Estimation of the Sobol indices in a linear functional multidimensional model

J.C. Fort, T. Klein, A. Lagnoux and B. Laurent*

* Institut de Mathématiques de Toulouse
  TOULOUSE - FRANCE

7th International Conference on Sensitivity Analysis of Model Output
Nice - July 1-4, 2013

This work has been partially supported by the French National Research Agency through COSINUS program: Costa-Brava project
Let $\mathbb{H}$ a separable Hilbert space endowed with the scalar product $\langle , \rangle$. Usually $\mathbb{H} = L^2$.
We consider the following linear model

$$Y = \mu + \sum_{k=1}^{p} \langle \beta^k, X^k \rangle + \varepsilon$$

- $X^k$ are centered stochastic processes $\in \mathbb{H}$ s.t $\mathbb{E}(\|X^k\|^4) < \infty$;
- $\beta^k$ are elements of $\mathbb{H}$;
- $\varepsilon$ is a centered noise independent of the $X^k$’s s.t $\mathbb{E}(\|\varepsilon\|^4) < \infty$.

Remark: such a model can arise for example when one wants to define a metamodel to replace an expensive black-box.
Our goal is to quantify the influence of $X^k$ on $Y$, for $k = 1 \ldots p$.

We use as the suggested by Hoeffding decomposition the Sobol index

$$S^{(k)} := \frac{\text{Var}(E(Y|X^k))}{\text{Var}(Y)}, \quad k = 1 \ldots p.$$
Our goal is to quantify the influence of $X^k$ on $Y$, for $k = 1 \ldots p$.

We use as the suggested by Hoeffding decomposition the Sobol index

$$S^{(k)} := \frac{\text{Var} \left( \mathbb{E}(Y|X^k) \right)}{\text{Var}(Y)}, \quad k = 1 \ldots p.$$  

The model: Let us restrict to $p = 1$ and consider

$$Y = \mu + < \beta, X > + \varepsilon$$ \hspace{1cm} (2)$$

In this setting, the quantity to estimate

$$S = \frac{\text{Var} \left( \mathbb{E}(Y|X) \right)}{\text{Var}(Y)}$$

is of less interest, but the computations then easily extend to the generic model.
Outline of the talk

Estimators considered
- A first estimation of $\text{Var}(\mathbb{E}(Y|X))$
- A second estimation of $\text{Var}(\mathbb{E}(Y|X))$

Asymptotic properties of the estimators

Numerical Applications

Conclusion
Outline of the talk

Estimators considered
- A first estimation of $\text{Var}(\mathbb{E}(Y|X))$
- A second estimation of $\text{Var}(\mathbb{E}(Y|X))$

Asymptotic properties of the estimators

Numerical Applications

Conclusion
A first estimation of $\text{Var}(\mathbb{E}(Y|X))$

Precisions on the framework

The observations consist in $n$ i.i.d. copies $(X_i, Y_i)$ of $(X, Y)$.

Since $\text{Var}(Y)$ is naturally estimated by the empirical variance based on $(Y_1, \ldots, Y_n)$

\[
\frac{1}{n} \sum_{i=1}^{n} \left( Y_i - \frac{1}{n} \sum_{i=1}^{n} Y_i \right)^2,
\]

the main purpose is to estimate the quantity $\text{Var}(\mathbb{E}(Y|X))$. 
A first estimation of $\text{Var}(\mathbb{E}(Y|X))$

Our approach is based on the so-called Karhunen-Loève decomposition of the processes $X$ and $\beta$:

$$X = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_j \varphi_j \quad \text{and} \quad \beta = \sum_{j=1}^{\infty} \gamma_j \varphi_j$$

with $\xi_j$ centered and uncorrelated random variables. Then

$$\langle X, \varphi_j \rangle = \sqrt{\lambda_j} \xi_j.$$
A first estimation of $\text{Var}(\mathbb{E}(Y|X))$

Notice that

$$
\mathbb{E}(YX) = \mathbb{E}(\langle X, \beta \rangle X) = \mathbb{E} \left[ \left( \sum_{l=1}^{\infty} \sqrt{\lambda_l} \gamma_l \xi_l \right) \left( \sum_{l=1}^{\infty} \sqrt{\lambda_l} \xi_l \varphi_l \right) \right]
$$

$$
= \mathbb{E} \left[ \left( \sum_{l=1}^{\infty} \lambda_l \gamma_l \xi_l^2 \varphi_l \right) \right] = \sum_{l=1}^{\infty} \lambda_l \gamma_l \varphi_l
$$
A first estimation of $\text{Var}(\mathbb{E}(Y|X))$

Notice that

$$\mathbb{E}(YX) = \mathbb{E}(<X, \beta > X) = \mathbb{E}\left[\left(\sum_{l=1}^{\infty} \sqrt{\lambda_l} \gamma_l \xi_l \right) \left(\sum_{l=1}^{\infty} \sqrt{\lambda_l} \xi_l \varphi_l \right)\right]$$

$$= \mathbb{E}\left[\left(\sum_{l=1}^{\infty} \lambda_l \gamma_l \xi_l^2 \varphi_l \right)\right] = \sum_{l=1}^{\infty} \lambda_l \gamma_l \varphi_l$$

As a consequence, $\gamma_j = \frac{1}{\lambda_j} < \mathbb{E}(YX), \varphi_j >$ that is naturally estimated by

$$\hat{\gamma}_j = \frac{1}{\lambda_j} \frac{1}{n} \sum_{i=1}^{n} <X_i, \varphi_j > Y_i.$$
A first estimation of $\text{Var}(\mathbb{E}(Y|X))$

First, we have

$$\hat{\gamma}_j = \frac{1}{\lambda_j} \frac{1}{n} \sum_{i=1}^{n} <X_i, \varphi_j> Y_i.$$
A first estimation of $\text{Var}(\mathbb{E}(Y|X))$

- First, we have

$$\hat{\gamma}_j = \frac{1}{\lambda_j n} \sum_{i=1}^{n} < X_i, \varphi_j > Y_i.$$  

- Second, expansion in the KL basis gives

$$\text{Var}(\mathbb{E}(Y|X)) = \mathbb{E}(< \beta, X >^2) = \sum_{j=1}^{\infty} \lambda_j \gamma_j^2.$$  

A natural estimation of $\text{Var}(\mathbb{E}(Y|X))$ is then

$$\hat{E}_m^1 = \sum_{l=1}^{m} \frac{1}{\lambda_l} \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} Y_i < X_i, \varphi_l > Y_j < X_j, \varphi_l >.$$
A second estimation of $\text{Var}(\mathbb{E}(Y|X))$

- We consider another design of experiment: let $\varepsilon'$ be a copy of $\varepsilon$, independent of $X$ and $\varepsilon$ and

$$\begin{cases} Y &= \mu + \langle X, \beta \rangle + \varepsilon \\ Y^X &= \mu + \langle X, \beta \rangle + \varepsilon' \end{cases}$$

- Now the observations consist in

  1. $n$-sample of $(X, Y)$: $(X_i, Y_i)$, $1 \leq i \leq n$.
  2. $n$-sample of $(X, Y^X)$: $(X_i, Y_i^X)$, $1 \leq i \leq n$. 


A second estimation of $\text{Var}(\mathbb{E}(Y|X))$

- We consider another design of experiment: let $\varepsilon'$ be a copy of $\varepsilon$, independent of $X$ and $\varepsilon$ and

$$
\begin{align*}
Y &= \mu + <X, \beta> + \varepsilon \\
Y^X &= \mu + <X, \beta> + \varepsilon'
\end{align*}
$$

- Now the observations consist in
  
  (1) $n$-sample of $(X, Y) : (X_i, Y_i), 1 \leq i \leq n$.
  
  (2) $n$-sample of $(X, Y^X) : (X_i, Y_i^X), 1 \leq i \leq n$.

- $\text{Var}(Y)$ is naturally estimated by the empirical variance based on $(Y_1, \ldots, Y_n)$ and $(Y_1^X, \ldots, Y_n^X)$

$$
\frac{1}{2n} \sum_{i=1}^{n} \left[ (Y_i)^2 + (Y_i^X)^2 \right] - \left( \frac{1}{2n} \sum_{i=1}^{n} [Y_i + Y_i^X] \right)^2.
$$
A second estimation of $\text{Var}(\mathbb{E}(Y|X))$

- It remains to estimate $\text{Var}(\mathbb{E}(Y|X))$ that can be rewritten as
  \[ \text{Var}(\mathbb{E}(Y|X)) = \text{Cov}(Y, Y^X). \]
- A natural estimation of $\text{Var}(\mathbb{E}(Y|X))$ is then:
  \[
  \hat{E}^2 = \frac{1}{n} \sum_{i=1}^{n} Y_i Y_i^X - \left( \frac{1}{2n} \sum_{i=1}^{n} [Y_i + Y_i^X] \right)^2.
  \]
Straighforwardly $\hat{E}_m^1$ is biased and

$$B_m = \mathbb{E}(\hat{E}_m^1) - \text{Var}(\mathbb{E}(Y|X)) = \sum_{l=m+1}^{\infty} \lambda_l \gamma_l^2$$

whereas $\hat{E}^2$ is unbiased.
Straighforwardly $\hat{E}_m^1$ is biased and

$$B_m = \mathbb{E}(\hat{E}_m^1) - \text{Var}(\mathbb{E}(Y|X)) = \sum_{l=m+1}^{\infty} \lambda_l \gamma^2_l$$

whereas $\hat{E}_m^2$ is unbiased.

**Some statistical questions:**

1. Are $\hat{E}_m^1$ and $\hat{E}_m^2$ “good” estimators for $\text{Var}(\mathbb{E}(Y|X))$?
2. Are they consistent? If yes, what is the rate of convergence?
   *Answer*: Central Limit Theorem (cv in $\sqrt{n}$).
3. Are they asymptotically efficient?
4. Can we measure their quality at a fixed $n$?
   *Answer*: Berry-Esseen and/or concentration inequalities.
5. Are the estimators and designs of experiment comparable?
Outline of the talk

Estimators considered
   A first estimation of $\text{Var}(\mathbb{E}(Y|X))$
   A second estimation of $\text{Var}(\mathbb{E}(Y|X))$

Asymptotic properties of the estimators

Numerical Applications

Conclusion
Asymptotic properties of $\hat{E}_m^1$

Consistency: $\hat{E}_m^1$ and $\hat{E}_m^2 \xrightarrow{P} n \to \infty$ are consistent.
Asymptotic properties of $\hat{E}^1_m$

Consistency: $\hat{E}^1_m$ and $\hat{E}^2 \xrightarrow{\mathbb{P}}$ are consistent.

Asymptotic normality

$$\hat{E}^1_m = \sum_{l=1}^{m} \frac{1}{\lambda_l} \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} Y_i < X_i, \varphi_l > Y_j < X_j, \varphi_l >$$

$$= U_n K + P_n L - B_m + \text{Var}(\mathbb{E}(Y|X))$$

with $U_n K = \sum_{l=1}^{m} \frac{1}{\lambda_l} \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} Z^c_{i,l}$

and $P_n L = \frac{2}{n} \sum_{l=1}^{m} \sum_{i=1}^{n} \gamma_l Z^c_{i,l}$. 
Asymptotic properties of $\hat{E}_m^1$

We want to show

$$B_m^2 = o\left(\frac{1}{n}\right), \quad U_nK = o_P\left(\frac{1}{\sqrt{n}}\right), \quad \sqrt{n}P_nL \xrightarrow{n \to \infty} N(0, C(\beta))$$
Asymptotic properties of $\hat{E}_m^1$

We want to show

$$B_m^2 = o\left(\frac{1}{n}\right), \quad U_nK = o_p\left(\frac{1}{\sqrt{n}}\right), \quad \sqrt{n}P_nL \xrightarrow{\mathcal{L}} \mathcal{N}(0, C(\beta)).$$

Assumptions :

- (A1) $\mathbb{E}(\|X\|^4) < +\infty$ and $\mathbb{E}(\varepsilon^4) < +\infty$.
- (A2) $\sup_{l \geq 1} \mathbb{E}(\xi_l^4) < +\infty$.
- (A3) there exist $C > 0$ and $\delta > 1$ such that

$$\forall l \geq 1, \quad \lambda_l \leq Cl^{-\delta}.$$  

Now let $m = m(n) = \sqrt{nh(n)}$, where $h(n)$ satisfies : $h(n) \to 0$ and $\forall \alpha > 0, \ n^\alpha h(n) \to +\infty$ as $n \to +\infty$. 
Theorem (Asymptotic normality)

(i) Since \( \hat{E}_m^1 - \text{Var}(E(Y|X)) = U_n K + P_n L - B_m \)

and assuming (A1-3) and \( n^{1/2}(\delta + 2s) << m << \sqrt{n} \), one gets

\[
\begin{align*}
B_m^2 &= o\left(\frac{1}{n}\right) \quad \mathbb{E}\left((U_n K)^2\right) = o\left(\frac{1}{n}\right) \\
\sqrt{n}P_n L &\xrightarrow{L} \mathcal{N}(0, 4\text{Var}(Y < X, \beta >))
\end{align*}
\]

then \( \sqrt{n}(\hat{E}_m^1 - \text{Var}(E(Y|X))) \xrightarrow{L} \mathcal{N}(0, 4\text{Var}(Y < X, \beta >)) \).

(ii) Since \( \mathbb{E}(Y^4) < \infty \),

\[
\sqrt{n}(\hat{E}^2 - \text{Var}(E(Y|X))) \xrightarrow{L} \mathcal{N}(0, \text{Var}((Y - \mathbb{E}(Y))(Y^X - \mathbb{E}(Y^X))))
\].
We may assume that $h(n) = 1/\log(n)$, and hence $m(n) = \sqrt{n}/\log n$, to fill the condition

$$\forall \alpha > 0, \lim_{n \to \infty} n^\alpha h(n) = +\infty.$$  

The estimator $\hat{V}_m^\chi$ converges at the parametric rate $1/\sqrt{n}$, for any $\beta$. We could have chosen a smaller value of $m$ leading to the same asymptotic efficiency, but depending on $\delta$. 

Asymptotic properties of $\hat{S}_m^1$ and $\hat{S}^2$

Using the so-called Delta method, one can extend these properties of the numerators to the estimators of the Sobol index $S$:

**Theorem (Asymptotic Normality)**

(i) Under the same assumptions as in the previous theorem, we have

$$\sqrt{n} \left( \hat{S}_m^1 - S \right) \xrightarrow{L} \mathcal{N} \left( 0, \frac{\text{Var}(U)}{(\text{Var}(Y))^2} \right)$$

where $U := 2Y < X, \beta > -S(Y - \mathbb{E}(Y))^2$.

(ii) Since $\mathbb{E}(Y^4) < \infty$,

$$\sqrt{n} \left( \hat{S}^2 - S \right) \xrightarrow{L} \mathcal{N} \left( 0, \frac{\text{Var}(V)}{(\text{Var}(Y))^2} \right)$$

where $V := (Y - \mathbb{E}(Y))(Y^X - \mathbb{E}(Y)) - S^X/2 \left( (Y - \mathbb{E}(Y))^2 + (Y^X - \mathbb{E}(Y))^2 \right)$. 
Remark

- For independent inputs, we establish more generally in the product space
  - the consistency
  - the asymptotic normality
  - the asymptotic efficiency

of \( \hat{S}_m^1 := (\hat{S}_{m}^{(1,1)}, \ldots, \hat{S}_{m}^{(1,p)}) \) and \( \hat{S}^2 := (\hat{S}^{(2,1)}, \ldots, \hat{S}^{(2,p)}) \) to the vector of Sobol indices

\[
S := (S^{(1)}, \ldots, S^{(p)}),
\]

the indices 1 and 2 refer to the first and second estimators.

- One can also generalize these results to Sobol indices defined for subsets \( I \subset \{1, \ldots, p\} \).
Outline of the talk

Estimators considered
A first estimation of $\text{Var}(\mathbb{E}(Y|X))$
A second estimation of $\text{Var}(\mathbb{E}(Y|X))$

Asymptotic properties of the estimators

Numerical Applications

Conclusion
We consider the model with \( p = 2, \mu = 0 \) and \( \varepsilon = 0 \):

\[
Y = \langle \beta^1, X^1 \rangle + \langle \beta^2, X^2 \rangle
\]

1. **First Model**: \( \gamma^i = (\gamma^i_1, \gamma^i_2, \gamma^i_3, \ldots) \) for \( i = 1, \ldots, 2 \)

\[
\gamma^i_l = l^{\delta^i_l} \quad \text{for} \quad 1 \leq l \leq L \quad \text{and} \quad \gamma^i_l = 0 \quad \text{for} \quad l > L;
\]

with \( i = 1 \ldots 2 \) and \( \delta^i_l = (-1/2 - 1/100) \).

2. **Second Model**: \( \gamma^i = (0, \gamma^i_2, \gamma^i_3, \ldots) \) for \( i = 1, \ldots, 2 \).

3. **Third Model**: \( \gamma^i = (\gamma^i_3, \gamma^i_4, \gamma^i_5, \ldots) \) for \( i = 1, \ldots, 2 \).

We perform \( N_{\text{sim}} = 5000 \) simulations and we study the influence of the parameter \( n \), where \( 3n \) observations are used for both methods. We set \( L = 500 \) and \( m = \lfloor \sqrt{3n}/ \log(3n) \rfloor \).
### First Model: \( S = (0.5107, 0.4893) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{RMSE}(\hat{S}_m) )</th>
<th>( \text{RMSE}(\hat{S}_{SPF}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^2 )</td>
<td>( 10^{-2}[7.17, 7.21] )</td>
<td>( 10^{-2}[8.95, 9.14] )</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>( 10^{-2}[2.26, 2.20] )</td>
<td>( 10^{-2}[2.79, 2.83] )</td>
</tr>
</tbody>
</table>

### Second Model: \( S = (0.7535, 0.2465) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{RMSE}(\hat{S}_m) )</th>
<th>( \text{RMSE}(\hat{S}_{SPF}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^2 )</td>
<td>( 10^{-2}[8.07, 5.45] )</td>
<td>( 10^{-2}[7.80, 9.90] )</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>( 10^{-2}[2.52, 1.71] )</td>
<td>( 10^{-2}[2.41, 3.13] )</td>
</tr>
</tbody>
</table>

### Third Model: \( S = (0.8655, 0.1345) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{RMSE}(\hat{S}_m) )</th>
<th>( \text{RMSE}(\hat{S}_{SPF}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^2 )</td>
<td>( 10^{-1}[3.01, 0.48] )</td>
<td>( 10^{-2}[7.12, 9.97] )</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>( 10^{-2}[4.67, 1.28] )</td>
<td>( 10^{-2}[2.24, 3.17] )</td>
</tr>
</tbody>
</table>
We consider the model with $p=4$, $\mu = 0$ and $\varepsilon = 0$:

$$Y = \sum_{k=1}^{4} \langle \beta^k, X^k \rangle$$

1. First Model: $\gamma^i = (\gamma_1^i, \gamma_2^i, \gamma_3^i, \ldots)$ for $i = 1, \ldots, 4$
   $$\gamma_j^i = (l + 1)^{\delta_i} \quad \text{for} \quad 1 \leq l \leq L \quad \text{and} \quad \gamma_j^i = 0 \quad \text{for} \quad l > L;$$
   with $i = 1 \ldots 4$ and $\delta_i = (-1/2 - 1/100, -1, -2, 3/2)$.

2. Second Model: $\gamma^i = (0, \gamma_2^i, \gamma_3^i, \ldots)$ for $i = 1, \ldots, 4$.

3. Third Model: $\gamma^i = (\gamma_3^i, \gamma_4^i, \gamma_5^i, \ldots)$ for $i = 1, \ldots, 4$.

We perform $N_{sim} = 5000$ simulations and we study the influence of the parameter $n$, where $5n$ observations are used for both methods. We set $L = 500$ and $m = \lfloor \sqrt{5n}/\log(5n) \rfloor$. 

First Model : \( S = (0.5438, 0.2639, 0.0635, 0.1288) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{RMSE}(\hat{S}_m) )</th>
<th>( \text{RMSE}(\hat{S}_{SPF}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^2 )</td>
<td>( 10^{-2}[5.55, 4.29, 2.35, 3.22] )</td>
<td>( 10^{-2}[9.92, 9.80, 9.75, 9.63] )</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>( 10^{-2}[1.82, 1.36, 0.72, 0.99] )</td>
<td>( 10^{-2}[3.13, 3.12, 3.11, 3.06] )</td>
</tr>
</tbody>
</table>

Second Model : \( S = (0.7080, 0.2085, 0.0200, 0.0635) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{RMSE}(\hat{S}_m) )</th>
<th>( \text{RMSE}(\hat{S}_{SPF}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^2 )</td>
<td>( 10^{-2}[6.35, 3.92, 1.47, 2.31] )</td>
<td>( 10^{-1}[1.04, 0.99, 0.99, 0.99] )</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>( 10^{-2}[1.92, 1.22, 0.41, 0.73] )</td>
<td>( 10^{-2}[3.29, 3.15, 3.19, 3.14] )</td>
</tr>
</tbody>
</table>

Third Model : \( S = (0.7561, 0.1871, 0.0112, 0.0456) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \text{RMSE}(\hat{S}_m) )</th>
<th>( \text{RMSE}(\hat{S}_{SPF}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^2 )</td>
<td>( 10^{-2}[6.14, 3.72, 1.22, 2.01] )</td>
<td>( 10^{-1}[1.07, 1.00, 1.01, 0.99] )</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>( 10^{-2}[1.97, 1.17, 0.33, 0.60] )</td>
<td>( 10^{-2}[3.36, 3.16, 3.14, 3.13] )</td>
</tr>
</tbody>
</table>
Outline of the talk

Estimators considered
  A first estimation of $\text{Var}(\mathbb{E}(Y|X))$
  A second estimation of $\text{Var}(\mathbb{E}(Y|X))$

Asymptotic properties of the estimators

Numerical Applications

Conclusion
We construct two different estimators of

\[ S := (S^{(1)}, \ldots, S^{(p)}), \]

based on two different designs of experiment for the functional linear regression.

1. The first one \( \hat{S}_m^1 \) is based on the Karhunen-Loève expansion of the covariance operator \( \Gamma(f) = \mathbb{E}(< X, f > X) \) and performs better for large values of \( p \).

2. Nevertheless, it is more complex and requires the knowledge of the \( \lambda_j \) and \( \varphi_j \) that can be estimated in a future work.

3. The second is more general and applies whatever the context but is performing as well.
Bibliography


