Master M2RI

An introduction to sensitivity analysis

Part I

Variance-based sensitivity analysis and beyond

Agnès Lagnoux **lagnoux@univ-tlse2.fr**

Contents

Introduction

Mathematical models are used in many fields (one can think of environmental risk assessment, nuclear safety, aeronautics...) to model real phenomena. This modeling gives birth to a computer code that is then used to perform simulations of the model. These codes representing physical phenomena take as inputs (many) numerical parameters, physical variables (that can be real numbers, vectors or even functions) and give several outputs in general. Nevertheless, in real applications, such codes may be very time consuming.

In general the inputs parameters are not well known, one said that they are uncertain. In the statistical approach, we model this uncertainty by considering the inputs as random objects (random variables, random vectors or even stochastic processes). One of the aim of sensitivity analysis (SA) is to study how the uncertainty in the output is related to the input uncertainty. Hence SA can be for example used to detect the most influential variables, to detect the variables that are not influential (and then fixed them to some nominal value), to calibrate model inputs. There exists many different techniques to perform SA; several are local while others are global.

To illustrate the context of interest, we briefly present an example involving a functional input: the heat flow in a uniform rod. This example is borrowed from [35]. Consider the heat equation describing the evolution of the temperature $T(t,s)$ as a function of the position $s \in [0,L]$ ($L>0$), and the time $t \geqslant 0$ in a uniform rod with length L:

$$\frac{\partial T}{\partial t} = X_1 \frac{\partial^2 T}{\partial s^2},\ T(t,0) = T(t,L) = X_2,\ T(0,s) = X_2 + X_3(s).$$

Let $t_0 > 0$ be given and consider the output Y depending on the inputs X_1, X_2, X_3 in the following way: $Y = G(X_1,X_2,X_3) = T(t_0, L/2)$. This means that we are interested with the temperature at time t_0 in $L/2$ as a function of X_1, X_2, X_3. We have $d = 3$, the diffusivity coefficient X_1 and the border condition X_2 both lay in $E_1 = E_2 = \mathbb{R}_+$ while the random non negative process $X_3(s)$ lies in E_3 the set of all non negative continuous functions on $[0,L]$ vanishing at 0 and L. Here, the quantities X_1, X_2, X_3 on which the solution of the previous equation depends are random and assumed to be stochastically independent.

In this lecture, we focus on global sensitivity analysis (GSA) that is related to the ANOVA or Hoeffding decomposition. This technique is based on a decomposition of the variance that gives raise to the definition of indices (called Sobol' indices). As it will be shown later on, these indices can be seen as indicators on the importance of some inputs parameters.

Here, we mainly restrain our presentation to the statistical analysis of Sobol' indices. The first part of these notes are strongly inspired from a lecture given by Thierry Klein. Another source of inspiration is the book [35].

Chapter 1

Hoeffding decomposition and variance-based indices

1.1 Introductory example: the linear model

Let $\mathbb{X} = (X_1, \ldots, X_d)$ be random objects (inputs) and let $Y = G(\mathbb{X}) = G(X_1, \ldots, X_d)$ be the random output. Here G is assumed to be unknown. In some applications G is a computer code seen as a black box: if one gives to the computer the values of the inputs, the code returns an answer. In others applications, G can be the measurement of a real experience once the inputs are fixed. To model G, the first method used by statisticians is to propose an approximation defined by a linear function, that is to consider that

$$Y = \sum_{j=1}^{d} \beta_j X_j.$$

In this case, if the inputs are independent, we simply have

$$\mathrm{Var}(Y) = \sum_{j=1}^{d} \beta_j^2 \mathrm{Var}(X_j).$$

Hence $\beta_j^2 \frac{\mathrm{Var}(X_j)}{\mathrm{Var}(Y)}$ represents the part of the variance of Y that is due to the input X_j.

Now if the model is not linear, one can proceed an ANOVA-type decomposition of the variance of the output Y in order to quantify the importance of an input.

1.2 The ANOVA-Hoeffding decomposition of the variance

1.2.1 A simple example

In order to understand this decomposition, we will first consider a very simple example. Let X_1 and X_2 be two independent random variables distributed uniformly on $\{0,1\}$ and $\{0,1,2\}$ respectively. Let G be an application from $\{0,1\} \times \{0,1,2\}$ to $\mathbb{R}$ and $Y = G(\mathbb{X}) = G(X_1, X_2)$. Then one may write

$$G(\mathbb{X}) = G(X_1, X_2) = G_\emptyset + G_{\{1\}}(X_1) + G_{\{2\}}(X_2) + G_{\{1,2\}}(X_1, X_2), \tag{1.1}$$

where

$$G_\emptyset = \frac{1}{6}(G(0,0) + G(0,1) + G(0,2) + G(1,0) + G(1,1) + G(1,2)), \quad \text{(mean value of } G\text{)}$$

$$G_{\{1\}}(x_1) = \frac{1}{3}(G(x_1,0) + G(x_1,1) + G(x_1,2)) - G_\emptyset, \forall x_1 \in \{0,1\}, \quad \text{(deviation to the global behavior when only } x_1 \text{ varies)}$$

$$G_{\{2\}}(x_2) = \frac{1}{2}(G(0,x_2) + G(1,x_2)) - G_\emptyset, \forall x_2 \in \{0,1,2\}, \quad \text{(deviation to the global behavior when only } x_2 \text{ varies)}$$

$$G_{\{1,2\}}(x_1,x_2) = G(x_1,x_2) - G_{\{1\}}(x_1) - G_{\{2\}}(x_2) - G_\emptyset, \forall (x_1,x_2) \in \{0,1\} \times \{0,1,2\}, \quad \text{(residual part)}.$$

One can see that

$$\begin{aligned} G_\emptyset &= \mathbb{E}[G(X_1,X_2)],\\ G_{\{1\}}(X_1) &= \mathbb{E}[G(X_1,X_2)|X_1] - \mathbb{E}[G(X_1,X_2)],\\ G_{\{2\}}(X_2) &= \mathbb{E}[G(X_1,X_2)|X_2] - \mathbb{E}[G(X_1,X_2)],\\ G_{\{1,2\}}(X_1,X_2) &= G(X_1,X_2) - G_\emptyset(X_1,X_2) - G_{\{1\}}(X_1,X_2) - G_{\{1\}}(X_1,X_2). \end{aligned}$$

In addition, we have

$$\mathbb{E}[G_\emptyset G_{\{1\}}(X_1)] = G_\emptyset\Big(\frac{1}{3}\sum_{j=0}^{2}\mathbb{E}[G(X_1,j)]\Big) - G_\emptyset^2 = G_\emptyset\Big(\frac{1}{6}\sum_{j=0}^{2}\sum_{i=0}^{1}G(i,j)\Big) - G_\emptyset^2 = 0 = \mathbb{E}[G_\emptyset G_{\{2\}}(X_2)] \quad \text{(by symmetry)}$$

$$\mathbb{E}[G_{\{1\}}(X_1)G_{\{2\}}(X_2)] = \Big(\frac{1}{2}\sum_{i=0}^{1}\mathbb{E}[G(i,X_2)]\Big)\Big(\frac{1}{3}\sum_{j=0}^{2}\mathbb{E}[G(X_1,j)]\Big) - G_\emptyset\Big(\frac{1}{3}\sum_{j=0}^{2}\mathbb{E}[G(X_1,j)]\Big) - G_\emptyset\Big(\frac{1}{2}\sum_{i=0}^{1}\mathbb{E}[G(i,X_2)]\Big) + G_\emptyset^2 = 0$$

$$\mathbb{E}[G_\emptyset G_{\{1,2\}}(X_1,X_2)] = 0$$

$$\mathbb{E}[G_{\{1\}}(X_1)G_{\{1,2\}}(X_1,X_2)] = 0 = \mathbb{E}[G_{\{2\}}(X_2)G_{\{1,2\}}(X_1,X_2)] \quad \text{(by symmetry)}.$$

Hence the variables appearing in (1.1) are orthogonal so that we can simply perform a L^2-decomposition of the variance:

$$\mathrm{Var}\,(G(X_1,X_2)) = \mathrm{Var}\left(G_{\{1\}}(X_1)\right) + \mathrm{Var}\left(G_{\{2\}}(X_2)\right) + \mathrm{Var}\left(G_{\{1,2\}}(X_1,X_2)\right). \tag{1.2}$$

In summary, G can be decomposed as an orthogonal sum of constant, marginal, and joint functions. Then $G_{\{1,2\}}$ represents the part of G that cannot be explained by either global or marginal effects.

In the next section, we generalize (1.2) without specifying the law and the supports of the inputs.

1.2.2 General case

Let $(X_1,\dots,X_d)$ be independent random variables denoted by $\mathbb{X}$ and such that X_i belongs to a measurable Polish space[1] $(E_i,\mathscr{B}(E_i))$.

Example 1.1. *For example, $d=4$, X_1 a Poisson random variable with parameter $\lambda>0$ (i.e. for all $k\in\mathbb{N}$, $\mathbb{P}(X_1=k) = e^{-\lambda}\frac{\lambda^k}{k!}$), X_2 a Gaussian random variable with parameters m and σ^2, X_3 an exponential random variable with parameter 1 (i.e. with density $f(t)=\exp(-t)$, for $t\geqslant 0$), and X_4 a Cauchy random variable on $\mathbb{R}$ (i.e. with density $h(t)=\frac{1}{\pi(1+x^2)}$).*

Example 1.2. *For example, $d=3$, X_1 a Poisson random variable with parameter $\lambda>0$, X_2 a centered Gaussian vector of dimension 3 with covariance matrix Σ, and X_3 a Brownian motion.*

We denote by $L^2(P_{\mathbb{X}})$ the set of all measurable function f on $(E,\mathscr{E})$ such that $\mathbb{E}[f^2(\mathbb{X})]<\infty$ where $E=\prod_i^d E_i$ and $\mathscr{E}=\otimes_{i=1}^d\mathscr{B}(E_i)$. The space $L^2(P_{\mathbb{X}})$ is an Hilbert space with inner product defined by for any $f\in L^2(P_{\mathbb{X}})$ and $g\in L^2(P_{\mathbb{X}})$

$$\langle f,g\rangle = \mathbb{E}[f(\mathbb{X})g(\mathbb{X})].$$

For any A, $B\subset\{1,\dots,d\}$, we set $\mathbb{X}_A=(X_i)_{i\in A}$ and L^2_A the subspace of functions in $L^2(P_{\mathbb{X}})$ that are E_A measurable $(E_A=\prod_{i\in A}E_i)$ and

$$L^2_{B\perp A} = \{f\in L^2_B \mid \forall g\in L^2_A,\ \mathbb{E}[f(\mathbb{X}_B)g(\mathbb{X}_A)]=0\}.$$

[1] A Polish space is a separable completely metrizable topological space; that is, a space homeomorphic to a complete metric space that has a countable dense subset.

Theorem 1.3 (Hoeffding decomposition). *Let $G \in L^2(P_{\mathbb{X}})$. Then G may be uniquely decomposed in $L^2(P_{\mathbb{X}})$ as the following orthogonal expansion*

$$G(\mathbb{X}) = \sum_{A \subset \{1,\dots d\}} G_A(\mathbb{X}_A) \quad a.s. \tag{1.3}$$

where

1. *$G_\emptyset$ constant a.s.,*
2. *$\forall\, A \subset \{1,\dots d\}$, $A \neq \emptyset$, $\forall\, i \in A$, $\int_{E_i} G_A(\mathbf{x}_A)\mathbb{P}_{X_i}(dx_i) = 0$.*

The unique solution writes, $\forall\, A \subset \{1,\dots d\}$,

$$G_A(\mathbf{x}_A) = \sum_{B \subset A} (-1)^{|A|-|B|}\mathbb{E}[G(\mathbb{X})|\mathbb{X}_B = \mathbf{x}_B], \quad a.s. \tag{1.4}$$

Note that the results stated in this theorem are useful to define the Sobol' indices in the next section but also in

- Chapter 3 - Section 3.2.1 for the generalization of the Sobol' indices to multivariate and functional outputs,
- Chapter 5 - Section 5.2 for the definition of the Cramér-von Mises indices,
- Chapter 6 - Section 6.2 for some kernel generalizations in sensitivity analysis.

The proof could be done using induction and the same arguments as in the two dimensional example developed at the beginning of this chapter. A smarter elegant method developed in a general context is also presented. Indeed, we follow the proof proposed in [67] in the much more general context of a commuting projector collection (see Theorem 2.1 in [67]). In the sequel, we propose both proofs.

Proof of Theorem 1.3 by induction. For $d = 2$, we have (1.1). The different terms can be interpreted as follows: $G_\emptyset$ is the projection of G on $\Delta_\emptyset$, the set of constant functions on E. $G_{\{1\}}$ (resp. $G_{\{2\}}$) is the projection of G on $\iota^{\emptyset}_{\{1\}}$, the set of functions depending only on the first coordinate (resp. $\iota^{\emptyset}_{\{2\}}$, the set of functions on E depending only on the second coordinate). Moreover, as we have assumed that X_1 and X_2 are independent, $\iota^{\emptyset}_{\{1\}}$ and $\iota^{\emptyset}_{\{2\}}$ are orthogonal subspaces of $L^2(\mathbb{P}_{\mathbb{X}})$. Finally, $G_{\{1,2\}}$ is the projection of G on $\iota^{\{1\}}_{\{1,2\}} \cap \iota^{\{2\}}_{\{1,2\}}$ which is the orthogonal subspace of $\iota^{\emptyset}_{\{1\}} \cup \iota^{\emptyset}_{\{2\}} \cup \Delta_\emptyset$ in $L^2(\mathbb{P}_{\mathbb{X}})$.

Assume now that (1.3) holds for any dimension $2 \leqslant p \leqslant d-1$ and consider a square integrable function G of $(X_1, X_2, \dots, X_p, X_{p+1})$. Here, as before the components $X_1, \cdots, X_{p+1}$ are assumed to be independent. Set first $\tilde{X}_1 := (X_1, X_2, \dots, X_p)$ and $\tilde{X}_2 := X_{p+1}$. Using the orthogonal decomposition (1.1) for these two variables we may write

$$\begin{aligned}
\tilde{G}(\tilde{X}_1, \tilde{X}_2) &:= G(X_1, X_2, \dots, X_p, X_{p+1}) \\
&= \tilde{G}_\emptyset + \tilde{G}_{\{1\}}(\tilde{X}_1) + \tilde{G}_{\{2\}}(\tilde{X}_2) + \tilde{G}_{\{1,2\}}(\tilde{X}_1, \tilde{X}_2) \\
&= G_\emptyset + \tilde{G}_{\{1\}}(X_1, X_2, \dots, X_d) + \tilde{G}_{\{2\}}(X_2) + \tilde{G}_{\{1,2\}}(\tilde{X}_1, \tilde{X}_2).
\end{aligned}$$

Now, we may conclude using the induction hypothesis on $\tilde{G}_{\{1\}}(X_1, X_2, \dots, X_p)$ and observing that, from the expansion (1.1), $\tilde{G}_{\{1,2\}}(\tilde{X}_1, \tilde{X}_2)$ is centered and uncorrelated both from any square integrable function of $\tilde{X}_1 = (X_1, X_2, \dots, X_p)$ and any square integrable function of $\tilde{X}_2 = X_{p+1}$. □

Before turning to the second proof of Theorem 1.3, let us introduce some tools. A classical very interesting operator acting on $L^2(\mathbb{P}_{\mathbb{X}})$ with values in L^2_A is the orthogonal projector Proj_A. For $f \in L^2(\mathbb{P}_{\mathbb{X}})$, $\mathrm{Proj}_A f$ is the closest element in L^2_A of f (closest means having least L^2 deviation). That is,

$$\inf_{h \in L^2_A} \mathbb{E}[(f(\mathbb{X}) - h(\mathbb{X}_A))^2] = \min_{h \in L^2_A} \mathbb{E}[(f(\mathbb{X}) - h(\mathbb{X}_A))^2] = \mathbb{E}[(f(\mathbb{X}) - (\mathrm{Proj}_A f)(\mathbb{X}_A))^2].$$

As Proj_A is a projection, we have

$$\mathrm{Proj}_A^2 f = \mathrm{Proj}_A(\mathrm{Proj}_A f) = \mathrm{Proj}_A f \text{ so that } \mathrm{Proj}_A(f - \mathrm{Proj}_A f) = \mathrm{Proj}_A f - \mathrm{Proj}_A f = 0.$$

The last equality implies that $f - \mathrm{Proj}_A f$ is perpendicular to L^2_A. So that, for any $h \in L^2_A$,

$$\mathbb{E}[(f(\mathbb{X}) - (\mathrm{Proj}_A f)(\mathbb{X}_A))h(\mathbb{X}_A)] = 0.$$

Notice that this last property also characterizes $\mathrm{Proj}_A f$. That is, $\mathrm{Proj}_A f$ is the unique function g of L^2_A such that $f-g$ is perpendicular to any member of L^2_A. In other words,

$$\forall h \in L^2_A, \langle h, g\rangle = \langle h, f\rangle \Longleftrightarrow g = \mathrm{Proj}_A f. \tag{1.5}$$

Notice further that (1.5) also implies that Proj_A is self-adjoint. Indeed, for f and g square integrable functions, we may write

$$\langle \mathrm{Proj}_A f, g\rangle = \langle \mathrm{Proj}_A f, \mathrm{Proj}_A g\rangle = \langle \mathrm{Proj}_A g, \mathrm{Proj}_A f\rangle = \langle \mathrm{Proj}_A g, f\rangle = \langle f, \mathrm{Proj}_A g\rangle.$$

Last but not least, let I denotes the identity operator on $L^2(\mathbb{P}_{\mathbb{X}})$ ($If = f$ for all f). Then, $I - \mathrm{Proj}_A$ is the orthogonal projector on $(L^2_A)^\perp$ the space of all square integrable functions that are orthogonal to all members of L^2_A. As a matter of fact, for $f \in L^2(\mathbb{P}_{\mathbb{X}})$ and $h \in (L^2_A)^\perp$, we may write

$$\langle h, (I - \mathrm{Proj}_A) f\rangle = \langle h, f\rangle - \langle h, \mathrm{Proj}_A f\rangle = \langle h, f\rangle,$$

and we may conclude using once more time (1.5) that $I - \mathrm{Proj}_A = \mathrm{Proj}_{A^\perp}$. A classical result of probability theory (see, e.g., [120]), is that $\mathrm{Proj}_A f$ is the conditional expectation with respect to the σ-field $\mathscr{E}_{\mathscr{A}}$. In other words, we have,

$$(\mathrm{Proj}_A f)(\mathbb{X}_A) = \mathbb{E}[f(\mathbb{X})|X_A].$$

Now, under the assumption that the components of $\mathbb{X}$ are independent, the computation of $\mathbb{E}[f(\mathbb{X})|X_A]$ becomes easier as described in the following lemma.

Lemma 1.4. *da2021basics Let $f \in L^2(\mathbb{P}_{\mathbb{X}})$ and $A, B \subset \{1,\dots d\}$. Recall that $\overline{A}$ denotes the complementary set of A. Then, under the assumption that the components of $\mathbb{X}$ are independent, we have, for almost $\mathbb{P}_{\mathbb{X}}$ all $\mathbf{x}$ in E,*

$$(Proj_A f)(\mathbf{x}) = \mathbb{E}[f(\mathbb{X})|\mathbb{X}_A = \mathbf{x}_A] = \int_{E_{\overline{A}}} f(\mathbf{x})\mathbb{P}(d\mathbf{x}_{\overline{A}}). \tag{1.6}$$

$$(Proj_B Proj_A f)(\mathbf{x}) = (Proj_{A\cap B} f)(\mathbf{x}) = (Proj_A Proj_B f)(\mathbf{x}). \tag{1.7}$$

Proof of Lemma **??**. For $h \in L^2_A$, as $E = E_A \times E_{\overline{A}}$ and by the independence assumption we may write,

$$\begin{aligned}
\langle h, f\rangle = \mathbb{E}[hf] &= \int_{E_A \times E_{\overline{A}}} h(\mathbf{x}_A) f(\mathbf{x}) \mathbb{P}_{\overline{A}}(d\mathbf{x}_{\overline{A}}) \mathbb{P}_A(d\mathbf{x}_A) \\
&= \int_{E_A} h(\mathbf{x}_A) \left(\int_{E_{\overline{A}}} f(\mathbf{x}) \mathbb{P}_{\overline{A}}(d\mathbf{x}_{\overline{A}}) \right) \mathbb{P}_A(d\mathbf{x}_A) \\
&= \int_{E_A} h(\mathbf{x}_A) \mathbb{E}[f(\mathbb{X})|\mathbb{X}_A = \mathbf{x}_A] \mathbb{P}_A(d\mathbf{x}_A) \\
&= \langle h, \mathbb{E}[f(\mathbb{X})|\mathbb{X}_A = \mathbf{x}_A]\rangle.
\end{aligned}$$

Here, the last equality comes from Fubini Theorem as $h(\mathbf{x}_A) f(\mathbf{x})$ is integrable. This implies (1.6) by using (1.5). Let now $h \in L^2_{A\cap B}$, as $L^2_{A\cap B} \subset L^2_A$ and $L^2_{A\cap B} \subset L^2_B$, we have, by (1.5),

$$\langle h, \mathrm{Proj}_B \mathrm{Proj}_A f\rangle = \langle h, \mathrm{Proj}_A f\rangle = \langle h, f\rangle.$$

So that, we may conclude that (1.7) holds as $A \cap B = B \cap A$ and using (1.5). □

Proof of Theorem 1.3 using the projection operators.

Existence To begin with, let us first prove that there exists a decomposition of G in the form of (1.3) such that *1.* and *2.* hold. Let $j \in \{1,\dots,d\}$. To alleviate the notation, we will write Proj_{-j} for $\mathrm{Proj}_{\overline{\{j\}}}$ where $\overline{\{j\}}$ is the complementary of $\{j\}$ in $\{1,\dots,d\}$. Notice that from (1.6) $(\mathrm{Proj}_{-j} G)(\mathbf{x})$ does not depend any more on x_j (a.s.). Moreover, from (1.7) for any $i, j \in \{1,\dots,d\}$, Proj_{-i} and Proj_{-j} are commuting. We may write

$$I = \prod_{j=1}^{d} \Big(\mathrm{Proj}_{-j} + (I - \mathrm{Proj}_{-j}) \Big).$$

Taking into account the commutation in expanding the last product leads to

$$I = \sum_{A\subset\{1,\dots d\}} \left(\prod_{j\in A}(I-\mathrm{Proj}_{-j})\right)\left(\prod_{j\in\overline{A}}\mathrm{Proj}_{-j}\right), \tag{1.8}$$

Considering the complementary set, one may easily check that $\bigcap_{j\in\overline{A}}\overline{\{j\}} = A$. So that, using (1.7), we have that $\left(\prod_{j\in\overline{A}}\mathrm{Proj}_{-j}\right) = \mathrm{Proj}_A$. Let us now define, for any $A\subset\{1,\dots d\}$, $G_A = \left(\prod_{j\in A}(I-\mathrm{Proj}_{-j})\right)\mathrm{Proj}_A G$. It follows directly from (1.8) that, a.s.,

$$G(\mathbf{x}) = \sum_{A\subset\{1,\dots d\}} G_A(\mathbf{x}).$$

By definition, we get $G_\emptyset = \mathrm{Proj}_\emptyset G = \int_E G(\mathbf{x})\mathbb{P}(d\mathbf{x})$ by (1.6), thus $G_\emptyset$ is a.s. constant. Moreover, by (1.7), we have a.s.

$$G_A(\mathbf{x}) = \left(\prod_{j\in A}(I-\mathrm{Proj}_{-j})\right)\mathrm{Proj}_A G(\mathbf{x}) = \mathrm{Proj}_A\left(\prod_{j\in A}(I-\mathrm{Proj}_{-j})\right)G(\mathbf{x}).$$

So that, $G_A(\mathbf{x})$ does not depend on $\mathbf{x}_{\overline{A}}$ and we may write $G_A(\mathbf{x}) = G_A(\mathbf{x}_A)$. Now, for $i\in A$ and $f\in L^2$, $\mathrm{Proj}_{-i}(I-\mathrm{Proj}_{-i})f = 0$. So that,

$$\int_{E_i} G_A(\mathbf{x}_A)\mathbb{P}_{X_i}(dx_i) = \mathrm{Proj}_{-i}G_A(\mathbf{x}) = \mathrm{Proj}_{-i}(I-\mathrm{Proj}_{-i})\prod_{j\in A, j\neq i}(I-\mathrm{Proj}_{-j})\mathrm{Proj}_A G(\mathbf{x}) = 0\ (a.s.).$$

Thus, we have proved that there exists a decomposition of G in the form (1.3) satisfying the properties *1.* and *2.*

Uniqueness Now, considering a decomposition of G in the form of (1.3)

$$G(\mathbf{x}) = \sum_{A\subset\{1,\dots d\}} G_A(\mathbf{x}_A),\quad a.s.$$

satisfying the properties *1.* and *2.*, we will prove that it also satisfies (1.4), leading to a.s. uniqueness. First, note that $\mathbb{E}[G(\mathbb{X})] = \int_E G(\mathbf{x})\mathbb{P}_{\mathbb{X}}(d\mathbf{x}) = G_\emptyset$. Indeed, as $G(\mathbf{x}) = G_\emptyset + \sum_{A\subset\{1,\dots d\},A\neq\emptyset} G_A(\mathbf{x}_A)$ and $\forall\, A\subset\{1,\dots d\}$, $A\neq\emptyset$, for any $i\in A$,

$$\int_E G_A(\mathbf{x}_A)\mathbb{P}_{\mathbb{X}}(d\mathbf{x}) = \int_{\prod_{j\neq i}E_j}\underbrace{\int_{E_i} G_A(\mathbf{x}_A)\mathbb{P}_{X_i}(dx_i)}_{=0}\prod_{j\neq i}\mathbb{P}_{X_j}(dx_j) = 0.$$

Now, let $i\in\{1,\dots,d\}$. We have $\int_{\prod_{j\neq i}E_j} G(\mathbf{x})\prod_{j\neq i}\mathbb{P}_{X_j}(dx_j) = G_\emptyset + G_{\{i\}}(X_i)$ since, by property 2.,

$$\sum_{A\neq i, A\neq\emptyset}\underbrace{\int_{\prod_{j\neq i}E_j} G_A(\mathbf{x}_A)\prod_{j\neq i}\mathbb{P}_{X_j}(dx_j)}_{=0} = 0.$$

Thus, by (1.6) for any $i\in\{1,\dots,d\}$,

$$G_{\{i\}}(x_i) = \int_{\prod_{j\neq i}E_j} G(\mathbf{x})\prod_{j\neq i}\mathbb{P}_{X_j}(dx_j) - G_\emptyset = \mathbb{E}[G(\mathbb{X})|X_i = x_i] - \mathbb{E}\,G(\mathbb{X}).$$

More generally, let $A\subset\{1,\dots d\}$. We can write

$$G(\mathbf{x}) = \sum_{B\subset\{1,\dots d\},\bar{A}\cap B\neq\emptyset} G_B(\mathbf{x}_B) + \sum_{B\subset A} G_B(\mathbf{x}_B)$$

Let $B\subset\{1,\dots d\}$, $\bar{A}\cap B\neq\emptyset$. For any $i\in\bar{A}\cap B$, we have using one more time property *2.*

$$\mathbb{E}[G_B(\mathbb{X}_B)|\mathbb{X}_A = \mathbf{x}_A] = \int_{\prod_{j\notin A}E_j} G_B(\mathbf{x}_B)\prod_{j\notin A}\mathbb{P}_{X_j}(dx_j) = \int_{\prod_{j\notin A, j\neq i}E_j}\underbrace{\int_{E_i} G_B(\mathbf{x}_B)\mathbb{P}_{X_i}(dx_i)}_{=0}\prod_{j\notin A, j\neq i}\mathbb{P}_{X_j}(dx_j) = 0.$$

Using the assumed expansion of G and the last equation, we thus have, for $\mathbb{P}_{\mathbb{X}}$-a.s. all $\mathbf{x}\in E$,

$$\mathbb{E}[G(\mathbb{X})|\mathbb{X}_A = \mathbf{x}_A] = \sum_{B\subset A}\mathbb{E}[G_B(\mathbb{X}_B)|\mathbb{X}_A = \mathbf{x}_A] = \sum_{B\subset A} G_B(\mathbf{x}_B). \tag{1.9}$$

Defining for $\mathbb{P}_{\mathbb{X}}$-a.s. all $\mathbf{x}\in E$, $g(\mathbf{x},A) = \mathbb{E}[G(\mathbb{X})|\mathbb{X}_A = \mathbf{x}_A]$ and $f(\mathbf{x},B) = G_B(\mathbf{x}_B)$, we get $g(\mathbf{x},A) = \sum_{B\subset A} f(\mathbf{x},B)$. Applying the Rota inversion formula [101], we get

$$f(\mathbf{x},A) = \sum_{B\subset A}(-1)^{|A|-|B|}g(\mathbf{x},B)\qquad\text{namely}\qquad G_A(\mathbf{x}_A) = \sum_{B\subset A}(-1)^{|A|-|B|}\mathbb{E}[G(\mathbb{X})|\mathbb{X}_B = \mathbf{x}_B].$$

□

Remark 1.5. *If we have a look to the proof of Theorem 1.3, we can see that, for any subset A of $\{1,\dots,d\}$, one has by induction:*

$$G_A(\mathbb{X}_A) = \mathbb{E}[G(\mathbb{X})|\mathbb{X}_A] - \sum_{B \subsetneq A} G_B(\mathbb{X}_B).$$

Example 1.6. *Let $\mathbb{X} = (X_1, X_2, X_3)$ be a vector of three independent random variables and G a square integrable random function of X_1, X_2, X_3. Then the Hoeffding decomposition is*

$$G(\mathbb{X}) = G(X_1, X_2, X_3) = G_\emptyset + G_1 + G_2 + G_3 + G_{1,2} + G_{1,3} + G_{2,3} + G_{1,2,3}$$

with

$$\begin{aligned}
G_\emptyset &= \mathbb{E}[G(\mathbb{X})] =: m \\
G_{\{1\}} &= \mathbb{E}[G(\mathbb{X})|X_1] - m \\
G_{\{2\}} &= \mathbb{E}[G(\mathbb{X})|X_2] - m \\
G_{\{3\}} &= \mathbb{E}[G(\mathbb{X})|X_3] - m \\
G_{\{1,2\}} &= \mathbb{E}\left[G(\mathbb{X})|\mathbb{X}_{\{1,2\}}\right] - G_{\{1\}} - G_{\{2\}} - m \\
G_{\{1,3\}} &= \mathbb{E}\left[G(\mathbb{X})|\mathbb{X}_{\{1,3\}}\right] - G_{\{1\}} - G_{\{3\}} - m \\
G_{\{2,3\}} &= \mathbb{E}\left[G(\mathbb{X})|\mathbb{X}_{\{2,3\}}\right] - G_{\{2\}} - G_{\{3\}} - m \\
G_{\{1,2,3\}} &= G(\mathbb{X}) - G_{\{1,2\}} - G_{\{1,3\}} - G_{\{2,3\}} - G_{\{1\}} - G_{\{2\}} - G_{\{3\}} - m.
\end{aligned}$$

Corollary 1.7. *Let $G \in L^2(\mathbb{P}_\mathbb{X})$. Then, under the assumptions of Theorem 1.3, the two following properties hold.*

1. *Let $A \in \{1,\dots,d\}$ such that $A \neq \emptyset$. Then, $\mathbb{E}[G_A(\mathbb{X}_A)] = 0$.*
2. *Let $A \neq B \in \{1,\dots,d\}$. Then G_A and G_B are orthogonal in $\mathbb{L}^2(\mathbb{P}_\mathbb{X})$ i.e. $\langle G_A, G_B\rangle = \mathbb{E}[G_A(\mathbb{X}_A)G_B(\mathbb{X}_B)] = 0$*

Proof of Corollary 1.7. 1. Let $A \in \{1,\dots,d\}$ such that $A \neq \emptyset$. Let $i \in A$. Using the point 2. of the previous theorem, we have

$$\mathbb{E}[G_A(\mathbb{X}_A)] = \int_{\prod_{j\in A, j\neq i} E_j} \underbrace{\int_{E_i} G_A(\mathbf{x}_A)\mathbb{P}_{X_i}(dx_i)}_{0} \prod_{j\in A, j\neq i} \mathbb{P}_{X_j}(dx_j).$$

2. Let $A \neq B \in \{1,\dots,d\}$. As $A \neq B$, either there exists $i \in A \cap \bar{B}$ or there exists $i \in B \cap \bar{A}$. Assume without loss of generality that there exists $i \in B \cap \bar{A}$, as the other case can be handled similarly. We have

$$\mathbb{E}[G_A(\mathbb{X}_A)G_B(\mathbb{X}_B)] = \int_{\prod_{j\neq i} E_j} \underbrace{\int_{E_i} G_B(\mathbf{x}_B)\mathbb{P}_{X_i}(dx_i)}_{0} G_A(\mathbf{x}_A) \prod_{j\neq i} \mathbb{P}_{X_j}(dx_j) = 0. \qquad \square$$

The first term in the Hoeffding decomposition in (1.3) corresponds to the mean behavior of the model G. Then, any higher-order term in this decomposition can be interpreted as the contribution on the initial function G of a group of variables $\mathbb{X}_A$ when removing all the lower-order interaction contributions related to $\mathbb{X}_B, B \subset A$. For example, the contribution $G_{\{i\}}$ $(i = 1,\dots,d)$ provides the centered effect of the variable X_i alone while $G_{\{i,j\}}$ $(1 \leqslant i < j \leqslant d)$ describes the centered interaction effect of the variables X_i and X_j removing the single effects.

From Corollary 1.7, we know that the summand in the Hoeffding decomposition are centered (except for $A = \emptyset$) and uncorrelated. Thus the variance of the function $G(\mathbb{X})$ can be computed as the sum of the variances of the non constant terms of the decomposition. This is the foundation stone for second-order methods in SA. The following corollary formally states the variance decomposition.

Corollary 1.8. *Under the assumptions of Theorem 1.3 and if we set* $V_A = \mathrm{Var}\,(G_A(\mathbb{X}_A)) = \mathbb{E}[G_A(\mathbb{X}_A)^2]$, *then*

$$V = \mathrm{Var}\,(G(\mathbb{X})) = \sum_{A\subset\{1,\dots,d\},A\neq\emptyset} V_A.$$

Furthermore,

$$V_A = \sum_{B\subset A} (-1)^{|A|-|B|}\mathrm{Var}(\mathbb{E}[G(\mathbb{X})|\mathbb{X}_B]).$$

Proof of Corollary 1.7. The first part of the corollary is straightforward from Corollary 1.7. Indeed, as $G(\mathbb{X}) - G_\emptyset = \sum_{A\in\{1,\dots,d\},A\neq\emptyset} G_A(\mathbb{X}_A)$ and from the results in Corollary 1.7, we get

$$V = \mathbb{E}[G(\mathbb{X}) - G_\emptyset]^2 = \sum_{A\in\{1,\dots,d\},A\neq\emptyset} \mathbb{E}\,G_A^2(\mathbb{X}_A) + \sum_{\emptyset\subsetneq A\neq B\in\{1,\dots,d\}} \mathbb{E}[G_A(\mathbb{X}_A)G_B(\mathbb{X}_B)] = \sum_{A\in\{1,\dots,d\},A\neq\emptyset} \mathrm{Var}(G_A(\mathbb{X}_A)) = \sum_{A\in\{1,\dots,d\},A\neq\emptyset} V_A.$$

Now, let $A\in\{1,\dots,d\}$, $A\neq\emptyset$. From (1.9) in the proof of Theorem 1.3, we have $\mathbb{P}_{\mathbb{X}_A}$-p.s. for all $\mathbf{x}_A$, $\mathbb{E}[G(\mathbb{X})|\mathbb{X}_A = \mathbf{x}_A] = \sum_{B\subset A} G_B(\mathbf{x}_B)$. Then, from Corollary 1.7, we get

$$\mathrm{Var}(\mathbb{E}[G(\mathbb{X})|\mathbb{X}_A]) = \mathrm{Var}\left(\sum_{B\subset A} G_B(\mathbb{X}_B)\right) = \sum_{B\subset A}\mathrm{Var}(G_B(\mathbb{X}_B)).$$

Once more, applying the Rota inversion formula [101], we get

$$\mathrm{Var}(G_A(\mathbb{X}_A)) = \sum_{B\subset A} (-1)^{|A|-|B|}\mathrm{Var}(\mathbb{E}[G(\mathbb{X})|\mathbb{X}_B]).$$ □

Remark 1.9. *It is a well known fact that for* L^2 *random variables the conditional expectation of* $\mathbb{E}[Z|W]$ *is a* W*- measurable random variable that is the best approximation in the* L^2 *sense of* Z *by a* W*-measurable random variable. Hence* G_A *is the best approximation of the function* G *in* L^2_A*. So* V_A *can be seen as the quantification of the sensitivity of* G *with respect to the inputs* $\mathbb{X}_A$*. Now the quantity* $V_A/\mathrm{Var}\,(G(\mathbb{X}))$ *would be the key quantity for the study of sensitivity analysis for* L^2 *random variables. In this lecture, we will restrict our study to these quantities.*

1.3 Sobol' indices

Definition 1.10. *Let* $A\subset\{1,\dots,d\}$, $\mathbb{X} = (X_1,\dots,X_d)$ *where the* X_i*'s are independent random variables, and* G *be a square integrable function of* $\mathbb{X}$*. We define the following indices.*

1. ***The Sobol index associated to*** A *is*
$$S^A = \frac{\mathrm{Var}\,(G_A(\mathbb{X}_A))}{\mathrm{Var}\,(G(\mathbb{X}))} = \frac{V_A}{V}.$$
2. ***The first order Sobol index associated to the input*** X_j *is*
$$S^j = S^{\{j\}} = \frac{V_{\{j\}}}{V}.$$
3. ***The total Sobol index associated to*** A *is*
$$S^{A,Tot} = 1 - S^{\overline{A}}.$$
In particular,
$$S^{j,Tot} = 1 - S^{\overline{j}} = 1 - \frac{V_{\overline{\{j\}}}}{V}. = 1 - \frac{V_{\overline{A}}}{V}$$
4. ***The closed Sobol index associated to*** A *is*
$$S^{A,Cl} = \sum_{B\subset A} S^B = \frac{\mathrm{Var}(\mathbb{E}[G(\mathbb{X})|\mathbb{X}_A])}{V}.$$

The Sobol' index is equivalent to the square of the correlation ratio (measure of the relationship between the statistical dispersion within individual categories and the dispersion across a whole population or a sample), well known in statistics and introduced by Pearson in [96] as part of analysis of variance. In the sensitivity analysis domain, it has

been introduced by Sobol' in [109] in a general context but has appeared before in some applicative papers for the case where A is a singleton (e.g. [31]). The Sobol' index has often been called the variance-based importance measure.

Example 1.11. *Let X_1, X_2, X_3 be three independent random variables $\mathcal{N}(0,1)$ distributed and let a_1, a_2, a_3, a_4 be four real numbers. Consider the following function*

$$G(\mathbb{X}) = G(X_1, X_2, X_3) = a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_1 X_2.$$

1. *Assume that $a_3 = a_4 = 0$. Then we have*

$$\mathbb{E}[G(\mathbb{X})|X_1] = a_1 X_1, \quad \mathbb{E}[G(\mathbb{X})|X_2] = a_2 X_2, \quad \text{and} \quad \mathbb{E}[G(\mathbb{X})|X_3] = 0,$$

and also

$$\mathbb{E}[G(\mathbb{X})|X_1, X_2] = a_1 X_1 + a_2 X_2, \quad \mathbb{E}[G(\mathbb{X})|X_1, X_3] = a_1 X_1, \quad \text{and} \quad \mathbb{E}[G(\mathbb{X})|X_2, X_3] = a_2 X_2.$$

Then the Sobol' indices are given by

$$S^1 = \frac{a_1^2}{a_1^2 + a_2^2}, \quad S^2 = \frac{a_2^2}{a_1^2 + a_2^2}, \quad S^3 = 0, \quad S^{1,2} = 0, \quad S^{1,3} = 0, \quad S^{2,3} = 0, \quad S^{1,2,3} = 0.$$

The closed Sobol' index for $\{1,2\}$ is $S^{\{1,2\},Cl} = S^{1,2} + S^1 + S^2 = 1$.

2. *General case $a_3 \neq 0$ and $a_4 \neq 0$. One has*

$$S^1 = \frac{a_1^2}{a_1^2 + a_2^2 + a_3^2 + a_4^2}, \quad S^2 = \frac{a_2^2}{a_1^2 + a_2^2 + a_3^2 + a_4^2}, \quad S^3 = \frac{a_3^2}{a_1^2 + a_2^2 + a_3^2 + a_4^2},$$
$$S^{1,2} = \frac{a_4^2}{a_1^2 + a_2^2 + a_3^2 + a_4^2}, \quad S^{1,3} = 0, \quad S^{2,3} = 0,$$
$$S^{1,2,3} = 0.$$

The closed Sobol' index for $\{1,2\}$ is $S^{\{1,2\},Cl} = S^1 + S^2 + S^{1,2} = \frac{a_1^2 + a_2^2 + a_4^2}{a_1^2 + a_2^2 + a_3^2 + a_4^2} \neq 1$.

To conclude this chapter, let us give some obvious properties of the Sobol' indices.

1. If the function G does not depend on the random variable X_i, then $S^A = 0$ for any A such that $i \in A$.
2. If $S^{A,\mathrm{Tot}} = 1$, then G only depends on the random variables, the indices of which are in A.
3. $S^i = S^{i,\mathrm{Cl}}$ quantifies the part of the variability that is due to the action of variable X_i alone. We speak of the first order importance of X_i. $S^{i,j}$ quantifies the part of the variability that is due to the interaction between the variables X_i and X_j when the first orders have been removed. To understand better this phenomenon, take $a_1 = a_2 = a_3 = 0$ in the previous example: $G(\mathbb{X}) = a_4 X_1 X_2$. Then $S^1 = S^2 = 0$, $S^{1,2} = 1$ meaning that alone X_1 and X_2 have no influence on the variability of the output, but together they are responsible of all the variability.
4. It is clear from Corollary 1.8 that
$$1 = \sum_{A \in \{1,\dots,d\}} S^A \quad \text{and} \quad 0 \leqslant S^A \leqslant 1.$$

In general, it is not possible to compute explicitly the Sobol' indices. Indeed, in most applications, G is unknown or very complicated, whence it is then impossible to perform analytic computations. Therefore the statistician wants to give an estimation of these indices. This is the topic of Chapters 2 and 4.

1.4 Exercises

Exercise 1 (Ishigami function)**.** *The Ishigami model is given by:*

$$Y = f(\mathbb{X}) = f(X_1, X_2, X_3) = \sin X_1 + 7\sin^2 X_2 + 0.1 X_3^4 \sin X_1 \tag{1.10}$$

where $(X_j)_{j=1,2,3}$ *are i.i.d. uniform random variables on* $[-\pi;\pi]$.

1. *Show that*
$$S^1 = 0.3139,\ S^2 = 0.4424,\ S^3 = 0.$$
Comment.
2. *Show that*
$$S^{\{1,2\},Cl} = 0.7563,\ S^{\{1,3\},Cl} = 0.5575,\ S^{\{1,3\},Cl} = 0.4424,\ S^{\{1,2,3\},Cl} = 1.$$

Exercise 2 (Sobol g-function)**.** *Assume that* $X_1, \ldots, X_d$ *are i.i.d. random variables uniformly distributed on* $[0,1]$. *Now take* d *real numbers* $a_1, \ldots, a_d$ *and define the Sobol g-function by*

$$Y = g_{sobol}(\mathbb{X}) = g_{sobol}(X_1, \ldots, X_d) = \prod_{k=1}^{d} g_k(X_k) = \prod_{k=1}^{d} \frac{|4X_k - 2| + a_k}{1 + a_k}. \tag{1.11}$$

Compute $S^{i,Cl}$ *for* $i \in \{1, \ldots, d\}$.

Chapter 2

First estimation of Sobol' indices: Pick-Freeze scheme

2.1 Pick-Freeze estimation of the Sobol' indices

As previously, we consider a (non necessarily linear) regression model connecting an output $Y \in \mathbb{R}$ to independent a random input $\mathbb{X} = (X_1, \dots X_d)$ with, for $i = 1, \dots d$, X_i belongs to the probability space $\mathscr{E}_i$. We denote

$$Y = f(\mathbb{X}) = f(X_1, \dots, X_d) \tag{2.1}$$

where f is a deterministic real-valued measurable function defined on $\mathscr{E} = \mathscr{E}_1 \times \dots \mathscr{E}_d$. We assume that Y is square integrable and non deterministic ($\mathrm{Var}(Y) \neq 0$).

Since the knowledge of all closed indices allows to recover all the other indices, we focus on the estimation of the closed indices that have in addition a simple expression: for any subset A of $I_d = \{1, \dots, d\}$,

$$S^{\mathbf{u},\mathrm{Cl}} = \frac{\mathrm{Var}\left[\mathbb{E}[Y|\mathbb{X}_{\mathbf{u}}]\right]}{\mathrm{Var}\left[Y\right]}. \tag{2.2}$$

In practice, it is often important to be able to estimate simultaneously several indices. For this purpose, let $\mathbf{u} = (\mathbf{u_1}, \dots, \mathbf{u_k})$ be k subsets of I_d. The vector of the closed Sobol' indices is then given by

$$S^{\mathbf{u},\mathrm{Cl}} = \left(\frac{\mathrm{Var}(\mathbb{E}[Y|\mathbb{X}_{\mathbf{u_1}}])}{\mathrm{Var}(Y)}, \dots, \frac{\mathrm{Var}(\mathbb{E}[Y|\mathbb{X}_{\mathbf{u_k}}])}{\mathrm{Var}(Y)}\right) = \left(\frac{\mathrm{Var}(\mathbb{E}[Y|X_i, i \in \mathbf{u_1}])}{\mathrm{Var}(Y)}, \dots, \frac{\mathrm{Var}(\mathbb{E}[Y|X_i, i \in \mathbf{u_k}])}{\mathrm{Var}(Y)}\right).$$

Example 2.1. *Assume $d = 5$, $k = 3$, and take $\mathbf{u} = (\{1\}, \{1,3,5\}, \{2,4\})$ in that case*

$$S^{\mathbf{u},Cl} = \left(\frac{\mathrm{Var}(\mathbb{E}[Y|X_1]}{\mathrm{Var}(Y)}, \frac{\mathrm{Var}(\mathbb{E}[Y|X_1, X_3, X_5])}{\mathrm{Var}(Y)}, \frac{\mathrm{Var}(\mathbb{E}[Y|X_2, X_4])}{\mathrm{Var}(Y)}\right).$$

It is easy to estimate $\mathrm{Var}(Y)$ by the empirical variance; the problem here is to estimate quantities like $\mathrm{Var}(\mathbb{E}[Y|\mathbb{X}_{\mathbf{u_j}}])$. Indeed in general, the estimation of conditional expectation is not an easy task.
In the sequel, we present a very nice trick allowing to transform the variance of the conditional expectation into a covariance. To do so, we need some additional definitions.

Definition 2.2. *For the input $\mathbb{X}$ and for any subset $\mathbf{u}$ of I_d, we define $\mathbb{X}^{\mathbf{u}}$ by the vector such that $X_i^{\mathbf{u}} = X_i$ if $i \in \mathbf{u}$ and $X_i^{\mathbf{u}} = X_i'$ if $i \notin \mathbf{u}$ where X_i' is an independent copy of X_i. Then we set*

$$Y^{\mathbf{u}} = f(\mathbb{X}^{\mathbf{u}}).$$

$\mathbb{X}^{\mathbf{u}}$ and $Y^{\mathbf{u}}$ are named the Pick-Freeze versions of $\mathbb{X}$ and Y respectively. The idea to keep the variable if the index is in $\mathbf{u}$ and to resample a new one if the index is not in $\mathbf{u}$.

Example 2.3. *Assume $d = 2$ and $Y = f(X_1, X_2)$ and assume $\mathbf{v} = \{1\}$. Then $\mathbb{X} = (X_1, X_2)$ and $\mathbb{X}^{\mathbf{v}} = (X_1, X_2')$ where X_2' is an independent copy of X_2 (X_2' is also independent of X_1). Finally,*

$$Y = (X_1, X_2) \text{ and } Y^{\mathbf{v}} = f(X_1, X_2').$$

The next lemma shows how to express the numerator of $S^{\mathbf{u},\mathrm{Cl}}$ in terms of a covariance. This will lead to a natural estimator.

Lemma 2.4. *For any $u \subset I_d$, one has*

$$\mathrm{Var}(\mathbb{E}[Y|\mathbb{X}_{\mathbf{u}}]) = \mathrm{Cov}\left(Y, Y^{\mathbf{u}}\right). \tag{2.3}$$

Proof of Lemma 2.4. It is clear that Y and $Y^{\mathbf{u}}$ have the same law. In addition, we can assume without loss of generality that $\mathbb{E}[Y] = 0$. Now conditioning on the variables X_i, for $i \in \mathbf{u}$, Y and $Y^{\mathbf{u}}$ are independent so

$$\mathrm{Cov}\left(Y, Y^{\mathbf{u}}\right) = \mathbb{E}[Y\,Y^{\mathbf{u}}] = \mathbb{E}[\mathbb{E}[Y\,Y^{\mathbf{u}}|\mathbb{X}_{\mathbf{u}}]] = \mathbb{E}[\mathbb{E}[Y|\mathbb{X}_{\mathbf{u}}]\mathbb{E}[Y^{\mathbf{u}}|\mathbb{X}_{\mathbf{u}}]] = \mathbb{E}[\mathbb{E}[Y|\mathbb{X}_{\mathbf{u}}]^2] = \mathrm{Var}(\mathbb{E}[Y|\mathbb{X}_{\mathbf{u}}]). \qquad \square$$

A first estimation for $S^{\mathbf{u},\mathbf{Cl}}$

In view of Lemma 2.4, we are now able to define a first natural estimator of $S^{\mathbf{u},\mathrm{Cl}}$ based on a N-sample $(\mathbb{X}_1, \ldots, \mathbb{X}_N)$ of $\mathbb{X}$ leading to a N-sample $(Y_1, \ldots, Y_N)$ of Y. The estimation scheme also requires the Pick-Freeze versions of $\mathbb{X}_j$ and Y_j for all the subsets $\mathbf{u_1}, \ldots \mathbf{u_k}$. Thus the observations consist in $(Y_j, Y_j^{u_1}, \ldots, Y_j^{u_k})_{(1 \leqslant j \leqslant N)}$ based on $(\mathbb{X}_j, \mathbb{X}_j^{u_1}, \ldots, \mathbb{X}_j^{u_k})_{(1 \leqslant j \leqslant N)}$. All sums are taken for j from 1 to N.

$$S_N^{\mathbf{u},\mathrm{Cl}} = \left(\frac{\frac{1}{N}\sum Y_j Y_j^{\mathbf{u_1}} - \left(\frac{1}{N}\sum Y_j\right)\left(\frac{1}{N}\sum Y_j^{\mathbf{u_1}}\right)}{\frac{1}{N}\sum Y_j^2 - \left(\frac{1}{N}\sum Y_j\right)^2}, \ldots, \frac{\frac{1}{N}\sum Y_j Y_j^{\mathbf{u_k}} - \left(\frac{1}{N}\sum Y_j\right)\left(\frac{1}{N}\sum Y_j^{\mathbf{u_k}}\right)}{\frac{1}{N}\sum Y_j^2 - \left(\frac{1}{N}\sum Y_j\right)^2} \right). \tag{2.4}$$

A second estimation for $S^{\mathbf{u},\mathbf{Cl}}$

Since the observations consist in $(Y_j, Y_j^{u_1}, \ldots, Y_j^{u_k})_{(1 \leqslant j \leqslant N)}$, a more precise estimation of the first and second moments can be done and we are able to define a second estimator of $S^{\mathbf{u},\mathrm{Cl}}$ taking into account all the available information. Define

$$Z_j^{\mathbf{u}} = \frac{1}{k+1}\left(Y_j + \sum_{\ell=1}^{k} Y_j^{\mathbf{u}_\ell}\right), \quad M_j^{\mathbf{u}} = \frac{1}{k+1}\left(Y_j^2 + \sum_{\ell=1}^{k} (Y_j^{\mathbf{u}_\ell})^2\right).$$

The second estimator is then given by

$$T_N^{\mathbf{u},\mathrm{Cl}} = \left(\frac{\frac{1}{N}\sum Y_j Y_j^{\mathbf{u_1}} - \left(\frac{1}{2N}\sum (Y_j + Y_j^{\mathbf{u_1}})\right)^2}{\frac{1}{N}\sum M_j^{\mathbf{u}} - \left(\frac{1}{N}\sum Z_j^{\mathbf{u}}\right)^2}, \ldots, \frac{\frac{1}{N}\sum Y_j Y_j^{\mathbf{u_k}} - \left(\frac{1}{2N}\sum (Y_j + Y_j^{\mathbf{u_k}})\right)^2}{\frac{1}{N}\sum M_j^{\mathbf{u}} - \left(\frac{1}{N}\sum Z_j^{\mathbf{u}}\right)^2} \right). \tag{2.5}$$

Remark 2.5. *Let us just explain why the second estimator is going to be a little better. In $S_N^{\mathbf{u},Cl}$ in order to estimate the expected value $\mathbb{E}[Y]$, we only use one of the sample we have that is we compute $\frac{1}{N}\sum Y_j$. Nevertheless, since we have a $2N$ sample, it seems reasonable to use all the information we have and consider rather $\frac{1}{2N}\sum (Y_j + Y_j^{\mathbf{u_1}})$. We see that in the second case the variance of the estimator of the mean is reduced by a factor 2.*

Here, we showed how to construct two estimators $S_N^{\mathbf{u},\mathrm{Cl}}$ and $T_N^{\mathbf{u},\mathrm{Cl}}$ of the Sobol' indices. In the sequel, we focus our study on $S_N^{\mathbf{u},\mathrm{Cl}}$; it is easy following the same road map to perform the same study for $T_N^{\mathbf{u},\mathrm{Cl}}$. Two natural statistical questions arises.

1. Are they consistent? That means do we have convergence almost sure of $S_N^{\mathbf{u},\mathrm{Cl}}$ and $T_N^{\mathbf{u},\mathrm{Cl}}$?
2. If yes, do we have a central limit theorem?

We recall in the next section the definition and properties of different stochastic convergences. Furthermore, the method developed to answer the second question is based on the so-called Delta-method, which is also recalled in the next section.

2.2 Preliminary technical results

Here we recall some basic facts about stochastic convergences together with a well known result allowing to transfer a Central Limit Theorem via differentiable functions.

2.2.1 Some basic facts about stochastic convergences

The results of this paragraph are well known results concerning stochastic convergences. The proofs can be found in the book written by Ad van der Vaart [122].

Theorem 2.6. *[Continuous mapping] Let $g: \mathbb{R}^k \mapsto \mathbb{R}^m$ be continuous at every point of a $C \subset C$ such that $\mathbb{P}(X \in C) = 1$.*
(i) If $X_n \xrightarrow[n]{\mathcal{L}} X$, then $g(X_n) \xrightarrow[n]{\mathcal{L}} g(X)$;
(ii) If $X_n \xrightarrow[n]{Pr} X$, then $g(X_n) \xrightarrow[n]{Pr} g(X)$;
(iii) If $X_n \xrightarrow[n]{a.s.} X$, then $g(X_n) \xrightarrow[n]{a.s.} g(X)$.

Theorem 2.7. *[Prohorov's theorem] Let X_n be random vectors in $\mathbb{R}^k$.*
(i) If $X_n \xrightarrow[n]{\mathcal{L}} X$ for some X, then $(X_n)_n$ is uniformly tight.
(ii) If X_n is uniformly tight, then there exists a subsequence with $X_{n_j} \xrightarrow[n]{\mathcal{L}} X$ as $j \to \infty$, for some X.

Theorem 2.8. *Let $(X_n)_n$, $(Y_n)_n$ and X, Y be some random vectors and c be a constant .*
i) If $X_n \xrightarrow[n]{a.s.} X$ then $X_n \xrightarrow[n]{Pr} X$.
ii) If $X_n \xrightarrow[n]{Pr} X$ then $X_n \xrightarrow[n]{\mathcal{L}} X$.
iii) $X_n \xrightarrow[n]{Pr} c$ if and only if $X_n \xrightarrow[n]{\mathcal{L}} c$.
iv) If $X_n \xrightarrow[n]{\mathcal{L}} X$ and $d(X_n, Y_n) \xrightarrow[n]{Pr} 0$ then $Y_n \xrightarrow[n]{\mathcal{L}} X$.
v) (Slutsky) If $X_n \xrightarrow[n]{\mathcal{L}} X$ and $Y_n \xrightarrow[n]{Pr} c$ then $(X_n, Y_n) \xrightarrow[n]{\mathcal{L}} (X, c)$.
vi) If $X_n \xrightarrow[n]{Pr} X$ and $Y_n \xrightarrow[n]{Pr} Y$ then $(X_n, Y_n) \xrightarrow[n]{Pr} (X, Y)$.

Lemma 2.9. *[Slutsky] Let X_n, X and Y_n be random vectors or variables. If $X_n \xrightarrow[n]{\mathcal{L}} X$ and $Y_n \xrightarrow[n]{\mathcal{L}} c$ for a constant c, then*
(i) $X_n + Y_n \xrightarrow[n]{\mathcal{L}} X + c$;
(ii) $Y_n X_n \xrightarrow[n]{\mathcal{L}} cX$;
(iii) $Y_n^{-l} X_n \xrightarrow[n]{\mathcal{L}} c^{-1} X$ provided $c \neq 0$.

We introduce here some useful notations

- $X_n = o_P(1)$ means that X_n converges to 0 in probability and $X_n = o_P(R_n)$ means that $X_n = Y_n R_n$ where Y_n converges to 0 in probability.
- $X_n = O_P(1)$ means that the family $(X_n)_n$ is uniformly tight and $X_n = O_P(R_n)$ means that $X_n = Y_n R_n$ where the family $(Y_n)_n$ is uniformly tight.

Lemma 2.10. *Let X_n be a sequence of random vectors going to zero in probability. Then for any $p > 0$, and any function R such that $R(0) = 0$,*

1. $R(h) = o(\|h\|^p) \implies R(X_n) = o_P(\|X_n\|^p)$.
2. $R(h) = O(\|h\|^p) \implies R(X_n) = O_P(\|X_n\|^p)$.

Theorem 2.11 (Classical CLT). *Let $(Z_j)_{i\in\mathbb{N}^*}$ be i.i.d. random variables having the same distribution as a random variable Z such that $\mathbb{E}[Z^2]<\infty$. Let $m=\mathbb{E}[Z]$, $\sigma^2=\mathrm{Var}(Z)$, and $\overline{Z}_n=\frac{1}{n}\sum_{i=1}^n Z_i$. Then*

$$\sqrt{n}\left(\overline{Z}_n-m\right)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,\sigma^2).$$

Remark 2.12. *If the random variables belong to $\mathbb{R}^k$ and have the same distribution as $Z=(Z_1,\dots,Z_k)$, the result is analogous and the limit distribution is the centered Gaussian vector with covariance matrix Σ defined for $1\leqslant i\leqslant k$ and $1\leqslant j\leqslant k$ by $\Sigma_{i,j}=\mathrm{Cov}(Z_i,Z_j)$.*

2.2.2 The Delta method

Assume that we are interested in some transformation of an unknown parameter θ; let's say $\phi(\theta)$. In addition, assume that we know for some reason[1] that $\sqrt{n}(T_n-\theta)\xrightarrow[n\to\infty]{\mathcal{L}}X$. The natural question would then be: do we still have something like $\sqrt{n}\big(\phi(T_n)-\phi(\theta)\big)\xrightarrow[n\to\infty]{\mathcal{L}}$?

The answer is obviously yes if ϕ is linear since the continuous mapping theorem insures that

$$\phi\left(\sqrt{n}(T_n-\theta)\right)\xrightarrow[n\to\infty]{\mathcal{L}}\phi(X)$$

and then by linearity

$$\sqrt{n}\big(\phi(T_n)-\phi(\theta)\big)\xrightarrow[n\to\infty]{\mathcal{L}}\phi(X).$$

The answer is not obvious in the general case. Nevertheless it seems reasonable to think that if ϕ is differentiable, ϕ behaves locally as a linear mapping and the result should be true.

Theorem 2.13 (Delta method). *Let ϕ be an application from $\mathbb{R}^k$ to $\mathbb{R}^m$ differentiable at the point θ. Let (T_n) be a sequence of random vectors in $\mathbb{R}^k$ and $(r_n)_n$ be a sequence of real numbers going to ∞. Then the difference $r_n\big(\phi(T_n)-\phi(\theta)\big)-D\phi(\theta)\left(r_n(T_n-\theta)\right)$ converges to zero in probability. Moreover,*

$$r_n\big(\phi(T_n)-\phi(\theta)\big)\xrightarrow[n\to\infty]{\mathcal{L}}D\phi(\theta)(T);$$

as soon as $r_n(T_n-\theta)\xrightarrow[n\to\infty]{\mathcal{L}}T$.

Proof of Theorem 2.13. Consider $R(h)=\phi(\theta+h)-\phi(\theta)-D\phi(\theta)(h)$. Since ϕ is differentiable, we know that $R(h)=o(\|h\|)$ as $h\to 0$. Now Slutsky's lemma (Lemma 2.9) shows that $T_n-\theta\xrightarrow[n\to\infty]{\mathbb{P}}0$. Applying Lemma 2.10,

$$\phi(T_n)-\phi(\theta)-D\phi(\theta)(T_n-\theta)=R(T_n-\theta)=o_P(\|T_n-\theta\|).$$

Multiplying both sides by r_n, one gets

$$r_n\phi(T_n)-r_n\phi(\theta)-r_nD\phi(\theta)(T_n-\theta)=r_no_P(\|T_n-\theta\|).$$

$r_no_P(\|T_n-\theta\|)=o_P(r_n\|T_n-\theta\|)$. In addition, using Prohorov's theorem (Theorem 2.7), since $r_n(T_n-\theta)\xrightarrow[n\to\infty]{\mathcal{L}}T$, the sequence $(r_n(T_n-\theta))_n$ is uniformly tight, whence $o_P(r_n\|T_n-\theta\|)=o_P(1)$ [2]. We have just proved the first part of the theorem.

Now $D\phi(\theta)$ being a continuous linear mapping, we have

$$r_nD\phi(\theta)(T_n-\theta)\xrightarrow[n\to\infty]{\mathcal{L}}D\phi(\theta)(T)$$

by the continuity mapping theorem (Theorem 2.6). We conclude using Theorem 2.8, item (iv). □

[1] For example, using the CLT.

[2] One shall write $o_P(r_n\|T_n-\theta\|)=r_n\|T_n-\theta\|Z_n$ with $Z_n=o_P(1)$ then for an $\epsilon>0$ fixed, we take M such that $\mathbb{P}(r_n\|T_n-\theta\|>M)<\epsilon$. It is then easy to see that $\forall\eta>0$, $\mathbb{P}\left(r_n\|T_n-\theta\|Z_n>\eta\right)\to 0$.

Example 2.14 (Fundamental example). *If* $\sqrt{n}\,(T_n-\theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\Sigma)$. *Then*

$$\sqrt{n}\left(\phi(T_n)-\phi(\theta)\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, D\phi(\theta)\Sigma D\phi(\theta)^T).$$

Example 2.15. *Let* (X_i) *be a sequence of i.i.d. random variables distributed as* $\mathcal{E}(\lambda)$, *here* λ *is an unknown parameter in* $(0,+\infty)$. *Then by the CLT we have*

$$\sqrt{n}\left(\overline{X}_n-\frac{1}{\lambda}\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\frac{1}{\lambda^2}).$$

Now applying the Delta method with $\phi(x)=\frac{1}{x}$ *we get*

$$\sqrt{n}\left(\frac{1}{\overline{X}_n}-\lambda\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,\lambda^2).$$

2.3 Asymptotic properties of the Pick-Freeze estimators

2.3.1 Consistency and CLT for $S_N^{\mathbf{u},\mathrm{Cl}}$ and $T_N^{\mathbf{u},\mathrm{Cl}}$

Theorem 2.16 (Consistency). *If* $\mathbb{E}[Y^2]<+\infty$ *then* $S_N^{\mathbf{u},Cl}$ *and* $T_N^{\mathbf{u},Cl}$ *converge a.s. to* $S^{\mathbf{u},Cl}$ *when goes to infinity.*

Proof of Theorem 2.16. It is a simple application of the strong law of large numbers and the continuity mapping theorem (Theorem 2.6). □

Theorem 2.17 (Central limit theorem). *Assume that* $\mathbb{E}[Y^4]<\infty$. *Then:*

1.
$$\sqrt{N}\left(S_N^{\mathbf{u},Cl}-S^{\mathbf{u},Cl}\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_k\left(0,\Gamma_{\mathbf{u},S}\right) \tag{2.6}$$

where $\Gamma_{\mathbf{u},S}=\left((\Gamma_{\mathbf{u},S})_{l,j}\right)_{1\leqslant l,j\leqslant k}$ *with*

$$(\Gamma_{\mathbf{u},S})_{l,j}=\frac{\mathrm{Cov}(YY^{u_l},YY^{u_j})-S^{u_l,Cl}\mathrm{Cov}(YY^{u_j},Y^2)-S^{u_j,Cl}\mathrm{Cov}(YY^{u_l},Y^2)+S^{u_j,Cl}S^{u_l,Cl}\mathrm{Var}(Y^2)}{(\mathrm{Var}(Y))^2}$$

2.
$$\sqrt{N}\left(T_N^{\mathbf{u},Cl}-S^{\mathbf{u},Cl}\right) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_k\left(0,\Gamma_{\mathbf{u},T}\right) \tag{2.7}$$

where $\Gamma_{\mathbf{u},T}=\left((\Gamma_{\mathbf{u},T})_{l,j}\right)_{1\leqslant l,j\leqslant k}$ *with*

$$(\Gamma_{\mathbf{u},T})_{l,j}=\frac{\mathrm{Cov}(YY^{u_l},YY^{u_j})-S^{u_l,Cl}\mathrm{Cov}(YY^{u_j},M^{\mathbf{u}})-S^{u_j,Cl}\mathrm{Cov}(YY^{u_l},M^{\mathbf{u}})+S^{u_j,Cl}S^{u_l,Cl}\mathrm{Var}(M^{\mathbf{u}})}{(\mathrm{Var}(Y))^2}.$$

Proof of Theorem 2.17. Since $S_N^{\mathbf{u},\mathrm{Cl}}$ and $T_N^{\mathbf{u},\mathrm{Cl}}$ are invariant by any centering (translation) of the Y_j's and $Y_j^{\mathbf{u_i}}$'s for $i=1,\dots,k$, we can simplify the next calculations translating by $\mathbb{E}[Y]$. For the sake of simplicity, Y_j and $Y_j^{\mathbf{u_i}}$ now denote the centered random variables.

Proof of (2.6). Recall that $S_N^{\mathbf{u},\mathrm{Cl}}-S^{\mathbf{u},\mathrm{Cl}}$ is equal to

$$\left(\frac{\frac{1}{N}\sum Y_jY_j^{\mathbf{u_1}}-\left(\frac{1}{N}\sum Y_j\right)\left(\frac{1}{N}\sum Y_j^{\mathbf{u_1}}\right)}{\frac{1}{N}\sum Y_j^2-(\frac{1}{N}\sum Y_j)^2}-S^{\mathbf{u_1},\mathrm{Cl}},\dots,\frac{\frac{1}{N}\sum Y_jY_j^{\mathbf{u_k}}-\left(\frac{1}{N}\sum Y_j\right)\left(\frac{1}{N}\sum Y_j^{\mathbf{u_k}}\right)}{\frac{1}{N}\sum Y_j^2-(\frac{1}{N}\sum Y_j)^2}-S^{\mathbf{u_k},\mathrm{Cl}}\right).$$

Let $W_j = (Y_j Y_j^{\mathbf{u_i}}, i = 1,\dots,k, Y_j, Y_j^{\mathbf{u_i}}, i = 1\dots,k, Y_j^2)^t$ ($j = 1,\dots,$ N) and g the mapping from $\mathbb{R}^{2k+2}$ to $\mathbb{R}^k$ defined by

$$g(x_1,\dots,x_k,y,y_1,\dots,y_k,z) = \left(\frac{x_1 - y y_1}{z-y^2},\dots,\frac{x_k - y y_k}{z-y^2}\right).$$

Let Σ denote the covariance matrix of W_j and set

$$E = \mathbb{E}[Y], \quad V = \mathrm{Var}(Y), \quad \text{and} \quad W = (Y Y^{\mathbf{u}}, Y, Y^{\mathbf{u}}, Y^2)^t$$

and

$$\Sigma = \begin{pmatrix} \mathrm{Var}(Y Y^{\mathbf{u}}) & \mathrm{Cov}(Y Y^{\mathbf{u}}, Y) & \mathrm{Cov}(Y Y^{\mathbf{u}}, Y^{\mathbf{u}}) & \mathrm{Cov}(Y Y^{\mathbf{u}}, Y^2) \\ \mathrm{Cov}(Y, Y Y^{\mathbf{u}}) & V & \mathrm{Cov}(Y, Y^{\mathbf{u}}) & \mathrm{Cov}(Y, Y^2) \\ \mathrm{Cov}(Y^{\mathbf{u}}, Y Y^{\mathbf{u}}) & \mathrm{Cov}(Y^{\mathbf{u}}, Y) & V & \mathrm{Cov}(Y^{\mathbf{u}}, Y^2) \\ \mathrm{Cov}(Y^2, Y Y^{\mathbf{u}}) & \mathrm{Cov}(Y^2, Y) & \mathrm{Cov}(Y^2, Y^{\mathbf{u}}) & \mathrm{Var}(Y^2) \end{pmatrix}$$

First, the following central limit theorem holds

$$\sqrt{N}\left(\frac{1}{N}\sum W_j - \mathbb{E}[W]\right) \xrightarrow[N\to\infty]{\mathscr{L}} \mathcal{N}_{2k+2}(0,\Sigma)$$

We then apply the so-called Delta method to W and g so that

$$\sqrt{N}\left(g(\overline{W}_N) - g(\mathbb{E}[W])\right) \xrightarrow[N\to\infty]{\mathscr{L}} \mathcal{N}\left(0, J_g(\mathbb{E}[W])\Sigma J_g(\mathbb{E}[W])^t\right).$$

with $J_g(\mathbb{E}[W])$ the Jacobian of g at point $\mathbb{E}(W)$. Define $g = (g_1,\dots,g_k)$. For $i,i' = 1,\dots,k$,

$$\begin{cases} \frac{\partial g_{i'}}{\partial x_i}(\mathbb{E}[W]) = \delta_{i,i'}/V \\ \frac{\partial g_{i'}}{\partial y}(\mathbb{E}[W]) = 0 \\ \frac{\partial g_{i'}}{\partial y_i}(\mathbb{E}[W]) = 0 \\ \frac{\partial g_{i'}}{\partial z}(\mathbb{E}[W]) = -S^{\mathbf{u_i},\mathrm{Cl}}/V \end{cases}$$

with $\delta_{i,i} = 1$ and $\delta_{i,i'} = 0$ if $i \neq i'$. Thus $\Gamma_{\mathbf{u},S} = J_g(\mathbb{E}[W])\Sigma J_g(\mathbb{E}[W])^t$ is as stated in Theorem 2.17.

Proof of (2.7). The proof is similar to the one of (2.6). We now define $W_j = (Y_j Y_j^{\mathbf{u_i}}, i = 1,\dots,k, Y_j, Y_j^{\mathbf{u_i}}, i = 1\dots,k, \overline{(Y_j^{\mathbf{u}})^2})^t$. We apply the delta method to g from $\mathbb{R}^{2k+2}$ into $\mathbb{R}^k$ defined by

$$g(x_1,\dots,x_k,y,y_1,\dots,y_k,z) = \left(\frac{x_1 - \left(\frac{y+y_1}{2}\right)^2}{z - \left(\frac{y+y_1+\dots+y_k}{k+1}\right)^2},\dots,\frac{x_k - \left(\frac{y+y_k}{2}\right)^2}{z - \left(\frac{y+y_1+\dots+y_k}{k+1}\right)^2}\right).$$

For $i,i' = 1,\dots,k$,

$$\begin{cases} \frac{\partial g_{i'}}{\partial x_i} u(\mathbb{E}[W]) = \delta_{i,i'}/V \\ \frac{\partial g_{i'}}{\partial y}(\mathbb{E}[W]) = 0 \\ \frac{\partial g_{i'}}{\partial y_i}(\mathbb{E}[W]) = 0 \\ \frac{\partial g_{i'}}{\partial z}(\mathbb{E}[W]) = -S^{\mathbf{u_i},\mathrm{Cl}}/V. \end{cases}$$

□

Remark 2.18. *Note that in Theorem 2.17, we have the stronger assumption $\mathbb{E}[Y^4] < \infty$. Indeed, since we want a CLT for sums of quantities like Y_i^2, it is necessary to impose that Y_i^2 has a second order moment that is $\mathbb{E}[Y^4] < \infty$.*

Example 2.19. *1. Assume $k=d$, $u=\big(\{1\},\dots,\{p\}\big)$ and $\mathbb{E}[Y^4]<\infty$. Here*

$$S^{\mathbf{u},Cl}=\left(\frac{\mathrm{Var}(\mathbb{E}[Y|X_1])}{\mathrm{Var}(Y)},\dots,\frac{\mathrm{Var}(\mathbb{E}[Y|X_d])}{\mathrm{Var}(Y)}\right)$$

and

$$T_N^{\mathbf{u},Cl}=\left(\frac{\frac{1}{N}\sum Y_jY_j^1-\left(\frac{1}{2N}\sum(Y_j+Y_j^1)\right)^2}{\frac{1}{N}\sum M_j^{\mathbf{u}}-\left(\frac{1}{N}\sum Z_j^{\mathbf{u}}\right)^2},\dots,\frac{\frac{1}{N}\sum Y_jY_j^p-\left(\frac{1}{2N}\sum(Y_j+Y_j^p)\right)^2}{\frac{1}{N}\sum M_j^{\mathbf{u}}-\left(\frac{1}{N}\sum Z_j^{\mathbf{u}}\right)^2}\right).$$

The CLT becomes

$$\sqrt{N}\left(T_N^{\mathbf{u},Cl}-S^{\mathbf{u},Cl}\right)\underset{N\to\infty}{\overset{\mathscr{L}}{\to}}\mathcal{N}_d\left(0,\Gamma_{\mathbf{u},T}\right)$$

where $\Gamma_{\mathbf{u},T}=\big((\Gamma_{\mathbf{u},T})_{i,i'}\big)_{1\leqslant i,i'\leqslant d}$ with

$$\begin{aligned}(\mathrm{Var}(Y))^2\,(\Gamma_{\mathbf{u},T})_{i,i'}=\mathrm{Cov}(YY^i,YY^{i'})-S^{i,Cl}\mathrm{Cov}(YY^{i'},M^{\mathbf{u}})-S^{i',Cl}\mathrm{Cov}(YY^l,M^{\mathbf{u}})\\+S^{i,Cl}S^{i',Cl}\mathrm{Var}(M^{\mathbf{u}}).\end{aligned}$$

2. We can obviously have a CLT for any index of order 2. Indeed if we take, $k=1$ and $(i,i')\in\{1,\dots,d\}^2$ with $i\neq i'$ and $u=\{i,i'\}$. We get $Z^{\mathbf{u}}=\frac{1}{2}\left(Y+Y^{\mathbf{u}}\right)$ and $M^{\mathbf{u}}=\frac{1}{2}\left(Y^2+(Y^{\mathbf{u}})^2\right)$; thus

$$S^{\mathbf{u},\mathrm{Cl}}=\frac{\mathrm{Var}(\mathbb{E}[Y|X_i,X_{i'}])}{\mathrm{Var}(Y)}\quad\text{and}\quad T_N^{\mathbf{u},\mathrm{Cl}}=\frac{\frac{1}{N}\sum Y_jY_j^{\mathbf{u}}-\left(\frac{1}{2N}\sum(Y_j+Y_j^{\mathbf{u}})\right)^2}{\frac{1}{2N}\sum\left(Y_j^2+(Y_j^{\mathbf{u}})^2\right)-\left(\frac{1}{2N}\sum(Y_j+Y_j^{\mathbf{u}})\right)^2}.$$

The CLT becomes

$$\sqrt{N}\left(T_N^{\mathbf{u},\mathrm{Cl}}-S^{\mathbf{u},\mathrm{Cl}}\right)\underset{N\to\infty}{\overset{\mathscr{L}}{\to}}\mathcal{N}_1\left(0,\Gamma_{\mathbf{u},T}\right)$$

with

$$(\mathrm{Var}(Y))^2\,(\Gamma_{\mathbf{u},T})=\mathrm{Var}(YY^{\mathbf{u}})-2S^{\mathbf{u},\mathrm{Cl}}\mathrm{Cov}(YY^{\mathbf{u}},Y^2)+\frac{\left(S^{\mathbf{u},\mathrm{Cl}}\right)^2}{2}\left(\mathrm{Var}(Y^2)+\mathrm{Cov}(Y^2,(Y^{\mathbf{u}})^2)\right).$$

3. One can also straightforwardly deduce the joint distribution of the vector of all indices of order 2. For example, if $p=3$ take $k=3$ and $\mathbf{u}=(\{1,2\},\{1,3\},\{2,3\})$ and apply Theorem 2.17.

Proposition 2.20. *Assume that $k=1$. The asymptotic variance σ_T^2 of $T_N^{\mathbf{u},Cl}$ is always less than or equal to the asymptotic variance σ_S^2 of $T_N^{\mathbf{u},Cl}$, with equality if and only if $S^{\mathbf{u},Cl}=0$ or $S^{\mathbf{u},Cl}=1$.*

To prove this Proposition, we need the following immediate Lemma:

Lemma 2.21. *Y and $Y^{\mathbf{u}}$ are exchangeable random variables, ie. $(Y,Y^{\mathbf{u}})\overset{\mathscr{L}}{=}(Y^{\mathbf{u}},Y)$.*

Proof of Lemma 2.21. Assume that $d=2$, $\mathbf{u}=\{1\}$ and denote $X=X_1$ and $Z=X_2$. The general case stems easily. For any $a,b\in\mathbb{R}$, we have:

$$\begin{aligned}\mathbb{P}(Y\leqslant a,Y^X\leqslant b)&=\mathbb{P}(f(X,Z)\leqslant a,f(X,Z')\leqslant b)\\&=\int\mathbb{P}(f(X,Z)\leqslant a,f(X,Z')|X=x)\mathrm{d}P(x)\\&=\int\mathbb{P}(f(X,Z)\leqslant a|X=x)\mathbb{P}(f(X,Z')\leqslant b|X=x)\mathrm{d}P(x),\text{ by independence conditonally on }X\\&=\int\mathbb{P}(f(X,Z)\leqslant b|X=x)\mathbb{P}(f(X,Z')\leqslant a|X=x)\mathrm{d}P(x),\text{ as }Z\overset{\mathscr{L}}{=}Z'\\&=\mathbb{P}(Y\leqslant b,Y^X\leqslant a).\end{aligned}$$

□

Proof of Proposition 2.20. The limiting variances σ_S^2 and σ_T^2 (given in Theorem 2.17) are translation-invariant, so that one may assume without loss of generality that $\mathbb{E}[Y]=0$. Expanding the variances and using the exchangeability of Y

and $Y^{\mathbf{u}}$, $(\mathrm{Var}(Y))^2\left(\sigma_S^2-\sigma_T^2\right)$ is equal to:

$$(\mathrm{Var}(Y))^2\left(\sigma_S^2-\sigma_T^2\right)=\frac{(S^{\mathbf{u},\mathrm{Cl}})^2}{2}\left(\mathrm{Var}(Y^2)-\mathrm{Cov}\left(Y^2,(Y^{\mathbf{u}})^2\right)\right).$$

We now use Cauchy-Schwarz inequality to see that:

$$\mathrm{Cov}\left(Y^2,(Y^{\mathbf{u}})^2\right)\leqslant\sqrt{\mathrm{Var}\left(Y^2\right)\mathrm{Var}\left((Y^{\mathbf{u}})^2\right)}=\mathrm{Var}\left(Y^2\right)$$

so the second term is always non-negative. This proves that the asymptotic variance of $S_N^{\mathbf{u},\mathrm{Cl}}$ is greater than the asymptotic variance of $T_N^{\mathbf{u},\mathrm{Cl}}$.
For the equality case, we notice that $S^{\mathbf{u},\mathrm{Cl}}=0$ implies the equality of the asymptotic variances. If $S^{\mathbf{u},\mathrm{Cl}}\neq 0$, equality holds if and only if there is equality in Cauchy-Schwarz, ie. there exists $k\in\mathbb{R}$ so that:

$$Y^2=k(Y^{\mathbf{u}})^2 \text{ almost surely}$$

by taking expectations and using $\mathrm{Var}(Y)=\mathrm{Var}(Y^{\mathbf{u}})$ we see that $k=1$ necessarily, hence $Y=Y^{\mathbf{u}}$ almost surely, and $S^{\mathbf{u},\mathrm{Cl}}=1$ thanks to (2.3). □

2.3.2 Asymptotic efficiency of $T_N^{\mathbf{u},\mathbf{Cl}}$

In this section we study the asymptotic efficiency of $S_N^{\mathbf{u},\mathrm{Cl}}$ and $T_N^{\mathbf{u},\mathrm{Cl}}$. This notion (see [122], Section 25 for its definition) extends the notion of Cramér-Rao bound to the semiparametric setting and enables to define a criteria of optimality for estimators, called asymptotic efficiency.
Let $\mathscr{P}$ be the set of all cumulative distribution functions (cdf) of exchangeable random vectors in $L^2(\mathbb{R}^2)$. It is clear that the cdf Q of a random vector of $L^2(\mathbb{R}^2)$ is in $\mathscr{P}$ if and only if Q is symmetric:

$$Q(a,b)=Q(b,a)\quad\forall(a,b)\in\mathbb{R}^2.$$

Let P be the cdf of (Y,Y^X). We have $P\in\mathscr{P}$ thanks to Lemma 2.21.

Proposition 2.22 (Asymptotic efficiency). *$(T_N^{\mathbf{u},Cl})_N$ is asymptotically efficient for estimating $S^{\mathbf{u},Cl}$ for $P\in\mathscr{P}$.*

We will use the following Lemma, which is also of interest in its own right:

Lemma 2.23 (Asymptotic efficiency in $\mathscr{P}$).

1. *Let $\Phi_1:\mathbb{R}\to\mathbb{R}$ be a function in $L^2(P)$. The sequence of estimators $(\Phi_N^1)_N$ given by:*
$$\Phi_N^1=\frac{1}{N}\sum\frac{\Phi_1(Y_i)+\Phi_1(Y_i^X)}{2}$$
is asymptotically efficient for estimating $\mathbb{E}[\Phi_1(Y)]$ for $P\in\mathscr{P}$.
2. *Let $\Phi_2:\mathbb{R}^2\to\mathbb{R}$ be a symmetric function in $L^2(P)$. The sequence $(\Phi_N^2)_N$ given by:*
$$\Phi_N^2=\frac{1}{N}\sum\Phi_2\left(Y_i,Y_i^X\right)$$
is asymptotically efficient for estimating $\mathbb{E}[\Phi_2(Y,Y^X))$ for $P\in\mathscr{P}$.

Proof of Lemma 2.23. Let, for $g\in L^2(P)$ and $t\in\mathbb{R}$, P_t^g be the cdf satisfying:

$$\mathrm{d}P_t^g=(1+tg)\mathrm{d}P.$$

It is clear that the tangent set of $\mathscr{P}$ at P is the closure of:

$$\dot{\mathscr{P}}_P=\{g \text{ bounded}, \mathbb{E}[g(Y,Y^X)]=0 \text{ and } g(a,b)=g(b,a)\ \forall(a,b)\in\mathbb{R}^2\}.$$

Let, for $Q\in\mathscr{P}$:

$$\Psi_1(Q)=\mathbb{E}_Q\left[\Phi_1(Y)\right]\quad\text{and}\quad\Psi_2(Q)=\mathbb{E}_Q\left[\Phi_2(Y,Y^X)\right].$$

We recall that $\mathbb{E}_Q$ denotes the expectation obtained by assuming that the random vector (Y, Y^X) follows the Q distribution.
Following [122] Section 25.3, we compute the efficient influence functions of Ψ_1 and Ψ_2 with respect to $\mathscr{P}$ and the tangent set $\dot{\mathscr{P}}_P$. These empirical influence functions are related to the minimal asymptotic variance of a regular estimator sequence whose observations lie in $\mathscr{P}$ (op.cit., Theorems 25.20 and 25.21). Let $g \in \dot{\mathscr{P}}_P$.

1. We have

$$\frac{\Psi_1(P_t^g) - \Psi_1(P)}{t} = \mathbb{E}_P\left[\Phi_1(Y) g(Y, Y^X)\right] = \mathbb{E}_P\left[\left(\frac{\Phi_1(Y) + \Phi_1(Y^X)}{2} - \mathbb{E}(\Phi_1(Y))\right) g(Y, Y^X)\right].$$

   As:

$$\widetilde{\Psi_{1,P}} = \frac{\Phi_1(Y) + \Phi_1(Y^X)}{2} - \mathbb{E}[\Phi_1(Y)] \in \dot{\mathscr{P}}_P,$$

   it is the efficient influence function of Ψ_1 at P. Hence the efficient asymptotic variance is:

$$\mathbb{E}_P\left[\left(\widetilde{\Psi_{1,P}}\right)^2\right] = \frac{\mathrm{Var}\left(\Phi_1(Y) + \Phi_1(Y^X)\right)}{4}.$$

   As, by the central limit theorem, (Φ_N^1) clearly achieves this efficient asymptotic variance, it is an asymptotically efficient estimator of $\Psi_1(P)$.

2. We have:

$$\frac{\Psi_2(P_t^g) - \Psi_2(P)}{t} = \mathbb{E}_P\left[\Phi_2(Y, Y^X) g(Y, Y^X)\right] = \mathbb{E}_P\left[\left(\Phi_2(Y, Y^X) - \mathbb{E}(\Phi_2(Y, Y^X))\right) g(Y, Y^X)\right].$$

   Thanks to the symmetry of Φ_2, we have that

$$\widetilde{\Psi_{2,P}} = \Phi_2(Y, Y^X) - \mathbb{E}(\Phi_2(Y, Y^X))$$

   belongs to $\dot{\mathscr{P}}_P$, hence it is the efficient influence function of Ψ_2. So the efficient asymptotic variance is:

$$\mathbb{E}_P\left[\left(\widetilde{\Psi_{2,P}}\right)^2\right] = \mathrm{Var}\left(\Phi_2(Y, Y^X)\right),$$

   and this variance is achieved by (Φ_N^2). □

2.3.3 Application to significance test

In order to simplify the notation, we write the vectors $S^{\mathbf{u},\mathrm{Cl}}$ as column vectors. In this section, we give a general procedure to build significance tests of level α. Then we illustrate this procedure on two examples.
Let $\mathbf{u} = (\mathbf{u_1}, \ldots, \mathbf{u_k})$ so that for any $i = 1, \ldots, k$, $\mathbf{u_j}$ is a subset of $I_d = \{1, \ldots, d\}$. Similarly, let $\mathbf{v} = (\mathbf{v_1}, \ldots, \mathbf{v_l})$ and $\mathbf{w} = (\mathbf{w_1}, \ldots, \mathbf{w_l})$ so that for any $i = 1, \ldots, l$, $\mathbf{v_i} I_d$ and $\mathbf{w_i}$ are subset of I_d.
Consider the following general testing problem:

$$H_0 : S^{\mathbf{u},\mathrm{Cl}} = 0 \text{ and } S^{\mathbf{v},\mathrm{Cl}} = S^{\mathbf{w},\mathrm{Cl}} \quad \text{against} \quad H_1 : H_0 \text{ is not true.}$$

Remark 2.24. *Note that one can also perform tests of the following form:*

$$H_0 : S^{\mathbf{u},Cl} \leqslant s \quad \textit{against} \quad H_1 : S^{\mathbf{u},Cl} > s,$$

or

$$H_0 : S^{\mathbf{u},Cl} \leqslant S^{\mathbf{v},Cl} \quad \textit{against} \quad H_1 : S^{\mathbf{u},Cl} > S^{\mathbf{v},Cl}.$$

Applying Theorem 2.17 we have

$$G_N = \sqrt{N}\left(\begin{pmatrix} S_N^{\mathbf{u},\mathrm{Cl}} \\ S_N^{\mathbf{v},\mathrm{Cl}} - S_N^{\mathbf{w},\mathrm{Cl}} \end{pmatrix} - \begin{pmatrix} S^{\mathbf{u},\mathrm{Cl}} \\ S^{\mathbf{v},\mathrm{Cl}} - S^{\mathbf{w},\mathrm{Cl}} \end{pmatrix}\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}_{k+l}(0, \Gamma). \tag{2.8}$$

Since we have an explicit expression of Γ in terms of variances and covariances, we can build an estimator Γ_N of Γ thanks to empirical means. Note that $(\Gamma_N)_N$ converges a.s. to Γ. Define

$$\widetilde{G}_N = \sqrt{N}\begin{pmatrix} S_N^{\mathbf{u},\mathrm{Cl}} \\ S_N^{\mathbf{v},\mathrm{Cl}} - S_N^{\mathbf{w},\mathrm{Cl}} \end{pmatrix}.$$

Then:

$$G_N = \widetilde{G}_N - \begin{pmatrix} S^{\mathbf{u},\mathrm{Cl}} \\ S^{\mathbf{v},\mathrm{Cl}} - S^{\mathbf{w},\mathrm{Cl}} \end{pmatrix}.$$

Corollary 2.25. *Under H_0, $\widetilde{G}_N \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_{k+l}(0,\Gamma)$. Under H_1, $|\widetilde{G}_N(1)| + |\widetilde{G}_N(2)| \underset{N\to\infty}{\overset{a.s.}{\to}} \infty$.*

This corollary allows us to construct several tests. It is a well-known fact that in the case of a vectorial null hypothesis "there exists no uniformly most powerful test, not even among the unbiased tests". In practice, we return to the dimension 1 introducing a function $F : \mathbb{R}^{k+l} \to \mathbb{R}$ and testing $H_0(F) : F(h) = 0$ (respectively $H_1(F) : F(h) \neq 0$) instead of $H_0 : h = 0$ (resp. $H_1 : h \neq 0$). The choice of a reasonable test "depends on the alternatives at which we wish a high power".

Remark 2.26. *If we take as test statistic $T_N = A\widetilde{G}_N$ where A is a linear form defined on $\mathbb{R}^{l+k}$, then, under H_0, $T_N \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, A\Gamma A'\right)$. Replacing Γ by Γ_N and using Slutsky's lemma, we get*

$$(A\Gamma_N A')^{-1/2} T_N \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,1).$$

Thus we reject H_0 if $(A\Gamma_N A')^{-1/2} T_N \geqslant z_\alpha$ where z_α is the $1-\alpha$ quantile of a standard Gaussian random variable. One can have a similar result when A is not anymore linear but only C^1 by applying the so-called Delta method.

Example 2.27. *We compare 5 different test statistics through their power function. Let $\mathbb{X} = (X_1, X_2) \sim \mathcal{N}(0, I_2)$, and*

$$Y = f(\mathbb{X}) = \lambda_1 X_1 + \lambda_1 X_2 + \lambda_2 X_1 X_2,$$

with $2\lambda_1^2 + \lambda_2^2 = 1$. We consider here the following testing problem

$$H_0 : S^{1,Cl} = S^{2,Cl} = \lambda_1^2 = 0 \quad \textit{against} \quad H_1 : \lambda_1 \neq 0.$$

Computations lead to

$$\Gamma(1,1) = \Gamma(2,2) = 3 - 2\lambda_1^2 - 11\lambda_1^4 + 24\lambda_1^6 - 24\lambda_1^8$$
$$\Gamma(2,1) = \Gamma(1,2) = -7\lambda_1^4 + 24\lambda_1^6 - 24\lambda_1^8.$$

The Gaussian limit in Theorem 2.17 is $\mathcal{N}_2(0, 3Id_2)$ under H_0 while it is asymptotically distributed as $\mathcal{N}_2(0,\Gamma)$ under H_1.

***Test 1**: we take as test statistic $T_{N,1} = \widetilde{G}_N(1) + \widetilde{G}_N(2)$.*
Under H_0, $T_{N,1} \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,6)$ so we reject H_0 if $T_{N,1} > z_\alpha$ where $z_\alpha/\sqrt{6}$ is the $(1-\alpha)$ quantile of a standard Gaussian random variable.While under H_1, following the procedure of Remark 2.26 with $A = (1\ 1)$.

$$\left(T_{N,1} - 2\sqrt{N}\lambda_1^2\right) / \left(2[\Gamma(1,1) + \Gamma(1,2)]\right)^{1/2} \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0,1).$$

It is then easy to compute the theoretical power function.

***Test 2**: since the Sobol indices are non negative, the testing problem is naturally unilateral. However in view of more general contexts we introduce the test statistic $T_{N,2} = |\widetilde{G}_N(1)| + |\widetilde{G}_N(2)|$. We reject H_0 if $T_{N,2} > z_\alpha$ where $z_\alpha/\sqrt{3}$ is the $(1-\alpha)$ quantile of the random variable having*

$$\frac{2}{\sqrt{\pi}} e^{-u^2/4} \Phi(u/\sqrt{2})\, \mathbb{1}_{\mathbb{R}_+}(u)$$

as density (Φ being the distribution function of a standard Gaussian random variable). Under H_1, the power function of $T_{N,2}$ and the limit variance are estimated using Monte Carlo techniques.

Test 3: *in the same spirit, we introduce the test statistic* $T_{N,3} = |\widetilde{G}_N(1) + \widetilde{G}_N(2)|$. *We reject* H_0 *if* $T_{N,3} > z_\alpha$ *where* $z_\alpha/\sqrt{6}$ *is the* $(1-\alpha/2)$ *quantile of a standard Gaussian random variable.Under* H_1, *the power function of* $T_{N,3}$ *and the limit variance are estimated using Monte Carlo techniques.*

Test 4: *we use the* L^2 *norm and consider* $T_{N,4} = (G_N(1))^2 + (G_N(2))^2$. *Under* H_0, $T_{N,4}/3 \underset{N\to\infty}{\overset{\mathcal{L}}{\to}} \chi_2(2)$ *so we reject* H_0 *if* $T_{N,4} > z_\alpha$ *where* $z_\alpha/3$ *is the* $(1-\alpha)$ *quantile of a* χ_2 *random variablewith 2 degrees of freedom. Under* H_1, *the power function of* $T_{N,4}$ *and the limit variance are estimated using Monte Carlo techniques.*

Test 5: *we use the infinity norm and consider* $T_{N,5} = \max(|G_N(1)|;|G_N(2)|)$. *We reject* H_0 *if* $T_{N,5} > z_\alpha$ *where* $z_\alpha/\sqrt{3}$ *is the* $[1+\sqrt{1-\alpha}]/2$ *quantile of a standard Gaussian random variable.Under* H_1, *the power function of* $T_{N,5}$ *and the limit variance are estimated using Monte Carlo techniques. In Figure 2.1, we thus present the plot of the different power functions for* $N = 100$, 500 *and* 1000. *Figure 2.1 shows, as expected, that increasing* N *leads to a steeper power function (hence, a better discrimination between the hypothesis), and that the estimated power function gets closer to the true one. We also see that no test is the most powerful, uniformly in* λ_1, *in accordance with the theory quoted above.*

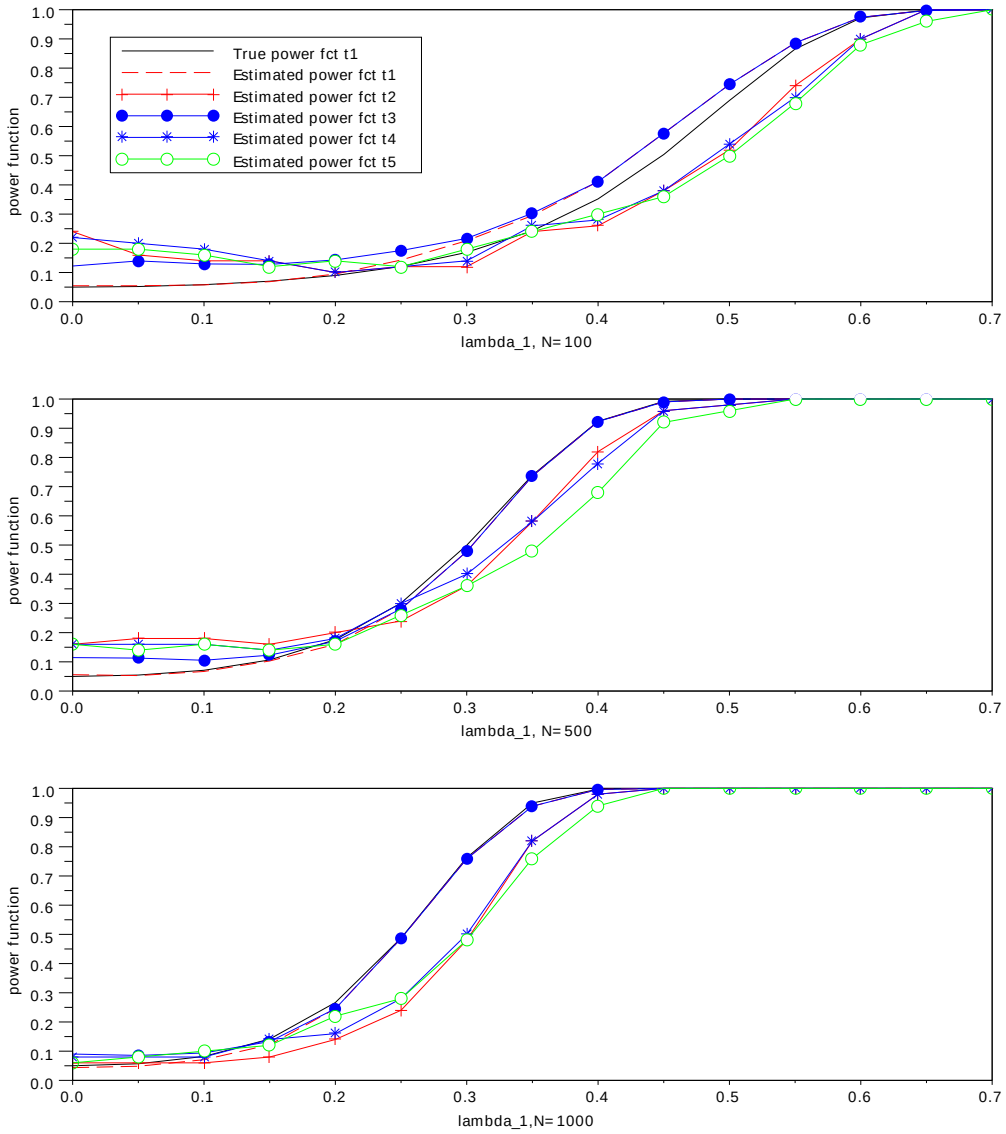


Figure 2.1: Estimated power functions for different values of N.

Numerical application: a real test case

It is customary in aeronautics to model the fuel mass needed to link two fixed countries with a commercial aircraft by the Bréguet formula:

$$M_{fuel} = \left(M_{empty} + M_{pload}\right)\left(e^{\frac{SFC \cdot g \cdot Ra}{V \cdot F}\,10^{-3}} - 1\right). \tag{2.9}$$

The fixed variables are

- M_{empty} : *Empty weight* = basic weight of the aircraft (excluding fuel and passengers)
- M_{pload} : *Payload* = maximal carrying capacity of the aircraft
- g : Gravitational constant
- Ra : *Range* = distance traveled by the aircraft

The uncertain variables are

- V : *Cruise speed* = aircraft speed between ascent and descent phase
- F : *Lift-to-drag ratio* = aerodynamic coefficient
- SFC : *Specific Fuel Consumption* = characteristic value of engines

We model the uncertainties as presented in Table 2.1.

variable	density	parameter
V	*Uniform*	(V_{min}, V_{max})
F	*Beta*	$(7, 2, F_{min}, F_{max})$
SFC	$\theta_2\, e^{-\theta_2(u-\theta_1)}\, \mathbb{1}_{[\theta_1,+\infty[}$	$\theta_1 = 17.23, \theta_2 = 3.45$

Table 2.1: Uncertainty modeling

The probability density function of a beta distribution on $[a,b]$ with shape parameters (α,β) is

$$g_{(\alpha,\beta,a,b)}(x) = \frac{(x-a)^{(\alpha-1)}(b-x)^{\beta-1}}{(b-a)^{\beta-1}B(\alpha,\beta)}\,\mathbb{1}_{[a,b](x)},$$

where $B(\cdot,\cdot)$ is the beta function. We take the nominal and extremal values of V and F as in Table 2.2.

variable	nominal value	min	max
V	**231**	226	234
F	**19**	18.7	19.05

Table 2.2: Minimal and maximal values of uncertain variables

The uncertainty on the cruise speed V represents a relative difference of arrival time of 8 minutes.
The airplane manufacturer may wonder whether he has to improve the quality of the engine (SFC) or the aerodynamical property of the plane (F). Thus we study the sensitivity of M_{fuel} with respect to F and SFC and we want to know if $H_0 : S^{SFC,\mathrm{Cl}} > S^{F,\mathrm{Cl}}$ or $H_1 : S^{SFC,\mathrm{Cl}} \leqslant S^{F,\mathrm{Cl}}$. Applying the test procedure described previously we can not reject H_0.

2.4 Concentration inequalities

2.4.1 Motivation

The starting point is the strong law of large numbers.

Theorem 2.28. *Assume $(X_N)_{N\geqslant 1}$ is a sequence of i.i.d. random variables such that $\mathbb{E}[|X_N|] < +\infty$ then*

$$\frac{X_1 + \ldots + X_N}{N} \xrightarrow[N\to\infty]{a.s.} \mathbb{E}[X_1].$$

For the statistician, $\mathbb{E}[X_1]$ represents an unknown quantity to be estimated and $(X_1+\ldots+X_N)/N$ is an natural estimator. In the real life N never goes to infinity, we only have a finite number of observations ($N=100$, $N=1000$). It is then natural to wonder for a fixed N if $(X_1+\ldots+X_N)/N$ is close or far from $\mathbb{E}[X_1]$. The speed of convergence is also unnatural question we can be interested in. The first answer concerning the rate of convergence is given by the central limit theorem.

Theorem 2.29. *Let $(X_N)_{N\geqslant 1}$ be a sequence of i.i.d. random variables such that the variance σ^2 exists (i.e. $\mathbb{E}[X_1^2]<+\infty$) then*

$$\sqrt{n}\Big(\frac{X_1+\ldots+X_N}{N}-\mathbb{E}[X_1]\Big)\xrightarrow[N\to\infty]{\mathscr{L}}\mathcal{N}(0,\sigma^2).$$

Roughly speaking this theorem, tells us that $(X_1+\ldots+X_N)/N$ goes at rate $\sqrt{N}$ to $\mathbb{E}[X_1]$. Nevertheless, this is an asymptotic result and gives no information when N is fixed (in particular if N is small).
The aim of concentration inequalities is to give non asymptotic results allowing to quantify the error $(X_1+\ldots+X_N)/N-\mathbb{E}[X_1]$ for a fixed N.

There exists several concentration inequalities, we only present the one needed for our purpose.

2.4.2 Bennett's inequality

For $u\in\mathbb{R}$, set $\phi(u)=e^u-u-1$ and for $u\geqslant -1$, $h(u)=(1+u)\log(1+u)-u$.

Theorem 2.30 (Bennett's inequality)**.** *Let $X_1,\ldots,X_N$ be N independent random variables with finite variance. Assume that for all index i, $X_i\leqslant b$. Set*

$$S=\sum_{j=1}^N\big(X_j-\mathbb{E}[X_j]\big)\quad\text{and}\quad v=\sum_{j=1}^N\mathbb{E}[X_j^2].$$

1. *For $t>0$*

$$\log\mathbb{E}[e^{tS}]\leqslant N\log\left(1+\frac{v}{Nb^2\phi(bt)}\right)\leqslant\frac{v}{b^2}\phi(bt) \tag{2.10}$$

2. *For $x>0$,*

$$\mathbb{P}(S\geqslant x)\leqslant\exp\left(-\frac{v}{b^2}h\left(\frac{bx}{v}\right)\right)\leqslant\exp\{-\frac{x^2}{2(v+bx/3)}\}. \tag{2.11}$$

Proof of Theorem 2.30. Left in Exercise 6. □

2.4.3 Concentration inequalities for $S_N^{\mathbf{u},\mathrm{Cl}}$

In this section we give concentration inequalities satisfied by the Sobol indices in one dimension (i.e. $k=1$). Classically, these exponential inequalities can then be used to construct non-asymptotic confidence intervals with a guaranteed coverage probability.

Let us introduce the following random variables

$$U_j^{\pm}=Y_jY_j^{\mathbf{u}}-(S^{\mathbf{u},\mathrm{Cl}}\pm y)(Y_j)^2\text{ et }J_j^{\pm}=(S^{\mathbf{u},\mathrm{Cl}}\pm y)Y_j-Y_j^{\mathbf{u}}$$

Set V_U^+ (resp. V_U^-, V_J^+ and V_J^-) the moment of order 2 of the variables U_j^+ (resp. U_j^-, J_j^+ and J_j^-).

Theorem 2.31. *Soit $b>0$ et $y>0$. We assume that Y_j and $Y_j^{\mathbf{u}}$ belongs to $[-b,b]$. Then*

$$\mathbb{P}\left(S_N^{\mathbf{u},Cl} \geqslant S^{\mathbf{u},Cl}+y\right) \leqslant M_1+2M_2+2M_3, \tag{2.12}$$

$$\mathbb{P}\left(S_N^{\mathbf{u},Cl} \leqslant S^{\mathbf{u},Cl}-y\right) \leqslant M_4+2M_2+2M_5, \tag{2.13}$$

where

$$M_1=\exp\left\{-\frac{NV_U^+}{b_U^2}h\left(\frac{b_U}{V_U^+}\frac{yV}{2}\right)\right\} \qquad M_3=\exp\left\{-\frac{NV_J^+b^2}{b_U^2}h\left(\frac{b_U}{bV_J^+}\sqrt{\frac{yV}{2}}\right)\right\}$$
$$M_2=\exp\left\{-\frac{NV}{b^2}h\left(\frac{b}{V}\sqrt{\frac{yV}{2}}\right)\right\} \qquad M_4=\exp\left\{-\frac{NV_U^-}{b_U^2}h\left(\frac{b_U}{V_U^-}\frac{yV}{2}\right)\right\}$$
$$M_5=\exp\left\{-\frac{NV_J^-b^2}{b_U^2}h\left(\frac{b_U}{bV_J^-}\sqrt{\frac{yV}{2}}\right)\right\}$$

and $b_U=b^2(1+S^{\mathbf{u},Cl}+y)$.

Proof of Theorem 2.31. Since $S^{\mathbf{u},\mathrm{Cl}}$ and $S_N^{\mathbf{u},\mathrm{Cl}}$ are invariant when one translates the variables Y and $Y^{\mathbf{u}}$ we can assume that $\mathbb{E}[Y]=0$.

1. U_j^+ et U_j^- are bounded by b_U, J_j^+ and J_j^- by b_U/b. Moreover

$$\mathbb{E}[U_j^+]=-yV \qquad \mathbb{E}[J_j^+]=0$$
$$\mathbb{E}[U_j^-]=yV \qquad \mathbb{E}[J_j^-]=0$$

and

$$V_U^{\pm}=\mathrm{Var}(YY^{\mathbf{u}})+(S^{\mathbf{u},\mathrm{Cl}}+y)^2\mathrm{Var}(Y^2)-2(S^{\mathbf{u},\mathrm{Cl}}\pm y)\mathrm{Cov}(YY^{\mathbf{u}},Y^2)+y^2V^2$$

$$V_J^{\pm}=((S^{\mathbf{u},\mathrm{Cl}}\pm y)^2+1)V-2(S^{\mathbf{u},\mathrm{Cl}}\pm y)C_u.$$

2. Proof of (2.12). As

$$\{a+b\geqslant c\}\subset\{a\geqslant c/2\}\cup\{b\geqslant c/2\} \quad \text{et} \quad \{ab\geqslant c\}\subset\{|a|\geqslant\sqrt{c}\}\cup\{|b|\geqslant\sqrt{c}\}$$

we have

$$\begin{aligned}
\mathbb{P}\left(S_N^{\mathbf{u},\mathrm{Cl}}\geqslant S^{\mathbf{u},\mathrm{Cl}}+y\right) &= \mathbb{P}\left(\frac{\frac{1}{N}\sum_{j=1}^N Y_jY_j^{\mathbf{u}}-\overline{Y}_N\overline{Y}_N^{\mathbf{u}}}{\frac{1}{N}\sum_{j=1}^N(Y_j)^2-\left(\overline{Y}_N\right)^2}\geqslant S^{\mathbf{u},\mathrm{Cl}}+y\right)\\
&= \mathbb{P}\left(\frac{1}{N}\sum_{j=1}^N\left(U_j^+-\mathbb{E}[U^+]\right)+\overline{Y}_N\overline{J}_N^+\geqslant yV\right)\\
&\leqslant \mathbb{P}\left(\sum_{j=1}^N\left(U_j^+-\mathbb{E}[U^+]\right)\geqslant N\frac{yV}{2}\right)+\mathbb{P}\left(\sum_{j=1}^N Y_j\geqslant N\sqrt{\frac{yV}{2}}\right)\\
&+\mathbb{P}\left(\sum_{j=1}^N(-Y_j)\geqslant N\sqrt{\frac{yV}{2}}\right)+\mathbb{P}\left(\sum_{j=1}^N J_j^+\geqslant N\sqrt{\frac{yV}{2}}\right)\\
&+\mathbb{P}\left(\sum_{j=1}^N(-J_j^+)\geqslant N\sqrt{\frac{yV}{2}}\right).
\end{aligned}$$

Inequality (2.12) comes from the application of Bennett's inequality (apply Bennett's result five time).

3. Proof (2.13). Similarly we have

$$\begin{aligned}\mathbb{P}\left(S_N^{\mathbf{u},\mathrm{Cl}} \leqslant S^{\mathbf{u},\mathrm{Cl}} - y\right) &= \mathbb{P}\left(\frac{1}{N}\sum_{j=1}^{N}\left(-U_j^- + \mathbb{E}[U^-]\right) + (-\overline{Y}_N)\overline{J}_N^- \geqslant yV\right) \\ &\leqslant \mathbb{P}\left(\sum_{j=1}^{N}\left(-U_j^- + \mathbb{E}[U^-]\right) \geqslant N\frac{yV}{2}\right) + \mathbb{P}\left(\sum_{j=1}^{N} Y_j \geqslant N\sqrt{\frac{yV}{2}}\right) \\ &+ \mathbb{P}\left(\sum_{j=1}^{N}(-Y_j) \geqslant N\sqrt{\frac{yV}{2}}\right) + \mathbb{P}\left(\sum_{j=1}^{N} J_j^- \geqslant N\sqrt{\frac{yV}{2}}\right) \\ &+ \mathbb{P}\left(\sum_{j=1}^{N}(-J_j^-) \geqslant N\sqrt{\frac{yV}{2}}\right).\end{aligned}$$

Inequality (2.13) comes from the application of Bennett's inequality (apply Bennett's result five time). □

2.5 Conclusion and remarks on the Pick-Freeze estimation

The Pick-Freeze estimators have desirable statistical properties such as consistency, central limit theorem with a rate of convergence in $\sqrt{n}$, concentration inequalities and Berry-Esseen bounds, and asymptotic efficiency. However, the Pick-Freeze scheme has two major drawbacks. First, it relies on a particular experimental design that may be unavailable in practice. Second, its cost may be prohibitive when estimating several indices. Naturally, the cost of an estimator depends on the cost of each evaluation of the code and on the number of evaluations. The number of model calls to estimate all first-order Sobol' indices grows linearly with the number of input parameters. For example, if we consider $d = 99$ input parameters and only $N = 1000$ calls are allowed, then only a sample of size $N/(d+1) = 10$ is available to estimate each single first-order Sobol' index. It is a poor amount of information to get a satisfying estimation of the Sobol' indices.

2.6 Exercises

Exercise 3. *Show that $S_N^{\mathbf{u},Cl}$ is invariant by any centering (translation) of the Y_i's and $Y_j^{\mathbf{u_i}}$'s for $i = 1,\dots,k$.*

Exercise 1 continued. Ishigami function *Recall that the Ishigami model is given by:*

$$Y = f(X) = f(X_1, X_2, X_3) = \sin X_1 + 7\sin^2 X_2 + 0.1 X_3^4 \sin X_1 \tag{2.14}$$

where $(X_j)_{j=1,2,3}$ are i.i.d. uniform random variables in $[-\pi;\pi]$. Remind that

$$S^1 = 0.3139,\ \ S^2 = 0.4424,\ \ S^3 = 0.$$

1. *Make a program, that gives the Pick and Freeze estimator of these indices (see Equations (2.4) and (2.5)).*
2. *Illustrate Theorem 2.17 in dimension 1.*
3. *Perform simulations in order to show that our test procedure allows us to recover the fact that $S^{3,Cl} = 0$, even for relatively small values of N.*

Exercise 2 continued. Sobol g-function *Recall that the Sobol g-function is given by*

$$Y = g_{sobol}(\mathbb{X}) = g_{sobol}(X_1,\dots,X_d) = \prod_{k=1}^{d} \frac{|4X_k - 2| + a_k}{1 + a_k},$$

where $X_1,\dots,X_d$ are i.i.d. random variables uniformly distributed on $[0,1]$.

1. *Make a program, that gives the Pick and Freeze estimator of these indices (see Equations (2.4) and (2.5)).*
2. *Illustrate Theorem 2.17 in dimension 1.*

Exercise 4. *Let $Y = X_1 + X_2$ with X_1 and X_2 i.i.d. $\mathcal{N}(0,1)$ distributed. Let $\mathbf{u} = (\{1\},\{2\})$ so that*

$$S^{\mathbf{u},Cl} = \left(\frac{\mathrm{Var}(\mathbb{E}[Y|X_1])}{\mathrm{Var}(Y)}, \frac{\mathrm{Var}(\mathbb{E}[Y|X_2])}{\mathrm{Var}(Y)}\right).$$

Compute $S^{\mathbf{u},Cl}$ and give an explicit formula for the covariance matrices of Theorem 2.17.

Exercise 5. *Let $Y = X_1 + X_2$ where X_1 and X_2 are i.i.d. uniformly distributed on $[0,1]$. Let $\mathbf{u} = \{1\}$. Compute $S^{\mathbf{u},Cl}$ and give the bounds M_1, M_2 and M_3 of Theorem 2.31.*

Exercise 6. *The aim here is to prove (2.10) and (2.11).*

1. Check that one can assume (without loss of generality) that $b = 1$.
2. Proof of (2.10).
- Show that $u \mapsto \phi(u)/u^2$ is an increasing function and that $u \mapsto \log(1+u)$ is a concave function.
- Deduce that $e^{tX_j} \leqslant 1 + tX_j + X_j^2\left(e^t - t - 1\right)$.
- Give then an upper bound of $\psi_S(t) = \log\mathbb{E}[e^{tS}]$.
- Conclude to (2.10) using the concavity of $u \mapsto \log(1+u)$.
3. Proof of (2.11).
- Using Markow inequality, prove Markov exponential inequality:

$$\mathbb{P}(X \geqslant x) \leqslant e^{-tx + \log\mathbb{E}[e^{tX}]}$$

for any nonnegative random variable X and $x \geqslant 0$.
- Deduce an upper bound of $\mathbb{P}(S \geqslant x)$.
- Verify that

$$h(u) = (1+u)\ln(1+u) - u \geqslant \frac{u^2}{2(1+u/3)}.$$

to conclude to (2.11).

Exercise 7. *Let $\mathbb{X} = (X_1, X_2, X_3) \sim \mathcal{N}(0, I_3)$, $2\lambda_1^2 + \lambda_2^2 = 1$ and*

$$Y = f(\mathbb{X}) = \lambda_1(X_2 + X_3) + \lambda_2 X_1 X_2.$$

Compute the Sobol' indices and test if X_1 has any influence on the output i.e. $H_0 : S^{1,Cl} = 0$, $S^{\{1,2\},Cl} = S^{2,Cl}$ and $S^{\{1,3\},Cl} = S^{3,Cl}$.

Chapter 3

Extension to multivariate and functional outputs

3.1 Motivation

We begin by considering two examples that enlighten the need for a proper definition of sensitivity indices for multivariate outputs.

Example 3.1. *Let us consider the following nonlinear model*

$$Y = f^{a,b}(X_1, X_2) = \begin{pmatrix} f_1^{a,b}(X_1, X_2) \\ f_2^{a,b}(X_1, X_2) \end{pmatrix} = \begin{pmatrix} X_1 + X_1 X_2 + X_2 \\ aX_1 + bX_1 X_2 + X_2 \end{pmatrix}$$

where X_1 and X_2 are assumed to be i.i.d. standard Gaussian random variables.
First, we compute the one-dimensional Sobol indices $S^j(f_i^{a,b})$ of $f_i^{a,b}$ with respect to X_j $(i, j = 1,2)$. We get

$$(S^1(f_1^{a,b}), S^1(f_2^{a,b})) = (1/3, a^2/(1 + a^2 + b^2))$$
$$(S^2(f_1^{a,b}), S^2(f_2^{a,b})) = (1/3, 1/(1 + a^2 + b^2)).$$

So that, the ratios

$$\frac{S^1(f_i^{a,b})}{S^2(f_i^{a,b})},\ i = 1,2$$

do not depend on b. Moreover, for $|a| > 1$, as this ratio is greater than or equal to 1, X_1 seems to have more influence on the output.
Now let us perform a sensitivity analysis on $\|Y\|^2$. Straightforward calculus lead to

$$S^1(\|Y\|^2) \geqslant S^2(\|Y\|^2) \iff (a-1)(a^3 + a^2 + 5a + 5 - 4b) \geqslant 0.$$

For the quantity $\|Y\|^2$, the region where X_1 is the most influential variable depends on the value of b. This region is not very intuitive and cannot be deduced by our first component-by-component sensitivity analysis. See Figure 3.1.

A second motivation to introduce new Sobol indices is related to the statistical problem of their estimation. As the dimension increases the statistical estimation of the whole vector of scalar Sobol indices becomes more and more expensive. Moreover, the interpretation of such a large vector is not easy. This strengthens the fact that one needs to introduce Sobol indices of small dimension, which condense all the information contained in a large collection of scalars.

In the next section we define new Sobol indices generalizing the scalar ones and containing all the information.

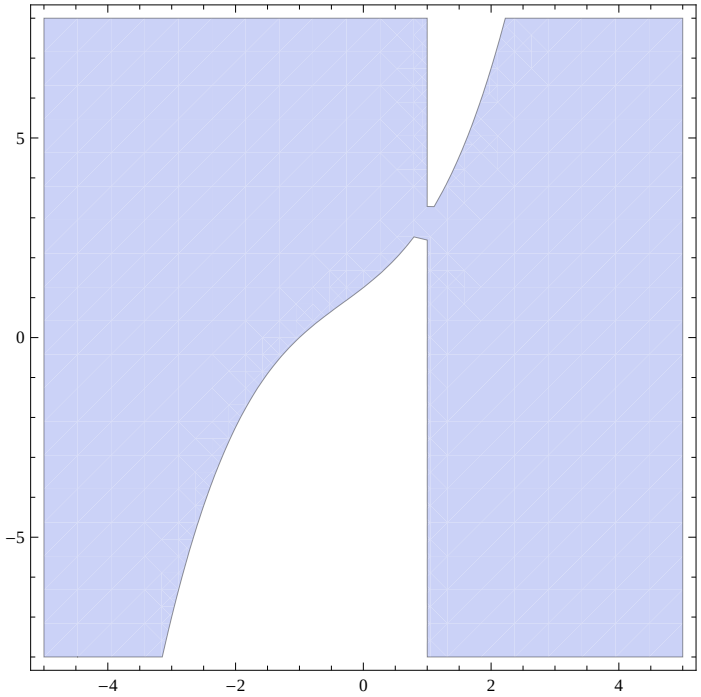


Figure 3.1: Plot of $(a-1)(a^3+a^2+5a+5-4b) \geqslant 0$. The blue corresponds to regions where $S^{\mathbb{1}}(\|Y\|^2) \geqslant S^{\mathbf{2}}(\|Y\|^2)$.

3.2 Case of multivariate outputs

3.2.1 Definition of the new indices

Let us recall the notation. We denote by $\mathbb{X} = (X_1, \dots, X_d)$ the random input, defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and valued in some measurable space $E = E_1 \times \cdots \times E_d$. We denote also by Y the output

$$Y = f(\mathbb{X}) = f(X_1, \dots, X_d),$$

where $f : E \to \mathbb{R}^k$ is an unknown measurable function (d and k are positive integers). We assume that $X_1, \dots, X_d$ are independent and that Y is square integrable (i.e. $\mathbb{E}[\|Y\|^2] < \infty$). We also assume, without loss of generality, that the covariance matrix of Y is positive definite.
Let $\mathbf{u}$ be a subset of $\{1, \dots, d\}$ and denote by $\sim\mathbf{u}$ its complement in $\{1, \dots, d\}$. Further, we set $X_{\mathbf{u}} = (X_i, i \in \mathbf{u})$ and $E_{\mathbf{u}} = \prod_{i \in \mathbf{u}} E_i$.
As the inputs $X_1, \dots, X_d$ are independent, f may be decomposed through the Hoeffding decomposition as in Theorem 1.3:

$$f(X) = c + f_{\mathbf{u}}(X_{\mathbf{u}}) + f_{\sim\mathbf{u}}(X_{\sim\mathbf{u}}) + f_{\mathbf{u},\sim\mathbf{u}}(X_{\mathbf{u}}, X_{\sim\mathbf{u}}), \tag{3.1}$$

where $c \in \mathbb{R}^k$, $f_{\mathbf{u}} : E_{\mathbf{u}} \to \mathbb{R}^k$, $f_{\sim\mathbf{u}} : E_{\sim\mathbf{u}} \to \mathbb{R}^k$ and $f_{\mathbf{u},\sim\mathbf{u}} : E \to \mathbb{R}^k$ are given by

$$c = \mathbb{E}[Y],\ f_{\mathbf{u}} = \mathbb{E}[Y|X_{\mathbf{u}}] - c,\ f_{\sim\mathbf{u}} = \mathbb{E}[Y|X_{\sim\mathbf{u}}] - c,\ f_{u,\sim\mathbf{u}} = Y - f_{\mathbf{u}} - f_{\sim\mathbf{u}} - c.$$

Thanks to L^2-orthogonality, computing the covariance matrix of both sides of (3.1) leads to

$$\Sigma = C_{\mathbf{u}} + C_{\sim\mathbf{u}} + C_{\mathbf{u},\sim\mathbf{u}}. \tag{3.2}$$

Here Σ, $C_{\mathbf{u}}$, $C_{\sim\mathbf{u}}$ and $C_{\mathbf{u},\sim\mathbf{u}}$ are denoting respectively the covariance matrices of Y, $f_{\mathbf{u}}(X_{\mathbf{u}})$, $f_{\sim\mathbf{u}}(X_{\sim\mathbf{u}})$ and $f_{\mathbf{u},\sim\mathbf{u}}(X_{\mathbf{u}}, X_{\sim\mathbf{u}})$.

Remark 3.2. *Notice that for scalar outputs (i.e. when $k = 1$), the covariance matrices are scalar (variances), so that (3.2) may be interpreted as the decomposition of the total variance of Y. The summands traduce the fluctuation induced by the input factors $X_{\mathbf{u}}$ and $X_{\sim\mathbf{u}}$, and the interactions between them. The (univariate) Sobol index $S^{\mathbf{u}}(f) = \mathrm{Var}(\mathbb{E}[Y|X_{\mathbf{u}}))/\mathrm{Var}(Y)$ is then interpreted as the sensibility of Y with respect to $X_{\mathbf{u}}$. Due to non-commutativity of the matrix product, a direct generalization of this index is not straightforward.*

In the general case ($k \geqslant 2$), for any square matrix M of size k, the equation (3.2) can be scalarized in the following way

$$\mathrm{Tr}(M\Sigma) = \mathrm{Tr}(MC_{\mathbf{u}}) + \mathrm{Tr}(MC_{\sim\mathbf{u}}) + \mathrm{Tr}(MC_{\mathbf{u},\sim\mathbf{u}}).$$

This suggests to define as soon as $\mathrm{Tr}(M\Sigma) \neq 0$ the M-sensitivity measure of Y with respect to $X_{\mathbf{u}}$ as

$$S^{\mathbf{u}}(M; f) = \frac{\mathrm{Tr}(MC_{\mathbf{u}})}{\mathrm{Tr}(M\Sigma)}.$$

Of course we can analogously define

$$S^{\sim\mathbf{u}}(M; f) = \frac{\mathrm{Tr}(MC_{\sim\mathbf{u}})}{\mathrm{Tr}(M\Sigma)}, \quad S^{\mathbf{u},\sim\mathbf{u}}(M; f) = \frac{\mathrm{Tr}(MC_{\mathbf{u},\sim\mathbf{u}})}{\mathrm{Tr}(M\Sigma)}.$$

The following lemma is obvious.

Lemma 3.3.

1. *The generalized sensitivity measures sum up to 1*

$$S^{\mathbf{u}}(M; f) + S^{\sim\mathbf{u}}(M; f) + S^{\mathbf{u},\sim\mathbf{u}}(M; f) = 1. \tag{3.3}$$

2. $0 \leqslant S^{\mathbf{u}}(M; f) \leqslant 1$.

3. *Left-composing f by a linear operator O of $\mathbb{R}^k$ changes the sensitivity measure accordingly to*

$$S^{\mathbf{u}}(M; Of) = \frac{\mathrm{Tr}(MOC_{\mathbf{u}}O^t)}{\mathrm{Tr}(MO\Sigma O^t)} = \frac{\mathrm{Tr}(O^tMOC_{\mathbf{u}})}{\mathrm{Tr}(O^tMO\Sigma)} = S^{\mathbf{u}}(O^tMO; f). \tag{3.4}$$

4. *For $k = 1$ and for any $M \neq 0$, we have $S^{\mathbf{u}}(M; f) = S^{\mathbf{u}}(f)$.*

3.2.2 The important identity case

We now consider the special case $M = \mathrm{Id}_k$ (the identity matrix of dimension k). We set $S^{\mathbf{u}}(f) = S^{\mathbf{u}}(\mathrm{Id}_k; f)$. The index $S^{\mathbf{u}}(f)$ has the following obvious properties.

Proposition 3.4.

1. *$S^{\mathbf{u}}(f)$ is invariant by left-composition of f by any isometry of $\mathbb{R}^k$ i.e.*

$$\text{for any square matrix } O \text{ of size } k \text{ s.t. } O^tO = Id_k, \ S^{\mathbf{u}}(Of) = S^{\mathbf{u}}(f);$$

2. *$S^{\mathbf{u}}(f)$ is invariant by left-composition by any nonzero scaling of f i.e.*

$$\text{for any } \lambda \in \mathbb{R}, \ S^{\mathbf{u}}(\lambda f) = S^{\mathbf{u}}(f);$$

Remark 3.5. *The properties in this proposition are natural requirements for a sensitivity measure. In the next section, we will show that these requirements can be fulfilled by $S^{\mathbf{u}}(M; f)$ only when $M = \lambda Id_k$ ($\lambda \in \mathbb{R}^*$). Hence, the canonical choice among indices of the form $S^{\mathbf{u}}(M; f)$ is the sensitivity index $S^{\mathbf{u}}(f)$.*

3.2.3 Identity is the only good choice

The following proposition can be seen as a kind of reciprocal of Proposition 3.4.

Proposition 3.6. *Let M be a square matrix of size k such that*

1. *M does not depend neither on f nor $\mathbf{u}$;*

2. *M has full rank;*

3. *$S^{\mathbf{u}}(M; f)$ is invariant by left-composition of f by any isometry of $\mathbb{R}^k$.*

Then $S^{\mathbf{u}}(M; \cdot) = S^{\mathbf{u}}(\cdot)$.

Proof of Proposition 3.6. We can write $M = M_{Sym} + M_{Antisym}$ where $M^t_{Sym} = M_{Sym}$ and $M^t_{Antisym} = -M_{Antisym}$. Since, for any symmetric matrix V, we have $\mathrm{Tr}(M_{Antisym}V) = 0$, we deduce that $S^{\mathbf{u}}(M;f) = S^{\mathbf{u}}(M_{Sym};f)$ (C_u and Σ being symmetric matrices). Thus we assume, without loss of generality, that M is symmetric.
We diagonalize M in an orthonormal basis: $M = PDP^t$, where $P^tP = \mathrm{Id}_k$ and D diagonal. We have

$$S^{\mathbf{u}}(M;f) = \frac{\mathrm{Tr}(PDP^tC_{\mathbf{u}})}{\mathrm{Tr}(PDP^t\Sigma)} = \frac{\mathrm{Tr}(DP^tC_{\mathbf{u}}P)}{\mathrm{Tr}(DP^t\Sigma P)} = S^{\mathbf{u}}(D;P^tf).$$

By assumption 1. and 3., M can be assumed to be diagonal.
Now we want to show that $M = \lambda\mathrm{Id}_k$ for some $\lambda \in \mathbb{R}^*$. Suppose, by contradiction, that M has two different diagonal coefficients $\lambda_1 \neq \lambda_2$. It is clearly sufficient to consider the case $k = 2$. Choose $f = \mathrm{Id}_2$ (hence, $p = 2$), and $\mathbf{u} = \{1\}$. We have $\Sigma = \mathrm{Id}_2$ and $C_{\mathbf{u}} = \left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)$. Hence on one hand $S^{\mathbf{u}}(M;f) = \frac{\lambda_1}{\lambda_1+\lambda_2}$. On the other hand, let O be the isometry which exchanges the two vectors of the canonical basis of $\mathbb{R}^2$. We have $S^{\mathbf{u}}(M;Of) = \frac{\lambda_2}{\lambda_1+\lambda_2}$. Thus 3. is contradicted if $\lambda_1 \neq \lambda_2$. The case $\lambda = 0$ is forbidden by 2. Finally, it is easy to check that, for any $\lambda \in \mathbb{R}^*$, $S^{\mathbf{u}}(\lambda\mathrm{Id}_k;\cdot) = S^{\mathbf{u}}(\mathrm{Id}_k;\cdot) = S^{\mathbf{u}}(\cdot)$. □

We now give a toy example to illustrate our definition.

Example 3.1 (continued) *We consider the nonlinear model*

$$Y = f^{a,b}(X_1,X_2) = \begin{pmatrix} f_1^{a,b}(X_1,X_2)\\ f_2^{a,b}(X_1,X_2)\end{pmatrix} = \begin{pmatrix} X_1 + X_1X_2 + X_2\\ aX_1 + bX_1X_2 + X_2\end{pmatrix}$$

where X_1 and X_2 are assumed to be i.i.d. standard Gaussian random variables.
We have

$$S^1(f^{a,b}) = \frac{1+a^2}{4+a^2+b^2} \quad \text{and} \quad S^2(f^{a,b}) = \frac{2}{4+a^2+b^2}$$

and obviously

$$S^1(f^{a,b}) \geqslant S^2(f^{a,b}) \iff a^2 \geqslant 1.$$

This result has the natural interpretation that, as X_1 is scaled by a, it has more influence if and only if this scaling enlarges X_1's support i.e. $|a| > 1$.

3.2.4 Estimation of $S^{\mathbf{u}}(f)$

The Pick-Freeze estimator

In practice, the covariance matrices $C_{\mathbf{u}}$ and Σ are not analytically available. So as in the scalar case ($k = 1$),we will estimate $S^{\mathbf{u}}(f)$ by using a Monte-Carlo Pick-Freeze method, which uses a finite sample of evaluations of f.
For this purpose we set $Y^{\mathbf{u}} = f(X_{\mathbf{u}}, X'_{\sim\mathbf{u}})$ where $X'_{\sim\mathbf{u}}$ is an independent copy of $X_{\sim\mathbf{u}}$ which is still independent of $X_{\mathbf{u}}$. Let N be an integer. We take N independent copies $Y_1,\dots,Y_N$ (resp. $Y^{\mathbf{u}}_1,\dots,Y^{\mathbf{u}}_N$) of Y (resp. $Y^{\mathbf{u}}$). For $l = 1,\dots,k$, and $j = 1,\dots,N$, we also denote by $Y_{j,l}$ (resp. $Y^{\mathbf{u}}_{j,l}$) the l^{th} component of Y_j (resp. $Y^{\mathbf{u}}_j$). We then define the following estimator of $S^{\mathbf{u}}(f)$:

$$S^{\mathbf{u}}_N(f) = \frac{\sum_{l=1}^k \left(\frac{1}{N}\sum_{j=1}^N Y_{j,l}Y^{\mathbf{u}}_{j,l} - \left(\frac{1}{N}\sum_{j=1}^N \frac{Y_{j,l}+Y^{\mathbf{u}}_{j,l}}{2}\right)^2\right)}{\sum_{l=1}^k \left(\frac{1}{N}\sum_{j=1}^N \frac{Y^2_{j,l}+(Y^{\mathbf{u}}_{j,l})^2}{2} - \left(\frac{1}{N}\sum_{j=1}^N \frac{Y_{j,l}+Y^{\mathbf{u}}_{j,l}}{2}\right)^2\right)}. \tag{3.5}$$

Remark 3.7. *Note that this estimator can be written*

$$S^{\mathbf{u}}_N(f) = \frac{\mathrm{Tr}\left(C_{\mathbf{u},N}\right)}{\mathrm{Tr}\left(\Sigma_N\right)} \tag{3.6}$$

where $C_{\mathbf{u},N}$ and Σ_N are the empirical estimators of $C_{\mathbf{u}} = \mathrm{Cov}(Y, Y^{\mathbf{u}})$ and $\Sigma = \mathrm{Var}(Y)$ defined by

$$C_{\mathbf{u},N} = \frac{1}{N}\sum_{j=1}^N Y^{\mathbf{u}}_j Y^t_j - \left(\frac{1}{N}\sum_{j=1}^N \frac{Y_j+Y^{\mathbf{u}}_j}{2}\right)\left(\frac{1}{N}\sum_{j=1}^N \frac{Y_j+Y^{\mathbf{u}}_j}{2}\right)^t$$

and

$$\Sigma_N = \frac{1}{N}\sum_{j=1}^N \frac{Y_jY^t_j + Y^{\mathbf{u}}_j(Y^{\mathbf{u}}_j)^t}{2} - \left(\frac{1}{N}\sum_{j=1}^N \frac{Y_j+Y^{\mathbf{u}}_j}{2}\right)\left(\frac{1}{N}\sum_{j=1}^N \frac{Y_j+Y^{\mathbf{u}}_j}{2}\right)^t.$$

Asymptotic properties

A straightforward application of the Strong Law of Large Numbers leads to the following proposition.

Proposition 3.8 (Consistency). *$S_N^{\mathbf{u}}(f)$ converges almost surely to $S^{\mathbf{u}}(f)$ when $N \to +\infty$.*

We now study to the asymptotic normality of $(S_N^{\mathbf{u}}(f))_N$.

Proposition 3.9 (Asymptotic normality). *Assume $\mathbb{E}[Y_l^4] < \infty$ for all $l = 1, \dots, k$. For $l = 1, \dots, k$, we set*

$$U_l = (Y_{1,l} - \mathbb{E}[Y_l])(Y_{1,l}^{\mathbf{u}} - \mathbb{E}[Y_l]), \qquad V_l = (Y_{1,l} - \mathbb{E}[Y_l])^2 + (Y_{1,l}^{\mathbf{u}} - \mathbb{E}[Y_l])^2.$$

Then

$$\sqrt{N}\left(S_N^{\mathbf{u}}(f) - S^{\mathbf{u}}(f)\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}_1\left(0, \sigma^2\right) \tag{3.7}$$

where

$$\sigma^2 = a^2 \sum_{l,l' \in \{1,\dots,k\}} \mathrm{Cov}(U_l, U_{l'}) + b^2 \sum_{l,l' \in \{1,\dots,k\}} \mathrm{Cov}(V_l, V_{l'}) + 2ab \sum_{l,l' \in \{1,\dots,k\}} \mathrm{Cov}(U_l, V_{l'}), \tag{3.8}$$

with

$$a = \frac{1}{\sum_{l=1}^k \mathrm{Var}(Y_l)}, \qquad b = -\frac{a}{2} S^{\mathbf{u}}(f).$$

3.3 Case of functional outputs

In many practical situations the output Y is functional. It is then useful to extend the vectorial indices to functional outputs. This is the aim of this section.

3.3.1 Definition

Let $\mathbb{H}$ be a separable Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$ and the norm $||\cdot||$. Let f be a $\mathbb{H}$-valued function, i.e. Y and $Y^{\mathbf{u}}$ are $\mathbb{H}$-valued random variables. Recall that $\mathbb{E}[Y]$ is defined by duality as the unique member of $\mathbb{H}$ satisfying

$$\mathbb{E}[\langle h, Y \rangle] = \langle h, \mathbb{E}[Y] \rangle \quad \text{for all} \quad h \in \mathbb{H}.$$

We assume that $\mathbb{E}[\|Y\|^2] < \infty$. Recall that the covariance operator associated with Y is the endomorphism Γ on $\mathbb{H}$ defined, for $h \in \mathbb{H}$ by $\Gamma(h) = \mathbb{E}[\langle Y, h \rangle Y]$. We also recall that it is a well known fact that $\mathbb{E}[\|Y\|^2] < \infty$ implies that Γ is then a Trace class operator and its trace is then well defined. We generalize the definition of $S^{\mathbf{u}}(f)$ introduced in Section 3.2 for functional outputs.

Definition 3.10. *$S^{\mathbf{u},\infty}(f) = \frac{\mathrm{Tr}(\Gamma_{\mathbf{u}})}{\mathrm{Tr}(\Gamma)}$, where $\Gamma_{\mathbf{u}}$ is the endomorphism on $\mathbb{H}$ defined by $\Gamma_{\mathbf{u}}(h) = \mathbb{E}[\langle Y^{\mathbf{u}}, h \rangle Y]$ for any $h \in \mathbb{H}$.*

In the next lemma, we give the so-called polar decomposition of the traces of Γ and $\Gamma_{\mathbf{u}}$.

Lemma 3.11. *We have*

$$\mathrm{Tr}(\Gamma) = \mathbb{E}[\|Y\|^2] - \|\mathbb{E}[Y]\|^2$$
$$\mathrm{Tr}(\Gamma_{\mathbf{u}}) = \frac{1}{4}\left[\mathbb{E}[\left\|Y + Y^{\mathbf{u}}\right\|^2 - \mathbb{E}[\left\|Y - Y^{\mathbf{u}}\right\|^2] - 4\,\|\mathbb{E}[Y]\|^2\right].$$

Let $(\varphi_l)_{1 \leqslant l}$ be an orthonormal basis of $\mathbb{H}$. Then

$$\|Y\|^2 = \sum_{i=1}^{\infty} \langle Y, \varphi_i \rangle^2.$$

Now, in order to proceed to the estimation of $S^{\mathbf{u},\infty}(f)$ (and thus first to the estimation of $\mathrm{Tr}(\Gamma)$ and $\mathrm{Tr}(\Gamma_{\mathbf{u}})$), we truncate the previous sum by setting

$$\|Y\|_m^2 = \sum_{i=1}^{m} \langle Y, \varphi_i \rangle^2.$$

Remark 3.12. *It amounts to truncate the expansion of Y to a certain level m. Let Y_m be the truncated approximation of Y:*

$$Y_m = \sum_{l=1}^{m} \langle Y, \varphi_i \rangle \varphi_l,$$

seen as a vector of dimension m. Thus the results of Section 3.2.4 can be applied to Y_m. Notice that Y_m is then the projection of Y onto $Span(\varphi_1, \ldots, \varphi_m)$.

3.3.2 Estimation of $S^{\mathbf{u},\infty}(f)$

As in Section 3.2.4, we define the following estimator of $S^{\mathbf{u},\infty}(f)$:

$$S_N^{\mathbf{u},m}(f) = \frac{\frac{1}{4N}\sum_{j=1}^{N}\left(\left\|Y_j + Y_j^{\mathbf{u}}\right\|_m^2 - \left\|Y_j - Y_j^{\mathbf{u}}\right\|_m^2 - \left\|\overline{Y} + \overline{Y^{\mathbf{u}}}\right\|_m^2\right)}{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\|Y_j\|_m^2 + \left\|Y_j^{\mathbf{u}}\right\|_m^2}{2} - \left\|\frac{\overline{Y}+\overline{Y^{\mathbf{u}}}}{2}\right\|_m^2\right)}.$$

Let T be a $\mathbb{H}$-valued random variable. For any sequence $(T_j)_{i\in\mathbb{N}^*}$ of iid variables distributed as T, we define

$$D_{N,m}(T) = \frac{1}{N}\sum_{j=1}^{N}\left(\|T_i\|_m^2 - \left\|\overline{T}\right\|_m^2\right)$$

and

$$\begin{aligned} e_j &= \mathbb{E}[\langle T_i, \varphi_j \rangle] \\ v_j &= \mathbb{E}[\langle T_i, \varphi_j \rangle^2] \\ Z_{i,j} &= \langle T_i, \varphi_j \rangle - e_j \\ W_{i,j} &= \langle T_i, \varphi_j \rangle^2 - v_j. \end{aligned}$$

In the spirit of [43], we decompose $D_{N,m}(T)$ and give asymptotics for each of the terms of the decomposition.

Proposition 3.13.

1. *$D_{N,m}(T)$ can be rewritten as the sum of a totally degenerated U-statistic of order 2, a centered linear term and a deterministic term in the following way*

$$D_{N,m}(T) - \mathbb{E}\left(\|T\|^2\right) + \|\mathbb{E}(T)\|^2 = -U_N K(T) + P_N L(T) - B_m(T) \tag{3.9}$$

where

$$U_N K(T) := \sum_{l=1}^{m} \frac{1}{N^2} \sum_{1 \leq i \neq j \leq N} Z_{i,l} Z_{j,l}$$

$$P_N L(T) := \frac{1}{N}\left(1 - \frac{1}{N}\right) \sum_{l=1}^{m} \sum_{i=1}^{N} \left(W_{i,l} - 2 e_l Z_{i,l}\right)$$

$$B_m(T) := \sum_{l>m} \left(v_l - e_l^2\right) + \frac{1}{N} \sum_{l=1}^{m} \left(v_l - e_l^2\right).$$

2. *Assume that there exists $\delta > 1$ so that*

$$v_l = \mathbb{E}\left(\langle T, \varphi_l \rangle^2\right) = O(l^{-(\delta+1)}) \tag{3.10}$$

and $\delta' > 1$ so that

$$\mathbb{E}\left(\langle T, \phi_l \rangle^4\right) = O(l^{-\delta'}). \tag{3.11}$$

Then for any $m = m(N)$ so that:

$$\frac{m(N)}{N^{\frac{1}{2\delta}}} \to +\infty, \quad \frac{m(N)}{\sqrt{N}} \to 0, \tag{3.12}$$

we have

   (a) $B_m^2(T) = o(1/N)$

   (b) $\mathbb{E}\left((U_N K(T))^2\right) = o(1/N)$

   (c) $P_N L(T) - P_N L'(T) = o_{\mathbb{P}}\left(\frac{1}{\sqrt{N}}\right)$

where $P_N L'(T) := \frac{1}{N}\left(1 - \frac{1}{N}\right) \sum_{l=1}^{\infty} \sum_{i=1}^{N} \left[W_{i,l} - 2 e_l Z_{i,l}\right]$.

Theorem 3.14. *Suppose that conditions (3.10), (3.11) and (3.12) of Proposition 3.13 are fulfilled. Then we have: Under some mild assumptions, one has*

$$\sqrt{N}(S_N^{\mathbf{u},m}(f) - S^{\mathbf{u},\infty}(f)) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}(0, \sigma^2) \tag{3.13}$$

with σ^2 depending on the moments of $(D_{N,m}(Y), D_{N,m}(Y + Y^{\mathbf{u}}), D_{N,m}(Y - Y^{\mathbf{u}}))$.

3.4 Exercises

Exercise 8. *We consider the following model:*

$$Y = f(X_1, X_2) = \begin{pmatrix} f_1(X_1, X_2) \\ f_2(X_1, X_2) \end{pmatrix} = \begin{pmatrix} X_1 + X_2 + X_1 X_2 \\ aX_1 + bX_1 X_2 + X_2 \end{pmatrix}$$

where X_1 and X_2 independent uniform random variables on $[-1, 1]$.
Compute

- *the sensitivity indices of each of the coordinates of Y: $S^1(f_1)$, $S^1(f_2)$ and $S^2(f_1)$ and $S^2(f_2)$;*
- *the global sensitivity indices of Y: $S^1(f)$ and $S^2(f)$.*

Exercise 9. *We study the following two-dimensional model*

$$Y = f(X_1, X_2) = \begin{pmatrix} f_1(X_1, X_2) \\ f_2(X_1, X_2) \end{pmatrix} = \begin{pmatrix} X_1 \cos X_2 \\ X_1 \sin X_2 \end{pmatrix}$$

with $(X_1, X_2) \sim Unif([0; 10]) \otimes Unif([0; \pi/2])$.
Compute

- *the sensitivity indices of each of the coordinates of Y: $S^1(f_1)$, $S^1(f_2)$ and $S^2(f_1)$ and $S^2(f_2)$;*
- *the sensitivity indices of $\|Y\|^2$: $S^1(\|Y\|^2)$ and $S^2(\|Y\|^2)$;*
- *the global sensitivity indices of Y: $S^1(f)$ and $S^2(f)$.*

Exercise 10. *We consider as first example*

$$Y = f^a(X_1, X_2) = \begin{pmatrix} aX_1 \\ X_2 \end{pmatrix},$$

with X_1 and X_2 i.i.d. standard Gaussian random variables.

Chapter 4

Another estimation of Sobol' indices: rank-based procedure

In a recent work [28], Chatterjee studies the dependence between two variables by introducing an empirical correlation coefficient based on rank statistics, see Section 4.1.1 below for the precise definition. Further, the quantification of the dependence has also been investigated in the bivariate case (namely, in the copula setting), see [4, 37, 121]. The striking point of [28] is that this empirical correlation coefficient converges almost surely (a.s.) to the Cramér-von-Mises index priorly introduced in [49] as the sample size goes to infinity.
In this paper, we show how to embed Chatterjee's method in the GSA framework, thereby eliminating the two drawbacks of the classical Pick-Freeze estimation mentioned above. Thus no particular design of experiment is needed for the estimation that can be done with a unique n-sample. In addition, we generalize Chatterjee's approach to allow the estimation of a large class of GSA indices which includes the Sobol' indices (defined in Section 1.3 of Chapter 1) and the higher-order moment indices proposed by Owen [91–93] (defined in Section 5.1 of Chapter 5). Using a single sample of size n, it is now possible to estimate at the same time all the first-order Sobol' indices, the Cramér-von-Mises indices, and other useful sensitivity indices. Furthermore, we show that this new procedure provides estimators also converging at rate $\sqrt{n}$ by proving a CLT in the estimation of the first-order Sobol' indices.

4.1 A novel generation of estimators based on rank statistics

4.1.1 Chatterjee's correlation coefficient

In [28], Chatterjee considers a pair of real-valued random variables (V, Y) and an i.i.d. sample $(V_j, Y_j)_{1\leqslant j\leqslant n}$. In order to simplify the presentation, we assume that the laws of V and Y are both diffuse (ties are excluded). The pairs $(V_{(1)}, Y_{(1)}), \ldots, (V_{(n)}, Y_{(n)})$ are rearranged in such a way that

$$V_{(1)} < \ldots < V_{(n)}.$$

Then let $\pi(j)$ be the rank of V_j in the sample $(V_1, \ldots, V_n)$ of V and define

$$N'(j) = \begin{cases} \pi^{-1}(\pi(j)+1) & \text{if } \pi(j)+1 \leqslant n, \\ j & \text{if } \pi(j) = n. \end{cases} \tag{4.1}$$

The new correlation coefficient defined by Chatterjee in [28] is denoted $\xi_n(V, Y)$ and given by

$$\frac{1}{n}\sum_{j=1}^{n}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{\{Y_k\leqslant Y_j\}}\mathbb{1}_{\{Y_k\leqslant Y_{N'(j)}\}} - \Big(\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{\{Y_j\leqslant Y_k\}}\Big)^2\Big)\Big/\frac{1}{n}\sum_{j=1}^{n}F_n(Y_j)(1-F_n(Y_j)) \tag{4.2}$$

where F_n stands for the empirical distribution function of Y: $F_n(t) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{\{Y_k\leqslant t\}}$.

The author proves that $\xi_n(V, Y)$ converges a.s. to a deterministic limit $\xi(V, Y)$ given by

$$\xi(V,Y) = \frac{\int_{\mathbb{R}^k}\mathbb{E}\left[\left(F(t)-F^{\mathrm{u}}(t)\right)^2\right]dF(t)}{\int_{\mathbb{R}^k}F(t)(1-F(t))dF(t)} = \frac{\mathbb{E}\left[\mathbb{E}\left[\left(F(Z)-F^{\mathrm{u}}(Z)\right)^2\right]\right]}{\int_{\mathbb{R}^k}F(t)(1-F(t))dF(t)}.$$

This quantity is equal to the Cramér-von-Mises sensitivity index $S^V_{2,CVM}$ with respect to V as soon as V is one of the random variables X_1, ..., X_p in the model (2.1) that are assumed to be real-valued. See Section 5.2 of Chapter 5 for the definition and properties of the Cramér-von Mises indices.

Further, he also proves a CLT when V and Y are independent. Observe that the analogue of the Pick-Freeze version Y^V with respect to V of Y becomes $Y_{N'}$ and

$$\mathrm{Var}(\mathbb{E}[\mathbb{1}_{\{Y\leqslant t\}}|X_{\mathbf{u}}]) = \mathrm{Cov}(\mathbb{1}_{\{Y\leqslant t\}}, \mathbb{1}_{\{Y^{\mathbf{u}}\leqslant t\}}).$$

is replaced by the formula

$$\mathbb{E}[\mathbb{1}_{\{Y_j\geqslant t\}}\mathbb{1}_{\{Y_{N'(j)}\geqslant t\}}|V_1,\dots,V_n] = G_{V_j}(t)G_{V_{N'(j)}}(t) \tag{4.3}$$

for all $j=1,\dots,n$ that is mentioned in the proof of Lemma 7.10 in [28, p.24], with G_V the conditional survival function: $G_V(t)=\mathbb{P}(Y\geqslant t|V)$. It is worth noticing that a unique n sample of input-output provides consistent estimations of the p first-order Cramér-von-Mises indices.

4.1.2 Generalization of Chatterjee's method

In this section, we propose a universal estimation procedure of expectations of the form

$$\mathbb{E}[\mathbb{E}[g(Y)|V]\mathbb{E}[h(Y)|V]],$$

for two integrable functions g and h. In fact, we consider a more general random element V (no longer assumed to be real) and a more general permutation denoted by τ_n. This result is a generalization of (4.3) and can be interpreted as an approximation of $\mathrm{Var}(\mathbb{E}[Y|X_{\mathbf{u}}]) = \mathrm{Cov}(Y, Y^{\mathbf{u}})$. To this end, we introduce the function Ψ_V defined by

$$\Psi_V(g) = \mathbb{E}[g(Y)|V] \tag{4.4}$$

for any integrable function g. Let $\mathscr{F}_n$ be the σ-algebra generated by $\{V_1,\dots,V_n\}$. Note that in Section 5.2 of Chapter 5 dealing with the Cramér-von Mises indices, we will consider $g(x)=g_t(x)=\mathbb{1}_{\{x\geqslant t\}}$ so that $\Psi_V(g)=\mathbb{P}(Y\geqslant t|V)=G_V(t)$.

Lemma 4.1. *Let g and h be two integrable functions such that gh is also integrable. Let $(V_j,Y_j)_{1\leqslant j\leqslant n}$ be an n-sample of (V,Y). Consider a $\mathscr{F}_n$-measurable random permutation τ_n such that $\tau_n(j)\neq j$, for all $j=1,\dots,n$. Then*

$$\mathbb{E}\left[g(Y_j)h(Y_{\tau_n(j)})|V_1,\dots,V_n\right] = \Psi_{V_j}(g)\Psi_{V_{\tau n(j)}}(h). \tag{4.5}$$

The previous lemma (the proof of which has been postponed to Appendix A.1) leads to a generalization of the first part of the numerator of ξ_n defined in (4.2). Following the same lines as in [28], one may prove that such a quantity converges a.s. as $n\to\infty$ under some mild conditions. The reader is referred to Appendix A.1 for the detailed proof of Proposition 4.2.

Proposition 4.2. *Let g and h be two bounded measurable functions. Consider a $\mathscr{F}_n$-measurable random permutation τ_n with no fix point (i.e. $\tau_n(j)\neq j$ for all $j=1,\dots,n$) and such that $V_{\tau_n(i)}\overset{\mathscr{L}}{=}V_{\tau_n(j)}$ for any i and $j=1,\dots,n$. In addition, we assume that for any $j=1,\dots,n$, $V_{\tau_n(j)}\to V_j$ as $n\to\infty$ a.s. Then $\chi_n(V,Y;g,h)$ defined by*

$$\chi_n(V,Y;g,h) = \frac{1}{n}\sum_{j=1}^{n} g(Y_j)h(Y_{\tau_n(j)}) \tag{4.6}$$

converges a.s. as $n\to\infty$ to $\chi(V,Y;g,h)=\mathbb{E}[\Psi_V(g)\Psi_V(h)]$, where Ψ_V has been defined in (4.4).

Notice that the permutation $\tau_n=N$ defined by

$$N(j) = \begin{cases} \pi^{-1}(\pi(j)+1) & \text{if } \pi(j)+1\leqslant n,\\ \pi^{-1}(1) & \text{if } \pi(j)=n. \end{cases} \tag{4.7}$$

satisfies the assumptions of Lemma 4.1 and Proposition 4.2. Observe that N only differs from N' defined in (4.1) at j such that $\pi(j)=n$.

4.2 The rank estimator of the first-order Sobol' indices

4.2.1 Estimation procedure based on rank statistics

We can now leverage the above results and construct a new family of estimators for Sobol' indices. More precisely, let us consider the model (2.1) and assume we want to estimate the first-order Sobol' index S^1 defined in (2.2) with respect

to $V = X_1$ assumed to be real-valued. We then define N as in (4.7) where π is the rank of X_1. Taking $g(x) = h(x) = x$ and $\tau_n = N$, (4.5) provides the analogue to ξ_n to estimate the classical Sobol' indices:

$$\xi_n^{\text{Sobol}}(X_1, Y) := \frac{\frac{1}{n}\sum_{j=1}^n Y_j Y_{N(j)} - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2}{\frac{1}{n}\sum_{j=1}^n (Y_j)^2 - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2}, \tag{4.8}$$

where the denominator is reduced to the empirical variance of Y. As the functions g and h are here unbounded, Proposition 4.2 does not apply and thus offers no asymptotic information. However, the quantity of interest Y being generally bounded in practice, appropriately truncated versions of g and h could be considered.

4.2.2 A central limit theorem

We establish a CLT for the estimator $\xi_n^{\text{Sobol}}(X_1, Y)$ of the first-order Sobol' index with respect to X_1 (assumed to be real-valued) under some mild assumptions on the model f and the random input X_1 in (2.1). The proof of the theorem is given in Appendix A.2.

Theorem 4.3. *Assume that X_1 is uniformly distributed on $[0,1]$ and f in (2.1) is a twice differentiable function with respect to its first coordinate. Further, we suppose that f and its two first derivatives (with respect to its first coordinate) are bounded. Then*

$$\sqrt{n}\left(\xi_n^{Sobol}(X_1, Y) - S^1\right)$$

is asymptotically Gaussian with zero mean and explicit variance σ^2 given in Appendix A.2.4.

Remark 4.4. *The boundedness of f implies that f has a fourth moment, that is the minimal assumption to get a CLT. Moreover, let us observe that Theorem 4.3 only implies the convergence in probability. Nevertheless, under the assumptions of Theorem 4.3 (f bounded so is Y), Proposition 4.2 applies to derive the almost sure convergence of $\xi_n^{Sobol}(X1, Y)$.*

The assumption on the distribution of X_1 can be relaxed as stated in the following corollary.

Corollary 4.5. *Let F_{X_1} be the cumulative distribution function of X_1. Assume that $f \circ F_{X_1}^{-1}$ is a twice differentiable function such that $f \circ F_{X_1}^{-1}$ and its two first derivatives are bounded. Then the conclusion of Theorem 4.3 still holds.*

Theorem 4.3 and Corollary 4.5 naturally allow to build statistical tests for testing $H_0 : S^1 = 0$ against $H_1 : S^1 \neq 0$. One can note that Chatterjee [28] result allows to test the independence of the input X_1 with respect to the output Y which is a stronger assumption than $S^1 = 0$, this was for example studied in [106]. In addition, our result allows to compute the power of the statistical test against any alternative of the kind $H_{1,0} : S^1 > s_0^1$ for any $s_0^1 > 0$.

Remark 4.6. *A careful reading of the different steps of the proof shows that Theorem 4.3 can be slightly extended to more general situations involving more than two successive order statistics and with more general second variable $(X_2, \ldots, X_p)$. See the forthcoming paper [50].*

The proof of our CLT is a bit long and technical and is postponed to the Appendix A.2. In a nutshell, this proof stands on three main ingredients. First, the regularity assumption on the function f allows to expand the statistic under study as a quadratic functional of the two independent sequences of random variables. The quadratic part for the first sequence involves order statistics of the uniform distribution and may be linearized. The second ingredient is the distribution representation of uniform order statistics by ratios of exponential convolution. The third ingredient is less classical and involves a conditional trick to show a central limit theorem for an empirical mean of a product. Let sketch the idea on a simple example. Let $(\xi_n)_n$ and $(\delta_n)_n$ be two independent sequences of centered square integrable random variables. We set $M_n = n^{-1/2}\sum_{j=1}^n \xi_j \delta_j$ and let $\mathcal{T}$ be the σ-field generated by the sequence (δ_n). Of course, the classical CLT gives that M_n converges in distribution towards a centered Gaussian distribution with variance $\text{Var}(\xi_1)\text{Var}(\delta_1)$. A less classical proof of this result consists in showing that, a.s., conditionally to $\mathcal{T}$ the same convergence in distribution holds. Indeed, this last result follows directly from the Lindeberg CLT and the strong law of large numbers for $n^{-1}\sum_{j=1}^n \delta_j^2$.

4.2.3 Comparison of the different estimation procedures

The estimator based on rank statistics $\xi_n^{\text{Sobol}}(X_1, Y)$ defined in (4.8) can be compared to the classical Pick-Freeze estimators S_n^1 and T_n^1 given in (2.4) and (2.5) respectively (with $\mathbf{u} = \{1\}$) but also to a sequence of estimators involving the estimators $\widehat{T}_n$ introduced in [34].

Required sample sizes With the rank-based procedure, a unique n-sample of input-output provides consistent and asymptotically normal estimations of the p first-order Sobol' indices (together with consistent and asymptotically normal estimations of the p first-order Cramér-von-Mises indices with no extra cost). In contrast, using the Pick-Freeze estimation, if one wants to estimate all the p first-order Sobol' indices and the p Cramér-von-Mises indices, $(p+2)n$ calls of the computer code are required. The number of calls grows linearly with respect to the number of input parameters. This is a practical issue for large input dimension domains. A second drawback of the Pick-Freeze estimation scheme comes from the need of the particular Pick-Freeze design that is not always available.

Limiting variances Since the empirical mean and variance are already known to be asymptotically efficient in the statistical sense[1] to estimate the expectation and the variance of the output, we restrict our study to the comparison of the limiting variances obtained via the Pick-Freeze and the rank-based procedures in the estimation of $\mathbb{E}[\mathbb{E}[Y|X_1]^2]$ only.

In view of the proof of [62, Proposition 2.2], the Pick-Freeze limiting variance obtained using both S_n^1 and T_n^1 in estimating $\mathbb{E}[\mathbb{E}[Y|X_1]^2] = \mathbb{E}[YY^1]$ is simply given by $\text{Var}(YY^1)$, where $Y^1 = f(X_1, W^1)$ is the Pick-Freeze version of $Y = f(X_1, X_2, \ldots, X_p) = f(X_1, W)$.

Using the above Lemmas A.1 and A.2 together with (A.19) leads to the rank-based limiting variance obtained using $\xi_n^{\text{Sobol}}(X_1, Y)$:

$$\begin{aligned}\Sigma_B^{1,1} + \Sigma_C^{1,1} = \mathbb{E}\left[\text{Var}\left(YY^1|X_1\right)\right] + \mathbb{E}\left[\text{Cov}\left(YY^1, YY^{11}|X_1\right)\right] - \mathbb{E}[(Y+Y^1)f_x(X_1,W)X_1]^2 \\ + \mathbb{E}[(Y+Y^1)(\tilde{Y}+\tilde{Y}^1)f_x(X_1,W)f_x(\tilde{X}_1,\tilde{W})(X_1 \wedge \tilde{X}_1)],\end{aligned} \tag{4.9}$$

where $Y = f(X_1, X_2, \ldots, X_p) = f(X_1, W)$, $Y^1 = f(X_1, W^1)$, $Y^{11} = f(X_1, W^{11})$, $\tilde{Y} = f(\tilde{X}_1, \tilde{W})$, and $\tilde{Y}^1 = f(\tilde{X}_1, \tilde{W}^1)$ with X_1 and $\tilde{X}_1$ i.i.d., W, $\tilde{W}$, W^1, and W^{11} i.i.d. also independent of X_1 and $\tilde{X}_1$. Note that Y^1 and Y^{11} (respectively $\tilde{Y}^1$) are Pick-Freeze versions of Y (resp. $\tilde{Y}$). The paragraph's aim is to compare the limiting variances obtained by the two methods (Pick-Freeze and rank-based).

To do so, we recall that the Pick-Freeze experiment requires $n(p+1)$ observations (or computations of the black-box code) to estimate the p first-order Sobol' indices. In order to have a fair comparison of both estimation methods, we then consider that we have $n(p+1)$ i.i.d. observations of Y given by model (2.1) to estimate the p first-order Sobol' indices using the rank statistics. With $n(p+1)$ observations instead of n, the asymptotic variance obtained using the rank-based methodology is divided by $(p+1)$, so that we want to compare

$$V_{\text{PF}} := (p+1)(\text{Var}(YY^1), \ldots, \text{Var}(YY^p))^\top \text{ to } V_{\text{Rank}} := (\Sigma_B^{1,1} + \Sigma_C^{1,1}, \ldots, \Sigma_B^{p,p} + \Sigma_C^{p,p})^\top$$

where Y^i is the Pick-Freeze version of Y with respect to X_i (for $i = 2, \ldots, p$) and $\Sigma_B^{i,i} + \Sigma_C^{i,i}$ has the same expression as $\Sigma_B^{1,1} + \Sigma_C^{1,1}$ in (4.9) replacing the superscripts and the subscripts 1 by i (for $i = 2, \ldots, p$).

Example. We consider the following linear model

$$Y = f(X_1, \ldots, X_p) = \alpha X_1 + X_2 + \ldots + X_p, \tag{4.10}$$

where $\alpha > 0$ is a fixed constant, X_1, X_2, …, and X_p are p independent and uniformly distributed random variables on $[0,1]$.

We denote by $m_{1,p}$ and $m_{2,p}$ the two first moments of $Z_p := X_2 + \ldots + X_p$ and $m_{1,p,\alpha}$ and $m_{2,p,\alpha}$ the two first moments of $Z_{p,\alpha} := \alpha X_1 + X_3 + \ldots + X_p$. In addition, let v_p and $v_{p,\alpha}$ be the variances of Z_p of $Z_{p,\alpha}$. Hence $v_p = m_{2,p} - m_{1,p}^2$, $v_{p,\alpha} = m_{2,p,\alpha} - m_{1,p,\alpha}^2$,

$$\begin{aligned}&m_{1,p} = \frac{1}{2}(p-1), \quad m_{2,p} = \frac{1}{12}(p-1)(3p-2), \quad m_{1,p,\alpha} = \frac{1}{2}(\alpha + m_{1,p-1}) = \frac{1}{2}(\alpha + p - 2), \\ &m_{2,p,\alpha} = \frac{1}{3}\alpha^2 + \alpha m_{1,p-1} + m_{2,p-1} = \frac{1}{3}\alpha^2 + \frac{1}{2}(p-2)\alpha + \frac{1}{12}(p-2)(3p-5).\end{aligned}$$

[1] The reader is referred to [122, Section 25] for the definition of the asymptotic efficiency and related results.

By symmetry, after obvious computations, one gets, for $i = 2, \ldots, p$,

$$\mathrm{Var}(YY^1) = \frac{4}{45}\alpha^4 + \frac{1}{3}m_{1,p}\alpha^3 + \frac{1}{3}\Big(2v_p + m_{1,p}^2\Big)\alpha^2 + 2m_{1,p}v_p\alpha + v_p(v_p + 2m_{1,p}^2),$$
$$\mathrm{Var}(YY^i) = \frac{4}{45} + \frac{1}{3}m_{1,p,\alpha} + \frac{1}{3}\Big(2v_{p,\alpha} + m_{1,p,\alpha}^2\Big) + 2m_{1,p,\alpha}v_{p,\alpha} + v_{p,\alpha}(v_{p,\alpha} + 2m_{1,p,\alpha}^2)$$

while

$$V_{\mathrm{Rank}}^1 = \frac{4}{45}\alpha^4 + \frac{1}{3}m_{1,p}\alpha^3 + \frac{1}{3}\Big(4v_p + m_{1,p}^2\Big)\alpha^2 + 4m_{1,p}v_p\alpha + v_p\Big(v_p + 4m_{1,p}^2\Big),$$
$$V_{\mathrm{Rank}}^i = \frac{4}{45} + \frac{1}{3}m_{1,p,\alpha} + \frac{1}{3}\Big(4v_{p,\alpha} + m_{1,p,\alpha}^2\Big) + 4m_{1,p,\alpha}v_{p,\alpha} + v_{p,\alpha}\Big(v_{p,\alpha} + 4m_{1,p,\alpha}^2\Big).$$

We compare these limiting variances in Figures 4.1 and 4.2. The results are clear and illustrate the fact that the rank-based methodology works much better for all value of $p \geqslant 2$. In addition, the more the value of p increases the greater the gain, as expected.

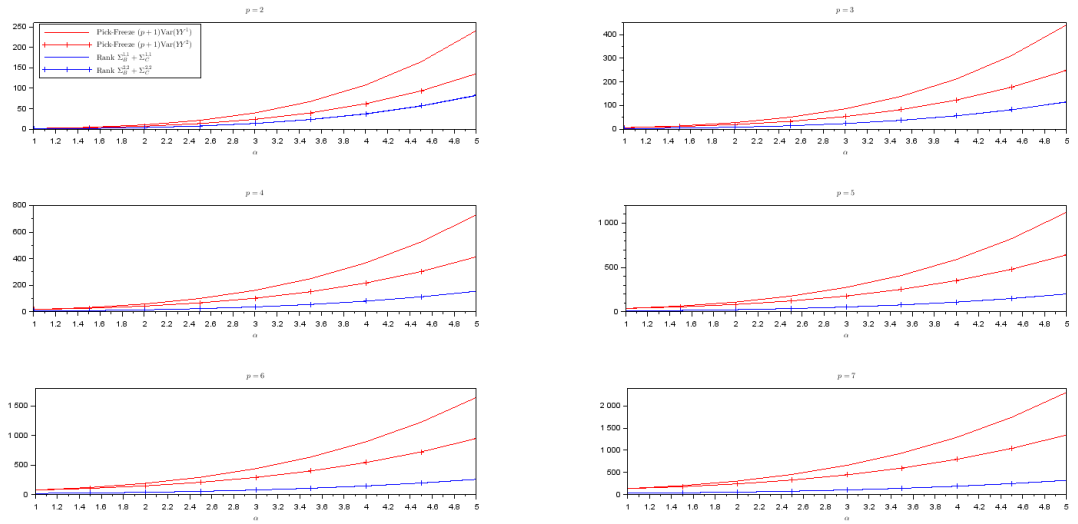


Figure 4.1: Linear model defined in (4.10). The limiting variances with respect to X_1 (plain lines) and to X_2 (plain lines with +) are plotted. The rank-based estimation procedure is represented in blue while the Pick-Freeze estimation procedure is represented in red. As explained, the Pick-Freeze estimation procedure has been weighted by $(p+1)$ to have a fair comparison. The number of variables involved in the model varies from $p = 2$ to $p = 7$.

Remark 4.7. *Observe that a more precise comparison should consists in comparing (via definite-positiveness) the limiting covariance-variance matrices involving both the limiting variances and the limiting covariances. If it is straightforward to compute the covariance terms for the Pick-Freeze methodology: for $i = 2, \ldots, p$,*

$$\mathrm{Cov}(YY^1, YY^i) = \frac{1}{24}\alpha^4 + \frac{1}{12}m_{1,p-1}\alpha^3 + \Big(\frac{7}{144} + \frac{1}{4}v_{p-1} + \frac{1}{6}\Big(m_{1,p-1} + \frac{1}{2}\Big)^2\Big)\alpha^2$$
$$+ \Big(\frac{1}{8} + \frac{1}{12}m_{1,p-1} + \frac{1}{2}v_{p-1} + v_{p-1}m_{1,p-1}\Big)\alpha + v_{p-1}\Big(m_{1,p-1} + \frac{1}{2}\Big)^2,$$

it is much more tricky to deal with the rank-based procedure. Indeed, to do so a joint CLT is required for the vector of all p first-order Sobol' indices whose proof is not a direct generalization of the proof of Theorem 4.3. Such an extension will be done in a forthcoming paper.

Asymptotic efficiency The two previous procedures do not rely on the same design of experiment so that it is not possible to determine which one is the more efficient in the sense of [122, Section 25].
By [48, Proposition 2.5], the sequence of estimators $(T_n^1)_n$ is asymptotically efficient to estimate S^1 when the distribution P of (Y, Y^1) belongs to $\mathscr{P}$, the set of all c.d.f. of exchangeable random vectors in $L^2(\mathbb{R}^2)$.
Using a unique n-sample, one may compare the rank-based estimators introduced in this paper and the procedure involving the estimators $\widehat{T}_n$ defined in [34, page 11]. Such estimator is particularly tricky to compute and not easily tractable in practice. More precisely, the initial n-sample is split into two samples of sizes n_1 and $n_2 = n - n_1$. The first

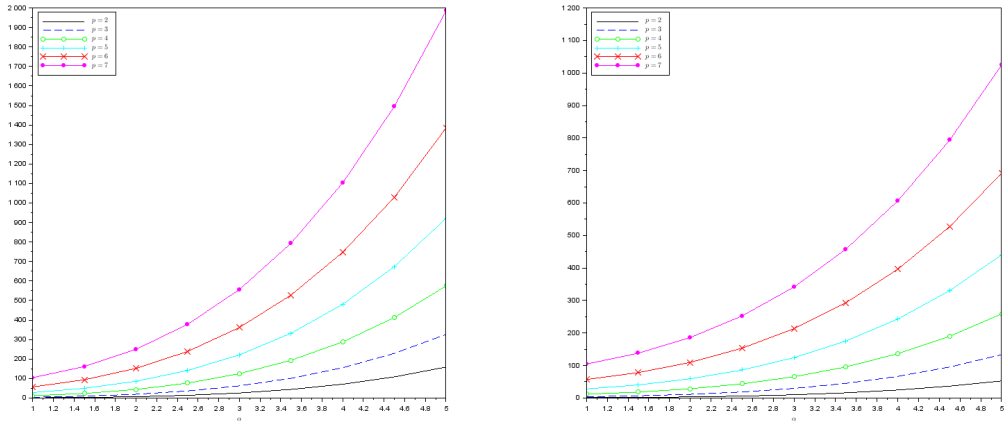


Figure 4.2: Linear model defined in (4.10). The difference between the limiting variances with respect to X_1 (left panel) and to X_2 (right panel) are plotted. As explained, the Pick-Freeze estimation procedure has been weighted by $(p+1)$ to have a fair comparison. The number of variables involved in the model varies from $p=2$ to $p=7$.

sample is dedicated to the estimation of the joint density of (X,Y) while the second one is used to compute a Monte-Carlo estimation of the integral involved in the quantity of interest. In a work under progress [52], another estimator based on kernels and the same design of experiment is proposed. This estimator is more tractable in practice.
By [34, Theorems 3.4 and 3.5], the sequence of estimators $(\widehat{T}_n)_n$ is asymptotically efficient to estimate $\mathbb{E}[\mathbb{E}[Y|X]^2]$ leading to an asymptotically efficient sequence of estimators of S^1. The proof of the following proposition has been postponed in Appendix A.3.

Proposition 4.8. *Consider the sequence of estimators $\widehat{T}_n$ introduced in [34, page 11]. Assume that the joint distribution P of (X,Y) is absolutely continuous with respect to the product probability $P_X \otimes P_Y$, namely $P(dx,dy) = f(x,y)P_X(dx)P_Y(dy)$. Then the sequence $(R^1_n)_n$*

$$R^1_n = \frac{\widehat{T}_n - \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)^2}{\frac{1}{n}\sum_{i=1}^n Y_i^2 - \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)^2}$$

is asymptotically efficient in estimating S^1. In addition, its (minimal) variance $\sigma^2_{\min}$ is

$$\sigma^2_{\min} := \frac{1}{\mathrm{Var}(Y)^2}\mathrm{Var}\left(2\mathbb{E}[Y](1-S^1)Y + S^1Y^2 + \mathbb{E}[Y|X](\mathbb{E}[Y|X]-2Y)\right).$$

Thus we are interested in the comparison of $\sigma^2_{\min}$ and σ^2 given in Theorem 4.3. Let us consider again the example of the linear model (4.10) introduced in the previous paragraph.
Example (continued). We consider the model defined in (4.10). As done in the previous paragraph, we only compare $V^1_{\mathrm{Eff}} := \mathrm{Var}(\mathbb{E}[Y|X_1](2Y-\mathbb{E}[Y|X_1]))$ to $\Sigma^{1,1}_B + \Sigma^{1,1}_C$ and $V^i_{\mathrm{Eff}} := \mathrm{Var}(\mathbb{E}[Y|X_i](2Y-\mathbb{E}[Y|X_i]))$ to $\Sigma^{i,i}_B + \Sigma^{i,i}_C$ for $i=2,\dots,p$. After some trivial computations, one gets

$$V^1_{\mathrm{Eff}} = \frac{4}{45}\alpha^4 + \frac{1}{3}m_{1,p}\alpha^3 + \frac{1}{3}\left(4v_p + m^2_{1,p}\right)\alpha^2 + 4m_{1,p}v_p\alpha + 4v_p m^2_{1,p},$$
$$V^i_{\mathrm{Eff}} = \frac{4}{45} + \frac{1}{3}m_{1,p,\alpha} + \frac{1}{3}\left(4v_{p,\alpha} + m^2_{1,p,\alpha}\right) + 4m_{1,p,\alpha}v_{p,\alpha} + 4v_{p,\alpha}m^2_{1,p,\alpha}.$$

We compare these limiting variances in Figure 4.3. We observe that the limiting variances obtained with the rank methodology do not differ much from the efficient variances.

4.2.4 Recovering other classical indices

In [44], the authors considered computer codes of the form (2.1) valued on a compact Riemannian manifold. In this framework, they proposed a sensitivity index in the flavour of the Cramé-von-Mises index and they used the Pick-Freeze scheme to provide a consistent estimator. The authors of [51] extend the previous indices to the context of

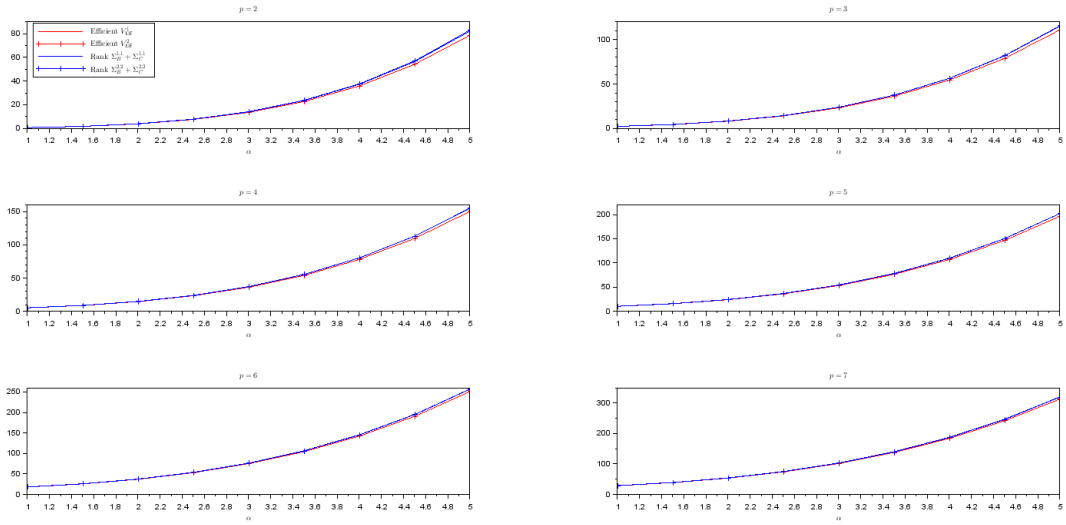


Figure 4.3: Linear model defined in (4.10). The limiting variances with respect to X_1 (plain lines) and to X_2 (plain lines with +) are plotted. The rank-based estimation procedure is represented in blue while the efficient variances are represented in red. The number of variables involved in the model varies from $p = 2$ to $p = 7$.

general metric spaces and propose U-statistics-based estimators improving the classical Pick-Freeze procedure. In light of Section 4.1.1, one may introduce a novel estimation of the indices introduced in [51] requiring a unique n-sample. The reader is referred to [42] for more details on the procedure.
Following [92, 93], extensions to Sobol' indices are obtained by replacing their numerator by higher-order moments. In [49], the authors construct a Pick-Freeze estimator for such extensions. One again, we are now able to propose another estimation scheme based on a unique n-sample. The reader is referred to [50] for the generalization of Lemma 4.1 and the corresponding asymptotic study.

Chapter 5

Beyond variance-based indices: the Cramér-von Mises index and its extensions

As pointed out before, Sobol' indices are based on L^2 decomposition. As a matter of fact, Sobol' indices are well adapted to measure the contribution of an input on the deviation around the mean of Y. We begin by considering two examples that enlighten the need for more general indices in a sense to be defined later.

Example 5.1. *Let X_1 and X_2 be two independent random variables with distinct distributions sharing the four first moments. Consider*

$$Y = X_1 + X_2 + X_1^2 X_2^2.$$

Then

$$\begin{aligned}\mathrm{Var}\left(\mathbb{E}\left[Y|X_1\right]\right) &= \mathrm{Var}(X_1 + X_1^2\mathbb{E}\left[X_2^2\right]) \\ &= \mathrm{Var}(X_2 + X_2^2\mathbb{E}\left[X_1^2\right]) = \mathrm{Var}\left(\mathbb{E}\left[Y|X_2\right]\right)\end{aligned}$$

so that $S^1 = S^2$. Nevertheless, Y is a symmetrical function of X_1, X_2 but if X_1 and X_2 have different distributions, X_1 and X_2 should act differently.

It seems important to consider sensitivity indices that take into account not only the two first moments but eventually on the whole distribution. it leads us to consider moment-independent importance measures.

5.1 A first approach to go beyond variance-based indices: higher order moment-based indices

Using the classical Hoeffding decomposition, for a singleton $v \in I_d$, the numerator of the classical Sobol index with respect to v is given by

$$H_v^2 = \mathbb{E}\left[\left(\mathbb{E}[Y|X^{\mathbf{v}}] - \mathbb{E}[Y]\right)^2\right]. \tag{5.1}$$

Following [93] and [92], we generalize this quantity by considering higher order moments. Indeed, for any integer $p \geqslant 2$, we set

$$H_v^p := \mathbb{E}\left[\left(\mathbb{E}[Y|X^{\mathbf{v}}] - \mathbb{E}[Y]\right)^p\right]. \tag{5.2}$$

$H_v = H_v^2$. The following lemma gives the Pick and Freeze representation of H_v^p for $p \geqslant 2$.

Lemma 5.2. *For any $v \in I_d$, one has*

$$\mathbb{E}\left[\left(\mathbb{E}[Y|X^{\mathbf{v}}] - \mathbb{E}[Y]\right)^p\right] = \mathbb{E}\left[\prod_{i=1}^{p}\left(Y^{v,i} - \mathbb{E}[Y]\right)\right]. \tag{5.3}$$

Here, $Y^{v,1} = Y$ and for $i = 2,\ldots,p$, $Y^{v,i}$ is constructed independently as Y^v defined in equation (5.18).

Obviously, H_v^p is non negative for even p and

$$\left|H_v^p\right| \leqslant \mathbb{E}\left[|Y-\mathbb{E}[Y]|^p\right].$$

Further, H_v^p is invariant by any translation of the output.

Estimation procedure In view of the estimation of H_v^p, we first expand the product in the right-hand side of (5.3) to get that

$$H_v^p = \sum_{l=0}^{p} \binom{p}{l} (-1)^{p-l} \mathbb{E}[Y]^{p-l} \mathbb{E}\left[\prod_{i=1}^{l} Y^{v,i}\right].$$

with the usual convention $\prod_{i=1}^{0} Y^{v,i} = 1$. Second, we use a Monte Carlo scheme and consider the following Pick and Freeze design constituted by the following $p \times N$-sample

$$\left(Y_j^{v,i}\right)_{(i,j)\in I_p \times I_N}.$$

We define for any any $N \in \mathbb{N}^*$, $j \in I_N$ and $l \in I_p$,

$$P_{l,j}^v = \binom{p}{l}^{-1} \sum_{k_1<\ldots<k_l \in I_p} \left(\prod_{i=1}^{l} Y_j^{v,k_i}\right) \quad \text{and} \quad \overline{P}_l^v = \frac{1}{N}\sum_{j=1}^{N} P_{l,j}^v.$$

The Monte Carlo estimator is then

$$H_{p,N}^v = \sum_{l=0}^{p} \binom{p}{l} (-1)^{p-l} \left(\overline{P}_1^v\right)^{p-l} \overline{P}_l^v. \tag{5.4}$$

Notice that we generalize the estimation procedure of [48] and use all the available information by considering the means over the set of indices $k_1, \ldots, k_l \in I_d$, $k_n \neq k_m$. The following theorem provides asymptotic properties of $H_{p,N}^v$.

Theorem 5.3. *$H_{p,N}^v$ is consistent and asymptotically Gaussian:*

$$\sqrt{N}\left(H_{p,N}^v - H_p^v\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}\left(0,\sigma^2\right) \tag{5.5}$$

where

$$\sigma^2 = p\left[\mathrm{Var}(Y) + (p-1)\mathrm{Cov}(Y, Y^{v,2})\right]\left(\sum_{l=1}^{p} a_l b_l\right)^2,$$

$$a_l = \frac{l}{p}\mathbb{E}[Y]^{l-1}, \qquad l = 1,\ldots,p$$

$$b_1 = (-1)^{p-1} p(p-1)\mathbb{E}[Y]^{p-1} + \sum_{l=2}^{p-1} \binom{p}{l} (-1)^{p-l}(p-l)\mathbb{E}[Y]^{p-l-1}\mathbb{E}\left[\prod_{i=1}^{l} Y^{v,i}\right]$$

and

$$b_l = \binom{p}{l}(-1)^{p-l}\mathbb{E}[Y]^{p-l}, \qquad l = 1,\ldots,p.$$

Proof of Theorem 5.3. The consistency follows from a straightforward application of the strong law of large numbers. The asymptotic normality is derived by two successive applications of the delta method [122] .

(1) Let $W_j^1 = (Y_j^{v,1},\ldots,Y_j^{v,p})^T$ $(j = 1,\ldots,N)$ and g^1 the mapping from $\mathbb{R}^p$ to $\mathbb{R}^p$ whose l-th coordinate is given by

$$g_l^1(x_1,\ldots,x_p) = \binom{p}{l}^{-1} \sum_{\substack{k_1<\ldots<k_l \\ k_i \in I_p, i=1,\ldots,l}} \left(\prod_{i=1}^{l} x_{k_i}\right).$$

Let Σ^1 be the covariance matrix of W_j^1. Clearly, one has $\Sigma^1_{ii} = \mathrm{Var}(Y)$ for $i \in I_p$ while $\Sigma^1_{ij} = \mathrm{Cov}(Y^{v,i}, Y^{v,j}) = \mathrm{Cov}(Y, Y^{v,2})$. The multidimensional central limit theorem gives with $m = (\mathbb{E}[Y], \dots, \mathbb{E}[Y])^T$

$$\sqrt{N}\left(\frac{1}{N}\sum_{j=1}^{N} W_j^1 - m\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}_p\left(0, \Sigma^1\right).$$

We then apply the so-called delta method to W^1 and g^1 so that

$$\sqrt{N}\left(g^1\left(\overline{W}_N^1\right) - g^1\left(\mathbb{E}\left[W^1\right]\right)\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}\left(0, J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)\Sigma^1 J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)^T\right)$$

with $J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)$ the Jacobian of g^1 at point $\mathbb{E}\left[W^1\right]$. Notice that for $i \in I_p$ and $k \in I_p$,

$$\frac{\partial g_l^1}{\partial x_k}\left(\mathbb{E}\left[W^1\right]\right) = \frac{\binom{p-1}{l-1}}{\binom{p}{l}} m^{l-1} = \frac{l}{p}\mathbb{E}[Y]^{l-1} =: a_l.$$

Thus $\Sigma^2 := J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)\Sigma^1 J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)^T$ is given by

$$\Sigma^2_{ij} = p a_i a_j \left(\Sigma^1_{11} + (p-1)\Sigma^1_{12}\right).$$

(2) Now consider $W_j^2 = (P_j^{v,1}, \dots P_j^{v,p})^T$ $(j = 1, \dots, N)$ and g^2 the mapping from $\mathbb{R}^p$ to $\mathbb{R}$ defined by

$$g^2(y_1, \dots, y_p) = \sum_{l=0}^{p} \binom{p}{l} (-1)^{p-l} y_1^{p-l} y_l.$$

We apply once again the delta method to W^2 so that

$$\sqrt{N}\left(g^2\left(\overline{W}_N^2\right) - g^2\left(\mathbb{E}\left[W^2\right]\right)\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}\left(0, J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)\Sigma^2 J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)^T\right)$$

with $J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)$ the Jacobian of g^2 at point $\mathbb{E}\left[W^2\right]$. Notice that for $k \in I_p$,

$$\begin{aligned}\frac{\partial g^2}{\partial y_1}\left(\mathbb{E}\left[W^2\right]\right) &= (-1)^{p-1} p(p-1)\mathbb{E}[Y]^{p-1} \\ &\quad + \sum_{l=2}^{p-1} \binom{p}{l} (-1)^{p-l}(p-l)\mathbb{E}[Y]^{p-l-1}\mathbb{E}\left[\prod_{i=1}^{l} Y^{v,i}\right]\end{aligned}$$

and

$$\frac{\partial g^2}{\partial y_l}\left(\mathbb{E}\left[W^2\right]\right) = \binom{p}{l}(-1)^{p-l}\mathbb{E}[Y]^{p-l}.$$

Thus the limiting variance is

$$\sigma^2 := J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)\Sigma^2 J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)^T = p\left(\Sigma^1_{11} + (p-1)\Sigma^1_{12}\right)\left(\sum_{i=1}^{p} a_i b_i\right)^2,$$

where b_i is the i-th coordinate of $\nabla g^2\left(\mathbb{E}\left[W^2\right]\right)$. □

5.2 Distribution-based indices: Cramér-von Mises indices

In this section, we consider indices based on the whole distribution: the Cramér-von Mises indices that are based on the conditional distribution of the output and requires only $3 \times N$.

5.2.1 Definition and properties of the Cramér-von Mises indices

The code will be denoted by $Z = f(X_1, \dots, X_d) \in \mathbb{R}^k$. Let F be the distribution function of Z. For any $t = (t_1, \dots, t_k) \in \mathbb{R}^k$,

$$F(t) = \mathbb{P}\left(Z \leqslant t\right) = \mathbb{E}\left[\mathbb{1}_{\{Z \leqslant t\}}\right]$$

and $F^{\mathrm{u}}(t)$ the conditional distribution function of Z conditionally on X_{u}:

$$F^{\mathrm{u}}(t) = \mathbb{P}\left(Z \leqslant t | X_{\mathrm{u}},\right) = \mathbb{E}\left[\mathbb{1}_{\{Z\leqslant t\}} | X_{\mathrm{u}}\right].$$

Notice that $\{Z \leqslant t\}$ means that $\{Z_1 \leqslant t_1, \dots, Z_k \leqslant t_k\}$. Obviously, $\mathbb{E}\left[F^{\mathrm{u}}(t)\right] = F(t)$. Now, we apply the previous framework with $Y(t) = \mathbb{1}_{\{Z\leqslant t\}}$ and $p = 2$. Since for any fixed $t \in \mathbb{R}^k$, $Y(t)$ is a real-valued random variable, we apply the framework presented in Chapter 2. More precisely, for any $v \in I_p$ let $\sim v$ be $I_p \setminus \{v\}$ and we first perform the Hoeffding decomposition of $Y(t)$:

$$Y(t) = \mathbb{1}_{\{Z\leqslant t\}} = \mathbb{E}[Y(t)] + (\mathbb{E}[Y(t)|X_v] - \mathbb{E}[Y(t)]) + (\mathbb{E}[Y(t)|X_{\sim v}] - \mathbb{E}[Y(t)]) + R(t,v), \tag{5.6}$$

where

$$R(t,v) = Y(t) - \mathbb{E}[Y(t)] - (\mathbb{E}[Y(t)|X_v] - \mathbb{E}[Y(t)]) - (\mathbb{E}[Y(t)|X_{\sim v}] - \mathbb{E}[Y(t)]).$$

As usually done, we compute the variance of both sides of (5.6) which leads to

$$\begin{aligned}
\mathrm{Var}(Y(t)) &= F(t)(1-F(t)) \\
&= \mathrm{Var}\left(\mathbb{E}[Y(t)|X_v] - \mathbb{E}[Y(t)]\right) + \mathrm{Var}\left(\mathbb{E}[Y(t)|X_{\sim v}] - \mathbb{E}[Y(t)]\right) + \mathrm{Var}(R(t,v)) \\
&= \mathrm{Var}\left(F^v(t)\right) + \mathrm{Var}\left(F^{\sim v}(t)\right) + \mathrm{Var}(R(t,v)) \\
&= \mathbb{E}\left[\left(F^v(t) - F(t)\right)^2\right] + \mathbb{E}\left[\left(F^{\sim v}(t) - F(t)\right)^2\right] + \mathrm{Var}(R(t,v))
\end{aligned} \tag{5.7}$$

by the decorrelation of the different terms involved in the Hoeffding decomposition.

Remark 5.4. *A straightforward application of the results of Chapter 2 provides for any fixed $t \in \mathbb{R}^k$ a consistent and asymptotically normal procedure for the estimation of*

$$\mathbb{E}\left[\left(F^v(t) - F(t)\right)^2\right] = \mathrm{Var}\left(F^v(t)\right) \quad \textit{and} \quad \mathbb{E}\left[\left(F^{\sim v}(t) - F(t)\right)^2\right] = \mathrm{Var}\left(F^{\sim v}(t)\right).$$

Now we integrate the terms in (5.7) in $t \in \mathbb{R}^k$ with respect to the distribution of Z:

$$\begin{aligned}
&\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t) \\
&= \int_{\mathbb{R}^k} \mathbb{E}\left[\left(F^v(t) - F(t)\right)^2\right] dF(t) + \int_{\mathbb{R}^k} \mathbb{E}\left[\left(F^{\sim v}(t) - F(t)\right)^2\right] dF(t) + \int_{\mathbb{R}^k} \mathrm{Var}(R(t,v))dF(t)
\end{aligned} \tag{5.8}$$

This integration has to be understood in the Riemmann-Stieltjes sense (see, e.g., [119]). Notice that the first term in the right hand side of (5.8) represents a Cramér-von Mises-type distance of order 2 between the distribution $\mathscr{L}(Z)$ of Z and the distribution $\mathscr{L}(Z|X_v)$ of Z given X_v.

Following the classical way of defining Sobol indices, we normalize the previous equation by

$$\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t)$$

leading to

$$1 = \frac{\int_{\mathbb{R}^k} \mathbb{E}\left[\left(F^v(t) - F(t)\right)^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t)} + \frac{\int_{\mathbb{R}^k} \mathbb{E}\left[\left(F^{\sim v}(t) - F(t)\right)^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t)} + \frac{\int_{\mathbb{R}^k} \mathrm{Var}(R(t,v))dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t)}. \tag{5.9}$$

Then we define the Cramér-von Mises indices with respect to v and $\sim v$ by

$$S^v_{2,CVM} := \frac{\int_{\mathbb{R}^k} \mathbb{E}\left[\left(F(t) - F^v(t)\right)^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t)} \quad \text{and} \quad S^{\sim v}_{2,CVM} := \frac{\int_{\mathbb{R}^k} \mathbb{E}\left[\left(F(t) - F^{\sim v}(t)\right)^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t)}. \tag{5.10}$$

Properties 5.5. *These new indices are naturally adapted to multivariate outputs and they share the same properties as the classical Sobol index, namely,*

1. *as seen in (5.9), the different contributions sum to 1;*
2. *they are invariant by translation, by any isometry, and by any nondegenerated scaling of the components of Y.*

Remark 5.6.

1. *We could have defined the following indices instead:*

$$\int_{\mathbb{R}^k} \frac{\mathbb{E}\left[(F(t)-F^{v}(t))^2\right]}{F(t)(1-F(t))} dF(t) \quad \text{and} \quad \int_{\mathbb{R}^k} \frac{\mathbb{E}\left[(F(t)-F^{\sim v}(t))^2\right]}{F(t)(1-F(t))} dF(t),$$

   normalizing by $F(t)(1-F(t))$ (like in the Anderson-Darling statistic) before the integration phase. Nevertheless, the previous integrals might not be defined. Moreover, even if the integrals are well-defined, one may encounter numerical explosion during the estimation procedure that might be produced for small and large values of t since the normalizing factor then cancels.

2. *In this paper, we only consider first-order sensitivity indices as well for the classical Sobol indices and for the Cramér-von Mises indices. Anyway, as well as for the Sobol indices, one may define higher order and total Cramér-von Mises indices. The construction of the former is straightforward taking v no longer a singleton. For example, if one is interested in the second order Cramér-von Mises index with respect to the first and second inputs, it suffices to take $v = \{1,2\}$. Concerning the latter, the total Cramér-von Mises index $S^{Tot,v}_{2,CVM}$ with repect to v is defined by*

$$S^{Tot,v}_{2,CVM} := 1 - S^{\sim v}_{2,CVM} = 1 - \frac{\int_{\mathbb{R}^k} \mathbb{E}\left[(F(t)-F^{\sim v}(t))^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t)) dF(t)}.$$

3. *To use the Hoeffding decomposition, the inputs are required to be independent. Howver, one can compute the Cramér-von Mises index when the inputs are dependent. Nevertheless, these are then difficult to interpret.*

5.2.2 Pick-Freeze estimation of the Cramér-von Mises indices and asymptotic properties

This section is dedicated to the estimation of $S^{v}_{2,CVM}$ (and $S^{\sim v}_{2,CVM}$). One has to estimate both the numerator and the denominator of the indices. Nevertheless, when the output Z has independent coordinates that are absolutely continuous with respect to the Lebesgue measure, we have

$$\int_{\mathbb{R}^k} F(t)(1-F(t)) dF(t) = \mathbb{E}[F(Z)(1-F(Z)] = \frac{1}{2^k} - \frac{1}{3^k}.$$

Thus the normalizing factor reduces to $\frac{1}{2^k} - \frac{1}{3^k}$. As a consequence, we propose two versions of central limit theorems: the first one deals with the numerator's estimator and can be applied when the output Z has independent coordinates that are absolutely continuous with respect to the Lebesgue measure whereas the second one concerns the general estimator and may apply in many other cases.

We denote the numerator of $S^{v}_{2,CVM}$ by $N^{v}_{2,CVM}$ defined as the Cramér-von Mises type distance of order 2 between $\mathscr{L}(Z)$ and $\mathscr{L}(Z|X_{\mathrm{u}})$. Notice that it can be rewritten as

$$N^{v}_{2,CVM} = \mathbb{E}_{\tilde{Z}}\left[\mathbb{E}_{X_v}\left[\left(F(\tilde{Z}) - F^{v}(\tilde{Z})\right)^2\right]\right],$$

where $\tilde{Z}$ is an independent copy of Z.

Numerator estimation of the Cramér-von Mises indices

Then we proceed to a double Monte Carlo scheme for the estimation of $N^{v}_{2,CVM}$ and consider the following design of experiment consisting in:

1. the classical pick and freeze sample, that is, two N-samples of Z: $(Z^{v,1}_j, Z^{v,2}_j)$, $1 \leqslant j \leqslant N$;
2. a third N-sample of Z independent of $(Z^{v,1}_j, Z^{v,2}_j)_{1\leqslant j\leqslant N}$: W_k, $1 \leqslant k \leqslant N$.

The empirical estimator of $N^{v}_{2,CVM}$ is then given by

$$\widehat{N}^{v}_{2,CVM} = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z^{v,1}_j \leqslant W_k\}}\mathbb{1}_{\{Z^{v,2}_j \leqslant W_k\}} - \left[\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z^{v,1}_j \leqslant W_k\}} + \mathbb{1}_{\{Z^{v,2}_j \leqslant W_k\}}\right)\right]^2\right\}. \tag{5.11}$$

Now we established the consistency of $\widehat{N}^{v}_{2,CVM}$ that follows directly from an auxiliary lemma (see Appendix B).

Corollary 5.7. *$\widehat{N}^{v}_{2,CVM}$ is strongly consistent as N goes to infinity.*

Now we turn to the asymptotic normality of $\widehat{N}^{\nu}_{2,CVM}$. We follow van der Vaart [122] to establish the following proposition (more precisely Theorems 20.8 and 20.9, Lemma 20.10, and Example 20.11).

Theorem 5.8. *The sequence of estimators $\widehat{N}^{\nu}_{2,CVM}$ is asymptotically Gaussian in estimating $N^{\nu}_{2,CVM}$. That is, $\sqrt{N}\left(\widehat{N}^{\nu}_{2,CVM} - N^{\nu}_{2,CVM}\right)$ converges in distribution towards the centered Gaussian law with a limiting variance ξ^2 whose explicit expression can be found in the proof.*

Remark 5.9. *Thanks to Theorem 5.8, we are now able to provide asymptotic confidence intervals for the estimation of $N^{\nu}_{2,CVM}$. They are of the form $(\widehat{N}^{\nu}_{2,CVM} \pm z_{\alpha}\xi/\sqrt{N})$, where z_{α} is the $1-\alpha/2$ quantile of a standard normal distribution. Unfortunately, the variance ξ^2 is unknown but thanks to its explicit form it is easy to replace it by a consistent estimator $\widehat{\xi}$ and use Slutsky's lemma to have an asymptotic confidence interval.*

Global estimation of the Cramér-von Mises indices

In order to estimate the Cramér-von Mises index $S^{\nu}_{2,CVM}$, it remains to estimate its denominator denoted by $D^{\nu}_{2,CVM}$. Notice that it can be rewritten as

$$D^{\nu}_{2,CVM} = \mathbb{E}\left[F(Z)(1-F(Z))\right]$$

and estimated using the design of experiment already introduced for the estimation of the numerator by

$$\widehat{D}^{\nu}_{2,CVM} = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z_j^{\nu,1}\leqslant W_k\}} + \mathbb{1}_{\{Z_j^{\nu,2}\leqslant W_k\}}\right) - \left[\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z_j^{\nu,1}\leqslant W_k\}} + \mathbb{1}_{\{Z_j^{\nu,2}\leqslant W_k\}}\right)\right]^2\right\}. \tag{5.12}$$

Proceeding as in Subsection 5.2.2, we have the following.

Corollary 5.10. *$\widehat{S}^{\nu}_{2,CVM}$ is strongly consistent as N goes to infinity.*

The following central limit theorem comes from the functional Delta method.

Theorem 5.11. *The sequence of estimators $\widehat{S}^{\nu}_{2,CVM}$ is asymptotically Gaussian in estimating $S^{\nu}_{2,CVM}$. That is, $\sqrt{N}\left(\widehat{S}^{\nu}_{2,CVM} - S^{\nu}_{2,CVM}\right)$ converges in distribution towards the centered Gaussian law with a limiting variance that can be computed.*

5.2.3 Practical advice and general comments on the Cramér-von Mises indices

Practical advice

In a general setting, for all the nice properties of the Cramér-von Mises indices and their easy to implement efficient estimation, we recommend using the Cramér-von Mises indices. As a consequence, considering a sample with the appropriate size, one can estimate one at a time the Cramér-von Mises indices and the Sobol indices. More precisely, if one wants to estimate p Sobol indices a sample size of $(p+1)N$ is required. With only N more output evaluations, we get both the p Sobol indices and the Cramér-von Mises ones. Furthermore, the theoretical theorems provide confidence intervals that controlled the accuracy of the estimations. Moreover, when the practitioner is interested in a specific feature (e.g., mean behavior or quantile) of the output, he should use more suited indices (e.g., the classical Sobol indices for the mean or the indices introduced in [43] for the quantile).

Cramér-von Mises indices versus Sobol indices

Cramér-von Mises and Sobol indices are both based on the Hoeffding decomposition and sum to 1. Nevertheless, the former are based on the whole distribution of the output, in contrast with the latter that are only based on the order two moments. Notice that two variables that have a different influence on the output may have the same Sobol indices (just as two random variables with different distributions can have the same variance). This point represents one limitation of Sobol indices and does not occur with the Cramér-von Mises indices as one can see in Exercise 12.

In contrast, the indices based on the whole distribution partially get rid of such limitations and pathological patterns. However, one can build an example based, e.g., on two input variables that leads to the same indices $S^1_{2,CVM}$ and $S^2_{2,CVM}$ once the integration with respect to t has been done.

In addition, remark that a null value for a Sobol index does not imply that the input is unimportant whereas a null value for a Cramér-von Mises index means that the input is unimportant. Moreover, by definition, a large Cramér-von Mises index means that the input variable under concern has a great influence on the output in regions where the output has a large distribution mass. That is why we advise the practitioner to use them in a general context. Nevertheless, when one is interested in the mean output behavior, the Sobol indices are more adapted. Indeed, as noted in [43], the Sobol indices minimize the contrast associated with the mean. In the same spirit, if one is interested in specific feature of the output (for example, an α-quantile), one should use the index based on the associated contrast. See [43] for more details on the notion of contrast and the results therein.

Cramér-von Mises indices versus moment independent indices

There already exist several moment-independent indices: some of them have been introduced by Borgonovo et *al.* (density-based indices [15], cumulative-distribution-function-based indices [17]). See also [13] for other indices and references therein. More recently, Da Veiga [32] shows that those indices are special cases of a class of sensitivity indices based on the Csizár f-divergence. A lot of classical "distances" between probability measures as, e.g., the Kullback-Leibler divergence, the Hellinger distance and the total variation distance belong to this family of divergences. Other dissimilarity measures exist to compare probability distributions; in particular, integral probability metrics [87]. See Chapter 6 for more details on such indices.

In comparison with the indices defined by (17) in [16], we can notice that the integration is done with respect to the distribution of the output in the former indices while the integration is done with respect to the Lebesgue measure in the latter indices. Our method represents at least two advantages: (i) the index always exist whatever the output distribution; (ii) such an integration weights the support of the output distribution.

Since the space of the probability measures on $\mathbb{R}^k$ is of infinite dimension, the different distances on this space are not equivalent; hence, they are very difficult to compare. Each index is constructed on a specific distance and has its own interest. Despite the fact that the Cramér-von Mises indices have no clear dual formulation, they present the following remarkable advantages. As we will see in the next sections, one can easily estimate them with a low simulation cost that does not depend on the dimension of the output. The sample required for their estimation also provide a Sobol indices estimation. In addition, these estimators are asymptotically normal and converge at the rate $\sqrt{N}$ which allows the practitioner to build confidence intervals.

5.2.4 Application: The Giant Cell Arthritis Problem

Context and goal

In this subsection, we consider the realistic problem of management of suspected giant cell arthritis posed by Bunchbinder and Detsky in [22]. More recently, this problem was also studied by Felli and Hazen [41] and Borgonovo *et al.* [14]. As explained in [22], "giant cell arthritis (GCA) is a vasculitis of unknown etiology that affects large and medium sized vessels and occurs almost exclusively in patients 50 years or older". This disease may lead to severe side effects (loss of visual accuity, fever, headache,...) whereas the risks of not treating it include the threat of blindness and major vessels occlusion. A patient with suspected GCA can receive a therapy based on prednisone. Unfortunately, a treatment with high prednisone doses may cause severe complications. Thus when confronted by a patient with suspected GCA, the clinician must adopt a strategy. There is a considerable literature on sensitivity analysis for these sorts of models, based on the utility of learning a model input before choosing a treatment strategy (see, e.g., [40] and [89]). In [22], the authors considered four different strategies:

A : treat none of the patients;

B : proceed to the biopsy and treat all the positive patients;

C : proceed to the biopsy and treat all the patients whatever their result;

D : treat all the patients.

The clinician wants to adopt the strategy optimizing the patient outcomes measured in terms of utility. The reader is referred to [88] for more details on the concept of utility. The basic idea is that a patient with perfect health is assigned a utility of 1 and the expected utility of the other patients (not perfectly healthy) is calculated subtracting

some "disutilities" from this perfect score of 1. These strategies are represented in Figures 5.1-5.4 with the different inputs involved in the computation of the utilities.

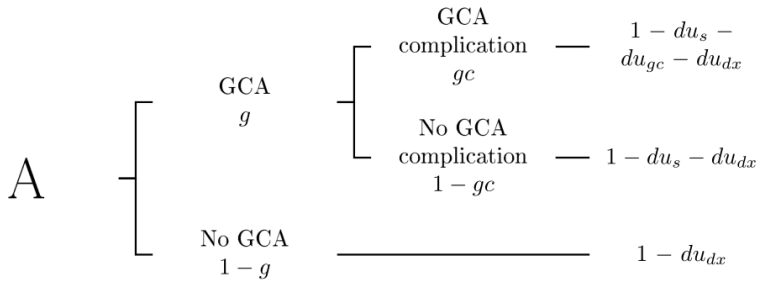


Figure 5.1: The decision tree for the treat none alternative.

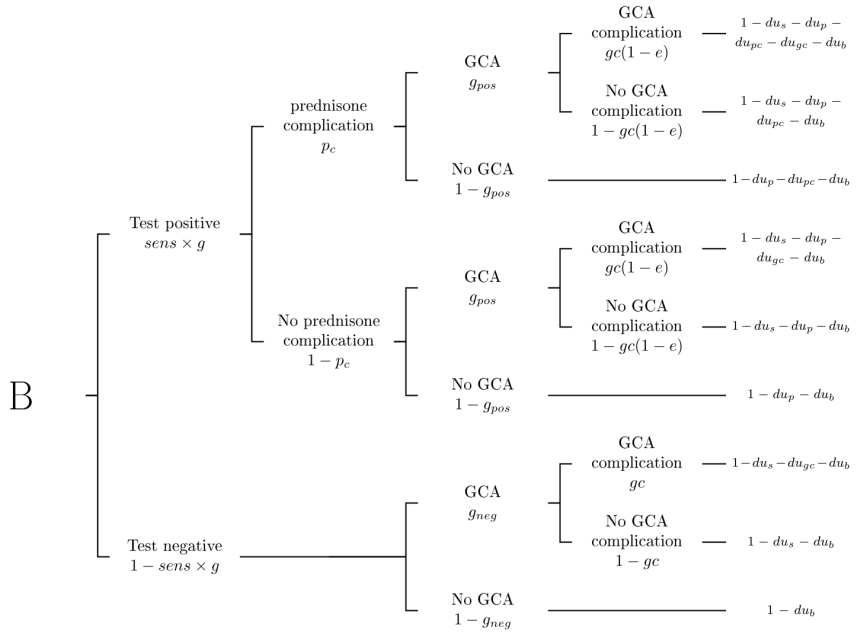


Figure 5.2: The decision tree for the biopsy and the treat positive alternative.

For example, in strategy A (see Figure 5.1), the utility of a patient having GCA and developing severe GCA complications is given by $1 - d_s - du_{gc} - du_{dx}$. His entire subpath is then

$$g \times gc \times (1 - d_s - du_{gc} - du_{dx}).$$

The input parameters and the modelisation of the random ones

As seen in Figures 5.1 to 5.4, the different strategies involve input parameters like, e.g., the proportion g of patients having GCA or the probability gc for a patient to develop severe GCA complications (fixed at 0.8 as done in [22]) or even the disutility associated having GCA symptoms. Table 5.1 summarizes the input parameters involved.

The values $\mathbb{P}[\cdot]$ and $D(\cdot)$ refer, respectively, to the probability of an event and to the disutility associated with an event. The minimum and maximum values m and M depict each parameter's range for the sensitivity analysis. The base values are provided by clinician expertise. The utilities of the different strategies when all the input parameters are set to their base value are summarized in Table 5.2.

The base value of some input parameters are reliable while the others are really uncertain which leads us to consider them as random. As a consequence, if Y_A, Y_B, Y_C, and Y_D represent the outcomes corresponding to the four different strategies A to D, the clinician aims to determine

$$\max\{\mathbb{E}[Y_A], \mathbb{E}[Y_B], \mathbb{E}[Y_C], \mathbb{E}[Y_D]\} \tag{5.13}$$

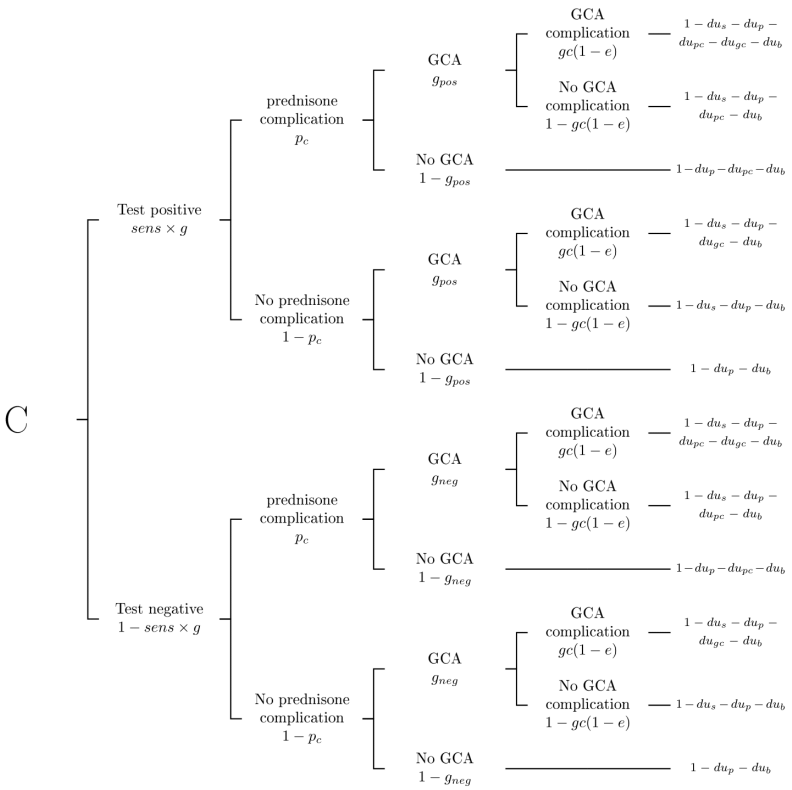


Figure 5.3: The decision tree for the biopsy and the treat all alternative.

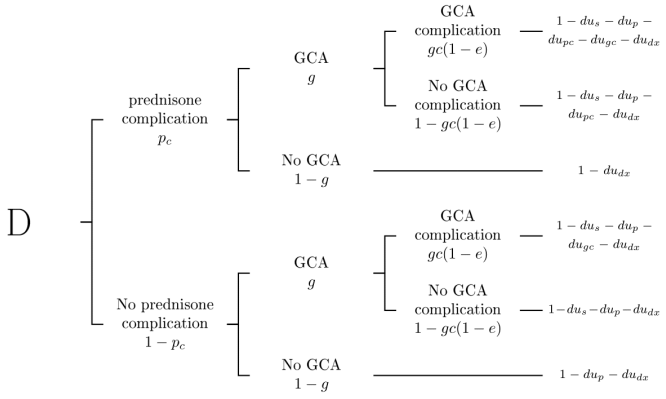


Figure 5.4: The decision tree for the treat all alternative.

with the uncertain model input presented in Table 5.1. A sensitivity analysis is then performed to determine the most influential input variables on the outcome.

As done in [14, 41], all the random inputs will be independently based on beta distributions. The beta density parameters corresponding to each random input are determined by fitting the base value as their mean and capturing 95% of the probability mass in the range defined by the minimum and maximum. The remaining 5% will be equally distributed to either side of this range if possible. Concretely, each random input will be distributed as

$$Z\mathbb{1}_{m\leqslant Z<M} + U\mathbb{1}_{m>Z} + V\mathbb{1}_{Z\geqslant M},$$

where Z, U, and V are independent random variables. Z is beta distributed with parameters (α, β). U and V are uniform random variables on $[0, m]$ and $[M, 1]$, respectively.

Fixed parameters	Symbols	Fixed value				
$\mathbb{P}$[having GCA]	g	0.8	–	–	–	–
D(having symptoms of GCA)	du_s	0.12	–	–	–	–
D(having a temporal artery biopsy)	du_b	0.005	–	–	–	–
D(not knowing the true diagnosis)	du_{dx}	0.025	–	–	–	–
					beta(α,β)	
Uncertain parameters	Symbols	Base	Min. m	Max. M	α	β
$\mathbb{P}$[developing severe complications of GCA]	gc	0.3	0.05	0.5	4.179	11.011
$\mathbb{P}$[developing severe iatrogenic side effects]	pc	0.2	0.05	0.5	2.647	10.589
Efficacy of high dose prednisone	e	0.9	0.8	1	27.787	3.087
Sensitivity of temporal artery biopsy	$sens$	0.83	0.6	1	7.554	1.547
D(major complication from GCA)	du_{gc}	0.8	0.3	0.9	27.454	6.864
D(prednisone therapy)	du_p	0.08	0.03	0.2	4.555	52.380
D(major iatrogenic side effect)	du_{pc}	0.3	0.2	0.9	15.291	35.680

Table 5.1: The data used by Buchbinder and Detsky [22] in their analysis.

Treatment alternative	Utilility	Expectation
A Treat none	0.6870	0.6991
B Biopsy and treat positive	0.7575	0.7570
C Biopsy and treat all	0.7398	0.7371
D Treat all	0.7198	0.7171

Table 5.2: The utilities of the different strategies when all the input parameters are set to their base value (second column) and their expectation when they are random (third column).

Sensitivity analysis

As already mentioned, the clinician wants to determine the highest utility. In [13], the authors then consider the highest utility as output and lead a sensitivity analysis to determine the input having the largest influence on this output. Since we are able to treat multivariate outputs, we consider a more general framework in this paper: the output is the four-dimensional random variable $Y = (Y_A, Y_B, Y_C, Y_D)$, where Y_S represents the outcome corresponding to strategy S.

We compare three different methodologies and indices. First, we consider the Sobol indices introduced in [47] (multivariate). Second, we consider the indices constructed in this paper, based on the Cramér-von Mises distance and estimated by the ratios of the numerator estimator (5.11) and the denominator estimator (5.12). Third, we consider the index presented in [13] and named β defined by

$$\beta_i = \mathbb{E}[\sup_{y \in \mathscr{Y}} \{|F_Y(y) - F_{Y|X_i}(y)|\}].$$

Then we use the estimator given in [14, Table 1] adapted to the multivariate case that is based on the tedious and costly estimation of conditional expectations.

Results

Table 5.3 summarizes the sensitivity measures of the seven random inputs on the multivariate output with the three different methodologies while Table 5.4 presents the associated ranks. It is worth mentioning that the same total sample size has been used to compare properly the three methodologies.

As a conclusion, in this example, unlike the indices defined by Borgonovo *et al.*, the multivariate sensitivity indices and the Cramér-von Mises indices provide the same ranking. The main advantage of the Cramér-von Mises sensitivity methodology with respect to the one of Borgonovo *et al.* is that one can use the pick and freeze estimation scheme which provides an accurate estimation (see (5.11)) simple to implement.

Notice that in [14], the authors study a slightly different model that explains the numerical differences between their results and the ones of the present paper. Furthermore, they perform a sensitivity analysis on the best alternative with the greater mean instead of considering the multivariate output.

	Sensitivity meas.	1	2	3	4	5	6	7	Cputime
$N = 10^2$	Multivariate	0.3690	0.0193	0.0105	-0.0821	-0.0617	0.1150	-0.0751	0.0624
	Borgonovo *et al.*	0.1195	0.1047	0.1064	0.1022	0.1046	0.1063	0.1027	1.5132
	Cramér-von Mises	0.3496	0.0745	0.0206	-0.0010	0.0084	0.1042	0.0105	0.9048
$N = 10^3$	Multivariate	0.4024	0.1201	0.0516	-0.0190	-0.0043	0.2403	0.0093	0.0156
	Borgonovo *et al.*	0.1788	0.1192	0.1009	0.1007	0.1044	0.1195	0.1028	57.8452
	Cramér-von Mises	0.3494	0.0750	0.0209	-0.0008	0.0086	0.1045	0.0109	10.1089
$N = 10^4$	Multivariate	0.3828	0.1333	0.0618	-0.0016	0.0100	0.3182	0.0217	0.0312
	Borgonovo *et al.*	0.3842	0.1572	0.1033	0.0930	0.0986	0.1775	0.1061	$5.1988\ 10^3$
	Cramér-von Mises	0.3494	0.0775	0.0232	0.0011	0.0108	0.1056	0.0124	436.8028

Table 5.3: Sensitivity measures. The estimation of the Cramér-von Mises indices is the ratio of (5.11) and (5.12).

	Sensitivity meas.	Ranking
$N = 10^2$	Multivariate	1 6 2 3 5 7 4
	Borgonovo *et al.*	1 3 6 2 5 7 4
	Cramér-von Mises	1 6 2 3 7 5 4
$N = 10^3$	Multivariate	1 6 2 3 7 5 4
	Borgonovo *et al.*	1 6 2 5 7 3 4
	Cramér-von Mises	1 6 2 3 7 5 4
$N = 10^4$	Multivariate	1 6 2 3 7 5 4
	Borgonovo *et al.*	1 6 2 7 3 5 4
	Cramér-von Mises	1 6 2 3 7 5 4

Table 5.4: Ranks. The estimation of the Cramér-von Mises indices is the ratio of (5.11) and (5.12).

5.3 Sensitivity indices for codes valued in general metric spaces

We consider a black-box code f defined on a product of measurable spaces $E = E_1 \times E_2 \times \ldots \times E_p$ ($p \in \mathbb{N}^*$) taking its values in a metric space $\mathscr{X}$. The output denoted by Z is then given by

$$Z = f(X_1, \ldots, X_p). \tag{5.14}$$

We denote by $\mathbb{P}$ the distribution of the output code Z.

The aim of this work is to give answers to the following questions.

Question 1 How can we perform Global Sensitivity Analysis (GSA) when the output space is the space of probability distribution functions (p.d.f.) on $\mathbb{R}$ or the space of cumulative distribution functions (c.d.f.)?

Question 2 How can we perform GSA for stochastic computer codes?

Question 3 How can we perform GSA with respect to the choice of the distributions of the input variables?

5.3.1 The general metric spaces sensitivity index

In this section, we recall the defintion and the properties of the general metric spaces sensitivity index introduced in [51]. We also discuss several ways of estimation: the Pick-Freeze estimation as introduced in [62], the estimation procedure based on U-statistics proposed in [51], and a rank-based procedure initiated in [45].
In [51], the authors performed GSA for codes f taking values in general metric spaces. To do so, they consider a family of test functions parameterized by $m \in \mathbb{N}^*$ elements of $\mathscr{X}$ and defined by

$$\begin{array}{ccc} \mathscr{X}^m \times \mathscr{X} & \to & \mathbb{R} \\ (a, x) & \mapsto & T_a(x). \end{array}$$

Let $\mathbf{u} \subset \{1, \ldots, p\}$ and $X_{\mathbf{u}} = (X_i, i \in \mathbf{u})$. Assuming that the test functions T_a are L^2-functions with respect to the product measure $\mathbb{P}^{\otimes m} \otimes \mathbb{P}$ (where $\mathbb{P}^{\otimes m}$ is the product m-times of the distribution of the output code Z) on $\mathscr{X}^m \times \mathscr{X}$, they allow to define the general metric space (GMS) sensitivity index with respect to $X_{\mathbf{u}}$ by

$$S^{\mathbf{u}}_{2,\mathrm{GMS}} = \frac{\int_{\mathscr{X}^m} \mathbb{E}\left[(\mathbb{E}[T_a(Z)] - \mathbb{E}[T_a(Z)|X_{\mathbf{u}}])^2\right] d\mathbb{P}^{\otimes m}(a)}{\int_{\mathscr{X}^m} \mathrm{Var}(T_a(Z)) d\mathbb{P}^{\otimes m}(a)} = \frac{\int_{\mathscr{X}^m} \mathrm{Var}\left(\mathbb{E}[T_a(Z)|X_{\mathbf{u}}]\right) d\mathbb{P}^{\otimes m}(a)}{\int_{\mathscr{X}^m} \mathrm{Var}(T_a(Z)) d\mathbb{P}^{\otimes m}(a)}. \tag{5.15}$$

Roughly speaking, the previous indices divided into two parts. First, for any value of a, we consider the numerator $\mathbb{E}\big[(\mathbb{E}[T_a(Z)]-\mathbb{E}[T_a(Z)|X_{\mathbf{u}}])^2\big]$ and the denominator $\mathrm{Var}(T_a(Z))$ of the classical Sobol index of $T_a(Z)$. This part is called the Sobol part. Second, we integrate each part with respect to the measure $\mathbb{P}^{\otimes m}$; it is called the integration part.

As explained in [51], by construction, the indices $S^{\mathbf{u}}_{2,\mathrm{GMS}}$ lie in $[0,1]$ and share the same properties as their Sobol counterparts:

- the different contributions sum to 1; (5.16)
- they are invariant by translation, by any isometry and by any non-degenerated scaling of Z. (5.17)

Particular examples By convention, when the test functions T_a do not depend on a, we set $m=0$.

1. For $\mathscr{X}=\mathbb{R}$, $m=0$, and T_a given by $T_a(x)=x$, one recovers the classical Sobol indices (see [109, 110]). In this case, it appears that the parameterized test functions do not depend on the parameter a. For $\mathscr{X}=\mathbb{R}^k$ and $m=0$, one can recover the index defined for vectorial outputs in [47, 68] by extending (5.15).
2. For $\mathscr{X}=\mathbb{R}^k$, $m=1$, and T_a given by $T_a(x)=\mathbb{1}_{\{x\leq a\}}$, one recovers the index based on the Cramér-von-Mises distance defined in [49] and defined in (5.10).
3. Consider that $\mathscr{X}=\mathscr{M}$ is a manifold, $m=2$ and T_a is given by $T_a(x)=\mathbb{1}_{\{x\in\widetilde{B}(a_1,a_2)\}}$, where $\widetilde{B}(a_1,a_2)$ stands for the ball whose center is the middle point between a_1 and a_2 with radius $\overline{a_1a_2}/2$. Here, one recovers the index defined in [44].

Estimation Three different estimation procedures are available in this context. The two first methods are based on the Pick-Freeze scheme. More precisely, the Pick-Freeze scheme, considered in [62], is a well tailored design of experiment. Namely, let $X^{\mathbf{u}}$ be the random vector such that $X^{\mathbf{u}}_i=X_i$ if $i\in\mathbf{u}$ and $X^{\mathbf{u}}_i=X'_i$ if $i\notin\mathbf{u}$ where X'_i is an independent copy of X_i. We then set

$$Z^{\mathbf{u}}:=f(X^{\mathbf{u}}). \tag{5.18}$$

Further, the procedure consists in rewriting the variance of the conditional expectation in terms of covariances as follows

$$\mathrm{Var}(\mathbb{E}[Z|X_{\mathbf{u}}])=\mathrm{Cov}(Z,Z^{\mathbf{u}}). \tag{5.19}$$

Alternatively, the third estimation procedure that can be seen as an ingenious and effective approximation of the Pick-Freeze scheme is based on rank statistics [45]. Until now, it is unfortunately only available to estimate first-order indices in the case of real-valued inputs.

- First method - Pick-Freeze. Introduced in [49], this procedure is based on a double Monte-Carlo scheme to estimate the Cramér-von-Mises (CVM) indices $S^{\mathbf{u}}_{2,\mathrm{CVM}}$. More precisely, to estimate $S^{\mathbf{u}}_{2,\mathrm{GMS}}$ in our context, we consider the following design of experiment consisting in

  1. a classical Pick-Freeze N-sample, that is two N-samples of Z: $(Z_j,Z^{\mathbf{u}}_j)$, $1\leqslant j\leqslant N$;
  2. m other N-samples of Z independent of $(Z_j,Z^{\mathbf{u}}_j)_{1\leqslant j\leqslant N}$: $W_{l,k}$, $1\leqslant l\leqslant m$, $1\leqslant k\leqslant N$.

  The empirical estimator of the numerator of $S^{\mathbf{u}}_{2,\mathrm{GMS}}$ is then given by

$$\begin{aligned}\widehat{N}^{\mathbf{u}}_{2,\mathrm{GMS,PF}}=&\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{N}\sum_{j=1}^{N}T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)\,T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z^{\mathbf{u}}_j)\Big]\\&-\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{2N}\sum_{j=1}^{N}\Big(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z^{\mathbf{u}}_j)\Big)\Big]^2\end{aligned}$$

  while the one of the denominator is

$$\begin{aligned}\widehat{D}^{\mathbf{u}}_{2,\mathrm{GMS,PF}}=&\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{2N}\sum_{j=1}^{N}\Big(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)^2+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z^{\mathbf{u}}_j)^2\Big)\Big]\\&-\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{2N}\sum_{j=1}^{N}\Big(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z^{\mathbf{u}}_j)\Big)\Big]^2.\end{aligned}$$

For $\mathscr{X} = \mathbb{R}^k$, $m = 1$, and T_a given by $T_a(x) = \mathbb{1}_{x \leqslant a}$, the index $S^{\mathbf{u}}_{2,\mathrm{GMS,PF}}$ is nothing more than the index $S^{\mathbf{u}}_{2,\mathrm{CVM}}$ defined in [49] based on the Cramér-von-Mises distance and on the whole distribution of the output. Its estimator $\widehat{S}^{\mathbf{u}}_{2,\mathrm{CVM}}$ defined as the ratio of $\widehat{N}^{\mathbf{u}}_{2,\mathrm{GMS,PF}}$ and $\widehat{D}^{\mathbf{u}}_{2,\mathrm{GMS,PF}}$ with $T_a(x) = \mathbb{1}_{x \leqslant a}$ has been proved to be asymptotically Gaussian [49, Theorem 3.8]. The proof relies on Donsker's theorem and the functional delta method [122, Theorem 20.8]. Analogously, in the general case of $S^{\mathbf{u}}_{2,\mathrm{GMS}}$, the central limit theorem is still valid as soon as the collection $(T_a)_{a \in \mathscr{X}^m}$ forms a Donsker's class of functions.

- Second method - U-statistics. As done in [51], this method allows the practitioner to get rid of the additional random variables $(W_{l,k})$ for $l \in \{1,\dots,m\}$ and $k \in \{1,\dots,N\}$. The estimator is now based on U-statistics and deals simultaneously with the Sobol part and the integration part with respect to $d\mathbb{P}^{\otimes m}(a)$. It suffices to rewrite $S^{\mathbf{u}}_{2,\mathrm{GMS}}$ as

$$S^{\mathbf{u}}_{2,\mathrm{GMS}} = \frac{I(\Phi_1) - I(\Phi_2)}{I(\Phi_3) - I(\Phi_4)}, \tag{5.20}$$

where,

$$\begin{aligned}
\Phi_1(\mathbf{z}_1,\dots,\mathbf{z}_{m+1}) &= T_{z_1,\dots,z_m}(z_{m+1})\, T_{z_1,\dots,z_m}(z^{\mathbf{u}}_{m+1}),\\
\Phi_2(\mathbf{z}_1,\dots,\mathbf{z}_{m+2}) &= T_{z_1,\dots,z_m}(z_{m+1})\, T_{z_1,\dots,z_m}(z^{\mathbf{u}}_{m+2}),\\
\Phi_3(\mathbf{z}_1,\dots,\mathbf{z}_{m+1}) &= T_{z_1,\dots,z_m}(z_{m+1})^2,\\
\Phi_4(\mathbf{z}_1,\dots,\mathbf{z}_{m+2}) &= T_{z_1,\dots,z_m}(z_{m+1})\, T_{z_1,\dots,z_m}(z_{m+2}),
\end{aligned} \tag{5.21}$$

denoting by $\mathbf{z}_i$ the pair $(z_i, z^{\mathbf{u}}_i)$ and, for $l = 1,\dots,4$,

$$I(\Phi_l) = \int_{\mathscr{X}^{m(l)}} \Phi_l(\mathbf{z}_1,\dots,\mathbf{z}_{m(l)})\, d\mathbb{P}_2^{u,\otimes m(l)}(\mathbf{z}_1\dots,\mathbf{z}_{m(l)}), \tag{5.22}$$

with $m(1) = m(3) = m+1$ and $m(2) = m(4) = m+2$. Finally, one considers the empirical version of (5.20) as estimator of $S^{\mathbf{u}}_{2,\mathrm{GMS}}$

$$\widehat{S}^{\mathbf{u}}_{2,\mathrm{GMS,Ustat}} = \frac{U_{1,N} - U_{2,N}}{U_{3,N} - U_{4,N}}, \tag{5.23}$$

where, for $l = 1,\dots,4$,

$$U_{l,N} = \binom{N}{m(l)}^{-1} \sum_{1 \leqslant i_1 < \dots < i_{m(l)} \leqslant N} \Phi^s_l\left(\mathbf{Z}_{i_1},\dots,\mathbf{Z}_{i_{m(l)}}\right) \tag{5.24}$$

and the function

$$\Phi^s_l(\mathbf{z}_1,\dots,\mathbf{z}_{m(l)}) = \frac{1}{(m(l))!} \sum_{\tau \in \mathscr{S}_{m(l)}} \Phi_l(\mathbf{z}_{\tau(1)},\dots,\mathbf{z}_{\tau(m(l))})$$

is the symmetrized version of Φ_l. In [51, Theorem 2.4], the estimator $\widehat{S}^{\mathbf{u}}_{2,\mathrm{GMS,U\text{-}stat}}$ has been proved to be consistent and asymptotically Gaussian.

- Third method - Rank-based. In [28], Chatterjee proposes an efficient way based on ranks to estimate a new coefficient of correlation. This estimation procedure can be seen as an approximation of the Pick-Freeze scheme and then has been exploited in [45] to perform a more efficient estimation of $S^{\mathbf{u}}_{2,\mathrm{GMS}}$. Nevertheless, this method is only well tailored for estimating first-order indices i.e. in the case of $\mathbf{u} = \{i\}$ for some $i \in \{1,\dots,p\}$ and when the input $X_i \in \mathbb{R}$.

  In our context, recall that $\mathbf{u} = \{i\}$ and let $Y = Z$. Let also $\pi_i(j)$ be the rank of $X_{i,j}$ in the sample $(X_{i,1},\dots,X_{i,N})$ of X_i and define

$$N_i(j) = \begin{cases} \pi_i^{-1}(\pi_i(j)+1) & \text{if } \pi_i(j)+1 \leqslant N,\\ \pi_i^{-1}(1) & \text{if } \pi_i(j) = N. \end{cases} \tag{5.25}$$

  Then the empirical estimator $\widehat{S}^i_{2,\mathrm{GMS,Rank}}$ of $S^i_{2,\mathrm{GMS}}$ only requires a N-sample $(Z_j)_{1 \leqslant j \leqslant N}$ of Z and is given by the ratio between

$$\begin{aligned}
\widehat{N}^i_{2,\mathrm{GMS,Rank}} = &\frac{1}{N^m} \sum_{1 \leqslant i_1,\dots,i_m \leqslant N} \left[\frac{1}{N}\sum_{j=1}^N T_{Z_{i_1},\cdots,Z_{i_m}}(Z_j)\, T_{Z_{i_1},\cdots,Z_{i_m}}(Z_{N_i(j)})\right]\\
&- \frac{1}{N^m} \sum_{1 \leqslant i_1,\dots,i_m \leqslant N} \left[\frac{1}{N}\sum_{j=1}^N T_{Z_{i_1},\cdots,Z_{i_m}}(Z_j)\right]^2
\end{aligned} \tag{5.26}$$

and $\widehat{D}^i_{2,\text{GMS,Rank}}$

$$\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{N}\sum_{j=1}^{N}T_{Z_{i_1},\cdots,Z_{i_m}}(Z_j)^2\Big]-\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{N}\sum_{j=1}^{N}T_{Z_{i_1},\cdots,Z_{i_m}}(Z_j)\Big]^2. \tag{5.27}$$

It is worth mentioning that $Z_{N_i(j)}$ plays the same role as Z^i_j (the Pick-Freeze version of Z_j with respect to X_i) in the Pick-Freeze estimation procedure.

Comparison of the estimation procedures
First, the Pick-Freeze estimation procedure allows the estimation of several sensitivity indices: the classical Sobol indices for real-valued outputs, as well as their generalization for vectorial-valued codes, but also the indices based on higher moments [93] and the Cramér-von-Mises indices which take the whole distribution into account [44,49]. Practically, this methodology is quite general and easy to implement. Moreover, the Pick-Freeze estimators have desirable statistical properties. More precisely, this estimation scheme has been proved to be consistent and asymptotically normal (i.e. the rate of convergence is $\sqrt{N}$) in [48,51,62]. The limiting variances can be computed explicitly, allowing the practitioner to build confidence intervals. In addition, for a given sample size N, exponential inequalities have been established. Last but not least, the sequence of estimators is asymptotically efficient from such a design of experiment (see, [122] for the definition of the asymptotic efficiency and [48] for more details on the result).
However, the Pick-Freeze estimators have two major drawbacks. First, they rely on a particular experimental design that may be unavailable in practice. Second, it can be unfortunately very time consuming in practice: the number of model calls to estimate all first-order Sobol indices grows linearly with the number of input parameters. For example, if we consider $p = 99$ input parameters and only $N = 1000$ calls are allowed, then only a sample of size $N/(p+1) = 10$ is available to estimate each single first-order Sobol index.

Secondly, the estimation procedure based on U-statistics has the same kind of asymptotic guarantees as the Pick-Freeze estimators (namely, consistency and asymptotic normality). Furthermore, the estimation scheme is reduced to $2N$ evaluations of the code. Last, using the results of Hoeffding [59] on U-statistics, the asymptotic normality is proved straightforwardly.

Finally, embedding Chatterjee's method in the GSA framework (called rank-based method in this framework) thereby eliminates the two drawbacks of the classical Pick-Freeze estimation. Indeed, the strength of the rank-based estimation procedure lies in the fact that only one N-sample of Z is required while $(m+2)$ samples of size N are necessary in the Pick-Freeze estimation of a single index (worse, $(m+1+p)$ samples of size N are required when one wants to estimate p indices). Using a single sample of size N, it is now possible to estimate at the same time all the first-order Sobol indices, first-order Cramér-von-Mises indices, and other useful first-order sensitivity indices as soon as all inputs are real valued. More generally, the rank-based method allows for the estimation of a large class of GSA indices which includes the Sobol indices and the higher-order moment indices proposed by Owen [91–93]. In addition, the rank-based estimator has nice theoretical properties. For instance, the estimator of the Sobol index S^i has been proved to be consistent and asymptotically Gaussian (see, e.g., Theorem 3.3 in [45]).

5.3.2 The universal sensitivity index

In this section, the aim is to generalize the already-known general metric space index $S^{\mathbf{u}}_{2,\text{GMS}}$ presented in the previous section. Here, we then define a new index that we call the "universal sensitivity index" and we denote by $S^{\mathbf{u}}_{2,\text{Univ}}$.
To do so, observe that Formula (5.15) can be generalized in the following ways.

1. The point a in the definition of the test functions can be allowed to belong to another measurable space than $\mathscr{X}^m$.

2. The probability measure $\mathbb{P}^{\otimes m}$ in (5.15) can be replaced by any "admissible" probability measure.

Such generalizations lead to the definition of a universal sensitivity index and its procedures of estimation.

Definition 5.12. *Let a belongs to some measurable space Ω endowed with some probability measure $\mathbb{Q}$. For any $\boldsymbol{u}\subset\{1,\cdots,p\}$, we define the* universal sensitivity index *with respect to $X_{\boldsymbol{u}}$ by*

$$S^{\boldsymbol{u}}_{2,Univ}(T_a,\mathbb{Q})=\frac{\int_\Omega \mathbb{E}\left[(\mathbb{E}[T_a(Z)]-\mathbb{E}[T_a(Z)|X_{\boldsymbol{u}}])^2\right]d\mathbb{Q}(a)}{\int_\Omega \text{Var}(T_a(Z))d\mathbb{Q}(a)}=\frac{\int_\Omega \text{Var}\left(\mathbb{E}[T_a(Z)|X_{\boldsymbol{u}}]\right)d\mathbb{Q}(a)}{\int_\Omega \text{Var}(T_a(Z))d\mathbb{Q}(a)}. \tag{5.28}$$

Notice that the index $S^{\mathbf{u}}_{2,\mathrm{Univ}}(T_a,\mathbb{Q})$ is obtained by the integration over a with respect to $\mathbb{Q}$ of the Hoeffding decomposition of $T_a(Z)$. Hence, by construction, this index lies in $[0,1]$ and shares the same properties as its Sobol counterparts, namely the two previously cited properties in (5.16) and (5.17).
The universality is twofold. First, it allows to consider more general relevant indices. Secondly, this definition encompasses, as particular cases, the classical sensitivity indices. Indeed,

- the so-called Sobol index $S^{\mathbf{u}}$ with respect to $X_{\mathbf{u}}$ is $S^{\mathbf{u}}_{2,\mathrm{Univ}}(\mathrm{Id},\mathbb{P})$, with Id the identity test function;
- the Cramér-von-Mises index $S^{\mathbf{u}}_{2,\mathrm{CVM}}$ with respect to $X_{\mathbf{u}}$ is $S^{\mathbf{u}}_{2,\mathrm{Univ}}(\mathbb{1}_{\cdot\leqslant a},\mathbb{P}^{\otimes d})$ where $\mathscr{X}=\mathbb{R}^d$ and $\Omega=\mathscr{X}$;
- the general metric space sensitivity index $S^{\mathbf{u}}_{2,\mathrm{GMS}}$ with respect to $X_{\mathbf{u}}$ is $S^{\mathbf{u}}_{2,\mathrm{Univ}}(\mathbb{1}_{\cdot\leqslant a},\mathbb{P}^{\otimes m})$ where $\Omega=\mathscr{X}^m$.

An example where $\mathbb{Q}$ is different from $\mathbb{P}$ will be considered in Section 5.4.

Estimation Here, we assume that $\mathbb{Q}$ is different from $\mathbb{P}^{\otimes m}$ and we follow the same tracks as for the estimation of $S^{\mathbf{u}}_{2,\mathrm{GMS}}$ in Section 5.3.1.

- First method - Pick-Freeze. We use the same design of experiment as in the first method of Section 5.3.1 but instead of considering that the m additional N-samples $(W_{l,k})$ for $l\in\{1,\dots,m\}$ and $k\in\{1,\dots,N\}$ are drawn with respect to the distribution $\mathbb{P}$ of the output, they are now drawn with respect to $\mathbb{Q}$. More precisely, we consider the following design of experiment consisting in
  1. a classical Pick-Freeze sample, that is two N-samples of Z: $(Z_j,Z^{\mathbf{u}}_j)$, $1\leqslant j\leqslant N$;
  2. m $\mathbb{Q}$-distributed N-samples $W_{l,k}$, $l\in\{1,\dots,m\}$ and $k\in\{1,\dots,N\}$ that are independent of $(Z_j,Z^{\mathbf{u}}_j)$ for $1\leqslant j\leqslant N$.

  The empirical estimator of the numerator of $S^{\mathbf{u}}_{2,\mathrm{Univ}}$ is then given by

$$\widehat{N}^{\mathbf{u}}_{2,\mathrm{Univ,PF}}=\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{N}\sum_{j=1}^{N}T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z^{\mathbf{u}}_j)\Big]$$
$$-\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{2N}\sum_{j=1}^{N}\Big(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z^{\mathbf{u}}_j)\Big)\Big]^2$$

  while the one of the denominator is

$$\widehat{D}^{\mathbf{u}}_{2,\mathrm{Univ,PF}}=\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{2N}\sum_{j=1}^{N}\Big(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)^2+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z^{\mathbf{u}}_j)^2\Big)\Big]$$
$$-\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{2N}\sum_{j=1}^{N}\Big(T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z_j)+T_{W_{1,i_1},\cdots,W_{m,i_m}}(Z^{\mathbf{u}}_j)\Big)\Big]^2.$$

  As previously, it is straightforward (as soon as the collection $(T_a)_{a\in\mathscr{X}^m}$ forms a Donsker's class of functions) to adapt the proof of Theorem [49, Theorem 3.8] to prove the asymptotic normality of the estimator.
- Second method - U-statistics. This method is not relevant in this case since $\mathbb{Q}\neq\mathbb{P}^{\otimes d}$.
- Third method - Rank-based. Here, the design of experiment reduces to
  1. a N-sample of Z: Z_j, $1\leqslant j\leqslant N$;
  2. a N-sample of W that is $\mathbb{Q}$-distributed: W_k, $1\leqslant k\leqslant N$, independent of Z_j, $1\leqslant j\leqslant N$.

  Assume as previously $\mathbf{u}=\{i\}$ and $N_i(\cdot)$ be defined in (5.25). The empirical estimator $\widehat{S}^i_{2,\mathrm{Univ,Rank}}$ of $S^i_{2,\mathrm{Univ}}$ is then given by the ratio between

$$\widehat{N}^i_{2,\mathrm{Univ,Rank}}=\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{N}\sum_{j=1}^{N}T_{W_{i_1},\cdots,W_{i_m}}(Z_j)T_{W_{i_1},\cdots,W_{i_m}}(Z_{N_i(j)})\Big]$$
$$-\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{N}\sum_{j=1}^{N}T_{W_{i_1},\cdots,W_{i_m}}(Z_j)\Big]^2 \tag{5.29}$$

  and $\widehat{D}^i_{2,\mathrm{Univ,Rank}}$

$$\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{N}\sum_{j=1}^{N}T_{W_{i_1},\cdots,W_{i_m}}(Z_j)^2\Big]-\frac{1}{N^m}\sum_{1\leqslant i_1,\dots,i_m\leqslant N}\Big[\frac{1}{N}\sum_{j=1}^{N}T_{W_{i_1},\cdots,W_{i_m}}(Z_j)\Big]^2. \tag{5.30}$$

We recall that this last method only applies for first-order sensitivity indices and real-valued input variables.

5.4 Sensitivity analysis in general Wasserstein spaces

5.4.1 Definition

For any $q \geqslant 1$, we define the q-Wasserstein distance between two probability distributions that are L^q-integrable and characterized by their c.d.f.'s F and G on $\mathbb{R}^p$ by

$$W_q(F,G) = \min_{X\sim F, Y\sim G} \left(\mathbb{E}[\|X-Y\|^q]^{1/q}\right),$$

where $X \sim F$ and $Y \sim G$ mean that X and Y are random variables with respective c.d.f.'s F and G. We define the Wasserstein space $\mathscr{W}_q(\mathbb{R}^p)$ as the space of all measures defined on $\mathbb{R}^p$ endowed with the q-Wasserstein distance W_q with finite q-moments. In the sequel, any measure is identified to its c.d.f. or in some cases to its p.d.f. In the unidimensional case ($p = 1$), it is a well known fact that $W_q(F,G)$ has an explicitly expression given by

$$W_q(F,G) = \left(\int_0^1 |F^-(v) - G^-(v)|^q \, dv\right)^{1/q} = \mathbb{E}[|F^-(U) - G^-(U)|^q]^{1/q}, \tag{5.31}$$

where F^- and G^- are the generalized inverses of the increasing functions F and G and U is a random variable uniformly distributed on $[0,1]$. Of course, $F^-(U)$ and $G^-(U)$ have c.d.f.'s F and G. The representation (5.31) of the q-Wasserstein distance when $p = 1$ can be generalized to a wider class of "contrast functions". For more details on Wasserstein spaces, one can refer to [123] and [11] and the references therein.

Definition 5.13. *We call* contrast function *any application c from $\mathbb{R}^2$ to $\mathbb{R}$ satisfying the "measure property" $\mathscr{P}$ defined by*

$$\mathscr{P} : \forall x \leqslant x' \text{ and } \forall y \leqslant y', c(x',y') - c(x',y) - c(x,y') + c(x,y) \leqslant 0,$$

meaning that c defines a negative measure on $\mathbb{R}^2$.

For instance, $c(x,y) = -xy$ satisfies $\mathscr{P}$. If c satisfies $\mathscr{P}$, any function of the form $a(x) + b(y) + c(x,y)$ also satisfies $\mathscr{P}$. If C is a convex real function, $c(x,y) = C(x-y)$ satisfies $\mathscr{P}$. In particular, $c(x,y) = (x-y)^2 = x^2 + y^2 - 2xy$ satisfies $\mathscr{P}$ and actually so does $c(x,y) = |x-y|^q$ as soon as $q \geqslant 1$.

Definition 5.14. *We define the* Skorokhod space $\mathscr{D} := \mathscr{D}([0,1])$ *of all distribution functions as the space of all non-decreasing functions from $\mathbb{R}$ to $[0,1]$ that are right-continuous with left-hand limits with limit 0 (resp. 1) in $-\infty$ (resp. $+\infty$) equipped with the supremum norm.*

Definition 5.15. *For any $F \in \mathscr{D}$, any $G \in \mathscr{D}$, and any positive contrast function c, we define the c-*Wasserstein cost *by*

$$W_c(F,G) = \min_{X\sim F, Y\sim G} \mathbb{E}[c(X,Y)] < +\infty.$$

Obviously, $W_q^q = W_c$ with $c(x,y) = |x-y|^q$. The following theorem has been established by Cambanis, Simon, and Stout in [24].

Theorem 5.16. *Let c be a contrast function. Then*

$$W_c(F,G) = \int_0^1 c(F^-(v), G^-(v)) \, dv = \mathbb{E}[c(F^-(U), G^-(U))],$$

where U is a random variable uniformly distributed on $[0,1]$.

5.4.2 Sensitivity analysis

In this section, we particularize the indices defined in Section 5.3.2 in the specific context of general Wasserstein spaces.

More precisely, we consider here that our computer code is $\mathscr{W}_q(\mathbb{R})$-valued; namely, the output of an experiment is the c.d.f. or the p.d.f. of a measure $\mu \in \mathscr{W}_q(\mathbb{R})$. For instance, in [21], [70] and [86], the authors deal with p.d.f.-valued

computer codes (and stochastic computer codes). In other words, they define the following application

$$\begin{aligned} f: \quad E &\to \mathscr{F} \\ x &\mapsto f_x \end{aligned} \tag{5.32}$$

where $\mathscr{F}$ is the set of p.d.f.'s

$$\mathscr{F} = \left\{ g \in L^1(\mathbb{R});\ g \geqslant 0,\ \int_{\mathbb{R}} g(t)\,dt = 1 \right\}.$$

Here, we choose to identify any element of $\mathscr{W}_q(\mathbb{R})$ with its c.d.f. In this framework, the output of the computer code is then a c.d.f. denoted by

$$\mathbb{F} = f(X_1, \ldots, X_p). \tag{5.33}$$

Moreover, $\mathbb{P}$ denotes the law of the c.d.f. $\mathbb{F}$ and we set $q = 2$. The case of a general value of q can be handled analogously. Consider F, F_1, and F_2 three elements of $\mathscr{W}_2(\mathbb{R})$ and, for $a = (F_1, F_2)$, the family of test functions

$$T_a(F) = T_{(F_1,F_2)}(F) = \mathbb{1}_{W_2(F_1,F) \leqslant W_2(F_1,F_2)}. \tag{5.34}$$

Then, for all $\mathbf{u} \subset \{1, \cdots, p\}$, the already known index $S^{\mathbf{u}}_{2,\mathrm{GMS}}$ of (5.15) becomes

$$\begin{aligned} S^{\mathbf{u}}_{2,W_2} &= S^{\mathbf{u}}_{2,\mathrm{Univ}}((F_1,F_2,F) \mapsto T_{F_1,F_2}(F), \mathbb{P}^{\otimes 2}) \\ &= \frac{\int_{\mathscr{W}_2(\mathbb{R})\times\mathscr{W}_2(\mathbb{R})} \mathbb{E}\left[\left(\mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)}] - \mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)} | X_{\mathbf{u}}] \right)^2 \right] d\mathbb{P}^{\otimes 2}(F_1,F_2)}{\int_{\mathscr{W}_2(\mathbb{R})\times\mathscr{W}_2(\mathbb{R})} \mathrm{Var}(\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)}) d\mathbb{P}^{\otimes 2}(F_1,F_2)} \\ &= \frac{\int_{\mathscr{W}_2(\mathbb{R})\times\mathscr{W}_2(\mathbb{R})} \mathrm{Var}\left(\mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)} | X_{\mathbf{u}}] \right) d\mathbb{P}^{\otimes 2}(F_1,F_2)}{\int_{\mathscr{W}_2(\mathbb{R})\times\mathscr{W}_2(\mathbb{R})} \mathrm{Var}(\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)}) d\mathbb{P}^{\otimes 2}(F_1,F_2)}. \end{aligned} \tag{5.35}$$

As explained in Section 5.3.1, $S^{\mathbf{u}}_{2,W_2}$ is obtained by integration over a with respect to $\mathbb{P}$ of the Hoeffding decomposition of $T_a(\mathbb{F})$. Hence, by construction, this index lies in $[0,1]$ and shares the two properties previously cited in (5.16) and (5.17).

5.4.3 Estimation procedure

As noticed in the previous section,

$$S^{\mathbf{u}}_{2,W_2} = S^{\mathbf{u}}_{2,\mathrm{Univ}}(T_a, \mathbb{P}^{\otimes 2})$$

with T_a defined in (5.34) is a particular case of indices of the form (5.28). When a belongs to the same space as the output and when $\mathbb{Q}$ is equal to $\mathbb{P}^{\otimes m}$, we first use the Pick-Freeze estimations of the indices given in (5.35). To do so, it is convenient once again to use (5.19) leading to

$$S^{\mathbf{u}}_{2,W_2} = \frac{\int_{\mathscr{W}_2(\mathbb{R})\times\mathscr{W}_2(\mathbb{R})} \mathrm{Cov}\left(\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)}, \mathbb{1}_{W_2(F_1,\mathbb{F}^{\mathbf{u}}) \leqslant W_2(F_1,F_2)} \right) d\mathbb{P}^{\otimes 2}(F_1,F_2)}{\int_{\mathscr{W}_2(\mathbb{R})\times\mathscr{W}_2(\mathbb{R})} \mathrm{Var}(\mathbb{1}_{W_2(F_1,\mathbb{F}) \leqslant W_2(F_1,F_2)}) d\mathbb{P}^{\otimes 2}(F_1,F_2)} \tag{5.36}$$

Secondly, one may resort to the estimations based on U-statistics together on the Pick-Freeze design of experiment. Thirdly, it is also possible and easy to obtain rank-based estimations in the vein of (4.2).

5.4.4 Numerical application

Example 5.17 (Toy model)**.** *Let X_1, X_2, X_3 be three independent and positive random variables. We consider the c.d.f.-valued code f for which the output is given by*

$$\mathbb{F}(t) = \frac{t}{1 + X_1 + X_2 + X_1 X_3} \mathbb{1}_{0 \leqslant t \leqslant 1 + X_1 + X_2 + X_1 X_3} + \mathbb{1}_{1 + X_1 + X_2 + X_1 X_3 < t}, \tag{5.37}$$

so that

$$\mathbb{F}^{-1}(v) = v\Big(1 + X_1 + X_2 + X_1 X_3\Big). \tag{5.38}$$

In (5.35), the distributions F_1 and F_2 can be either $\mathscr{U}([0,1])$, $\mathscr{U}([0,2])$, $\mathscr{U}([0,3])$, or $\mathscr{U}([0,4])$ with respective probabilities $q_1 = (1-p_1)(1-p_2)$, $q_2 = (1-p_1)p_2 + p_1(1-p_2)(1-p_3)$, $q_3 = p_1((1-p_2)p_3 + p_2(1-p_3))$, and $q_4 = p_1 p_2 p_3$. In the sequel, we give, for all sixteen possibilities for the distribution of (F_1, F_2), the corresponding contributions for the numerator and for the denominator of (5.35).

With probability $p_{1,1} = (1-p_1)^2(1-p_2)^2$, F_1 *and* $F_2 \sim \mathscr{U}([0,1])$. *Then* $W_2^2(F_1,F_2) = 0$, $W_2^2(F_1,\mathbb{F}) = \frac{1}{3}(X_1+X_2+X_1X_3)^2$, *and* $W_2^2(F_1,\mathbb{F}) \leqslant W_2^2(F_1,F_2)$ *if and only if* $X_1+X_2+X_1X_3 = 0$. *Since* $\mathbb{P}(X_1+X_2+X_1X_3=0) = (1-p_1)(1-p_2)$, *the contribution* $d_{1,1}$ *to the denominator is thus*

$$d_{1,1} = q_{1,1}(1-q_{1,1}) \quad \text{with } q_{1,1} = (1-p_1)(1-p_2).$$

Moreover,

$$\mathbb{E}[\mathbb{1}_{X_1+X_2+X_1X_3=0}|X_1] = \mathbb{P}\Big(X_1+X_2+X_1X_3=0|X_1\Big) = \mathbb{1}_{X_1=0}\mathbb{P}(X_2=0) = (1-p_2)\mathbb{1}_{X_1=0}.$$

so that, the contribution to the numerator is given by

$$n_{1,1}^1 = \mathrm{Var}(\mathbb{E}[\mathbb{1}_{X_1+X_2+X_1X_3=0}|X_1]) = p_1(1-p_1)(1-p_2)^2.$$

Similarly, one gets

$$n_{1,1}^2 = \mathrm{Var}(\mathbb{E}[\mathbb{1}_{X_1+X_2+X_1X_3=0}|X_2]) = p_2(1-p_2)(1-p_1)^2 \quad \text{and} \quad n_{1,1}^3 = 0.$$

Moreover, regarding the indices with respect to X_1 *and* X_3,

$$\mathbb{E}[\mathbb{1}_{X_1+X_2+X_1X_3=0}|X_1,X_3] = \mathbb{P}\Big(X_1+X_2+X_1X_3=0|X_1,X_3\Big) = \mathbb{1}_{X_1=0}\mathbb{P}(X_2=0) = (1-p_2)\mathbb{1}_{X_1=0}$$

and the contribution to the numerator is given by

$$n_{1,1}^{1,3} = \mathrm{Var}(\mathbb{E}[\mathbb{1}_{X_1+X_2+X_1X_3=0}|X_1,X_3]) = p_1(1-p_1)(1-p_2)^2.$$

The remaining fifteen cases can be treated similarly and are gathered (with the first case developed above) in the following table. Finally, one may compute the explicit expression of $S_{2,W_2}^{\boldsymbol{u}}$

$$S_{2,W_2}^{\boldsymbol{u}} = \frac{\int_{\mathscr{W}_2(\mathbb{R})\times\mathscr{W}_2(\mathbb{R})} \mathrm{Cov}\left(\mathbb{1}_{W_2(F_1,\mathbb{F})\leqslant W_2(F_1,F_2)}, \mathbb{1}_{W_2(F_1,\mathbb{F}^{\boldsymbol{u}})\leqslant W_2(F_1,F_2)}\right) d\mathbb{P}^{\otimes 2}(F_1,F_2)}{\int_{\mathscr{W}_2(\mathbb{R})\times\mathscr{W}_2(\mathbb{R})} \mathrm{Var}(\mathbb{1}_{W_2(F_1,\mathbb{F})\leqslant W_2(F_1,F_2)}) d\mathbb{P}^{\otimes 2}(F_1,F_2)} = \frac{\sum_{k,l} p_{k,l} n_{k,l}^{\boldsymbol{u}}}{\sum_{k,l} p_{k,l} d_{k,l}}.$$

Some numerical values have not been explicited in the table but given below

Case 2 $\mathrm{Var}(\mathbb{1}_{X_1=1}(1-(1-p_2)\mathbb{1}_{X_3=0})) = p_1(1-p_1)(1-(1-p_2)(1-p_3))^2 + p_1(1-p_2)^2p_3(1-p_3),$

Case 6 $\mathrm{Var}(\mathbb{1}_{X_1=1}(p_2-(1-p_2)\mathbb{1}_{X_3=0})) = p_1(1-p_1)(p_2-(1-p_2)(1-p_3))^2 + p_1(1-p_2)^2p_3(1-p_3),$

Case 11 $\mathrm{Var}(\mathbb{1}_{X_1=1}(p_2+(1-2p_2)\mathbb{1}_{X_3=1})) = p_1(1-p_1)(p_2+(1-2p_2)p_3)^2 + p_1(1-2p_2)^2p_3(1-p_3),$

Case 15 $\mathrm{Var}(\mathbb{1}_{X_1=1}(p_2+(1-p_2)\mathbb{1}_{X_3=1})) = p_1(1-p_1)(p_2+(1-p_2)p_3)^2 + p_1(1-p_2)^2p_3(1-p_3).$

Direct representations of the indices $S_{2,W_2}^{\mathbf{u}}$ *In Figure 5.5, we have represented the indices* S_{2,W_2}^1, S_{2,W_2}^2, S_{2,W_2}^3, *and* $S_{2,W_2}^{1,3}$ *given by (5.35) is provided in Figure 5.5 with respect to the values of* p_1 *and* p_2 *varying from 0 to 1 for a fixed value of* p_3. *We have considered three different values of* p_3*:* $p_3 = 0.01$ *(first row),* 0.5, *(second row) and* 0.99 *(third row).*

Regions of predominance of the indices $S_{2,W_2}^{\mathbf{u}}$ *In addition, the regions of predominance of each index* S_{2,W_2}^1, S_{2,W_2}^2, S_{2,W_2}^3, *and* $S_{2,W_2}^{1,3}$ *given by (5.35) is provided in Figure 5.5. The values of* p_1 *and* p_2 *still vary from 0 to 1 and the fixed values of* p_3 *considered are:* $p_3 = 0.01$ *(first row),* 0.5, *(second row) and* 0.99 *(third row).*

Comparison of the estimation procedures (rank-based and Pick-Freeze) *In order to compare the accuracy of the Pick-Freeze method and the accuracy of the rank-based method at a fixed size, we assume that only* 450 *calls of the computer code are allowed to estimate the indices* $S_{2,W_2}^{\boldsymbol{u}}$ *for* $\boldsymbol{u} = \{1\}, \{2\}$, *and* $\{3\}$. *Hence, the sample size allowed in the rank-based procedure is* $N = 450$. *In the Pick-Freeze methodology, the estimation of the Wasserstein indices* $S_{2,W_2}^{\boldsymbol{u}}$ *requires one initial output sample, three extra output samples, and two extra samples to handle the integration to get the Pick-Freeze versions (one for each index) leading to an allowed sample size* $N = \lfloor 450/6 \rfloor = 75$. *We only focus on the first-order indices since, as explained previously, the rank-based procedure has not been developed yet for higher-order indices. We repeat the estimation procedure* $n_r = 200$ *times. The boxplots of the mean square errors for the estimation of the Wasserstein indices* $S_{2,W_2}^{\boldsymbol{u}}$ *have been plotted in Figure 5.7. We observe that, for a fixed total number of calls* 450 *to the code* f *(corresponding to a rank-based sample size* $N = 450$ *and to a Pick-Freeze sample size* $N = 74$ *for the Wasserstein indices* $S_{2,W_2}^{\boldsymbol{u}}$*), the rank-based estimation procedure performs much better than the Pick-Freeze method with significantly lower mean errors.*

Case 1	$F_1 \sim \mathcal{U}([0,1])$, $F_2 \sim \mathcal{U}([0,1])$	Case 2	$F_1 \sim \mathcal{U}([0,1])$, $F_2 \sim \mathcal{U}([0,2])$
Prob.	q_1^2	Prob.	$q_1 q_2$
Num. 1	$p_1(1-p_1)(1-p_2)^2$	Num. 1	$p_1(1-p_1)(p_2+p_3-p_2p_3)^2$
Num. 2	$(1-p_1)^2 p_2(1-p_2)$	Num. 2	$p_1^2 p_2(1-p_2)(1-p_3)^2$
Num. 3	0	Num. 3	$p_1^2(1-p_2)^2 p_3(1-p_3)$
Num. 1,3	$p_1(1-p_1)(1-p_2)^2$	Num. 1,3	$\mathrm{Var}(\mathbb{1}_{X_1=1}(1-(1-p_2)\mathbb{1}_{X_3=0})$
q Den.	$(1-p_1)(1-p_2)$	q Den.	$(1-p_1)+p_1(1-p_2)(1-p_3)$
Case 3	$F_1 \sim \mathcal{U}([0,1])$, $F_2 \sim \mathcal{U}([0,3])$	Case 4	$F_1 \sim \mathcal{U}([0,1])$, $F_2 \sim \mathcal{U}([0,4])$
Prob.	$q_1 q_3$	Prob.	$q_1 q_4$
Num. 1	$p_1(1-p_1)p_2^2 p_3^2$	Num. 1	0
Num. 2	$p_1^2 p_2(1-p_2)p_3^2$	Num. 2	0
Num. 3	$p_1^2 p_2^2 p_3(1-p_3)$	Num. 3	0
Num. 1,3	$p_1 p_2^2 p_3(1-p_1p_3)$	Num. 1,3	0
q Den.	$1-p_1p_2p_3$	q Den.	0
Case 5	$F_1 \sim \mathcal{U}([0,2])$, $F_2 \sim \mathcal{U}([0,1])$	Case 6	$F_1 \sim \mathcal{U}([0,2])$, $F_2 \sim \mathcal{U}([0,2])$
Prob.	$q_1 q_2$	Prob.	q_2^2
Num. 1	$p_1(1-p_1)p_2^2 p_3^2$	Num. 1	$p_1(1-p_1)(p_2-(1-p_2)(1-p_3))^2$
Num. 2	$p_1^2 p_2(1-p_2)p_3^2$	Num. 2	$p_2(1-p_2)(p_1(1-p_3)-(1-p_1))^2$
Num. 3	$p_1^2 p_2^2 p_3(1-p_3)$	Num. 3	$p_1^2(1-p_2)^2 p_3(1-p_3)$
Num. 1,3	$p_1 p_2^2 p_3(1-p_1p_3)$	Num. 1,3	$\mathrm{Var}(\mathbb{1}_{X_1=1}(p_2-(1-p_2)\mathbb{1}_{X_3=0}))$
q Den.	$1-p_1p_2p_3$	q Den.	$(1-p_1)p_2+p_1(1-p_2)(1-p_3)$
Case 7	$F_1 \sim \mathcal{U}([0,2])$, $F_2 \sim \mathcal{U}([0,3])$	Case 8	$F_1 \sim \mathcal{U}([0,2])$, $F_2 \sim \mathcal{U}([0,4])$
Prob.	$q_2 q_3$	Prob.	$q_2 q_4$
Num. 1	$p_1(1-p_1)p_2^2 p_3^2$	Num. 1	0
Num. 2	$p_1^2 p_2(1-p_2)p_3^2$	Num. 2	0
Num. 3	$p_1^2 p_2^2 p_3(1-p_3)$	Num. 3	0
Num. 1,3	$p_1 p_2^2 p_3(1-p_1p_3)$	Num. 1,3	0
q Den.	$1-p_1p_2p_3$	q Den.	0
Case 9	$F_1 \sim \mathcal{U}([0,3])$, $F_2 \sim \mathcal{U}([0,1])$	Case 10	$F_1 \sim \mathcal{U}([0,3])$, $F_2 \sim \mathcal{U}([0,2])$
Prob.	$q_1 q_3$	Prob.	$q_2 q_3$
Num. 1	0	Num. 1	$p_1(1-p_1)(1-p_2)^2$
Num. 2	0	Num. 2	$(1-p_1)^2 p_2(1-p_2)$
Num. 3	0	Num. 3	0
Num. 1,3	0	Num. 1,3	$p_1(1-p_1)(1-p_2)^2$
q Den.	0	q Den.	$(1-p_1)p_2+p_1$
Case 11	$F_1 \sim \mathcal{U}([0,3])$, $F_2 \sim \mathcal{U}([0,3])$	Case 12	$F_1 \sim \mathcal{U}([0,3])$, $F_2 \sim \mathcal{U}([0,4])$
Prob.	q_3^2	Prob.	$q_3 q_4$
Num. 1	$p_1(1-p_1)(p_2(1-p_3)+(1-p_2)p_3)^2$	Num. 1	$p_1(1-p_1)(1-p_2)^2$
Num. 2	$p_1^2 p_2(1-p_2)(2p_3-1)^2$	Num. 2	$(1-p_1)^2 p_2(1-p_2)$
Num. 3	$p_1^2(2p_2-1)^2 p_3(1-p_3)$	Num. 3	0
Num. 1,3	$\mathrm{Var}(\mathbb{1}_{X_1=1}(p_2+(1-2p_2)\mathbb{1}_{X_3=1})$	Num. 1,3	$p_1(1-p_1)(1-p_2)^2$
q Den.	$p_1(p_2(1-p_3)+(1-p_2)p_3)$	q Den.	$(1-p_1)p_2+p_1$
Case 13	$F_1 \sim \mathcal{U}([0,4])$, $F_2 \sim \mathcal{U}([0,1])$	Case 14	$F_1 \sim \mathcal{U}([0,4])$, $F_2 \sim \mathcal{U}([0,2])$
Prob.	$q_1 q_4$	Prob.	$q_2 q_4$
Num. 1	0	Num. 1	$p_1(1-p_1)(1-p_2)^2$
Num. 2	0	Num. 2	$(1-p_1)^2 p_2(1-p_2)$
Num. 3	0	Num. 3	0
Num. 1,3	0	Num. 1,3	$p_1(1-p_1)(1-p_2)^2$
q Den.	0	q Den.	$(1-p_1)p_2+p_1$
Case 15	$F_1 \sim \mathcal{U}([0,4])$, $F_2 \sim \mathcal{U}([0,3])$	Case 16	$F_1 \sim \mathcal{U}([0,4])$, $F_2 \sim \mathcal{U}([0,4])$
Prob.	$q_3 q_4$	Prob.	q_4^2
Num. 1	$p_1(1-p_1)(p_2+(1-p_2)p_3)^2$	Num. 1	$p_1(1-p_1)p_2^2 p_3^2$
Num. 2	$p_1^2 p_2(1-p_2)(1-p_3)^2$	Num. 2	$p_1^2 p_2(1-p_2)p_3^2$
Num. 3	$p_1^2(1-p_2)^2 p_3(1-p_3)$	Num. 3	$p_1^2 p_2^2 p_3(1-p_3)$
Num. 1,3	$\mathrm{Var}(\mathbb{1}_{X_1=1}(p_2+(1-p_2)\mathbb{1}_{X_3=1})$	Num. 1,3	$p_1 p_2^2 p_3(1-p_1p_3)$
q Den.	$p_1(p_2+(1-p_2)p_3)$	q Den.	$p_1p_2p_3$

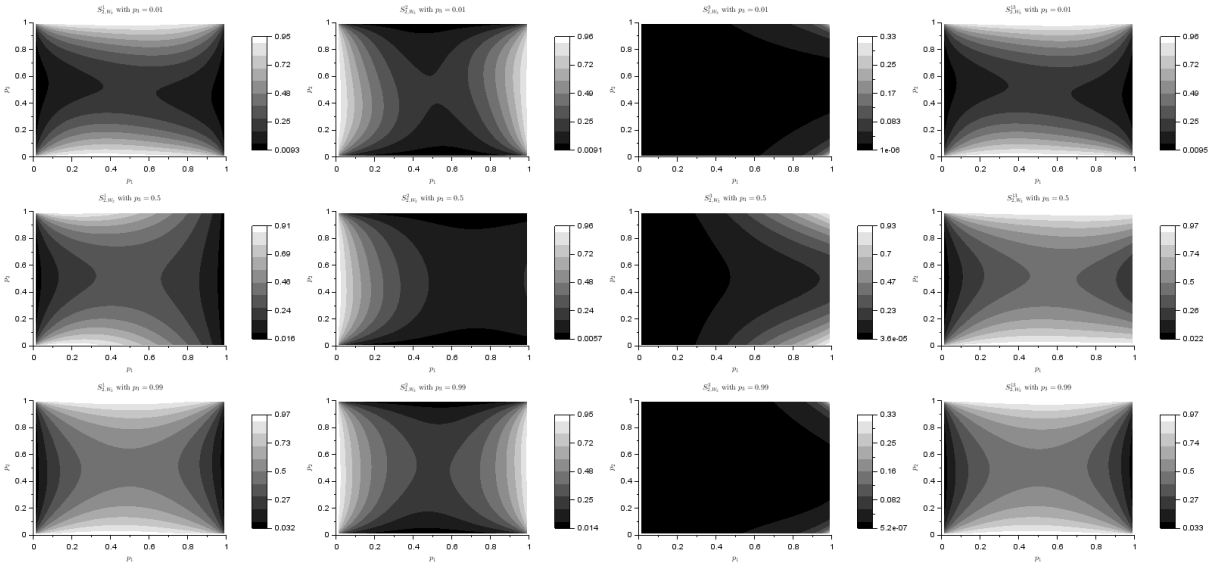


Figure 5.5: Model (5.37). Values of the indices S^1_{2,W_2}, S^2_{2,W_2}, S^3_{2,W_2}, and $S^{1,3}_{2,W_2}$ given by (5.35) (from left to right) with respect to the values of p_1 and p_2 (varying from 0 to 1). In the first row (resp. second and third), p_3 is fixed to $p_3 = 0.01$ (resp. 0.5 and 0.99).

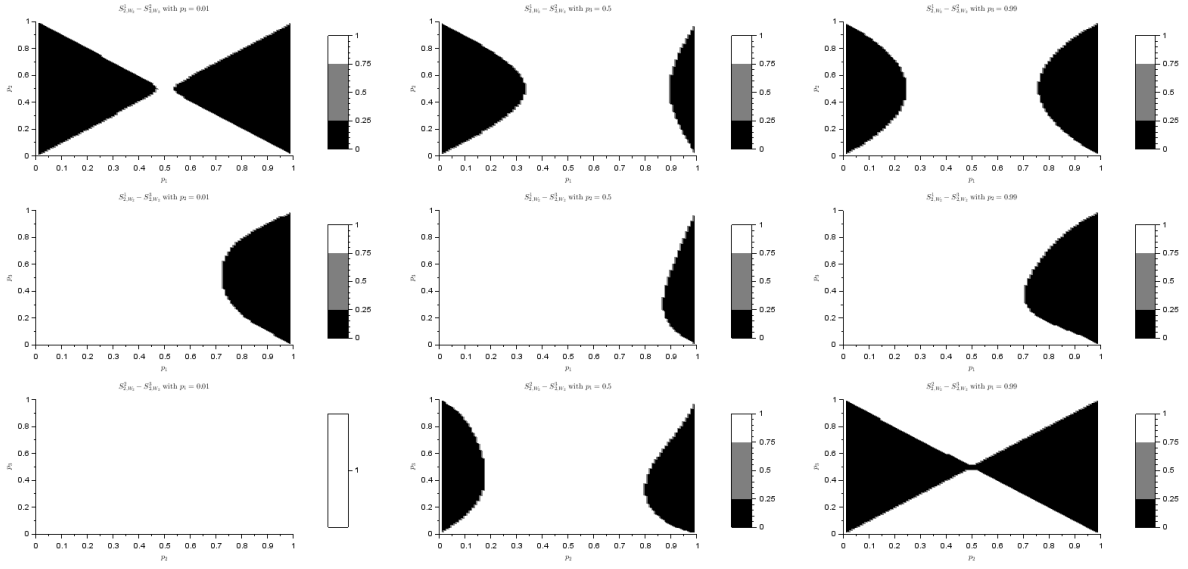


Figure 5.6: Model (5.37). In the first row of the figure, regions where $S^1_{2,W_2} \geqslant S^2_{2,W_2}$ (black), $S^1_{2,W_2} \leqslant S^2_{2,W_2}$ (white), and $S^1_{2,W_2} = S^2_{2,W_2}$ (gray) with respect to p_1 and p_2 varying from 0 to 1 and, from left to right, $p_3 = 0.01$, 0.5, and 0.99. Analogously, the second (resp. last) row considers the regions with S^1_{2,W_2} and S^3_{2,W_2} (resp. S^2_{2,W_2} and S^3_{2,W_2}) with respect to p_1 and p_3 (resp. p_2 and p_3) varying from 0 to 1 and, from left to right, $p_2 = 0.01$, 0.5, and 0.99 (resp. $p_1 = 0.01$, 0.5, and 0.99).

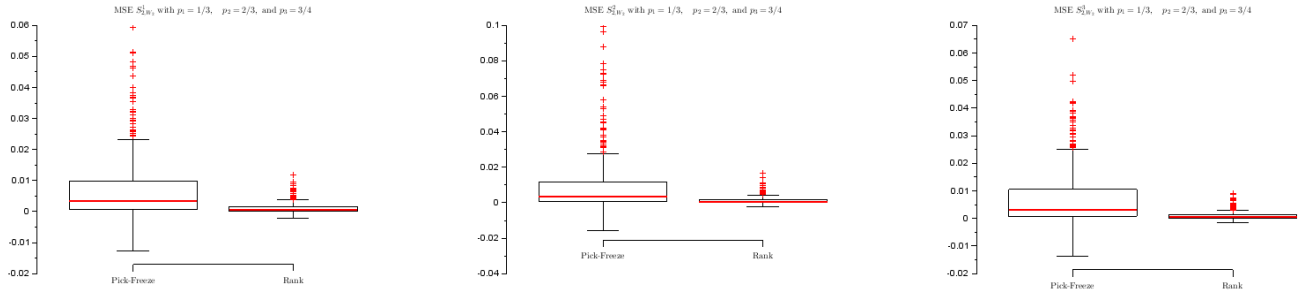


Figure 5.7: Model (5.37) with $p_1 = 1/3$, $p_2 = 2/3$, and $p_3 = 3/4$. Boxplots of the mean square errors of the estimation of the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ with a fixed sample size N and $n_r = 200$ replications. The indices with respect to $\mathbf{u} = \{1\}$, $\{2\}$, and $\{3\}$ are displayed from left to right. The results of the Pick-Freeze estimation procedure with $N = 75$ for the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ are provided in the left side of each graphic. The results of the rank-based methodology with $N = 450$ are provided in the right side of each graphic.

5.5 Sensitivity analysis for stochastic computer codes

This section deals with stochastic computer codes in the sense that two evaluations of the code for the same input lead to different outputs. Before performing a SA in this context, let us briefly describe the state of the art in this setting.

5.5.1 State of the art

A first natural way to handle stochastic computer codes is definitely to consider the expectation of the output code. Indeed, as mentioned in [21], previous works dealing with stochastic simulators together with robust design or optimization and SA consist mainly in approximating the mean and the variance of the stochastic output [1,23,36,65] and then performing a GSA on the expectation of the output code [78].
As pointed out by [60], another approach amounts to consider that the stochastic code is of the form $f(X,D)$ where the random element X contains the classical input variables and the variable D is an extra unobserved random input. Such an idea was exploited in [62] to compare the estimation of the Sobol indices in an "exact" model to the estimation of the Sobol indices in an associated metamodel. In this framework, the metamodel is considered as a random perturbation of the "exact" model and the perturbation is a function of the inputs and of an extra independent random variable. Analogously, the author of [81] assumes the existence of an extra random variable D which is not chosen by the practitioner but rather generated at each computation of the output independently of X. In this setting, the author builds two different indices. The first index is obtained by substituting $f(X,D)$ for $f(X)$ in the classical definition of the first-order Sobol index $S^i = \mathrm{Var}(\mathbb{E}[f(X)|X_i])/\mathrm{Var}(f(X))$. In this case, D is considered as another input, even though it is not observable. The second index is obtained by substituting $\mathbb{E}[f(X,D)|X]$ for $f(X)$ in the Sobol index. The noise is then smoothed out. Similarly, the authors of [57] traduce the randomness of the computer code using such an extra random variable. In practice, they approximate the statistical properties of the first-order Sobol indices by using a sample of the extra random variable and the associated sample of estimates of the Sobol indices. In [126], the expectation of these random Sobol indices is investigated in the case of stochastic simulator and generalized lambda models. In the same vein, the authors of [6] propose to deal with the differential entropy of the output of a stochastic simulator.

5.5.2 The space $\mathscr{W}_q$ as an ideal version of stochastic computer codes

When dealing with stochastic computer codes, the practitioner is generally interested in the distribution μ_x of the output for a given input x. As previously seen, one can translate this type of codes in terms of a deterministic code by considering an extra input which is not chosen by the practitioner himself but which is a latent variable generated randomly by the computer code and independently of the classical input. As usual in the framework of SA, one considers the input as a random variable. All the random variables (the one chosen by the practitioner and the one generated by the computer code) are built on the same probability space, leading to the function f_s

$$\begin{aligned} f_s : \quad E \times \mathscr{D} \quad &\to \mathbb{R} \\ (x,D) \quad &\mapsto f_s(x,D), \end{aligned} \tag{5.39}$$

where D is the extra random variable lying in some space $\mathscr{D}$. We naturally denote the output random variable $f_s(x,\cdot)$ by $f_s(x)$.

Hence, one may define another (deterministic) computer code associated with f_s for which the output associated to x is the probability measure μ_x

$$\begin{aligned} f: \quad E &\to \mathscr{W}_q(E) \\ x &\mapsto \mu_x. \end{aligned} \tag{5.40}$$

The framework of (5.40) is exactly the one of Section 5.4.2 and has already been handled. Obviously, in practice, one does not assess the output of the code f but one can only obtain an empirical approximation of the measure μ_x given by n evaluations of f_s at x, namely,

$$\mu_{x,n} = \frac{1}{n}\sum_{k=1}^{n} \delta_{f_s(x,D_k)}$$

where $\delta_.$ is the Dirac function. Further, (5.40) can be seen as an ideal version of (5.39). Concretely, for a single random input $\overrightarrow{X} = (X_1,\dots,X_p) \in E = E_1 \times \cdots \times E_p$, we will evaluate n times the code f_s defined by (5.39) (so that the code will generate independently n hidden variables $D_1,\dots,D_n$) and one may observe

$$f_s(\overrightarrow{X},D_1),\dots,f_s(\overrightarrow{X},D_n)$$

leading to the measure $\mu_{\overrightarrow{X},n} = \sum_{k=1}^{n} \delta_{f_s(\overrightarrow{X},D_k)}/n$ that approximates the distribution μ_x of $f_s(\overrightarrow{X})$. We emphasize on the fact that the random variables $D_1,\dots,D_n$ are not observed.

5.5.3 Sensitivity analysis

Let us now present the methodology we adopt in the sequel. In order to study the sensitivity of the distribution μ_x, one can use the framework introduced in Section 5.4.2 and the index $S^{\mathbf{u}}_{2,W_q}$ given by (5.35).
In an ideal scenario which corresponds to the framework of (5.40), one may assess the probability measure μ_x for any x. Then following the estimation procedure of Section 5.4.3, one gets an estimation of the sensitivity index $S^{\mathbf{u}}_{2,W_q}$ with good asymptotic properties [51, Theorem 2.4].
In the more realistic framework presented above in (5.39), we only have access to the approximation $\mu_{x,n}$ of μ_x rendering more complex the estimation procedure and the study of the asymptotic properties. In this case, the general design of experiments is the following

$$\begin{aligned}
(\overrightarrow{X}_1, D_{1,1},\dots,D_{1,n}) &\to f_s(\overrightarrow{X}_1,D_{1,1}),\dots,f_s(\overrightarrow{X}_1,D_{1,n}),\\
(\overrightarrow{X}^{\mathbf{u}}_1, D'_{1,1},\dots,D'_{1,n}) &\to f_s(\overrightarrow{X}^{\mathbf{u}}_1,D'_{1,1}),\dots,f_s(\overrightarrow{X}^{\mathbf{u}}_1,D'_{1,n}),\\
&\vdots\\
(\overrightarrow{X}_N, D_{N,1},\dots,D_{N,n}) &\to f_s(\overrightarrow{X}_N,D_{N,1}),\dots,f_s(\overrightarrow{X}_N,D_{N,n}),\\
(\overrightarrow{X}^{\mathbf{u}}_N, D'_{N,1},\dots,D'_{N,n}) &\to f_s(\overrightarrow{X}^{\mathbf{u}}_N,D'_{N,1}),\dots,f_s(\overrightarrow{X}^{\mathbf{u}}_N,D'_{N,n}),
\end{aligned}$$

where $\overrightarrow{X}_j$ is the j-th realization of $\overrightarrow{X}$ with $j = 1,\cdots,N$, $\overrightarrow{X}^{\mathbf{u}}_j$ is the associated Pick-Freeze version, and $2\times N\times n$ is the total number of evaluations of the stochastic computer code (5.39). Then, we construct the approximations $\mu_{\overrightarrow{X}_j,n}$ of $\mu_{\overrightarrow{X}_j}$ for any $j = 1,\dots,N$ given by

$$\mu_{\overrightarrow{X}_j,n} = \frac{1}{n}\sum_{k=1}^{n} \delta_{f_s(\overrightarrow{X}_j,D_{j,k})}. \tag{5.41}$$

From there, one may use one of the three estimation procedures presented in Section 5.3.1.

- First method - Pick-Freeze. It suffices to plug the empirical version μ_n of each measure μ under concern in (5.36) to get $\widehat{S}^{\mathbf{u}}_{2,W_q,\mathrm{PF},n}$.
- Second method - U-statistics. For $l = 1,\dots,4$, let

$$U_{l,N,n} = \binom{N}{m(l)}^{-1} \sum_{1\le i_1<\cdots<i_{m(l)}\le N} \Phi^s_l\left(\boldsymbol{\mu}_{i_1,n},\dots,\boldsymbol{\mu}_{i_{m(l)},n}\right) \tag{5.42}$$

  where as previously seen $\Phi^s_.$ is the symmetrized version of $\Phi_.$ defined in (5.21) and $\boldsymbol{\mu} = (\mu,\mu^{\mathbf{u}})$. Then, we estimate $S^{\mathbf{u}}_{2,W_q}$ by

$$\widehat{S}^{\mathbf{u}}_{2,W_q,\mathrm{Ustat},n} = \frac{U_{1,N,n} - U_{2,N,n}}{U_{3,N,n} - U_{4,N,n}}. \tag{5.43}$$

- Third method - Rank-based. The rank-based estimation procedure may also easily be extended to this context by using the empirical version μ_n of each measure μ under concern instead of the true one μ, as explained into more details in the numerical study developed in Section 5.5.5. This procedure leads to $\widehat{S}^{\mathbf{u}}_{2,W_q,\mathrm{Rank},n}$.

Actually, these estimators are easy to compute since, for two discrete measures supported on a same number of points and given by

$$\nu_1 = \frac{1}{n}\sum_{k=1}^{n}\delta_{x_k},\ \nu_2 = \frac{1}{n}\sum_{k=1}^{n}\delta_{y_k},$$

the Wasserstein distance between ν_1 and ν_2 simply writes

$$W_q^q(\nu_1,\nu_2) = \frac{1}{n}\sum_{k=1}^{n}(x_{(k)}-y_{(k)})^q, \tag{5.44}$$

where $z_{(k)}$ is the k-th order statistics of z.

5.5.4 Central limit theorem for the estimator based on U-statistics

In this section, we focus on the computationnally less expensive estimator: the one based on U-statistics. For statistical purposes, we establish a central limit theorem for $\widehat{S}^{\mathbf{u}}_{2,W_q,\mathrm{Ustat},n}$ inspired from [51]. In addition, we consider several examples and study when the conditions of Proposition 5.18 hold.

Proposition 5.18. *Consider three i.i.d. copies X_1, X_2 and X_3 of a random variable X. Let $\delta(N)$ be a sequence tending to 0 as N goes to infinity and such that*

$$\mathbb{P}\left(\left|W_q(\mu_{X_1},\mu_{X_3}) - W_q(\mu_{X_1},\mu_{X_2})\right| \leqslant \delta(N)\right) = o\left(\frac{1}{\sqrt{N}}\right).$$

Let n be such that $\mathbb{E}[W_q(\mu_X,\mu_{X,n})] = o(\delta(N)/\sqrt{N})$. Under the assumptions of Theorem 2.4 in [51], we get, for any $\mathbf{u}\subset\{1,\cdots,p\}$,

$$\sqrt{N}\left(\widehat{S}^{\mathbf{u}}_{2,W_q,Ustat,n} - S^{\mathbf{u}}_{2,W_q}\right) \xrightarrow[n\to+\infty]{\mathscr{L}} \mathcal{N}(0,\sigma^2) \tag{5.45}$$

where the asymptotic variance σ^2 is given by (13) in the proof of Theorem 2.4 in [51].

In some particular frameworks, one may derive easily a suitable value of $\delta(N)$. Two examples are given in the following.

Example 5.19. *If the inverse of the random variable $W = \left|W_q(\mu_{X_1},\mu_{X_3}) - W_q(\mu_{X_1},\mu_{X_2})\right|$ has a finite expectation, then, by Markov inequality,*

$$\mathbb{P}(W\leqslant\delta(N)) = \mathbb{P}\left(W^{-1}\geqslant\delta(N)^{-1}\right) \leqslant \frac{1}{\delta(N)}\mathbb{E}\left[\frac{1}{W}\right]$$

and it suffices to choose $\delta(N)$ so that $\delta(N)^{-1} = o\left(N^{-1/2}\right)$ as N goes to infinity.

Example 5.20 (Uniform example)**.** *Assume that X is uniformly distributed on $[0,1]$ and that μ_X is a Gaussian distribution centered at X with unit variance. Then the Wasserstein distance $W_2(\mu_{X_1},\mu_{X_2})$ rewrites as $(X_1-X_2)^2$ so that the random variable $W = \left|W_2(\mu_{X_1},\mu_{X_3}) - W_2(\mu_{X_1},\mu_{X_2})\right|$ is given by*

$$\left|(X_1-X_3)^2 - (X_1-X_2)^2\right| = |(X_3-X_2)(X_2+X_3-2X_1)|.$$

Consequently,

$$\mathbb{P}(W\leqslant\delta(N)) \leqslant \mathbb{P}(|X_3-X_2|\leqslant\sqrt{\delta(N)}) + \mathbb{P}(|X_2+X_3-2X_1|\leqslant\sqrt{\delta(N)}).$$

Notice that $|X_3-X_2|$ is triangularly distributed with parameter $a=0$, $b=1$, and $c=0$ leading to

$$\mathbb{P}(|X_3-X_2|\leqslant\alpha) = \alpha(2-\alpha), \quad \text{for all } \alpha\in[0,1].$$

In addition,

$$\begin{aligned}\mathbb{P}(|X_2+X_3-2X_1|\leqslant\sqrt{\delta(N)}) &\leqslant \mathbb{P}(||X_2-X_1|-|X_3-X_1||\leqslant\sqrt{\delta(N)})\\ &= \int_0^1 \mathbb{P}(||X_2-u|-|X_3-u||\leqslant\sqrt{\delta(N)})\,du.\end{aligned}$$

Now, X_2-u and X_3-u are two independent random variables uniformly distributed on $[-u,-u]$. Then (see Figure 5.8), one has

$$\mathbb{P}(||X_2-u|-|X_3-u||\leqslant\alpha)\leqslant 4\alpha,$$

whence

$$\mathbb{P}(|X_2+X_3-2X_1|\leqslant\sqrt{\delta(N)})\leqslant 4\sqrt{\delta(N)}.$$

Thus it turns out that $\mathbb{P}(W\leqslant\delta(N))=O(\sqrt{\delta(N)})$. Consequently, a suitable choice for $\delta(N)$ is $\delta(N)=o(1/N)$.

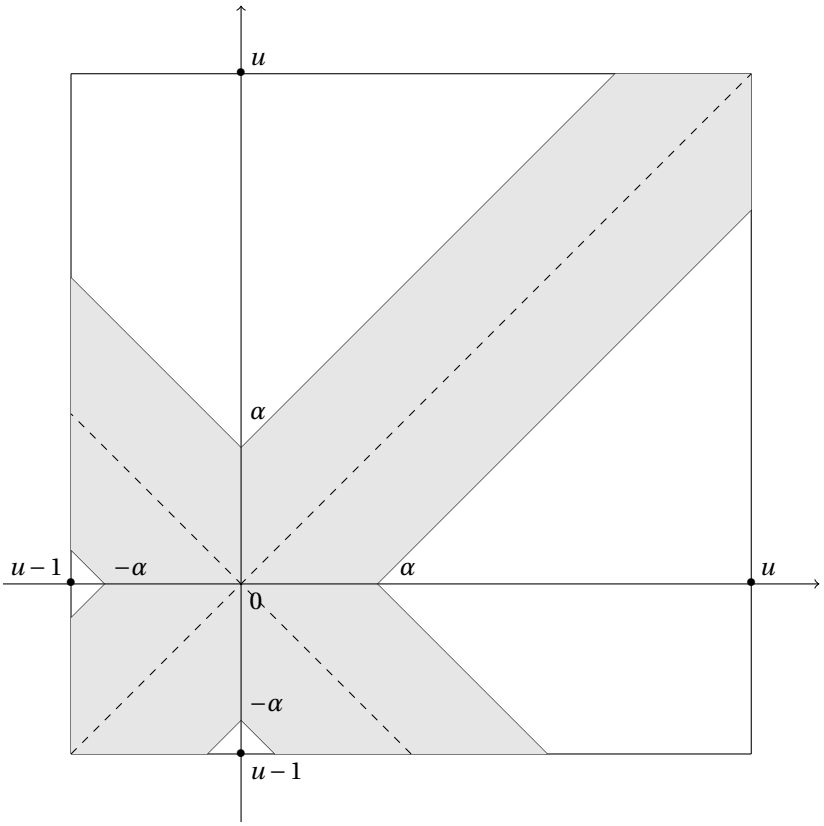


Figure 5.8: Domain $\Gamma_{u,\alpha}=\{(x_1,x_2)\in[0,1];\ ||x_1-u|-|x_2-u||\leqslant\alpha\}$ (in grey).

Analogously, one may derive suitable choices for n in some particular cases. For instance, we refer the reader to [11] to get upper bounds on $\mathbb{E}[W_q(\mu_X,\mu_{X,n})]$ for several values of $q\geqslant 1$ and several assumptions on the distribution on μ_X: general, uniform, Gaussian, beta, log concave, etc. Here are some results.

- In the general framework, the upper bound for $q\geqslant 1$ relies on the functional

$$J_q(\mu_X)=\int_{\mathbb{R}}\frac{\left(F_{\mu_X}(x)(1-F_{\mu_X}(x))\right)^{q/2}}{f_{\mu_X}(x)^{q-1)}}dx$$

  where F_{μ_X} is the c.d.f. associated to μ_X and f_{μ_X} its p.d.f. See [11, Theorems 3.2, 5.1 and 5.3].

- Assume that μ_X is uniformly distributed on $[0,1]$. Then by [11, Theorems 4.7, 4.8 and 4.9], for any $n\geqslant 1$,

$$\mathbb{E}[W_2(\mu_X,\mu_{X,n})^2]\leqslant\frac{1}{6n},$$

  for any $q\geqslant 1$ and for any $n\geqslant 1$,

$$\mathbb{E}[W_q(\mu_X,\mu_{X,n})^q]^{1/q}\leqslant(Const)\sqrt{\frac{q}{n}}.$$

  and for any $n\geqslant 1$,

$$\mathbb{E}[W_\infty(\mu_X,\mu_{X,n})]\leqslant\frac{(Const)}{n}.$$

  E.g. $(Const)=\sqrt{\pi/2}$.

- Assume that μ_X is a log-concave distribution with standard deviation σ. Then by [11, Corollaries 6.10 and 6.12], for any $1 \leqslant q < 2$ and for any $n \geqslant 1$,

$$\mathbb{E}[W_q(\mu_X, \mu_{X,n})^q] \leqslant \frac{(Const)}{2-q} \left(\frac{\sigma}{\sqrt{n}}\right)^q,$$

for any $n \geqslant 1$,

$$\mathbb{E}[W_2(\mu_X, \mu_{X,n})^2] \leqslant \frac{(Const)\sigma^2 \log n}{n},$$

and for any $q > 2$ and for any $n \geqslant 1$,

$$\mathbb{E}[W_q(\mu_X, \mu_{X,n})^q] \leqslant \frac{C_q \sigma^q}{n},$$

where C_q depends on q, only. Furthermore, if μ_X supported on $[a,b]$, then for any $n \geqslant 1$,

$$\mathbb{E}[W_2(\mu_X, \mu_{X,n})^2] \leqslant \frac{(Const)(b-a)^2}{n+1}.$$

E.g. $(Const) = 4/\ln 2$. Cf. [11, Corollary 6.11].

Example 5.20 - continued. We consider that X is uniformly distributed on $[0,1]$ and μ_X is a Gaussian distribution centered at X with unit variance. Then, by [11, Corollary 6.14], we have, for any $n \geqslant 3$,

$$\mathbb{E}[W_2(\mu_X, \mu_{X,n})^2] \leqslant \frac{(Const)\log\log n}{n},$$

and for any $q > 2$ and for any $n \geqslant 3$,

$$\mathbb{E}[W_q(\mu_X, \mu_{X,n})^q] \leqslant \frac{C_q}{n(\log n)^{q/2}},$$

where C_q depends only on q. Since we have already chosen $\delta(N) = o(N^{-1})$, it remains to take n so that $\log\log n/n = o(N^{-2})$ to fulfill the condition $\mathbb{E}[W_2(\mu_X, \mu_{X,n})] = o(\delta(N)/\sqrt{N})$.

5.5.5 Numerical study

Example 5.17 - continued. Here, we consider again the code given by (5.37) and we set $\overrightarrow{X} = (X_1, X_2, X_3)$ and $p = 3$. Having in mind the notation of Section 5.5.2, we consider the ideal code

$$\begin{array}{rccc} f: & E & \to & \mathcal{W}_q(E) \\ & (X_1, X_2, X_3) & \mapsto & \mu_{(X_1,X_2,X_3)} \end{array}$$

where $\mu_{(X_1,X_2,X_3)}$ is the uniform distribution on $[0, 1 + X_1 + X_2 + X_1X_3]$ for which the c.d.f. is $\mathbb{F}$ given by (5.37) and its stochastic counterpart

$$\begin{array}{rccc} f_s: & E \times D & \to & \mathbb{R} \\ & (X_1, X_2, X_3, D) & \mapsto & f_s(X_1, X_2, X_3, D) \end{array} \tag{5.46}$$

where $f_s(X_1, X_2, X_3, D)$ is a realization of $\mu_{(X_1,X_2,X_3)}$.
Hence, we no longer assume that one may observe N realizations of $\mathbb{F}$ associated to the N initial realizations of (X_1, X_2, X_3). Instead, for any of the N initial realizations of (X_1, X_2, X_3), we assess n realizations of a uniform random variable on $[0, 1 + X_1 + X_2 + X_1X_3]$.
In order to compare the estimation accuracy of the Pick-Freeze method and the rank-based method at a fixed size, we assume once again that only 450 calls of the computer code f are allowed to estimate the Fréchet indices $S^{\mathbf{u}}(\mathbb{F})$ and the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ for $\mathbf{u} = \{1\}$, $\{2\}$, and $\{3\}$. As in Example 5.17 of Section 5.4.4, the sample size allowed in the rank-based procedure is then $N = 450$ while, in the Pick-Freeze methodology, it is only $N = 112$ for the Fréchet indices $S^{\mathbf{u}}(\mathbb{F})$ and $N = 75$ for the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$. We only focus on the first-order indices since, as explained previously, the rank-based procedure has not been developed yet for higher-order indices. The empirical c.d.f. based on the empirical measures $\mu_{X_j,n}$ for $j = 1, \ldots, N$ in (5.41) are constructed with $n = 500$ evaluations. We repeat the estimation procedure $n_r = 200$ times. The boxplots of the mean square errors for the estimation of the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ have been plotted in Figure 5.9. We observe that, for a fixed total number of calls 450 to the code f (corresponding to a rank-based sample size $N = 450$ and to a Pick-Freeze sample size $N = 74$ for the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$), the rank-based estimation procedure performs much better than the Pick-Freeze method with significantly lower mean errors.

Another numerical study, in the particular setting of stochastic computer codes and inspired by [58], is considered in Section 5.6.3.

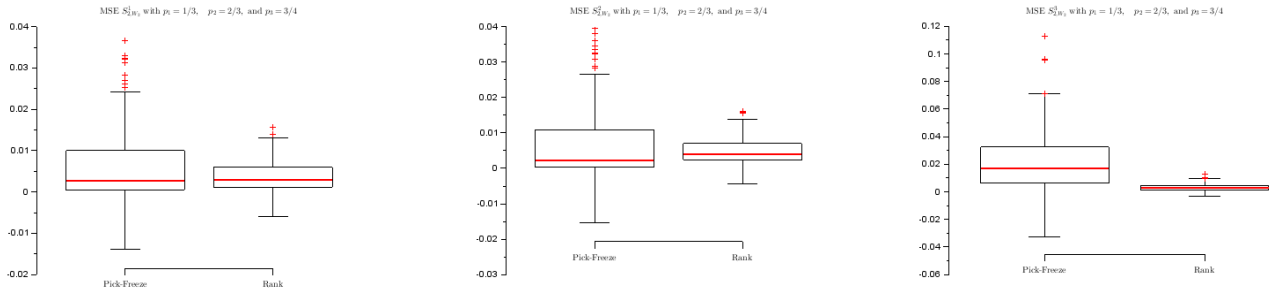


Figure 5.9: Model (5.46) with $p_1 = 1/3$, $p_2 = 2/3$, and $p_3 = 3/4$. Boxplot of the mean square errors of the estimation of the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ with a fixed sample size N, an approximation size n fixed at $n = 500$, and a number $n_r = 200$ of replications. The indices with respect to $\mathbf{u} = \{1\}$, $\{2\}$, and $\{3\}$ are displayed from left to right. The results of the Pick-Freeze estimation procedure with $N = 75$ for the Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ are provided in the left side of each graphic. The results of the rank-based methodology with $N = 450$ are provided in the right side of each graphic.

5.6 Sensitivity analysis with respect to the law of the inputs

This section deals with what is called second-level analysis and that corresponds to the SA with respect to the input distributions (rather than the inputs themselves). Before explaining our contributions in this framework, let us briefly describe its state of the art.

5.6.1 State of the art

The paper [83] is devoted to second-level uncertainty which corresponds to the uncertainty on the input distributions and/or on the parameters of the input distributions. As mentioned by the authors, such uncertainties can be handled in two different manners: (1) aggregating them with no distinction [26, 27] or (2) separating them [83]. In [26], the uncertainty concerns the parameters of the input distributions. The authors study the expectation with respect to the distribution of the parameters of the conditional output. In [27], the second-level uncertainties are transformed into first-level uncertainties considering the aggregated vector containing the input random variables vector together with the vector of uncertain parameters. Alternatively, in [83], the uncertainty brought by the lack of knowledge of the input distributions and the uncertainty of the random inputs are treated separately. A double Monte-Carlo algorithm is first considered. In the outer loop, a Monte-Carlo sample of input distribution is generated, while the inner loop proceeds to a GSA associated to each distribution. A more efficient algorithm is also proposed with a unique Monte-Carlo loop. The SA is then performed using the so-called Hilbert-Schmidt dependence measures (HSIC indices) on the input distributions rather than the input random variables themselves. See, e.g., [54] for the definition of the HSIC indices and more details on the algorithms.
In [84], a different approach is adopted. A failure probability is studied while the uncertainty concerns the parameters of the input distributions. An algorithm with low computational cost is proposed to handle such uncertainty together with the rare event setting. A single initial sample allows to compute the failure probabilities associated to different parameters of the input distributions. A similar idea is exploited in [72] in which the authors consider input perturbations and Perturbed-Law based Indices that are used to quantify the impact of a perturbation of an input p.d.f. on a failure probability. Analogously, the authors of [56, 58] are interested in (marginal) p.d.f. perturbations and the aim is to study the "robustness of the Sobol indices to distributional uncertainty and to marginal distribution uncertainty" which correspond to second-level uncertainty. For instance, the basic idea of the approach proposed in [56] is to view the total Sobol index as an operator which inputs the p.d.f. and returns the Sobol index. Then the analysis of robustness is done computing and studying the Fréchet derivative of this operator. The same principle is used in [58] to treat the robustness with respect to the marginal distribution uncertainty. Recently, [117] proposes a very clever approach of second-level SA when some moments of the distribution of the inputs are fixed. Its approach characterizes among all compactly supported input distribution with fixed first moments the range of variability of the Sobol indices.

5.6.2 Link with stochastic computer codes

We propose a new procedure that stems from the methodology in the context of stochastic computer codes described in Section 5.5. We denote by η_i $(i = 1, \dots, p)$ the distribution of the input X_i $(i = 1, \dots, p)$ in the model given by (5.14). There are several ways to model the uncertainty with respect to the choice of each η_i. Here we adopt the following framework. We assume that each η_i belongs to some family $\mathscr{P}_i$ of probability measures endowed with the probability

measure $\mathbb{P}_{\eta_i}$. In general, there might be measurability issues and the question of how to define a $\sigma-$field on some general spaces $\mathscr{P}_i$ can be tricky. We will restrict our study to the simple case where the existence of the probability measure $\mathbb{P}_{\eta_i}$ on $\mathscr{P}_i$ is given by the construction of the set $\mathscr{P}_i$. More precisely, we proceed as follows.

- First, for $1 \leqslant i \leqslant p$, let d_i be an integer and let $\Theta_i \subset \mathbb{R}^{d_i}$. Then consider the probability space $\left(\Theta_i, \mathscr{B}(\Theta_i), \nu_{\Theta_i}\right)$ where $\mathscr{B}(\Theta_i)$ is the Borel $\sigma-$field and ν_{Θ_i} is a probability measure on $(\Theta_i, \mathscr{B}(\Theta_i))$.

- Second, for $1 \leqslant i \leqslant p$, we consider an identifiable parametric set of probability measure $\mathscr{P}_i$ on E_i: $\mathscr{P}_i := \{\eta_\theta, \theta \in \Theta_i\}$. Let us denote by π_i the one-to-one mapping from Θ_i to $\mathscr{P}_i$ defined by $\pi_i(\theta) := \eta_\theta \in \mathscr{P}_i$ and define the $\sigma-$field $\mathscr{F}_i$ on $\mathscr{P}_i$ by

$$A \in \mathscr{F}_i \iff \exists B \in \mathscr{B}(\Theta_i),\ A = \pi_i(B).$$

  Then we endow this measurable space with the probability Π_i defined, for any $A \in \mathscr{F}_i$, by

$$\Pi_i(A) = \nu_{\Theta_i}\left(\pi_i^{-1}(A)\right).$$

- Third, in order to perform a second-level SA on (5.14), we introduce the stochastic mapping f_s from $\mathscr{P}_1 \times \ldots \times \mathscr{P}_p$ to $\mathscr{X}$ defined by

$$f_s\left(\eta_1, \ldots, \eta_p\right) = f(X_1, \ldots, X_p) \tag{5.47}$$

  where $(X_1, \ldots, X_p)$ is a random vector distributed as $\mu_1 \otimes \ldots \otimes \mu_p$. Hence f_s is a stochastic computer code from $\mathscr{P}_1 \times \ldots \times \mathscr{P}_p$ to $\mathscr{X}$ and once the probability measures $\mathbb{P}_{\eta_i}$ on each $\mathscr{P}_i$ are defined, we can perform SA using the framework of Section 5.5.

5.6.3 Numerical study

As in [58], let us consider the synthetic example defined on $[0,1]^3$ by

$$f(X_1, X_2, X_3) = 2X_2 e^{-2X_1} + X_3^2. \tag{5.48}$$

We are interested in the uncertainty in the support of the random variables X_1, X_2, and X_3. To do so, we follow the notation and framework of [58]. For $i = 1, 2$, and 3, we assume that X_i is uniformly distributed on the interval $[A_i, B_i]$, where A_i and B_i are themselves uniformly distributed on $[0, 0.1]$ and $[0.9, 1]$ respectively. As remarked in [58], it seems natural that f will vary more in the X_2-direction when X_1 is close to 0 and less when X_1 is close to 1.
As mentioned in Section 5.6.1, the authors of [58] view the total Sobol index as an operator which inputs the p.d.f. and returns the total Sobol index. Then they study the Fréchet derivative of this operator and determine the most influential p.d.f., which depends on a parameter denoted by δ. Finally, they make the parameter δ vary.
Here, we adopt the methodology explained in the previous section (Section 5.6.2). Namely, we consider the stochastic computer code given by

$$f_s(\eta_1, \eta_2, \eta_3) = 2X_2 e^{-2X_1} + X_3^2, \tag{5.49}$$

where the X_i's are independently drawn according to the uniform measure η_i on $[A_i, B_i]$ with A_i and B_i themselves uniformly distributed on $[0, 0.1]$ and $[0.9, 1]$ respectively. Then to estimate the indices $S^{\mathbf{u}}_{2,W_2}$, for $\mathbf{u} = \{1\}$, $\{2\}$, $\{3\}$, $\{1,2\}$, $\{1,3\}$, and $\{2,3\}$, we proceed as follows.

1. For $i = 1$, 2, and 3,
   (a) we produce a N-sample $\left([A_{i,j}, B_{i,j}]\right)_{j=1,\ldots,N}$ of intervals $[A_i, B_i]$.
   (b) for $j = 1, \ldots, N$,
      i. we generate a n-sample $\left(X_{i,j,k}\right)_{k=1,\ldots,n}$ of X_i, where $X_{i,j,k}$ is uniformly distributed on $[A_{i,j}, B_{i,j}]$.
      ii. we compute the n-sample $\left(Z_{j,k}\right)_{k=1,\ldots,n}$ of the output using

$$Z = f(X_1, X_2, X_3) = 2X_2 e^{-2X_1} + X_3^2.$$

   Thus we get a N-sample of the empirical measures of the distribution of the output Z given by

$$\mu_{Z_j,n} := \frac{1}{n}\sum_{k=1}^{n} \delta_{Z_{j,k}}, \quad \text{for } j = 1, \ldots, N.$$

   (c) We order the intervals $\left([A_{i,j}, B_{i,j}]\right)_{j=1,\ldots,N}$ and we get the Pick-Freeze versions of Z to treat the SA regarding the input $\mathbf{u}$.

2. Finally, it remains to compute the indicators of the empirical version of (5.36) using (5.44) and their means to get the Pick-Freeze estimators of $S^{\mathbf{u}}_{2,W_2}$.

Notice that we only consider the estimators based on the Pick-Freeze method since we allow for both bounds of the interval to vary and, as explained previously, the rank-based procedure has not been developed yet, neither for higher-order indices nor in higher dimensions.

Simulations First, we compute the estimators of $S^{\mathbf{u}}_{2,W_2}$ following the previous procedure with a sample size $N = 500$ and an approximation size $n = 500$. The results are displayed in Table 5.5 (first row). We also perform another batch of simulations allowing for higher variability on the bounds: for $i = 1$, 2, and 3, A_i is now uniformly distributed on $[0, 0.45]$ while B_i is now uniformly distributed on $[0.55, 1]$. The results are displayed in Table 5.5 (second row).

	$\mathbf{u}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
$A_i \in [0, 0.1]$ $B_i \in [0.9, 1]$	$\hat{S}^{\mathbf{u}}_{2,W_2}$	0.07022	0.08791	0.09236	0.14467	0.21839	0.19066
$A_i \in [0, 0.45]$ $B_i \in [0.55, 1]$	$\hat{S}^{\mathbf{u}}_{2,W_2}$	0.11587	0.06542	0.169529	0.22647	0.40848	0.34913

Table 5.5: Model (5.48). GSA on the parameters of the input distributions. Estimations of $S^{\mathbf{u}}_{2,W_2}$ with a sample size $N = 500$ and an approximation size $n = 500$. In the first row, for $i = 1$, 2, and 3, A_i is uniformly distributed on $[0, 0.1]$ while B_i is uniformly distributed on $[0.9, 1]$. In the second row, we allow for more variability: for $i = 1$, 2, and 3, A_i is uniformly distributed on $[0, 0.45]$ while B_i is uniformly distributed on $[0.55, 1]$.

Second, we run another simulation allowing for more variability on the upper bound related to the third input X_3 only: B_3 is uniformly distributed on $[0.5, 1]$ (instead of $[0.9, 1]$). For $i = 1$ and 2, A_i is still uniformly distributed on $[0, 0.1]$ while B_i is still uniformly distributed on $[0.9, 1]$. The results are displayed in Table 5.6. We still use a sample size $N = 500$ and an approximation size $n = 500$.

$\mathbf{u}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
$\hat{S}^{\mathbf{u}}_{2,W_2}$	0.01196	0.06069	0.56176	-0.01723	0.63830	0.59434

Table 5.6: Model (5.48). GSA on the parameters of the input distributions. Estimations of $S^{\mathbf{u}}_{2,W_2}$ with a sample size $N = 500$ and an approximation size $n = 500$ and more variability on B_3, now uniformly distributed on $[0.5, 1]$. For $i = 1$ and 2, A_i is still uniformly distributed on $[0, 0.1]$ while B_i is still uniformly distributed on $[0.9, 1]$.

Third, the aim is to highlight the fact that performing a classical GSA differs from performing a second-level SA. In that view, we perform a classical GSA on the inputs rather than on the parameters of their distributions (corresponding to a second-level analysis). Namely, we consider the index $S^{\mathbf{u}}_{2,\mathrm{CVM}}$ and proceed to its estimation with a sample size $N = 10^4$. The reader is referred to [49, Section 3] for the definition of this index $S^{\mathbf{u}}_{2,\mathrm{CVM}}$ and its Pick-Freeze estimator together with their properties. The results are displayed in Table 5.7.

$\mathbf{u}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
$\hat{S}^{\mathbf{u}}_{2,\mathrm{CVM}}$	0.13717	0.15317	0.33889	0.33405	0.468163	0.53536

Table 5.7: Model (5.48). Direct GSA on the inputs. Estimations of $S^{\mathbf{u}}_{2,\mathrm{CVM}}$ with a sample size $N = 10^4$. The reader is referred to [49, Section 3] for the definition of the index $S^{\mathbf{u}}_{2,\mathrm{CVM}}$ and its Pick-Freeze estimator together with their properties.

Comments When one is interested in the choice of the input distributions of X_1, X_2, and X_3, the first row in Table 5.5 shows that each choice is equally important. Now, if we give more freedom to the space where the distribution lives, the relative importance may change as one can see in Table 5.5 (second row) and in Table 5.6. More precisely, in Table 5.6, the variability of the third input distribution (namely, the variability of its upper bound) is five times larger than the other variabilities. Not surprisingly, it results that the importance of the choice of the third input distribution is then much more important than the choices of the distributions of the two first inputs.

As said in the previous paragraph, when one is interested in the choice of the input distributions of X_1, X_2, and X_3, the first row in Table 5.5 shows that each choice is equally important. Nevertheless, performing a classical GSA on the inputs using the Cramér-von-Mises index for example, we see, in Table 5.7, that the index related to X_3 is more than twice as important as X_1 and X_2 (when considering only first-order effects). Hence, here, the classical GSA largely differs numerically from a second-level SA as expected.

5.7 Exercices

Exercise 11. *Consider X_1 and X_2 two independent standard Gaussian variable and*

$$Z = f(X_1, X_2) = 2X_1 + 3X_1X_2 + X_2.$$

Make a program that computes $\widehat{N}^u_{2,CVM}$ defined in (5.11).

Exercise 12. *Let us consider the quite simple linear model*

$$Y = \alpha X_1 + X_2, \ \alpha > 0,$$

where X_1 has a Bernoulli distribution with success probability $0 < p < 1$ and X_1, X_2 are independent. Assume further that X_2 has a continuous distribution F_2 on $\mathbb{R}$ such that $\mathbb{E}[X_2] = \alpha p$ and with finite variance $\mathrm{Var}(X_2) = \alpha^2 p(1-p)$.

1. *Compute the first order Sobol indices of X_1 and X_2.*
2. *Determine the the distribution of Y given $X_1 = 0$ and the distribution of Y given $X_1 = 1$. Deduce the distribution of Y.*
3. *Compute $S^1_{2,CVM}$ and $S^2_{2,CVM}$. Comment.*
4. *Compute H^q_1 and H^q_2.*
5. *Application in the following particular cases:*
   - *(i) X_2 is a centered Gaussian random variable with variance $\mathrm{Var}(X_2) = \alpha^2 p(1-p)$;*
   - *(ii) X_2 is uniformly distributed on $[0, b]$ with $b = 2\alpha\sqrt{3p(1-p)}$;*
   - *(iii) X_2 is exponentially distributed with mean $1/\lambda = \alpha\sqrt{p(1-p)}$.*
6. *Propose a numerical illustration with $N = 100$ and 500.*

Exercise 13. *Let us consider the following nonlinear model*

$$Y = \exp\{X_1 + 2X_2\}, \tag{5.50}$$

where X_1 and X_2 are independent standard Gaussian random variables. The distribution of Y is said to be log-normal.

1. *Derive both its density and its distribution functions.*
2. *Compute the indices of order q ($q \geqslant 2$):*

$$H^q_1 = \mathbb{E}\left[(e^{X_1+2} - e^{5/2})^q\right] \quad \text{and} \quad H^q_2 = \mathbb{E}\left[(e^{2X_1+1/2} - e^{5/2})^q\right].$$

3. *Prove that the Cramér-von Mises indices $S^1_{2,CVM}$ and $S^2_{2,CVM}$ are given by*

$$S^1_{2,CVM} = \frac{6}{\pi}\arctan 2 - 2 \approx 0.1145 \quad \text{and} \quad S^2_{2,CVM} = \frac{6}{\pi}\arctan\sqrt{19} - 2 \approx 0.5693.$$

Chapter 6

Beyond variance-based indices: other extensions

As emphasized previously, Sobol' indices are based on the second-order moment (i.e. the variance) of the output Y. However, in some cases, variance poorly represents the variability of the distribution. In Chapter 5, we investigate In Sections 6.1 and 6.2, we present different ways to consider other features on the output, providing generalization frameworks for Sobol' indices. It also happens that the quantity of interest is not related to the overall variability of the model output. In such cases, sensitivity analysis techniques have then to be adapted: goal-oriented sensitivity analysis tools are presented in Section 6.3.

6.1 Moment-independent importance measures

6.1.1 Introduction

As seen in Chapter 5, a first attempt to define a moment-independent importance measure has been addressed using the Cramér-von Mises distance and leading to the Cramér-von Mises indices defined in (5.10).

Alternative definitions for measuring the strength of the statistical dependence of Y on one input X_i and relying on distribuaitons rather than moments have been proposed by Borgonovo and his co-authors giving rise to the class of distribution-based sensitivity measures. Assuming all considered random variables have an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{R}$, they define the importance of X_i as the distance between the unconditional distribution of Y and its conditional distribution (see [13] for a review). The first representative of this class is the δ-importance measure, which is based on the L^1-norm between densities [12]:

$$\delta_i = \mathbb{E}\Big[\int \big|f_Y(y) - f_{Y|X_i}(y)\big|\, dy\Big] = \iint \big|f_Y(y) f_{X_i}(x_i) - f_{Y,X_i}(y,x_i)\big|\, dy dx_i, \tag{6.1}$$

where f_{X_i}, f_Y, and f_{Y,X_i} stand for the marginal densities of X_i and Y and the joint density of the pair (Y, X_i). Clearly, $\delta_i = 0$ if and only if X_i and Y are independent.
Observe that the integrand $\big|f_Y(y) - f_{Y|X_i}(y)\big|$ is the total variation distance between the distribution $\mathbb{P}_Y$ of Y and the conditional distribution $\mathbb{P}_{Y|X_i}$ of Y by X_i.

The integral $\int \big|f_Y(y) - f_{Y|X_i}(y)\big|\, dy$ can be rewritten as

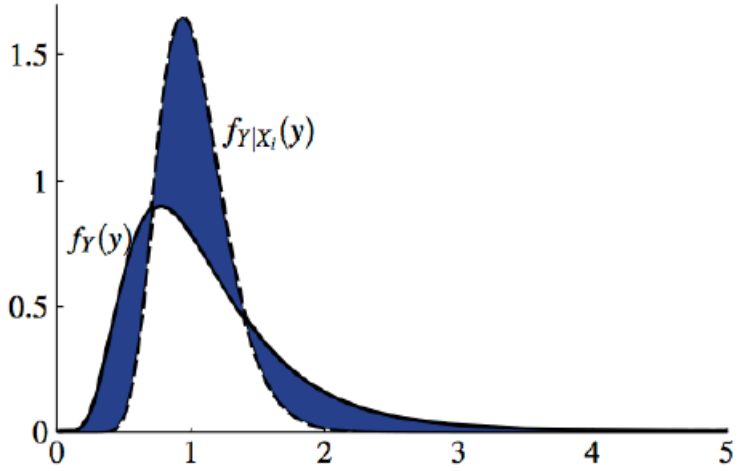
$$\int g\Big(\frac{f_Y(y)}{f_{Y|X_i}(y)}\Big) f_{Y|X_i}(y) dy \tag{6.2}$$

in terms of the convex function g defined by $g(t) = |t-1|$ and verifying $g(1) = 0$. Then one may consider another convex functions g such that $g(1) = 0$. For instance, $g(t) = -\ln t$ or $g(t) = t \ln t$ that both leads to Kullback-Leibler divergences. In the first case, one gets the following index:

$$\theta_i = \mathbb{E}\Big[\int \big(\ln f_{Y|X_i}(y) - \ln f_Y(y)\big) f_{Y|X_i}(y) dy\Big] = \mathbb{E}\Big[D_{KL}\big(\mathbb{P}_{Y|X_i} || \mathbb{P}_Y\big)\Big]. \tag{6.3}$$

In the latter case, one gets

$$\theta_i' = \mathbb{E}\Big[\int \big(\ln f_Y(y) - \ln f_{Y|X_i}(y)\big) f_Y(y) dy\Big] = \mathbb{E}\Big[D_{KL}\big(\mathbb{P}_Y || \mathbb{P}_{Y|X_i}\big)\Big]. \tag{6.4}$$



Another representative is the β_i sensitivity measure, based on the Kolmogorov-Smirnov separation between cumulative distribution functions:

$$\beta_i = \mathbb{E}\Big[\sup_y \big|F_Y(y) - F_{Y|X_i}(y)\big|\Big]. \tag{6.5}$$

The previous indices are defined in terms of the f-divergence between the distribution of Y and the conditional distribution of Y by X_i.

6.1.2 New focus on Sobol' indices: general indices built on a dissimilarity measure

Let us first come back to the unnormalized first-order Sobol' sensitivity index for a given input variable X_i and rewrite it as

$$\mathrm{Var}\,\mathbb{E}[Y|X_i) = \mathbb{E}\Big[\mathbb{E}Y - \mathbb{E}[Y|X_i)\Big]^2. \tag{6.6}$$

This alternative formulation provides a genuine intuition of what first-order Sobol' indices measure: they compare the output distribution P_Y and the output distribution conditionally to X_i, $\mathrm{P}_{Y|X_i}$, through their mean values and integrate the squared difference to get the unnormalized index.
A straightforward extension then readily consists in replacing the comparison of the means by some general dissimilarity measure $d(\cdot,\cdot)$ between probability distributions to propose a new sensitivity index defined as

$$\mathbb{E}\Big[d\left(\mathrm{P}_Y, \mathrm{P}_{Y|X_i}\right)\Big]. \tag{6.7}$$

Finally, the advantage of such a general formulation is that many choices for d are available: we see in what follows that using natural dissimilarity measures yields sensitivity indices related to well known quantities, while others possess theoretical properties that make it possible to build efficient estimators as well as enjoyable additional characteristics.

Taking the Cramér-von Mises distance in (6.7) leads to the Cramér-von Mises indices already studied in Chapter 5.

6.1.3 The F-divergence case

Coming back to the choice of the dissimilarity measure $d(\cdot,\cdot)$, a popular choice is the class of Csiszár f-divergences $d(\cdot,\cdot) = d_f(\cdot||\cdot)$ [29]. Assuming all considered random variables have an absolutely continuous distribution with respect to the Lebesgue measure on $\mathbb{R}$, the f-divergence between P_Y and $\mathrm{P}_{Y|X_i}$ is given by

$$d_f(\mathrm{P}_Y||\mathrm{P}_{Y|X_i}) = \int_{\mathbb{R}} f\left(\frac{p_Y(y)}{p_{Y|X_i}(y)}\right) f_{Y|X_i}(y)\,dy \tag{6.8}$$

where f is any convex function such that $f(1) = 0$. Standard choices for function f include for example:

- the Kullback-Leibler divergence $f(t) = -\ln(t)$ or $f(t) = t\ln(t)$,
- the Hellinger distance: $f(t) = \left(\sqrt{t}-1\right)^2$,

- the total variation distance: $f(t) = |t-1|$,
- the Pearson χ^2 divergence: $f(t) = (t-1)^2$ or $f(t) = t^2 - 1$,
- the Neyman χ^2 divergence: $f(t) = (t-1)^2/t$ or $f(t) = (1-t^2)/t$.

Plugging this dissimilarity measure in (6.7) and deconditioning with respect to X_i yields the following sensitivity index

$$\int_{\mathbb{R}^2} f\left(\frac{p_Y(y) f_{X_i}(x)}{p_{X_i,Y}(x,y)}\right) p_{X_i,Y}(x,y)\,dx\,dy \tag{6.9}$$

Note that inequalities on Csiszár f-divergences imply that such sensitivity indices are positive and equal zero when Y and X_i are independent. Also, it is important to note that, given the form of $Sr_i^{d_f}$, it is invariant under any smooth and uniquely invertible transformation of the variables X_i and Y (see the proof for mutual information in Kraskov [66]). This is an enjoyable property over variance-based Sobol' sensitivity indices, which are only invariant under linear transformations for a scalar output.

It is easy to see that the total variation distance with $f(t) = |t-1|$ gives a sensitivity index equal to δ_i defined in (6.1). In addition, the Kullback-Leibler divergence with $f(t) = -\ln(t)$ yields θ_i' defined in (6.4) that is the *m*utual information between X_i and Y, denoted $I(X_i;Y)$. A normalized version of this sensitivity index was extensively studied [66, 74, 95]. Similarly, the Neyman χ^2 divergence with $f(t) = (1-t^2)/t$ leads to

$$\int_{\mathbb{R}^2} \left(\frac{p_{X_i,Y}(x,y)}{p_Y(y)p_{X_i}(x)} - 1\right)^2 p_Y(y)p_{X_i}(x)\,dx\,dy,$$

which is the so-called *s*quared-loss mutual information between X_i and Y (or *mean square contingency*). These results show that some previously proposed sensitivity indices are actually special cases of more general indices defined through Csiszár f-divergences. An extensive study of their properties is given in Rahman [99]. Moreover, the specific structure of (6.9) makes it possible to envision more efficient tools for the estimation of these sensitivity indices. Indeed, it only involves approximating a density ratio rather than full densities. This point is investigated below. But more importantly, we see that special choices for f define sensitivity indices that are actually well-known dependence measures such as the mutual information. This paves the way for looking at new sensitivity indices relying on recent kernel-based dependence measures (see Section 6.2).

Estimation

Coming back to (6.9), the goal is to estimate

$$\int_{\mathbb{R}^2} f\left(\frac{1}{r(x,y)}\right) p_{X_i,Y}(x,y)\,dx\,dy = \mathbb{E}_{(X_i,Y)} f\left[\frac{1}{r(X_i,Y)}\right]$$

where $r(x,y) = p_{X_i,Y}(x,y)/(p_Y(y)p_{X_i}(x))$ is the ratio between the joint density of (X_i,Y) and the marginals. Of course, straightforward estimation is possible if one estimates the densities $p_{X_i,Y}(x,y)$, $p_{X_i}(x)$ and $p_Y(y)$ with, e.g., kernel density estimators. Assuming that we have a sample $\mathbf{Z}^n = (\mathbf{X}_i^n, \mathbf{Y}^n) = \left(X_i^{(j)}, Y^{(j)}\right)_{j=1\ldots n}$, the kernel density estimate of $p_{X_i,Y}(x,y)$ is given by

$$\hat{p}_{X_i,Y}(x,y) = \frac{1}{n}\sum_{j=1}^{n} K_H(z - Z^{(j)})$$

for $z = (x,y)$ and $K_H(z) = |H|^{-1/2}K(H^{-1/2}z)$ where K is a multivariate kernel and H is the bandwidth matrix. We similarly get $\hat{p}_{X_i}(x)$ and $\hat{p}_Y(y)$, the kernel density estimates of $p_{X_i}(x)$ and $p_Y(y)$, respectively. We can deduce the following plug-in estimators for the ratio and the sensitivity index

$$\hat{r}(x,y) = \frac{\hat{p}_{X_i,Y}(x,y)}{\hat{p}_Y(y)\hat{p}_{X_i}(x)}, \qquad \hat{S}_{X_i}^{d_f} = \frac{1}{n}\sum_{j=1}^{n} f\left(\frac{1}{\hat{r}\left(X_j^{(j)}, Y^{(j)}\right)}\right).$$

If we only consider the case of a scalar output and each input factor separately, this plug-in estimator can easily be computed. However kernel density estimation suffers from the curse of dimensionality and the estimator proposed above cannot be used when dealing with multidimensional outputs or group of inputs. For such a general setting, one can observe that only the ratio function $r(x,y)$ is needed: we can then expect more robust estimates by focusing only on it. Powerful estimation methods for such ratios include, e.g., maximum-likelihood estimation [118], unconstrained least-squares importance fitting [63]. A k-nearest neighbors strategy dedicated to mutual information is also discussed in Kraskov [66].

Unfortunately, all proposed estimation procedures lead to sub-optimal estimators which exhibit a non-parametric rate of convergence. A potential progress towards the design of efficient estimators may come from adaptation of the work by Laurent [69] or Giné and Nickl [53].

6.2 Kernel-based sensitivity analysis

The dissimilarity point of view introduced above, although appealing (especially with the Cramér-von Mises distance properties), still only provides a complementary perspective to Sobol' indices. In particular, none of the importance measures introduced so far enjoys the highly desirable FANOVA decomposition property. Such a decomposition not only provides a proper normalization constant for sensitivity indices, but also appropriately defines the concept of higher-order effects.
In this section, starting from another general class of dissimilarities between probability distributions, we introduce recent advances on distances based on kernel representations of probability distributions. This framework can be seen as a straightforward generalization of Sobol' indices with profitable characteristics: estimation is far less prone to the curse of dimensionality and a FANOVA-like decomposition is readily available.

6.2.1 Integral probability metrics and MMD distance

Integral Probability Metrics (IPM) [87] are a popular family of distance measures on probabilities given by

$$\gamma_{\mathscr{F}}(\mathrm{P},\mathrm{Q}) = \sup_{f\in\mathscr{F}} \left| \int_S f\,d\mathrm{P} - \int_S f\,d\mathrm{Q} \right| \tag{6.10}$$

for two probability measures P and Q defined on a measurable space S and where $\mathscr{F}$ is a class of real-valued bounded measurable functions on S. Just as the choice of function F in Csiszár F-divergences gives rise to different measures, the choice of $\mathscr{F}$ generates different IPM, e.g., the Wasserstein distance, the Dudley metric or the total variation distance. It is interesting to note that Csiszár F-divergences and IPM are very distinct classes of dissimilarity measures, since they only intersect at the total variation distance [115]. Notable examples of $\mathscr{F}$ and associated distances include:

- The space of bounded continuous functions for the Dudley metric;
- The space of functions with bounded variations for the Kolmogorov metric;
- The space of Lipschitz bounded functions for the Wasserstein distance;
- The space of characteristic functions on Borel sets for the total variation distance.

Except for the total variance case, plugging the general expression (6.10) of an IPM in (6.7) no longer yields a closed-form expression for a sensitivity index as in (6.9).
Recently, in a rich sequence of papers, several authors focused on another function class $\mathscr{F}$, namely the space of functions in the unit ball of a characteristic Reproducing Kernel Hilbert Space (RKHS) which gives rise to the Maximum Mean Discrepancy (MMD) distance [108].

6.2.2 Reproducing Kernel Hilbert Space

Let $\mathscr{X}$ be an arbitrary set and $\mathscr{H}$ a Hilbert space of real-valued functions $f : \mathscr{X} \to \mathbb{R}$ on $\mathscr{X}$ with inner product $\langle\cdot,\cdot\rangle_{\mathscr{H}}$ (denoted $\langle\cdot,\cdot\rangle$ in what follows). For every $x \in \mathscr{X}$, we define the *evaluation functional* $L_x : \mathscr{H} \to \mathbb{R}$ as $f \mapsto L_x(f) = f(x)$.

Definition 6.1. *A Hilbert space $\mathscr{H}$ is a Reproducing Kernel Hilbert Space (RKHS) if the evaluation functionals are continuous.*

More details on RKHS can be found in the monograph of Berlinet and Thomas-Agnan [9].

Definition 6.2. *A function $k : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ is a kernel if*

1. *k is symmetric, i.e. $\forall x, x' \in \mathscr{X},\ k(x,x') = k(x',x)$*
2. *k is positive definite, i.e. for any $n \in \mathbb{N}$, $x_1,\dots,x_n \in \mathscr{X}$ and $\alpha_1,\dots,\alpha_n \in \mathbb{R}$*

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j k(x_i,x_j) \geqslant 0$$

Definition 6.3. *The reproducing kernel of a RKHS $\mathcal{H}$ is a function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that*

1. $\forall x \in \mathcal{X},\ k(x,\cdot) \in \mathcal{H}$
2. $\forall f \in \mathcal{H}$ *and* $\forall x \in \mathcal{X},\ \langle f, k(x,\cdot)\rangle = f(x)$ *(reproducing property).*

In particular, the reproducing kernel is a kernel, i.e. it is symmetric and positive definite.

Theorem 6.4. *Every symmetric positive definite kernel defines a unique RHKS, of which it is the reproducing kernel.*

Theorem 6.5 (Mercer, see [3])**.** *Suppose k is a symmetric positive definite kernel on a compact set $\mathcal{X}$ and consider the integral operator $T_k : \mathbb{L}^2(\mathcal{X}) \to \mathbb{L}^2(\mathcal{X})$ defined by*

$$\left(T_k f\right)(x) = \int_{\mathcal{X}} k(x,u) f(u)\, du.$$

Then there is an orthonormal basis $\{e_i\}$ of $\mathbb{L}^2(\mathcal{X})$ consisting of eigenfunctions of T_k such that the corresponding sequence of eigenvalues $\{\lambda_i\}$ are non-negative. The eigenfunctions corresponding to non-zero eigenvalues are continuous on $\mathcal{X}$ and k has the following representation

$$k(x,x') = \sum_{i=1}^{\infty} \lambda_i e_i(x) e_i(x')$$

where the convergence is absolute and uniform.

Theorem 6.6 (Representer's theorem, see [107])**.** *Let $\mathcal{H}$ be a RKHS with kernel k and consider the minimization problem*

$$\min_{f \in \mathcal{H}} \sum_{i=1}^{n} l(f(x_i), y_i) + \lambda \|f\|^2$$

where $l(\cdot,\cdot)$ is any loss function and $\lambda > 0$. Then, if $l(\cdot,\cdot)$ is convex with respect to its second argument, the solution of the minimization problem is unique and writes

$$f^*(x) = \sum_{i=1}^{n} a_i k(x_i, x)$$

Definition 6.7. *Let ξ denote a random variable on $\mathcal{X}$ with probability distribution P_ξ and $\mathcal{H}$ a RKHS with kernel k. The kernel embedding of the distribution P_ξ is the function in $\mathcal{H}$ given by the kernel mean*

$$\mu_{\mathrm{P}_\xi} = \mathbb{E}_\xi[k(\xi,\cdot)]$$

Definition 6.8. *A kernel is characteristic if the kernel embedding $\mu : \mathcal{M}_1^+ \to \mathcal{H}$ is injective.*

MMD distance via kernel embedding of distributions

When particularizing $\mathscr{F}$ as the unit ball of a RKHS $\mathscr{H}$ in (6.10), we can write

$$\begin{aligned}
\mathrm{MMD}^2(\mathrm{P},\mathrm{Q};\mathscr{H}) := \gamma_{\mathscr{F}}(\mathrm{P},\mathrm{Q})^2 &= \left(\sup_{f\in\mathscr{H},\ \|f\|\leqslant 1} \int_S f d\mathrm{P} - \int_S f d\mathrm{Q}\right)^2 \\
&= \left(\sup_{f\in\mathscr{H},\ \|f\|\leqslant 1} \mathbb{E}_{\xi\sim\mathrm{P}}[f(\xi)] - \mathbb{E}_{\zeta\sim\mathrm{Q}}[f(\zeta)]\right)^2 \\
&= \left(\sup_{f\in\mathscr{H},\ \|f\|\leqslant 1} \mathbb{E}_{\xi\sim\mathrm{P}}[\langle f, k(\xi,\cdot)\rangle] - \mathbb{E}_{\zeta\sim\mathrm{Q}}[\langle f, k(\zeta,\cdot)\rangle]\right)^2 \ \begin{pmatrix}\text{reproducing}\\ \text{property}\end{pmatrix} \\
&= \left(\sup_{f\in\mathscr{H},\ \|f\|\leqslant 1} \langle f, \mathbb{E}_{\xi\sim\mathrm{P}}[k(\xi,\cdot)] - \mathbb{E}_{\zeta\sim\mathrm{Q}}[k(\zeta,\cdot)]\rangle\right)^2 \ \begin{pmatrix}\text{linearity of}\\ \text{inner product}\end{pmatrix} \\
&= \left(\sup_{f\in\mathscr{H},\ \|f\|\leqslant 1} \langle f, \mu_{\mathrm{P}} - \mu_{\mathrm{Q}}\rangle\right)^2 \ \begin{pmatrix}\text{kernel embedding}\\ \text{definition}\end{pmatrix} \\
&= \|\mu_{\mathrm{P}} - \mu_{\mathrm{Q}}\|^2 \ \text{(definition of the norm)} \\
&= \mathbb{E}_{\xi,\xi'\sim\mathrm{P}}[k(\xi,\xi')] + \mathbb{E}_{\zeta,\zeta'\sim\mathrm{Q}}[k(\zeta,\zeta')] - 2\mathbb{E}_{\xi\sim\mathrm{P},\zeta\sim\mathrm{Q}}[k(\xi,\zeta)] \ \begin{pmatrix}\text{reproducing}\\ \text{property}\end{pmatrix}
\end{aligned} \tag{6.11}$$

assuming $\mathbb{E}_{\xi,\xi'\sim\mathrm{P}}[k(\xi,\xi')] < \infty$ and $\mathbb{E}_{\zeta,\zeta'\sim\mathrm{Q}}[k(\zeta,\zeta')] < \infty$. Here, $\|\cdot\|$ is the Hilbertian norm (under $\mathscr{H}$) and $\mu_{\mathrm{P}} = \mathbb{E}_{\xi\sim\mathrm{P}}[k(\xi,\cdot)]$ is the kernel embeddings of P.

The MMD is a distance if, and only if, the kernel k is characteristic (i.e., the kernel embedding is injective) in order to ensure that two different probability distributions are not mapped to the same function in the RKHS. Most kernels used in the machine learning community such as the Gaussian kernel, the exponential kernel or the class of Matérn kernels are characteristic. However, polynomial kernels $k(y,y') = \left(yy' + \alpha\right)^d$ are not characteristic, since probability distributions with the same moments up to order d but with different higher-order ones are mapped to the same function. In particular, the vanilla linear kernel $k(y,y') = yy'$ is worth mentioning, since in this case the MMD boils down to a simple comparison of the means, i.e., $\mathrm{MMD}^2(\mathrm{P},\mathrm{Q};\mathscr{H}) = \left(\mathbb{E}_{\xi\sim\mathrm{P}}[\xi] - \mathbb{E}_{\zeta\sim\mathrm{Q}}[\zeta]\right)^2$, which obviously does not define a distance on probability distributions.

MMD estimation

The kernel formulation (6.11) is easily used to define an estimator for $\mathrm{MMD}^2(\mathrm{P},\mathrm{Q};\mathscr{H})$ given by

$$\widehat{\mathrm{MMD}}^2(\mathrm{P},\mathrm{Q};\mathscr{H}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \left[k(\xi_i,\xi_j) - k(\xi_i,\zeta_j) - k(\zeta_i,\xi_j) + k(\zeta_i,\zeta_j)\right]$$

where $\{\xi_i\}_{i=1\ldots n} \sim \mathrm{P}$ and $\{\zeta_i\}_{i=1\ldots n} \sim \mathrm{Q}$. This is an unbiased one-sample U-statistic estimate, but a biased V-statistic has also been proposed.

Sensitivity index based on the MMD - distance between $\mathbf{P}_Y$ and $\mathbf{P}_{Y|\mathbb{X}_B}$

Recall that the variance decomposition states that the variance of the output can be decomposed as $\mathrm{Var}\, Y = \sum_{A\subset\{1,\ldots,d\}} V_A$ where each term is given by

$$V_A = \sum_{B\subset A} (-1)^{|A|-|B|} \mathrm{Var}(\mathbb{E}[Y|\mathbb{X}_B]).$$

The MMD-based equivalent decomposition is described in the following theorem [33].

Assumption 6.2.1. *$\forall A \subset \{1,\ldots,d\}$ and $\mathrm{P}_{\mathbb{X}_A}$-almost all $\mathbf{x}_A \in \mathscr{X}_A$, $\mathbb{E}_{\xi\sim\mathrm{P}_{Y|\mathbb{X}_A=\mathbf{x}_A}}[k_{\mathscr{Y}}(\xi,\xi)] < \infty$ with the convention $\mathrm{P}_{Y|\mathbb{X}_A} = \mathrm{P}_Y$ if $A = \emptyset$.*

Theorem 6.9. *Assume the same assumptions as the one assumed in Theorem 1.3 (in particular, the random vector $\mathbb{X}$ has independent components). Assume further that Assumption 6.2.1 and Mercer's theorem holds. Let denote $\mathrm{MMD}^2_{tot} = \mathbb{E}[k(Y,Y)] - \mathbb{E}[k(Y,Y')]$ where Y' is an independent copy of Y. Then, the total MMD can be decomposed as*

$$\mathrm{MMD}^2_{tot} = \sum_{A \subset \{1,\dots,d\}} \mathrm{MMD}^2_A$$

where each term is given by

$$\mathrm{MMD}^2_A = \sum_{B \subset A} (-1)^{|A|-|B|} \mathbb{E}_{\mathbb{X}_B}\left[\mathrm{MMD}^2(\mathrm{P}_Y, \mathrm{P}_{Y|\mathbb{X}_B}; \mathscr{H})\right].$$

Theorem 6.9 is very similar to the FANOVA one (Corollary 1.8). One can note that the total variance of the output is replaced by a generalized variance MMD^2_{tot} defined by the kernel, and that each subset effect is obtained by removing lesser order ones in the MMD distance of the conditional distributions (instead of the variance of the conditional expectations in the FANOVA). The following corollary states that these two decompositions coincide when using the linear kernel (Sobol' decomposition).

Corollary 6.10. *When $k(y,y') = yy'$ in Theorem 6.9, the decomposition is the same as the one provided in Corollary 1.8). In other words,*

$$\mathrm{MMD}^2_{tot} = V \text{ and } \forall B \subset \{1,\dots,d\},\ \mathbb{E}_{\mathbb{X}_B}\left[\mathrm{MMD}^2(\mathrm{P}_Y, \mathrm{P}_{Y|\mathbb{X}_B}; \mathscr{H})\right] = \mathrm{Var}(\mathbb{E}[Y|\mathbb{X}_B]).$$

It further implies $\forall A \subset \{1,\dots,d\}$, $\mathrm{MMD}^2_A = V_A$.

Thanks to Theorem 6.9 we can now define properly normalized MMD-based indices.

Definition 6.11. *In the frame of Theorem 6.9, let $A \subset \{1,\dots,d\}$ and $j \in \{1,\dots,d\}$.*

- *The normalized MMD-based sensitivity index associated to A is defined as*

$$S^{\mathrm{MMD}}_A = \frac{\mathrm{MMD}^2_A}{\mathrm{MMD}^2_{tot}}.$$

- *$S^{\mathrm{MMD}}_j = S^{\mathrm{MMD}}_{\{j\}}$ is the MMD-based index associated to the singleton $\{j\}$. It is called the first-order MMD-based index for the input variable X_j. More generally, if $l = |A|$, S^{MMD}_A is called the MMD-based index of order l associated to X_A.*
- *The total MMD-based index associated to X_A is defined as $S^{T,\mathrm{MMD}}_A = 1 - S^{\mathrm{MMD}}_{\bar{A}}$. In particular*

$$S^{T,\mathrm{MMD}}_j = S^{T,\mathrm{MMD}}_{\{j\}} = 1 - \frac{\mathrm{MMD}^2_{\sim\{j\}}}{\mathrm{MMD}^2_{tot}}.$$

From Theorem 6.9, we have the fundamental normalization identity

$$\sum_{A \subset \{1,\dots,d\}} S^{\mathrm{MMD}}_A = 1.$$

6.2.3 HSIC dependence measure

Coming back to moment-independent sensitivity indices, we have seen in Section 6.1 that some special cases of f-divergences lead to sensitivity indices which are dependence measures, such as the mutual information

$$S^{MI}_j = I(X_j; Y) = \int_{\mathbb{R}^2} p_{X_j,Y}(x,y) \ln\left(\frac{p_{X_j,Y}(x,y)}{p_Y(y)\, p_{X_j}(x)}\right) dxdy$$

or the squared-loss mutual information

$$S_j^{SMI} = SMI(X_j;Y) = \int_{\mathbb{R}^2} \left(\frac{p_{X_j,Y}(x,y)}{p_Y(y)p_{X_j}(x)} - 1 \right)^2 p_Y(y)p_{X_j}(x)dxdy.$$

To measure the dependence between any two random variables $(U,V) \sim \mathrm{P}_{UV}$ with marginals $U \sim \mathrm{P}_U$ and $V \sim \mathrm{P}_V$, it is common practice to measure the dissimilarity between the joint distribution P_{UV} and the product of the marginals $\mathrm{P}_U \times \mathrm{P}_V$. In the examples above, the dissimilarity measures are the Kullback-Leibler divergence and the χ^2 divergence, respectively. However, the kernel-based dissimilarity measure MMD has been introduced in the previous section: it is then possible to define a dependence measure equal to the MMD distance between P_{UV} and $\mathrm{P}_U \times \mathrm{P}_V$. This measure is the so-called *Hilbert-Schmidt Independence Criterion* or HSIC and is defined below.

Definition 6.12. *Consider a RKHS $\mathcal{H}_\mathcal{X}$ of functions $\mathcal{X} \to \mathbb{R}$ with kernel $k_\mathcal{X}$ and a second RKHS $\mathcal{H}_\mathcal{Y}$ of functions $\mathcal{Y} \to \mathbb{R}$ with kernel $k_\mathcal{Y}$. The Hilbert-Schmidt Independence Criterion* $\mathrm{HSIC}(U,V)$ *between two random vectors (U,V) with joint distribution P_{UV} on $\mathcal{X} \times \mathcal{Y}$ and marginal distributions P_U on $\mathcal{X}$ and P_V on $\mathcal{Y}$ is given by*

$$\mathrm{HSIC}(U,V) = \|\mu_{\mathrm{P}_{UV}} - \mu_{\mathrm{P}_U \times \mathrm{P}_V}\|^2_{\mathcal{H}_\mathcal{X} \times \mathcal{H}_\mathcal{Y}}$$

where

$$\mu_{\mathrm{P}_{UV}} = \int_{\mathcal{X}\times\mathcal{Y}} k_\mathcal{X}(\cdot,u)k_\mathcal{Y}(\cdot,v)d\mathrm{P}_{UV}(u,v)$$

is the kernel embedding of the joint distribution and $\mu_{\mathrm{P}_U \times \mathrm{P}_V} = \int_{\mathcal{X}\times\mathcal{Y}} k_\mathcal{X}(\cdot,u)k_\mathcal{Y}(\cdot,v)d\mathrm{P}_U(u)d\mathrm{P}_V(v)$ is the kernel embedding of the product of the marginals. Furthermore, we have the following formulation with kernels

$$\mathrm{HSIC}(U,V) = \mathbb{E}_{U,U',V,V'}[k_\mathcal{X}(U,U')k_\mathcal{Y}(V,V')] + \mathbb{E}_{U,U'}[k_\mathcal{X}(U,U')]\mathbb{E}_{V,V'}[k_\mathcal{Y}(V,V')] - 2\mathbb{E}_{U,V}\left[\mathbb{E}_{U'}[k_\mathcal{X}(U,U')]\mathbb{E}_{V'}[k_\mathcal{Y}(V,V')]\right]$$

where (U,V) and (U',V') are independent copies distributed as P_{UV}, provided $\mathbb{E}_{U,U'}[k_\mathcal{X}(U,U')] < \infty$ and $\mathbb{E}_{V,V'}[k_\mathcal{Y}(V,V')] < \infty$.

If both kernels $k_\mathcal{X}$ and $k_\mathcal{Y}$ are characteristic, the HSIC dependence measure equals 0 if, and only if, the variables are independent (just like for the mutual information). Concerning estimation, assume that $(U_i,V_i)_{i=1\ldots n}$ is a sample of the random vector (U,V) and denote $K_\mathcal{X}$ and $K_\mathcal{Y}$ the Gram matrices with entries $K_\mathcal{X}(i,j) = k_\mathcal{X}(U_i,U_j)$ and $K_\mathcal{Y}(i,j) = k_\mathcal{Y}(V_i,V_j)$, and H the centering matrix such that $H(i,j) = \delta_{ij} - \frac{1}{n}$. Gretton et al [54] propose the following consistent estimator for $\mathrm{HSIC}(U,V)$

$$\begin{aligned}\widehat{\mathrm{HSIC}}(U,V) &= \frac{1}{n^2}\mathrm{Tr}(K_\mathcal{X} H K_\mathcal{Y} H) \\ &= \frac{1}{n^2}\sum_{i,j=1}^n k_\mathcal{X}(U_i,U_j)k_\mathcal{Y}(V_i,V_j) \\ &+ \frac{1}{n^2}\sum_{i,j=1}^n k_\mathcal{X}(U_i,U_j)\frac{1}{n^2}\sum_{i,j=1}^n k_\mathcal{Y}(V_i,V_j) \\ &- \frac{2}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n k_\mathcal{X}(U_i,U_j)\frac{1}{n}\sum_{j=1}^n k_\mathcal{Y}(V_i,V_j)\right].\end{aligned} \tag{6.12}$$

Notice that an unbiased estimator is also introduced in [111]. HSIC has been widely used in a large panel of applications, such as feature selection [111] or independence testing [55].

Sensitivity index based on HSIC - distance between $\mathbf{P}_{\mathbb{X}_B,Y}$ and $\mathbf{P}_{\mathbb{X}_B} \times \mathbf{P}_Y$

Now, HSIC-based indices also enjoy a FANOVA-like decomposition. Nevertheless, it requires assumptions on the kernel structure assigned to the inputs.

Assumption 6.2.2. $\forall A \subset \{1,\ldots,d\}, \mathbb{E}_{\xi\sim\mathrm{P}_{\mathbb{X}_A}}[k_{\mathcal{X}_A}(\xi,\xi)] < \infty$ *and* $\mathbb{E}_{\xi\sim\mathrm{P}_Y}[k_\mathcal{Y}(\xi,\xi)] < \infty$.

Assumption 6.2.3. *The reproducing kernel $k_{\mathscr{X}}$ is of the form*

$$k_{\mathscr{X}}(\mathbf{x},\mathbf{x}') = \prod_{j=1}^{d}\left(1 + k_j(x_j, x_j')\right) \tag{6.13}$$

where for each $j = 1,\dots,d$, $k_j(\cdot,\cdot)$ is the reproducing kernel of a RKHS $\mathscr{H}_j$ of real functions depending only on variable x_j and such that $1 \notin \mathscr{H}_j$.
Furthermore, for all $j = 1,\dots,d$ and $\forall x_j \in \mathscr{X}_j$, we have

$$\int_{\mathscr{X}_j} k_j(x_j, x_j')\, d\mathrm{P}_{X_j}(x_j') = 0. \tag{6.14}$$

Observe that this assumption involves using univariate kernels associated to a RKHS which does not include constant functions. This may seem to be an original setting, but this is actually what is usually done in standard support vector machines where a bias is explicitly introduced when using RBF kernels. Indeed, it is a well-known fact that they define a RKHS which does not contain constant functions [116]. With Assumption 6.2.3, we can now state a decomposition for HSIC-based sensitivity indices.

Theorem 6.13. *Under the same assumptions of Theorem 1.3 (in particular, the random vector $\mathbb{X}$ has independent components), with Assumptions 6.2.2, 6.2.3 and assuming Mercer's theorem holds, the HSIC dependence measure between $\mathbb{X} = (X_1,\dots,X_d)$ and Y can be decomposed as*

$$\mathrm{HSIC}(\mathbb{X}, Y) = \sum_{A \subset \{1,\dots,d\}} \mathrm{HSIC}_A$$

where each term is given by

$$\mathrm{HSIC}_A = \sum_{B \subset A} (-1)^{|A|-|B|} \mathrm{HSIC}(\mathbb{X}_B, Y)$$

and $\mathrm{HSIC}(\mathbb{X}_B, Y)$ is defined with kernel $k_B(\mathbf{x}_B, \mathbf{x}_B') = \prod_{j \in B}\left(1 + k_j(x_j, x_j')\right)$ on the inputs as in (6.13).

The proof relies on Theorem 4.1 from Kuo et al. [67] and is given in Da Veiga [33]. Normalized HSIC-based indices can finally be defined.

Definition 6.14. *In the frame of Theorem 6.13, let $A \subset \{1,\dots,d\}$. The normalized HSIC-based sensitivity index associated to a subset A of input variables is defined as*

$$S_A^{\mathrm{HSIC}} = \frac{\mathrm{HSIC}_A}{\mathrm{HSIC}(\mathbb{X}, Y)},$$

while the total HSIC-based index associated to A is

$$S_A^{T,\mathrm{HSIC}} = \sum_{B \subset \{1,\dots,d\},\, B \cap A \neq \emptyset} S_B^{\mathrm{HSIC}} = 1 - \frac{\mathrm{HSIC}(\mathbb{X}_{-A}, Y)}{\mathrm{HSIC}(\mathbb{X}, Y)}.$$

From Theorem 6.13, we have the fundamental normalizing identity providing the interpretation of HSIC-based indices as percentage of the explained HSIC dependence measure between $\mathbb{X} = (X_1,\dots,X_d)$ and Y:

$$\sum_{A \subset \{1,\dots,d\}} S_A^{\mathrm{HSIC}} = 1.$$

Notice that assumption 6.2.3 requires the use of kernels defining a so-called RKHS of *zero-mean functions* [124]. A general procedure to build such kernels was introduced by [38]: a zero-mean kernel $k_0^D(\cdot,\cdot)$ is given by

$$k_0^D(x,x') = k(x,x') - \frac{\int k(x,t)\,d\mathrm{P}(t) \int k(x',t)\,d\mathrm{P}(t)}{\iint k(s,t)\,d\mathrm{P}(s)\,d\mathrm{P}(t)}$$

from an arbitrary univariate $k(\cdot,\cdot)$, where $k_0^D(\cdot,\cdot)$ satisfies $\forall x,\ \int k_0^D(x,t)\,d\mathrm{P}(t) = 0$.

6.3 Goal-oriented sensitivity analysis

The Sobol indices have been widely used in many contexts. Application studies generally show a common drawback : they do not emphasized a capital point, namely that the efficiency of an index has to be ranked *w.r.t.* the statistical parameter(s) or features that have to be estimated.
It seems very intuitive that to estimate a mean or a median (central parameter) could involve very different variables than estimating extreme quantiles. Thus the same index should not be used for these two different tasks. So we need to adapt the indices to each particular goal we track, that we may call a "goal oriented" sensitivity study. As a matter of fact the Sobol indices are well suited to quantify the sensitivity of an estimator based on a variance criterion : a mean. Shortly speaking, we propose to define an index for each statistical purpose. Of course it may happen that several goals are to be reached, then one can adopt a mixed strategy i.e compute various indices related to each goal and combine them to define some importance criteria of the input variables.

6.3.1 Target SA via Monte Carlo thresholding

Let us define $\mathscr{R} \subset \mathscr{Y}$ a region of interest of the output domain $\mathscr{Y}$. From a SA perspective, our objective here is to measure the impact of each input on the fact that the output reaches $\mathscr{R}$. By a simple change of variable

$$Z = \mathbb{1}_{\{Y \in \mathscr{R}\}} \tag{6.15}$$

(for example $Z = \mathbb{1}_{\{Y>t\}}$ for a given threshold t), this implies that we have to evaluate the impact of the inputs on level sets of the output. Such an objective has been called *target SA* (TSA) by Raguet and Marrel [98] and Marrel and Chabridon [77]. Note that other changes of variable than (6.15) are possible in order to smooth the discontinuity of the indicator function.

Remark 6.15. *Another objective (that is not discussed here), called "conditional SA", aims at evaluating the influence of inputs within the region of interest only, ignoring what happens outside [77, 98].*

The first straightforward TSA approach just consists in recycling a first-order Sobol' index with the new output Z

$$V_j = \mathrm{Var}\mathbb{E}[Z|X_j]. \tag{6.16}$$

where the unnormalized version of Sobol' index is used (as in (6.6) of Section 6.1.2) in order to be consistent with Sections 6.1 and 6.2 of this chapter. From (6.16), it is easy to see that [30, 73]

$$V_j = \mathbb{E}[(p_j - p)^2] \tag{6.17}$$

where $p = \mathbb{E}[Z] = \mathbb{P}(Y \in \mathscr{R})$ (resp. $p_j = \mathbb{E}[Z|X_j] = \mathbb{P}(Y \in \mathscr{R}|X_j)$) is called the probability of failure (resp. the conditional probability of failure). Different estimation schemes of this quantity have been considered for instance in Wei et al [125].
Interestingly, by developing (6.17) via the Bayes' theorem, it can also be shown that we have [97]

$$V_j = \mathbb{E}\left[\left(\frac{p f_{X_j|Y\in\mathscr{R}}(X_j)}{f_{X_j}(X_j)} - p\right)^2\right] = p^2\mathbb{E}\left[\left(\frac{f_{X_j|Y\in\mathscr{R}}(X_j)}{f_{X_j}(X_j)} - 1\right)^2\right]$$

that is proportional (up to the factor p^2) to the Pearson χ^2 divergence between $f_{X_j|Y\in\mathscr{R}}$ and f_{X_j}. This perspective demonstrates that this simple sensitivity index on an indicator function operates as a measure comparing the initial input distribution and its conditional counterpart when the output is forced to lie in the region of interest. In this context, several research works (see, e.g, [105]) try to develop some given-data estimation procedures.
In a similar way but without this connection in mind, Spear and Hornberger [114] proposed to perform the same comparison between the probability distributions, via the Kolmogorov distance. More precisely, they define a sensitivity index as

$$\sup_{x\in\mathscr{X}} |F_{X_j|Y\in\mathscr{R}}(x) - F_{X_j}(x)|.$$

Saltelli et al [104] has called it the Monte Carlo filtering method.
Finally, we can also define the following HSIC-based sensitivity index [112]

$$S_j^{\mathrm{HSIC,unnorm}} = \mathrm{HSIC}(X_j, \mathbb{1}_{\{Y\in\mathscr{R}\}}) = \mathrm{HSIC}(X_j, Z)$$

where $Z = \mathbb{1}_{\{Y\in\mathscr{R}\}}$. If we set $k(z,z') = \delta_{zz'}$ and choose any kernel $k_{\mathscr{X}}$ for the inputs, it can be shown that

$$S_j^{\mathrm{HSIC,unnorm}} = \mathbb{P}(Y \in \mathscr{R})^2\, \mathrm{MMD}^2\left(\mathrm{P}_{X_j}, \mathrm{P}_{X_j|Y\in\mathscr{R}}; \mathscr{H}\right),$$

where $\mathscr{H}$ is the RKHS with kernel $k_{\mathscr{X}}$, see Spagnol et al [113] for a proof in a goal-oriented context related to optimization. In other words, such a sensitivity index detects an influential variable if, when conditioned on the fact that the output reaches the region of interest, its resulting conditional distribution varies from the original one. This is the same idea as in the Monte Carlo filtering approach presented in Section 6.3.1, except that the distributions are compared with the MMD instead of the Kolmogorov distance.

6.3.2 Reliability-oriented sensitivity analysis

In a general context, structural reliability aims at determining the failure probability (or a quantile) of the numerical model output $Y = G(\mathbf{X})$, by considering the input variables as random. To solve this problem, several authors from the structural reliability community have largely contributed to the development and promotion of Monte Carlo techniques and geometrically-based approximation approaches, such as the *first/second-order reliability methods* (FORM/SORM). These techniques are now widely used in other physical domains, from hydraulics to aerospace engineering. A large number of more sophisticated methods have also been developed to compute a failure probability and, more generally, to infer rare events with a small number of runs of the model $G(\cdot)$. The two main classes of advanced techniques are the variance-reduction Monte Carlo methods [71, 85] and the metamodel-based approaches [7, 8]. From a practical point of view, it appears that importance measures of the inputs are particular useful for a post-hoc analysis of rare event algorithm results or even as an help for accelerating the rare event inference algorithm (see, e.g., [61]).
The reliability-oriented SA aims at studying the dependence of a rare-event related QoI to the model input variables [27]. One major difficulty is that the QoI can be very restrictive on the output (due to the rareness of the event). It is then necessary to find methods able to analyze correctly the influence of the input uncertainties on this critical domain of the output. In the Section 6.3.1, several importance measures have been introduced. however, their computations require the use of a Monte Carlo (i.i.d.) sample with a sufficient data number in the two classes of Z in order to compute significant statistics. In the structural reliability community, sensitivity indices embedded within different failure probability estimation methods have then been proposed (see Chabridon [25] for a complete overview). As an example, one can mention:

- the local indices based on partial derivatives of the failure probability with respect to the parameters of the pdf of the inputs [10], that can easily deduced from a sample via the so-called score-function [102],
- the FORM/SORM and reliability-index-based importance factors [64, 75, 76, 94],
- the global indices based on conditional failure probability (as in Section 6.3.1) and obtained by conditional samples, which are by-products of rare event simulation techniques (e.g., importance sampling or subset simulation) [2, 127].

6.3.3 Contrast-function-based indices

In this section, a general methodology is introduced and discussed to build sensitivity indices. For the sake of clarity, let us consider that the output of interest Y is scalar. Let us underline that, in the probabilistic framework of SA, the output Y is random. Indeed, the uncertainty on the inputs is propagated to the output via the model. The output is thus described by a probability distribution whose support is included in $\mathbb{R}$. Let us now come back to the definition of first-order Sobol' indices

$$S_j = \frac{\mathrm{Var}(\mathbb{E}[Y|X_j])}{\mathrm{Var}(Y)} = 1 - \frac{\mathbb{E}[\mathrm{Var}(Y|X_j)]}{\mathrm{Var}(Y)}.$$

It is well known that

$$\mathrm{Var}(Y) = \inf_{\theta\in\mathbb{R}} \mathbb{E}[(Y-\theta)^2],$$
$$\mathrm{Var}\,(Y|X_j) = \inf_{\theta\in\mathbb{R}} \mathbb{E}[(Y-\theta)^2|X_j].$$

Thus we can write

$$S_j = 1 - \frac{\mathbb{E}[\inf_{\theta\in\mathbb{R}} \mathbb{E}[(Y-\theta)^2|X_j]]}{\inf_{\theta\in\mathbb{R}} \mathbb{E}[(Y-\theta)^2]} = 1 - \frac{\mathbb{E}[\inf_{\theta\in\mathbb{R}} \mathbb{E}[l_2(Y;\theta)|X_j]]}{\inf_{\theta\in\mathbb{R}} \mathbb{E}[l_2(Y;\theta)]},$$

with $l_2(y;\theta) = (y-\theta)^2$ the quadratic loss function. So that S_j appears to compare the optimal value of the function $\mathbb{E}(Y-\theta)^2$ to the expected optimal value of the conditional function $\mathbb{E}[(Y-\theta)^2|X_j]$. A function $\theta \mapsto \mathbb{E}\Psi(Y;\theta)$ is a contrast function if it admits a unique minimizer θ^*. The quadratic loss function is a contrast function as $\theta \mapsto l_2(Y;\theta) = \mathbb{E}\,(Y-\theta)^2$ admits a unique minimizer, namely $\mathbb{E}[Y]$.
As noticed in Fort et al [43], a contrast function is a very useful object in *statistical learning theory* (see, e.g., Massart [79]) where it defines estimation procedures of some feature $\theta^* \in \Theta$ (scalar or functional) associated to a random

variable Y. For instance, when observing a n-sample $Y^{(1)},\dots,Y^{(n)}$ of the random variable Y, an estimator of θ^* is given by the minimizer of $\theta \mapsto \Psi_n(\theta)$, where Ψ_n is obtained by substituting the expectation w.r.t. the variable Y by the expectation w.r.t. the empirical measure of the sample. For example, by considering the quadratic loss function $\theta \mapsto l_2(Y;\theta) = \mathbb{E}(Y-\theta)^2$, one gets $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n Y^{(j)}$, namely the empirical mean to estimate $\theta^* = \mathbb{E}Y$.
Thus it seems that first-order Sobol' indices are well tailored when one is interested in the influence of input parameters for the estimation of the mean behavior of the output Y. If one is interested in estimating other QoI (i.e., characteristics of the probability distribution of Y), such as quantiles, one has to consider different loss functions leading to different contrast functions. It leads to new sensitivity indices called goal-oriented sensitivity indices in Fort et al [43]. More precisely, let us state Definition 6.16 below.

Definition 6.16. *Let $\theta \mapsto \mathbb{E}\Psi(Y;\theta)$ be a contrast function such that $\mathbb{E}\min_\theta \Psi(Y;\theta)$ is finite. The sensitivity indices associated to Ψ are then defined as*

$$1 - \frac{\mathbb{E}[\inf_{\theta\in\mathbb{R}} \mathbb{E}[l(Y,\theta)|X_j]]}{\inf_{\theta\in\mathbb{R}} \mathbb{E}[l(Y,\theta)]},\ j = 1,\dots,d.$$

Examples

1. If the QoI is the mean of Y, one considers

$$\Psi(y;\theta) = l_2(y;\theta) = (y-\theta)^2.$$

2. If the QoI is the median of Y, one considers

$$\Psi(y;\theta) = l_1(y;\theta) = |y-\theta|.$$

3. If the QoI is the quantile of order $\alpha \in (0,1)$ of Y, one considers

$$\Psi(y;\theta) = l_\alpha(y,\theta) = \begin{cases} (1-\alpha)(\theta - y), & \text{if } y \leqslant \theta \\ \alpha(y-\theta), & \text{if } y > \theta. \end{cases}$$

   In this case, the literature refers to QOSA for quantile-oriented SA. We refer to Maume-Deschamps et al [80] and Browne et al [20] for a complete study including the statistical estimation of the corresponding indices. Another line of research explores the use of random forest given-data estimation procedures [39].

Exercices

In the sequel, we present two examples for which the indices based on contrast can be analytically computed and we show how they differ from the classical Sobol' indices.

Exercise 14. *Let $Y = X_1 + X_2$, with $X_1 \sim Exp(1), X_2 \sim -X_1$ and X_1 and X_2 independent.*

1. *Determine the distribution of Y.*
2. *Compute the Sobol' indices.*
3. *Now we aim at defining sensitivity indices with respect to the α-quantile $q_Y(\alpha)$ of Y. To do so, we use the following contrast*

$$\Psi(\theta) = \mathbb{E}(Y-\theta)(\alpha - \mathbb{1}_{Y\leq\theta}),$$

that characterizes the α-quantile.

   (a) *Check that $\mathbb{E}[\min_\theta \psi(Y;\theta)] = 0$.*

   (b) *Compute the indices*

$$S_\psi^k = \frac{\mathbb{E}_{(X_k,Y)}\left[\psi(Y;\theta^*) - \psi(Y;\theta_k(X_k))\right]}{\mathbb{E}[\psi(Y;\theta^*)]} = \frac{\mathbb{E}[\psi(Y;\theta^*)] - \mathbb{E}_{(X_k,Y)}\left[\psi(Y;\theta_k(X_k))\right]}{\mathbb{E}[\psi(Y;\theta^*)]},$$

   for $k = 1, 2$, where $\theta^* = q_Y(\alpha)$, $\theta_1(X_1) = q_{Y/X_1}(\alpha)$ and $\theta_2(X_2) = q_{Y/X_2}(\alpha)$.

   (c) *Comment the results.*

   (d) *Compute the limits as $\alpha \to 0$ of the sensitivity indices.*

Exercise 15. *Let $Y = X_1 + X_2$, with $X_1 \sim Exp(1), X_2 \sim Exp(a), a > 0$, two independent variables.*

1. *Determine the distribution of Y.*
2. *Compute the Sobol' indices.*
3. *Now we aim at providing a sensitivity index with respect to the probability of Y to exceed $t \geq 0$, i.e $\mathbb{P}(Y \geqslant t)$. A contrast which characterizes such quantity of interest can be the following*

$$\Psi(\theta) = \mathbb{E}[|\mathbb{1}_{Y\geq t} - \theta|^2] \tag{6.18}$$

which in fact turns to be a quadratic contrast.

   (a) *Check that $\mathbb{E}[\min_\theta \psi(Y;\theta)] = 0$.*

   (b) *Compute the indices*

$$S_\psi^k = \frac{\mathbb{E}[\psi(Y;\theta^*)] - \mathbb{E}_{(X_k,Y)}\left[\psi(Y;\theta_k(X_k))\right]}{\mathbb{E}[\psi(Y;\theta^*)]},$$

   for $k = 1, 2$, where $\theta^* = \mathbb{P}(Y \geqslant t)$, $\theta_1(X_1) = \mathbb{P}(X_2 \geqslant t - X_1/X_1)$ and $\theta_2(X_2) = \mathbb{P}(X_1 \geqslant t - X_2/X_2)$.

   (c) *Comment the results.*

Chapter 7

Practical

The model considered in this practical is the following:

$$Y = f(X_1, X_2) = \exp\{X_1 + 2X_2\}$$

where the X_i's are i.i.d. $\mathcal{N}(0,1)$.

☞ Code the function f.

7.1 Exponential model and Sobol' indices

Recall that the (closed) Sobol index S^i with respect to X_i is defined by:

$$S^i = \frac{\mathrm{Var}\left(\mathbb{E}[Y|X_i]\right)}{\mathrm{Var}(Y)}.$$

☞ Compute the exact values of the Sobol' indices.

7.1.1 Pick-Freeze estimation of the Sobol' indices

To estimate by the Pick-Freeze procedure the Sobol indiex S^i with respect to X_i, we rewrite the variance of the conditional expectation in terms of a covariance between the output Y and its Pick-Freeze version Y_{PF} given by:

$$Y_{PF,i} = f(X_{PF,i})$$

where $X_{PF,i}$ from X by freezing the coordinate i and resampling the other coordinates independently. In other words,

$$S^i = \frac{\mathrm{Cov}\left(Y, Y_{PF,i}\right)}{\mathrm{Var}(Y)},$$

Then it remains to compute the ration between the empirical covariance and the empirical variance to get the Pick-Freeze estimation of S^1 (for example $i = 1$):

$$S^1_{N,PF} = \frac{\frac{1}{N}\sum Y^i Y^i_{PF,1} - \left(\frac{1}{N}\sum Y^i\right)\left(\frac{1}{N}\sum Y^i_{PF,1}\right)}{\frac{1}{N}\sum (Y^i)^2 - \left(\frac{1}{N}\sum Y^i\right)^2}$$

or the efficient estimation:

$$T^1_{N,PF} = \frac{\frac{1}{N}\sum Y^i Y^i_{PF,1} - \left(\frac{1}{2N}\sum \left(Y^i + Y^i_{PF,1}\right)\right)^2}{\frac{1}{N}\sum (Y^i)^2 - \left(\frac{1}{N}\sum Y^i\right)^2}.$$

☞ Construct the Pick-Freeze samples.

☞ Compute the Pick-Freeze estimations of the Sobol' indices.

☞ Compute the quadratic error on n samples of size N.

7.1.2 Rank estimation of the Sobol' indices

To estimate by the rank procedure the Sobol index S^1 with respect to X_1, we rearrange the initial sample $(X_1^1, Y^1), \ldots, (X_1^N, Y^N)$ with respect to the first coordinate: $X_1^{(1)} < \ldots < X_1^{(N)}$.
We get

$$(X_1^{(1)}, Y^{(1)}), \ldots, (X_1^{(N)}, Y^{(N)})$$

The rank estimation is then given by

$$S^1_{N,Rank} = \frac{\frac{1}{N}\sum_{i=1}^{N-1} Y^{(i)} Y^{(i+1)} - \left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right)^2}{\frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right)^2}.$$

☞ Construct the rearranged samples.

☞ Compute the rank estimations of the Sobol' indices.

☞ Compute the quadratic error on n samples of size N.

7.1.3 Comparison of the estimation methods

7.2 Exponential model and Cramér-von Mises indices

Recall that the Cramér-von Mises $S^i_{2,CVM}$ with respect to X_i is given by:

$$S^i_{2,CVM} = \frac{\int_{\mathbb{R}^k} \mathbb{E}\left[\left(F(t) - F^i(t)\right)^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t)) dF(t)}.$$

☞ Compute the exact values of the Cramér-von Mises indices.

7.2.1 Pick-Freeze estimation of the Cramér-von Mises indices

To estimate by the Pick-Freeze procedure the Cramér-von Mises $S^i_{2,CVM}$ with respect to X_i, we rewrite the numerator in terms of a covariance:

$$\int_{\mathbb{R}^k} \mathbb{E}\left[\left(F(t) - F^i(t)\right)^2\right] dF(t) = \mathbb{E}\left[\mathrm{Cov}\left(\mathbb{1}_{Y \leqslant W}, \mathbb{1}_{Y_{PF,i} \leqslant W}\right)\right],$$

where W is a random variable independent of Y and distributed as Y.
Then it remains to proceed by a Monte-Carlo scheme to estimate both the expectation and the covariance. To do so, we generate:

- two N-samples of Y: $(Y_j^{i,1}, Y_j^{i,2})$, $1 \leqslant j \leqslant N$ (Pick-Freeze) ;
- a third N-sample of Y: W_k, $1 \leqslant k \leqslant N$

and we get the Pick-Freeze estimation of the numerator of the Cramér-von Mises index $S^i_{2,CVM}$:

$$N^i_{2,CVM,PF} = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\{Z_j^{i,1} \leqslant W_k\}} \mathbb{1}_{\{Z_j^{i,2} \leqslant W_k\}} - \left[\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z_j^{i,1} \leqslant W_k\}} + \mathbb{1}_{\{Z_j^{i,2} \leqslant W_k\}}\right)\right]^2\right\}.$$

☞ Construct the Pick-Freeze samples.

☞ Compute the Pick-Freeze estimations of the Cram'r-von Mises indices.

☞ Compute the quadratic error on n samples of size N.

7.2.2 Rank estimation of the Cramér-von Mises indices

From a unique N-sample of Y, we estimate the numerator by:

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\left[\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\{Y_{(j)} \leqslant Y_i\}} \mathbb{1}_{\{Y_{(j+1)} \leqslant Y_i\}}\right] - \left[\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\{Y_j \leqslant Y_i\}}\right]^2\right\}$$

where $Y_{(j)}$ is the output corresponding to the j-th input ordered input X_1 and the denomiator by

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\left[\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\{Y_j \leqslant Y_i\}}\right] - \left[\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}_{\{Y_j \leqslant Y_i\}}\right]^2\right\}.$$

☞ Construct the rearranged samples.

☞ Compute the rank estimations of the Cramér-von Mises indices.

☞ Compute the quadratic error on n samples of size N.

7.2.3 Comparison of the estimation methods

7.3 Stochastic exponential model - Wasserstein indices

The model considered is still:

$$Y = f(X_1, X_2) = \exp\{X_1 + 2X_2\}$$

where the X_i's are i.i.d. distributed as $\mathcal{N}(0,1)$. Then we write

$$X_2 = \frac{G_1 + G_2}{\sqrt{2}},$$

where G_1 and G_2 are independent standard Gaussian distributed random variables and also independent of X_1.

⚠ Here, we assume that the practitioner has only access to X_1 and G_1.

The Wasserstein indices $S^{\mathbf{u}}_{2,W_2}$ are given by:

$$\frac{\int_{\mathscr{W}_2(\mathbb{R})^2} \mathbb{E}\left[\left(\mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F})\leqslant W_2(F_1,F_2)}] - \mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F})\leqslant W_2(F_1,F_2)}|X^{\mathbf{u}}]\right)^2\right] d\mathbb{P}^{\otimes 2}(F_1,F_2)}{\int_{\mathscr{W}_2(\mathbb{R})^2} \mathrm{Var}(\mathbb{1}_{W_2(F_1,\mathbb{F})\leqslant W_2(F_1,F_2)}) d\mathbb{P}^{\otimes 2}(F_1,F_2)}.$$

7.3.1 Pick-Freeze estimation of the Wasserstein indices

We generate a N-sample of inputs: $(X_{1,1},\ldots,X_{1,N})$ to get a N-sample of F_1.
For each input $X_{1,i}$ $(i = 1,\ldots,N)$, we compute n times the output and we approximate $F_{1,i}$ by

$$F_{1,i,n} = \frac{1}{n}\sum_{k=1}^{n} \delta_{f_s(X_{1,j},D_{1,j,k})}.$$

The same is done for F_2, $\mathbb{F}$ and the Pick-Freeze version of $\mathbb{F}$.
Then the numerator of the Wasserstein index is estimated by:

$$\frac{1}{N^2}\sum_{i,j}^{N}\Big[\frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{k,n})\leqslant W_2(F_{1,i,n},F_{2,j,n})}\,\mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{PF,k,n})\leqslant W_2(F_{1,i,n},F_{2,j,n})} - \left(\frac{1}{2N}\sum_{k=1}^{N}\left(\mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{k,n})\leqslant W_2(F_{1,i,n},F_{2,j,n})} + \mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{PF,k,n})\leqslant W_2(F_{1,i,n},F_{2,j,n})}\right)\right)^2\Big]$$

and its denominator by:

$$\frac{1}{N^2}\sum_{i,j}^{N}\left[\frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{k,n})\leqslant W_2(F_{1,i,n},F_{2,j,n})} - \left(\frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{k,n})\leqslant W_2(F_{1,i,n},F_{2,j,n})}\right)^2\right].$$

To compute explicitly our estimator, it remains to compute terms of the form:

$$W_2(F_{1,i,n},F_{2,j,n}) = \frac{1}{n}\sum_{l=1}^{n}(Y_{1,i,(k)} - Y_{2,j,(k)})^2,$$

where $Y_{1,i,k} = f_s(X_{1,i},D_{1,i,k})$ and $Y_{1,i,(k)}$ is the k-th order statistics and the same for $Y_{2,j,k}$ and $Y_{2,j,(k)}$.

☞ Construct the Pick-Freeze samples.

☞ Compute the Pick-Freeze estimation of the Wasserstein indices.

7.3.2 Rank estimation of the Wasserstein indices

☞ Compute the rank estimation of the Wasserstein indices.

7.3.3 Comparison of the estimation methods

Chapter 8

Practical in French

Le modèle considéré dans ce TP est le suivant :

$$Y = f(X_1, X_2) = \exp\{X_1 + 2X_2\}$$

où les X_i sont i.i.d. $\mathcal{N}(0,1)$.

☞ Coder la fonction f.

8.1 Modèle exponentiel et indices de Sobol

Nous rappellons que l'indice de Sobol S^i par rapport à l'entrée X_i est défini par :

$$S^i = \frac{\mathrm{Var}\left(\mathbb{E}[Y|X_i]\right)}{\mathrm{Var}(Y)}.$$

☞ Calculer les valeurs des indices de Sobol.

8.1.1 Calcul de l'estimateur Pick-Freeze des indices de Sobol

Pour estimer par la méthode Pick-Freeze l'indice de Sobol S^i par rapport à l'entrée X_i, nous utilisons le fait que la variance de l'espérance conditionnelle peut s'écrire en termes de covariance entre la sortie Y et sa version Pick-Freeze Y_{PF} donnée par

$$Y_{PF,i} = f(X_{PF,i})$$

où $X_{PF,i}$ s'obtient en gelant l'entrée i et en générant de manière indépendante les autres entrées.
En d'autres termes,

$$S^i = \frac{\mathrm{Cov}\left(Y, Y_{PF,i}\right)}{\mathrm{Var}(Y)},$$

Il reste ensuite à utiliser la variance et la covariance empiriques et à en faire le ratio pour obtenir l'estimation Pick-Freeze de S^1 (par exemple i=1) :

$$S^1_{N,PF} = \frac{\frac{1}{N}\sum Y^i Y^i_{PF,1} - \left(\frac{1}{N}\sum Y^i\right)\left(\frac{1}{N}\sum Y^i_{PF,1}\right)}{\frac{1}{N}\sum (Y^i)^2 - \left(\frac{1}{N}\sum Y^i\right)^2}$$

ou encore l'estimation efficace :

$$T^1_{N,PF} = \frac{\frac{1}{N}\sum Y^i Y^i_{PF,1} - \left(\frac{1}{2N}\sum \left(Y^i + Y^i_{PF,1}\right)\right)^2}{\frac{1}{N}\sum (Y^i)^2 - \left(\frac{1}{N}\sum Y^i\right)^2}.$$

☞ Construire les échantillons Pick-Freeze.

☞ Calculer les estimations Pick-Freeze des indices de Sobol.

☞ Calculer l'erreur quadratique sur n échantillons de taille N.

8.1.2 Calcul de l'estimateur des rangs des indices de Sobol

Pour estimer par la méthode des rangs l'indice de Sobol S^1 par rapport à l'entrée X_1, nous réordonnons l'échantillon initial $(X_1^1, Y^1), \ldots, (X_1^N, Y^N)$ selon la première coordonnée : $X_1^{(1)} < \ldots < X_1^{(N)}$.
Nous obtenons ainsi l'échantillon

$$(X_1^{(1)}, Y^{(1)}), \ldots, (X_1^{(N)}, Y^{(N)})$$

L'estimateur des rangs est alors donné par

$$S^1_{N,Rank} = \frac{\frac{1}{N}\sum_{i=1}^{N-1} Y^{(i)} Y^{(i+1)} - \left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right)^2}{\frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right)^2}.$$

☞ Construire les échantillons réordonnés.

☞ Calculer les estimations par la méthode des rangs des indices de Sobol.

☞ Calculer l'erreur quadratique sur n échantillons de taille N.

8.1.3 Comparaison des méthodes d'estimation

8.2 Modèle exponentiel et indices de Cramér-von Mises

Nous rappellons que l'indice de Cramér-von Mises $S^i_{2,CVM}$ par rapport à l'entrée X_i est défini par :

$$S^i_{2,CVM} = \frac{\int_{\mathbb{R}^k} \mathbb{E}\left[\left(F(t) - F^i(t)\right)^2\right] dF(t)}{\int_{\mathbb{R}^k} F(t)(1-F(t))dF(t)}.$$

☞ Calculer les valeurs des indices de Cramér-von Mises.

8.2.1 Calcul de l'estimateur Pick-Freeze des indices de Cramér-von Mises

Pour estimer par la méthode Pick-Freeze l'indice de Cramér-von Mises $S^i_{2,CVM}$ par rapport à l'entrée X_i, nous écrivons ensuite le numérateur en termes de covariance :

$$\int_{\mathbb{R}^k} \mathbb{E}\left[\left(F(t) - F^i(t)\right)^2\right] dF(t) = \mathbb{E}\left[\mathrm{Cov}\left(\mathbb{1}_{Y \leqslant W}, \mathbb{1}_{Y_{PF,i} \leqslant W}\right)\right],$$

où W est une variable indépendante de Y et de même loi que Y.
Il reste ensuite à faire un schéma de Monte Carlo double pour estimer l'espérance et la covariance.
Nous générons:

- deux N-échantillons de Y: $(Y_j^{i,1}, Y_j^{i,2})$, $1 \leqslant j \leqslant N$ (Pick-Freeze) ;
- un troisième N-échantillon de Y: W_k, $1 \leqslant k \leqslant N$

et nous obtenons l'estimation suivante du numérateur de l'indice de Cramér-von Mises $S^i_{2,CVM}$:

$$N^i_{2,CVM,PF} = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j^{i,1} \leqslant W_k\}}\mathbb{1}_{\{Z_j^{i,2} \leqslant W_k\}} - \left[\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z_j^{i,1} \leqslant W_k\}} + \mathbb{1}_{\{Z_j^{i,2} \leqslant W_k\}}\right)\right]^2\right\}.$$

☞ Construire les échantillons Pick-Freeze.

☞ Calculer les estimations Pick-Freeze des indices de Cramér-von Mises.

☞ Calculer l'erreur quadratique sur n échantillons de taille N.

8.2.2 Calcul de l'estimateur des rangs des indices de Cramér-von Mises

A partir d'un unique échantillon de Y, nous estimons le numérateur par

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\Big[\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Y_{(j)}\leq Y_i\}}\mathbb{1}_{\{Y_{(j+1)}\leq Y_i\}}\Big]-\Big[\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Y_j\leq Y_i\}}\Big]^2\right\}$$

où $Y_{(j)}$ est la sortie correspondant à la j-ème entrée réordonnée de X_1 et le dénominateur par

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\Big[\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Y_j\leq Y_i\}}\Big]-\Big[\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Y_j\leq Y_i\}}\Big]^2\right\}.$$

☞ Construire les échantillons réordonnés.

☞ Calculer les estimations par la méthode des rangs des indices de Cramér-von Mises.

☞ Calculer l'erreur quadratique sur n échantillons de taille N.

8.2.3 Comparaison des méthodes d'estimation

8.3 Modèle exponentiel stochastique - Indices de Wasserstein

Le modèle est toujours

$$Y = f(X_1, X_2) = \exp\{X_1 + 2X_2\}$$

où les X_i sont i.i.d. de loi $\mathcal{N}(0,1)$. Nous pouvons alors écrire

$$X_2 = \frac{G_1 + G_2}{\sqrt{2}},$$

où G_1 et G_2 sont des gaussiennes standard indépendantes entre elles et indépendante de X_1.

⚠ Ici, nous supposons que le praticien a seulement accès à X_1 et G_1.

Les indices de sensibilités $S^{\mathbf{u}}_{2,W_2}$ sont alors

$$\frac{\int_{\mathscr{W}_2(\mathbb{R})^2}\mathbb{E}\left[\left(\mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F})\leq W_2(F_1,F_2)}]-\mathbb{E}[\mathbb{1}_{W_2(F_1,\mathbb{F})\leq W_2(F_1,F_2)}|X^{\mathbf{u}}]\right)^2\right]d\mathbb{P}^{\otimes 2}(F_1,F_2)}{\int_{\mathscr{W}_2(\mathbb{R})^2}\mathrm{Var}(\mathbb{1}_{W_2(F_1,\mathbb{F})\leq W_2(F_1,F_2)})d\mathbb{P}^{\otimes 2}(F_1,F_2)}.$$

8.3.1 Calcul de l'estimateur Pick-Freeze des indices de Wasserstein

Nous générons un échantillon de taille N d'entrées $(X_{1,1},\dots,X_{1,N})$ pour avoir un échantillon de taille N de F_1.
Pour chaque entrée $X_{1,i}$ $(i=1,\dots,N)$, on calcule n fois la sortie et on approche $F_{1,i}$ par

$$F_{1,i,n} = \frac{1}{n}\sum_{k=1}^{n}\delta_{f_s(X_{1,j},D_{1,j,k})}.$$

Nous faisons de même pour F_2, $\mathbb{F}$ et la version Pick-Freeze de $\mathbb{F}$.
Le numérateur de l'indice de Wasserstein est alors estimé par

$$\frac{1}{N^2}\sum_{i,j}^{N}\Big[\frac{1}{N}\sum_{k=1}^{N}\mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{k,n})\leq W_2(F_{1,i,n},F_{2,j,n})}\mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{PF,k,n})\leq W_2(F_{1,i,n},F_{2,j,n})}$$
$$-\left(\frac{1}{2N}\sum_{k=1}^{N}\left(\mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{k,n})\leq W_2(F_{1,i,n},F_{2,j,n})}+\mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{PF,k,n})\leq W_2(F_{1,i,n},F_{2,j,n})}\right)\right)^2\Big]$$

et son dénominateur par

$$\frac{1}{N^2}\sum_{i,j}^{N}\left[\frac{1}{N}\sum_{k=1}^{N}\mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{k,n})\leq W_2(F_{1,i,n},F_{2,j,n})}-\left(\frac{1}{N}\sum_{k=1}^{N}\mathbb{1}_{W_2(F_{1,i,n},\mathbb{F}_{k,n})\leq W_2(F_{1,i,n},F_{2,j,n})}\right)^2\right].$$

Afin de caluler de manière explicite notres estimateur, il reste à calculer des termes de la forme :

$$W_2(F_{1,i,n}, F_{2,j,n}) = \frac{1}{n}\sum_{l=1}^{n}(Y_{1,i,(k)} - Y_{2,j,(k)})^2,$$

où $Y_{1,i,k} = f_s(X_{1,i}, D_{1,i,k})$ et $Y_{1,i,(k)}$ la k-ème statistique d'ordre et de même pour $Y_{2,j,k}$ et $Y_{2,j,(k)}$.

☞ Fonction de création des échantillons Pick-Freeze des entrées.

☞ Calcul des estimations Pick-Freeze des indices de Wasserstein.

8.3.2 Calcul de l'estimateur des rangs des indices de Wasserstein

☞ Calculer les estimations des indices de Wasserstein par la méthode des rangs.

8.3.3 Comparaison des méthodes d'estimation

Appendix A

Proof of the results of Chapter 4

A.1 Proof of the consistency

Proof of Lemma 4.1. Since τ_n has no fix point, and using the measurability of τ_n and the independence, we have

$$\begin{aligned}\mathbb{E}\left[g(Y_j)h(Y_{\tau_n(j)})|\mathscr{F}_n\right] &= \mathbb{E}\Big[g(Y_j)\sum_{\substack{l=1,\\ l\neq j}}^{n} h(Y_l)\mathbb{1}_{\{\tau_n(j)=l\}}|\mathscr{F}_n\Big] = \sum_{\substack{l=1,\\ l\neq j}}^{n}\mathbb{1}_{\{\tau_n(j)=l\}}\mathbb{E}\Big[g(Y_j)h(Y_l)|\mathscr{F}_n\Big]\\ &= \sum_{\substack{l=1,\\ l\neq j}}^{n}\mathbb{1}_{\{\tau_n(j)=l\}}\mathbb{E}\Big[g(Y_j)|\mathscr{F}_n\Big]\mathbb{E}\Big[h(Y_l)|\mathscr{F}_n\Big] = \mathbb{E}\big[g(Y_j)|V_j\big]\sum_{\substack{l=1,\\ l\neq j}}^{n}\mathbb{1}_{\{\tau_n(j)=l\}}\mathbb{E}\big[h(Y_l)|V_l\big]\\ &= \Psi_{V_j}(g)\sum_{\substack{l=1,\\ l\neq j}}^{n}\mathbb{1}_{\{\tau_n(j)=l\}}\Psi_{V_l}(h) = \Psi_{V_j}(g)\Psi_{V_{\tau_n(j)}}(h).\end{aligned}$$

□

Proof of Proposition 4.2. We follow the steps of the proof of Corollary 7.12 in [28]. Our proof is significantly simpler since τ_n is assumed to have no fix points and V is continuous so that there are no ties in the sample. To simplify the notation, we denote $\chi_n(V,Y;g,h)$ and $\chi(V,Y;g,h)$ by χ_n and χ respectively.
We first prove that, for any measurable function φ,

$$\varphi(V_1)-\varphi(V_{\tau_n(1)})\to 0 \tag{A.1}$$

in probability as $n\to\infty$. Let $\varepsilon>0$. By the special case of Lusin's theorem (see [28, Lemma 7.5]), there exists a compactly supported continuous function $\tilde{\varphi}\colon\mathbb{R}\to\mathbb{R}$ such that $\mathbb{P}(\{x;\varphi(x)\neq\tilde{\varphi}(x)\})<\varepsilon$, where $\mathbb{P}$ stands for the distribution of V. Then for any $\delta>0$,

$$\begin{aligned}\mathbb{P}\Big(\big|\varphi(V_1)-\varphi(V_{\tau_n(1)})\big|>\delta\Big) &\leqslant \mathbb{P}\left(\big|\tilde{\varphi}(V_1)-\tilde{\varphi}(V_{\tau_n(1)})\big|>\delta\right)\\ &\quad+\mathbb{P}\left(\varphi(V_1)\neq\tilde{\varphi}(V_1)\right)+\mathbb{P}\left(\varphi(V_{\tau_n(1)})\neq\tilde{\varphi}(V_{\tau_n(1)})\right).\end{aligned} \tag{A.2}$$

By continuity of $\tilde{\varphi}$ and since $V_{\tau_n(1)}\to V_1$ as $n\to\infty$ with probability one, the first term in the right hand side of (A.2) converges to 0 as $n\to\infty$. By construction of $\tilde{\varphi}$, the second term is lower than ε. Turning to the third one, we have thus

$$\begin{aligned}\mathbb{E}[\varphi(V_{\tau_n(1)})] &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[\varphi(V_{\tau_n(j)})] = \frac{1}{n}\sum_{j=1}^{n}\sum_{\substack{l=1\\ l\neq j}}^{n}\mathbb{E}[\varphi(V_l)\mathbb{1}_{\{\tau_n(j)=l\}}]\\ &= \frac{1}{n}\sum_{l=1}^{n}\sum_{\substack{j=1\\ j\neq l}}^{n}\mathbb{E}[\varphi(V_l)\mathbb{1}_{\{\tau_n(j)=l\}}] = \frac{1}{n}\sum_{l=1}^{n}\mathbb{E}[\varphi(V_l)\sum_{\substack{j=1\\ j\neq l}}^{n}\mathbb{1}_{\{\tau_n(j)=l\}}] = \frac{1}{n}\sum_{l=1}^{n}\mathbb{E}[\varphi(V_l)] = \mathbb{E}[\varphi(V_1)]\end{aligned}$$

where we have used the fact that τ_n has no fix point, $V_{\tau_n(i)}\overset{\mathscr{L}}{=}V_{\tau_n(j)}$ for any i and $j=1,\dots,n$, and the V_i's have no ties. This yields

$$\mathbb{P}(\varphi(V_{\tau_n(1)})\neq\tilde{\varphi}(V_{\tau_n(1)})) = \mathbb{P}(\varphi(V_1)\neq\tilde{\varphi}(V_1))<\varepsilon,$$

and, since ε and δ are arbitrary, (A.1) is therefore proved. Now, since $x\mapsto\Psi_x$ is a measurable and bounded function and applying (A.1), we have

$$\begin{cases}\Psi_{V_1}(g)-\Psi_{V_{\tau_n(1)}}(g) & \to 0,\\ \Psi_{V_1}(h)-\Psi_{V_{\tau_n(1)}}(h) & \to 0,\end{cases}\qquad\text{in probability as } n\to\infty. \tag{A.3}$$

Lemma 4.1 and the dominated convergence theorem lead to

$$\mathbb{E}[\chi_n] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[g(Y_j)h(Y_{\tau_n(j)})] = \mathbb{E}[g(Y_1)h(Y_{\tau_n(1)})] = \mathbb{E}[\Psi_{V_1}(g)\Psi_{V_{\tau_n(1)}}(h)] \to \mathbb{E}[\Psi_V(g)\Psi_V(h)] = \chi \tag{A.4}$$

where we have taken into account the fact that $\Psi_V(g)$ and $\Psi_V(h)$ are bounded (due to the boundedness of g and h) and used (A.3).
The last step of the proof consists in comparing χ_n with $\mathbb{E}[\chi_n]$ using Mc Diarmid's concentration inequality [82]. Sharper constants can be obtained in Mc Diarmid's inequality by using the inequalities from [18, 19]. As we are interested in asymptotic results the accuracy of the constant has no impact on the result. Following the same lines as in the proof of [28, Lemma 7.11], Mc Diarmid's concentration inequality in [82] then implies

$$\mathbb{P}(\left|\chi_n - \mathbb{E}[\chi_n]\right| \geqslant t) \leqslant 2\exp\{-2n^2t^2/C^2\}, \tag{A.5}$$

where C is a universal constant and we conclude the proof by combining (A.4) and (A.5). □

A.2 Proof of the asymtotic normality

Framework and goal We consider the model defined in (2.1) that can be rewritten as $Y = f(X,W)$ where $X = X_1$ and $W = (X_2,\dots,X_p)$ are two independent inputs of the numerical code f that is assumed to be bounded.
The random variables X and W are defined on a product space $\Omega = \Omega_X \times \Omega_W$; so that for any $\omega \in \Omega$, there exists $\omega_X \in \Omega_X$ and $\omega_W \in \Omega_W$ and we have $(X,W)(\omega) = (X(\omega_X), W(\omega_W))$. Further, we consider π_W the projection on Ω_W and the product measure $\mathbb{P} = \mathbb{P}_X \otimes \mathbb{P}_W = \mathscr{L}_X \otimes \mathscr{L}_W$, where $\mathscr{L}_X$ is the distribution of X and $\mathscr{L}_W$ is the distribution of W. Naturally, $\mathbb{P}_W = \mathbb{P}\circ\pi_W^{-1}$.
We aim to prove a CLT for the estimator $\xi_n^{\text{Sobol}}(X,Y)$ of the classical first-order Sobol' index with respect to X given by (2.2), the estimator of which defined in (4.8) is given by

$$\xi_n^{\text{Sobol}}(X_1,Y) = \frac{\frac{1}{n}\sum_{j=1}^{n} Y_j Y_{N(j)} - \left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right)^2}{\frac{1}{n}\sum_{j=1}^{n} Y_j^2 - \left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right)^2}$$

where N is defined in (4.7). Notice that the denominator is reduced to the empirical variance of Y. As explained in Section 4.1.1 of Chapter 4, we denote by $Y_{(j)}$ the output associated to $X_{(j)}$ where $X_{(j)}$ stands for the j-th order statistics of $(X_1,\dots,X_n)$. Then observing that

$$\sum_{j=1}^{n} Y_j Y_{N(j)} = \sum_{j=1}^{n} Y_{(j)} Y_{(j+1)} =: \sum_{j=1}^{n} Y_{\sigma_n(j)} Y_{\sigma_n(j+1)}$$

where, to avoid any confusion, σ_n stands for the permutation that rearranges the sample $(X_1,\dots,X_n)$, the estimator $\xi_n^{\text{Sobol}}(X_1,Y)$ can be written as

$$\xi_n^{\text{Sobol}}(X_1,Y) = \frac{\frac{1}{n}\sum_{j=1}^{n-1} Y_{\sigma_n(j)} Y_{\sigma_n(j+1)} - \left(\frac{1}{n}\sum_{j=1}^{n} Y_{\sigma_n(j)}\right)^2}{\frac{1}{n}\sum_{j=1}^{n} Y_{\sigma_n(j)}^2 - \left(\frac{1}{n}\sum_{j=1}^{n} Y_{\sigma_n(j)}\right)^2}. \tag{A.6}$$

A.2.1 Proof of Theorem 4.3

The proof will proceed as follows. First, in view of (A.6), we prove a CLT for

$$\left(\frac{1}{n}\sum_{j=1}^{n-1} Y_{\sigma_n(j)} Y_{\sigma_n(j+1)}, \frac{1}{n}\sum_{j=1}^{n} Y_{\sigma_n(j)}, \frac{1}{n}\sum_{j=1}^{n} Y_{\sigma_n(j)}^2\right).$$

that amounts to prove a CLT for

$$\left(\frac{1}{n}\sum_{j=1}^{n-1} Y_{\sigma_n(j)} Y_{\sigma_n(j+1)}, \frac{1}{n}\sum_{j=1}^{n-1} Y_{\sigma_n(j)}, \frac{1}{n}\sum_{j=1}^{n-1} Y_{\sigma_n(j)}^2\right),$$

since f is bounded. Secondly, we use the so-called delta method [122, Theorem 3.1] to conclude to Theorem 4.3.
It is worth noticing that the permutation on the W's do not affect the result as seen in the sequel. For $j = 1,\dots n-1$, introducing

$$\Delta_{n,j} := f\left(X_{\sigma_n(j)}, W_j\right) - f\left(\frac{j}{n+1}, W_j\right), \quad W_{n,j} := \left(\frac{j}{n+1}, W_j\right) \tag{A.7}$$

leads to $Y_{\sigma_n(j)} = f\left(X_{\sigma_n(j)}, W_{\sigma_n(j)}\right) \stackrel{\mathscr{L}}{=} f\left(X_{\sigma_n(j)}, W_j\right) = \Delta_{n,j} + f\left(W_{n,j}\right)$ and

$$\begin{aligned} Y_{\sigma_n(j)} Y_{\sigma_n(j+1)} &= f\left(X_{\sigma_n(j)}, W_{\sigma_n(j)}\right) f\left(X_{\sigma_n(j+1)}, W_{\sigma_n(j+1)}\right) \\ &\stackrel{\mathscr{L}}{=} f\left(X_{\sigma_n(j)}, W_j\right) f\left(X_{\sigma_n(j+1)}, W_{j+1}\right) \\ &= \left(f\left(W_{n,j}\right) + \Delta_{n,j}\right)\left(f\left(W_{n,j+1}\right) + \Delta_{n,j+1}\right) \\ &= f\left(W_{n,j}\right) f\left(W_{n,j+1}\right) + \Delta_{n,j} f\left(W_{n,j+1}\right) + \Delta_{n,j+1} f\left(W_{n,j}\right) + \Delta_{n,j}\Delta_{n,j+1}. \end{aligned}$$

Thus we are led to establish a CLT for

$$Z_n = \frac{1}{n}\sum_{j=1}^{n-1} \begin{pmatrix} f(W_{n,j}) f(W_{n,j+1}) + \Delta_{n,j} f\left(W_{n,j+1}\right) + \Delta_{n,j+1} f\left(W_{n,j}\right) + \Delta_{n,j}\Delta_{n,j+1} \\ f(W_{n,j}) + \Delta_{n,j} \\ \left(f(W_{n,j}) + \Delta_{n,j}\right)^2 \end{pmatrix}. \tag{A.8}$$

Let us discard the negligible terms in the CLT for Z_n. In that view, noticing that

$$\mathbb{E}\left[X_{\sigma_n(j)}\right] = \frac{j}{n+1} \quad \text{and} \quad \mathrm{Var}(X_{\sigma_n(j)}) = \frac{j(n-j+1)}{(n+1)^2(n+2)} = \mathbb{E}\left[\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right)^2\right] \leqslant \frac{4}{n+2},$$

we first establish

$$X_{\sigma_n(j)} - \frac{j}{n+1} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right). \tag{A.9}$$

As explained below, (A.9) will imply

$$\frac{1}{n}\sum_{j=1}^{n-1} \Delta_{n,j}^2 = O_{\mathbb{P}}\left(\frac{1}{n}\right) \quad \text{and} \quad \frac{1}{n}\sum_{j=1}^{n-1} \Delta_{n,j}\Delta_{n,j+1} = O_{\mathbb{P}}\left(\frac{1}{n}\right). \tag{A.10}$$

First of all, we expand $\Delta_{n,j}$ (resp. $\Delta_{n,j+1}$) using the Taylor-Lagrange formula, for any $j = 1, \dots n-1$ and we obtain

$$\Delta_{n,j} = \left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) f_x\left(W_{n,j}\right) + \frac{1}{2}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right)^2 f_{xx}\left(\delta_{n,j}, W_{\sigma_n(j)}\right), \tag{A.11}$$

where $\delta_{n,j}$ (resp. $\delta_{n,j+1}$) lies in the unordered segment $(X_{\sigma_n(j)}, j/(n+1))$ (resp. $(X_{\sigma_n(j+1)}, (j+1)/(n+1))$) and where f_x and f_{xx} are the first and second derivatives of f with respect to the first coordinate. This leads to expansions for $\Delta_{n,j}^2$ and $\Delta_{n,j}\Delta_{n,j+1}$:

$$\begin{aligned} \Delta_{n,j}^2 &= \left(X_{\sigma_n(j)} - \frac{j}{n+1}\right)^2 \left(f_x\left(W_{n,j}\right) + \frac{1}{2}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) f_{xx}\left(\delta_{n,j}, W_{\sigma_n(j)}\right)\right)^2 \\ \Delta_{n,j}\Delta_{n,j+1} &= \left(X_{\sigma_n(j)} - \frac{j}{n+1}\right)\left(X_{\sigma_n(j+1)} - \frac{j+1}{n+1}\right) \\ &\quad \times \left(f_x\left(W_{n,j}\right) + \frac{1}{2}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) f_{xx}\left(\delta_{n,j}, W_{\sigma_n(j)}\right)\right) \\ &\quad \times \left(f_x\left(W_{n,j+1}\right) + \frac{1}{2}\left(X_{\sigma_n(j+1)} - \frac{j+1}{n+1}\right) f_{xx}\left(\delta_{n,j+1}, W_{\sigma_n(j+1)}\right)\right). \end{aligned}$$

Finally, using the boundedness of f, f_x, and f_{xx}, together with (A.9), (A.10) follows.
Remark that the proof of (A.10) yields also

$$\frac{1}{n}\sum_{j=1}^{n-1} \Delta_{n,j} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \tag{A.12}$$

from which it is clear that this term will contribute in the CLT on Z_n. Then (A.10) entails that the asymptotic study reduces to that of the empirical mean of $Z_{n,j} = B_{n,j} + C_{n,j}$ where

$$B_{n,j} := \begin{pmatrix} f\left(W_{n,j}\right) f\left(W_{n,j+1}\right) \\ f(W_{n,j}) \\ f(W_{n,j})^2 \end{pmatrix} \text{ and } C_{n,j} := \begin{pmatrix} \Delta_{n,j} f\left(W_{n,j+1}\right) + \Delta_{n,j+1} f\left(W_{n,j}\right) \\ \Delta_{n,j} \\ 2\Delta_{n,j} f(W_{n,j}) \end{pmatrix}. \tag{A.13}$$

First, we consider $B_{n,j}$ in (A.13) and we establish the following result, the proof of which has been postponed to Appendix A.2.2.

Lemma A.1. *As $n \to \infty$, the random vector B_n given by*

$$\frac{1}{n}\sum_{j=1}^{n-1} B_{n,j} = \frac{1}{n}\sum_{j=1}^{n-1}\left(f\left(W_{n,j}\right) f\left(W_{n,j+1}\right), f\left(W_{n,j}\right), f\left(W_{n,j}\right)^2\right)^\top$$

satisfies a CLT. More precisely, $\sqrt{n}\left(B_n - m_B\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}_3(0,\Sigma_B)$, where

$$m_B := \left(\mathbb{E}[YY'], \mathbb{E}[Y], \mathbb{E}[Y^2]\right)^\top, \tag{A.14}$$

$Y' = f(X,W')$, W' is an independent copy of W, and Σ_B has an explicit expression given in Appendix A.2.2.

Remark that Y' is the so-called Pick-Freeze version of Y with respect to X. Secondly, we establish a conditional CLT for the empirical mean of the $C_{n,j}$'s defined in (A.13). The reader is referred to Appendix A.2.3 for the proof of this result.

Lemma A.2. *There exists a measurable set $\Pi \in \Omega_W$ having $\mathbb{P}_W$-probability one such that, for any $\omega_W \in \Pi$, we have*

$$\sqrt{n} C_n(\cdot, \omega_W) \xrightarrow[n\to\infty]{\mathscr{L}_X} \mathcal{N}_3(0,\Sigma_C).$$

Moreover, Σ_C does not depend on ω_W and has an explicit expression given Appendix A.2.3.

Considering the characteristic function of the vector $\sqrt{n}(B_n - \mathbb{E}[B_n], C_n)$, one may write

$$\mathbb{E}\left[e^{i\left(\sqrt{n}\langle s,(B_n-\mathbb{E}[B_n])\rangle + \sqrt{n}\langle t, C_n\rangle\right)}\right] = \mathbb{E}\left[e^{i\sqrt{n}\langle s,(B_n-\mathbb{E}[B_n])\rangle}\mathbb{E}\left[e^{i\sqrt{n}\langle t,C_n\rangle}\middle|\mathscr{F}_W\right]\right]$$

for any s and $t \in \mathbb{R}^3$. On the one hand, $\mathbb{E}\left[e^{i\sqrt{n}\langle t,C_n\rangle}\middle|\mathscr{F}_W\right]$ converges a.s. to $\exp\{-t^\top \Sigma_C t/2\}$ which is not random. On the other hand, $\sqrt{n}\langle s,(B_n - \mathbb{E}[B_n])\rangle$ converges in distribution to a Gaussian random variable denoted by B_s. By Slutsky's lemma,

$$\left(\sqrt{n}\langle s,(B_n-\mathbb{E}[B_n])\rangle, \mathbb{E}\left[e^{i\sqrt{n}\langle t,C_n\rangle}\middle|\mathscr{F}_W\right]\right)$$

converges in distribution to $(B_s, \exp\{-t^\top\Sigma_C t/2\})$. We consider the application $h\colon (u,v) \in \mathbb{R}\times D(0,1) \mapsto e^{iu}v \in \mathbb{C}$ where $D(0,1)$ is the unit disc in $\mathbb{C}$. The continuity and the boundedness of h lead to the convergence in distribution of $e^{i\sqrt{n}\langle s,(B_n-\mathbb{E}[B_n])\rangle}\left[e^{i\sqrt{n}\langle t,C_n\rangle}\middle|\mathscr{F}_W\right]$ and we conclude to the asymptotic normality of $\sqrt{n}(B_n - \mathbb{E}[B_n], C_n)$ to a six-dimensional Gaussian random vector with zero mean and variance-covariance matrix $\begin{pmatrix}\Sigma_B & 0\\ 0 & \Sigma_C\end{pmatrix}$. It remains to apply the so-called delta method [122, Theorem 3.1] and Slutsky's lemma to get the required result. The details of the computation of the asymptotic variance σ^2 can be found in Appendix A.2.4.

A.2.2 Proof of Lemma A.1

One has

$$\mathbb{E}[B_n] = \frac{1}{n}\sum_{j=1}^{n-1}\left(\mathbb{E}\left[f\left(W_{n,j}\right) f\left(W_{n,j+1}\right)\right], \mathbb{E}\left[f\left(W_{n,j}\right)\right], \mathbb{E}\left[f\left(W_{n,j}\right)^2\right]\right)^\top,$$

the first coordinate of which converges as $n \to \infty$ to

$$\begin{aligned}\int \mathbb{E}\left[f(x,W) f\left(x',W'\right)\right] \mathrm{d}\mathscr{L}_{(X,X)}(x,x') &= \int_0^1 \mathbb{E}\left[f(x,W) f\left(x,W'\right)\right]\mathrm{d}x\\ &= \mathbb{E}\left[\mathbb{E}\left[f(X,W) f\left(X,W'\right)|X\right]\right]\\ &= \mathbb{E}\left[f(X,W) f\left(X,W'\right)\right] = \mathbb{E}\left[YY'\right].\end{aligned}$$

The two other coordinates can be handled similarly leading to

$$\mathbb{E}[B_n] \underset{n\to\infty}{\to} \left(\mathbb{E}[YY'], \mathbb{E}[Y], \mathbb{E}[Y^2]\right)^\top = m_B.$$

We apply the CLT for dependent variables proved in [90] to $\widetilde{B}^1_{n,j}$, the centered version of the random variables $f(W_{n,j})f(W_{n,j+1})/\sqrt{n}$ with $m=1$, $\alpha=0$, and because f is bounded (so is $\widetilde{B}^1_{n,j}$). Assumptions (1) and (2) in [90] obviously hold, the assumption (3) is naturally fulfilled and assumption (4) is a mere consequence of Chebyshev's inequality and the boundedness of f. Now, it remains to check that assumption (5) holds. We have

$$\begin{aligned}\sum_{i,j=1}^{n-1}\mathrm{Cov}(\widetilde{B}^1_{n,i},\widetilde{B}^1_{n,j}) &= \frac{1}{n}\sum_{i,j=1}^{n-1}\mathrm{Cov}\big(f\left(W_{n,i}\right)f\left(W_{n,i+1}\right),f\left(W_{n,j}\right)f\left(W_{n,j+1}\right)\big)\\ &= \frac{1}{n}\sum_{j=1}^{n-1}\mathrm{Var}\big(f\left(W_{n,j}\right)f\left(W_{n,j+1}\right)\big)+\frac{2}{n}\sum_{j=1}^{n-2}\mathrm{Cov}\big(f\left(W_{n,j}\right)f\left(W_{n,j+1}\right),f\left(W_{n,j+1}\right)f\left(W_{n,j+2}\right)\big).\end{aligned}$$

On the one hand, by [46, Lemma 1.1],

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n-1}\mathrm{Var}\big(f\left(W_{n,j}\right)f\left(W_{n,j+1}\right)\big)\underset{n\to\infty}{\to}\int\mathrm{Var}\big(f\left(x,W\right)f\left(x',W'\right)\big)\,\mathrm{d}\mathscr{L}_{(X,X)}(x,x')\\ &\quad=\int_0^1\mathrm{Var}\big(f\left(x,W\right)f\left(x,W'\right)\big)\,\mathrm{d}x=\mathbb{E}\left[\mathrm{Var}\big(f\left(X,W\right)f\left(X,W'\right)|X\big)\right]=\mathbb{E}\left[\mathrm{Var}\left(YY'|X\right)\right],\end{aligned}$$

where W' is an independent copies of W, $Y=f(X,W)$, and $Y'=f(X,W')$. On the other hand, by [46, Lemma 1.1],

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n-2}\mathrm{Cov}\big(f\left(W_{n,j}\right)f\left(W_{n,j+1}\right),f\left(W_{n,j+1}\right)f\left(W_{n,j+2}\right)\big)\\ &\underset{n\to\infty}{\to}\mathbb{E}\left[\mathrm{Cov}\big(f\left(X,W\right)f\left(X,W'\right),f\left(X,W'\right)f\left(X,W''\right)|X\big)\right]=\mathbb{E}\left[\mathrm{Cov}\left(YY',YY''|X\right)\right],\end{aligned}$$

where W' and W'' are two independent copies of W. Further, $Y=f(X,W)$, $Y'=f(X,W')$, and $Y''=f(X,W'')$. Actually, notice that all linear combination of the coordinates of

$$\left(f(W_{n,j})f(W_{n,j+1}),f(W_{n,j}),f(W_{n,j})^2\right)^\top \tag{A.15}$$

is a one-dependent random variable. In addition, following the same lines as above, one may check that any linear combination still satisfies the assumptions of [90]. Hence, any linear combination of the coordinates of B_n satisfies a CLT so that Lemma A.1 is proved, up to the computation of the asymptotic variance-covariance matrix Σ_B done in what follows.

Computation of the asymptotic covariance matrix Σ_B

We consider a linear combination of the random vector in (A.15) given by

$$uf(W_{n,j})f(W_{n,j+1})+vf(W_{n,j})+wf(W_{n,j})^2,$$

where $(u,v,w)\in\mathbb{R}^3$. This one-dimensional random vector is one-dependent and its centered version normalized by $\sqrt{n}$, denoted by $\widetilde{B}_{n,j}$, satisfies the assumptions of [90]. To calculate the asymptotic variance-covariance matrix Σ_B, we compute explicitly the limit of

$$\sum_{i,j=1}^{n-1}\mathrm{Cov}(\widetilde{B}_{n,i},\widetilde{B}_{n,j}),$$

as $n\to\infty$ using [46, Lemma 1.1]. It remains to take $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ to get the diagonal terms of the asymptotic variance-covariance matrix and to solve a three-dimensional system of equations to get the remaining terms. Finally, as computed previously and using notation of [46, Lemma 1.1], the first diagonal term of Σ_B is :

$$\begin{aligned}\Sigma_B^{1,1}&=\int\mathrm{Var}\big(f\left(x,W\right)f\left(x',W'\right)\big)\,\mathrm{d}\mathscr{L}_{(X,X)}(x,x')\\ &\quad+2\int\mathrm{Cov}\big(f\left(x,W\right)f\left(x',W'\right),f\left(x',W'\right)f\left(x'',W''\right)\big)\,\mathrm{d}\mathscr{L}_{(X,X,X)}(x,x',x'')\\ &=\int_0^1\mathrm{Var}\big(f\left(x,W\right)f\left(x,W'\right)\big)\,\mathrm{d}x+2\int_0^1\mathrm{Cov}\big(f\left(x,W\right)f\left(x,W'\right),f\left(x,W'\right)f\left(x,W''\right)\big)\,\mathrm{d}x\\ &=\mathbb{E}\left[\mathrm{Var}\big(f\left(X,W\right)f\left(X,W'\right)|X\big)\right]+2\mathbb{E}\left[\mathrm{Cov}\big(f\left(X,W\right)f\left(X,W'\right),f\left(X,W'\right)f\left(X,W''\right)|X\big)\right]\\ &=\mathbb{E}\left[\mathrm{Var}\left(YY'|X\right)\right]+2\mathbb{E}\left[\mathrm{Cov}\left(YY',YY''|X\right)\right],\end{aligned}$$

where we remind that $Y = f(X,W)$, $Y' = f(X,W')$, and $Y'' = f(X,W'')$ with W' and W'' independent copies of W. The other terms are

$$\begin{aligned}
\Sigma_B^{2,2} &= \int_0^1 \mathrm{Var}\big(f(x,W)\big)\,\mathrm{d}x = \mathbb{E}\big[\mathrm{Var}\big(f(X,W)\,|X\big)\big] = \mathbb{E}\left[\mathrm{Var}(Y|X)\right],\\
\Sigma_B^{3,3} &= \int_0^1 \mathrm{Var}\big(f(x,W)^2\big)\,\mathrm{d}x = \mathbb{E}\big[\mathrm{Var}\big(Y^2|X\big)\big],\\
\Sigma_B^{1,2} = \Sigma_B^{2,1} &= 2\int_0^1 \mathrm{Cov}\big(f(x,W)\,f\big(x,W'\big), f(x,W)\big)\,\mathrm{d}x = 2\mathbb{E}\big[\mathrm{Cov}\big(YY', Y|X\big)\big],\\
\Sigma_B^{1,3} = \Sigma_B^{3,1} &= 2\int_0^1 \mathrm{Cov}\big(f(x,W)\,f\big(x,W'\big), f(x,W)^2\big)\,\mathrm{d}x = 2\mathbb{E}\big[\mathrm{Cov}\big(YY', Y^2|X\big)\big],\\
\Sigma_B^{2,3} = \Sigma_B^{3,2} &= \int_0^1 \mathrm{Cov}\big(f(x,W), f(x,W)^2\big)\,\mathrm{d}x = \mathbb{E}\big[\mathrm{Cov}(Y, Y^2|X)\big].
\end{aligned}$$

A.2.3 Proof of Lemma A.2

Let $\omega_W \in \Pi$ as defined in [46, Lemma 1.1]. The aim is to establish a CLT for $\sqrt{n}C_{n,j}(\cdot,\omega_W)$. To ease the reading, we omit the notation $(\cdot,\omega_W)$ as classically done in probability. First, dealing with the first coordinate $f\left(W_{n,j+1}\right)\Delta_{n,j} + f\left(W_{n,j}\right)\Delta_{n,j+1}$ of $C_{n,j}$ defined in (A.13), one has

$$\begin{aligned}
f\left(W_{n,j+1}\right)\Delta_{n,j} = &\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) f\left(W_{n,j+1}\right) f_x\left(W_{n,j}\right)\\
&+ \frac{1}{2}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right)^2 f\left(W_{n,j+1}\right) f_{xx}\left(\delta_{n,j}, W_j\right)
\end{aligned}$$

using the expansion of $\Delta_{n,j}$ given in (A.11). By (A.9) and using the boundedness of f and f_{xx}, we get that

$$\frac{1}{n}\sum_{j=1}^{n-1}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right)^2 f\left(W_{n,j+1}\right) f_{xx}\left(\delta_{n,j}, W_j\right)$$

is $O_{\mathbb{P}}(1/n)$. We follow the same lines to treat the term $f\left(W_{n,j}\right)\Delta_{n,j+1}$ and thus

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n-1} f\left(W_{n,j+1}\right)\Delta_{n,j} + f\left(W_{n,j}\right)\Delta_{n,j+1} &= \frac{1}{n}\sum_{j=1}^{n-1}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) f\left(W_{n,j+1}\right) f_x\left(W_{n,j}\right)\\
&+ \frac{1}{n}\sum_{j=1}^{n-1}\left(X_{\sigma_n(j+1)} - \frac{j+1}{n+1}\right) f\left(W_{n,j}\right) f_x\left(W_{n,j+1}\right) + O_{\mathbb{P}}\left(\frac{1}{n}\right)\\
&= \frac{1}{n}\sum_{j=1}^{n-1}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) f_x\left(W_{n,j}\right)\left(f\left(W_{n,j-1}\right) + f\left(W_{n,j+1}\right)\right) + O_{\mathbb{P}}\left(\frac{1}{n}\right).
\end{aligned}$$

So that, using again the expansion of $\Delta_{n,j}$ given in (A.11), (A.9), and the boundedness of f and f_{xx} to handle the second and third coordinate of $C_{n,j}$, the study of C_n reduces to that of the random vector

$$\frac{1}{n}\sum_{j=1}^{n-1}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) f_x\left(W_{n,j}\right)\begin{pmatrix} f\left(W_{n,j-1}\right) + f\left(W_{n,j+1}\right)\\ 1\\ 2f\left(W_{n,j+1}\right)\end{pmatrix} \tag{A.16}$$

by the independence between σ_n and $W_1,\ldots,W_n$. In that view, let us consider the following linear combination $u(f(W_{n,j-1}) + f(W_{n,j+1})) + v + 2wf(W_{n,j+1})$, where $(u,v,w) \in \mathbb{R}^3$ and the empirical mean

$$\frac{1}{n}\sum_{j=1}^{n-1}\left(X_{\sigma_n(j)} - \frac{j}{n+1}\right) f_x\left(W_{n,j}\right) \times \left(u(f(W_{n,j-1}) + f(W_{n,j+1})) + v + 2wf(W_{n,j+1})\right). \tag{A.17}$$

Now it remains to apply [46, Lemma 1.4] [1] with $\chi_j = \left(W_{j-1}, W_j, W_{j+1}\right)$ and $\psi = \psi_{uvw}$ with

$$\psi_{uvw}\left(\frac{j-1}{n+1}, \frac{j}{n+1}, \frac{j+1}{n+1}, \chi_j\right) = f_x\left(W_{n,j}\right)\left(u(f(W_{n,j-1}) + f(W_{n,j+1})) + v + 2wf(W_{n,j+1})\right), \tag{A.18}$$

[1] A slightly generalization of this lemma is required to handle the pair $(j/(n+1),(j+1)/(n+1))$ rather than the quantity j/n. Its proof comes directly following the same lines as in the proof of this lemma

noticing that, as $n \to \infty$, $(1/n)\sum_{j=1}^{n-1} \delta_{(j-1)/(n+1),j/(n+1),(j+1)/(n+1),\chi_j}$ converges in distribution to $Q = \mathscr{L}_{(X,X,X)} \otimes \mathscr{L}_W \otimes \mathscr{L}_W \otimes \mathscr{L}_W$ by [46, Lemma 1.1]. Thus we deduce that the empirical mean in (A.17) converges in distribution for any 3-uplet (u,v,w). Since any linear combination of the components of the random vector defined in (A.16) satisfies a CLT, so does the random vector itself. The proof of Lemma A.2 is now complete, up to the computation of the asymptotic variance-covariance matrix Σ_C done in the paragraph that follows.

Computation of the asymptotic covariance matrix Σ_C

We use the explicit expression (4) in the proof of [46, Lemma 1.4] of the asymptotic variance σ_ψ^2 (actually a slightly generalized version of the lemma) with $Q = \mathscr{L}_{(X,X,X)} \otimes \mathscr{L}_W \otimes \mathscr{L}_W \otimes \mathscr{L}_W$ and with ψ given by (A.18). Then taking the values $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ leads to the diagonal terms of the asymptotic variance-covariance matrix Σ_C while solving a three-dimensional system of equations provides the remaining terms. For instance, reminding that $\chi_j = (W_{j-1}, W_j, W_{j+1})$ and $W_{n,j} = (j/(n+1), W_j)$ and

$$\psi_{100}\left(\frac{j-1}{n+1}, \frac{j}{n+1}, \frac{j+1}{n+1}, \chi_j\right) = f_x\left(W_{n,j}\right)\left(f(W_{n,j-1}) + f(W_{n,j+1})\right)$$

(namely, ψ_{uvw} with $(u,v,w) = (1,0,0)$), we have

$$\begin{aligned}\Sigma_C^{1,1} = &\int \psi_{100}(x_1, x_1', x_1'', \chi_1)\psi_{100}(x_2, x_2', x_2'', \chi_2) x_1 \wedge x_2 \wedge x_1' \wedge x_2' \wedge x_1'' \wedge x_2'' \\ &\times \mathrm{d}Q(x_1, x_1', x_1'', \chi_1)\mathrm{d}Q(x_2, x_2', x_2'', \chi_2) - \left(\int \psi_{100}(x, x', x'', \chi) x \wedge x' \wedge x'' \mathrm{d}Q(x, x', x'', \chi)\right)^2 \\ =&\mathbb{E}[(Y_1 + Y_1')(Y_2 + Y_2') f_x(X_1, W_1) f_x(X_2, W_2)(X_1 \wedge X_2)] - \mathbb{E}[(Y + Y') f_x(X, W) X]^2,\end{aligned}$$

where we remind that $Y = f(X, W)$ and $Y' = f(X, W')$ with W' an independent copy of W (and analogously for Y_1 and Y_2). Finally, the remaining terms of Σ_C are:

$$\begin{aligned}
\Sigma_C^{2,2} &= \mathbb{E}[f_x(X_1, W_1) f_x(X_2, W_2)(X_1 \wedge X_2)] - \mathbb{E}[f_x(X, W) X]^2 \\
\Sigma_C^{3,3} &= 4\mathbb{E}[Y_1' Y_2' f_x(X_1, W_1) f_x(X_2, W_2)(X_1 \wedge X_2)] - 4\mathbb{E}[Y' f_x(X, W) X]^2 \\
\Sigma_C^{1,2} = \Sigma_C^{2,1} &= \mathbb{E}[(Y_1 + Y_1') f_x(X_1, W_1) f_x(X_2, W_2)(X_1 \wedge X_2)] - \mathbb{E}[(Y + Y') f_x(X, W) X]\mathbb{E}[f_x(X, W) X] \\
\Sigma_C^{1,3} = \Sigma_C^{3,1} &= 2\mathbb{E}[(Y_1 + Y_1') f_x(X_1, W_1) Y_2' f_x(X_2, W_2)(X_1 \wedge X_2)] - 2\mathbb{E}[(Y + Y') f_x(X, W) X]\mathbb{E}[Y' f_x(X, W) X] \\
\Sigma_C^{2,3} = \Sigma_C^{3,2} &= 2\mathbb{E}[f_x(X_1, W_1) Y_2' f_x(X_2, W_2)(X_1 \wedge X_2)] - 2\mathbb{E}[f_x(X, W) X]\mathbb{E}[Y' f_x(X, W) X].
\end{aligned}$$

A.2.4 Asymptotic variance σ^2 of Theorem 4.3

We have proved yet that

$$\sqrt{n}\left(\begin{pmatrix} B_n \\ C_n \end{pmatrix} - \begin{pmatrix} m_B \\ 0 \end{pmatrix}\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}_6\left(0, \begin{pmatrix} \Sigma_B & 0 \\ 0 & \Sigma_C \end{pmatrix}\right),$$

where the explicit expressions of m_B, Σ_B and Σ_C are given in (A.14) of Lemma A.1, Appendices A.2.2 and A.2.3 respectively. Applying the so-called delta method [122, Theorem 3.1] to the linear function $f(x,y) = x + y$, we conclude that

$$\sqrt{n}(Z_n - m_B) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}_3(0, \Sigma_B + \Sigma_C) \tag{A.19}$$

Further, we notice that $\xi_n^{\mathrm{Sobol}}(X,Y) \overset{\mathscr{L}}{=} \Psi(Z_n)$ with $\Psi(x,y,z) = (x - y^2)/(z - y^2)$. The so-called delta method [122, Theorem 3.1] then gives

$$\sqrt{N}\left(\xi_n^{\mathrm{Sobol}}(X,Y) - S^X\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathscr{N}_1(0, \sigma^2)$$

where $S^X = \mathrm{Var}(\mathbb{E}[Y|X])/\mathrm{Var}(Y)$ is the first-order Sobol' index with respect to X and $\sigma^2 = g^\top(\Sigma_B + \Sigma_C)g$ with $g = \nabla\Psi(m_B)$. By assumption $\mathrm{Var}(Y) \neq 0$, Ψ is differentiable at m_B and we will see in the sequel that $g^\top(\Sigma_B + \Sigma_C)g \neq 0$, so that the application of the delta method is justified. By differentiation, we get that, for any x, y, and z so that $z \neq y^2$:

$$\nabla\Psi(x,y,z) = \left(\frac{1}{z - y^2}, -2y\frac{z - x}{(z - y^2)^2}, -\frac{x - y^2}{(z - y^2)^2}\right)^\top \tag{A.20}$$

so that

$$g = \nabla\Psi(m_B) = \left(\frac{1}{\mathrm{Var}(Y)}, 2\mathbb{E}[Y]\frac{\mathbb{E}[YY'] - \mathbb{E}[Y^2]}{\mathrm{Var}(Y)^2}, -\frac{S^X}{\mathrm{Var}(Y)}\right)^\top = \frac{1}{\mathrm{Var}(Y)}\left(1, 2\mathbb{E}[Y](S^X - 1), -S^X\right)^\top.$$

Hence the asymptotic variance σ^2 in Theorem 4.3 is finally given by $\sigma^2 = g^\top(\Sigma_B + \Sigma_C)\, g$ where Σ_B and Σ_C have been defined in Appendices A.2.2 and A.2.3 respectively. The matrix Σ_B rewrites as

$$\Sigma_B = \begin{pmatrix} v_{01} + 2c_{01,02} & 2c_{01,03} & 2c_{01,00} \\ 2c_{01,03} & \mathrm{Var}(Y)(1 - S^X) & 2c_{03,00} \\ 2c_{01,00} & 2c_{03,00} & v_{00} \end{pmatrix}$$

where $v_{ij} = \mathbb{E}[\mathrm{Var}(A_i A_j | X)]$, $c_{ij,kl} = \mathbb{E}[\mathrm{Cov}(A_i A_j, A_k A_l | X)]$, $A_0 = Y$, $A_1 = Y'$, $A_2 = Y''$, and $A_3 = 1$ (Y and Y'' have been defined just before (A.15)). The matrix Σ_C rewrites as

$$\Sigma_C = \begin{pmatrix} s^2_{\psi_{100}} & s^2_{\psi_{110}} & s^2_{\psi_{101}} \\ s^2_{\psi_{110}} & s^2_{\psi_{010}} & s^2_{\psi_{011}} \\ s^2_{\psi_{101}} & s^2_{\psi_{011}} & s^2_{\psi_{001}} \end{pmatrix}$$

where s^2_ψ and ψ_{uvw} have been defined in [46, Equation (4)] and (A.18) respectively.

A.3 Proof of the asymtotic efficiency of R^1_n

Proof of Proposition 4.8. By [34, Theorems 3.4 and 3.5] and classical results on efficiency, observe that

$$U_n = \left(\widehat{T}_n, \frac{1}{n}\sum_{i=1}^n Y_i, \frac{1}{n}\sum_{i=1}^n Y_i^2 \right)^\top$$

is asymptotically efficient, componentwise, for estimating $U = \left(\mathbb{E}[\mathbb{E}[Y|X]^2], \mathbb{E}[Y], \mathbb{E}[Y^2]\right)^\top$. The efficiency in product space [122, Theorem 25.50] yields the joint efficiency from this componentwise efficiency. Now, we consider once again the function Ψ introduced in the proof of Theorem 4.3. Since Ψ is differentiable on $\mathbb{R}^3 \setminus \left\{(x,y,z) \,\middle|\, z \neq y^2\right\}$, the efficiency and delta method result [122, Theorem 25.47] implies that $(\Psi(U_n))_n$ is asymptotically efficient for estimating $\Psi(U)$. The conclusion follows as $\Psi(U) = S^X$.

Let us compute the minimal variance. To do so, assume that the joint distribution P of (X,Y) is absolutely continuous with respect to the Cartesian product $P_X \otimes P_Y$, namely $P(dx,dy) = f(x,y)P_X(dx)P_Y(dy)$. Then

$$\mathbb{E}[Y|X=x] = \int y f_{Y|X=x}(y) P_Y(dy) = \int y \frac{f(x,y)}{\int f(x,y)P_Y(dy)} P_Y(dy).$$

For any $t \in (0,1)$, let us introduce $f_t(x,y) := (1 + t h(x,y)) f(x,y)$ and

$$P_t(dx,dy) := (1 + t h(x,y)) f(x,y) P_X(dx) P_Y(dy)$$

where $h(x,y) > -1$ and $\int h(x,y) f(x,y) P_x(dx) P_Y(dy) = 0$. Now we consider the function

$$F(t) := \iint_{x,y'} \left(\frac{\int y f_t(x,y) P_Y(dy)}{\int f_t(x,y) P_Y(dy)} \right)^2 P_t(dx, dy').$$

Denoting by $G(x,t) := \int y f_t(x,y) P_Y(dy) / \int f_t(x,y) P_Y(dy)$, one gets

$$F'(t) = \iint_{x,y'} \left[2G(x,t) \frac{\partial}{\partial t} G(x,t) f_t(x,y') + G(x,t)^2 h(x,y') f(x,y') \right] P_X(dx) P_Y(dy')$$

so that $F'(0) = \langle \mathbb{E}[Y|X=x](2y - \mathbb{E}[Y|X=x]), h \rangle_P$. The interest function $I := \mathbb{E}[Y|X](2Y - \mathbb{E}[Y|X])$ has $\mathbb{E}[\mathbb{E}[Y|X]^2]$ and variance $\mathrm{Var}(\mathbb{E}[Y|X](2Y - \mathbb{E}[Y|X]))$. Hence it remains to apply the delta method to get the final (minimal) variance

$$g^\top \begin{pmatrix} \mathrm{Var}(I) & \mathrm{Cov}(I,Y) & \mathrm{Cov}(I,Y^2) \\ \mathrm{Cov}(I,Y) & \mathrm{Var}(Y) & \mathrm{Cov}(Y,Y^2) \\ \mathrm{Cov}(I,Y^2) & \mathrm{Cov}(Y,Y^2) & \mathrm{Var}(Y^2) \end{pmatrix} g$$

where $g := \nabla\Psi(U)$, and by (A.20),

$$g = \left(\frac{1}{\mathrm{Var}(Y)}, 2\mathbb{E}[Y] \frac{\mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[Y^2]}{\mathrm{Var}(Y)^2}, -\frac{S^X}{\mathrm{Var}(Y)} \right)^\top = \frac{1}{\mathrm{Var}(Y)} \left(1, 2\mathbb{E}[Y](S^X - 1), -S^X\right)^\top.$$

Finally, one gets the minimal variance mentioned in Proposition 4.8. □

Remark A.3. *This result can be also obtained making a LAN perturbation of the functional derivative on the tangent space. In this setting and following the notation of [122, Chapitre 25], let us consider the functional Φ defined by*

$$\Phi(P) := \frac{\mathbb{E}_P[\mathbb{E}_P[Y|X]] - \mathbb{E}_P[Y]^2}{\mathbb{E}_P[Y^2] - \mathbb{E}_P[Y]^2}.$$

Then, with the notation P_t for $t \in (0,1)$ introduced in the above proof, one gets

$$\frac{d}{dt}\Phi(P_t)_{|t=0} = \frac{1}{\mathrm{Var}(Y)}\langle \mathbb{E}[Y|X](2Y - \mathbb{E}[Y|X]) - 2\mathbb{E}[Y]Y - S^X(Y^2 - 2\mathbb{E}[Y]Y), h\rangle_P$$

leading to $\tilde{\Phi} := \frac{1}{\mathrm{Var}(Y)}\left(2\mathbb{E}[Y]Y(1-S^X) + S^X Y^2 - \mathbb{E}[Y|X](\mathbb{E}[Y|X] - 2Y)\right)$ *and the minimal variance is given by* $\sigma^2_{\min} = \mathrm{Var}(\tilde{\Phi}) = \frac{1}{\mathrm{Var}(Y)^2}\mathrm{Var}\left(2\mathbb{E}[Y](1-S^X)Y + S^X Y^2 + \mathbb{E}[Y|X](\mathbb{E}[Y|X] - 2Y)\right)$ *that coincides with the expression obtained via the delta method in Proposition 4.8.*

A.4 Technical results

A.4.1 Convergence of random measures

In the sequel, we will denote by $\mathscr{L}_Z$ the law of a random vector Z.

Lemma A.4. *Let k and $\ell \in [\![0, n]\!]$. There exists a measurable set $\Pi \subset \Omega_W$ with $\mathbb{P}_W$-probability one such that for any $\omega_W \in \Pi$,*

$$\pi_n(\omega_W) := \frac{1}{n}\sum_{j=1}^{n-k\vee\ell} \delta_{\left(\frac{j}{n+1},\dots,\frac{j+k}{n+1},W_j(\omega_W),\dots,W_{j+\ell}(\omega_W)\right)} \Rightarrow \pi := \mathscr{L}_{(X,\dots,X)} \otimes \mathscr{L}_W \otimes \cdots \otimes \mathscr{L}_W,$$

as $n \to \infty$ where as before X is uniformly distributed on $[0,1]$ and $\Rightarrow$ stands for the weak convergence of measures. Here $\mathscr{L}_{(X,\dots,X)}$ stands for the joint distribution of the vector $(X,\dots,X)$ of length k and $\mathscr{L}_W \otimes \cdots \otimes \mathscr{L}_W$ stands for the tensorial product of the distribution $\mathscr{L}_W$ ℓ times.

Proof of Lemma A.4. Let $\omega_W \in \Omega_W$. Let us consider the continuous and bounded functions defined on $\mathbb{R}^{k+\ell}$ by

$$g_{s,t}(x,w) = \exp\{i\langle s, x\rangle_{\mathbb{R}^k} + i\langle t, w\rangle_{\mathbb{R}^\ell}\},$$

for any $s \in \mathbb{R}^k$ and $t \in \mathbb{R}^\ell$. To prove the weak convergence of the measures $(\pi_n(\omega_W))_n$, we show that $\pi_n(\omega_W)(g_{s,t})$ converges almost surely for any $s \in \mathbb{Q}^k$ and $t \in \mathbb{Q}^\ell$ as $n \to \infty$. Finally, we will conclude by density of rational numbers in $\mathbb{R}$.
Let $(s,t) = (s_0,\dots,s_{k-1},t_0,\dots,t_{\ell-1}) \in \mathbb{Q}^{k+\ell}$ be fixed. To ease the reading, we use the shorthand notation g for $g_{s,t}$ and $\langle\cdot,\cdot\rangle$ for $\langle\cdot,\cdot\rangle_{\mathbb{R}^m}$ whatever the value of m and we omit the notation ω_W as classically done in probability.

One has

$$\pi_n(g) = \int g\, d\pi_n = \frac{1}{n}\sum_{j=1}^{n-k\vee\ell} e^{i\langle s,\vec{x}_j\rangle + i\langle t,\vec{W}_j\rangle}$$

where $\vec{x}_j := (j/(n+1),\dots,(j+k-1)/(n+1))$ and $\vec{W}_j = (W_j,\dots,W_{j+\ell-1})$. Obviously, by the independence of the sequence $(W_n)_n$ and the convergence theorem of Riemann sums,

$$\mathbb{E}\left[\pi_n(g)\right] = \prod_{l=0}^{\ell-1}\mathbb{E}\left[e^{it_l W}\right] \times \frac{1}{n}\sum_{j=1}^{n-k\vee\ell} e^{\langle s,\vec{x}_j\rangle} \underset{n\to\infty}{\to} \prod_{l=0}^{\ell-1}\mathbb{E}\left[e^{it_l W}\right] \times \int_0^1 e^{ix\langle s,\vec{1}\rangle}dx,$$

where the vector $\vec{1}$ stands for $(1,\dots,1) \in \mathbb{R}^k$. Observe that the almost sure convergence of π_n is equivalent to the almost sure convergence of its real part and that of its imaginary part. Setting

$$U_{n,j} = \langle s, \vec{x}_j\rangle + \langle t, \vec{W}_j\rangle,$$

we have $\Re(\pi_n(g)) = \frac{1}{n}\sum_{j=1}^{n-2}\cos\left(U_{n,j}\right)$. In order to apply the Borel-Cantelli lemma, we need to control the fourth moment

$$\mathbb{E}\left[\left(\Re(\pi_n(g)) - \mathbb{E}[\Re(\pi_n(g))]\right)^4\right] = \frac{1}{n^4}\mathbb{E}\left[\left(\sum_{j=1}^{n-k\vee\ell}\cos\left(U_{n,j}\right) - \mathbb{E}[\cos\left(U_{n,j}\right)]\right)^4\right].$$

The random variables $\cos\left(U_{n,j}\right) - \mathbb{E}\left[\cos\left(U_{n,j}\right)\right]$ are real-valued, centered, and bounded so that we can apply inequality (2.14) page 37 in [100]. Then we obtain

$$\mathbb{E}\left[\left(\sum_{j=1}^{n-k\vee\ell}\cos\left(U_{n,j}\right) - \mathbb{E}\left[\cos\left(U_{n,j}\right)\right]\right)^4\right] \leqslant 224n^2\left(\Lambda_2(\alpha^{-1})\right)^2 \tag{A.21}$$

where

$$\Lambda_2(\alpha^{-1}) = \sup_{0\leqslant m<n}(m+1)(\alpha_m)^{\frac{1}{2}},$$

where $(\alpha_m)_m$ is the sequence of the strong mixing coefficients of the sequence $(U_{n,j})$. Now since the random variable $U_{n,j}$ only depends on $\overrightarrow{W}_j$, α_m equal zero as soon as $m \geqslant \ell$. Hence, there exists a positive constant K such that

$$\frac{1}{n^4}\mathbb{E}\left[\left(\sum_{j=1}^{n-k\vee\ell}\cos\left(U_{n,j}\right) - \mathbb{E}\left[\cos\left(U_{n,j}\right)\right]\right)^4\right] \leqslant \frac{K}{n^2}.$$

It follows by Borel-Cantelli lemma that the real part of $\pi_n(g)$ converges almost surely. Since the imaginary part can be treated using the exact same steps, the proof of Lemma A.4 is almost complete. Hence, there exists a Borel set $N_{s,t}$ with $\mathbb{P}(N_{s,t}) = 1$ so that the previous convergence holds on $\Omega_W \backslash N_{s,t}$. It remains to define $\Pi := \Omega_W \backslash \cup_{(s,t)\in\mathbb{Q}^{k+\ell}} N_{s,t}$. Obviously, one has $\mathbb{P}(\Pi) = 1$ and the almost sure convergence holds on Π for all functions $g_{s,t}$ with $(s,t) \in \mathbb{Q}^{k+\ell}$.

Finally, the result holding for any uplet $(s,t) \in \mathbb{Q}^{k+\ell}$, we conclude to the required result by density of rational numbers in $\mathbb{R}$. □

A.4.2 Generalized L-Statistics

Lemma A.5. *Let $(E_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables with standard exponential distribution and let ψ be a bounded measurable function on $[0,1]$. We assume that the set of discontinuity points of ψ has null Lebesgue measure. Then, the sequence*

$$\left(n^{-1/2}\sum_{j=1}^{n-1}\psi(j/n)(E_j - 1)\right)_{n\in\mathbb{N}^*}$$

converges in distribution to a centered Gaussian law with asymptotic variance: $\sigma_\psi^2 := \int_{[0,1]}\psi^2(x)dx$.

Proof of Lemma A.5. For $k \in \mathbb{N}^*$, let cum_k denotes the cumulant of order k of

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\psi(j/n)(E_j - 1).$$

Obviously, $\mathrm{cum}_1 = 0$ and, for $k \geqslant 2$, $\mathrm{cum}_k = n^{-k/2}\sum_{j=1}^{n-1}\left(\psi(j/n)\right)^k$. So that, $\lim_{n\to\infty}\mathrm{cum}_2 = \int\psi^2(x)dx$ while, for $k \geqslant 3$, $\lim_{n\to\infty}\mathrm{cum}_k = 0$. □

Remark A.6. *The previous lemma obviously extends to the case of a continuous function $\Psi = (\psi_i)$ valued in $\mathbb{R}^d$ ($d \geqslant 1$). In this case, the asymptotic covariance matrix Σ_Ψ is the Gram matrix $\left(\int_{[0,1]}\psi_i(x)\psi_j(x)dx;\ 1 \leqslant i,j \leqslant d\right)$. Indeed, the previous lemma holds for any linear combination of such random vector sequence. A direct computation of the asymptotic variance leads to the quadratic form built on Σ_Ψ.*

The next lemma is a generalization of the CLT for a L-statistics (see, e.g., [122, Chapter 22]).

Lemma A.7. *Let $(U,\mathbb{B}(U))$ be a Polish space where $\mathbb{B}(U)$ denotes the Borel σ algebra of U. We consider a sequence $(\chi_j)_{1\leqslant j\leqslant n,\, n\in\mathbb{N}^*}$ valued in U and Q a probability measure on $U\times[0,1]$. We assume that the sequence of empirical measures $\left(\frac{1}{n}\sum_{j=1}^{n-1}\delta_{j/n,\chi_j}\right)_{n\in\mathbb{N}^*}$ converges in distribution to Q.*
Let ψ be a bounded measurable real function on $U\times[0,1]$. We assume that the set of discontinuity points of ψ has null Q-probability. Then,

$$D_n := \frac{1}{\sqrt{n}}\sum_{j=1}^{n-1}\psi\left(j/n,\chi_j\right)\left(X_{\sigma_n(j)}-\frac{j}{n+1}\right)\xrightarrow[n\to\infty]{\mathscr{L}}\mathcal{N}\left(0,s_\psi^2\right),$$

where the asymptotic variance s_ψ^2 is given in (A.23).

Proof of Lemma A.7. Recall that the sequence (E_i) has been defined in Lemma A.5. We have

$$\begin{aligned}X_{\sigma_n(j)}-\frac{j}{n+1}&\overset{\mathscr{L}}{=}\frac{\sum_{i=1}^{j}E_i}{\sum_{i=1}^{n+1}E_i}-\frac{j}{n+1}=\frac{1}{\frac{1}{n+1}\sum_{i=1}^{n+1}E_i}\left(\frac{1}{n+1}\sum_{i=1}^{j}E_i-\frac{j}{(n+1)^2}\sum_{i=1}^{n+1}E_i\right)\\&=\frac{1}{\frac{1}{n+1}\sum_{i=1}^{n+1}E_i}\left(\frac{1}{n+1}\sum_{i=1}^{j}(E_i-1)-\frac{j}{(n+1)^2}\sum_{i=1}^{n+1}(E_i-1)\right),\end{aligned}$$

so that,

$$D_n\overset{\mathscr{L}}{=}\frac{1}{\sqrt{n}(n+1)}\frac{1}{\frac{1}{n+1}\sum_{i=1}^{n+1}E_i}\sum_{j=1}^{n-1}\psi\left(j/n,\chi_j\right)\left(\sum_{i=1}^{j}(E_i-1)-\frac{j}{n+1}\sum_{i=1}^{n+1}(E_i-1)\right).$$

Using the assumption on the empirical measure, we get

$$\frac{1}{n}\sum_{j=1}^{n}\psi\left(j/n,\chi_j\right)\frac{j}{n+1}\to I:=\int_{U\times[0,1]}x\psi(x,\chi)dQ(x,\chi).$$

Further, by the weak law of large numbers, $(1/(n+1))\sum_{i=1}^{n+1}E_i$ converges in probability to $\mathbb{E}[E_1]=1$. Hence, by Slutsky's lemma, we are led to consider the random vector

$$V_n:=\frac{1}{\sqrt{n}}\begin{pmatrix}\frac{1}{n+1}\sum_{j=1}^{n-1}\psi\left(j/n,\chi_j\right)\sum_{i=1}^{j}(E_i-1)\\\sum_{i=1}^{n+1}(E_i-1)\end{pmatrix}.$$

Notice that the first coordinate of V_n can be rewritten as (up to the normalizing factor $n^{-1/2}$)

$$\sum_{i=1}^{n-1}\left(\frac{1}{n+1}\sum_{j=1}^{n-1}\psi\left(j/n,\chi_j\right)\mathbb{1}_{i\leqslant j}\right)(E_i-1).$$

For $t\in[0,1]$, let $\phi(t):=\int_{U\times[t,1]}\psi(x,\chi)dQ(x,\chi)$. We will show below that

$$\lim_n\sup_{t\in[0,1]}\left|\left(\frac{1}{n+1}\sum_{j=1}^{n-1}\psi\left(j/n,\chi_j\right)\mathbb{1}_{i\leqslant j}\right)-\phi(t)\right|=0. \tag{A.22}$$

Let assume for a while that this result holds. Then, in our study, we may replace V_n by

$$\widehat{V}_n:=\frac{1}{\sqrt{n}}\begin{pmatrix}\frac{1}{n+1}\sum_{i=1}^{n-1}\phi(i/n)(E_i-1)\\\sum_{i=1}^{n+1}(E_i-1)\end{pmatrix}$$

since (A.22) implies that $\lim_{n\to\infty}\mathbb{E}\|V_n-\widehat{V}_n\|^2=0$. Using Remark A.6, we obtain that the sequence $(\widehat{V}_n)_{n\in\mathbb{N}^*}$ converges in distribution to a centered Gaussian vector with covariance matrix

$$\begin{pmatrix}\int_0^1\phi^2(t)dt & \int_0^1\phi(t)dt\\ \int_0^1\phi(t)dt & 1\end{pmatrix}.$$

Finally, using the so-called delta method [122, Theorem 3.1], $(D_n)_{n\in\mathbb{N}^*}$ converges in distribution to a centered Gaussian variable with variance

$$s_\psi^2=\int_0^1(\phi(t)-I)^2dt. \tag{A.23}$$

It remains to show that (A.22) holds. First let assume that $\psi \geqslant 0$. Set, for $j = 1, \dots n$, $\phi_n(j/n) := (1/(n+1)) \sum_{j=1}^{n-1} \psi\left(j/n, \chi_j\right)$ and consider the piece-wise linear extension ϕ_n defined on $[0,1]$. The second Dini's theorem [103] allows to conclude that the sequence of functions $(\phi_n)_{n\in\mathbb{N}^*}$ converges uniformly to ϕ yielding the result. In the general case, we may mimic this reasoning on $\psi^+ = \sup(\psi, 0)$ and $\psi^- = \sup(-\psi, 0)$ and so conclude. □

Notice that, using the definitions of ϕ and I and applying Fubini's theorem, s_ψ^2 can be explicited as follows:

$$\begin{aligned}
s_\psi^2 &= \int_0^1 (\phi(t) - I)^2 dt = \int_0^1 \left(\int_{U\times[0,1]} \psi(x,\chi)(\mathbb{1}_{t\leqslant x} - x) dQ(x,\chi) \right)^2 dt \\
&= \int_0^1 \iint_{(U\times[0,1])^2} \psi(x_1,\chi_1)\psi(x_2,\chi_2)(\mathbb{1}_{t\leqslant x_1} - x_1)(\mathbb{1}_{t\leqslant x_2} - x_2) dQ(x_1,\chi_1) dQ(x_2,\chi_2) dt \\
&= \iint_{(U\times[0,1])^2} \psi(x_1,\chi_1)\psi(x_2,\chi_2) \int_0^1 (\mathbb{1}_{t\leqslant x_1} - x_1)(\mathbb{1}_{t\leqslant x_2} - x_2) dt dQ(x_1,\chi_1) dQ(x_2,\chi_2) \\
&= \iint_{(U\times[0,1])^2} \psi(x_1,\chi_1)\psi(x_2,\chi_2)(x_1 \wedge x_2 - x_1 x_2) dQ(x_1,\chi_1) dQ(x_2,\chi_2). \qquad \text{(A.24)}
\end{aligned}$$

Appendix B

Proofs of the results of Chapter 5

B.1 Proof of Theorem 5.3

Proof of Theorem 5.3. The consistency follows from a straightforward application of the strong law of large numbers. The asymptotic normality is derived by two successive applications of the delta method [122] .

(1) Let $W_j^1 := (Y_j^{\nu,1},\dots,Y_j^{\nu,p})^T$ $(j=1,\dots,N)$ and g^1 be the mapping from $\mathbb{R}^p$ to $\mathbb{R}^p$ whose l-th coordinate is given by

$$g_l^1(x_1,\dots,x_p) = \binom{p}{l}^{-1} \sum_{\substack{k_1<\dots<k_l \\ k_i\in I_p, i=1,\dots,l}} \left(\prod_{i=1}^{l} x_{k_i}\right).$$

Then $(W_j^1)_{j=1,\dots,N}$ is an i.i.d. sample distributed as $W^1 := (Y^{\nu,1},\dots,Y^{\nu,p})^T$.
Let Σ^1 be the covariance matrix of W_j^1. Clearly, one has $\Sigma^1_{ii} = \mathrm{Var}(Y)$ for $i\in I_p$ while for $i\neq j$, $\Sigma^1_{ij} = \mathrm{Cov}(Y^{\nu,i},Y^{\nu,j}) = \mathrm{Cov}(Y,Y^{\nu,2})$. The multidimensional central limit theorem gives that

$$\sqrt{N}\left(\frac{1}{N}\sum_{j=1}^{N} W_j^1 - m\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}_p\left(0,\Sigma^1\right),$$

where $m := (\mathbb{E}[Y],\dots,\mathbb{E}[Y])^T$. We then apply the so-called delta method to W^1 and g^1 so that

$$\sqrt{N}\left(g^1\left(\overline{W}_N^1\right) - g^1\left(\mathbb{E}\left[W^1\right]\right)\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}\left(0, J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)\Sigma^1 J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)^T\right),$$

where $J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)$ is the Jacobian of g^1 at point $\mathbb{E}\left[W^1\right]$. Notice that for $i\in I_p$ and $k\in I_p$,

$$\frac{\partial g_l^1}{\partial x_k}\left(\mathbb{E}\left[W^1\right]\right) = \frac{\binom{p-1}{l-1}}{\binom{p}{l}} m^{l-1} = \frac{l}{p}\mathbb{E}[Y]^{l-1} =: a_l.$$

Thus $\Sigma^2 := J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)\Sigma^1 J_{g^1}\left(\mathbb{E}\left[W^1\right]\right)^T$ is given by

$$\Sigma^2_{ij} = p a_i a_j \left(\Sigma^1_{11} + (p-1)\Sigma^1_{12}\right).$$

(2) Now consider $W_j^2 := (P_j^{\nu,1},\dots P_j^{\nu,p})^T$ $(j=1,\dots,N)$ and g^2 the mapping from $\mathbb{R}^p$ to $\mathbb{R}$ defined by

$$g^2(y_1,\dots,y_p) = \sum_{l=0}^{p}\binom{p}{l}(-1)^{p-l} y_1^{p-l} y_l.$$

Then $(W_j^2)_{j=1,\dots,N}$ is an i.i.d. sample distributed as $W^2 := (P^{\nu,1},\dots P^{\nu,p})^T$.
We apply once again the delta method to W^2 so that

$$\sqrt{N}\left(g^2\left(\overline{W}_N^2\right) - g^2\left(\mathbb{E}\left[W^2\right]\right)\right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathcal{N}\left(0, J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)\Sigma^2 J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)^T\right),$$

where $J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)$ is the Jacobian of g^2 at point $\mathbb{E}\left[W^2\right]$. Notice that for $k \in I_p$,

$$\frac{\partial g^2}{\partial y_1}\left(\mathbb{E}\left[W^2\right]\right) = (-1)^{p-1} p(p-1)\mathbb{E}[Y]^{p-1} + \sum_{l=2}^{p-1} \binom{p}{l} (-1)^{p-l}(p-l)\mathbb{E}[Y]^{p-l-1}\mathbb{E}\left[\prod_{i=1}^{l} Y^{v,i}\right]$$

and

$$\frac{\partial g^2}{\partial y_l}\left(\mathbb{E}\left[W^2\right]\right) = \binom{p}{l}(-1)^{p-l}\mathbb{E}[Y]^{p-l}.$$

Thus the limiting variance is

$$\sigma^2 := J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)\Sigma^2 J_{g^2}\left(\mathbb{E}\left[W^2\right]\right)^T = p\left(\Sigma_{11}^1 + (p-1)\Sigma_{12}^1\right)\left(\sum_{i=1}^{p} a_i b_i\right)^2,$$

where b_i is the i-th coordinate of $\nabla g^2\left(\mathbb{E}\left[W^2\right]\right)$. □

B.2 An auxiliary result and the proofs of the results of Section 5.2 of Chapter 5

Lemma B.1. *Let G and H be two measurable functions. Let $(U_j)_{j\in I_N}$ and $(V_k)_{k\in I_N}$ be two independent samples of i.i.d. random variables. Assume that $G(U_1, V_1)$ and $H(U_1, U_2, V_1)$ are both integrable and centered. We define S_N and T_N by*

$$S_N = \frac{1}{N^2}\sum_{j,k=1}^{N} G(U_j, V_k) \quad \text{and} \quad T_N = \frac{1}{N^3}\sum_{i,j,k=1}^{N} H(U_i, U_j, V_k).$$

Then S_N and T_N converge almost surely to 0 as N goes to infinity.

Proof of Lemma B.1. Notice that if $\mathbb{E}[S_N^4] = O\left(\frac{1}{N^2}\right)$ then by the Borel - Cantelli lemma, S_N converges almost surely to 0. Now,

$$\mathbb{E}[S_N^4] = \frac{1}{N^8}\sum \mathbb{E}[G(U_{i_1}, V_{j_1})G(U_{i_2}, V_{j_2})G(U_{i_3}, V_{j_3})G(U_{i_4}, V_{j_4})],$$

where the sum is taken over all the indices i_1, i_2, i_3, i_4, j_1, j_2, j_3, j_4 from 1 to N. The only cases leading to terms in $O\left(\frac{1}{N}\right)$ or even in $O(1)$ appear when we sum over indices that are all different except two i's or two j's or over indices that are all different. Nevertheless, in those cases, at least one term of the form $\mathbb{E}[G(U_i, V_j)]$ appears. Since the function G is centered, those cases are then discarded.

The proof of the result concerning T_N follows the same tracks. □

Proof of Corollary 5.7. The proof is based on Lemma B.1. First, we define $Z_j = \left(Z_j^{v,1}, Z_j^{v,2}\right)$,

$$G(Z_j, W_k) = \mathbb{1}_{\{Z_j^{v,1} \leqslant W_k\}}\mathbb{1}_{\{Z_j^{v,2} \leqslant W_k\}},$$
$$F(Z_j, W_k) = \frac{1}{2}\left(\mathbb{1}_{\{Z_j^{v,1} \leqslant W_k\}} + \mathbb{1}_{\{Z_j^{v,2} \leqslant W_k\}}\right),$$
$$H(Z_i, Z_j, W_k) = F(Z_i, W_k)F(Z_j, W_k).$$

Second, we proceed to the following decomposition:

$$\begin{aligned}
\widehat{N}^{v}_{2,CVM} &= \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}} - \left[\frac{1}{2N}\sum_{j=1}^{N}\left(\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}+\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}}\right)\right]^2\right\}\\
&= \frac{1}{N^2}\sum_{j,k=1}^{N}\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}} - \frac{1}{4N^3}\sum_{i,j,k=1}^{N}\left(\mathbb{1}_{\{Z_i^{v,1}\leqslant W_k\}}+\mathbb{1}_{\{Z_i^{v,2}\leqslant W_k\}}\right)\left(\mathbb{1}_{\{Z_j^{v,1}\leqslant W_k\}}+\mathbb{1}_{\{Z_j^{v,2}\leqslant W_k\}}\right)\\
&= \frac{1}{N^2}\sum_{j,k=1}^{N}G(Z_j,W_k) - \frac{1}{N^3}\sum_{i,j,k=1}^{N}H(Z_i,Z_j,W_k)\\
&= \frac{1}{N^2}\sum_{j,k=1}^{N}\{G(Z_j,W_k)-\mathbb{E}[G(Z_j,W_k)]\} - \frac{1}{N^3}\sum_{i,j,k=1}^{N}\{H(Z_i,Z_j,W_k)-\mathbb{E}[H(Z_i,Z_j,W_k)]\}\\
&\quad + \frac{1}{N^2}\sum_{j,k=1}^{N}\mathbb{E}[G(Z_j,W_k)] - \frac{1}{N^3}\sum_{i,j,k=1}^{N}\mathbb{E}[H(Z_i,Z_j,W_k)]\\
&= \frac{1}{N^2}\sum_{j,k=1}^{N}\{G(Z_j,W_k)-\mathbb{E}[G(Z_j,W_k)]\} - \frac{1}{N^3}\sum_{i,j,k=1}^{N}\{H(Z_i,Z_j,W_k)-\mathbb{E}[H(Z_i,Z_j,W_k)]\}\\
&\quad + \mathbb{E}[G(Z_1,W_1)] - \left(1-\frac{1}{N}\right)\mathbb{E}[H(Z_1,Z_2,W_1)] - \frac{1}{N}\mathbb{E}[H(Z_1,Z_1,W_1)].
\end{aligned}$$

The two first sums converge almost surely to 0 by Lemma B.1. The remaining term goes to $\mathbb{E}[G(Z_1,W_1)]-\mathbb{E}[H(Z_1,Z_2,W_1)]$ as N goes to infinity.

It remains to show that $N^{v}_{2,CVM} = \mathbb{E}[G(Z_1,W_1)] - \mathbb{E}[H(Z_1,Z_2,W_1)]$. On the one hand,

$$\begin{aligned}
N^{v}_{2,CVM} &= \int_{\mathbb{R}}\mathbb{E}[(F(t)-F^{v}(t))^2]dF(t) = \mathbb{E}[H_v^2(W_1)]\\
&= \mathbb{E}[\mathrm{Cov}(\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}},\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}})]\\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}] - \mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}]^2].
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}[G(Z_1,W_1)] - \mathbb{E}[H(Z_1,Z_2,W_1)]\\
&= \mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}] - \frac{1}{4}\mathbb{E}[(\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}+\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}})(\mathbb{1}_{\{Z_2^{v,1}\leqslant W_1\}}+\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}})]\\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}]\\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}|W_1]]\\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}|W_1]\mathbb{E}[\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}|W_1]]\\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}|W_1]]\mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}|W_1]]\\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}]\mathbb{E}[\mathbb{1}_{\{Z_2^{v,2}\leqslant W_1\}}]\\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}]] - \mathbb{E}[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}]^2\\
&= \mathbb{E}_W[\mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}\mathbb{1}_{\{Z_1^{v,2}\leqslant W_1\}}] - \mathbb{E}_Z[\mathbb{1}_{\{Z_1^{v,1}\leqslant W_1\}}]^2]
\end{aligned}$$

which completes the proof. □

Proof of Theorem 5.8. We define for $t\in\mathbb{R}$,

$$\begin{aligned}
\mathbb{G}_N^{1,2}(t,t) &= \frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j^{v,1}\leqslant t\}}\mathbb{1}_{\{Z_j^{v,2}\leqslant t\}},\\
\mathbb{G}_N^{i}(t) &= \frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{\{Z_j^{v,i}\leqslant t\}},\ i=1,2,\\
\mathbb{F}_N(t) &= \frac{1}{N}\sum_{k=1}^{N}\mathbb{1}_{\{W_k\leqslant t\}}
\end{aligned}$$

and we rewrite $\widehat{N}^{v}_{2,CVM}$ as a regular function depending on the four empirical processes defined above:

$$\widehat{N}^{v}_{2,CVM} = \int \left[\mathbb{G}^{1,2}_N - \left(\frac{\mathbb{G}^1_N + \mathbb{G}^2_N}{2} \right)^2 \right] d\mathbb{F}_N.$$

By Donsker's theorem,

$$\sqrt{N}\left(\mathbb{G}^{1,2}_N - \widetilde{G}, \mathbb{G}^1_N - F, \mathbb{G}^2_N - F, \mathbb{F}_N - F \right) \underset{N\to\infty}{\overset{\mathscr{L}}{\to}} \mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3, \mathbb{G}_4),$$

where $G(t,s) = \mathbb{P}\left(Z^{v,1} \leqslant t,\ Z^{v,2} \leqslant s\right)$, $\widetilde{G}(t) = G(t,t)$, and $\mathbb{G}$ is a centered Gaussian process of dimension 4 with covariance function defined by

$$\Pi(t,s) = \mathbb{E}\left[A_t A_s^T\right] - \mathbb{E}\left[A_t\right]\mathbb{E}\left[A_s\right]^T \text{ for } \quad (t,s) \in \mathbb{R}^2$$

and $A_t := \left(\mathbb{1}_{\{Z^{v,1}\leqslant t\}}\mathbb{1}_{\{Z^{v,2}\leqslant t\}}, \mathbb{1}_{\{Z^{v,1}\leqslant t\}}, \mathbb{1}_{\{Z^{v,2}\leqslant t\}}, \mathbb{1}_{\{W\leqslant t\}}\right)^T$.

Since these processes are càd làg functions of bounded variation, we introduce the maps $\psi_1,\ \psi_2 : BV_1[-\infty,+\infty]^2 \mapsto \mathbb{R}$ and $\Psi : BV_1[-\infty,+\infty]^4 \mapsto \mathbb{R}$ defined by

$$\psi_i(F_1,F_2) = \int (F_1)^i dF_2,\ i = 1,2, \quad \text{and} \quad \Psi(F_1,F_2,F_3,F_4) = \psi_1(F_1,F_4) - \psi_2\left(\frac{F_2+F_3}{2}, F_4\right),$$

where $BV_M[a,b]$ is the set of càd làg functions of variation bounded by M. Hence,

$$\widehat{N}^{v}_{2,CVM} = \Psi\left(\mathbb{G}^{1,2}_N, \mathbb{G}^1_N, \mathbb{G}^2_N, \mathbb{F}_N\right).$$

Now using the chain rule 20.9 and Lemma 20.10 in [122], the map Ψ is Hadamard-differentiable from the domain $BV_1[-\infty,+\infty]^4$ into $\mathbb{R}$ whose derivative is given by

$$(h_1,h_2,h_3,h_4) \mapsto D\psi_1(F_1,F_4)(h_1,h_4) - D\psi_2\left(\frac{F_2+F_3}{2}, F_4\right)\left(\frac{h_2+h_3}{2}, h_4\right),$$

where the derivatives of ψ_i are given by Lemma 20.10

$$(h_1,h_2) \mapsto h_2\varphi_i \circ F_1|^{+\infty}_{-\infty} - \int h_{2-} d\varphi_i \circ F_1 + \int \varphi_i'(F_1) h_1 dF_2$$

with $\varphi_i(x) = x^i$ and h_- is the left-continuous version of a càd làg function h.

Applying the functional delta method 20.8 in [122] we get the weak convergence of $\sqrt{N}\left(\widehat{N}^{v}_{2,CVM} - N^{v}_{2,CVM}\right)$ to the following limit distribution

$$\int \mathbb{G}_{4-} d(F^2 - \widetilde{G}) + \int \mathbb{G}_1 dF - \int F(\mathbb{G}_2 + \mathbb{G}_3) dF.$$

Since the map Ψ is continuous on the whole space $BV_1[-\infty,+\infty]^4$, the delta method in its stronger form 20.8 in [122] implies that the limit variable is the limit in distribution of the sequence

$$\begin{aligned} &D\Psi(\widetilde{G},F,F,F)\left(\sqrt{N}\left(\mathbb{G}^{1,2}_N - \widetilde{G}, \mathbb{G}^1_N - F, \mathbb{G}^2_N - F, \mathbb{F}_N - F\right)\right) \\ &= \sqrt{N}\left[\int (\mathbb{F}_N - F)_- d\left(F^2 - \widetilde{G}\right) + \int \left(\mathbb{G}^{1,2}_N - \widetilde{G} - F\left(\mathbb{G}^1_N + \mathbb{G}^2_N - 2F\right)\right) dF\right]. \end{aligned}$$

We define

$$\begin{aligned} U &:= \int \mathbb{1}_{\{W<t\}} d(F^2(t) - \widetilde{G}(t) = \widetilde{G}(W) - F(W)^2, \\ V &:= \int \left[\mathbb{1}_{\{Z^{v,1}\leqslant t\}}\mathbb{1}_{\{Z^{v,2}\leqslant t\}} - \left(\mathbb{1}_{\{Z^{v,1}\leqslant t\}} + \mathbb{1}_{\{Z^{v,2}\leqslant t\}}\right) F(t)\right] dF(t) \\ &= \frac{1}{2}\left(F(Z^{v,1})^2 + F(Z^{v,2})^2\right) - F(Z^{v,1} \vee Z^{v,2}). \end{aligned}$$

By independence, the limiting variance ξ^2 is

$$\xi^2 = \mathrm{Var}U + \mathrm{Var}V. \tag{B.1}$$

□

Bibliography

[1] B. Ankenman, B. L. Nelson, and J. Staum. Stochastic kriging for simulation metamodeling. In *2008 Winter Simulation Conference*, pages 362–370. IEEE, 2008.

[2] S. Au. Reliability-based design sensitivity by efficient simulation. *Computers & Structures*, 83:1048–1061, 2005.

[3] J. P. Aubin. *Applied Functional Analysis*. New York: Wiley-Interscience, 2^{nd} edition, 2000.

[4] M. Azadkia and S. Chatterjee. A simple measure of conditional dependence. *arXiv preprint arXiv:1910.12327*, 2019.

[5] J. Azaïs and M. Wschebor. *Level sets and extrema of random processes and fields*. John Wiley & Sons, Inc., Hoboken, NJ, 2009.

[6] S. Azzi, B. Sudret, and J. Wiart. Sensitivity analysis for stochastic simulators using differential entropy. *International Journal for Uncertainty Quantification*, 10(1), 2020.

[7] J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22:773–793, 2012.

[8] J. Bect, L. Li, and E. Vazquez. Bayesian subset simulation. *SIAM/ASA Journal on Uncertainty Quantification*, 5:762–786, 2017.

[9] A. Berlinet and C. Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer, 2004.

[10] P. Bjerager and S. Krenk. Parametric sensitivity in first order reliability theory. *Journal of Engineering Mechanics*, 115:1577–1582, 1989.

[11] S. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics, and kantorovich transport distances. *Memoirs of the American Mathematical Society*, To appear, 2019.

[12] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007.

[13] E. Borgonovo and M. Baucells. Invariant probabilistic sensitivity analysis. *Management Science*, 59(11):2536–2549, 2013.

[14] E. Borgonovo, G. Hazen, and E. Plischke. Probabilistic sensitivity measures: Foundations and estimation. *Submitted*, pages 1–24, 2014.

[15] E. Borgonovo, G. B. Hazen, V. R. R. Jose, and E. Plischke. Probabilistic sensitivity measures as information value. *European Journal of Operational Research*, 289(2):595–610, 2021.

[16] E. Borgonovo and B. Iooss. Moment independent importance measures and a common rationale. 2015.

[17] E. Borgonovo and B. Iooss. Moment independent and reliability-based importance measures. 2016.

[18] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[19] S. Boucheron, G. Lugosi, P. Massart, et al. On concentration of self-bounding functions. *Electronic Journal of Probability*, 14:1884–1899, 2009.

[20] T. Browne, J. Fort, B. Iooss, and L. L. Gratiet. Estimate of quantile-oriented sensitivity indices. *Working paper*, 2017, URL https://hal.archives-ouvertes.fr/hal-01450891.

[21] T. Browne, B. Iooss, L. Le Gratiet, J. Lonchampt, and E. Remy. Stochastic simulators based optimization by gaussian process metamodels - application to maintenance investments planning issues. *Quality and Reliability Engineering International*, 32:2067–2080, 2016.

[22] R. Buchbinder and A. S. Detsky. Management of suspected giant cell arteritis: A decision analysis. *J. Rheumatology*, 19(9):1220–1228, 1992.

[23] D. Bursztyn and D. M. Steinberg. Screening experiments for dispersion effects. In *Screening*, pages 21–47. Springer, 2006.

[24] S. Cambanis, G. Simons, and W. Stout. Inequalities for ek (x, y) when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36(4):285–294, 1976.

[25] V. Chabridon. *Reliability-oriented sensitivity analysis under probabilistic model uncertainty*. PhD thesis, UniversitÃ© Clermont-Auvergne, 2018.

[26] V. Chabridon, M. Balesdent, J.-M. Bourinet, J. Morio, and N. Gayton. Reliability-based sensitivity estimators of rare event probability in the presence of distribution parameter uncertainty. *Reliability Engineering & System Safety*, 178:164–178, 2018.

[27] V. Chabridon, M. Balesdent, G. Perrin, J. Morio, J.-M. Bourinet, and N. Gayton. "Global Reliability-oriented Sensitivity Analysis under Distribution Parameter Uncertainty". In C. Gogu, editor, *Mechanical Engineering Under Uncertainties: From Classical Approaches to Some Recent Developments*, pages 237–277. Wiley - ISTE Ltd, 2021.

[28] S. Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, pages 1–26, 2020.

[29] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[30] L. Cui, Z. Lu, and X. Zhao. Moment-independent importance measure of basic random variable and its probability density evolution solution. *SCIENCE CHINA Technological Sciences*, 53:1138–1145, 2010.

[31] R. Cukier, C. Fortuin, K. Shuler, A. Petschek, and J. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i theory. *The Journal of Chemical Physics*, 59(8):3873, 1973.

[32] S. Da Veiga. Global sensitivity analysis with dependence measures. *J. Stat. Comput. Simul.*, 85(7):1283–1305, 2015.

[33] S. Da Veiga. Kernel-based anova decomposition and shapley effects–application to global sensitivity analysis. *Preprint*, 2021, arXiv:2101.05487.

[34] S. Da Veiga and F. Gamboa. Efficient estimation of sensitivity indices. *Journal of Nonparametric Statistics*, 25(3):573–595, 2013.

[35] S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur. *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. SIAM, 2021.

[36] G. Dellino and C. Meloni. *Uncertainty management in simulation-optimization of complex systems*. Springer, 2015.

[37] H. Dette, K. F. Siburg, and P. A. Stoimenov. A copula-based non-parametric measure of regression dependence. *Scandinavian Journal of Statistics*, 40(1):21–41, 2013.

[38] N. Durrande, D. G. O., Roustant, and L. Carraro. ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis. *Journal of Multivariate Analysis*, 155:57–67, 2013.

[39] K. Elie-Dit-Cosaque. *Développement de mesures d'incertitudes pour le risque de modèle dans des contextes incluant de la dépendance stochastique (in English)*. Thèse de l'Université Claude Bernard Lyon 1, 2020.

[40] J. Felli and G. Hazen. Sensitivity analysis and the expected value of perfect information. *Med. Decis. Making*, 18(1):95–109, 1998.

[41] J. Felli and G. Hazen. Javelin diagrams: A graphical tool for probabilistic sensitivity analysis. *Decision Analysis*, 1(2):93–107, 2004.

[42] J.-C. Fort, T. Klein, and A. Lagnoux. Global sensitivity analysis and wasserstein spaces. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):880–921, 2021.

[43] J.-C. Fort, T. Klein, and N. Rachdi. New sensitivity analysis subordinated to a contrast. *ArXiv e-prints*, May 2013.

[44] R. Fraiman, F. Gamboa, and L. Moreno. Sensitivity indices for output on a Riemannian manifold. *arXiv e-prints*, page arXiv:1810.11591, Oct 2018.

[45] F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374, 2022.

[46] F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. Supplementary material to "global sensitivity analysis: A novel generation of mighty estimators based on rank statistics". *Bernoulli*, 28(4), 2022.

[47] F. Gamboa, A. Janon, T. Klein, and A. Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8:575–603, 2014.

[48] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. Statistical inference for Sobol Pick-Freeze Monte Carlo method. *Statistics*, 50(4):881–902, 2016.

[49] F. Gamboa, T. Klein, and A. Lagnoux. Sensitivity analysis based on Cramér von Mises distance. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):522–548, Apr. 2018.

[50] F. Gamboa, T. Klein, and A. Lagnoux. A central limit theorem for generalized L-statistics. Preprint, 2021.

[51] F. Gamboa, T. Klein, A. Lagnoux, and L. Moreno. Sensitivity analysis in general metric spaces. *Reliability Engineering & System Safety*, 212:107611, 2021.

[52] F. Gamboa, T. Klein, A. Lagnoux, C. Prieur, and S. da Veiga. New estimations of sensitivity indices using kernels. Work in progress, 2021.

[53] E. Giné, R. Nickl, et al. A simple adaptive estimator of the integrated square of a density. *Bernoulli*, 14(1):47–61, 2008.

[54] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer, 2005.

[55] A. Gretton, K. Fukumizu, C. Teo, L. Song, B. Schölkopf, and A. Smola. A kernel statistical test of independence. In *Advances in neural information processing systems*, pages 585–592, 2008.

[56] J. Hart and P. A. Gremaud. Robustness of the Sobol' indices to distributional uncertainty. *International Journal for Uncertainty Quantification*, 9(5):453–469, 2019.

[57] J. L. Hart, A. Alexanderian, and P. A. Gremaud. Efficient computation of Sobol' indices for stochastic models. *SIAM Journal on Scientific Computing*, 39(4):A1514–A1530, 2017.

[58] J. L. Hart and P. A. Gremaud. Robustness of the Sobol' indices to marginal distribution uncertainty. *SIAM/ASA Journal on Uncertainty Quantification*, 7(4):1224–1244, 2019.

[59] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948.

[60] B. Iooss, T. Klein, and A. Lagnoux. Sobol' sensitivity analysis for stochastic numerical codes. In *Proceedings of the SAMO 2016 Conference, Reunion Island, France*, pages 48–49, 2016.

[61] B. Iooss and A. Marrel. Advanced methodology for uncertainty propagation in computer experiments with large number of inputs. *Nuclear Technology*, 205:1588–1606, 2019.

[62] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, 1 2014.

[63] T. Kanamori, S. Hido, and M. Sugiyama. A least-squares approach to direct importance estimation. *The Journal of Machine Learning Research*, 10:1391–1445, 2009.

[64] T. Kim and J. Song. Generalized reliability importance measure (GRIM) using Gaussian mixture. *Reliability Engineering & System Safety*, 173:105–115, 2018.

[65] J. P. Kleijnen. Design and analysis of simulation experiments. In *International Workshop on Simulation*, pages 3–22. Springer, 2015.

[66] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138.1–066138.16, 2004.

[67] F. Kuo, I. Sloan, G. Wasilkowski, and H. Woźniakowski. On decompositions of multivariate functions. *Mathematics of computation*, 79(270):953–966, 2010.

[68] M. Lamboni, H. Monod, and D. Makowski. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96(4):450–459, 2011.

[69] B. Laurent. Efficient estimation of integral functionals of a density. *Ann. Statist.*, 24(2):659–681, 1996.

[70] L. Le Gratiet. Asymptotic normality of a sobol index estimator in gaussian process regression framework. *Preprint*, 2013.

[71] M. Lemaire, A. Chateauneuf, and J.-C. Mitteau. *Structural reliability*. Wiley, 2009.

[72] P. Lemaître, E. Sergienko, A. Arnaud, N. Bousquet, F. Gamboa, and B. Iooss. Density modification-based reliability sensitivity analysis. *Journal of Statistical Computation and Simulation*, 85(6):1200–1223, 2015.

[73] L. Li, Z. Lu, J. Feng, and B. Wang. Moment-independent importance measure of basic variable and its state dependent parameter solution. *Structural Safety*, 38:40–47, 2012.

[74] H. Liu, W. Chen, and A. Sudjianto. Relative entropy based method for probabilistic sensitivity analysis in engineering design. *ASME Journal of Mechanical Design*, 128:326–336, 2006.

[75] H. Madsen. Omission sensitivity factors. *Structural Safety*, 5.1:35–45, 1988.

[76] H. Madsen, S. Krenk, and N. Lind, editors. *Methods of structural safety*. Prentice Hall, 1986.

[77] A. Marrel and V. Chabridon. Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor. *Preprint*, 2020, URL https://hal-cea.archives-ouvertes.fr/hal-02541142.

[78] A. Marrel, B. Iooss, S. Da Veiga, and M. Ribatet. Global sensitivity analysis of stochastic computer models with joint metamodels. *Statistics and Computing*, 22(3):833–847, 2012.

[79] P. Massart. *Concentration inequalities and model selection*. Springer, 2007.

[80] V. Maume-Deschamps and I. Niang. Estimation of quantile-oriented sensitivity indices. *Statistics and Probability Letters*, 134:122–127, 2018.

[81] G. Mazo. An optimal tradeoff between explorations and repetitions in global sensitivity analysis for stochastic computer models. 2019.

[82] C. McDiarmid. On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188, 1989.

[83] A. Meynaoui, A. Marrel, and B. Laurent. New statistical methodology for second level global sensitivity analysis. working paper or preprint, Feb. 2019.

[84] J. Morio. Influence of input pdf parameters of a model on a failure probability estimation. *Simulation Modelling Practice and Theory*, 19(10):2244–2255, 2011.

[85] J. Morio and M. Balesdent. *Estimation of rare event probabilities in complex aerospace and other systems*. Woodhead Publishing, 2016.

[86] V. Moutoussamy, S. Nanty, and B. t. Pauwels. Emulators for stochastic simulation codes. In *CEMRACS 2013—modelling and simulation of complex systems: stochastic and deterministic approaches*, volume 48 of *ESAIM Proc. Surveys*, pages 116–155. EDP Sci., Les Ulis, 2015.

[87] A. Müller. Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.*, 29(2):429–443, 1997.

[88] J. v. Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton, NJ. Princeton University Press, 1953.

[89] J. E. Oakley. Decision-theoretic sensitivity analysis for complex computer models. *Technometrics*, 51(2):121–129, 2009.

[90] S. Orey et al. A central limit theorem for m-dependent random variables. *Duke Mathematical Journal*, 25(4):543–546, 1958.

[91] A. B. Owen. Better estimation of small sobol' sensitivity indices. *ACM Trans. Model. Comput. Simul.*, 23(2):11:1–11:17, May 2013.

[92] A. B. Owen. Variance components and generalized Sobol' indices. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):19–41, 2013.

[93] A. B. Owen, J. Dick, and S. Chen. Higher order Sobol' indices. *Information and Inference*, 3(1):59–81, 2014.

[94] I. Papaioannou and D. Straub. Variance-based reliability sensitivity analysis and the FORM α-factors. *Reliability Engineering & System Safety, In press*, 206(107496), 2021.

[95] C. Park and K. Ahn. A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment. *Reliability Engineering & System Safety*, 46:253–261, 1994.

[96] K. Pearson. On the partial correlation ratio. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 91(632):492–498, 1915.

[97] G. Perrin and G. Defaux. Efficient estimation of reliability-oriented sensitivity indices. *Journal of Scientific Computing*, 79:1433–1455, 2019.

[98] H. Raguet and A. Marrel. Target and conditional sensitivity analysis with emphasis on dependence measures. *Working paper*, 2018, URL https://arxiv.org/abs/1801.10047.

[99] S. Rahman. The f-sensitivity index. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):130–162, 2016.

[100] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31. Springer Science & Business Media, 1999.

[101] G.-C. Rota. On the foundations of combinatorial theory I. theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 2(4):340–368, 1964.

[102] R. Rubinstein. The score function approach for sensitivity analysis of computer simulation models. *Mathematics and Computers in Simulation*, 28:351–379, 1986.

[103] W. Rudin. Real and complex analysis. 1987. *Cited on*, 156, 1987.

[104] A. Saltelli, K. Chan, and E. Scott. *Sensitivity analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2000.

[105] G. Sarazin, J. Morio, A. Lagnoux, M. Balesdent, and L. Brevault. Reliability-oriented sensitivity analysis in presence of data-driven epistemic uncertainty. *Reliability Engineering & System Safety, In press*, 2021.

[106] H. Shi, M. Drton, and F. Han. On the power of chatterjee rank correlation, 2020.

[107] B. Shölkopf, R. Herbrich, and A. Smola. A generalized representer theorem. In *Proceedings of the Annual Conference of Computational Learning Theory*, pages 416–426. London: Springer, 2001.

[108] A. Smola, A. Gretton, L. Song, and B. Schölkopf. A Hilbert space embedding for distributions. In *International Conference on Algorithmic Learning Theory*, pages 13–31. Springer, 2007.

[109] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment*, 1(4):407–414 (1995), 1993.

[110] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[111] L. Song, A. Smola, A. Gretton, J. Bedo, and K. Borgwardt. Feature selection via dependence maximization. *The Journal of Machine Learning Research*, 13:1393–1434, 2012.

[112] A. Spagnol. *Kernel-based sensitivity indices for high-dimnensional optimization problems*. PhD thesis, University Lyon, France, 2020.

[113] A. Spagnol, R. Le Riche, and S. Da Veiga. Global sensitivity analysis for optimization with variable selection. *SIAM/ASA Journal on Uncertainty Quantification*, 7(2):417–443, 2019.

[114] R. Spear and G. Hornberger. Eutrophication in peel inlet II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, 14:43–49, 2019.

[115] B. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. Lanckriet. On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6:1550–1599, 2012.

[116] I. Steinwart, D. Hush, and C. Scovel. An explicit description of the reproducing kernel hilbert spaces of gaussian rbf kernels. *IEEE Transactions on Information Theory*, 52(10):4635–4643, 2006.

[117] J. Stenger, F. Gamboa, and M. Keller. Optimization of quasi-convex function over product measure sets, 2019.

[118] T. Suzuki, M. Sugiyama, J. Sese, and T. Kanamori. Approximating mutual information by maximum likelihood density ratio estimation. *Journal of Machine Learning Research-Proceedings Track*, 4:5–20, 2008.

[119] H. J. ter Horst. On Stieltjes integration in Euclidean space. *J. Math. Anal. Appl.*, 114(1):57–74, 1986.

[120] T. Tjur. *Conditional probability distributions*, volume 2. Institute of Mathematical Statistics, University of Copenhagen, 1974.

[121] W. Trutschnig. On a strong metric on the space of copulas and its induced dependence measure. *Journal of mathematical analysis and applications*, 384(2):690–705, 2011.

[122] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[123] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

[124] G. Wahba, Y. Wang, C. Gu, R. Klein, B. Klein, et al. Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy: the 1994 Neyman Memorial Lecture. *The Annals of Statistics*, 23(6):1865–1895, 1995.

[125] P. Wei, Z. Lu, W. Hao, J. Feng, and B. Wang. Efficient sampling methods for global reliability sensitivity analysis. *Computer Physics Communication*, 183:1728–1743, 2012.

[126] X. Zhu and B. Sudret. Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models. *arXiv preprint arXiv:2005.01309*, 2020.

[127] E. Zio and N. Pedroni. Monte Carlo simulation-based sensitivity analysis of the model of a thermal-hydraulic passive system. *Reliability Engineering & System Safety*, 107:90–106, 2012.