2003.
[138] H. Kumamoto, K. Tanaka, K. Inoue, Efficient evaluation of system reliability by Monte Carlo method, Reliability, IEEE Transactions on R-26 (5) (1977) 311–315.
[139] P. K. Sarkar, M. A. Prasad, Variance reduction in Monte-Carlo radiation transport using antithetic variates, Annals of Nuclear Energy 19 (5) (1992) 253–265.
[140] E. Saliby, R. J. Paul, A farewell to the use of antithetic variates in monte carlo simulation., JORS 60 (7) (2009) 1026–1035.
[141] H. Kumamoto, K. Tanaka, K. Inoue, E. J. Henley, Dagger-sampling Monte Carlo for system unavailability evaluation, Reliability, IEEE Transactions on R-29 (2) (1980) 122–125.

[142] S. Rongfu, S. Chanan, C. Lin, S. Yuanzhang, Short-term reliability evaluation using control variable based dagger sampling method, Electric Power Systems Research 80 (6) (2010) 682 – 689.
[143] C. Bucher, U. Bourgund, A fast and efficient response surface approach for structural reliability problems, Structural Safety 7 (1990) 57–66.
[144] H. Gomes, A. Awruch, Comparison of response surface and neural network with other methods for structural reliability analysis, Structural Safety 26 (2004) 49–67.
[145] L. Schueremans, D. Van Gemert, Use of Kriging as Meta-model in simulation procedures for structural reliability, in: 9th International conference on structural safety and reliability, Rome, 2005, pp. 2483–2490.
[146] J. Li, D. Xiu, Evaluation of failure probability via surrogate models, Journal of Computational Physics 229 (2010) 8966–8980.
[147] A. Basudhar, S. Missoum, A. Sanchez, Limit state function identification using Support Vector Machines for discontinuous responses and disjoint failure domains, Probabilistic Engineering Mechanics 23 (2008) 1–11.
[148] J.-M. Bourinet, F. Deheeger, M. Lemaire, Assessing small failure probabilities by combined subset simulation and support vector machines, Structural Safety 33 (2011) 343–353.
[149] G. Matheron, Principles of geostatistics, Economic Geology 58 (8) (1963) 1246.
[150] M. K. Sasena, Flexibility and efficienvy enhancements for constrained global design optimization with Kriging approximation, Ph.D. thesis, University of Michigan (2002).
[151] T. J. Santner, B. J. Williams, N. I. Notz, The design and analysis of computer experiments, 2003.
[152] B. Echard, N. Gayton, M. Lemaire, AK-MCS : An active learning reliability method combining Kriging and Monte-Carlo Simulation, Structural Safety 33 (2011) 145–154.
[153] J. Janusevskis, R. Le Riche, Simultaneous Kriging-based estimation and optimization of mean response, Journal of Global Optimization 55 (2) (2012) 313–336.
[154] M. Balesdent, J. Morio, J. Marzat, Kriging-based adaptive importance sampling algorithms for rare event estimation, Structural Safety 13 (2013) 1–10.
[155] V. Dubourg, E. Deheeger, B. Sudret, Metamodel-based importance sampling for the simumation of rare events, in: Faber, M. J. Kohler and K. Nishilima (Eds.), Proceedings of the 11th International Conference of Statistics and Probability in Civil Engineering (ICASP2011), Zurich, Switzerland, 2011.
[156] C. Cannamela, J. Garnier, B. Iooss, Controlled stratification for quantile estimation, Annals of Applied Stats 2 (4) (2008) 1554–1580.
[157] E. Vazquez, J. Bect, A Sequential Bayesian algorithm to estimate a probability of failure, in: 15th IFAC, Symposium on System Identification (SYSID'09), Saint-Malo, France, July 6-8, 2009.
[158] L. Li, J. Bect, E. Vazquez, Bayesian Subset Simulation: a Kriging-based subset simulation algorithm for the estimation of small probabilities of failure, in: Proceedings of PSAM 11 and ESREL 2012, 25-29 June 2012, Helsinki, Finland, 2012.
[159] J. Bect, D. Ginsbourger, L. Li, V. Picheny, E. Vazquez, Sequential design of computer experiments for the estimation of a probability of failure, Statistics and Computing 22 (3) (2012) 773–793.
[160] V. Picheny, Improving accuracy and compensating for uncertainty in surrogate modeling, Ph.D. thesis, University of Florida (2009).
[161] V. Baudoui, P. Klotz, J.-B. Hiriart-Urruty, S. Jan, F. Morel, LOcal Uncertainty Processing (LOUP) method for multidisciplinary robust design optimization, Structural and Multidisciplinary Optimization 46 (5) (2012) 1–16.
[162] L. Li, J. Bect, E. Vazquez, A numerical comparison of two sequential Kriging-based algorithms to estimate a probability of failure, in: Uncertainty in Computer Model Conference, Sheffield, UK, July, 12-14, 2010.