



HAL
open science

A nonparametric importance sampling estimator for moment independent importance measures

Pierre Derennes, Jérôme Morio, Florian Simatos

► **To cite this version:**

Pierre Derennes, Jérôme Morio, Florian Simatos. A nonparametric importance sampling estimator for moment independent importance measures. *Reliability Engineering and System Safety*, 2019, 187, pp.3-16. 10.1016/j.ress.2018.02.009 . hal-02133955

HAL Id: hal-02133955

<https://hal.science/hal-02133955v1>

Submitted on 20 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A nonparametric importance sampling estimator for moment independent importance measures

Pierre Derennes^{a,b}, Jérôme Morio^b, Florian Simatos^c

^a*Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse*

^b*ONERA - The French Aerospace Lab, BP 74025, 31055 Toulouse Cedex, France*

^c*ISAE-SUPAERO and Université de Toulouse, Toulouse, France*

Abstract

Moment independent importance measures have been proposed by E. Borgonovo [1] in order to alleviate some of the drawbacks of variance-based sensibility indices. They have gained increasing attention over the last years but their estimation remains a challenging issue. An effective estimation scheme in the case of correlated inputs, referred to as single-loop method, has been proposed by Wei et al. [2]. In this paper we show via simulation that this method may be inaccurate, making for instance 40% error in the simplest possible Gaussian case. We then propose a new estimation scheme which greatly improves the accuracy of the single-loop method, up to a factor 10 in some simple numerical examples. We prove that our estimator is strongly consistent and several simulation results are presented to demonstrate the advantages of the proposed method.

Keywords: Monte Carlo simulation, Importance Sampling, density-based sensitivity analysis, importance measures.

1. Introduction

The reliability and safety analysis of large and complex systems is a subject of current interest in various fields such as engineering [3, 4], data processing [5], aeronautics [6] or nuclear domain [7]. This analysis may be performed by quantitative methods which aim to provide estimation about a failure probability or any indicator of an undesired state or behavior of the system.

In this paper, we focus on the study of input-output models where the output is a deterministic function, called “black-box”, of the input assumed to be random. In this context, the following fundamental question formulated by Saltelli [8] arises: *how does uncertainty in the different inputs X_i contribute to the uncertainty in the model output?* This type of analysis, which belongs to the *sensitivity analysis* framework, presents two main objectives: on the one hand, to identify the most influential inputs, that we may then for instance seek to know with the greatest possible accuracy in order to reduce the output variability; and on the other hand, to determine non-influential inputs which for instance then makes it possible to decrease the model complexity.

There are essentially two frameworks for sensitivity analysis: local sensitivity analysis and global sensitivity analysis, see [9] and the associated references for a review of these methods. The local approach corresponds to the assessment of the local impact of an input on the model’s response by concentrating on the sensitivity in the vicinity of a set of nominal values. It may be defined as the partial derivatives of the model output. In

Email addresses: pierre.derennes@onera.fr (Pierre Derennes), jerome.morio@onera.fr (Jérôme Morio), florian.simatos@isae.fr (Florian Simatos)

contrast, global sensitivity analysis methods consider the whole variation range of the inputs: there are various techniques such as screening methods, graphical and smoothing tools, variance-based and moment-independent methods. The interested reader can refer to [9, 10] and the associated references for a review of these methods. Variance-based importance measures [11, 12] are one of the most widely used importance measures. They are based on Sobol’s indices which express the share of variance of the output that is due to a given input or input combination. However, this method focuses on the second-order moment of the output distribution which is not always sufficient to represent the entire variability of the distribution, as illustrated by Borgonovo [13]. To overcome this drawback, Borgonovo [1] proposed distribution-based sensitivity indices that are currently gaining increasing attention [14]. However, estimating these indices while minimizing the number of calls to the model response raises a challenging problem because these indices involve L_1 norms of differences of conditional and unconditional probability density functions (PDFs) of the output.

In the original article, Borgonovo proposed a PDF-based method based on a double loop Monte Carlo procedure coupled with kernel density estimation (KDE) which leads to a significant computational budget. Several other methodologies have been recently developed. A pseudo-double loop design involving a partition of the input spaces is proposed in [15]. Liu and Homma [16] propose to express Borgonovo’s indices in terms of unconditional and conditional cumulative distributions functions of the model output. However, as pointed by Liu *”for a computationally intensive model, when the total computational time is mainly due to the time of running the model, the improvement of the computational efficiency by the CDF-based method can be negligible”*. Furthermore, the density estimation cost is replaced by the necessity to find the intersection points of the unconditional and conditional PDFs of the output which leads to additional computational time and approximation errors. We also recall other recent works such as [17, 18] dealing with copula-based method sensitivity analysis, [19] which describes a new method combining the principle of fractional moment-based maximum entropy and the use of Nataf transformation but only suitable for models with independent inputs and [20] where the use of the Edgeworth series is proposed to avoid the density estimation. The present article focuses on PDF-based methods. In [2] a single-loop and a double-loop Monte Carlo simulation schemes adopting the KDE technique have been proposed and have already begun to be implemented in practical cases, see for instance [21, 22, 23]. The single-loop method has the following advantages: it is suitable in the case of correlated inputs and it substantially improves the computational burden in comparison to the PDF-based method initially introduced.

The present paper proposes a new method, close to the single-loop Monte Carlo simulation method proposed in [2], which improves the accuracy of the estimation of Borgonovo’s indices. After recalling some definitions and facts about Borgonovo’s indices in Section 2, we show in Section 3 through numerical simulations on a simple toy case that the single-loop method of [2] may provide inaccurate estimates, and we also discuss potential explanations for this inaccuracy. Our new estimation scheme is presented in Section 4 where some of its theoretical properties, and in particular its consistency, are derived. Section 5 presents numerical results which highlight the increase in accuracy of our method and Section 6 gives the conclusion of this work.

2. Borgonovo’s moment independent importance measure

Throughout the paper we consider a general input-output model $Y = \Phi(\mathbf{X})$ where the output Y depends on a d -dimensional real valued random variable $\mathbf{X} = (X_1, \dots, X_d)$

through a deterministic scalar function $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ called “black box” or “model response”. For $I \subset \{1, \dots, d\}$ a subset of indices we write $\mathbf{X}_I = (X_i, i \in I)$. We assume throughout that for every $I \subset \{1, \dots, d\}$ a strict subset (i.e., $I \neq \{1, \dots, d\}$), the pair (\mathbf{X}_I, Y) is absolutely continuous with respect to Lebesgue measure with PDF $f_{\mathbf{X}_I, Y}$. This implies in particular that the random variables \mathbf{X}_I , Y and Y conditioned on $\mathbf{X}_I = \mathbf{x}_I$ for any $I \subset \{1, \dots, d\}$ a strict subset, and any $\mathbf{x}_I \in \mathbb{R}^{\text{Card}(I)}$ are also absolutely continuous with respect to Lebesgue measure, and we will denote by $f_{\mathbf{X}_I}$, f_Y and $f_Y^{\mathbf{X}_I = \mathbf{x}_I}$ their respective PDFs.

The idea of Borgonovo’s global sensitivity analysis method introduced in [1] is to measure how fixing the input X_i at a value x_i modifies the entire distribution of the output Y . In [1], this modification of the output distribution is quantified by the *shift* $s(x_i)$ defined as the L_1 norm between f_Y and $f_Y^{X_i = x_i}$:

$$s(x_i) = \frac{1}{2} \left\| f_Y - f_Y^{X_i = x_i} \right\|_{L^1(\mathbb{R})} = \frac{1}{2} \int \left| f_Y(y) - f_Y^{X_i = x_i}(y) \right| dy . \quad (1)$$

So as to consider the whole range of values the random variable X_i can take into account, the sensibility of the output Y with respect to the input X_i is defined by averaging the shift over X_i , i.e., Borgonovo’s index is given by

$$\delta_i := \mathbb{E} [s(X_i)] . \quad (2)$$

For further references and more details on properties of Borgonovo’s indices the reader can for instance consult [1, 24].

Remark 1. *This definition can be generalized to a strict group of inputs $I \subset \{1, \dots, d\}$ by*

$$\delta_I := \mathbb{E} [s(\mathbf{X}_I)] \quad \text{with} \quad s(\mathbf{x}_I) = \frac{1}{2} \left\| f_Y - f_Y^{\mathbf{X}_I = \mathbf{x}_I} \right\|_{L^1(\mathbb{R})} = \frac{1}{2} \int \left| f_Y(y) - f_Y^{\mathbf{X}_I = \mathbf{x}_I}(y) \right| dy .$$

For ease of exposition we restrict our attention throughout the paper to the case of singletons, but all our results can be generalized to groups of inputs.

3. Estimation of Borgonovo’s indices

Estimating Borgonovo’s indices is a challenging task because of the unknown unconditional and conditional PDF f_Y and $f_Y^{X_i = x_i}$ that intervene in a convoluted way (i.e., through an L_1 -norm) in their definitions (1) and (2). The present article focuses on a single-loop Monte Carlo simulation method available in [2] which presents several advantages: it is notably effective for models with correlated inputs and implies fewer call to the model than the double-loop designs. This method has received increasing attention and has begun to be used in practice, see for instance [21, 22, 23].

3.1. The single-loop Monte Carlo simulation

The method that we propose here is close to the single-loop method of [2] which we recall first. The main idea behind this method is to re-interpret Borgonovo’s index δ_i as an L_1 -difference between the joint distribution $f_{X_i, Y}$ and the density of the random variables X_i and Y if they were independent. Indeed, from (1) and (2) it follows immediately that

$$\begin{aligned} \delta_i &= \frac{1}{2} \int f_{X_i}(x) \left(\int \left| f_Y(y) - f_Y^{X_i = x}(y) \right| dy \right) dx \\ &= \frac{1}{2} \int f_{X_i}(x) \left| f_Y(y) - \frac{f_{X_i, Y}(x, y)}{f_{X_i}(x)} \right| dx dy \\ &= \frac{1}{2} \| f_{X_i} f_Y - f_{X_i, Y} \|_{L_1(\mathbb{R}^2)} . \end{aligned}$$

This interpretation opens the way for various estimation procedures of δ_i , one of them being the single-loop Monte Carlo estimator based on rewriting the above expression as

$$\delta_i = \frac{1}{2} \mathbb{E} \left[\left| \frac{f_{X_i}(X_i) f_Y(Y)}{f_{X_i, Y}(X_i, Y)} - 1 \right| \right]. \quad (3)$$

where the (X_i, Y) are i.i.d. drawn according to $f_{X_i, Y}$.

Step SL1. Generate $(\mathbf{X}^1, \dots, \mathbf{X}^N)$ i.i.d. with common distribution \mathbf{X} , and then obtain N observations of the model by $Y^k = \Phi(\mathbf{X}^k)$ for $k = 1, \dots, N$.

Step SL2. Use the sample (\mathbf{X}^k, Y^k) to estimate the PDFs f_Y and $f_{X_i, Y}$ by kernel density estimation (KDE):

$$\hat{f}_Y(y) := \frac{1}{Nh} \sum_{k=1}^N K\left(\frac{y - Y^k}{h}\right), \quad y \in \mathbb{R}, \quad (4)$$

and

$$\hat{f}_{X_i, Y}(x, y) := \frac{1}{Nh_1 h_2} \sum_{k=1}^N K\left(\frac{x - X_i^k}{h_1}\right) K\left(\frac{y - Y^k}{h_2}\right), \quad (x, y) \in \mathbb{R}^2, \quad (5)$$

where K is the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$ and where the bandwidths h , h_1 and h_2 are estimated with the diffusion-based method proposed in [25]. This method chooses the bandwidth parameters optimally without ever using or assuming a parametric model for the data or any “rules of thumb”.

Step SL3. Estimate δ_i by

$$\hat{\delta}_i^{SL} = \frac{1}{2N} \sum_{k=1}^N \left| \frac{f_{X_i}(X_i^k) \hat{f}_Y(Y^k)}{\hat{f}_{X_i, Y}(X_i^k, Y^k)} - 1 \right| = \frac{1}{2N} \sum_{k=1}^N \left| \frac{f_{X_i}(X_i^k) \hat{f}_Y(Y^k) - \hat{f}_{X_i, Y}(X_i^k, Y^k)}{\hat{f}_{X_i, Y}(X_i^k, Y^k)} \right|. \quad (6)$$

This method combines one Monte Carlo loop and KDE procedure and it requires only N calls to the black-box function Φ for the estimation of all the indices.

3.2. Application to a Gaussian toy case

In this section, we apply the single-loop Monte Carlo scheme on a simple example of black-box system for which the unconditional and conditional PDFs of the output are known. We assume that the model output is given by

$$Y = X_1 + X_2,$$

where the input variables are independent and follow normal distribution $N(0, 1)$ and $N(0, 5)$ respectively, the second argument corresponding to the variance. In this case the unconditional and conditional output distributions are respectively given by

$$Y \sim N(0, 6), \quad (Y | X_1 = x_1) \sim N(x_1, 5) \quad \text{and} \quad (Y | X_2 = x_2) \sim N(x_2, 1),$$

and so the moment independent measures δ_i are given by

$$\delta_1 = \frac{1}{4\pi} \int_{\mathbb{R}^2} e^{-\frac{x^2}{2}} \left| \frac{1}{\sqrt{6}} e^{-\frac{y^2}{12}} - \frac{1}{\sqrt{5}} e^{-\frac{(y-x)^2}{10}} \right| dx dy,$$

and

$$\delta_2 = \frac{1}{4\pi} \int_{\mathbb{R}^2} e^{-\frac{x^2}{10}} \left| \frac{1}{\sqrt{6}} e^{-\frac{y^2}{12}} - \frac{1}{\sqrt{5}} e^{-\frac{(y-x)^2}{2}} \right| dx dy.$$

Table 1. Single-loop estimates of the importance measures δ_i in the Gaussian toy case. The sample size is $N = 5,000$ and the mean and standard deviation of the single-loop estimators $\hat{\delta}_i^{SL}$ are estimated with 100 simulations.

Input	Theoretical value δ_i	$\bar{\delta}_i^{SL}$	Relative difference	$\bar{\sigma}_i^{SL}$	$\text{CV}(\hat{\delta}_i^{SL})$
X_1	0.1436	0.1217	-0.1522	0.0055	0.0449
X_2	0.5382	0.3297	-0.3873	0.0046	0.0138

Using numerical integration, we get the approximations $\delta_1 \approx 0.1436$ and $\delta_2 \approx 0.5382$. We may compare them with the single-loop estimates $\hat{\delta}_1^{SL}$ and $\hat{\delta}_2^{SL}$ of both importance measures δ_1 and δ_2 . In [2], the sample size N is set to be $N = 3,000$ for all the numerical applications.

As an indicator of the accuracy of the single-loop estimator $\hat{\delta}_i^{SL}$, we consider the coefficient of variation $\text{CV}(\hat{\delta}_i^{SL})$ given by the ratio of the standard deviation to the mean, i.e.,

$$\text{CV}(\hat{\delta}_i^{SL}) = \frac{\sqrt{\text{Var}(\hat{\delta}_i^{SL})}}{\mathbb{E}(\hat{\delta}_i^{SL})}.$$

Since the true values of the standard deviation and the mean of $\hat{\delta}_i^{SL}$ are unknown, in the subsequent analysis we replace them by their estimations obtained with regular Monte Carlo methods with 100 simulations. We will write $\bar{\delta}_i^{SL}$ for the estimation of $\mathbb{E}(\hat{\delta}_i^{SL})$ and $\bar{\sigma}_i^{SL}$ for the estimation of $\sqrt{\text{Var}(\hat{\delta}_i^{SL})}$, so that the estimation of $\text{CV}(\hat{\delta}_i^{SL})$ is given by $\frac{\bar{\delta}_i^{SL}}{\bar{\sigma}_i^{SL}}$ (see the beginning of Section 5).

Table 1 reports numerical results obtained with $N = 5,000$ and $m = 100$ simulations to compute the estimate of $\text{CV}(\hat{\delta}_i^{SL})$. With coefficients of variation estimated around 1% and 5%, it may be thought that $\hat{\delta}_i^{SL}$ has converged and should thus provide an accurate estimate of δ_i . This intuition is corroborated by Figure 1 which displays one trajectory of both single-loop Monte Carlo estimators as the sample size N varies: in both cases, the trajectory becomes flat and seems to have converged.

However, these estimators present a significant bias. Indeed, since the theoretical values of the importance measures δ_i are available, the $(1 - \alpha)\%$ confidence interval of $\mathbb{E}(\hat{\delta}_i^{SL})$

$$\left[\bar{\delta}_i^{SL} - \frac{z_{1-\alpha/2} \bar{\sigma}_i^{SL}}{\sqrt{m}}, \bar{\delta}_i^{SL} + \frac{z_{1-\alpha/2} \bar{\sigma}_i^{SL}}{\sqrt{m}} \right],$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -quantile of the standard normal distribution and the relative difference

$$\frac{\bar{\delta}_i^{SL} - \delta_i}{\delta_i},$$

may be computed in order to appreciate the bias of the single-loop estimator $\hat{\delta}_i^{SL}$. We can note that both single-loop estimators have non negligible relative differences of the order of 15% and 40%, notably $\bar{\delta}_2^{SL}$ which strongly underestimates the measure δ_2 . Furthermore the reference values [0.1436 0.5382] do not belong to the 95% confidence intervals

$$\left[0.1217 \pm \frac{1.96 \times 0.0055}{\sqrt{100}} \right] = [0.1206, 0.1228],$$

and

$$\left[0.3297 \pm \frac{1.96 \times 0.0046}{\sqrt{100}} \right] = [0.3288, 0.3306].$$

If the single-loop Monte Carlo estimates may show significant differences in a simple model case where the KDE procedure is very efficient since the unconditional and conditional output PDFs are Gaussian, the question of the reliability of the estimation for complex models may be asked. In particular, we stress that because the sample (\mathbf{X}^k, Y^k) is used both in the Monte Carlo step **SL3** and also in the KDE step **SL2**, it is not clear at all that the estimator $\hat{\delta}_i^{SL}$ is consistent, i.e., that $\hat{\delta}_i^{SL} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \delta_i$. Actually, we believe that the simulation results presented in this section legitimately casts doubt on this claim, which would require further investigation.

We now present our new estimation scheme, which is provably consistent and also exhibits more accurate results presented in Section 5.

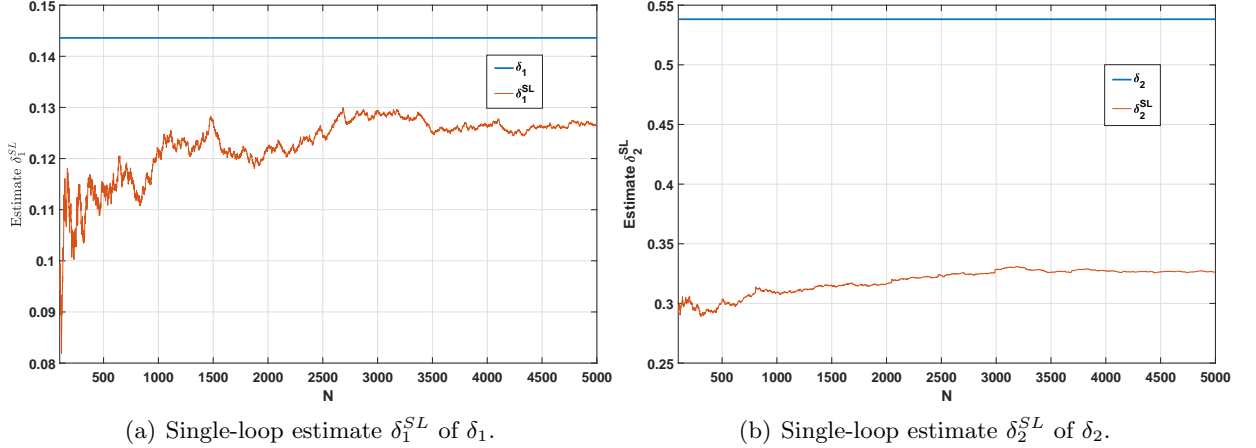


FIGURE 1. SINGLE-LOOP ESTIMATES OF THE IMPORTANCE MEASURES δ_i OF THE GAUSSIAN TOY CASE AS THE SAMPLE SIZE N VARIES.

4. Nonparametric importance sampling estimator

The estimation scheme that we propose will be exposed in Section 4.4 below. To explain the main idea, we first present in Section 4.1 a general importance sampling scheme based on an arbitrary auxiliary function g (see for instance [26] and the associated references). The theoretical properties of this scheme are analyzed in Section 4.2 while the influence of the auxiliary function g and various natural choices are discussed in Section 4.3.1. Once these preliminary results are set up, we finally propose our scheme in Section 4.4: it amounts to using a particular auxiliary function g in the general importance sampling scheme of Section 4.1.

4.1. Importance sampling estimator

We now present a general importance sampling estimator of δ_i based on its expression (3).

Step IS1. Same as step **SL1**, i.e., generate $(\mathbf{X}^1, \dots, \mathbf{X}^N)$ i.i.d. with common distribution \mathbf{X} , and then obtain N observations of the model by $Y^k = \Phi(\mathbf{X}^k)$ for $k = 1, \dots, N$.

Step IS2. Same as step **SL2**, in particular get the two estimators $\hat{f}_{\mathbf{X}, Y}$ and \hat{f}_Y from (\mathbf{X}^k, Y^k) as in (4) and (5).

Step IS3. Let g be any sampling distribution on \mathbb{R}^2 which is allowed to depend on the sample (\mathbf{X}^k, Y^k) . Let $(\mathbf{U}^1, \dots, \mathbf{U}^{N'})$ be N' i.i.d. random variables drawn according to g

with $\mathbf{U}^k = (U_1^k, U_2^k) \in \mathbb{R}^2$. Our estimator $\hat{\delta}_i^{IS,g}$ of δ_i is

$$\hat{\delta}_i^{IS,g} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{\left| \hat{f}_Y(U_2^k) f_{X_i}(U_1^k) - \hat{f}_{X_i,Y}(\mathbf{U}^k) \right|}{g(\mathbf{U}^k)}.$$

The study of the theoretical properties of $\hat{\delta}_i^{IS,g}$ will be carried over in the next section, but we immediately make two important remarks:

1. the proposed approach keeps the advantages of the single-loop estimator, namely only N model evaluations are needed to compute all Borgonovo's importance measures and it can be applied to models with correlated inputs;
2. we stress that the error induced by the Monte Carlo estimation in step **IS3** is inexpensive, in the sense that it can be made arbitrarily small without further call to the possibly expensive black-box function Φ . Thus essentially, the only estimation error is due to the kernel approximation of step **IS2**.

Furthermore, note that g is an arbitrary PDF on \mathbb{R}^2 and is allowed to depend on the first sample (\mathbf{X}^k, Y^k) . In particular, we can take $g = \hat{f}_{X_i,Y}$ in which case $\hat{\delta}_i^{IS,g}$ is given by

$$\hat{\delta}_i^{IS,g} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{|f_{X_i}(U_1^k) \hat{f}_Y(U_2^k) - \hat{f}_{X_i,Y}(\mathbf{U}^k)|}{\hat{f}_{X_i,Y}(\mathbf{U}^k)}.$$

This expression is at first sight similar to the definition (6) of the single-loop estimator $\hat{\delta}_i^{SL}$. Nonetheless, the key difference is that in our scheme, we use different samples (\mathbf{U}^k) and (\mathbf{X}^k, Y^k) in the KDE step **IS2** and in the Monte Carlo step **IS3**. Although, as mentioned above, the consistency of the single-loop estimator $\hat{\delta}_i^{SL}$ is not obvious, this key difference will actually make it possible to prove that our importance sampling estimator $\hat{\delta}_i^{IS,g}$ is consistent as we now show.

4.2. Theoretical properties of $\hat{\delta}_i^{IS,g}$

As any auxiliary function used to do importance sampling, the convergence of our estimator hinges upon the assumption that the support of the auxiliary distribution contains the support of the function being integrated. Rigorously, we are using importance sampling within the KDE framework where $\hat{f}_{X_i,Y}$ has unbounded support. However, these estimators are meant to converge to the true PDF f_Y and $f_{X_i,Y}$ which may have bounded support. Throughout this section, we assume that

$$\text{Supp}(|f_{X_i} f_Y - \hat{f}_{X_i,Y}|) \subset \text{Supp}(g), \quad (\text{A1})$$

When Assumption (A1) holds and upon standard assumptions on the bandwidths h , h_1 and h_2 which ensure convergence of the KDE \hat{f}_Y and $\hat{f}_{X_i,Y}$, we now show that $\hat{\delta}_i^{SL}$ is asymptotically unbiased and converges to δ_i when we let first $N' \rightarrow \infty$ and then $N \rightarrow \infty$.

Proposition 2. *Assume that Assumption (A1) holds and that $Nh, Nh_1h_2 \rightarrow \infty$ with $h, h_1h_2 \rightarrow 0$. Then $\sup_{N'} \mathbb{E}(\hat{\delta}_i^{IS,g}) \rightarrow \delta_i$ as $N \rightarrow \infty$ and $\hat{\delta}_i^{IS,g} \xrightarrow[N', N \rightarrow \infty]{\text{a.s.}} \delta_i$, where the convergence $\xrightarrow[N', N \rightarrow \infty]{\text{a.s.}}$ means that we let first $N' \rightarrow \infty$ and then $N \rightarrow \infty$.*

Proof. Let

$$\hat{\Delta} = \frac{1}{2} \left\| f_{X_i} \hat{f}_Y - \hat{f}_{X_i,Y} \right\|_{L_1(\text{Supp}(g))}.$$

Then $\mathbb{E}(\hat{\delta}_i^{IS,g}) = \mathbb{E}(\hat{\Delta})$, independent of N' , and the strong law of large numbers implies that $\hat{\delta}_i^{IS,g} \xrightarrow[N' \rightarrow \infty]{\text{a.s.}} \hat{\Delta}$. Furthermore, the assumptions on the bandwidth imply that $\hat{\Delta} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \delta_i$, see for instance [27, Theorem 1]. This shows the second claim of the statement, from which the first one follows by dominated convergence since $\hat{\Delta} \leq 1$. \square

Since the variables \mathbf{U}^k are i.i.d given the variables $(\mathbf{X}^k, \mathbf{Y}^k)$, the law of total variance gives the variance decomposition

$$\begin{aligned} & \text{Var} \left(\hat{\delta}_i^{IS,g} \right) & (7) \\ &= \text{Var} \left(\mathbb{E} \left[\hat{\delta}_i^{IS,g} | (\mathbf{X}^k, Y^k) \right] \right) + \mathbb{E} \left[\text{Var} \left(\hat{\delta}_i^{IS,g} | (\mathbf{X}^k, Y^k) \right) \right], \\ &= \text{Var} \left(\mathbb{E} \left[\frac{1}{2N'} \sum_{k=1}^{N'} \hat{h}(\mathbf{U}^k) | (\mathbf{X}^k, Y^k) \right] \right) + \mathbb{E} \left[\text{Var} \left(\frac{1}{2N'} \sum_{k=1}^{N'} \hat{h}(\mathbf{U}^k) | (\mathbf{X}^k, Y^k) \right) \right], \\ &= \text{Var} \left(\frac{1}{2N'} \sum_{k=1}^{N'} \mathbb{E} \left[\hat{h}(\mathbf{U}^k) | (\mathbf{X}^k, Y^k) \right] \right) + \frac{1}{4(N')^2} \mathbb{E} \left[\sum_{k=1}^{N'} \text{Var} \left(\hat{h}(\mathbf{U}^k) | (\mathbf{X}^k, Y^k) \right) \right], \\ &= \text{Var} \left(\hat{\Delta} \right) + \frac{1}{4N'} \mathbb{E} \left[\text{Var} \left(\hat{h}(\mathbf{U}) | (\mathbf{X}^k, Y^k) \right) \right], & (8) \end{aligned}$$

with

$$\hat{h}(x, y) = 1_{g(x,y)>0} \frac{\left| \hat{f}_Y(y) f_{X_i}(x) - \hat{f}_{X_i,Y}(x, y) \right|}{g(x, y)}.$$

This decomposition clearly highlights the two errors made in the estimation of $\hat{\delta}_i^{IS,g}$: the term $\text{Var}(\hat{\Delta})$ corresponds to the error induced by the KDE procedure of step **IS2** and the second term to the error induced by the importance sampling step **IS3**.

According to [27, Theorem 1] and Lebesgue's dominated convergence theorem, the first term tends to 0 when N tends to $+\infty$. As far as the second one is concerned, the variance term

$$\text{Var} \left(\hat{h}(\mathbf{U}) | (\mathbf{X}^k, Y^k) \right) = \int \frac{\left| \hat{f}_Y f_{X_i} - \hat{f}_{X_i,Y} \right|^2}{g} - \left(\int \left| \hat{f}_Y f_{X_i} - \hat{f}_{X_i,Y} \right| \right)^2,$$

may be infinite if the distribution g is not well chosen. Nevertheless in practice, assuming that g is nearly proportional to $\left| \hat{f}_Y f_{X_i} - \hat{f}_{X_i,Y} \right|$, this term can be made as small as desired because of the factor $\frac{1}{N'}$ without further calls to the possibly expensive black-box function Φ .

We conclude this theoretical section with a more detailed study of the consistency of $\hat{\delta}_i^{IS,g}$. The convergence result $\hat{\delta}_i^{IS,g} \xrightarrow[N', N \rightarrow \infty]{\text{a.s.}} \delta_i$ is not completely satisfactory because in practice, given a certain finite budget (in time, CPU or else), this budget has to be divided between the KDE step **IS2** and the Monte Carlo step **IS3**. In other words, in practice N and N' grow large together while the convergence $\hat{\delta}_i^{IS,g} \xrightarrow[N', N \rightarrow \infty]{\text{a.s.}} \delta_i$ would amount to first give all the budget to the Monte Carlo step **IS3**. Assuming that the auxiliary function g does not depend on the sample (\mathbf{X}^k, Y^k) , the following results, proved in [Appendix A](#) and [Appendix B](#), provide partial results when N and N' grow large simultaneously.

Proposition 3. Assume that:

1. f_{X_i} is bounded;
2. f_Y and $f_{X_i, Y}$ are bounded, twice differentiable and with uniformly continuous second derivative;
3. Condition (A1) holds and g has bounded support;
4. the bandwidths satisfy $h_1 = h_2 = h$, $h \rightarrow 0$ and $h \gg \ln \ln N \sqrt{\ln N/N}$, i.e.,

$$\frac{(\ln \ln N)^2 \ln N}{Nh^2} \xrightarrow{N \rightarrow \infty} 0. \quad (9)$$

Assume finally that $N' = N'(N)$ depends on N in such a way that $N' \rightarrow \infty$ as $N \rightarrow \infty$. Then $\hat{\delta}_i^{IS, g} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \delta_i$.

Proposition 4. Assume that:

1. f_{X_i} is bounded;
2. f_Y and $f_{X_i, Y}$ are bounded, twice differentiable and with uniformly continuous second derivative;
3. there exists $q > 2$ such that

$$\int \|y\|^{2q} f_Y(y) dy < \infty \quad \text{and} \quad \int \|(x, y)\|^{2q} f_{X_i, Y}(x, y) dx dy < \infty ;$$

4. g has unbounded support and there exists $\alpha < 1$ such that $\int g^\alpha < \infty$;
5. the bandwidths satisfy $h_1 = h_2 = h$, $h \rightarrow 0$ and (9).

Assume finally that $N' = N'(N)$ depends on N in such a way that there exists $\beta > \frac{2}{1-\alpha}$ with

$$(N')^\beta \frac{\sqrt{\ln N}}{\sqrt{N}h} \rightarrow 0.$$

Then $\hat{\delta}_i^{IS, g} \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \delta_i$.

4.3. Study of the accuracy of the proposed procedure

4.3.1. Influence of the sampling distribution g

In this section, we study the influence of the sampling distribution g involved at step **IS3**. To achieve this, we apply steps **IS1** and **IS2** on the previous Gaussian toy case with the parameter $N = 5,000$, getting an approximation $\frac{1}{2} \int |f_{X_i} \hat{f}_Y - \hat{f}_{X_i, Y}|$ of the index δ_i from the sample (X_i^k, Y^k) . Then, we perform $m = 100$ runs of step **IS3** with $N' = 2 \times 10^4$ using the different sampling distributions defined as follows:

1. The kernel estimator

$$g_1 = \hat{f}_{X_i, Y} ,$$

of the joint distribution of X_i and Y . The following procedure can be used to generate one realization \mathbf{S} from g_1 :

Step 1. Generate Z uniformly on $\{1, \dots, N\}$.

Step 2. Exit with $\mathbf{S} \sim N \left(\begin{pmatrix} X_i^Z \\ Y^Z \end{pmatrix}, \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix} \right)$.

2. The distribution

$$g_2(x, y) = f_{X_i}(x) \times \hat{f}_Y(y) ,$$

where \hat{f}_Y is the kernel density estimator of the output distribution. The following procedure can be used for generating a random variate \mathbf{S} from g_2 :

- Step 1. Generate Z uniformly on $\{1, \dots, N\}$.
 Step 2. Exit with $\mathbf{S} \sim f_{X_i} \otimes N(Y^Z, h)$.
 3. The PDF g_3 of the Gaussian distribution $N(\hat{m}_i, \hat{\Sigma}_i)$ where

$$\hat{m}_i = \left(\frac{1}{N} \sum_{k=1}^N X_i^k, \frac{1}{N} \sum_{k=1}^N Y^k \right) := (\bar{x}_i, \bar{y}),$$

and

$$\hat{\Sigma}_i = \frac{1}{N-1} \begin{pmatrix} \sum_{k=1}^N (X_i^k - \bar{x}_i)^2 & \sum_{k=1}^N (X_i^k - \bar{x}_i)(Y^k - \bar{y}) \\ \sum_{k=1}^N (X_i^k - \bar{x}_i)(Y^k - \bar{y}) & \sum_{k=1}^N (Y^k - \bar{y})^2 \end{pmatrix},$$

are the regular Monte Carlo estimators of the mean and the covariance matrix of (X_i, Y) respectively.

4. The uniform distribution g_4 on the domain provided by the KDE toolbox used in **IS2**. However, we may notice that this domain contains the samples (X_i^k, Y^k) but the distribution g_4 does not verify the assumption (A1) if the input X_i is not bounded.

As an indicator of the accuracy of the sampling distribution g_j , we consider the coefficient of variation CV_{g_j} given by

$$CV_{g_j} = \frac{\sqrt{\text{Var}(\hat{\delta}_i^{IS, g_j} | (X_i^k, Y^k))}}{\mathbb{E}(\hat{\delta}_i^{IS, g_j} | (X_i^k, Y^k))}.$$

As in Section 3.2, since the true values of the standard deviation and the mean of $\hat{\delta}_i^{IS, g_j}$ given (X_i^k, Y^k) are unknown, we replace them by their regular Monte Carlo estimators (see the beginning of Section 5). For each importance sampling procedure with the distributions g_j defined above, we display in Table 2 the corresponding estimates of the coefficient of variation CV_{g_j} and the mean. The results of Table 2 highlight the impact of the sampling distribution choice. On the one hand, the estimates of δ_1 obtained with the sampling distributions g_j are quite similar with 1% – 2% relative deviation. On the other hand, we can observe that the estimates of δ_2 obtained with g_1 and g_3 show significant underestimation of the theoretical values while those obtained with the other distributions are more accurate. Furthermore, we notice that the variability associated to the use of the distributions g_1 and g_3 is significant.

Those shortcomings are due to a sampling problem as illustrated in Figures 2 and 3. Indeed, the accuracy of the estimator $\hat{\delta}_2^{IS, g}$ depends on how the distribution g samples well on all the support of the integrand function $|f_{X_2} \hat{f}_Y - \hat{f}_{X_2, Y}|$. This is confirmed by Figure 3 where we observe the samples obtained using g_1 (Figure 3(c)) and g_3 (Figure 3(e)) are not distributed wherever the function $|f_{X_2} \hat{f}_Y - \hat{f}_{X_2, Y}|$ takes large values (Figure 3(a)). On the contrary, Figures 3(d) and 3(f) indicate that the distributions g_2 and g_4 are more appropriate than g_1 and g_3 , which corroborates the higher accuracy of the estimates of δ_2 obtained with g_4 and mostly g_2 which presents the lowest coefficients of variation among the distributions $(g_j)_{1 \leq j \leq 4}$. As far as the first delta index is concerned, it can be observed (see Figures 2(c), 2(d), 2(e), 2(f) and 2(g)) that the distributions of the different samples are quite similar, which may explain their similar performances regarding the estimation of the index δ_1 .

Thus, we notice on this simple model that the kernel estimator $\hat{f}_{X_i, Y}$ of the joint distribution of X_i and Y is not the most appropriate sampling distribution to estimate the importance measure δ_i , which may explain the inaccuracy of the single-loop estimator

$\hat{\delta}_2^{SL}$ observed in Section 3.2 since, as explained above, these two estimators have some similarity. Moreover this estimation error is not due to the KDE since the KDE procedure is very efficient in this simple model since the involved PDFs are Gaussian (see Figures 3(a) and 3(b)).

This numerical test case has shown the sampling distribution may have to be chosen with caution. Therefore, the search for the optimal sampling distribution is discussed in the section 4.4.

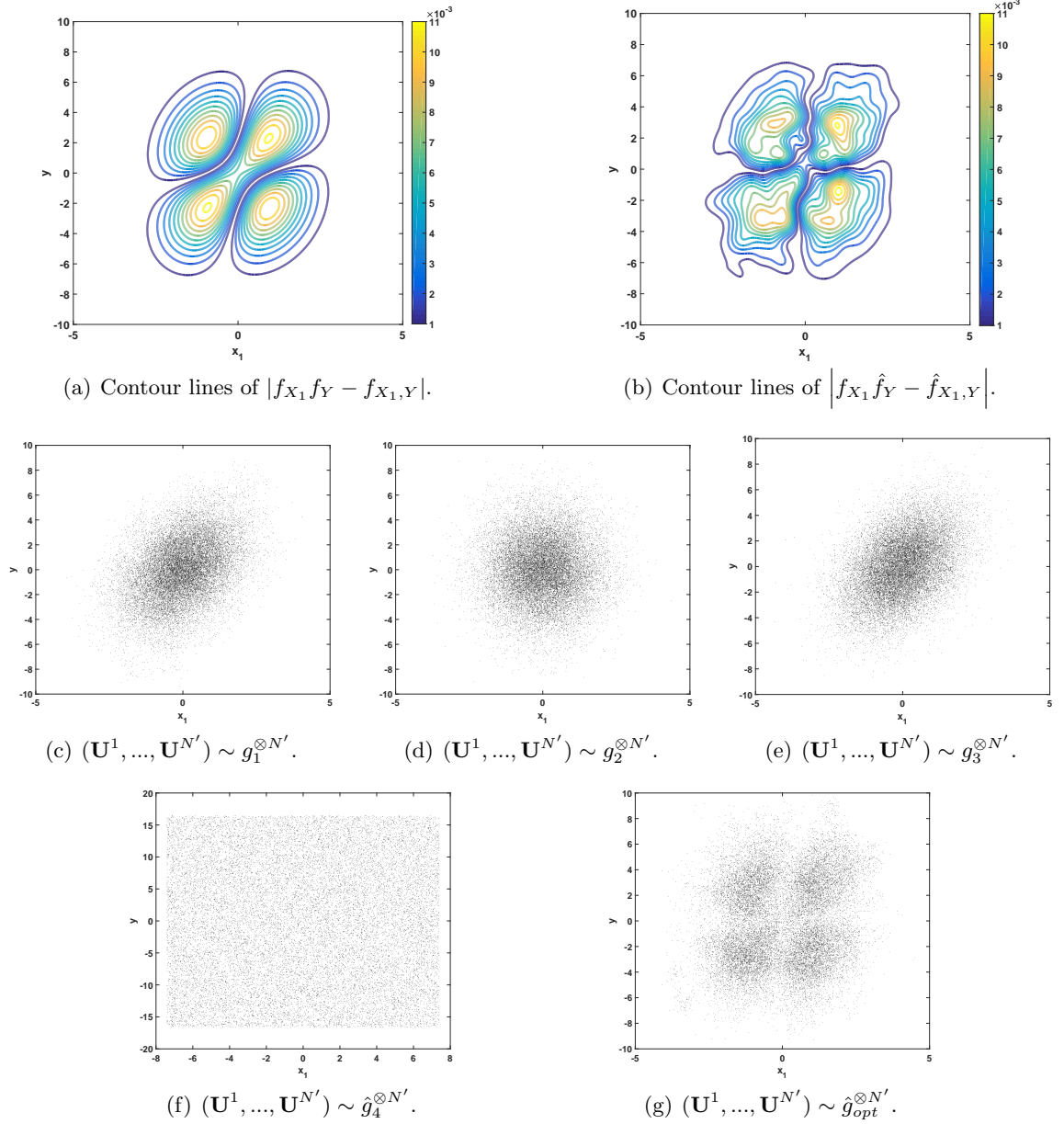
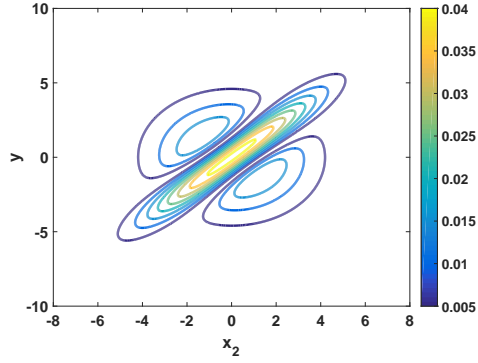


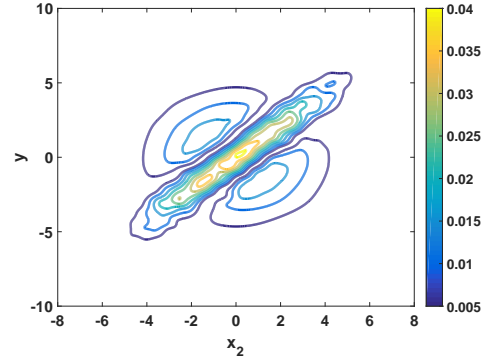
FIGURE 2. DISPLAY OF THE DIFFERENT SAMPLES $(\mathbf{U}^1, \dots, \mathbf{U}^{N'})$ USED FOR THE ESTIMATIONS OF THE MEASURE δ_1 OF THE GAUSSIAN TOY CASE IN COMPARISON TO THE CONTOUR LINES OF THE INTEGRAND FUNCTION $|f_{X_1} f_Y - f_{X_1, Y}|$ AND ITS KDE ESTIMATION $|f_{X_1} \hat{f}_Y - \hat{f}_{X_1, Y}|$.

Table 2. Estimates of the indices of the Gaussian toy case by performing $m = 100$ runs of step **IS3** with different sampling distributions.

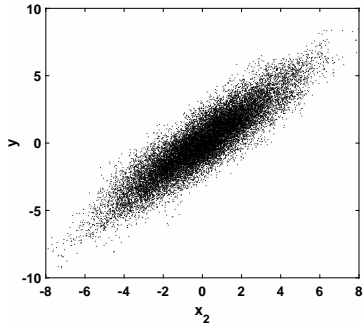
Input	δ_i	g_1		g_2		g_3		g_4		\hat{g}_{opt}	
		Mean	CV_{g_1}	Mean	CV_{g_2}	Mean	CV_{g_3}	Mean	CV_{g_4}	Mean	$CV_{\hat{g}_{opt}}$
X_1	0.1436	0.1465	0.0278	0.1474	0.0083	0.1470	0.0172	0.1471	0.0191	0.1473	0.0038
X_2	0.5382	0.4685	1.1882	0.5175	0.0097	0.4644	0.4822	0.5166	0.0237	0.5174	0.0050



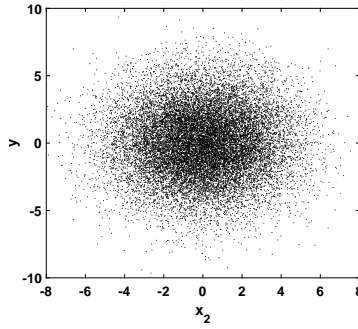
(a) Contour lines of $|f_{X_2} f_Y - f_{X_2, Y}|$.



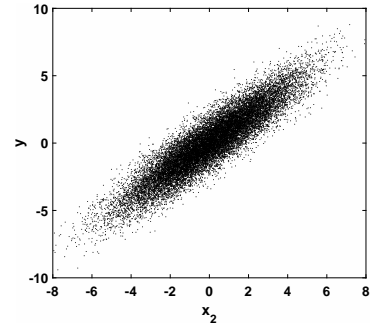
(b) Contour lines of $|f_{X_2} \hat{f}_Y - \hat{f}_{X_2, Y}|$.



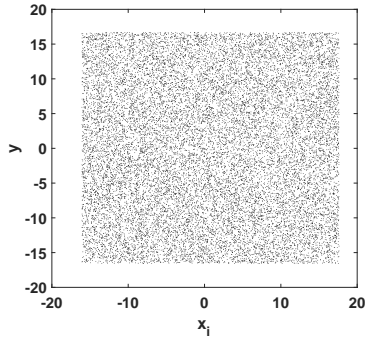
(c) $(\mathbf{U}^1, \dots, \mathbf{U}^{N'}) \sim g_1^{\otimes N'}$.



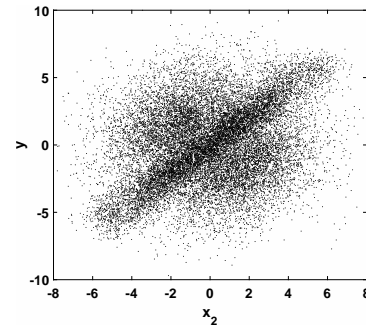
(d) $(\mathbf{U}^1, \dots, \mathbf{U}^{N'}) \sim g_2^{\otimes N'}$.



(e) $(\mathbf{U}^1, \dots, \mathbf{U}^{N'}) \sim g_3^{\otimes N'}$.



(f) $(\mathbf{U}^1, \dots, \mathbf{U}^{N'}) \sim \hat{g}_4^{\otimes N'}$.



(g) $(\mathbf{U}^1, \dots, \mathbf{U}^{N'}) \sim \hat{g}_{opt}^{\otimes N'}$.

FIGURE 3. DISPLAY OF THE DIFFERENT SAMPLES $(\mathbf{U}^1, \dots, \mathbf{U}^{N'})$ USED FOR THE ESTIMATIONS OF THE MEASURE δ_2 OF THE GAUSSIAN TOY CASE IN COMPARISON TO THE CONTOUR LINES OF THE INTEGRAND FUNCTION $|f_{X_2} f_Y - f_{X_2, Y}|$ AND ITS KDE ESTIMATION $|f_{X_2} \hat{f}_Y - \hat{f}_{X_2, Y}|$.

Table 3. Estimates of the indices of the Gaussian toy case by performing $m = 100$ runs of the proposed procedure with different sampling distributions.

Input	δ_i	g_2		g_4		\hat{g}_{opt}	
		Mean	CV	Mean	CV	Mean	CV
X_1	0.1436	0.1358	0.0439	0.1357	0.0503	0.1356	0.0429
X_2	0.5382	0.5188	0.0158	0.5177	0.0302	0.5182	0.0086

4.3.2. Global accuracy

In order to appreciate the global accuracy of the entire proposed procedure with respect to the choice of the sampling distribution g , we consider in this subsection the coefficient of variation of the IS estimator $\hat{\delta}_i^{IS,g}$ associated to the distribution g given by

$$CV(\hat{\delta}_i^{IS,g}) = \frac{\sqrt{\text{Var}(\hat{\delta}_i^{IS,g})}}{\mathbb{E}(\hat{\delta}_i^{IS,g})}.$$

In the previous subsection, it has been noted that the use of the sampling distribution g_1 and g_3 produced a strong variability for a fixed KDE estimation as far as the estimation of the delta indices of the Gaussian toy model is concerned. Therefore, only the distributions g_2 and g_4 are considered here. Since the true values of the standard deviation and the mean of $\hat{\delta}_i^{IS,g}$ are unknown, we replace them by their regular Monte Carlo estimators (see the beginning of Section 5) by performing $m = 100$ runs of the proposed procedure with the parameters $N = 5,000$ and $N' = 2 \times 10^4$.

We display in Table 3 the corresponding results of the Gaussian toy case. It can be observed that the performance of g_2 and g_4 to estimate δ_1 are quite similar with the same 4% relative deviation. As far as the estimation of δ_2 is concerned, g_2 presents a coefficient of variation twice as small as g_4 .

Let us apply the same approach with a second numerical toy case whose the model output is given by

$$Y = \frac{X_1}{10^{-2} + |X_2|},$$

where $X_1, X_2 \stackrel{i.i.d}{\sim} N(0, 1)$.

We display in Table 4 the corresponding results of this model. On this example, the variability associated to the use of the distributions g_4 is significant with 13% and 15% relative deviation while one obtains coefficients variation lower than 5% with the distribution g_2 . This may be due to a sampling problem as illustrated in Figure 4. Indeed, we can observe that the samples distributed according to g_4 (Figure 4(c)) do not fit with the regions where the function $|f_{X_2} \hat{f}_Y - \hat{f}_{X_2,Y}|$ takes large values (Figure 4(a)) whereas the samples obtained with g_2 are better distributed (see Figure 4(b)).

Following the example of the accuracy study of the section 4.3.1, this numerical example has shown that the sampling distribution g may have strong influence on the global accuracy of the proposed procedure. In order to ensure an acceptable accuracy, the distribution g has to be chosen so that its support verifies the theoretical assumption (A1) and is close as possible to the support of $|f_{X_i} f_Y - f_{X_i,Y}|$.

Table 4. Estimates of the indices of the second numerical toy case by performing $m = 100$ runs of the proposed procedure with different sampling distributions.

Input	g_2		g_4		\hat{g}_{opt}	
	Mean	CV	Mean	CV	Mean	CV
X_1	0.2980	0.0300	0.2982	0.1332	0.2966	0.0293
X_2	0.4901	0.0436	0.4935	0.1584	0.5186	0.0467

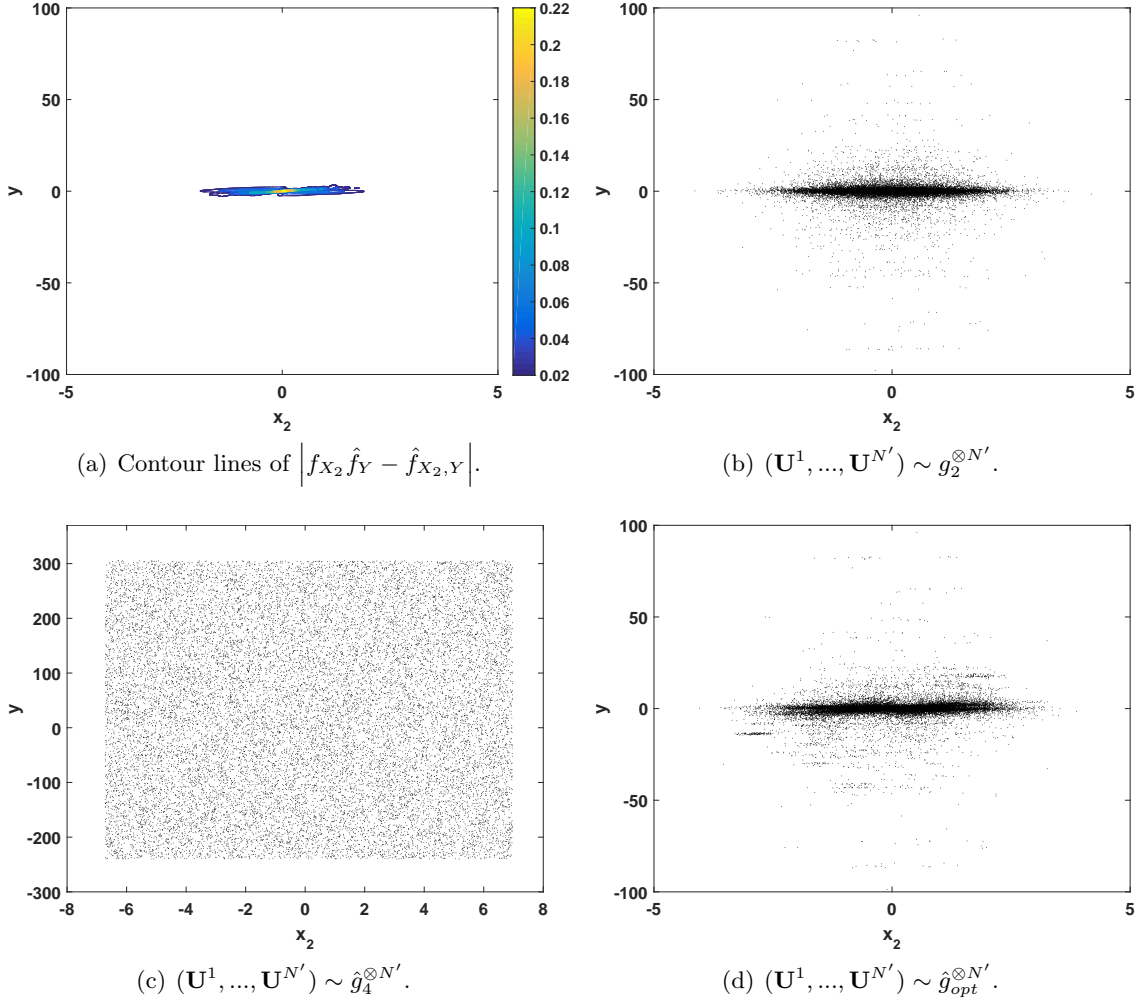


FIGURE 4. DISPLAY OF THE DIFFERENT SAMPLES $(\mathbf{U}^1, \dots, \mathbf{U}^{N'})$ USED FOR THE ESTIMATIONS OF THE MEASURE δ_2 OF THE SECOND NUMERICAL TOY CASE IN COMPARISON TO THE CONTOUR LINES OF THE INTEGRAND FUNCTION $|f_{X_2} \hat{f}_Y - \hat{f}_{X_2, Y}|$.

4.4. The optimal sampling distribution

Let us consider the second term $\text{Var}(\hat{h}(\mathbf{U}) | (\mathbf{X}^k, Y^k))$ of the variance decomposition of $\hat{\delta}_i^{IS, g}$ derived in (8). It is equal to zero when g is given by the function

$$g_{opt} = \frac{|f_{X_i} \hat{f}_Y - \hat{f}_{X_i, Y}|}{\|f_{X_i} \hat{f}_Y - \hat{f}_{X_i, Y}\|_{L_1(\mathbb{R}^2)}},$$

called *optimal sampling distribution*. Unfortunately g_{opt} cannot be used directly in practice because of the unknown normalization constant, but it can be approximated by some \hat{g}_{opt} using the nonparametric importance sampling procedure described in [28]. Assuming that **Steps IS1** and **IS2** have been performed and that we have the KDE \hat{f}_Y and $\hat{f}_{X_i, Y}$ at our disposal, \hat{g}_{opt} is derived by the following implementation steps:

- Generate a sample $\{(\tilde{X}_i^1, \tilde{Y}^1), \dots, (\tilde{X}_i^{N''}, \tilde{Y}^{N''})\}$ according to an initial distribution g_0 .

For the further numerical applications, we will set $g_0 = f_{X_i} \times \hat{f}_Y = g_2$ which has presented good results on the Gaussian toy case studied in the previous section. It has to be noticed that no additional calls to the model output are needed.

- Compute the weights

$$\omega(\tilde{X}_i^k, \tilde{Y}^k) = \frac{|f_{X_i}(\tilde{X}_i^k) \hat{f}_Y(\tilde{Y}^k) - \hat{f}_{X_i, Y}(\tilde{X}_i^k, \tilde{Y}^k)|}{g_0(\tilde{X}_i^k, \tilde{Y}^k)}, \quad k = 1, \dots, N''.$$

- Estimate g_{opt} by the weighted kernel estimator

$$\hat{g}_{opt}(x, y) = \frac{1}{N'' \tilde{h}_1 \tilde{h}_2 \tilde{\omega}} \sum_{k=1}^{N''} \omega(\tilde{X}_i^k, \tilde{Y}^k) K\left(\frac{x - \tilde{X}_i^k}{\tilde{h}_1}\right) K\left(\frac{y - \tilde{Y}^k}{\tilde{h}_2}\right), \quad (10)$$

where $\tilde{\omega} = \frac{1}{N''} \sum_{k=1}^{N''} \omega(\tilde{X}_i^k, \tilde{Y}^k)$. The convergence of the estimator \hat{g}_{opt} is studied in [28], even if we do not verify all its assumptions.

We can now explain our estimation procedure:

Step Opt1. Same as steps **SL1** and **IS1**, i.e., generate $(\mathbf{X}^1, \dots, \mathbf{X}^N)$ i.i.d. with common distribution \mathbf{X} , and then obtain N observations of the model by $Y^k = \Phi(\mathbf{X}^k)$ for $k = 1, \dots, N$.

Step Opt2. Same as steps **SL2** and **IS2**, in particular get the two estimators $\hat{f}_{\mathbf{X}, Y}$ and \hat{f}_Y from (\mathbf{X}^k, Y^k) as in (4) and (5).

Step Opt3. Our estimator $\hat{\delta}_i^{Opt}$ of δ_i is given by

$$\hat{\delta}_i^{Opt} = \hat{\delta}_i^{IS, \hat{g}_{opt}} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{|\hat{f}_Y(U_2^k) f_{X_i}(U_1^k) - \hat{f}_{X_i, Y}(\mathbf{U}^k)|}{\hat{g}_{opt}(\mathbf{U}^k)}, \quad (11)$$

where the random variables $\mathbf{U}^k = (U_1^k, U_2^k)$ are i.i.d according to the distribution \hat{g}_{opt} defined above. The following procedure can be used for generating one realization \mathbf{U} from \hat{g}_{opt} :

Step 1. Generate $Z \sim \sum_{k=1}^N \frac{\omega(\tilde{X}_i^k, \tilde{Y}^k)}{\tilde{\omega}} \delta_k$.

Step 2. Exit with $\mathbf{U} \sim N\left(\begin{pmatrix} \tilde{X}_i^Z \\ \tilde{Y}^Z \end{pmatrix}, \begin{pmatrix} \tilde{h}_1 & 0 \\ 0 & \tilde{h}_2 \end{pmatrix}\right)$.

We perform the same studies of the section 4.3 with this algorithm and with the parameters $N = 5,000$ and $N' = N'' = 2 \times 10^4$. With this set of parameters, the computation of \hat{g}_{opt} takes 3 s to compute using our own computer.

Firstly, let us appreciate the influence of the distribution \hat{g}^{Opt} . It can be seen in Table 2 that $\hat{\delta}^{Opt}$ presents the best robustness with 0.5% relative deviation, nearly twice smaller than that for $\hat{\delta}^{IS, g_2}$. Thus, for a given KDE estimation, the use of the approximation \hat{g}_{opt}

of the optimal sampling distribution seems to be the better choice in term of estimation error, which complies with the theory.

Secondly, let us focus on the global estimation error of the procedure. It can be observed on both numerical toy cases that $\hat{\delta}^{Opt}$ provides good estimates with coefficients of variation slightly lower than those of g_2 . At last, Figure 3(g) and Figure 4(d) shows that \hat{g}_{opt} samples well on the integration area of interest.

Thus, those results confirms the advantages of using \hat{g}_{opt} , albeit the performance of g_2 and $\hat{\delta}^{Opt}$ are quite similar. This similarity may be due to the fact that the error associated to the KDE estimation predominates in the global error. The distribution g_2 may be a good candidate since the use of \hat{g}_{opt} adds some computation time. However, this difference between the computational burden becomes negligible when the black-box function Φ is expensive. Henceforth, the estimator $\hat{\delta}_i^{Opt}$ obtained using the approximation \hat{g}_{opt} of the *optimal sampling distribution* is designated as the *nonparametric importance sampling estimator* in the following of this article.

5. Numerical examples

In this section, several numerical examples are considered in order to illustrate the application of the importance sampling based estimation scheme using the approximation \hat{g}_{opt} of the *optimal sampling distribution* discussed in Section 4.4. The results are compared with those performed with the single-loop Monte Carlo simulation.

The model output of the first example is expressed as an affine transformation of a Gaussian vector and is a generalization of the Gaussian toy case previously studied. The second example is cantilever beam structure model. The third and last example concern a risk assessment model.

An indicator of the efficiency of an estimator $\hat{\delta}_i$ (such as $\hat{\delta}_i^{SL}$ or $\hat{\delta}_i^{Opt}$) of the importance measure δ_i is the coefficient of variation

$$CV(\hat{\delta}_i) = \frac{\sqrt{\text{Var}(\hat{\delta}_i)}}{\mathbb{E}(\hat{\delta}_i)}.$$

For each estimator $\hat{\delta}_i$ of the importance measure δ_i , we approximate its mean and its standard deviation using Monte Carlo procedure. Considering m estimates $(\hat{\delta}_i^1, \dots, \hat{\delta}_i^m)$, we compute the respective unbiased estimators of the mean and the standard deviation (STD):

$$\bar{\delta}_i = \frac{1}{m} \sum_{k=1}^m \hat{\delta}_i^k \quad \text{and} \quad \bar{\sigma}_{\hat{\delta}_i} := \sqrt{\frac{1}{m-1} \sum_{k=1}^m (\hat{\delta}_i^k - \bar{\delta}_i)^2},$$

which provides the following estimator of the coefficient of variation:

$$\frac{\bar{\delta}_i}{\bar{\sigma}_{\hat{\delta}_i}}.$$

When the unconditional and conditional output distributions are known, the theoretical values of the importance measures are available using numerical integration. In this case, the relative difference (RD)

$$\frac{\bar{\delta}_i - \delta_i}{\delta_i},$$

may be computed in order to appreciate the error of the estimator $\tilde{\delta}$.

5.1. Example 1: an affine transformation of a Gaussian vector

In this subsection we assume that the expression of the model output is

$$Y_1 = \mathbf{A}\mathbf{X} ,$$

where $\mathbf{A} = [1.7 \ 1.8 \ 1.9 \ 2]$ and where the input \mathbf{X} follows the Gaussian distribution $N(0, \Sigma)$ where the covariance matrix Σ is defined as follows:

$$\Sigma = \begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1 & 1/2 & 1/3 \\ 1/3 & 1/2 & 1 & 1/2 \\ 1/4 & 1/3 & 1/2 & 1 \end{pmatrix} .$$

Classical results on Gaussian vector enable to determine the unconditional and conditional output distributions:

$$Y_1 \sim N(0, \mathbf{A}\Sigma^t\mathbf{A}) ,$$

and

$$Y_1|X_i = x_i \sim N(m_i, \sigma_i^2) ,$$

where the mean m_i and the variance σ_i^2 are defined by:

$$m_i = A_i x_i + \mathbf{A}_{-i} \mathbf{C}_i \Sigma_{ii}^{-1} x_i ,$$

$$\sigma_i^2 = \mathbf{A}_{-i} (\Sigma_{-i} - {}^t \mathbf{C}_i \mathbf{C}_i \Sigma_{ii}^{-1}) \mathbf{A}_{-i}^t ,$$

where \mathbf{A}_{-i} is the vector \mathbf{A} private of its i -th component A_i , Σ_{-i} is the matrix Σ private of its i -th row and column and \mathbf{C}_i the i -th column of Σ private of its i -th component Σ_{ii} . Thus, the theoretical values of the indices are known and can be computed using numerical integration.

We compare these reference values with the estimates obtained with $m = 100$ runs of the single-loop Monte Carlo simulation and the nonparametric importance sampling based method with the parameters $N'' = N = 5,000$ and $N' = 5 \times 10^4$ (the number of calls to the model is the same for both methods). The single-loop method and the proposed method take respectively 4 s and 54 s to compute the four delta indices δ_i with our own computer, among which 3.5 s for the estimation of the optimal sampling distribution. This difference between the computational burden of the proposed method and the one of the single-loop design, notably due to the fact that $N' \gg N$, becomes negligible when the black-box function Φ is expensive. The results are listed in Table 5. We can notice that both methods respect the good importance ranking, i.e., $X_3 > X_2 > X_4 > X_1$, and converge since the estimates are obtained with around 2% relative deviation. However, the estimates of $(\hat{\delta}_i^{IS})_{1 \leq i \leq 4}$ are approximately equal to the theoretical values with relative differences around 5% while the single-loop estimates show a lower precision with non negligible relative differences. Also, as far as the single-loop design is concerned, it seems that the higher is δ_i , the higher is the relative difference associated to the estimate δ_i^{SL} .

These values of relative differences highlight the bias of the nonparametric importance sampling estimator and, as discussed in Section 4.2, it is due to the kernel estimation step contribution and can be decreased by increasing the parameter N . In this example, it is inexpensive in term of computation since the black-box function is analytic but this is not generally the case.

Table 5. Estimates of the indices of the example 1.

Input	Theoretical	Single loop $\hat{\delta}_i^{SL}$				Proposed method $\hat{\delta}_i^{Opt}$			
	value δ_i	Mean	STD	CV	RD	Mean	STD	CV	RD
X_1	0.2857	0.2201	0.0062	0.0280	-0.2297	0.2707	0.0059	0.0218	-0.0526
X_2	0.3620	0.2620	0.0059	0.0226	-0.2761	0.3444	0.0054	0.0157	-0.0486
X_3	0.3792	0.2690	0.0059	0.0218	-0.2907	0.3607	0.0057	0.0157	-0.0487
X_4	0.3176	0.2383	0.0052	0.0219	-0.2497	0.3010	0.0060	0.0199	-0.0522

5.2. Example 2: a cantilever beam structure

This subsection deals with a structural system reliability problem introduced in [2]. We consider a rectangular cantilever beam structure subjected to two random forces F_1 and F_2 . The modulus of elasticity E and the dimensions of the beam denoted by ω (width), t (height) and L (length) are also assumed to be random. The output model of interest Y_2 is the maximum displacement of the structure which is expressed as follows

$$Y_2 = \frac{4L^3}{E\omega t} \sqrt{\left(\frac{F_1}{\omega^2}\right)^2 + \left(\frac{F_2}{t^2}\right)^2}.$$

All the six inputs F_1 , F_2 , E , ω , t and L are normally distributed whose the parameters are listed in Table 6. The correlations between the inputs are characterized by the following correlation matrix:

$$\begin{pmatrix} 1 & 0.2 & 0 & 0 & 0 & 0 \\ 0.2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.1 & 0.1 \\ 0 & 0 & 0 & 0.1 & 1 & 0.1 \\ 0 & 0 & 0 & 0.1 & 0.1 & 1 \end{pmatrix}.$$

We estimate all the six indices $(\delta_i)_{1 \leq i \leq 6}$ with $m = 100$ runs of the single-loop method and the scheme proposed in this article with the parameters $N'' = N = 5,000$ and $N' = 2 \times 10^4$. The single-loop method and the proposed method take respectively 5 s and 35 s to compute the six delta indices δ_i with our own computer, of which 5.4 s for the estimation of the optimal sampling distribution. The results are displayed in Table 7.

We observe that both methods lead to the same influence ranking, mainly $L > \omega > t > E > F_1 > F_2$ and converge since the estimates are obtained with coefficients of variation less than 10%. Also, we may note that each single-loop estimate is always slightly lower than the nonparametric importance sampling estimate.

5.3. Example 3

5.3.1. A risk assessment model

In this part we illustrate the application of the nonparametric importance sampling scheme to a probabilistic risk assessment model introduced in [30] and based on fault tree analysis. The principle of this method is that the unavailability of a system is evaluated using a fault tree. The system's failure, referred as 'top event', is then expressed in terms of the different sequences of events leading to the failure of the system, referred as 'cut sets'.

Table 6. Distributions of the input variables of the example 2.

Input	Distribution	Mean	coefficient of variation
F_1	Normal	500 lb	0.08
F_2	Normal	1000 lb	0.08
E	Normal	2.9×10^7 psi	0.08
ω	Normal	2.4487 in	0.08
t	Normal	3.8884 in	0.08
L	Normal	100 in	0.08

Table 7. Estimates of the indices of the example 2.

Input	Single loop $\hat{\delta}_i^{SL}$			Proposed method $\hat{\delta}_i^{Opt}$		
	Mean	STD	CV	Mean	STD	CV
F_1	0.0556	0.0045	0.0809	0.0600	0.0046	0.0767
F_2	0.0472	0.0046	0.0975	0.0507	0.0047	0.0927
E	0.0742	0.0053	0.0714	0.0809	0.0055	0.0680
ω	0.1566	0.0056	0.0358	0.1788	0.0066	0.0369
t	0.1245	0.0057	0.0458	0.1397	0.0064	0.0458
L	0.1985	0.0057	0.0287	0.2384	0.0060	0.0252

Here, the probability of the top event is written as [30]

$$Y_3 = X_1 X_3 X_5 + X_1 X_3 X_6 + X_1 X_4 X_5 + X_1 X_4 X_6 + X_2 X_3 X_4 \\ + X_2 X_3 X_5 + X_2 X_4 X_5 + X_2 X_5 X_6 + X_2 X_4 X_7 + X_2 X_6 X_7 ,$$

where all the inputs X_i are independent random variables following lognormal distribution and whose the parameters are listed in Table 8.

We estimate all the seven indices with $m = 100$ runs of both single-loop Monte Carlo and nonparametric importance sampling schemes with the parameters $N'' = N = 5,000$ and $N' = 2 \times 10^4$. The single-loop method and the proposed method take respectively 5 s and 41 s to compute the seven delta indices δ_i with our own computer, of which 6.3 s for the estimation of the optimal sampling distribution. The results are reported in Table 9.

Following example 2, we observe that both methods provide the same importance ranking, mainly $X_2 > X_6 > X_5 > X_4 > X_7 > X_1 > X_3$. Furthermore, it seems that each single-loop estimate is always slightly lower than the nonparametric importance sampling estimate and they are obtained with similar coefficient of variation.

Nonetheless, its difficult to measure the precision of these estimates since the unconditional and conditional output distributions are unknown. In order to get an idea of this accuracy, we consider in the next subsection a close model for which the theoretical values of the importance measures δ_i are available.

5.3.2. Study of a theoretical multiplicative model

Let us consider the following multiplicative model output:

$$Y_4 = \prod_{i=1}^d \Lambda_i ,$$

Table 8. Distributions of the input variables of the example 3.

Input	Distribution	Mean of $\ln(X_i)$	Variance of $\ln(X_i)$
X_1	Lognormal	0.6044	0.1776
X_2	Lognormal	1.0098	0.1776
X_3	Lognormal	-6.9965	0.1776
X_4	Lognormal	-6.3034	0.1776
X_5	Lognormal	-5.6103	0.1776
X_6	Lognormal	-5.3871	0.1776
X_7	Lognormal	-5.89792	0.1776

Table 9. Estimates of the indices of the example 3.

Input	Single loop $\hat{\delta}_i^{SL}$			Proposed method $\hat{\delta}_i^{Opt}$		
	Mean	STD	CV	Mean	STD	CV
X_1	0.0754	0.0051	0.0676	0.0828	0.0052	0.0628
X_2	0.1908	0.0051	0.0267	0.2237	0.0053	0.0237
X_3	0.0588	0.0045	0.0765	0.0647	0.0046	0.0711
X_4	0.1023	0.0048	0.0469	0.1137	0.0053	0.0466
X_5	0.1363	0.0053	0.0389	0.1532	0.0063	0.0411
X_6	0.1534	0.0051	0.0332	0.1751	0.0056	0.0320
X_7	0.0778	0.0048	0.0617	0.0855	0.0051	0.0596

where all the inputs Λ_i are independent and distributed according to lognormal distributions $L(m_i, \sigma_i^2)$ where m_i and σ_i^2 are the mean value and the variance of $\ln(\Lambda_i)$ respectively. The unconditional and conditional output distributions are known and given by

$$Y_4 \sim L \left(\sum_{i=1}^d m_i, \sum_{i=1}^d \sigma_i^2 \right),$$

$$(Y_4 | \Lambda_i = x_i) \sim L \left(\ln(x_i) + \sum_{j \neq i} m_j, \sum_{j \neq i} \sigma_j^2 \right),$$

and the theoretical values of the importance measures δ_i are available by performing numerical integration.

When $d = 3$ the output model Y_3 of the risk assessment model is expressed as a sum of terms similar to the output Y_4 . Thus, we may appreciate the accuracy of the estimates of the importance measures of the model output Y_3 by estimating those of Y_4 for some input $(\Lambda_1, \Lambda_2, \Lambda_3)$ equal in distribution to a triplet $(X_{i_1}, X_{i_2}, X_{i_3})$ chosen among the seven inputs X_i of the risk assessment model. It amounts to estimate the indices δ_i of the multiplicative model Y_4 by setting different values of the means m_i and fixing $(\sigma_i^2)_{1 \leq i \leq 3} = [0.1776 \ 0.1776 \ 0.1776]$. Using numerical integration, we get that the three sensitivity indices $(\delta_1, \delta_2, \delta_3)$ are equal to 0.2239, whatever the chosen triplet $(X_{i_1}, X_{i_2}, X_{i_3})$. Table 10 show the estimates obtained by performing $m = 100$ repeated simulations of both single-loop Monte Carlo and nonparametric importance sampling methods with parameters $N'' = N = 5,000$ and $N' = 2 \times 10^4$. We find that in each case, the nonparametric importance

Table 10. Estimates of the indices of the multiplicative model Y_4 for different values of the means m_i [parameter setup: $d = 3$, $(\sigma_i^2)_{1 \leq i \leq 3} = [0.1776 \ 0.1776 \ 0.1776]$].

Input ($\Lambda_1, \Lambda_2, \Lambda_3$)	Theoretical value	Single loop $\hat{\delta}_i^{SL}$			Proposed method $\hat{\delta}_i^{Opt}$		
		Mean	STD	RD	Mean	STD	RD
(X_1, X_3, X_5)	0.2239	0.1880	0.0049	-0.1603	0.2225	0.0063	-0.0063
(X_1, X_4, X_6)	0.2239	0.1862	0.0058	-0.1684	0.2198	0.0066	-0.0184
(X_2, X_6, X_7)	0.2239	0.1889	0.0048	-0.1561	0.2228	0.0059	-0.0047

Table 11. Estimates of the indices of the multiplicative model Y_4 for different values of the input dimension d [parameter setup: $m_i = 0$, $\sigma_i^2 = 1$].

Input	Theoretical value	Single loop $\hat{\delta}_i^{SL}$			Proposed method $\hat{\delta}_i^{Opt}$		
		Mean	STD	RD	Mean	STD	RD
$d = 4$	0.1846	0.8757	0.4570	3.7440	0.3972	0.0396	1.1517
$d = 5$	0.1604	2.0302	1.6746	11.6574	0.4570	0.0329	1.8492
$d = 6$	0.1436	3.6376	1.9370	24.3318	0.5036	0.1267	2.5068

sampling estimates present an higher accuracy than the single-loop estimates with highly better relative differences. Therefore, heuristically, one can think that the nonparametric importance sampling method is more efficient than the single-loop method to estimate the importance measures of the model output Y_3 .

Let us vary the dimension parameter d and fix for instance $m_i = 0$, $\sigma_i^2 = 1$ for all i in such way it is sufficient to compute the index δ_1 since all inputs has the same influence. We report in Table 11 the theoretical values of δ_1 computed by Borgonovo [24] and the estimates of the index δ_1 obtained by applying both methods $m = 100$ times with the parameters $N'' = N = 5,000$ and $N' = 2 \times 10^4$ for different values of the input's dimension d . Although the results are maybe not completely satisfactory yet, we have improved the accuracy of the single-loop estimator by a factor 3 when $d = 4$ and by a factor 10 when $d = 5$ and $d = 6$. In addition, the single-loop method may provides estimates greater than 1 and with large standard deviations.

Nevertheless, it has to be pointed out that the current problem is difficult since the unconditional and conditional outputs are lognormally distributed. Indeed, the lognormal distribution has a heavy right tail so that the Gaussian kernel K used at **Step SL2.** and **Step Opt2.** may not be the most suitable kernel. There exist some alternative approaches in the literature which avoid this issue for one dimensional densities, see for instance the estimator based on the modified Champernowne distribution [31] or the inverse beta transformation based method [32]. Also, this shortcoming induced by the heavy tail framework may be avoided by using the copula-based definition of the delta indice [17][18] or its CDF-based definition [16].

Thus, the bias observed in this example is only due to the KDE procedure which affects the quality of the estimator $\hat{\Delta} = \frac{1}{2} \left\| f_{X_i} \hat{f}_Y - \hat{f}_{X_i, Y} \right\|_{L_1(\mathbb{R}^2)}$. Indeed, let us take $d = 4$ for instance. We see on Figure 5(b) that the sample distributed according to \hat{g}_{opt} do not fit with the regions where the function $|f_{\Lambda_1} f_Y - f_{\Lambda_1, Y}|$ takes large values (Figure 5(a)), notably as far as the coordinate corresponding to the output is concerned.

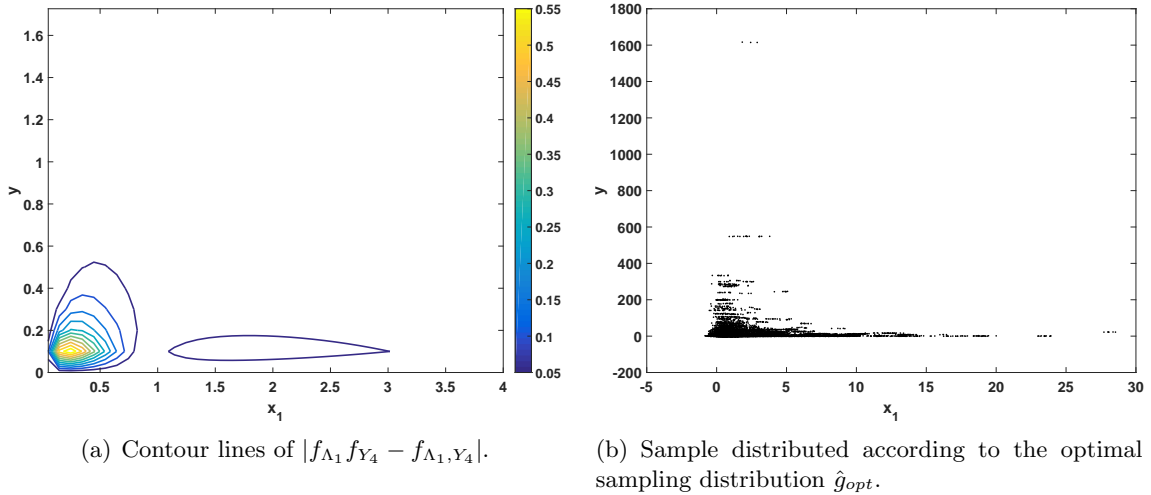


FIGURE 5. DISPLAY OF THE SAMPLE DISTRIBUTED ACCORDING TO THE OPTIMAL SAMPLING DISTRIBUTION \hat{g}_{opt} IN COMPARISON TO THE CONTOUR LINES OF THE INTEGRAND FUNCTION $|f_{\Lambda_1} f_{Y_4} - f_{\Lambda_1, Y_4}|$ [PARAMETER SETUP: $d = 4$, $m_i = 0$, $\sigma_i^2 = 1$].

6. Conclusion

In this paper, we focus on establishing a new estimation scheme of Borgonovo’s sensitivity indices inspired from the single-loop Monte Carlo design [2] whose implementations steps are detailed in Section 3.1. We illustrate with a numerical test case that even though the single-loop estimates may respect the influence ranking, they may present significant differences with the theoretical values. In order to obtain an higher accuracy, we developed in Section 4 a nonparametric importance sampling method which preserves the same advantages of the single-loop estimator, namely only one set of output’s observations are needed to compute all the importance measures δ_i and it can be applied to high dimensional models with correlated inputs. We prove the convergence of the scheme and the bias and the variance of the nonparametric importance sampling estimator are analysed. In Section 5, several numerical applications are performed to illustrate the gain of accuracy of the proposed method. Furthermore, the last example also highlights the limitations of the methods using the kernel estimation procedure in the heavy tail framework which may substantially affect the accuracy of the estimates of moment independent sensitivity measures.

7. Acknowledgements

The authors are grateful to the anonymous reviewers for their relevant remarks that helped improve the paper. The first author is also grateful to Université Paul Sabatier for providing the financial support for the study.

Appendix A. Proof of Proposition 3

The proof of Proposition 3 uses Theorem 7 in [29], which in our case reads as follows.

Theorem 5 (Theorem 7 in [29]). *Assume that f_{X_i} is bounded, that f_Y and $f_{X_i,Y}$ are bounded, twice differentiable and with uniformly continuous second derivative, and that (9) holds with $h = h_1 = h_2$. Then for any $q > 0$ it holds that*

$$\sup_{|y| \leq c(N)} \left| \hat{f}_Y(y) - f_Y(y) \right| = O \left(\left(\frac{\ln N}{Nh} \right)^{1/2} + h^2 \right) := O(\epsilon_1(N)) ,$$

$$\sup_{\|(x,y)\| \leq c(N)} \left| \hat{f}_{X_i,Y}(x,y) - f_{X_i,Y}(x,y) \right| = O \left(\left(\frac{\ln N}{Nh^2} \right)^{1/2} + h^2 \right) := O(\epsilon_2(N)) .$$

almost surely, where

$$c(N) = O \left(\ln \ln N \sqrt{\ln N} N^{1/(2q)} \right) .$$

We now prove Proposition 3. First of all, let

$$h(x,y) = 1_{g(x,y) > 0} \frac{|f_Y(y)f_{X_i}(x) - f_{X_i,Y}(x,y)|}{g(x,y)}, \quad x, y \in \mathbb{R}^2.$$

Since

$$\frac{1}{2N'} \sum_{k=1}^{N'} h(\mathbf{U}^k) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} \delta_i$$

by the strong law of large numbers, it is enough to prove that

$$\hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(\mathbf{U}^k) \xrightarrow[N \rightarrow \infty]{\text{a.s.}} 0.$$

The triangular inequality gives

$$\begin{aligned} & \left| \hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(\mathbf{U}^k) \right| \\ & \leq \frac{1}{2N'} \sum_{k=1}^{N'} \frac{1}{g(\mathbf{U}^k)} \left| \left| f_{X_i}(U_1^k) \hat{f}_Y(U_2^k) - \hat{f}_{X_i,Y}(\mathbf{U}^k) \right| - \left| f_{X_i}(U_1^k) f_Y(U_2^k) - f_{X_i,Y}(\mathbf{U}^k) \right| \right| \end{aligned}$$

and using the reversed triangular inequality $\|a\| - \|b\| \leq \|a - b\|$ and then again the triangular inequality we obtain

$$\begin{aligned} \left| \hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(\mathbf{U}^k) \right| & \leq \frac{1}{2N'} \sum_{k=1}^{N'} \frac{f_{X_i}(U_1^k)}{g(\mathbf{U}^k)} \left| \hat{f}_Y(U_2^k) - f_Y(U_2^k) \right| \\ & \quad + \frac{1}{2N'} \sum_{k=1}^{N'} \frac{1}{g(\mathbf{U}^k)} \left| \hat{f}_{X_i,Y}(\mathbf{U}^k) - f_{X_i,Y}(\mathbf{U}^k) \right|. \quad (\text{A.1}) \end{aligned}$$

Since g has bounded support, the random variables \mathbf{U}^k live in a bounded set and so we can use Theorem 5 to get the existence of random constants $C_1, C_2 < \infty$ such that

$$\left| \hat{f}_Y(U_2^k) - f_Y(U_2^k) \right| \leq C_1 \epsilon_1(N), \quad \left| \hat{f}_{X_i,Y}(\mathbf{U}^k) - f_{X_i,Y}(\mathbf{U}^k) \right| \leq C_2 \epsilon_2(N)$$

which gives

$$\left| \hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(\mathbf{U}^k) \right| \leq \frac{C_1 \epsilon_1(N)}{2N'} \sum_{k=1}^{N'} \frac{f_{X_i}(U_1^k)}{g(\mathbf{U}^k)} + \frac{C_2 \epsilon_2(N)}{2N'} \sum_{k=1}^{N'} \frac{1}{g(\mathbf{U}^k)}.$$

Since

$$\frac{1}{N'} \sum_{k=1}^{N'} \frac{f_{X_i}(U_1^k) + 1}{g(\mathbf{U}^k)} \xrightarrow[N' \rightarrow \infty]{\text{a.s.}} \int_{\text{Supp}(g)} (f_{X_i}(x) + 1) dx dy$$

and $\text{Supp}(g)$ is bounded, we obtain the result since $\epsilon_1(N), \epsilon_2(N) \rightarrow 0$.

Appendix B. Proof of Proposition 4

For the proof of Proposition 4 we need the following slight extension of Theorem 7 in [29]. The proof of this result is the same as the proof of [29, Theorem 7] but uses [29, Theorem 5] instead of [29, Theorem 3].

Theorem 6 (Extension of Theorem 7 in [29]). *Assume that f_{X_i} is bounded, that f_{X_i} and $f_{X_i,Y}$ are differentiable with uniformly continuous derivative, that there exists $q > 2$ such that*

$$\int \|y\|^{2q} f_Y(y) dy < \infty \quad \text{and} \quad \int \|(x,y)\|^{2q} f_{X_i,Y}(x,y) dx dy < \infty,$$

and that (9) holds with $h = h_1 = h_2$. Then

$$\sup_{y \in \mathbb{R}} \left| \hat{f}_Y(y) - f_Y(y) \right| \leq O(\epsilon_1(N)),$$

$$\sup_{(x,y) \in \mathbb{R}^2} \left| \hat{f}_{X_i,Y}(x,y) - f_{X_i,Y}(x,y) \right| \leq O(\epsilon_2(N)),$$

almost surely, where $\epsilon_1(N), \epsilon_2(N)$ are as in Theorem 5.

We now prove Proposition 4. We start from (A.1) to get

$$\begin{aligned} \left| \hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(\mathbf{U}^k) \right| &\leq \frac{C_1 \epsilon_1(N)}{2N'} \sum_{k=1}^{N'} \frac{f_{X_i}(U_1^k)}{g(\mathbf{U}^k)} + \frac{C_2 \epsilon_2(N)}{2N'} \sum_{k=1}^{N'} \frac{1}{g(\mathbf{U}^k)}, \\ &\leq \frac{C(\epsilon_1(N) + \epsilon_2(N))}{2N'} \sum_{k=1}^{N'} \frac{1}{g(\mathbf{U}^k)}. \end{aligned}$$

Now the problem is that

$$\frac{1}{N'} \sum_{k=1}^{N'} \frac{1}{g(\mathbf{U}^k)} \xrightarrow[N' \rightarrow \infty]{\text{a.s.}} \int_{\text{Supp}(g)} dx dy$$

which is infinite since $\text{Supp}(g)$ is unbounded. However, for $\beta > \frac{2}{1-\alpha}$ we have

$$\sup_{n \geq 1} \frac{1}{n^{1+\beta}} \sum_{k=1}^n \frac{1}{g(\mathbf{U}^k)} < \infty \text{ a.s.} \quad (\text{B.1})$$

Indeed, for every $\eta > 0$ we have by the union bound

$$\mathbb{P} \left(\frac{1}{n^{1+\beta}} \sum_{k=1}^n \frac{1}{g(\mathbf{U}^k)} \geq \eta \right) \leq n \mathbb{P} \left(\frac{1}{g(U)} \geq \eta n^\beta \right)$$

and so Markov inequality gives

$$\mathbb{P}\left(\frac{1}{n^\beta} \sum_{k=1}^n \frac{1}{g(\mathbf{U}^k)} \geq \eta\right) \leq n \frac{1}{(\eta n^\beta)^{1-\alpha}} \mathbb{E}\left[\left(\frac{1}{g(U)}\right)^{1-\alpha}\right] = \frac{1}{\eta^{1-\alpha} n^{\beta(1-\alpha)-1}} \int g^\alpha.$$

Since by assumptions we have $\int g^\alpha < \infty$ and $\beta(1-\alpha)-1 > 1$, we see that

$$\sum_{n \geq 1} \mathbb{P}\left(\frac{1}{n^{1+\beta}} \sum_{k=1}^n \frac{1}{g(\mathbf{U}^k)} \geq \eta\right) < \infty$$

which implies, by Borel–Cantelli lemma, that $\frac{1}{n^{1+\beta}} \sum_{k=1}^n \frac{1}{g(\mathbf{U}^k)} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0$, and in particular (B.1) holds. Up to random multiplicative constant, we thus have

$$\left| \hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(\mathbf{U}^k) \right| \leq C(\epsilon_1(N) + \epsilon_2(N))(N')^\beta$$

and the assumption on N' makes this bound vanish.

References

- [1] E. Borgonovo, A new uncertainty importance measure, *Reliability Engineering & System Safety* 92 (2007) 771–784.
- [2] P. Wei, Z. Lu, X. Yuan, Monte Carlo simulation for moment-independent sensitivity analysis, *Reliability Engineering & System Safety* 110 (2013) 60–67.
- [3] F. N. Catbas, M. Susoy, D. M. Frangopol, Structural health monitoring and reliability estimation: Long span truss bridge application with environmental monitoring data, *Engineering Structures* 30 (2008) 2347–2359.
- [4] E. Vaidogas, V. Juocevičius, Reliability of a timber structure exposed to fire: estimation using fragility function, *Mechanics* 73 (2016) 35–42.
- [5] J. Jaffari, M. Anis, Adaptive sampling for efficient failure probability analysis of sram cells, in: *Proceedings of the 2009 International Conference on Computer-Aided Design*, ACM, pp. 623–630.
- [6] C. Liu, Z. Lu, Y. Xu, Z. Yue, Reliability analysis for low cycle fatigue life of the aeronautical engine turbine disc structure under random environment, *Materials Science and Engineering: A* 395 (2005) 218–225.
- [7] I. Davison, N. Fairburn, Practical experience of failure analysis on nuclear facility or how I learned to love a well-structured fmea, in: *Safety and Reliability*, volume 36, Taylor & Francis, pp. 24–34.
- [8] A. Saltelli, Sensitivity analysis for importance assessment, *Risk Analysis* 22 (2002) 579–590.
- [9] B. Iooss, P. Lemaître, A review on global sensitivity analysis methods, in: *Uncertainty Management in Simulation-Optimization of Complex Systems*, Springer, 2015, pp. 101–122.
- [10] P. Wei, Z. Lu, J. Song, Variable importance analysis: a comprehensive review, *Reliability Engineering & System Safety* 142 (2015) 399–432.

- [11] I. M. Sobol', On sensitivity estimation for nonlinear mathematical models, *Matematicheskoe Modelirovanie* 2 (1990) 112–118.
- [12] J. Morio, Influence of input pdf parameters of a model on a failure probability estimation, *Simulation Modelling Practice and Theory* 19 (2011) 2244–2255.
- [13] E. Borgonovo, Measuring uncertainty importance: investigation and comparison of alternative approaches, *Risk analysis* 26 (2006) 1349–1361.
- [14] E. Borgonovo, E. Plischke, Sensitivity analysis: a review of recent advances, *European Journal of Operational Research* 248 (2016) 869–887.
- [15] E. Plischke, E. Borgonovo, C. L. Smith, Global sensitivity measures from given data, *European Journal of Operational Research* 226 (2013) 536–550.
- [16] Q. Liu, T. Homma, A new computational method of a moment-independent uncertainty importance measure, *Reliability Engineering & System Safety* 94 (2009) 1205–1211.
- [17] P. Wei, Z. Lu, J. Song, Moment-independent sensitivity analysis using copula, *Risk Analysis* 34 (2014) 210–222.
- [18] E. Plischke, E. Borgonovo, Copula-based sensitivity measures of computer experiments, *Safety and Reliability of Complex Engineered Systems* (2015).
- [19] L. Zhang, Z. Lu, L. Cheng, C. Fan, A new method for evaluating borgonovo moment-independent importance measure with its application in an aircraft structure, *Reliability Engineering & System Safety* 132 (2014) 163–175.
- [20] X. Luo, Z. Lu, X. Xu, A fast computational method for moment-independent uncertainty importance measure, *Computer Physics Communications* 185 (2014) 19–27.
- [21] M. M. Rajabi, B. Ataie-Ashtiani, C. T. Simmons, Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations, *Journal of Hydrology* 520 (2015) 101–122.
- [22] L. Zhang, Z. Lu, L. Cheng, D. Hong, Moment-independent regional sensitivity analysis of complicated models with great efficiency, *International Journal for Numerical Methods in Engineering* 103 (2015) 996–1014.
- [23] C. Zhou, Z. Lu, W. Li, Sparse grid integration based solutions for moment-independent importance measures, *Probabilistic Engineering Mechanics* 39 (2015) 46–55.
- [24] E. Borgonovo, W. Castaings, S. Tarantola, Moment independent importance measures: new results and analytical test cases, *Risk Analysis* 31 (2011) 404–428.
- [25] Z. I. Botev, J. F. Grotowski, D. P. Kroese, et al., Kernel density estimation via diffusion, *The Annals of Statistics* 38 (2010) 2916–2957.
- [26] J. Morio, M. Balesdent, D. Jacquemart, C. Vergé, A survey of rare event simulation methods for static input–output models, *Simulation Modelling Practice and Theory* 49 (2014) 287–304.
- [27] L. Devroye, L. Györfi, *Nonparametric density estimation: the L1 view*, John Wiley & Sons Incorporated, 1985.

- [28] P. Zhang, Nonparametric importance sampling, *Journal of the American Statistical Association* 91 (1996) 1245–1253.
- [29] B. E. Hansen, Uniform convergence rates for kernel estimation with dependent data, *Econometric Theory* 24 (2008) 726–748.
- [30] R. L. Iman, A matrix-based approach to uncertainty and sensitivity analysis for fault trees, *Risk analysis* 7 (1987) 21–33.
- [31] T. Buch-Larsen, J. P. Nielsen, M. Guillén, C. Bolancé, Kernel density estimation for heavy-tailed distributions using the Champernowne transformation, *Statistics* 39 (2005) 503–516.
- [32] C. Bolancé, M. Guillén, J. P. Nielsen, Inverse beta transformation in kernel density estimation, *Statistics & Probability Letters* 78 (2008) 1757–1764.