Sensitivity analysis in general metric spaces

Fabrice Gamboa, Thierry Klein, Agnès Lagnoux, Leonardo Moreno

Sensitivity analysis in general metric spaces

Fabrice Gamboa[1], Thierry Klein[2], Agnès Lagnoux[3], and Leonardo Moreno[4]

[1]Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. UT3, F-31062 Toulouse, France.
[3]Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; CNRS. UT2J, F-31058 Toulouse, France.
[2]Institut de Mathématiques de Toulouse; UMR5219. Université de Toulouse; ENAC - Ecole Nationale de l'Aviation Civile , Université de Toulouse, France
[4]Centro de Matemática, Universidad de la República, Uruguay

February 21, 2019

Abstract

Keywords:

1 Introduction

2 General setting

2.1 Notation

It is convenient to have short expressions for terms that converge in probability to zero. We follow [13]. The notation $o_{\mathrm{Pr}}(1)$ (respectively $O_{\mathrm{Pr}}(1)$) stands for a sequence of random variables that converges to zero in probability (resp. is bounded in probability) as $n \to \infty$. More generally, for a sequence of random variables R_n,

$$X_n = o_{\mathrm{Pr}}(R_n) \quad \text{means} \quad X_n = Y_n R_n \quad \text{with} \quad Y_n \overset{\mathrm{Pr}}{\to} 0$$
$$X_n = O_{\mathrm{Pr}}(R_n) \quad \text{means} \quad X_n = Y_n R_n \quad \text{with} \quad Y_n = O_{\mathrm{Pr}}(1).$$

For deterministic sequences X_n and R_n, the stochastic notation reduce to the usual o and O.

In the paper, c stands for a generic constant that may differ from one line to another.

2.2 A new index

We consider a black-box code f from $E := E_1 \times E_2 \times \cdots \times E_d$ valued in some separable metric space $(\mathcal{X}, d)$. The output is denoted by Z given by

$$Z = f(X^{(1)}, \ldots, X^{(d)}). \tag{1}$$

In [5], the authors perform a sensitivity analysis when $\mathcal{X} = \mathbb{R}^k$ on Z based on the whole distribution of Z (instead of considering only its second moment as usually via the so-called Sobol indices). In that view, they introduce a class of test functions parametrized by a single index $t \in \mathbb{R}^k$ and defined by

$$Y_t(Z) = \mathbb{1}_{\{Z \leqslant t\}}.$$

Then they compute

$$\mathbb{E}\left[(\mathbb{E}[Y_t(Z)]-\mathbb{E}[Y_t(Z)|X^v])^2\right]=\mathbb{E}\left[(F(t)-F^v(t))^2\right] \tag{2}$$

and $\mathrm{Var}(Y_t(Z))=F(t)(1-F(t))$ as for the classical Sobol indices. Finally, they integrate both (2) and $\mathrm{Var}(Y_t(Z))$ with respect to the distribution of the output code Z to obtain the Cramér Von Mises index with respect to v by

$$S^v_{2,CVM}:=\frac{\int_{\mathbb{R}^k}\mathbb{E}\left[(F(t)-F^v(t))^2\right]dF(t)}{\int_{\mathbb{R}^k}F(t)(1-F(t))dF(t)}. \tag{3}$$

In this example, the collection of the expectations $\mathbb{E}[Y_t(Z)]=\mathbb{E}[\mathbb{1}_{\{Z\leqslant t\}}]$ is parametrized by a single parameter t. Since its knowledge characterizes the distribution of Z, the previous indices depend as expected on the whole distribution of the output computer code. Using the Pick and Freeze methodology, they propose an estimator which requires $3N$ evaluations of the code for a rate of convergence of $\sqrt{N}$.

This approach has been generalized in [] to compact manifolds replacing the indicator function of half-spaces $\mathbb{1}_{\{Z\leqslant t\}}$ parametrized by t by the indicator function of balls $\mathbb{1}_{\{Z\in B(a,b)\}}$ indexed by two parameters a and b. In their work, $B(a,b)$ stands for the ball of diameter $\overline{ab}$. They also propose a procedure scheme based on $3N$ evaluations of the computer code.

In this paper, we generalize this methodology to separable metric spaces and to classes of test functions parametrized by a fixed number of indices. We prove a central limit theorem for an estimator based on a U-statistics that only requires $2N$ evaluations of the computer code. We also consider a V-statistics and study its asymptotic behavior. This technology can be applied to the framework considered in [5] reducing the computational cost with a U-statistics estimator whose asymptotic behavior can be deduced in an easier way. Similarly, the computational cost is reduced with respect to that in [] and the asymptotic behavior of the estimator is established.

More precisely, we assume that the test functions are parametrized by $m\in\mathbb{N}^*$ elements of $\mathcal{X}$. Hence for any $a=(a_i)_{i=1,\dots,m}\in\mathcal{X}^m$, the test functions

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

are L^2-functions with respect to the product measure $\mathbb{P}^{\otimes m}\otimes\mathbb{P}$ on $\mathcal{X}^m\times\mathcal{X}$. Then we define the general metric space sensitivity index with respect to v by

$$S^v_{2,GMS}:=\frac{\int_{\mathcal{X}^m}\mathbb{E}\left[(\mathbb{E}[Y_a(Z)]-\mathbb{E}[Y_a(Z)|X^v])^2\right]d\mathbb{P}^{\otimes m}(a)}{\int_{\mathcal{X}^m}\mathrm{Var}(Y_a(Z))d\mathbb{P}^{\otimes m}(a)}, \tag{4}$$

where $\mathbb{P}^{\otimes m}$ is the product m-times of the distribution of the output code Z.

Particular cases

1. For $\mathcal{X}=\mathbb{R}$, $m=1$ and Y_a is given by $Y_a(x)=x$, one recovers the classical Sobol indices (see [12, 11]).

2. For $\mathcal{X}=\mathbb{R}^k$ and $m=1$, one can recover the index defined for vectorial outputs in [3, 8] by extending (4) in the following way. We allow the function Y_a to take its values in $\mathcal{X}=\mathbb{R}^k$ so that we set $Y_a(x)=x$ and using (7), we define

$$S^v_{2,GMS}:=\frac{\int_{\mathcal{X}^m}tr\left(\mathrm{Cov}\left(Y_a(Z),Y_a(Z^v)\right)\right)d\mathbb{P}^{\otimes m}(a)}{\int_{\mathcal{X}^m}tr\left(\mathrm{Var}(Y_a(Z))\right)d\mathbb{P}^{\otimes m}(a)}. \tag{5}$$

3. For $\mathcal{X}=\mathbb{R}^k$, $m=1$ and Y_a is given by $Y_a(x)=\mathbb{1}_{\{x\leqslant a\}}$, one recovers the index based on Cramér von Mises distance defined in [5] and recalled in (3).

4. Now consider that $\mathcal{X}=\mathcal{M}$ a manifold, $m=2$ and Y_a is given by $Y_a(x)=\mathbb{1}_{\{x\in B(a_1,a_2)\}}$, where $B(a_1,a_2)$ will stand for the ball of diameter $\overline{a_1a_2}$. Here, one recovers the index defined in []. In some other examples, $B(a_1,a_2)$ will stand for the ball centered at a_1 with radius $\overline{a_1a_2}$.

2.3 Estimation procedure via U-statistics

Following the so-called Pick and Freeze scheme, let X^v be the random vector such that $X^v_v = X_v$ and $X^v_i = X'_i$ if $i \neq v$ where X'_i is an independent copy of X_i. Then, setting

$$Z^v := f(X^v), \tag{6}$$

an obvious computation leads to the following relationship (see, e.g., [7])

$$\mathrm{Var}(\mathbb{E}[Y_a(Z)|X^v]) = \mathrm{Cov}\left(Y_a(Z), Y_a(Z^v)\right).$$

Let us define $\mathbf{Z} = (Z, Z^v)^\top$ and we consider $(\mathbf{Z}_i, i = 1, \dots, m+2)$ $(m+2)$ i.i.d. copies of $\mathbf{Z}$. We denote by $\mathbb{P}^v_2$ the law of $\mathbf{Z} = (Z, Z^v)^\top$. Then the numerator rewrites as

$$\mathbb{E}_{Z_1,\dots,Z_m}\left[\mathrm{Var}(\mathbb{E}[Y_a(Z_{m+1})|X^v])\right] = \mathbb{E}_{Z_1,\dots,Z_m}\left[\mathrm{Cov}_{\mathbf{Z}_{m+1}}(Y_{Z_1,\dots,Z_m}(Z_{m+1}), Y_{Z_1,\dots,Z_m}(Z^v_{m+1}))\right]. \tag{7}$$

Here the notation $\mathbb{E}_Z$ stands for the expectation with respect to the random variable Z.

Now for any $1 \leqslant i \leqslant m+2$, we let $\mathbf{z}_i = (z_i, z^v_i)$ and we define

$$\begin{aligned}
\Phi_1(\mathbf{z}_1, \dots, \mathbf{z}_{m+1}) &:= Y_{z_1,\dots,z_m}(z_{m+1}) Y_{z_1,\dots,z_m}(z^v_{m+1})\\
\Phi_2(\mathbf{z}_1, \dots, \mathbf{z}_{m+2}) &:= Y_{z_1,\dots,z_m}(z_{m+1}) Y_{z_1,\dots,z_m}(z^v_{m+2})\\
\Phi_3(\mathbf{z}_1, \dots, \mathbf{z}_{m+1}) &:= Y_{z_1,\dots,z_m}(z_{m+1})^2\\
\Phi_4(\mathbf{z}_1, \dots, \mathbf{z}_{m+2}) &:= Y_{z_1,\dots,z_m}(z_{m+1}) Y_{z_1,\dots,z_m}(z_{m+2}).
\end{aligned}$$

We set

$$m(1) = m(3) = m+1 \quad \text{and} \quad m(2) = m(4) = m+2 \tag{8}$$

and we define for $j = 1, \dots, 4$,

$$I(\Phi_j) := \int_{\mathcal{X}^{m(j)}} \Phi_j(\mathbf{z}_1, \dots, \mathbf{z}_{m(j)}) d\mathbb{P}_2^{v, \otimes m(j)}(\mathbf{z}_1 \dots, \mathbf{z}_{m(j)}). \tag{9}$$

Finally, we introduce the application Ψ from $\mathbb{R}^4$ to $\mathbb{R}$ defined by

$$\Psi : \begin{array}{ccc} \mathbb{R}^4 & \to & \mathbb{R} \\ (x, y, z, t) & \mapsto & \frac{x-y}{z-t}. \end{array} \tag{10}$$

Then one can express $S^v_{2,GMS}$ in the following way

$$S^v_{2,GMS} := \Psi\left(I(\Phi_1), I(\Phi_2), I(\Phi_3), I(\Phi_4)\right). \tag{11}$$

Following the framework of Hoeffding [6], we replace the functions Φ_1, Φ_2, Φ_3 and Φ_4 by their symmetrized version Φ^s_1, Φ^s_2, Φ^s_3 and Φ^s_4:

$$\Phi^s_j(\mathbf{z}_1, \dots, \mathbf{z}_{m(j)}) = \frac{1}{(m(j))!} \sum_{\tau \in \mathcal{S}_{m(j)}} \Phi_j(\mathbf{z}_{\tau(1)}, \dots, \mathbf{z}_{\tau(m(j))})$$

for $j = 1, \dots, 4$ where $\mathcal{S}_k$ is the symmetric group of degree k. For $j = 1, \dots 4$, the integrals $I(\Phi^s_j)$ are naturally estimated by U-statistics of order $m(j)$. More precisely, we consider a N i.i.d. sample $(\mathbf{Z}_1, \dots, \mathbf{Z}_N)$ with distribution $\mathbb{P}^v_2$ and for $j = 1, \dots, 4$, we define

$$U_{j,N} := \binom{N}{m(j)}^{-1} \sum_{1 \leqslant i_1 < \dots < i_{m(j)} \leqslant N} \Phi^s_j\left(\mathbf{Z}_{i_1}, \dots, \mathbf{Z}_{i_{m(j)}}\right). \tag{12}$$

Theorem 7.1 in [6] ensures that $U_{j,N}$ converges in probability to $I(\Phi_j)$ for any $j = 1, \dots, 4$. Moreover, one may also prove that the convergence holds almost surely proceeding as in the proof of Lemma 6.1 in [5].

We estimate $S^v_{2,GMS}$ by

$$\widehat{S}^v_{2,GMS} := \frac{U_{1,N} - U_{2,N}}{U_{3,N} - U_{4,N}} = \Psi(U_{1,N}, U_{2,N}, U_{3,N}, U_{4,N}). \tag{13}$$

Remark 2.1. Notice that we consider $(m+2)$ copies of $\mathbf{Z}$ in the definition of $S^v_{2,GMS}$ (see (11)). Nevertheless, the estimation procedure only requires a N sample of $\mathbf{Z}$ (see (13)) which means only $2N$ evaluations of the black-box code.

Theorem 2.2. *If for* $j = 1, \ldots, 4$, $\mathbb{E}\left[\Phi^s_j\left(\mathbf{Z}_1, \ldots, \mathbf{Z}_{m(j)}\right)^2\right] < \infty$ *then*

$$\sqrt{N}\left(\widehat{S}^v_{2,GMS} - S^v_{2,GMS}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,\sigma^2) \tag{14}$$

where the asymptotic variance σ^2 *is given by* (22) *in the proof.*

2.4 Comments and particular cases

Considering an output code f, one may consider different choices of the family $(Y_a)_{a\in\mathcal{X}^m}$ of functions indexed by $a \in \mathcal{X}^m$ leading to very different indices. The choice is induced by the aim of the practitioner. To quantify the output sensitivity around the mean, one should consider the classical Sobol indices based on the variance and corresponding to the previous particular case 1. Otherwise, interested in the sensitivity of the whole distribution, one should take a family of functions that characterizes the distribution. For instance, in the previous particular case 3., the functions Y_a are the indicator functions of half-lines and yield the Cramér von Mises indices.
Moreover, since in the estimation procedure the number of output calls is independent of the choice of the family $(Y_a)_{a\in\mathcal{X}^m}$, one can consider and estimate simultaneously several indices with no-extra cost. In fact, the only computational challenge relies in our capability to evaluate the functions Φ at the sample points.

Particular cases

1. For $\mathcal{X} = \mathbb{R}$, $m = 1$ and Y_a is given by $Y_a(x) = x$, we provide a new estimator based on U-statistics of the classical Sobol index. In that case, the estimator is given by (13) and $U_{j,N}$ are given by

$$\begin{aligned}
U_{1,N} &= \frac{1}{N}\sum_{i=1}^N Z_i Z_i^v \\
U_{2,N} &= \frac{1}{N(N-1)}\left(\sum_{i=1}^N Z_i \sum_{i=1}^N Z_i^v - \sum_{i=1}^N Z_i Z_i^v\right) =: U^1_{2,N} - U^2_{2,N} \\
U_{3,N} &= \frac{1}{N}\sum_{i=1}^N Z_i^2 \\
U_{4,N} &= \frac{1}{N(N-1)}\left(\left(\sum_{i=1}^N Z_i\right)^2 - \sum_{i=1}^N Z_i^2\right) =: U^1_{4,N} - U^2_{4,N}
\end{aligned}$$

while in [4], the estimator S^v_N is given by

$$S^v_N := \frac{U_{1,N} - U^1_{2,N}}{U_{3,N} - U^1_{4,N}} = \Psi(U_{1,N}, U^1_{2,N}, U_{3,N}, U^1_{4,N}) \tag{15}$$

that takes into account the diagonal terms. Both procedures require $2N$ evaluations of the black-box code and have the same rate of convergence. The estimators are slightly different which induces different asymptotic variances. Note that if the computer is centered, the procedures are the same. Finally, one may improve the procedures using the information of the whole sample leading to the analog version of the estimation $\widehat{T}^v_N$ given in [4, Eq.(6)].

2. For $\mathcal{X} = \mathbb{R}^k$ and $m = 1$, one may realize the same analogy between the estimation procedure proposed in this paper and that in [3].

3. For $\mathcal{X} = \mathbb{R}^k$, $m = 1$ and Y_a is given by $Y_a(x) = \mathbb{1}_{\{x \leqslant a\}}$, we outperform the CLT proved in [5]. Indeed, the estimator proposed in [5] requires $3N$ evaluations of the computer code while only $2N$ are required in our procedure. In addition, their proof is based on the powerful but complex functional Delta method while the proof of Theorem 2.2 is an elementary application of Theorem 7.1 in [6] combined with the classical Delta method.

3 Beyond the applications in classical frameworks

3.1 Compact manifolds

3.2 Computer codes whose outputs are distribution functions

General setting In some applications, we deal with stochastic codes in the sense that two evaluations of the code for the same input x lead to different outputs. The practitioner is interested in the distribution μ_x of the output for a given x. This type of codes can be traduced in terms of a deterministic code by considering an extra input which is not chosen by the practitioner but which is a latent variable generated randomly by the computer code. In the framework of sensitivity analysis, one consider the inputs as random variables. Then we will construct all the random variables (the one chosen by the practitioner and those generated by the computer code) on the same probability space leading to the application:

$$\begin{aligned} f_s : \quad E \times D \quad &\to \mathbb{R} \\ (x, d) \quad &\mapsto f_s(x, d) \end{aligned} \tag{16}$$

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 whose output is a probability measure:

$$\begin{aligned} f : \quad E \quad &\to \mathcal{M}_2(\mathbb{R}) \\ x \quad &\mapsto \mu_x \end{aligned} \tag{17}$$

where $\mathcal{M}_2(\mathbb{R})$ is the set of the probability measures μ such that $\int x^2 d\mu(x) < +\infty$. Obviously, in practice, one does not assess the output code f but one only obtains a natural approximation of the measure μ_x given by n evaluations of f_s at x, namely,

$$\mu_{x,n} := \frac{1}{n} \sum_{j=1}^{n} \delta_{f_s(x, d_j)}.$$

Concretely, for a single random input $X \in E = E_1 \times \cdots \times E_d$ whose distribution is denoted by $\mathcal{L}$, we will evaluate n times the code f_s defined by (16) so that the the code will generate n variables $D_1, \ldots, D_n$ and one may observe

$$f_s(X, D_1), \ldots, f_s(X, D_n)$$

leading to the measure $\mu_{X,n} = \frac{1}{n} \sum_{j=1}^{n} \delta_{f_s(X, D_j)}$ approximating the distribution of $f_s(X)$. Note that the random variables $D_1, \ldots, D_n$ are not observed.

Sensitivity analysis In order to study the sensitivity of the distribution μ_x, one can use the framework introduced in Section 2.2. In that view, we endowed $\mathcal{M}_2(\mathbb{R})$ with the Wasserstein distance W_2 of order 2. Then (4) becomes

$$S^v_{2,GMS} = \frac{\int_{\mathcal{X}^m} \mathbb{E}\left[\left(\mathbb{E}[\mathbb{1}_{W_2(\mu_1,\mu_X) \leqslant W_2(\mu_1,\mu_2)}] - \mathbb{E}[\mathbb{1}_{W_2(\mu_1,\mu_X) \leqslant W_2(\mu_1,\mu_2)} | X^v]\right)^2\right] d\mathbb{P}^{\otimes 2}(\mu_1, \mu_2)}{\int_{\mathcal{X}^m} \mathrm{Var}(\mathbb{1}_{W_2(\mu_1,\mu_X) \leqslant W_2(\mu_1,\mu_2)}) d\mathbb{P}^{\otimes 2}(\mu_1, \mu_2)}$$

or even $S^v_{2,GMS} = \Psi\left(I(\Phi_1), I(\Phi_2), I(\Phi_3), I(\Phi_4)\right)$ with Ψ and I defined in (10) and (9) and

$$\begin{aligned} \Phi_1(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_3) &= \mathbb{1}_{W_2(\mu_1,\mu_3) \leqslant W_2(\mu_1,\mu_2)} \mathbb{1}_{W_2(\mu_1,\mu_3^v) \leqslant W_2(\mu_1,\mu_2)} \\ \Phi_2(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_4) &= \mathbb{1}_{W_2(\mu_1,\mu_3) \leqslant W_2(\mu_1,\mu_2)} \mathbb{1}_{W_2(\mu_1,\mu_4^v) \leqslant W_2(\mu_1,\mu_2)} \\ \Phi_3(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_3) &= \mathbb{1}_{W_2(\mu_1,\mu_3) \leqslant W_2(\mu_1,\mu_2)} \\ \Phi_4(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_4) &= \mathbb{1}_{W_2(\mu_1,\mu_3) \leqslant W_2(\mu_1,\mu_2)} \mathbb{1}_{W_2(\mu_1,\mu_4) \leqslant W_2(\mu_1,\mu_2)} \end{aligned}$$

where $\boldsymbol{\mu}_i = \boldsymbol{\mu}_{X_i}$ is the concatenation of the measure μ_{X_i} and its Pick and Freeze version denoted by $\mu_{X_i^v}$.

Indices estimation In an ideal scenario which corresponds to the framework of (17), one may asses to the probability measure μ_x for any x. Then following the estimation procedure of Section 2.3, one gets an estimation of the sensitivity index $S^v_{2,GMS}$ with a nice asymptotic behavior given in Theorem 2.2.

In the more realistic framework presented above in (16), 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:

$$\begin{array}{rcl} (X_1, D_{1,1}, \ldots, D_{1,n}) & \to & f_s(X_1, D_{1,1}), \ldots, f_s(X_1, D_{1,n}) \\ (X_1^v, D'_{1,1}, \ldots, D'_{1,n}) & \to & f_s(X_1^v, D'_{1,1}), \ldots, f_s(X_1^v, D'_{1,n}) \\ & \vdots & \\ (X_N, D_{N,1}, \ldots, D_{N,n}) & \to & f_s(X_N, D_{N,1}), \ldots, f_s(X_N, D_{N,n}) \\ (X_N^v, D'_{N,1}, \ldots, D'_{N,n}) & \to & f_s(X_N^v, D'_{N,1}), \ldots, f_s(X_N^v, D'_{N,n}) \end{array}$$

where $2 \times N \times n$ is the total number of evaluations of the stochastic code (16). Then we construct the approximations of μ_i (standing for μ_{X_i}) given by

$$\mu_{i,n} = \frac{1}{n} \sum_{j=1}^{n} \delta_{f_s(X_i, D_{i,j})},$$

for any $i = 1, \ldots, N$. Now for $j = 1, \ldots, 4$, let

$$U_{j,N,n} := \binom{N}{m(j)}^{-1} \sum_{1 \leqslant i_1 < \cdots < i_{m(j)} \leqslant N} \Phi_j^s \left(\boldsymbol{\mu}_{i_1,n}, \ldots, \boldsymbol{\mu}_{i_{m(j)},n} \right) \tag{18}$$

where as previously done Φ^s is the symmetrized version of Φ. Then we estimate $S^v_{2,GMS}$ by

$$\widehat{S}^v_{2,GMS,n} := \frac{U_{1,N,n} - U_{2,N,n}}{U_{3,N,n} - U_{4,N,n}} = \Psi(U_{1,N,n}, U_{2,N,n}, U_{3,N,n}, U_{4,N,n}). \tag{19}$$

Remark 3.1.

1. The estimator in (18) is 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

$$W_2^2(\nu_1, \nu_2) = \frac{1}{n} \sum_{j=1}^{n} (x_{(j)} - y_{(j)})^2,$$

   where $x_{(j)}$ is the j-th order statistics of x.

2. In [2], [9] and [10], the authors deal with stochastic computer codes with probability density functions as outputs. In other words, they define the following application:

$$\begin{aligned} f: \ E &\to \mathcal{F} \\ x &\mapsto f(x) \end{aligned} \tag{20}$$

   where $\mathcal{F}$ is the set of probability density functions:

$$\mathcal{F} := \left\{ g \in L^1(\mathbb{R}); \ g \geqslant 0, \ \int_{\mathbb{R}} g(x)dx = 1 \right\}.$$

Proposition 3.2. *Consider three i.i.d. copies X_1, X_2 and X_3 of X distributed according to $\mathcal{L}$. Let $\delta(N)$ be a sequence tending to 0 as N goes to infinity and such that*

$$\mathbb{P}\left(|W_2(\mu_{X_1}, \mu_{X_3}) - W_2(\mu_{X_1}, \mu_{X_2})| \leqslant \delta(N) \right) = o\left(\frac{1}{\sqrt{N}} \right).$$

We choose n such that $\mathbb{E}[W_2(\mu_X, \mu_{X,n})] = o(\delta(N)/\sqrt{N})$. Under the assumptions of Theorem 2.2, we get

$$\sqrt{N} \left(\widehat{S}^v_{2,GMS,n} - S^v_{2,GMS} \right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2) \tag{21}$$

where the asymptotic variance σ^2 is given by (22) in the proof of Theorem 2.2.

Practical choices of $\delta(N)$. In some particular frameworks, one may derive easily a suitable value of $\delta(N)$. Two examples are given in the following.

- If the inverse of the random variable $W := |W_2(\mu_{X_1}, \mu_{X_3}) - W_2(\mu_{X_1}, \mu_{X_2})|$ has a finite expectation, then, by Markov inequality,

$$\mathbb{P}\left(W \leqslant \delta(N)\right) = \mathbb{P}\left(W^{-1} \geqslant \delta(N)^{-1}\right) \leqslant \frac{1}{\delta(N)}\mathbb{E}\left[\frac{1}{W}\right]$$

  and it suffices to choose $\delta(N)$ so that $\delta(N)^{-1} = o\left(N^{-1/2}\right)$ as N goes to infinity.

- 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 radom variable $W = |W_2(\mu_{X_1}, \mu_{X_3}) - W_2(\mu_{X_1}, \mu_{X_2})|$ 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 $(X_2 + X_3)/2$ and X_1 are two independent random variables uniformly distributed on $[0,1]$. Hence it remains to compute $\mathbb{P}(|U_1 - U_2| \leqslant \alpha)$ for U_1 and U_2 two independent random variables uniformly distributed on $[0,1]$ and $\alpha = \sqrt{\delta(N)}$ and $\alpha = \sqrt{\delta(N)}/2$. It turns out that

$$\mathbb{P}(|U_1 - U_2| \leqslant \alpha) = \alpha(2 - \alpha)$$

  leading to $\mathbb{P}(W \leqslant \delta(N)) = O\left(\sqrt{\delta(N)}\right)$. Consequently, a suitable choice for $\delta(N)$ is $\delta(N) = o(1/N)$.

Practical choices of n. Analogously, one may derive easily suitable choices of the value of n in some particular cases. For instance, we refer the reader to [1] to get upper bounds on $\mathbb{E}[W_p(\mu_X, \mu_{X,n})]$ for several values of $p > 1$ and several assumptions on the distribution on μ_X: géneral, uniform, Gaussian, beta, log concave... Here are some results.

- In the general framework, the upper bound for $p \geqslant 1$ relies on the functional

$$J_p(\mu_X) := \int_{\mathbb{R}} \frac{(F_{\mu_X}(x)(1 - F_{\mu_X}(x)))^{p/2}}{f_{\mu_X}(x)^{p-1)}} dx$$

  where F_{μ_X} is the cumulative distribution function associated to μ_X and f_{μ_X} its probability distribution function. See Cf. [1, Theorems 3.2, 5.1 and 5.3].

- Assume that μ_X is uniformly distributed on $[0,1]$. Then by [1, 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 $p \geqslant 1$ and for any $n \geqslant 1$,

$$\mathbb{E}[W_p(\mu_X, \mu_{X,n})^p]^{1/p} \leqslant (Const)\sqrt{\frac{p}{n}}.$$

  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 [1, Corollaries 6.10 and 6.12], for any $1 \leqslant p < 2$ and for any $n \geqslant 1$,

$$\mathbb{E}[W_p(\mu_X, \mu_{X,n})^p] \leqslant \frac{(Const)}{2-p}\left(\frac{\sigma}{\sqrt{n}}\right)^p,$$

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 $p > 2$ and for any $n \geqslant 1$,

$$\mathbb{E}[W_p(\mu_X, \mu_{X,n})^p] \leqslant \frac{C_p \sigma^p}{n},$$

where C_p depends on p, 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. [1, Corollary 6.11].

Example 3.3. We consider the previous example in which X is uniformly distributed on $[0, 1]$ and μ_X is a Gaussian distribution centered at X with unit variance. Then by [1, 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 $p > 2$ and for any $n \geqslant 3$,

$$\mathbb{E}[W_p(\mu_X, \mu_{X,n})^p] \leqslant \frac{C_p}{n(\log n)^{p/2}},$$

where C_p depends on p, only. Since we already chose $\delta(N) = o(N^{-1})$, it remains to take $\log \log n/n = o(N^{-2})$ to fulfill the condition $\mathbb{E}[W_2(\mu_X, \mu_{X,n})] = o(\delta(N)/\sqrt{N})$.

4 Numerical applications

IN PROCESS

5 Proof

5.1 Proof of Theorem 2.2

The first step of the proof is to apply Theorem 7.1 of [6] to the random vector $(U_{1,N}, U_{2,N}, U_{3,N}, U_{4,N})^\top$. By Theorem 7.1 and Equations (6.1)-(6.3) in [6], it follows that

$$\sqrt{N} \left(\begin{pmatrix} U_{1,N} \\ U_{2,N} \\ U_{3,N} \\ U_{4,N} \end{pmatrix} - \begin{pmatrix} I(\Phi_1^s) \\ I(\Phi_2^s) \\ I(\Phi_3^s) \\ I(\Phi_4^s) \end{pmatrix} \right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

where Γ is the square matrix of size 4 given by

$$\Gamma(i, j) := m(i)m(j)\text{Cov}(\mathbb{E}[\Phi_i^s(\mathbf{Z}_1, \ldots, \mathbf{Z}_{m(i)})|\mathbf{Z}_1], \mathbb{E}[\Phi_j^s(\mathbf{Z}_1, \ldots, \mathbf{Z}_{m(j)})|\mathbf{Z}_1]).$$

Now, it remains to apply the so-called Delta method (see [13]) with the function Ψ defined by (10). We get the asymptotic behavior in (2.2) with σ^2 given by

$$\sigma^2 := g^\top \Gamma g \tag{22}$$

with $g = \nabla\Psi(I(\Phi_1^s), I(\Phi_2^s), I(\Phi_3^s), I(\Phi_4^s))$ and $\nabla\Psi = (z-t)^{-2}\,(z - t, -z + t, -x + y, x - y)^\top$.

5.2 Proof of Proposition 3.2

One has

$$\sqrt{N}\left(\widehat{S}^v_{2,GMS,n} - S^v_{2,GMS}\right) = \sqrt{N}\left(\widehat{S}^v_{2,GMS,n} - \widehat{S}^v_{2,GMS}\right) + \sqrt{N}\left(\widehat{S}^v_{2,GMS} - S^v_{2,GMS}\right).$$

By Theorem 2.2, the second term in the right hand side of the previous equation is asymptotically Gaussian. If we prove that the first term in the right hand side is $o_{\mathbb{P}}(1)$, then by Slutsky's Lemma [13, Lemma 2.8], $\sqrt{N}\left(\widehat{S}^v_{2,GMS,n} - S^v_{2,GMS}\right)$ is asymptotically Gaussian.
Now we prove that $\sqrt{N}\left(\widehat{S}^v_{2,GMS,n} - \widehat{S}^v_{2,GMS}\right) = o_{\mathbb{P}}(1)$. We write

$$\begin{aligned}
&\widehat{S}^v_{2,GMS,n} - \widehat{S}^v_{2,GMS} = \Psi(U_{1,N,n}, U_{2,N,n}, U_{3,N,n}, U_{4,N,n}) - \Psi(U_{1,N}, U_{2,N}, U_{3,N}, U_{4,N}) \\
&= \frac{[(U_{1,N,n} - U_{1,N}) - (U_{2,N,n} - U_{2,N})]\,(U_{3,N} - U_{4,N}) - [(U_{3,N,n} - U_{3,N}) - (U_{4,N,n} - U_{4,N})]\,(U_{1,N} - U_{2,N})}{[(U_{3,N,n} - U_{3,N}) - (U_{4,N,n} - U_{4,N}) + (U_{3,N} - U_{4,N})]\,(U_{3,N} - U_{4,N})}.
\end{aligned}$$

Since $(U_{i,N,n} - U_{i,N,n})$, for $i = 3, 4$ and $(U_{3,N} - U_{4,N})$ converges almost surely respectively to 0 and $I(\Phi_3) - I(\Phi_4)$, the denominator converges almost surely. Thus it suffices to prove that the numerator is $o_{\mathbb{P}}(1/\sqrt{N})$ which reduces to prove that $\sqrt{N}\,(U_{i,N,n} - U_{i,N}) = o_{\mathbb{P}}(1)$ for $i = 1, \ldots, 4$, where $U_{i,N,n}$ (respectively $U_{i,N}$) has been defined in (18) (resp. (12)). Let $i = 1$ for example. The other terms can be treated analogously. Here, $m(1) = 3$. We write

$$\begin{aligned}
&\mathbb{E}\left[|U_{1,N,n} - U_{1,N}|\right] \\
&\leqslant \binom{N}{3}^{-1} (3!)^{-1} \sum_{\substack{1\leqslant i_1<i_2<i_3\leqslant N \\ \tau\in\mathcal{S}_3}} \mathbb{E}\left[\left|\Phi_1\left(\boldsymbol{\mu}_{\tau(i_1),n}, \boldsymbol{\mu}_{\tau(i_2),n}, \boldsymbol{\mu}_{\tau(i_3),n}\right) - \Phi_1\left(\boldsymbol{\mu}_{\tau(i_1)}, \boldsymbol{\mu}_{\tau(i_2)}, \boldsymbol{\mu}_{\tau(i_3)}\right)\right|\right] \\
&= \mathbb{E}\left[\left|\Phi_1\left(\boldsymbol{\mu}_{1,n}, \ldots \boldsymbol{\mu}_{2,n}, \boldsymbol{\mu}_{3,n}\right) - \Phi_1\left(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3\right)\right|\right] \\
&\leqslant 2\mathbb{E}\left[\left|\mathbb{1}_{W_2(\mu_1,\mu_3)\leqslant W_2(\mu_1,\mu_2)} - \mathbb{1}_{W_2(\mu_{1,n},\mu_{3,n})\leqslant W_2(\mu_{1,n},\mu_{2,n})}\right|\right] \\
&=: 2\mathbb{E}\left[B_n\right]
\end{aligned}$$

where the random variable B_n in the expectation in the right hand side of the previous inequality is a Bernoulli random variable whose distribution does not depend on $(1, 2, 3)$. Let $\Delta(N)$ be the following event

$$\Delta(N) := \left\{\left|W_2(\mu_{\tau(1)}, \mu_{\tau(3)}) - W_2(\mu_{\tau(1)}, \mu_{\tau(2)})\right| \geqslant \delta(N)\right\}.$$

Obviously, we get $\mathbb{E}\left[B_n \mathbb{1}_{\Delta(N)^c}\right] \leqslant \mathbb{P}(\Delta(N)^c)$, where A^c stands for the complementary of A in Ω. Furthermore,

$$\begin{aligned}
\mathbb{E}\left[B_n \mathbb{1}_{\Delta(N)}\right] &\leqslant \mathbb{E}\left[B_n|\Delta(N)\right] = \mathbb{P}\left(B_n = 1|\Delta(N)\right) \\
&\leqslant \sum_{i=1}^{3} \mathbb{P}\left(W_2(\mu_i, \mu_{i,n}) \geqslant \frac{\delta(N)}{4}\right) \\
&\leqslant \frac{12}{\delta(N)} \mathbb{E}[W_2(\mu_i, \mu_{i,n})].
\end{aligned}$$

Finally, we introduce $\varepsilon > 0$ and study:

$$\begin{aligned}
\mathbb{P}\left(\sqrt{N}\,|U_{1,N,n} - U_{1,N}| \geqslant \varepsilon\right) &\leqslant \frac{\sqrt{N}}{\varepsilon} \mathbb{E}\left[|U_{1,N,n} - U_{1,N}|\right] \\
&\leqslant 2\frac{\sqrt{N}}{\varepsilon} \mathbb{E}\left[B_n\right] \\
&\leqslant \frac{\sqrt{N}}{\varepsilon} \frac{24}{\delta(N)} \mathbb{E}[W_2(\mu_i, \mu_{i,n})] + 2\frac{\sqrt{N}}{\varepsilon} \mathbb{P}(\Delta(N)^c)
\end{aligned}$$

It remains to choose first, $\delta(N)$ so that $\mathbb{P}(\Delta(N)^c) = o\left(1/\sqrt{N}\right)$ and second, n such that $\mathbb{E}[W_2(\mu_i, \mu_{i,n})] = o(\delta(N)/\sqrt{N})$. Consequently, $\sqrt{N}(U_{1,N,n} - U_{1,N}) = o_{\mathbb{P}}(1)$. Analogously, one gets $\sqrt{N}(U_{i,N,n} - U_{i,N}) = o_{\mathbb{P}}(1)$ for i=2, 3 and 4.

References

[1] S. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics, and kantorovich transport distances. *Memoirs of the American Mathematical Society*, To appear, 2019.

[2] 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.

[3] F. Gamboa, A. Janon, T. Klein, and A. Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8:575–603, 2014.

[4] 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.

[5] F. Gamboa, T. Klein, and A. Lagnoux. Sensitivity analysis based on Cramér–von Mises distance. *SIAM/ASA J. Uncertain. Quantif.*, 6(2):522–548, 2018.

[6] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948.

[7] 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.

[8] 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.

[9] L. Le Gratiet. Asymptotic normality of a sobol index estimator in gaussian process regression framework. *Preprint*, 2013.

[10] 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.

[11] I. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[12] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment*, 1(4):407–414, 1993.

[13] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.