A nonparametric importance sampling estimator for moment independent importance measures

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

\textsuperscript{a}Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse
\textsuperscript{b}ONERA - The French Aerospace Lab, BP 74025, 31055 Toulouse Cedex, France
\textsuperscript{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
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 $Y = \Phi(X)$ where the output $Y$ depends on a $d$-dimensional real valued random variable $X = (X_1, ..., X_d)$
through a deterministic scalar function $\Phi : \mathbb{R}^d \to \mathbb{R}$ called “black box” or “model response”. For $I \subset \{1, \ldots, d\}$ a subset of indices we write $X_I = (X_i, i \in I)$. We assume throughout that for every $I \subset \{1, \ldots, d\}$ a strict subset (i.e., $I \neq \{1, \ldots, d\}$), the pair $(X_I, Y)$ is absolutely continuous with respect to Lebesgue measure with PDF $f_{X_I, Y}$. This implies in particular that the random variables $X_I$, $Y$ and $Y$ conditioned on $X_I = x_I$ for any $I \subset \{1, \ldots, d\}$ a strict subset, and any $x_I \in \mathbb{R}^{\text{Card}(I)}$ are also absolutely continuous with respect to Lebesgue measure, and we will denote by $f_{X_I, Y}$, $f_Y$ and $f_{X_I=x_I}$ 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 $f_{Y|X_i=x_i}$:

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

(1)

So as to consider the whole range of values the random variable $X_i$ can take into account, the sensitivity 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

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

(2)

For further references and more details on and 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 $I \subset \{1, \ldots, d\}$ by

$$\delta_I := \mathbb{E}[s(X_I)] \quad \text{with} \quad s(x_I) = \frac{1}{2} \left\| f_Y - f_{Y|X_I=x_I} \right\|_{L^1(\mathbb{R})} = \frac{1}{2} \int \left| f_Y(y) - f_{Y|X_I=x_I}(y) \right| \, dy .$$

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 $\delta_i$ 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

$$\delta_i = \frac{1}{2} \int f_{X_i}(x) \left( \int \left| f_Y(y) - f_{Y|X_i=x_i}(y) \right| \, dy \right) \, dx$$

$$= \frac{1}{2} \int f_{X_i}(x) \left| f_Y(y) - \frac{f_{X_i,Y}(x,y)}{f_{X_i}(x)} \right| \, dxdy$$

$$= \frac{1}{2} \left\| f_{X_i} f_Y - f_{X_i,Y} \right\|_{L^1(\mathbb{R}^2)} .$$
This interpretation opens the way for various estimation procedures of \( \delta_i \), one of them being the single-loop Monte-Carlo estimator based on rewriting the above expression as

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

(3)

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

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

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

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

(4)

and

\[
\hat{f}_{X_i,Y}(x,y) := \frac{1}{Nh_1h_2} \sum_{k=1}^{N} K \left( \frac{x - X^k}{h_1} \right) K \left( \frac{y - Y^k}{h_2} \right), \quad (x,y) \in \mathbb{R}^2,
\]

(5)

where \(K\) is the Gaussian kernel \( K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)\) 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 \( \delta_i \) by

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

(6)

This method combines one Monte–Carlo loop and KDE procedure and it requires only \(N\) calls to the black-box function \( \Phi \) for the estimation of the indices.

### 3.2. Application on 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

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

and so the moment independent measures \( \delta_i \) 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

\[
\delta_2 = \frac{1}{4\pi} \int_{\mathbb{R}^2} e^{-\frac{x^2}{10}} \left| \frac{1}{\sqrt{6}} e^{-\frac{y^2}{12}} - \frac{1}{\sqrt{5}} e^{-\frac{(y-x)^2}{10}} \right| \, dx \, dy.
\]
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 $\hat{\delta}_{SL}^1$ and $\hat{\delta}_{SL}^2$ of both importance measures $\delta_1$ and $\delta_2$. In [2], the sample size $N$ is set to be $N = 3,000$ for all numerical applications.

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

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

Since the true values of the standard deviation and the mean of $\hat{\delta}_{SL}^i$ are unknown, in the subsequent analysis we replace them by their estimations obtained with regular Monte-Carlo methods with 100 simulations. We will write $\bar{\delta}_{SL}^i$ for the estimation of $E(\hat{\delta}_{SL}^i)$ and $\bar{\sigma}_{SL}^i$ for the estimation of $\sqrt{\text{Var}(\hat{\delta}_{SL}^i)}$, so that the estimation of $\text{CV}(\hat{\delta}_{SL}^i)$ is given by $\frac{\bar{\delta}_{SL}^i}{\bar{\sigma}_{SL}^i}$ (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 $\text{CV}(\hat{\delta}_{SL}^i)$. With coefficients of variation estimated around 1% and 5%, it may be thought that $\hat{\delta}_{SL}^i$ has converged and should thus provide an accurate estimate of $\delta_i$. 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, it is not clear from these results that this estimator converges to the correct value. Indeed, since the theoretical values of the importance measures $\delta_i$ are available, the $(1 - \alpha)\%$ confidence interval of $E(\hat{\delta}_{SL}^i)$

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

where $z_\beta$ is the $\beta$-quantile of the standard normal distribution, and the relative difference

$$\frac{\hat{\delta}_{SL}^i - \delta_i}{\delta_i},$$

may be computed in order to appreciate the accuracy of the single-loop estimator $\hat{\delta}_{SL}^i$. We can note that both single-loop estimators have non negligible relative differences of the order of 15% and 40%, notably $\hat{\delta}_{SL}^2$ which strongly underestimates the measure $\delta_2$.

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

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Theoretical value $\delta_i$</th>
<th>$\hat{\delta}_{SL}^i$</th>
<th>Relative difference</th>
<th>$\bar{\sigma}_{SL}^i$</th>
<th>CV($\hat{\delta}_{SL}^i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>0.1436</td>
<td>0.1217</td>
<td>-0.1522</td>
<td>0.0055</td>
<td>0.0449</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.5382</td>
<td>0.3297</td>
<td>-0.3873</td>
<td>0.0046</td>
<td>0.0138</td>
</tr>
</tbody>
</table>
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 \((X^k, Y^k)\) 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 \(\hat{\delta}^{SL}_i\) is consistent, i.e., that \(\hat{\delta}^{SL}_i \xrightarrow{a.s.} \delta_i\). 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](image1.png)

**Figure 1.** Single-loop estimates of the importance measures \(\delta_i\) 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 \(\delta_i\) based on its expression (3).

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

**Step IS2.** Same as step \(SL2\), in particular get the two estimators \(\hat{f}_{X,Y}\) and \(\hat{f}_Y\) from \((X^k, Y^k)\) 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 \((X^k, Y^k)\). Let \((U^1, \ldots, U^{N'})\) be \(N'\) i.i.d. random variables drawn according to \(g\).
with \( U^k = (U^k_1, U^k_2) \in \mathbb{R}^2 \). Our estimator \( \delta_i^{IS,g} \) of \( \delta_i \) is

\[
\delta_i^{IS,g} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{\left| \hat{f}_Y(U^k_2) f_{X_i}(U^k_1) - \hat{f}_{X_i,Y}(U^k) \right|}{g(U^k)}.
\]

The study of the theoretical properties of \( \delta_i^{IS,g} \) 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 \( \Phi \). 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 \((X^k, Y^k)\). In particular, we can take \( g = f_{X_i,Y} \) in which case \( \delta_i^{IS,g} \) is given by

\[
\delta_i^{IS,g} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{|f_{X_i}(U^k_1) \hat{f}_Y(U^k_2) - \hat{f}_{X_i,Y}(U^k)|}{\hat{f}_{X_i,Y}(U^k)}.
\]

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

### 4.2. Theoretical properties of \( \delta_i^{IS,g} \)

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

\[
\text{Supp}(f_{X_i} \times f_Y - f_{X_i,Y}) \subset \text{Supp}(g), \quad (A1)
\]

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 \( \delta_i^{IS,g} \) is asymptotically unbiased and converges to \( \delta_i \) when we let first \( N' \to \infty \) and then \( N \to \infty \).

**Proposition 2.** Assume that Assumption (A1) holds and that \( Nh, Nh_1 h_2 \to \infty \) with \( h, h_1 h_2 \to 0 \). Then \( \sup_{N'} \mathbb{E}(\hat{\delta}_i^{IS,g}) \to \delta_i \) as \( N \to \infty \) and \( \delta_i^{IS,g} \xrightarrow{a.s.} \delta_i \) as \( N', N \to \infty \), where the convergence \( \xrightarrow{a.s.} \) means that we let first \( N' \to \infty \) and then \( N \to \infty \).

**Proof.** Let

\[
\hat{\Delta} = \frac{1}{2} \left\| f_{X_i,Y} - \hat{f}_{X_i,Y} \right\|_{L_1(\text{Supp}(g))}.
\]
Then $E(\hat{\delta}^{IS,g}) = E(\hat{\Delta})$, independent of $N'$, and the strong law of large numbers implies that $\hat{\delta}^{IS,g} \xrightarrow{a.s.,N \to \infty} \Delta$. Furthermore, the assumptions on the bandwidth imply that $\hat{\Delta} \xrightarrow{a.s.,N \to \infty} \delta_i$, see for instance [27, Theorem 1]. This shows the second claim of the statement, from which the first one follows by dominated convergence since $\hat{\Delta} \leq 1$. □

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

$$\text{Var}(\hat{\delta}^{IS,g}) \quad (7) \quad = \text{Var}(E[\hat{\delta}^{IS,g} | (X^k,Y^k)]) + E[\text{Var}(\hat{\delta}^{IS,g} | (X^k,Y^k))],$$

$$= \text{Var}(E\left[\frac{1}{2N'} \sum_{k=1}^{N'} \hat{h}(U^k) | (X^k,Y^k)\right]) + E\left[\text{Var}\left(\frac{1}{2N'} \sum_{k=1}^{N'} \hat{h}(U^k) | (X^k,Y^k)\right)\right],$$

$$= \text{Var}(\hat{\Delta}) + \frac{1}{4N'} E\left[\text{Var}(\hat{h}(U) | (X^k,Y^k))\right], \quad (8)$$

with

$$\hat{h}(x,y) = 1_{g(x,y) > 0} \frac{|\hat{f}_Y(y)f_X(x) - \hat{f}_X,y(x,y)|}{g(x,y)}.$$

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

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

$$\text{Var}(\hat{h}(U) | (X^k,Y^k)) = \int \frac{|\hat{f}_Y f_X - \hat{f}_X,y|^2}{g} - \left(\int |\hat{f}_Y f_X - \hat{f}_X,y|\right)^2,$$

may be infinite if the distribution $g$ is not well chosen. Nevertheless, in practice, assuming that $g$ is nearly proportional to $g_{opt} = |\hat{f}_Y f_X - \hat{f}_X,y|$, 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 $\Phi$.

We conclude this theoretical section with a more detailed study of the consistency of $\hat{\delta}^{IS,g}$. The convergence result $\hat{\delta}^{IS,g} \xrightarrow{a.s.,N' \to \infty} \delta_i$ 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 $\text{IS2}$ and the Monte-Carlo step $\text{IS3}$. In other words, in practice $N$ and $N'$ grow large together while the convergence $\hat{\delta}^{IS,g} \xrightarrow{a.s.,N' \to \infty} \delta_i$ would amount to first give all the budget to the Monte-Carlo step $\text{IS3}$. Assuming that the auxiliary function $g$ does not depend on the sample $(X^k,Y^k)$, 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. \( f_X \) is bounded;
2. \( f_Y \) and \( f_{X,Y} \) are bounded, twice differentiable and with uniformly continuous second derivative;
3. Condition (A1) holds and \( g \) has bounded support;
4. the bandwidths satisfy \( h_1 = h_2 = h, \ h \to 0 \) and \( h \gg \ln \ln N \sqrt{\ln N/N} \), i.e.,
\[
\frac{(\ln \ln N)^2 \ln N}{Nh^2} \xrightarrow{N \to \infty} 0.
\] (9)
Assume finally that \( N' = N'(N) \) depends on \( N \) in such a way that \( N' \to \infty \) as \( N \to \infty \). Then \( \delta_{IS,g} \xrightarrow{N \to \infty} \delta_1 \).

Proposition 4. Assume that:
1. \( f_X \) is bounded;
2. \( f_Y \) and \( f_{X,Y} \) are bounded, twice differentiable and with uniformly continuous second derivative;
3. there exists \( q > 2 \) such that
\[
\int \|y\|^q f_Y(y)dy < \infty \quad \text{and} \quad \int \|(x,y)\|^q f_{X,Y}(x,y)dxdy < \infty;
\]
4. \( g \) has unbounded support and there exists \( \alpha < 1 \) such that \( \int g^\alpha < \infty \);
5. the bandwidths satisfy \( h_1 = h_2 = h, \ h \to 0 \) and (9).
Assume finally that \( N' = N'(N) \) depends on \( N \) in such a way that there exists \( \beta > \frac{2}{1-\alpha} \) with
\[
(N')^\beta \frac{\ln N}{\sqrt{Nh}} \xrightarrow{N \to \infty} 0.
\]
Then \( \delta_{IS,g} \xrightarrow{N \to \infty} \delta_1 \).

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 \( \text{IS3} \). To achieve this, we apply steps \( \text{IS1} \) and \( \text{IS2} \) on the previous Gaussian toy case with parameter \( N = 5,000 \), getting an approximation \( \frac{1}{2} \int |f_X, \hat{f}_Y - \hat{f}_{X,Y}| \) of the indice \( \delta_i \) from the sample \((X^k_i, Y^k_i)\). Then, we perform \( m = 100 \) runs of step \( \text{IS3} \) with \( N' = 2 \times 10^4 \) using the different sampling distributions defined as follows:

1. The kernel estimator \( g_1 = \hat{f}_{X,Y} \), of the joint distribution of \( X_i \) and \( Y \). The following procedure can be used to generate a sample \( S \) from \( g_1 \):

   Step 1. Generate \( Z \) uniformly on \( \{1, \ldots, N\} \).

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

2. The distribution \( g_2(x,y) = f_X(x) \times \hat{f}_Y(y) \), where \( \hat{f}_Y \) is the kernel density estimator of the output distribution. The following procedure can be used for generating a random variate \( S \) from \( g_2 \):
Step 1. Generate $Z$ uniformly on $\{1, ..., N\}$.

Step 2. Exit with $S \sim f_X \otimes N(Y^Z, h)$.

3. The PDF $g_3$ of the Gaussian distribution $N(\hat{m}_i, \hat{\Sigma}_i)$ where

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

and

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

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

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

As in Section 3.2, since the true values of the standard deviation and the mean of $\hat{\delta}^{IS,g_j}$ 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 $\delta_1$ obtained with the sampling distributions $g_1$ are quite similar with coefficients of variation between 1 and 2%. On the other hand, we observe that the estimates of $\delta_2$ 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 $\hat{\delta}^{IS,g_j}$ depends on how the distribution $g$ samples well on all the support of the integrand function $|f_{X_i} \hat{f}_Y - \hat{f}_{X_i,Y}|$. This is confirmed by Figure 3 where we observe that the samples obtained using $g_1$ (Figure 3(c)) and $g_3$ (Figure 3(e)) are not distributed wherever the function $|f_{X_i} \hat{f}_Y - \hat{f}_{X_i,Y}|$ 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 $\delta_2$ obtained with $g_4$ and mostly $g_2$ which presents the lowest coefficient 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 indice $\delta_1$.

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 $\delta_i$, which may explain the inaccuracy of the single-loop estimator.
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 Section 4.4.

Figure 2. Display of the different samples \((U^1, ..., U^{N'})\) used for the estimations of the measure \(\delta_1\) 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 \(|\hat{f}_{X_1}f_Y - \hat{f}_{X_1,Y}|\).
Table 2. Estimates of the indices of the Gaussian toy case by performing \( m = 100 \) runs of step IS3 with different sampling distributions.

<table>
<thead>
<tr>
<th>Input</th>
<th>( \delta_i )</th>
<th>( g_1 ) Mean</th>
<th>( CV_{g_1} )</th>
<th>( g_2 ) Mean</th>
<th>( CV_{g_2} )</th>
<th>( g_3 ) Mean</th>
<th>( CV_{g_3} )</th>
<th>( g_4 ) Mean</th>
<th>( CV_{g_4} )</th>
<th>( g_{\text{opt}} ) Mean</th>
<th>( CV_{g_{\text{opt}}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 )</td>
<td>0.1436</td>
<td>0.1465</td>
<td>0.0278</td>
<td>0.1474</td>
<td>0.0083</td>
<td>0.1470</td>
<td>0.0172</td>
<td>0.1471</td>
<td>0.0191</td>
<td>0.1473</td>
<td>0.0038</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>0.5382</td>
<td>0.4685</td>
<td>1.1882</td>
<td>0.5175</td>
<td>0.0097</td>
<td>0.4644</td>
<td>0.4822</td>
<td>0.5166</td>
<td>0.0237</td>
<td>0.5174</td>
<td>0.0050</td>
</tr>
</tbody>
</table>

(a) Contour lines of \( |f_{X_2} f_Y - f_{X_2} | \).
(b) Contour lines of \( |f_{X_2} \hat{f}_Y - f_{X_2} | \).
(c) \((U^1, ..., U^{N'}) \sim g_1 \otimes N'\).
(d) \((U^1, ..., U^{N'}) \sim g_2 \otimes N'\).
(e) \((U^1, ..., U^{N'}) \sim g_3 \otimes N'\).
(f) \((U^1, ..., U^{N'}) \sim g_4 \otimes N'\).
(g) \((U^1, ..., U^{N'}) \sim g_{\text{opt}} \otimes N'\).

Figure 3. Display of the different samples \((U^1, ..., U^{N'})\) used for the estimations of the measure \( \delta_2 \) of the Gaussian toy case in comparison to the contour lines of the integrand function \( |f_{X_2} f_Y - f_{X_2} | \) and its KDE estimation \( |f_{X_2} \hat{f}_Y - f_{X_2} | \).
Table 3. Estimates of the indices of the Gaussian toy case by performing \( m = 100 \) runs of the proposed procedure with different sampling distributions.

<table>
<thead>
<tr>
<th>Input ( \delta_i )</th>
<th>( g_2 )</th>
<th>( g_4 )</th>
<th>( \hat{g}_{opt} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Mean} )</td>
<td>( CV )</td>
<td>( \text{Mean} )</td>
<td>( CV )</td>
</tr>
<tr>
<td>( X_1 )</td>
<td>0.1436</td>
<td>0.1358</td>
<td>0.0439</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>0.5382</td>
<td>0.5188</td>
<td>0.0158</td>
</tr>
</tbody>
</table>

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 \( \hat{\delta}^{IS,g}_i \) associated to the distribution \( g \) given by

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

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 parameters \( N = 5000 \) 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 \( \delta_1 \) are quite similar with the same coefficient of variation around 4%. As far as the estimation of \( \delta_2 \) 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 model output is given by

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

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

We display in Table 4 the corresponding results of this model. On this example, the variability associated to the use of the distribution \( g_4 \) is significant with coefficients of variation of 13% and 15% while one obtains a coefficient 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 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 as close as possible to the support of \( [f_{X_i} \times f_Y - f_{X_i,Y}] \).
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.

<table>
<thead>
<tr>
<th>Input</th>
<th>$g_2$ Mean</th>
<th>$g_4$ Mean</th>
<th>$\hat{g}_{opt}$ Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>0.2980 0.0300</td>
<td>0.2982 0.1332</td>
<td>0.2966 0.0293</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.4901 0.0436</td>
<td>0.4935 0.1584</td>
<td>0.5186 0.0467</td>
</tr>
</tbody>
</table>

4.4. The optimal sampling distribution

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

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

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 $\hat{g}_{opt}$.
Appendix C

4.3

The following procedure can be used for generating a sample

\[ \left\{ \tilde{X}_i, \tilde{Y}_i \right\} \]

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.

- Generate a sample \( \left\{ (X_1, Y_1), \ldots, (X_{N''}, Y_{N''}) \right\} \) according to an initial distribution \( g_0 \).

- Compute the weights

\[ \omega(X_i, Y_i) = \frac{f_{X_i}(X_i) \hat{f}_Y(Y_i) - \hat{f}_{X_i,Y}(X_i, Y_i)}{g_0(X_i, Y_i)}, \quad k = 1, \ldots, N''. \]

- Estimate \( g_{opt} \) by the weighted kernel estimator

\[ \hat{g}_{opt}(x, y) = \frac{1}{N''h_1h_2\hat{\omega}} \sum_{k=1}^{N''} \omega(X_i, Y_i)K \left( \frac{x - \tilde{X}_i}{h_1} \right) K \left( \frac{y - \tilde{Y}_i}{h_2} \right), \quad (10) \]

where \( \hat{\omega} = \frac{1}{N''} \sum_{k=1}^{N''} \omega(X_i, Y_i) \). The convergence of the estimator \( \hat{g}_{opt} \) is studied in the next proposition whose proof is deferred to Appendix C.

**Proposition 5.** Let \( \hat{s} = |f_{X_i} \hat{f}_Y - \hat{f}_{X_i,Y}| \) and assume that:

1. \( \hat{s} / g_0 \) are bounded;
2. \( \int \hat{s}^\alpha / g_0^{\alpha - 1} < \infty \) for some \( \alpha > 2 \);
3. the bandwidths satisfy \( h_1 = h_2 = h, \quad \hat{h} \rightarrow 0 \) and \( \sqrt{N''}h \gg \ln \ln N'' \sqrt{\ln N''} \).

Then as \( N'' \rightarrow \infty \), we have \( \hat{g}_{opt}(x, y) \overset{a.s.}{\to} g_{opt}(x, y) \) for every \( x, y \).

We can now explain our estimation procedure:

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

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

**Step Opt3.** Our estimator \( \hat{\delta}_{opt} \) of \( \delta_i \) is given by

\[ \hat{\delta}_{opt}^i = \hat{\delta}_{opt}^{IS,g_{opt}} = \frac{1}{2N'} \sum_{k=1}^{N'} \frac{f_Y(U^k_2)f_{X_i}(U^k_1) - \hat{f}_{X_i,Y}(U^k)}{g_{opt}(U^k)}, \quad (11) \]

where the random variables \( U^k = (U^k_1, U^k_2) \) are i.i.d according to the distribution \( \hat{g}_{opt} \) defined above. The following procedure can be used for generating a sample \( U \) from \( \hat{g}_{opt} \):

Step 1. Generate \( Z \sim \sum_{k=1}^{N'} \frac{\omega(X_i, Y_i)}{\hat{\omega}} \epsilon_k \) with \( \epsilon_k \) the Dirac measure at \( k \).

Step 2. Exit with \( U \sim \mathcal{N} \left( \begin{pmatrix} X_i \\ Y_i \end{pmatrix}, \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix} \right) \).

We perform the same studies as in Section 4.3 with this algorithm and with the parameters \( N = 5000 \) 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 $\hat{\delta}^{Opt}$. It can be seen in Table 2 that $\hat{\delta}^{Opt}$ presents the best robustness with a coefficient of variation of 0.5%, nearly twice smaller than that for $\hat{\delta}^{2}$. 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 $g_{2}$ is 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 $\Phi$ is expensive. Henceforth, the estimator $\hat{\delta}^{Opt}_{i}$ 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 linear 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 concerns 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 $\delta_{i}$ is the coefficient of variation

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

For each estimator $\hat{\delta}_{i}$ of the importance measure $\delta_{i}$, we approximate its mean and its standard deviation using Monte–Carlo procedure. Considering $m$ estimates ($\hat{\delta}_{i}^{1}, ..., \hat{\delta}_{i}^{m}$), 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 $\hat{\delta}$.
5.1. Example 1: an linear transformation of a Gaussian vector

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

\[ Y_1 = AX, \]

where \( A = [1.7 \ 1.8 \ 1.9 \ 2] \) and where the input \( X \) follows the Gaussian distribution \( N(0, \Sigma) \) where the covariance matrix \( \Sigma \) 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:

\[ Y_1 \sim N(0, A \Sigma A^t), \]

and

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

where the mean \( m_i \) and the variance \( \sigma_i^2 \) are defined by:

\[
m_i = A_i x_i + A_{-i} C_i \Sigma_{ii}^{-1} x_i,
\]

\[
\sigma_i^2 = A_{-i} (\Sigma_{-i} - iC_i C_i \Sigma_{ii}^{-1}) A_{-i}^t,
\]

where \( A_{-i} \) is the vector \( A \) with its \( i \)-th component \( A_i \) removed, \( \Sigma_{-i} \) is the matrix \( \Sigma \) with its \( i \)-th row and column removed, and \( C_i \) the \( i \)-th column of \( \Sigma \) with its \( i \)-th component \( \Sigma_{ii} \) removed. 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 parameters \( N'' = N = 5000 \) 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 \( \delta_i \) 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 \( \Phi \) is expensive. The results are listed in Table 5. We can notice that both methods respect the good ranking, i.e., \( X_3 > X_2 > X_4 > X_1 \), and converge since the estimates are obtained with a coefficient of variation around 2%. However, the estimates of \( (\hat{\delta}^{IS}_i)_{1 \leq i \leq 4} \) 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 \( \delta_i \), the higher the relative difference associated to the estimate \( \delta^{SL}_i \).

These values of relative differences highlight the accuracy 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.

<table>
<thead>
<tr>
<th>Input</th>
<th>Theoretical value δ_i</th>
<th>Single loop δ^SL_i</th>
<th>Proposed method δ^Opt_i</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>STD</td>
</tr>
<tr>
<td>X_1</td>
<td>0.2857</td>
<td>0.2201</td>
<td>0.0062</td>
</tr>
<tr>
<td>X_2</td>
<td>0.3620</td>
<td>0.2620</td>
<td>0.0059</td>
</tr>
<tr>
<td>X_3</td>
<td>0.3792</td>
<td>0.2690</td>
<td>0.0059</td>
</tr>
<tr>
<td>X_4</td>
<td>0.3176</td>
<td>0.2383</td>
<td>0.0052</td>
</tr>
</tbody>
</table>

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 \( \omega \) (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

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

The six inputs \( F_1, F_2, E, \omega, t \) and \( L \) are normally distributed with their parameters 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 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 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 \(\delta_i\) with our own computer, among 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 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.

<table>
<thead>
<tr>
<th>Input</th>
<th>Distribution</th>
<th>Mean</th>
<th>coefficient of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>Normal</td>
<td>500 lb</td>
<td>0.08</td>
</tr>
<tr>
<td>$F_2$</td>
<td>Normal</td>
<td>1000 lb</td>
<td>0.08</td>
</tr>
<tr>
<td>$E$</td>
<td>Normal</td>
<td>$2.9 \times 10^7$ psi</td>
<td>0.08</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Normal</td>
<td>2.4487 in</td>
<td>0.08</td>
</tr>
<tr>
<td>$t$</td>
<td>Normal</td>
<td>3.8884 in</td>
<td>0.08</td>
</tr>
<tr>
<td>$L$</td>
<td>Normal</td>
<td>100 in</td>
<td>0.08</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Single loop $\delta_i^{SL}$</th>
<th>Proposed method $\delta_i^{Opt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>STD</td>
</tr>
<tr>
<td>$F_1$</td>
<td>0.0556</td>
<td>0.0045</td>
</tr>
<tr>
<td>$F_2$</td>
<td>0.0472</td>
<td>0.0046</td>
</tr>
<tr>
<td>$E$</td>
<td>0.0742</td>
<td>0.0053</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.1566</td>
<td>0.0056</td>
</tr>
<tr>
<td>$t$</td>
<td>0.1245</td>
<td>0.0057</td>
</tr>
<tr>
<td>$L$</td>
<td>0.1985</td>
<td>0.0057</td>
</tr>
</tbody>
</table>

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

$$Y_3 = X_1X_3X_5 + X_1X_3X_6 + X_1X_4X_6 + X_2X_3X_4 + X_2X_3X_5 + X_2X_4X_5 + X_2X_5X_6 + X_2X_4X_7 + X_2X_6X_7$$,

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

We estimate the seven indices with $m = 100$ runs of both single-loop Monte-Carlo and nonparametric importance sampling schemes with parameters $N'' = N = 5000$ 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 $\delta_i$ with our own computer, whose 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 ranking, mainly $X_2 > X_6 > X_3 > X_4 > X_7 > X_1 > X_3$. 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, it 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 $\delta_i$ are available.

5.3.2. Study of a theoretical multiplicative model

Let us consider the following multiplicative model output:

$$Y_d = \prod_{i=1}^{d} A_i$$,
Table 8. Distributions of the input variables of the example 3.

<table>
<thead>
<tr>
<th>Input</th>
<th>Distribution</th>
<th>Mean of $\ln(X_i)$</th>
<th>Variance of $\ln(X_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>Lognormal</td>
<td>0.6044</td>
<td>0.1776</td>
</tr>
<tr>
<td>$X_2$</td>
<td>Lognormal</td>
<td>1.0098</td>
<td>0.1776</td>
</tr>
<tr>
<td>$X_3$</td>
<td>Lognormal</td>
<td>-6.9965</td>
<td>0.1776</td>
</tr>
<tr>
<td>$X_4$</td>
<td>Lognormal</td>
<td>-6.3034</td>
<td>0.1776</td>
</tr>
<tr>
<td>$X_5$</td>
<td>Lognormal</td>
<td>-5.6103</td>
<td>0.1776</td>
</tr>
<tr>
<td>$X_6$</td>
<td>Lognormal</td>
<td>-5.3871</td>
<td>0.1776</td>
</tr>
<tr>
<td>$X_7$</td>
<td>Lognormal</td>
<td>-5.89792</td>
<td>0.1776</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Single loop $\hat{\delta}_{i}^{SL}$</th>
<th>Proposed method $\hat{\delta}_{i}^{Opt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>STD</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.0754</td>
<td>0.0051</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.1908</td>
<td>0.0051</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0.0588</td>
<td>0.0045</td>
</tr>
<tr>
<td>$X_4$</td>
<td>0.1023</td>
<td>0.0048</td>
</tr>
<tr>
<td>$X_5$</td>
<td>0.1363</td>
<td>0.0053</td>
</tr>
<tr>
<td>$X_6$</td>
<td>0.1534</td>
<td>0.0051</td>
</tr>
<tr>
<td>$X_7$</td>
<td>0.0778</td>
<td>0.0048</td>
</tr>
</tbody>
</table>

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

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

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

and the theoretical values of the importance measures $\delta_i$ 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 $\delta_i$ of the multiplicative model $Y_4$ by setting different values of the means $m_i$ and fixing $(\sigma^2_i)_{1 \leq i \leq 3} = [0.1776 0.1776 0.1776]$. 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 shows the estimates obtained by performing $m = 100$ repeated simulations of both single-loop Monte–Carlo and nonparametric importance sampling methods with parameters $N'' = N = 5000$ and $N' = 2 \times 10^4$. We find that in each case, the nonparametric importance sampling estimates present an higher accuracy than the single-loop estimates.
Table 10. Estimates of the indices of the multiplicative model $Y_4$ for different values of the means $m_i$ [parameter setup: $d = 3$, $(\sigma_i^2)_{1 \leq i \leq 3} = [0.1776 \ 0.1776 \ 0.1776]$].

<table>
<thead>
<tr>
<th>Input</th>
<th>Theoretical value</th>
<th>Single loop $\delta_i^{SL}$</th>
<th>Proposed method $\delta_i^{Opt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(A_1, A_2, A_3)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(X_1, X_3, X_5)$</td>
<td>0.2239</td>
<td>0.1880 0.0049 -0.1603</td>
<td>0.2225 0.0063 -0.0063</td>
</tr>
<tr>
<td>$(X_1, X_4, X_6)$</td>
<td>0.2239</td>
<td>0.1862 0.0058 -0.1684</td>
<td>0.2198 0.0066 -0.0184</td>
</tr>
<tr>
<td>$(X_2, X_6, X_7)$</td>
<td>0.2239</td>
<td>0.1889 0.0048 -0.1561</td>
<td>0.2228 0.0059 -0.0047</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Theoretical value</th>
<th>Single loop $\delta_i^{SL}$</th>
<th>Proposed method $\delta_i^{Opt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean  STD  RD</td>
<td>Mean  STD  RD</td>
</tr>
<tr>
<td>$d = 4$</td>
<td>0.1846</td>
<td>0.8757 0.4570 3.7440</td>
<td>0.3972 0.0396 1.1517</td>
</tr>
<tr>
<td>$d = 5$</td>
<td>0.1604</td>
<td>2.0302 1.6746 11.6574</td>
<td>0.4570 0.0329 1.8492</td>
</tr>
<tr>
<td>$d = 6$</td>
<td>0.1436</td>
<td>3.6376 1.9370 24.3318</td>
<td>0.5036 0.1267 2.5068</td>
</tr>
</tbody>
</table>

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 $i$ so that it is sufficient to compute the index $\delta_1$ since all inputs has the same influence. We report in Table 11 the theoretical values of $\delta_1$ computed by Borgonovo [24] and the estimates of the index $\delta_1$ obtained by applying both methods $m = 100$ times with parameters $N'' = N = 5000$ 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 $\hat{\Delta} = \frac{1}{2} \left\| f_{X_i} f_Y - f_{X_i, Y} \right\|_{L_1(B_2)}$. 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 $|f_{X_i} f_Y - f_{X_i, Y}|$ takes large values (Figure 5(a)), notably as far as the coordinate corresponding to the output is concerned.

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 implementation
Appendix A. Proof of Proposition 3

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

**Theorem 6** (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

\[
c(N) = O \left( \ln \ln N \sqrt{\ln N} \right)^{1/2q}.
\]
We now prove Proposition 3. First of all, let
\[
h(x, y) = 1_{g(x, y) > 0} \frac{|f_Y(y)f_{X_i}(x) - f_{X_i,Y}(x, y)|}{g(x, y)}, \; x, y \in \mathbb{R}^2.
\]
Since
\[
\frac{1}{2N'} \sum_{k=1}^{N'} h(U^k) \xrightarrow{a.s.} N \to \infty \delta_i
\]
by the strong law of large numbers, it is enough to prove that
\[
\hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(U^k) \xrightarrow{a.s.} N \to \infty 0.
\]
The triangular inequality gives
\[
\left| \hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(U^k) \right|
\leq \frac{1}{2N'} \sum_{k=1}^{N'} \frac{1}{g(U^k)} \left| f_{X_i}(U^k_1)f_Y(U^k_2) - \hat{f}_{X_i,Y}(U^k) - \left| f_{X_i}(U^k_1)f_Y(U^k_2) - f_{X_i,Y}(U^k) \right| \right|
\]
and using the reversed triangular inequality \(||a|| - ||b|| \leq ||b - a||\) and then again the triangular inequality we obtain
\[
\left| \hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(U^k) \right|
\leq \frac{1}{2N'} \sum_{k=1}^{N'} \frac{f_{X_i}(U^k_1)}{g(U^k)} \left| \hat{f}_Y(U^k_2) - f_Y(U^k_2) \right|
\leq \frac{1}{2N'} \sum_{k=1}^{N'} \frac{1}{g(U^k)} \left| \hat{f}_{X_i,Y}(U^k) - f_{X_i,Y}(U^k) \right|. \quad \text{(A.1)}
\]
Since \(g\) has bounded support, the random variables \(U^k\) live in a bounded set and so we can use Theorem 6 to get the existence of random constants \(C_1, C_2 < \infty\) such that
\[
\left| \hat{f}_Y(U^k_2) - f_Y(U^k_2) \right| \leq C_1 \epsilon_1(N), \; \left| \hat{f}_{X_i,Y}(U^k) - f_{X_i,Y}(U^k) \right| \leq C_2 \epsilon_2(N)
\]
which gives
\[
\left| \hat{\delta}_i^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(U^k) \right|
\leq \frac{C_1 \epsilon_1(N)}{2N'} \sum_{k=1}^{N'} \frac{f_{X_i}(U^k_1)}{g(U^k)} + \frac{C_2 \epsilon_2(N)}{2N'} \sum_{k=1}^{N'} \frac{1}{g(U^k)}.
\]
Since
\[
\frac{1}{N'} \sum_{k=1}^{N'} \frac{f_{X_i}(U^k)}{g(U^k)} \xrightarrow{a.s.} N \to \infty \int_{\text{Supp}(g)} (f_{X_i}(x) + 1)dx dy
\]
and \(\text{Supp}(g)\) is bounded, we obtain the result since \(\epsilon_1(N), \epsilon_2(N) \to 0\).

**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].

23
Theorem 7 (Extension of Theorem 7 in [29]). Assume that \( f_X \) is bounded, that \( f_X \) and \( f_{X,Y} \) are differentiable with uniformly continuous derivative, that there exists \( q > 2 \) such that
\[
\int \|y\|^{2q} f_Y(y) \, dy < \infty \quad \text{and} \quad \int \|(x,y)\|^{2q} f_{X,Y}(x,y) \, dx \, dy < \infty ,
\]
and that (9) holds with \( h = h_1 = h_2 \). Then
\[
\sup_{y \in \mathbb{R}} |\hat{f}_Y(y) - f_Y(y)| \leq O(\epsilon_1(N)) ,
\]
\[
\sup_{(x,y) \in \mathbb{R}^2} |f_{X,Y}(x,y) - f_{X,Y}(x,y)| \leq O(\epsilon_2(N)) ,
\]
almost surely, where \( \epsilon_1(N), \epsilon_2(N) \) are as in Theorem 6.

We now prove Proposition 4. We start from (A.1) to get
\[
\left| \hat{\delta}^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(U^k) \right| \leq \frac{C_1 \epsilon_1(N)}{2N'} \sum_{k=1}^{N'} \frac{f_{X,Y}(U^k)}{g(U^k)} + \frac{C_2 \epsilon_2(N)}{2N'} \sum_{k=1}^{N'} \frac{1}{g(U^k)} ,
\]
\[
\leq \frac{C(\epsilon_1(N) + \epsilon_2(N))}{2N'} \sum_{k=1}^{N'} \frac{1}{g(U^k)} .
\]

Now the problem is that
\[
\frac{1}{N'} \sum_{k=1}^{N'} \frac{1}{g(U^k)} \xrightarrow{\text{a.s.}} N' \rightarrow \infty \int_{\text{Supp}(g)} \, dx \, dy
\]
which is infinite since \( \text{Supp}(g) \) is unbounded. However, for \( \beta > \frac{2}{1-\alpha} \) we have
\[
\sup_{n \geq 1} \left( \frac{1}{n^{1+\beta}} \sum_{k=1}^{n} \frac{1}{g(U^k)} \right) < \infty \text{ a.s.} \quad (B.1)
\]
Indeed, for every \( \eta > 0 \) we have by the union bound
\[
P \left( \frac{1}{n^{1+\beta}} \sum_{k=1}^{n} \frac{1}{g(U^k)} \geq \eta \right) \leq nP \left( \frac{1}{g(U)} \geq \eta n^\beta \right)
\]
and so Markov inequality gives
\[
P \left( \frac{1}{n} \sum_{k=1}^{n} \frac{1}{g(U^k)} \geq \eta \right) \leq n \frac{1}{(\eta n^\beta)^{1-\alpha}} \mathbb{E} \left( \left( \frac{1}{g(U)} \right)^{1-\alpha} \right) = \frac{1}{\eta^{1-\alpha} n^{\beta(1-\alpha)-1}} \int g^\alpha .
\]
Since by assumptions we have \( \int g^\alpha < \infty \) and \( \beta(1-\alpha) - 1 > 1 \), we see that
\[
\sum_{n \geq 1} P \left( \frac{1}{n^{1+\beta}} \sum_{k=1}^{n} \frac{1}{g(U^k)} \geq \eta \right) < \infty
\]
which implies, by Borel–Cantelli lemma, that \( \frac{1}{n^{1+\beta}} \sum_{k=1}^{n} \frac{1}{g(U^k)} \xrightarrow{n \rightarrow \infty} 0 \), and in particular (B.1) holds. Up to random multiplicative constant, we thus have
\[
\left| \hat{\delta}^{IS,g} - \frac{1}{2N'} \sum_{k=1}^{N'} h(U^k) \right| \leq C(\epsilon_1(N) + \epsilon_2(N)) (N')^\beta
\]
and the assumption on \( N' \) makes this bound vanish.
Appendix C. Proof of Proposition 5

Let us define $V^k = (\tilde{X}^k, \tilde{Y}^k)$, $W^k = \frac{\hat{s}(V^k)}{g_0(V^k)}$ and

$$
\hat{g}(x, y) = \frac{1}{N''h_1h_2} \sum_{k=1}^{N''} W^k K \left( \frac{x - \tilde{X}^k}{h_1} \right) K \left( \frac{y - \tilde{Y}^k}{h_2} \right).
$$

Since the random variables $V^k$ are i.i.d, the sequence $\{V^k, W^k\}$ is strictly stationary and strong mixing. Furthermore, the normal Kernel belongs to the Schwartz space. Then, under the assumptions of Proposition 5, Theorem 7 in [29] ensures that for any $q > 0$

$$\sup_{||x,y)|| \leq c_{N''} |\hat{g}(x, y) - \hat{s}(x, y)| = O \left( \left( \frac{\ln(N'')}{N''h_1h_2} \right)^{1/2} + \tilde{h}_1 \tilde{h}_2 \right),$$

where $c_{N''} = O(\ln(\ln(N'')) \sqrt{\ln(N'')N''^{1/2q}})$. Thus as $N'' \to \infty$ we have $\hat{g}(x, y) \xrightarrow{a.s.} \hat{s}(x, y)$ for every $x, y$. Eventually, as $N'' \to \infty$ we have $\hat{g}_{opt}(x, y) \xrightarrow{a.s.} g_{opt}(x, y)$ for every $x, y$ according to the strong law of large numbers.


