Efficient estimation of Sobol' indices of any order from a single input/output sample

Joint work with Sébastien Da Veiga, Fabrice Gamboa, Thierry Klein, and Clémentine Prieur

Agnès Lagnoux
Institut de Mathématiques de Toulouse
TOULOUSE - FRANCE

Lancement ANR GATSBII, Toulouse, January 30-31, 2025

Outline of the talk

Framework

In this talk, we consider the following black-box model :

$$Y = f(V_1, \ldots, V_p),$$

where $f : \mathscr{E}^p \to \mathbb{R}^k$ is an deterministic and unknown function.

Main assumptions

1. $V_1, \ldots, V_p \in \mathscr{E}$ are independent.
2. $\mathbb{E}[\| Y \|^2] < \infty$.
3. Y is scalar (here, for sake of simplicity).

The so-called Sobol' indices

Quantification of the amount of randomness that a variable or a group of variables bring to Y => so-called Sobol' indices.

For instance, the first order Sobol' and the total Sobol' indices with respect to $V_{\mathbf{u}} = (V_i, i \in \mathbf{u})$ is given by (assuming Y is scalar)

$$S^{\mathbf{u}} = \frac{\mathrm{Var}(\mathbb{E}[Y|V_{\mathbf{u}}])}{\mathrm{Var}(Y)} \quad \textit{and} \quad S^{\mathbf{u},Tot} = 1 - S^{\sim\mathbf{u}} = 1 - \frac{\mathrm{Var}(\mathbb{E}[Y|V_{\sim\mathbf{u}}])}{\mathrm{Var}(Y)}$$

with $\mathbf{u} \subset \{1, \cdots, p\}$ and $\sim \mathbf{u} = \{1, \cdots, p\} \setminus \mathbf{u}$.

Such indices stem from the Hoeffding decomposition of the variance of f (or equivalently Y) that is assumed to lie in L^2.

Pick-Freeze estimation of Sobol' indices (I)

To fix ideas assume for example $p = 5$, $\mathbf{u} = \{1, 2\}$ so that $\sim \mathbf{u} = \{3, 4, 5\}$.
We consider the Pick-Freeze variable $Y^{\mathbf{u}}$ defined as follows :

- draw $V = (V_1, V_2, V_3, V_4, V_5)$,
- build $V^{\mathbf{u}} = (V_1, V_2, V_3', V_4', V_5')$.

Then, we compute

- $Y = f(V)$,
- $Y^{\mathbf{u}} = f(V^{\mathbf{u}})$.

A small miracle

$$\mathrm{Var}(\mathbb{E}[Y|V_{\mathbf{u}}]) = \mathrm{Cov}(Y, Y^{\mathbf{u}}) \text{ so that } S^{\mathbf{u}} = \frac{\mathrm{Cov}(Y, Y^{\mathbf{u}})}{\mathrm{Var}(Y)}.$$

Pick-Freeze estimation of Sobol' indices (II)

In practice, generate two n-samples :

- one n-sample of V : $(V_j)_{j=1,\ldots,n}$,
- one n-sample of $V^{\mathbf{u}}$: $\left(V_j^{\mathbf{u}}\right)_{j=1,\ldots,n}$.

Compute the code on both samples :

- $Y_j = f(V_j)$ for $j = 1, \ldots, n$,
- $Y_j^{\mathbf{u}} = f(V_j^{\mathbf{u}})$ for $j = 1, \ldots, n$.

Then estimate $S^{\mathbf{u}}$ by

$$S_{n,PF}^{\mathbf{u}} = \frac{\frac{1}{n}\sum_{j=1}^{n} Y_j Y_j^{\mathbf{u}} - \left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right)\left(\frac{1}{n}\sum_{j=1}^{n} Y_j^{\mathbf{u}}\right)}{\frac{1}{n}\sum_{j=1}^{n}(Y_j)^2 - \left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right)^2}$$

Pick-Freeze scheme (III) : some statistical properties

Is the Pick-Freeze estimator of the Sobol' index is "good"?

- Is it consistent ? YES SLLN.
- If yes, at which rate of convergence ? YES CLT (cv in $\sqrt{n}$).
- Is it asymptotically efficient ? YES.
- Is it possible to measure its performance for a fixed n ? YES Berry-Esseen and/or concentration inequalities.

Ref. : A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. " Asymptotic normality et efficiency of a Sobol' index estimator", *ESAIM P&S*, 2013.

F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. " Statistical Inference for Sobol' Pick Freeze Monte Carlo method", *Statistics*, 2015.

Drawbacks of the Pick-Freeze estimation

- The cost ($=$ number of evaluations of the function f) of the estimation of the p first-order Sobol' indices is quite expensive : $(p+1)n$.
- This methodology is based on a particular design of experiment that may not be available in practice. For instance, when the practitioner only has access to real data.

☞ *We are interested in an estimator based on a n-sample only.*

Mighty estimation based on ranks (I)

Here we assume that

the inputs V_i for $i = 1, \ldots, p$ are scalar ($\dim(\mathscr{E}) = d = 1$)

and we want to estimate the Sobol' index with respect to $X = V_i$:

$$S^i = \frac{\mathrm{Var}(\mathbb{E}[Y|V_i])}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}(\mathbb{E}[Y|X])}{\mathrm{Var}(Y)}.$$

To do so, we consider a n-sample of the input/output pair (X, Y) given by

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n).$$

The pairs $(X_{(1)}, Y_{(1)}), (X_{(2)}, Y_{(2)}), \ldots, (X_{(n)}, Y_{(n)})$ are rearranged in such a way that

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

Mighty estimation based on ranks (II)

We introduce

$$S^i_{n,Rank} = \frac{\frac{1}{n}\sum_{j=1}^{n-1} Y_{(j)} Y_{(j+1)} - \left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right)^2}{\frac{1}{n}\sum_{j=1}^{n} Y_j^2 - \left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right)^2}.$$

Statistical properties - only for $d = 1$ and first-order Sobol' indices

- Consistency : OK.
- Central Limit Theorem : OK.

Ref. : S. Chatterjee. "A new coefficient of Correlation", *JASA*, 2020.

F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. " Global Sensitivity Analysis : a new generation of mighty estimators based on rank statistics", *Bernoulli.* 2022.

Efficient estimation based on kernels

Here again we assume that the inputs V_i for $i = 1, \ldots, p$ are scalar.

To do so, the initial n-sample is split into two samples of sizes

- $n_1 = \lfloor n / \log n \rfloor \Rightarrow$ estimation of the joint density of (X, Y)
- $n_2 = n - n_1 \approx n \Rightarrow$ Monte-Carlo estimation of the integral involved in the quantity of interest.

Statistical properties - only for $d = 1$ and first-order Sobol' indices

- Consistency : OK.
- Central Limit Theorem : OK.
- Asymptotic efficiency : OK.

Ref. : S. Da Veiga and F. Gamboa. "Efficient estimation of sensitivity indices", *Journal of Nonparametric Statistics*, 2013.

Estimation based on nearest neighbors

Here the input X with respect we want to compute the Sobol' index is allowed to have dimension $d \geqslant 1$.

To do so, the initial n-sample is split into two samples of sizes

- $n/2 \Rightarrow$ estimation of the regression function $m(x) = \mathbb{E}[Y|X = x]$ using the first NN of x among the points of the first sample ;
- $n/2 \Rightarrow$ plug-in estimator.

Statistical properties

- Consistency : OK.
- Central Limit Theorem : OK only for $d \leqslant 3$.

Ref. : L. Devroye, L. Györfi, G. Lugosi, and H. Walk. "A nearest neighbor estimate of the residual variance", *EJS*, 2018.

Outline of the talk

Introduction

Recall that

$$S^X = \frac{\mathrm{Var}(\mathbb{E}[Y|X])}{\mathrm{Var}(Y)} = \frac{\mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[Y]^2}{\mathrm{Var}(Y)}$$

allowing a multidimensional X : $X \in \mathscr{D} = [0,1]^d$.

To estimate $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$ from the n-sample $(Y_j)_{j=1,\dots,n}$ of the output Y, we will naturally use the classical empirical mean and variance respectively.

☞ *Thus we focus on the estimation of* $T = \mathbb{E}[\mathbb{E}[Y|X]^2]$ *from the n-sample* $(X_j, Y_j)_{j=1,\dots,n}$ *of the pair* (X, Y).

Ingredients and our estimator

We propose an estimator of T based on two main ingredients :

1. estimation based on the efficient influence function of T,
2. mirror-type kernel estimators.

Let us start by giving the final form of our estimators :

$$\widehat{T}_n = \frac{1}{n} \sum_{i=1}^{n} (2Y_i - \widehat{m}_n(X_i))\widehat{m}_n(X_i),$$

with $\widehat{m}_n$ a kernel-based smoothed est. of m : $m(x) = \mathbb{E}[Y|X = x]$.

There are two equivalent ways for deriving such a formulation, both of them relying on the efficient influence function of T.

Ingredients and our estimator

In our case, the efficient influence function at any $P \in \mathscr{P}$ writes

$$\widetilde{\psi}_P(x,y) = (2y - m(x))m(x) - \psi(P).$$

where m is the regression function under P : $m(x) = \mathbb{E}_P[Y|X = x]$, see Klein, Lagnoux, Rochet (2024).

Then, if m under P_0 is known, taking

$$T_{n,oracle} = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - m(X_i))m(X_i)$$

leads to an asymptotically efficient estimator of T.

Regression function estimation

The domain of the inputs being compact, the crucial point is to handle possible boundary effects.

To do so, Doksum and Samorov (1995) estimate a truncated version of T defined as

$$T^{\text{trunc},\varepsilon} = \mathbb{E}[\mathbb{E}[Y|X]^2 \mathbb{1}_{X\in(\varepsilon,1-\varepsilon)^d}].$$

Even if $T^{\text{trunc},\varepsilon} \to T$ as $\varepsilon \to 0$ under mild assumptions, the practical tuning of the parameter ε depends on the unknown function f and its choice has a large impact.

Here, we therefore focus on mirror-type kernel estimators to estimate T rather than a truncated version of it. Such mirror-type estimators have been proposed recently to efficiently handle boundary effects inherent to kernel estimation.

Multi-index notation and smoothness

For any d and $\beta = (\beta_1, \ldots, \beta_d) \in \mathbb{R}_+^d$, we define its integer part γ by

$$\gamma := \lfloor \beta \rfloor = (\lfloor \beta_1 \rfloor, \ldots, \lfloor \beta_d \rfloor) \in \mathbb{N}^d.$$

In addition, we introduce, for any $v \in \mathbb{R}^d$,

$$|\gamma| = \gamma_1 + \cdots + \gamma_d, \quad \gamma! = \gamma_1! \ldots \gamma_d!, \quad \text{and} \quad v^\beta = v_1^{\beta_1} \ldots v_d^{\beta_d}.$$

Let $\alpha > 0$. We define $\mathscr{C}^\alpha(\mathscr{D}) = \{\phi : \mathscr{D} \to \mathbb{R}$ with derivatives up to order $\lfloor \alpha \rfloor$ and partial derivative of order $\lfloor \alpha \rfloor$ is $\alpha - \lfloor \alpha \rfloor$-Hölder$\}$.
Namely, there exists $C_\phi > 0$ such that, for any x and $x' \in \mathscr{D}$, one has

$$\left| \frac{\partial^\beta \phi}{\partial x^\beta}(x) - \frac{\partial^\beta \phi}{\partial x^\beta}(x') \right| \leqslant C_\phi \left\| x - x' \right\|_\infty^{\alpha - \lfloor \alpha \rfloor}$$

for any $\beta \in \mathbb{N}^d$ such that $|\beta| = \lfloor \alpha \rfloor$.

Assumptions

($\mathscr{A}1$) Support - The support of $(V_1, \ldots, V_p)$ is $[0,1]^p$ and that of X is $[0,1]^d$.

($\mathscr{A}2$) Absolute continuity - X is absolutely continuous with respect to the Lebesgue measure on $[0,1]^d$ with density function f_X and $\exists \delta > 0$ such that $\inf_{x \in [0,1]^d} f_X(x) \geqslant \delta$ for some $\delta > 0$.

($\mathscr{A}3$) Bounded moments - $\mathbb{E}[Y^4] < \infty$ and $\sigma^2(x) = \mathrm{Var}(Y|X = x)$ is bounded on $[0,1]^d$.

($\mathscr{A}4$) Smoothness of f_X - The density f_X of X belongs to $\mathscr{C}^\alpha([0,1]^d)$ for some $\alpha > 0$.

($\mathscr{A}5$) Smoothness of m - The regression function m belongs to $\mathscr{C}^\alpha([0,1]^d)$.

Assumptions

($\mathscr{A}$6) **Kernel** - Let $k : [0,1] \to \mathbb{R}$ be a bounded : $\|k\|_\infty < \infty$, univariate kernel of order $(\lfloor\alpha\rfloor + 1)$:

$$\int_0^1 u^\ell k(u)du = 0, \text{ for any } \ell \in \mathbb{N} \text{ such that } 0 < \ell \le \lfloor\alpha\rfloor$$
$$\int_0^1 u^{\lfloor\alpha\rfloor+1} k(u)du \neq 0, \quad \text{and} \quad \int_0^1 k(u)du = 1.$$

Finally,

$$K_h(u) = \frac{1}{h^d} K\Big(\frac{u}{h}\Big) = \frac{1}{h^d} \prod_{k=1}^d k\Big(\frac{u_k}{h}\Big), \forall u = (u_1, \dots, u_d) \in [0,1]^d.$$

($\mathscr{A}$7) **Bandwidth** - The sequence $(h_n)_{n\in\mathbb{N}}$ of bandwidths is positive and such that $h_n \to 0$ as $n \to \infty$.

First mirror-type transformation

As Bertin and co-authors (2020), for $x \in [0,1]^d$, we consider

$$A_x : \begin{cases} \mathbb{R}^d & \to & \mathbb{R}^d \\ u = (u_1, \ldots, u_d) & \mapsto & A_x(u) = (a_1(x_1)u_1, \ldots, a_d(x_d)u_d) \end{cases}$$

with $a_i(s) := 1 - 2\mathbb{1}_{(\frac{1}{2},1]}(s) \in \{-1, 1\}$.

Observe that $\mathscr{A} = \{A_x, x \in [0,1]^d\}$ is a finite subset of $GL_d(\mathbb{R})$ (where $GL_d(\mathbb{R})$ is the general linear group on $\mathbb{R}$), $\mathscr{A} = \{A_1, \ldots, A_\kappa\}$, with cardinality $\kappa = 2^d$. Moreover, it satisfies

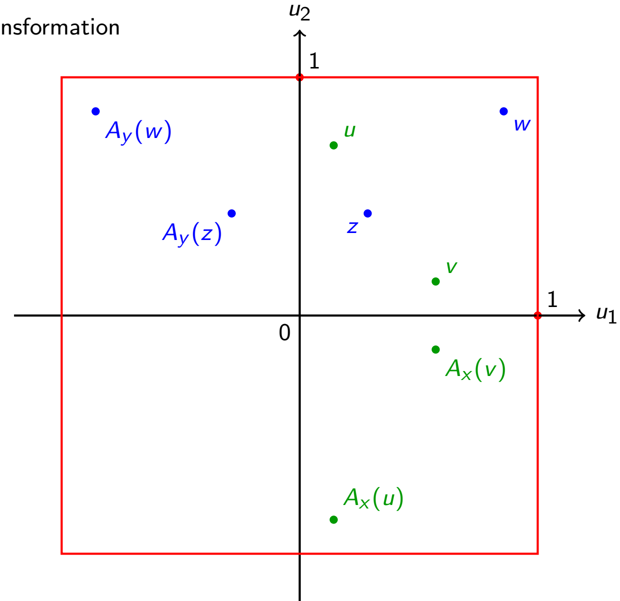
(i) for any $\ell = 1, \ldots, \kappa$, $\left|\det(A_\ell)\right| = 1$;

(ii) Mirror property :

$$\forall x \in [0,1]^d,\ x + A_x^{-1}([0,1/2]^d) \subset [0,1]^d.$$

First mirror-type transformation

$\mathscr{D} = [0,1]^2$
$x = (1/3, 3/4)$
$y = (2/3, 1/5)$



First mirror-type estimation

To estimate the regression function m, we consider a leave-one-out kernel estimator :

$$\widehat{m}_{n,h_n,i}(X_i) = \frac{\sum_{j\neq i} Y_j K_{h_n} \circ A_{X_i}(X_j - X_i)}{\sum_{j\neq i} K_{h_n} \circ A_{X_i}(X_j - X_i)} = \frac{\widehat{g}_{n,h_n,i}(X_i)}{\widehat{f}_{n,h_n,i}(X_i)}$$

Then, our first estimator is given by

$$\widehat{T}_{n,h_n} = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - \widehat{m}_{n,h_n,i}(X_i))\widehat{m}_{n,h_n,i}(X_i).$$

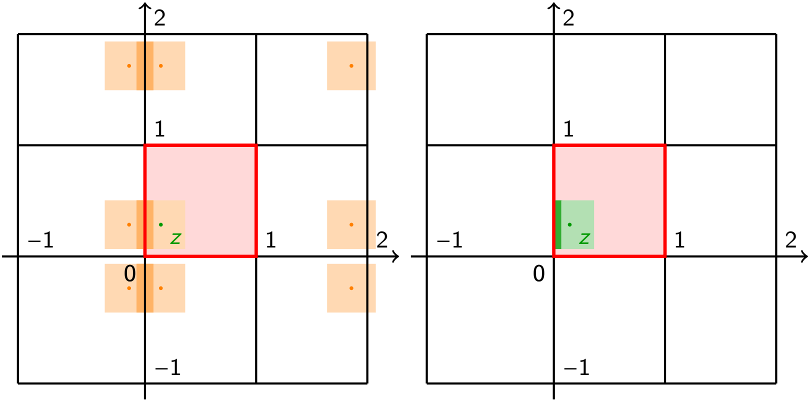
Second mirror-type transformation

As Pujol (2022), we consider the following transformations : for any $z \in [0,1]$,

$$m^{-1}(z) = -z, \quad m^0(z) = z, \quad \textit{and} \quad m^1(z) = 2 - z$$

and, for any $a \in \{-1,0,1\}^d$ and $x \in [0,1]^d$, the d-dimensional vector

$$M^a(x) = (m^{a_1}(x_1), \cdots, m^{a_d}(x_d)).$$


2
1
−1
1
2
0
−1
z
2
1
−1
1
2
0
−1
z

Second mirror-type transformation

($\mathscr{A}'4$) Smoothness of f_X - $f_X \in \mathscr{C}^{\alpha}([0,1]^d)$ for some $\alpha > 0$ and its derivatives of order β $(0 < \beta \leq \lfloor \alpha \rfloor)$ vanish near the boundary.

($\mathscr{A}'6$) Kernel - Let $\widetilde{k} : [-1,1] \to \mathbb{R}$ be a bounded : $\|\widetilde{k}\|_{\infty} < \infty$, univariate kernel of order $(\lfloor \alpha \rfloor + 1)$:

$$\int_{-1}^{1} u^{\ell} \widetilde{k}(u) du = 0, \textit{ for any } \ell \in \mathbb{N} \textit{ such that } 0 < \ell \leq \lfloor \alpha \rfloor$$

$$\int_{-1}^{1} u^{\lfloor \alpha \rfloor + 1} \widetilde{k}(u) du \neq 0, \quad \textit{and} \quad \int_{-1}^{1} \widetilde{k}(u) du = 1.$$

Finally,

$$\widetilde{K}_h(u) = \frac{1}{h^d} \widetilde{K}\left(\frac{u}{h}\right) = \frac{1}{h^d} \prod_{k=1}^{d} \widetilde{k}\left(\frac{u_k}{h}\right), \forall u \in [-1,1]^d.$$

Second mirror-type transformation

Now we propose the following regression function estimator :

$$\widetilde{m}_{n,h_n,i}(X_i) = \frac{\sum_{j\neq i} Y_j \sum_{a\in\{-1,0,1\}^d} \widetilde{K}_{h_n}(M^a(X_j) - X_i)}{\sum_{j\neq i} \sum_{a\in\{-1,0,1\}^d} \widetilde{K}_{h_n}(M^a(X_j) - X_i)} = \frac{\widetilde{g}_{n,h_n,i}(X_i)}{\widetilde{f}_{n,h_n,i}(X_i)}.$$

The associated plug-in estimator then becomes :

$$\widetilde{T}_{n,h_n} = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - \widetilde{m}_{n,h_n,i}(X_i))\widetilde{m}_{n,h_n,i}(X_i).$$

Under the previous assumptions and an additional technical one, for all $i \in \{1, \cdots, d\}$, we get :

- bias and variance controls

$$\left\| \mathbb{E}\left[\widehat{f}_{n,h_n,i}\right] - f_X \right\|_\infty = O\left(h_n^\alpha\right),$$

$$\mathbb{E}\left[\int_{[0,1]^d} (\widehat{f}_{n,h_n,i}(x) - f_X(x))^2 dx\right] = o(n^{-1/2}),$$

- lower control

$$\frac{1}{\inf_{x \in [0,1]^d} \left|\widehat{f}_{n,h_n,i}(x)\right|} = O_{\mathbb{P}}(1),$$

when $nh_n^{2d} \to \infty$ and $nh_n^{4\alpha} \to 0$ as $n \to \infty$. The same holds for $\widetilde{f}_{n,h_n,i}$.

Theorem (Central Limit Theorem and asymptotic efficiency)

Under the previous assumptions, one has (i)

$$\sqrt{n}\left(\widehat{T}_{n,h_n} - \mathbb{E}[\mathbb{E}[Y|X]^2]\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}\left(0, \mathrm{Var}((2Y - m(X))m(X))\right)$$

as soon as $\alpha > d/2$ *and* $h_n = n^{-\gamma}$ *with* $1/(4\alpha) < \gamma < 1/(2d)$ *;*

(ii) $\widehat{T}_{n,h_n}$ *is asymptotically efficient to estimate* $\mathbb{E}[\mathbb{E}[Y|X]^2]$ *from an i.i.d. sample* $(X_i, Y_i)_{i=1,\cdots,n}$ *of the pair* (X, Y)*.*

The same holds for $\widetilde{T}_{n,h_n}$*.*

Ref. : S. Da Veiga, F. Gamboa, T. Klein, A. Lagnoux, C. Prieur. "Efficient estimation of Sobol' indices of any order from a single input/output sample.". Available on Hal and Arxiv (2024). https://hal.science/hal-04052837v2.

Using the delta method, we are now able to get the asymptotic behaviour of the estimation of S^X, letting

$$\widehat{S}_{n,h_n} = \frac{\widehat{T}_{n,h_n} - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2}{\frac{1}{n}\sum_{j=1}^n Y_j^2 - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2} \quad \text{and} \quad \widetilde{S}_{n,h_n} = \frac{\widetilde{T}_{n,h_n} - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2}{\frac{1}{n}\sum_{j=1}^n Y_j^2 - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2}.$$

Corollary (CLT & AE for the estimation of the Sobol' indices)

Under all the assumptions of the theorem, one has (i)

$$\sqrt{n}\left(\widehat{S}_{n,h_n} - S^X\right) \quad \text{and} \quad \sqrt{n}\left(\widetilde{S}_{n,h_n} - S^X\right) \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(0,\sigma^2),$$

where the limit variance σ^2 has an explicit expression.

(ii) $\widehat{S}_{n,h_n}$ and $\widetilde{S}_{n,h_n}$ are asymptotically efficient to estimate S^X from an i.i.d. sample $(X_i,Y_i)_{i=1,\cdots,n}$ of the pair (X,Y).

Let us denote S^i the first-order Sobol index associated to the i-th input and its estimator $\widehat{S}^i$ given by :

$$\widehat{S}^i_{n,h_n} = \frac{\widehat{T}^i_{n,h_n} - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2}{\frac{1}{n}\sum_{j=1}^n Y_j^2 - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2} \quad \text{and} \quad \widetilde{S}^i_{n,h_n} = \frac{\widetilde{T}^i_{n,h_n} - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2}{\frac{1}{n}\sum_{j=1}^n Y_j^2 - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)^2}.$$

Corollary (CLT & AE for the global estimation of the p first-order Sobol' indices)

Under all the assumptions of the theorem, one has

$$\sqrt{n}\Big((\widehat{S}^1_{n,h_n},\dots,\widehat{S}^p_{n,h_n})^T - (S^1,\dots,S^p)^T\Big) \xrightarrow[n\to\infty]{\mathscr{D}} \mathcal{N}(0,\Sigma),$$
$$\sqrt{n}\Big((\widetilde{S}^1_{n,h_n},\dots,\widetilde{S}^p_{n,h_n})^T - (S^1,\dots,S^p)^T\Big) \xrightarrow[n\to\infty]{\mathscr{D}} \mathcal{N}(0,\Sigma),$$

where the limit variance Σ has an explicit expression. Furthermore, such estimations are asymptotically efficient.

Outline of the talk

Sketch of the proof : CLT

Following the same lines as in the proof of Theorem 2.1 in Doksum (1995), we aim at proving that

$$\widehat{T}_{n,h} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(2Y_i - m(X_i))m(X_i)}_{=T_{n,oracle}} + o_{\mathbb{P}}(n^{-1/2}). \qquad (1)$$

The conclusion of the theorem will then follow directly applying the standard central limit theorem for the sum of i.i.d. random variables to the right-hand side of the previous display together with Slutsky's lemma.

Sketch of the proof : asymptotic efficiency

The influence efficient function of ψ at P, as stated in Doksum (1995), is given by (see Klein (2024) for the details) :

$$\widetilde{\psi}_P(x,y) = (2y - m(x))m(x) - \mathbb{E}[Ym(X)].$$

Moreover, we deduce from (1) that

$$\widehat{T}_{n,h} = \psi(P) + \frac{1}{n}\sum_{i=1}^{n} \widetilde{\psi}_P(X_i, Y_i) + o_{\mathbb{P}}(n^{-1/2})$$

and conclude using Condition (25.22) of Van der Vaart (2000).

Outline of the talk

For all test cases :

- first-order and total-order Sobol' indices for each input variable V_i (i.e. $X = V_i$ and $X = V_{\sim i}$ resp.).
- second mirror-type estimator with an Epanechnikov kernel of order 2 and 4 (kernel bandwidth optimized via LOO on m).
- concurrent estimators :
  - nearest-neighbour estimator (Devroye 2018) ("NN")
  - PF estimator studied (Janon 2012) ("PF1")
  - replicated PF estimator (Tissot 2015) ("PF2")
  - rank estimator (Gamboa'20) ("Rank") for 1st-order indices
  - lag estimator (Klein 2024) ("Lag") for first-order indices.
- we generate a n-sample $(X_1, Y_1), \cdots, (X_n, Y_n)$ (except for PF).
- each experiment is repeated 50 times with $n = 500$.
- the reference value is obtained from a PF estimation with very large sample size.

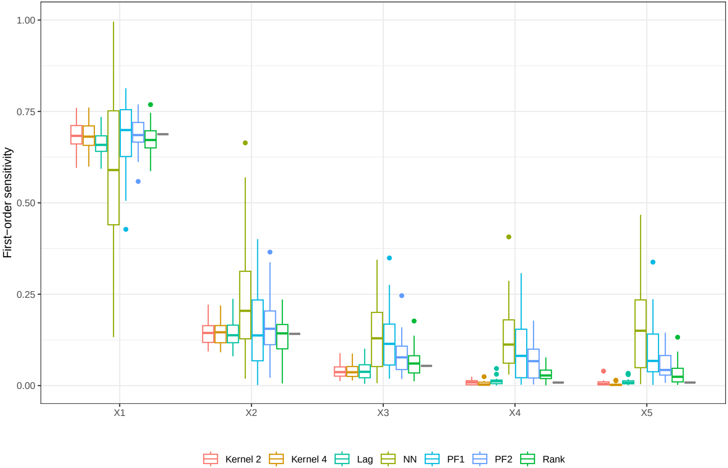
The Bratley function

First, we consider the Bratley function defined by :

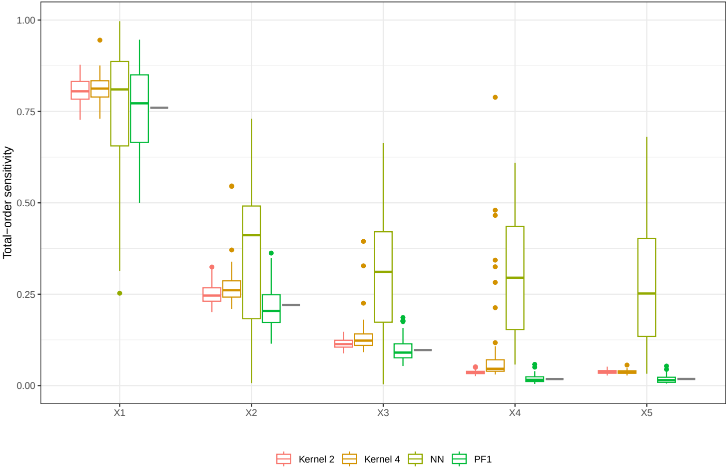
$$g_{\text{Bratley}}(V_1,\ldots,V_p) = \sum_{i=1}^{p}(-1)^i \prod_{j=1}^{i} V_j,$$

with $V_i \sim \mathscr{U}([0,1])$ i.i.d. and $p=5$.

The Bratley function - first-order indices - $n = 500$



The Bratley function - total-order indices - $n = 500$



The g-Sobol function

We investigate the g-Sobol function defined by

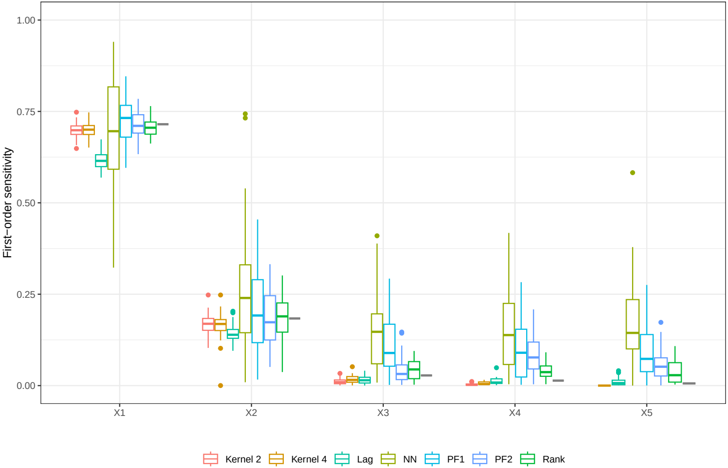
$$g_{\text{g-Sobol}}(V_1,\ldots,V_p)=\prod_{i=1}^{p}\frac{|4V_i-2|+a_i}{1+a_i},$$

with $V_i\sim\mathscr{U}([0,1])$ i.i.d., $p=5$ and $a=(0,1,4.5,9,99)$.

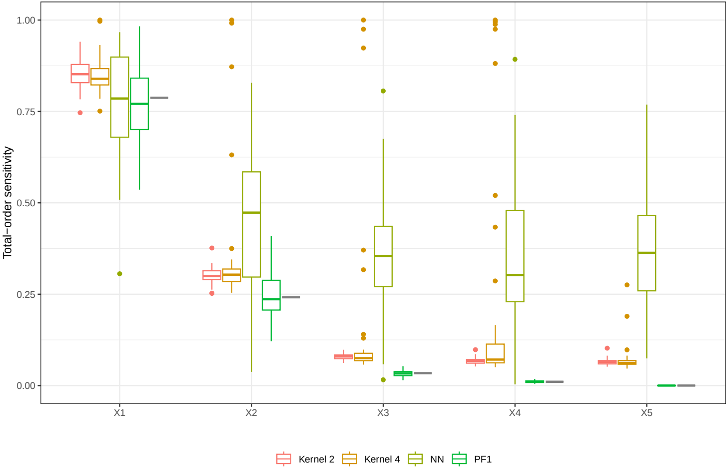
Notice that it is non-differentiable at any input value with a component equal to 0.5, but the impact on our estimator performance is negligible for first-order indices.

Except for the degraded performance of the lag estimator, the conclusions are the same as for the Bratley function, even for total indices.

The g-Sobol function - first-order indices - $n = 500$



The g-Sobol function - total-order indices - $n = 500$



Tuning of parameter ϵ

We illustrate numerically that the choice of the ϵ tuning parameter of the estimator proposed in Doksum (1995) is very sensitive, thus limiting its practical use as opposed to our mirror-type estimator.

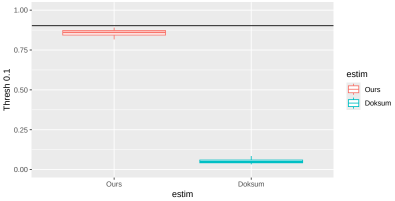
We consider Example 3.2 from Doksum (1995) :

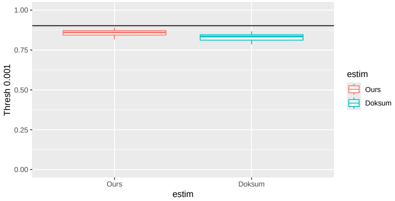
$$Y = \frac{1}{2} + 4X_1 + 4(X_2 - \frac{1}{2})^2 + 4X_3^{1/2} + \tau e,$$

with X_1, X_2, and X_3 i.i.d. $\sim \mathscr{U}([0,1])$ and $e \sim \mathscr{N}(0,1)$.

We test $\epsilon = 10^{-1}$ and 10^{-3}.

Tuning of parameter ϵ





When ϵ is equal to 10^{-3}, the performance of both estimators are similar. However when $\epsilon = 10^{-1}$, the bias of Doksum and Samarov (1995) can be very large. Since in practice such an estimation problem is unsupervised, the tuning of ϵ seems highly difficult and the non-robustness of the final estimator with respect to this parameter limits its practical use.

Thanks for your attention !
Questions ?

Reference

S. Da Veiga, F. Gamboa, T. Klein, A. Lagnoux, C. Prieur.
"Efficient estimation of Sobol' indices of any order from a single input/output sample.". Available on Hal and Arxiv (2024).
https://hal.science/hal-04052837v2.

Efficient influence function and asymptotic efficiency

Let $\mathscr{P}$ be the set of absolutely continuous probability distributions on $[0,1]^d \times \mathbb{R}$ and $P_0 \in \mathscr{P}$ be the probability distribution of (X, Y), such that we can write our target $T = \psi(P_0)$ where $\psi : \mathscr{P} \to \mathbb{R}$.

If ψ is differentiable at all $P \in \mathscr{P}$, the efficient influence function $\widetilde{\psi}_P : [0,1]^d \times \mathbb{R} \to \mathbb{R}$ is the gradient with smallest variance among all gradients of ψ at P with zero mean w.r.t. to P.

The link with efficient estimators is the following : a sequence of estimators T_n of $T = \psi(P_0)$ is asymptotically efficient if

$$T_n - T = T_n - \psi(P_0) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{\psi}_{P_0}(X_i, Y_i) + o_{P_0}\left(\frac{1}{\sqrt{n}}\right),$$

See Eq.(25.22) in van der Vaart (2000).

Efficient influence function and asymptotic efficiency

In our case, the efficient influence function at any $P \in \mathscr{P}$ writes

$$\tilde{\psi}_P(x,y) = (2y - m(x))m(x) - \psi(P).$$

where m is the regression function under P : $m(x) = \mathbb{E}_P[Y|X = x]$,
see Klein, Lagnoux, Rochet (2024).

Then, if m under P_0 is known, taking

$$T_{n,oracle} = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - m(X_i))m(X_i)$$

leads to an asymptotically efficient estimator of T.

Plug-in estimation

A first point of view consists in seeing

$$\widehat{T}_n = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - \widehat{m}_n(X_i))\widehat{m}_n(X_i),$$

as a plug-in version of

$$T_{n,oracle} = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - m(X_i))m(X_i)$$

where the difference $m - \widehat{m}_n$ needs to be controlled to still have

$$\widehat{T}_n = \psi(P_0) + \frac{1}{n}\sum_{i=1}^{n}\widetilde{\psi}_{P_0}(X_i, Y_i) + o_{P_0}\Big(\frac{1}{\sqrt{n}}\Big).$$

One-step estimation

A second point of view relies on one-step estimators, that consider a first-order bias correction of an initial estimator $\psi(\widehat{P})$ where $\widehat{P}$ is a smoothed estimate of P_0.

More precisely, a simple Taylor expansion of $\psi(P_0)$ around $\psi(\widehat{P})$ involves the efficient influence function $\widetilde{\psi}$ at $\widehat{P}$:

$$\psi(P_0) - \psi(\widehat{P}) = \mathbb{E}_{P_0}[\widetilde{\psi}_{\widehat{P}}] - \overbrace{\mathbb{E}_{\widehat{P}}[\widetilde{\psi}_{\widehat{P}}]}^{=0} + r_2(\widehat{P}, P) = \mathbb{E}_{P_0}[\widetilde{\psi}_{\widehat{P}}] + r_2(\widehat{P}, P)$$

since by definition, $\mathbb{E}_P[\widetilde{\psi}_P] = 0$ for all P. Thus, if $r_2(\widehat{P}, P) = o(1)$,

$$\psi(\widehat{P}) + \mathbb{E}_{P_0}[\widetilde{\psi}_{\widehat{P}}] \sim \psi(P_0).$$

One-step estimation

Thus it is possible to improve $\psi(\widehat{P})$ by considering an estimate of this first-order bias $\mathbb{E}_{P_0}[\widetilde{\psi}_{\widehat{P}}]$: for instance, $\mathbb{E}_{P_n}[\widetilde{\psi}_{\widehat{P}}]$ where P_n is the empirical distribution of the observations $(X_i, Y_i)_{i=1,\dots,n}$.

In our particular case, this induces an estimator given by

$$\widehat{T}_n = \psi(\widehat{P}) + \mathbb{E}_{P_n}[\widetilde{\psi}_{\widehat{P}}] = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - \widehat{m}(X_i))\widehat{m}(X_i)$$

where $\widehat{m}$ is the regression function under $\widehat{P}$, that is precisely a smoothing estimate of m. We can then hope that $\widehat{T}_n$ will be asymptotically efficient if the difference $\widehat{P} - P_0$ converges to 0 at an appropriate rate.

Construction of high-order kernels

The kernel k is typically chosen as a symmetric second-order kernel (Epanechnikov, Gaussian, ...) with the following properties :

$$\int k(u)du = 1, \quad \int uk(u)du = 0, \quad \int u^2 k(u) > 0.$$

The terminology second-order refers to the fact that the first non-zero moment of k is the second one (except for the zero-th order one which ensures the kernel is normalized).

Construction of high-order kernels

More generally, a high-order kernel of order r satisfies

$$\int k(u)du = 1, \quad \int u^j k(u)du = 0, \ \forall j = 1, \ldots, r-1, \quad \int u^r k(u) > 0.$$

Here, we will focus on high-order kernels with compact support, which are used together with mirror-type transformations to avoid boundary effects appearing when the domain is compact.

In particular, we will study symmetric kernels on $[-1,1]$ and non-symmetric ones on $[0,1]$.

Construction of high-order kernels

In order to build a kernel of order r with compact support $[-1,1]$, there are at least two approaches, which are described below.

Legendre orthonormal polynomials. The first construction relies on the (normalized) Legendre orthonormal polynomials on $[-1,1]$ denoted by $\{P_m(\cdot)\}_{m\in\mathbb{N}}$. Then we define the kernel k as

$$k(u) = \sum_{m=0}^{r+1} P_m(0)P_m(u)\mathbb{1}_{u\in[-1,1]}, \qquad (2)$$

see Comte (2017).

Construction of high-order kernels

High-order Epanechnikov kernel. Hansen (2005) proposes a high-order generalization of smooth and second-order kernels on $[-1,1]$ including the uniform, biweight, and Epanechnikov ones. Focusing on the latter, the kernel

$$k(u) = B_r(u) k_e(u) \tag{3}$$

where $k_e(u) = \frac{3}{4}(1-u^2)\mathbb{1}_{u\in[-1,1]}$ and

$$B_r(u) = \frac{\left(\frac{3}{2}\right)_{r/2-1}\left(\frac{5}{2}\right)_{r/2-1}}{(2)_{r/2-1}} \sum_{k=0}^{r/2-1} \frac{(-1)^k \left(\frac{r+3}{2}\right)_k u^{2k}}{k!(r/2-1-k)!\left(\frac{3}{2}\right)_k}$$

is of order r for odd r where $(x)_a$ is the Pochhammer's symbol.

Construction of high-order kernels

As for kernels with compact support $[0,1]$, the two following methods can be envisioned.

Shifted Legendre orthonormal polynomials. Similarly to the first construction above, we can also consider the shifted Legendre orthonormal polynomials on $[0,1]$, denoted by $\{Q_m(\cdot)\}_{m\in\mathbb{N}}$, leading to

$$k(u) = 2\sum_{m=0}^{r+1} Q_m(0)Q_m(u)\mathbb{1}_{u\in[0,1]}. \tag{4}$$

Construction

Dilatation. Another approach, due to Kerkyacharian (2001), relies on dilatations of an integrable function $g : \mathbb{R} \to \mathbb{R}$:

$$k(u) = \sum_{k=1}^{r} \binom{r}{k} (-1)^{k+1} \frac{1}{k} g\left(\frac{u}{k}\right). \tag{5}$$

If g has support $[a,b]$, then k has support $[a,rb]$ and is of order r.

To obtain a kernel with support $[0,1]$, one can for example take a shifted Epanechnikov kernel k_{shift} on $[0,1/r]$:

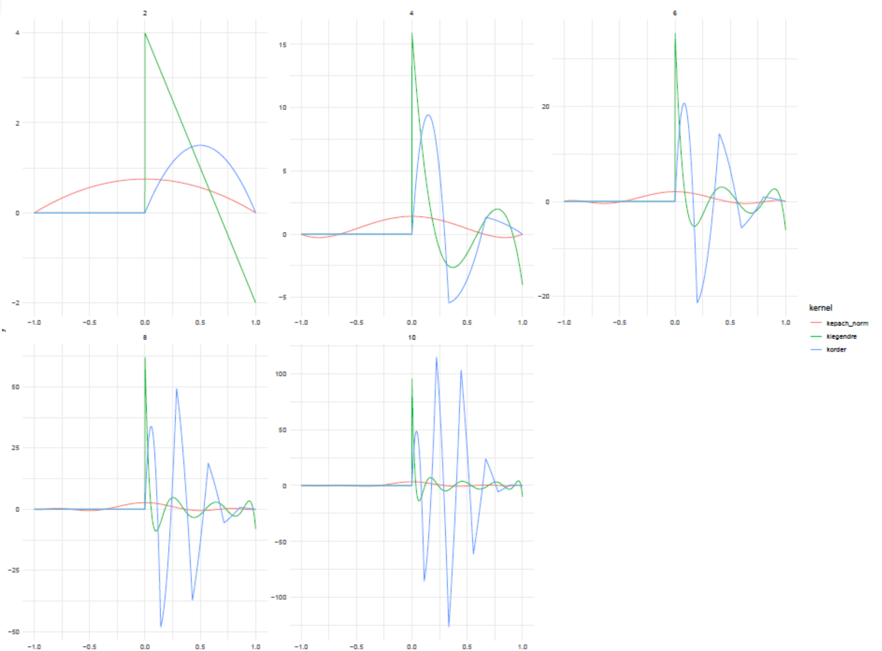
$$k_{\text{shift}}(u) = 6u(1-ru)r^2 \mathbb{1}_{u \in [0,1/r]}.$$

Numerical stability - Kernel values versus order

In what follows, we investigate numerically the high-order kernels introduced above.

Since kernels in (2) and (4) are identical up to a shift, we only focus on kernels as defined in (3) for $[-1,1]$ and (4) and (5) for $[0,1]$.

They are coded below, note that they all take as input a parameter h which corresponds to the kernel bandwidth.


2
4
6
8
10
kernel
kepach_norm
klegendre
korder

Numerical stability - Kernel values versus order

It appears clearly that non-symmetric kernels with support $[0,1]$ exhibit large variations which increase with the order, as opposed to the symmetric kernel on $[-1,1]$. This implies that numerical instabilities when computing estimators are to be expected, as illustrated below on a simple regression case.

Regression with mirror transformations

Now we consider a standard regression setting : we have access to a n-sample (X_i, Y_i) for $i = 1, \dots, n$ with

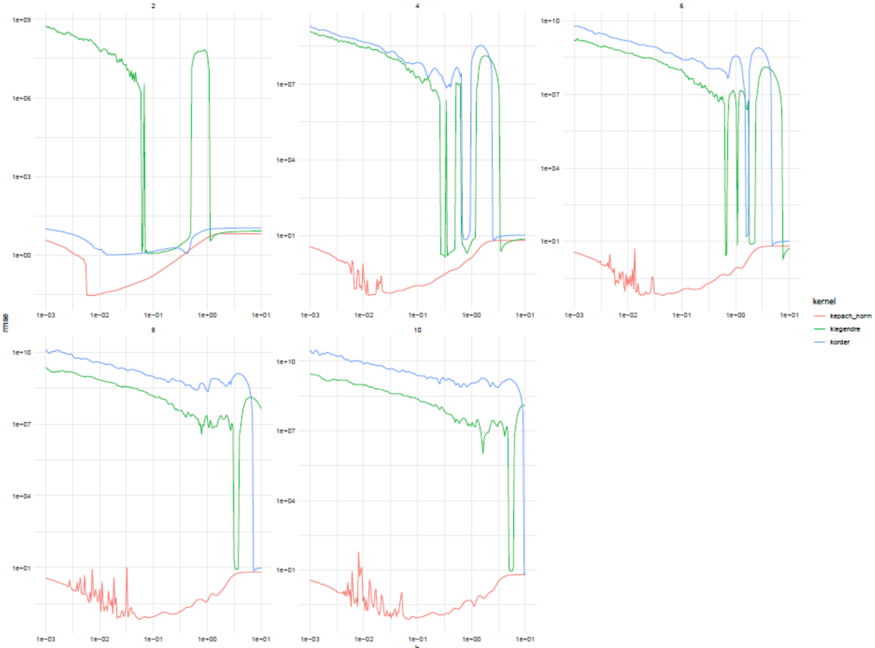
$$Y_i = m(X_i) + \epsilon_i$$

where the X_i's are i.i.d. r.v. with domain $[0,1]$ and ϵ_i is a centered noise.

We consider both regression estimators denoted by $\widehat{m}^1$ and $\widehat{m}^2$ and the Bratley function.

The only parameter which needs to be tuned is the bandwidth h.

Here, we will consider a grid of evenly-spaced values on a logarithmic scale, and compute the leave-one-out mean square error for each of them.


2
4
6
8
10
rmse
h
kernel
kepach_norm
klegendre
korder

Regression with mirror transformations

We clearly see a very high numerical instability for the first estimator with kernels supported on $[0,1]$, even on a simple regression example in dimension 1.