

# A survey on sensitivity analysis - Part I

Agnès Lagnoux

Institut de Mathématiques de Toulouse  
TOULOUSE - FRANCE

**Workshop - Inverse calibration of DEB numerical  
models for Indian Ocean tunas**

**Victoria, 26-29 May 2015**

## Travaux en commun avec

- F. Gamboa, T. Klein et B. Laurent (IMT);
- J.C. Fort (Paris 5);
- A. Janon (Paris 11);
- C. Prieur (LJK).

Travail partiellement financé par l'ANR programme COSINUS :  
projet Costa-Brava.

# Plan de l'exposé

## Introduction

- Le contexte

- Les indices de Sobol

- L'estimation Pick-Freeze

## Modèle à sortie vectorielle

- Motivation

- Le cas vectoriel

## Modèle à sortie fonctionnelle

## Le contexte

Nous considérons une sortie  $y$  définie par

$$y = f(x) := f(x^1, \dots, x^p)$$

où

- les entrées  $x^i$  pour  $i = 1, \dots, p$  sont des objets ;
- $f$  est la fonction inconnue déterministe appelée **boîte noire**.

## Le contexte

Ce que nous avons/savons :

- $f$  n'est pas analytiquement connue.
- Pour chaque valeur du  $p$ -uplet  $(x^1, \dots, x^p)$ , le code renvoie la valeur de la réponse  $y = f(x^1, \dots, x^p)$ .
- Cependant, chaque calcul de  $y$  est coûteux en temps.

## Le contexte

Ce que nous avons/savons :

- $f$  n'est pas analytiquement connue.
- Pour chaque valeur du  $p$ -uplet  $(x^1, \dots, x^p)$ , le code renvoie la valeur de la réponse  $y = f(x^1, \dots, x^p)$ .
- Cependant, chaque calcul de  $y$  est coûteux en temps.

Ce que nous voulons et ce dont nous avons besoin :

- Evaluer  $y$  pour chaque valeur du  $p$ -uplet  $x^1, \dots, x^p$  n'est pas raisonnable/possible.
- Fixer les variables non influentes à leur valeur nominale.
- Déterminer quelles sont les variables influentes.

## Le cadre probabiliste

De façon à quantifier l'importance d'une variable, il est courant de considérer les entrées comme des variables aléatoires indépendantes. Ainsi la sortie devient elle aussi aléatoire.

## Le cadre probabiliste

De façon à quantifier l'importance d'une variable, il est courant de considérer les entrées comme des variables aléatoires indépendantes. Ainsi la sortie devient elle aussi aléatoire.

**La question est maintenant** : comment mesurer l'importance d'une variable (ou d'un groupe de variables) ?

L'indicateur le plus simple de la variabilité d'une va est la variance ; nous allons donc construire un indicateur basé sur la variance.



## De la décomposition de Hoeffding aux indices de Sobol

Le point clé est que  $L^2$  est un espace de Hilbert et nous allons utiliser les projections orthogonales pour obtenir la **décomposition de Hoeffding** de la variance.

### Théorème

*f peut être écrite de manière unique comme la somme de termes décorrelés :*

$$Y = f(X^1, \dots, X^p) = \sum_{A \subset \{1, \dots, d\}} f_A(X^A),$$

avec  $X^A := (X^i, i \in A)$ .

Nous prenons ensuite la variance de chaque côté.

## Etude en dimension 2 pour fixer les idées

$$Y = f(X^1, X^2)$$

$$= \mathbb{E}(f(X))$$

$$+ \mathbb{E}(f(X)|X^1) - \mathbb{E}(f(X)) + \mathbb{E}(f(X)|X^2) - \mathbb{E}(f(X))$$

$$+ f(X^1, X^2) - [\mathbb{E}(f(X)) + \mathbb{E}(f(X)|X^1) - \mathbb{E}(f(X)) + \mathbb{E}(f(X)|X^2) - \mathbb{E}(f(X))]$$

effet moyen,

effets du premier ordre,

effets du second ordre.

## Etude en dimension 2 pour fixer les idées

$$\begin{aligned}
 Y &= f(X^1, X^2) \\
 &= \mathbb{E}(f(X)) \\
 &+ \mathbb{E}(f(X)|X^1) - \mathbb{E}(f(X)) + \mathbb{E}(f(X)|X^2) - \mathbb{E}(f(X)) \\
 &+ f(X^1, X^2) - [\mathbb{E}(f(X)) + \mathbb{E}(f(X)|X^1) - \mathbb{E}(f(X)) + \mathbb{E}(f(X)|X^2) - \mathbb{E}(f(X))]
 \end{aligned}$$

effet moyen,  
 effets du premier ordre,  
 effets du second ordre.

En prenant la variance de chaque côté,

$$\begin{aligned}
 \text{Var}(Y) &= \text{Var}(\mathbb{E}(f(X)|X^1)) + \text{Var}(\mathbb{E}(f(X)|X^2)) \\
 &+ \text{Var}(f(X^1, X^2) - [\mathbb{E}(f(X)|X^1) + \mathbb{E}(f(X)|X^2)]) .
 \end{aligned}$$

En normalisant par la variance totale de la sortie  $Y$ , nous obtenons

$$\begin{aligned}
 1 &= \frac{\text{Var}(\mathbb{E}(f(X)|X^1))}{\text{Var} Y} + \frac{\text{Var}(\mathbb{E}(f(X)|X^2))}{\text{Var} Y} \\
 &\quad + \frac{\text{Var}(f(X^1, X^2) - [\mathbb{E}(f(X)|X^1) + \mathbb{E}(f(X)|X^2)])}{\text{Var} Y} \\
 &:= S^1 + S^2 + S^{1,2} \quad \Rightarrow \text{les indices de Sobol}
 \end{aligned}$$

En normalisant par la variance totale de la sortie  $Y$ , nous obtenons

$$\begin{aligned}
 1 &= \frac{\text{Var}(\mathbb{E}(f(X)|X^1))}{\text{Var} Y} + \frac{\text{Var}(\mathbb{E}(f(X)|X^2))}{\text{Var} Y} \\
 &\quad + \frac{\text{Var}(f(X^1, X^2) - [\mathbb{E}(f(X)|X^1) + \mathbb{E}(f(X)|X^2)])}{\text{Var} Y} \\
 &:= S^1 + S^2 + S^{1,2} \quad \Rightarrow \text{les indices de Sobol}
 \end{aligned}$$

$S^1 = \frac{\text{Var}(\mathbb{E}(f(X)|X^1))}{\text{Var} Y}$  quantifie l'effet du premier ordre de  $X^1$ ,

tandis que  $S^1 + S^{1,2}$  quantifie l'effet total de  $X^1$ .

## Un exemple très simple

Considérons le modèle suivant

$$Y = f(X) = X^1 + X^1 X^2$$

avec  $X = (X^1, X^2) \sim \mathcal{N}_2(0, I_2)$ . Ainsi

$$(S^1, S^2, S^{1,2}) = (1/2, 0, 1/2).$$

## Estimation Pick-Freeze des indices de Sobol

Pour plus de simplicité, nous nous plaçons dans le cas de deux variables d'entrée. La sortie  $Y$  est alors donnée par

$$Y = f(X^1, X^2).$$

Supposons que nous souhaitons mesurer l'importance de la première entrée  $X^1$ . Dans cette optique, nous

- définissons de nouvelles variables aléatoires :

$$Y^1 = f(X^1, X^{2'}),$$

où  $X^{2'}$  est une copie de  $X^2$ , indépendante de  $X^1$  et  $X^2$ .

## Estimation Pick-Freeze des indices de Sobol

Pour plus de simplicité, nous nous plaçons dans le cas de deux variables d'entrée. La sortie  $Y$  est alors donnée par

$$Y = f(X^1, X^2).$$

Supposons que nous souhaitons mesurer l'importance de la première entrée  $X^1$ . Dans cette optique, nous

- définissons de nouvelles variables aléatoires :

$$Y^1 = f(X^1, X^{2'}),$$

où  $X^{2'}$  est une copie de  $X^2$ , indépendante de  $X^1$  et  $X^2$ .

- considérons le nouveau plan d'expérience suivant :
  - (1) un  $N$ -échantillon de  $Y = f(X^1, X^2)$ ,  $Y_i$ ,  $1 \leq i \leq N$ .
  - (2) un  $N$ -échantillon de  $Y^1 = f(X^1, X^{2'})$ ,  $Y_i^1$ ,  $1 \leq i \leq N$ .



L'estimation de l'indice de Sobol  $S^1$  est basée sur le résultat suivant

### Corollaire

$$\text{Var}(\mathbb{E}(Y|X^1)) = \text{Cov}(Y, Y^1).$$

L'estimation de l'indice de Sobol  $S^1$  est basée sur le résultat suivant

### Corollaire

$$\text{Var}(\mathbb{E}(Y|X^1)) = \text{Cov}(Y, Y^1).$$

Alors

$$S^1 = \frac{\text{Var}(\mathbb{E}(Y|X^1))}{\text{Var}(Y)} = \frac{\text{Cov}(Y, Y^1)}{\text{Var}(Y)}$$

est naturellement estimé par

$$S_N^1 = \frac{\frac{1}{N} \sum Y_i Y_i^1 - \left( \frac{1}{N} \sum \left[ \frac{Y_i + Y_i^1}{2} \right] \right)^2}{\frac{1}{N} \sum \left[ \frac{(Y_i)^2 + (Y_i^1)^2}{2} \right] - \left( \frac{1}{N} \sum \left[ \frac{Y_i + Y_i^1}{2} \right] \right)^2}.$$

## Quelques questions statistiques

- L'estimateur Pick-Freeze est-il un "bon" estimateur de l'indice de Sobol ?
- Est-il consistant ? **Réponse** : LFGN.
- Si oui, à quelle vitesse de convergence ?  
**Réponse** : TCL (cv en  $\sqrt{n}$ ). La variance limite est donnée par

$$\frac{\text{Var}(V^1)}{\text{Var}(Y)^2}, \quad \text{avec} \quad V^1 := Y Y^1 - S^1 \left[ \frac{Y^2 + (Y^1)^2}{2} \right].$$

- Est-il asymptotiquement efficace (en un sens à préciser) ?  
**Réponse** : oui.
- Peut-on mesurer sa performance à  $n$  fixé ?  
**Réponse** : Berry-Esseen et/ou inégalités de concentration.

# Plan de l'exposé

## Introduction

Le contexte

Les indices de Sobol

L'estimation Pick-Freeze

## Modèle à sortie vectorielle

Motivation

Le cas vectoriel

## Modèle à sortie fonctionnelle

## Premier exemple introductif

Considérons le modèle non linéaire suivant

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

où  $X_1$  et  $X_2$  sont des Gaussiennes standard indépendantes.

## Premier exemple introductif

Calculons les indices de Sobol scalaires  $S^j(f_i^{a,b})$  de  $f_i^{a,b}$  par rapport à  $X_j$  ( $i, j = 1, 2$ ) :

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

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

Ainsi les ratios

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

ne dépendent pas de  $b$ . De plus, pour  $|a| > 1$ ,  $X_1$  semble être la plus influente sur la sortie.

## Premier exemple introductif

Maintenant nous faisons l'analyse de sensibilité sur  $\|Y\|^2$ .

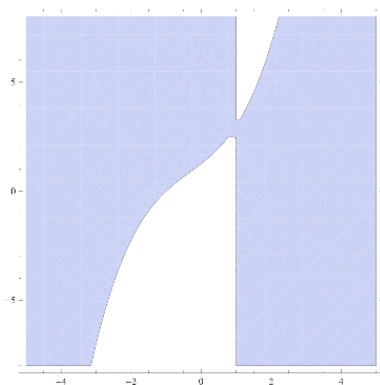
Facilement, nous avons

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

Pour  $\|Y\|^2$ , la région où  $X_1$  est la plus influente dépend de  $b$ .

Cette région n'est pas très intuitive...

## Premier exemple introductif



**Figure :** Graphe de  $(a - 1)(a^3 + a^2 + 5a + 5 - 4b) \geq 0$ . Le bleu correspond aux régions où  $S^1(\|Y\|^2) \geq S^2(\|Y\|^2)$ .



## Deuxième exemple introductif

Nous étudions le modèle 2-dimensionnel suivant

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

avec  $(X_1, X_2) \sim \text{Unif}([0; 10]) \otimes \text{Unif}([0; \pi/2])$ .

## Deuxième exemple introductif

Nous avons

$$S^1(f_1^{a,b}) = S^1(f_2^{a,b}) = \frac{10}{5\pi^2 - 30} \approx 0.52$$

$$S^2(f_1^{a,b}) = S^2(f_2^{a,b}) = \frac{3(\pi^2 - 8)}{4(\pi^2 - 6)} \approx 0.36.$$

Ainsi  $X_1$  semble avoir plus d'influence sur la sortie que  $X_2$ .

Cependant, nous obtenons  $\|Y\|^2 = X_1^2$  qui ne dépend pas de  $X_2$ .

## Une autre motivation

Lorsque la dimension augmente, l'estimation du vecteur complet des indices de Sobol scalaire devient de plus en plus coûteuse.

De plus, l'interprétation d'un tel vecteur n'est pas facile.

Ceci renforce la nécessité d'introduire de nouveaux indices de Sobol qui condenseront toute l'information contenue dans une large collection de scalaires.

## Généralisation de l'indice de Sobol au cas vectoriel

Tout d'abord, nous introduisons,  $\forall \mathbf{u} \subset \{1, \dots, p\}$ ,

$$X_{\mathbf{u}} = (X_i, i \in \mathbf{u}) \quad \text{et} \quad X_{\sim \mathbf{u}} = (X_i, i \in \{1, \dots, p\} \setminus \mathbf{u});$$

puis

$$Y^{\mathbf{u}} = f(X_{\mathbf{u}}, X'_{\sim \mathbf{u}}),$$

où  $X'_{\sim \mathbf{u}}$  est une copie indépendante de  $X_{\sim \mathbf{u}}$ .

## Généralisation de l'indice de Sobol au cas vectoriel

Tout d'abord, nous introduisons,  $\forall \mathbf{u} \subset \{1, \dots, p\}$ ,

$$X_{\mathbf{u}} = (X_i, i \in \mathbf{u}) \quad \text{et} \quad X_{\sim \mathbf{u}} = (X_i, i \in \{1, \dots, p\} \setminus \mathbf{u});$$

puis

$$Y^{\mathbf{u}} = f(X_{\mathbf{u}}, X'_{\sim \mathbf{u}}),$$

où  $X'_{\sim \mathbf{u}}$  est une copie indépendante de  $X_{\sim \mathbf{u}}$ .

Partant de

$$\text{Tr}(\text{Var} Y) = \text{Tr}(\text{Cov}(Y, Y^{\mathbf{u}})) + \text{Tr}(\text{Cov}(Y, Y^{\sim \mathbf{u}})) + \text{Reste.}$$

nous définissons ensuite

$$S^{\mathbf{u}}(f) = \frac{\text{Tr}(\text{Cov}(Y, Y^{\mathbf{u}}))}{\text{Tr}(\text{Var}(Y))}.$$

## Illustration sur un troisième exemple

Considérons le modèle suivant

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

avec  $X_1$  et  $X_2$  des va Gaussiennes standard iid. Nous avons

$$S^1(f^a) = \frac{a^2}{a^2 + 1} \quad \text{et} \quad S^2(f^a) = \frac{1}{a^2 + 1} = 1 - S^1(f),$$

ce qui paraît raisonnable.

## Retour sur le premier exemple

Considérons de nouveau le modèle

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

Nous avons

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

et

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

Ce résultat a une interprétation naturelle :

si  $X_1$  est multiplié par  $a$ , il a plus d'influence

ssi cette homothétie agrandit son support ie ssi  $|a| > 1$ .

## Estimation de $S^u(f)$

Nous estimons naturellement

$$S^u(f) = \frac{\text{Tr}(\text{Cov}(Y, Y^u))}{\text{Tr}(\text{Var}(Y))}.$$

par

$$S_{u,N} = \frac{\sum_{l=1}^k \left( \frac{1}{N} \sum_{i=1}^N Y_{i,l} Y_{i,l}^u - \left( \frac{1}{N} \sum_{i=1}^N \frac{Y_{i,l} + Y_{i,l}^u}{2} \right)^2 \right)}{\sum_{l=1}^k \left( \frac{1}{N} \sum_{i=1}^N \frac{Y_{i,l}^2 + (Y_{i,l}^u)^2}{2} - \left( \frac{1}{N} \sum_{i=1}^N \frac{Y_{i,l} + Y_{i,l}^u}{2} \right)^2 \right)}.$$



## Estimation de $S^u(f)$

Notons que cet estimateur peut être réécrit sous la forme

$$S_{\mathbf{u},N} = \frac{\text{Tr}(C_{\mathbf{u},N})}{\text{Tr}(\Sigma_N)}$$

où  $C_{\mathbf{u},N}$  et  $\Sigma_N$  sont les estimateurs empiriques de  $C_{\mathbf{u}} = \text{Cov}(Y, Y^{\mathbf{u}})$  et  $\Sigma = \text{Var}(Y)$  définis par

$$C_{\mathbf{u},N} = \frac{1}{N} \sum_{i=1}^N Y_i^{\mathbf{u}} Y_i^t - \left( \frac{1}{N} \sum_{i=1}^N \frac{Y_i + Y_i^{\mathbf{u}}}{2} \right) \left( \frac{1}{N} \sum_{i=1}^N \frac{Y_i + Y_i^{\mathbf{u}}}{2} \right)^t$$

et

$$\Sigma_N = \frac{1}{N} \sum_{i=1}^N \frac{Y_i Y_i^t + Y_i^{\mathbf{u}} (Y_i^{\mathbf{u}})^t}{2} - \left( \frac{1}{N} \sum_{i=1}^N \frac{Y_i + Y_i^{\mathbf{u}}}{2} \right) \left( \frac{1}{N} \sum_{i=1}^N \frac{Y_i + Y_i^{\mathbf{u}}}{2} \right)^t.$$

## Propriétés asymptotiques de $S_{\mathbf{u},N}$

### Consistance

Par une application directe de la loi forte des grands nombres, nous obtenons que  $S_{\mathbf{u},N}$  est consistant.

### Normalité asymptotique

Nous pouvons établir que  $\sqrt{N}(S_{\mathbf{u},N} - S^{\mathbf{u}}(f))$  converge en loi vers une gaussienne centré.

### Inégalité de concentration

Nous pouvons établir des bornes précises pour

$$\mathbb{P}(|S_{\mathbf{u},N} - S^{\mathbf{u}}(f)| \geq t).$$

## Application numérique : exemple jouet (I)

Nous considérons l'exemple de l'introduction avec  $k = p = 2$ ,  $a = 2$  et  $b = 3$  :

$$Y = f(X_1, X_2) = \begin{pmatrix} X_1 + X_2 + X_1X_2 \\ 2X_1 + 3X_1X_2 + X_2 \end{pmatrix}.$$

Nous considérons deux cas :  $X_1$  et  $X_2$  va indépendantes

- ① normales standard ;
- ② uniformes sur  $[0, 1]$ .

Dans les deux cas, les indices de Sobol  $S^1(f)$  et  $S^2(f)$  peuvent être calculés.

## Application numérique : exemple jouet (II)

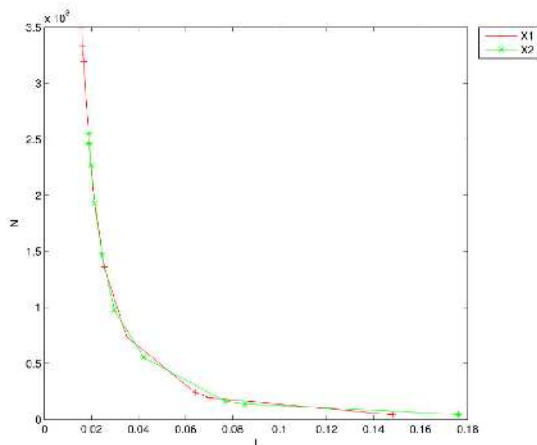
Nous générons 100 simulations de la méthode Pick-Freeze pour  $N = 100, 200$  et  $10000$ .

Dans chaque cas, nous notons le nb d'intervalles estimés contenant la vraie valeur.

		$N=100$	$N=2000$	$N=10000$	Vraie valeur
Cas Gaussien	$S^1(f)$	0.97	0.94	0.97	0.2941
	$S^2(f)$	0.94	0.93	0.93	0.1176
Cas uniforme	$S^1(f)$	1	1	1	0.6084
	$S^2(f)$	0.97	0.98	0.97	0.3566

## Application numérique : exemple jouet (III)

Pour différentes valeurs de  $t$ , nous déterminons le plus petit  $N$  (estimé) tel que la borne de  $\mathbb{P}(|S_{u,N} - S^u(f)| \geq t)$  soit inférieure à 5%.



## Application numérique : modèle ressort (I)

Nous considérons le déplacement d'une masse  $x(t)$  connecté à un ressort pour des temps  $t \in [0; 40]$  donné par :

$$mx''(t) + cx'(t) + kx(t) = 0,$$

avec les CI  $x(0) = l, x'(0) = 0$ .

La sortie vectorielle est définie par

$$Y = f(X) = (x(t_1), x(t_2), \dots, x(t_{800}))$$

pour  $t_i = 0.05i$  et  $k = 800$ .

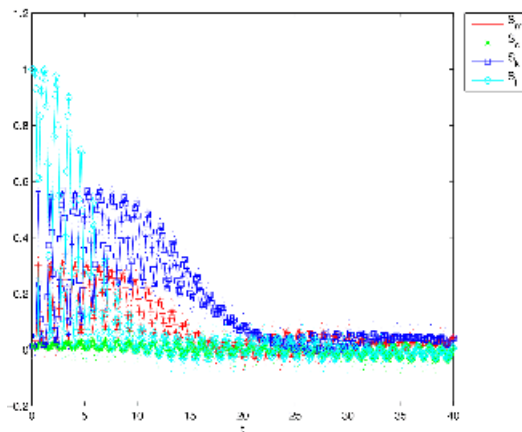
## Application numérique : modèle ressort (II)

Les paramètres d'entrée sont  $X = (m, c, k, l)$ ,  $p = 4$  :

Variable	Interprétation (unité SI)	Distribution
$m$	poids (kg)	Unif([10 ;12])
$c$	constante d'amortissement ( $\text{N} \cdot \text{m}^{-1} \cdot \text{s}$ )	Unif([0.4 ; 0.8])
$k$	constante du ressort ( $\text{N} \cdot \text{m}^{-1}$ )	Unif([70 ;90])
$l$	longueur initiale (m)	Unif([-1 ; -0.25])

## Application numérique : modèle ressort (III)

En considérant les composantes de  $Y$  indépendemment, il est possible d'estimer les indices de Sobol (unidimensionnels) de  $Y(t_i)$  pour  $i = 1, \dots, 800$  et chaque variable d'entrée.





## Application numérique : modèle ressort (IV)

Nous calculons maintenant les indices de Sobol de  $Y$ , pour  $\mathbf{u} = \{m\}, \{c\}, \{k\}$  ou  $\{l\}$  ainsi que les IC à 95% pour  $N = 2000$ .

Variable $\mathbf{u}$	Estimée ponctuelle de $S^{\mathbf{u}}(Y)$	IC 95% pour $S^{\mathbf{u}}(Y)$
$m$	0.0826	[0.0600 ; 0.1052]
$c$	0.0020	[-0.0181 ; 0.0222]
$k$	0.2068	[0.1835 ; 0.2301]
$l$	0.0561	[0.0328 ; 0.0794]

# Plan de l'exposé

## Introduction

Le contexte

Les indices de Sobol

L'estimation Pick-Freeze

## Modèle à sortie vectorielle

Motivation

Le cas vectoriel

## Modèle à sortie fonctionnelle

## Généralisation de l'indice de Sobol au cas fonctionnel

Tout d'abord, nous introduisons,  $\forall \mathbf{u} \subset \{1, \dots, p\}$ ,

$$X_{\mathbf{u}} = (X_i, i \in \mathbf{u}) \quad \text{et} \quad X_{\sim \mathbf{u}} = (X_i, i \in \{1, \dots, p\} \setminus \mathbf{u});$$

puis

$$Y^{\mathbf{u}} = f(X_{\mathbf{u}}, X'_{\sim \mathbf{u}}),$$

où  $X'_{\sim \mathbf{u}}$  est une copie indépendante de  $X_{\sim \mathbf{u}}$ .

## Généralisation de l'indice de Sobol au cas fonctionnel

Tout d'abord, nous introduisons,  $\forall \mathbf{u} \subset \{1, \dots, p\}$ ,

$$X_{\mathbf{u}} = (X_i, i \in \mathbf{u}) \quad \text{et} \quad X_{\sim \mathbf{u}} = (X_i, i \in \{1, \dots, p\} \setminus \mathbf{u});$$

puis

$$Y^{\mathbf{u}} = f(X_{\mathbf{u}}, X'_{\sim \mathbf{u}}),$$

où  $X'_{\sim \mathbf{u}}$  est une copie indépendante de  $X_{\sim \mathbf{u}}$ .

Nous définissons alors

$$S^{\mathbf{u}, \infty} = \frac{\text{Tr}(\Gamma_{\mathbf{u}})}{\text{Tr}(\Gamma)},$$

où

$$\begin{aligned} \Gamma(h) &= \mathbb{E}[\langle Y - \mathbb{E}(Y), h \rangle (Y - \mathbb{E}(Y))] \\ \Gamma_{\mathbf{u}}(h) &= \mathbb{E}[\langle Y - \mathbb{E}(Y), h \rangle (Y^{\mathbf{u}} - \mathbb{E}(Y^{\mathbf{u}}))]. \end{aligned}$$

## Estimation de l'indice de Sobol

Nous avons

$$\text{Tr}(\Gamma) = \mathbb{E} (\|Y\|^2) - \|\mathbb{E}(Y)\|^2$$

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

## Estimation de l'indice de Sobol

Nous avons

$$\text{Tr}(\Gamma) = \mathbb{E} (\|Y\|^2) - \|\mathbb{E}(Y)\|^2$$

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

Soit  $(\varphi_l, 1 \leq l)$  une base orthonormée de  $\mathbb{H}$ , alors

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

Maintenant en vue de l'estimation, nous tronquons la somme et définissons

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

## Estimation de l'indice de Sobol

Plan d'expérience Pick and Freeze :

- (1) un éch. de taille  $N$  de  $Y : Y_i, 1 \leq i \leq N$ ;
- (2) un éch. de taille  $N$  de  $Y^u : Y_i^u, 1 \leq i \leq N$ .

## Estimation de l'indice de Sobol

Plan d'expérience Pick and Freeze :

- (1) un éch. de taille  $N$  de  $Y : Y_i, 1 \leq i \leq N$ ;
- (2) un éch. de taille  $N$  de  $Y^{\mathbf{u}} : Y_i^{\mathbf{u}}, 1 \leq i \leq N$ .

Un estimateur naturel de  $S^{\mathbf{u},\infty}$  est

$$S_{\mathbf{u},m,N} = \frac{\frac{1}{4N} \sum_{i=1}^N (\|Y_i + Y_i^{\mathbf{u}}\|_m^2 - \|Y_i - Y_i^{\mathbf{u}}\|_m^2 - \|\bar{Y} + \bar{Y}^{\mathbf{u}}\|_m^2)}{\frac{1}{N} \sum_{i=1}^N \left( \frac{\|Y_i\|_m^2 + \|Y_i^{\mathbf{u}}\|_m^2}{2} - \left\| \frac{\bar{Y} + \bar{Y}^{\mathbf{u}}}{2} \right\|_m^2 \right)}.$$



# Propriétés asymptotiques de $S_{\mathbf{u},\mathbf{m},N}$

## Consistance

Par une application de la loi forte des grands nombres, nous obtenons que  $S_{\mathbf{u},\mathbf{m},N}$  est consistant.

## Normalité asymptotique

Nous pouvons établir que  $\sqrt{N}(S_{\mathbf{u},\mathbf{m},N} - S^{\mathbf{u},\infty}(f))$  converge en loi vers une gaussienne centré dont on peut déterminer explicitement la variance.

## Bibliographie

J .C. Fort, T. Klein, A. Lagnoux, B. Laurent. "Estimation of the Sobol indices in a linear functional multidimensional model", *JSPI*, 2013.

N. Hilgert, A. Mas, N. Verzelen. "Minimax adaptive tests for the functional linear model", *Annals of Statistics*, in press, Arxiv : 1206.1094.

A. Janon, T. Klein, A. Lagnoux, M. Nodet, C. Prieur. " Asymptotic normality et efficiency of a Sobol index estimator", *ESAIM P&S*, 2013.

F. Gamboa, A. Janon, T. Klein et A. Lagnoux. "Sensitivity indices for multivariate outputs", *CRAS*, pp307-310, 2013.

M. Lamboni, H. Monod and D. Makowski. "Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models", *Reliability Engineering & System Safety*, pp450-459, 2011.

I.M. Sobol. "Sensitivity estimates for nonlinear mathematical models", *Math. Mod. Comput. Exp.*, 1993.