École chercheurs ASPEN 2014
TP2 : Compléments de correction

Laurent Gilquin

6 Mai 2014

1 Etude de l'échantillonnage

Pour les fonction *sobolEff*,*sobolroalhs* (package *sensitivity*) ainsi que pour la fonction *rbd2006* (répertoire UTIL) on observera:

- le nombre d'évaluations du modèle,
- comment se répartissent les échantillons dans les espaces 2D?

Remarques
Pour cette section, on désigne par d le nombre de variables (paramètres) du modèle dont on souhaite estimer les indices de Sobol.

```
# ----------------------------
# Methode sobolEff
# ----------------------------
```

Afin d'illustrer la méthode on prend $N_s = 5$.
On commence par créer deux échantillons X_1, X_2 de quatre variables (nombres de colonnes) et de taille 5 (nombre de lignes).

$$X_1 = \begin{pmatrix} 0.92182057 & 0.49846228 & 0.62587006 & 0.728057408 \\ 0.64800010 & 0.54171384 & 0.01844564 & 0.652517769 \\ 0.02674223 & 0.39318287 & 0.75735374 & 0.938739943 \\ 0.43251820 & 0.33829621 & 0.63181015 & 0.271918610 \\ 0.06378569 & 0.68250909 & 0.19638286 & 0.187706074 \end{pmatrix}$$

$$X_2 = \begin{pmatrix} 0.10260336 & 0.43979332 & 0.4694285 & 0.4554818 \\ 0.02408165 & 0.55616500 & 0.6772088 & 0.3815535 \\ 0.21045131 & 0.78038702 & 0.8842801 & 0.9544125 \\ 0.06782885 & 0.06207149 & 0.9640411 & 0.1132256 \\ 0.19106040 & 0.20856520 & 0.8882240 & 0.1402232 \end{pmatrix}$$

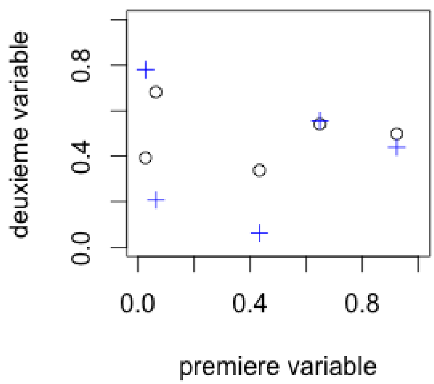
Le plan d'expérience crée par la méthode SobolEff recquiert : $N_s(d+1)$ évaluations du modèle, soit ici $5*(4+1) = 25$. Le plan d'expérience a la forme suivante:

$$X = \begin{pmatrix} X_1 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{pmatrix}$$

où X_1 est le premier échantillon précédent et les matrices C_i sont crées en remplaçant la i$^{\text{eme}}$ colonne de X_2 par la i$^{\text{eme}}$ colonne de X_1. Exemple pour C_1 on obtient:

$$C_1 = \begin{pmatrix} 0.92182057 & 0.43979332 & 0.4694285 & 0.4554818 \\ 0.64800010 & 0.55616500 & 0.6772088 & 0.3815535 \\ 0.02674223 & 0.78038702 & 0.8842801 & 0.9544125 \\ 0.43251820 & 0.06207149 & 0.9640411 & 0.1132256 \\ 0.06378569 & 0.20856520 & 0.8882240 & 0.1402232 \end{pmatrix}$$

On vient de "geler" la première variable. Graphiquement, lorsqu'on plot les colonnes 1 et 2 de X_1 et C_1 sur un même graphe, on observe que les points noir et les croix bleues sont alignées suivant l'axe des abscisse (axe correspondant à la première variable).



On remarque le même phénomène si l'on plot la colonne 1 et n'importe qu'elle autre colonne de X_1 et C_1.
On observe le même phénomène pour la i$^{\text{eme}}$ variable en utilisant X_1 et C_i et la colonne i.

```
# ----------------------------
# Methode sobolroalhs
# ----------------------------
```

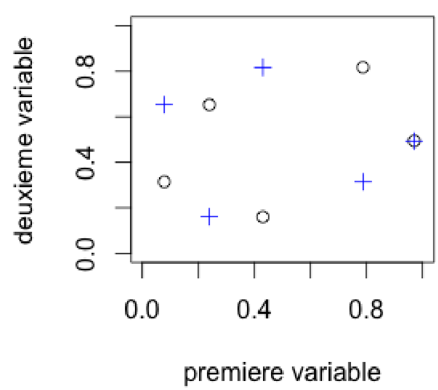
La méthode sobolroalhs ne recquiert pas d'échantillons X_1, X_2 ils sont automatiquement générés par la méthode. Il faut cependant préciser le nombre de variables ($factors$) et le nombre de lignes ($levels$) de l'échantillon.
Cette méthode génère alors deux plans lhs répliqués. Elle crée un premier plan lhs et le réplique en permutant les éléments de chaque colonne du plan. Le plan d'expérience total consiste en l'ensemble des deux plans répliqués et recquiert donc un totale de $2 \times levels$ évaluations du modèle.

Voici deux plans répliqués obtenus par la méthode pour $levels = 5$ et $factors = 4$:

$$X_1 = \begin{pmatrix} 0.0782587 & 0.3146720 & 0.1320852 & 0.99706681 \\ 0.4299305 & 0.1602540 & 0.6688786 & 0.03617451 \\ 0.9696113 & 0.4937460 & 0.2632904 & 0.25969159 \\ 0.7870688 & 0.8172876 & 0.4417141 & 0.79108133 \\ 0.2390640 & 0.6530176 & 0.9418843 & 0.59223919 \end{pmatrix}$$

$$X_2 = \begin{pmatrix} 0.0782587 & 0.6530176 & 0.1320852 & 0.25969159 \\ 0.9696113 & 0.4937460 & 0.6688786 & 0.03617451 \\ 0.7870688 & 0.3146720 & 0.4417141 & 0.99706681 \\ 0.2390640 & 0.1602540 & 0.9418843 & 0.79108133 \\ 0.4299305 & 0.8172876 & 0.2632904 & 0.59223919 \end{pmatrix}$$

On observe bien un réarangement des éléments de la première colonne, il en va de même pour les autre colonnes. Graphiquement, le principe de réplication se traduit de la manière suivante: pour chaque ensemble de deux colonnes de X_1, X_2 tout point est aligné avec un autre point suivant chaque axe.



Pour une valeur de $levels$ petite, il peut arriver que l'on observe une superposition (point le plus à droite).

```
# ----------------------------
# Methode rbd
# ----------------------------
```

Cette méthode fait partie des méthodes spectrales présentées dans le cours 2. C'est une méthode qui n'est pas adaptée à des fonctions irrégulières au sens de Fourier.

2 Analyse de sensibilité d'un modèle additif et/ou multiplicatif

2.1 Modèle additif

On effectue les analyses de sensibilité du modèle:

$$Y = X1 + X2$$

- Coder le modèle dans une fonction *somme2* (fonction somme2.R dans le répertoire UTIL du TP2)
- Calculer les indices de sobol d'ordre un et les indices totaux avec une méthode de Monte Carlo, vous utiliserez la fonction *sobolEff* du package *sensitivity* pour les indices d'ordre 1 et la fonction *soboljansen* du même package pour les indices totaux.
- Faire un plot des résultats, de l'échantillonnage et des réponses en fonctions des entrées.

Remarques
Pour le cas du modèle additif la somme des indices d'ordre principaux doit valoir 1, il n'y a pas d'interactions. Pour chacun des différents cas, si l'on augmente la taille de l'échantillon (n) on approche de plus en plus les valeurs théoriques des indices.

2.1.1 Cas 1

$$X_1 \sim U(-1, 1)$$
$$X_2 \sim U(-1, 1)$$

Les valeurs théoriques sont $S_1 = S_2 = \frac{1}{2}$.

2.1.2 Cas 2

$$X_1 \sim U(0, 2)$$
$$X_2 \sim U(-1, 1)$$

Les valeurs théoriques sont $S_1 = S_2 = \frac{1}{2}$.

2.1.3 Cas 3

$$\begin{aligned} X_1 &\sim U(-1, 1) \\ X_2 &\sim U(-2, 2) \end{aligned}$$

Les valeurs théoriques sont $S_1 = \frac{1}{5},\ S_2 = \frac{4}{5}$.

2.1.4 Cas 4

$$\begin{aligned} X_1 &\sim U(-1, 1) \\ X_2 &\sim \mathcal{N}\left(0, \tfrac{4}{3}\right) \end{aligned}$$

Les valeurs théoriques sont $S_1 = \frac{1}{5},\ S_2 = \frac{4}{5}$.

Conclusion

Ces quatre différents cas mettent en avant l'importance de la variance de chaque facteur sur la valeur de l'indice de sensibilité; la variance étant directement liée à l'incertitude que l'on a sur chaque facteur.

2.2 Modèle multiplicatif

$$Y = X1 * X2$$

- Coder le modèle dans une fonction *produit2* (fonction produit2.R dans le répertoire Util du TP2).
- Calculer les indices de sobol d'ordre un et les indices totaux avec une méthode de Monte Carlo, vous utiliserez la fonction *sobolEff* du package *sensitivity* pour les indices d'ordre 1 et la fonction *soboljansen* du même package pour les indices totaux.
- Faire un plot des résultats, de l'échantillonnage et des réponses en fonctions des entrées.
- On traitera les mêmes cas que précédemment.

Remarques
Pour le cas du modèle additif la somme des indices d'ordre principaux doit valoir 1, il n'y a pas d'interactions. Pour chacun des différents cas, si l'on augmente la taille de l'échantillon (n) on approche de plus en plus les valeurs théoriques des indices.

2.2.1 Cas 1

$$X_1 \sim U(-1, 1)$$
$$X_2 \sim U(-1, 1)$$

Les valeurs théoriques sont $S_1 = S_2 = 0$.

2.2.2 Cas 2

$$X_1 \sim U(0, 2)$$
$$X_2 \sim U(-1, 1)$$

Les valeurs théoriques sont $S_1 = 0,\ S_2 = \frac{3}{4}$.

2.2.3 Cas 3

$$X_1 \sim U(-1,1)$$
$$X_2 \sim \mathcal{N}\left(0, \frac{1}{3}\right)$$

Les valeurs théoriques sont $S_1 = 0,\ S_2 = \frac{3}{4}$.

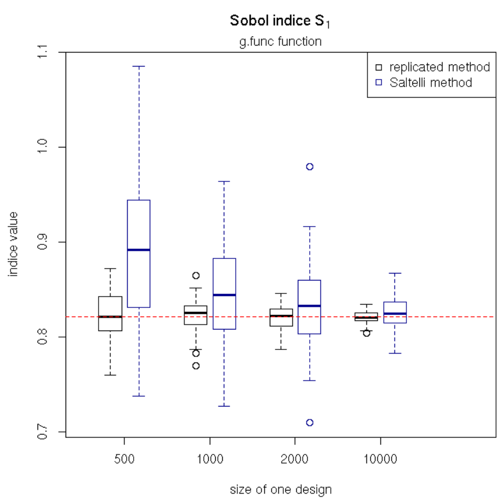
Conclusion

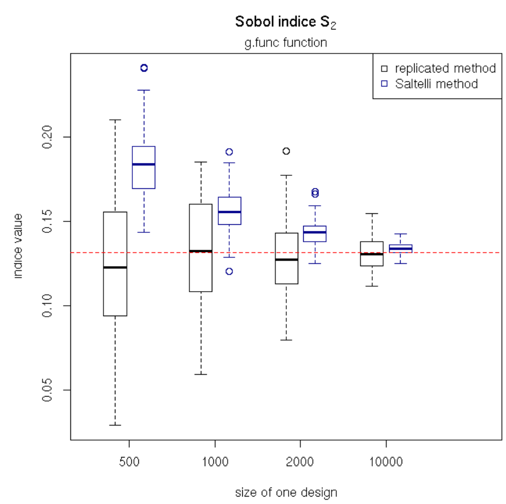
Bien que le modèle soit multiplicatif, suivant les lois des facteurs on peut obtenir des indices principaux non nuls.
La somme des effets principaux est strictement inférieure à 1. Dans le cas 2, l'effet de translation à également un impact sur la valeur de l'indice.

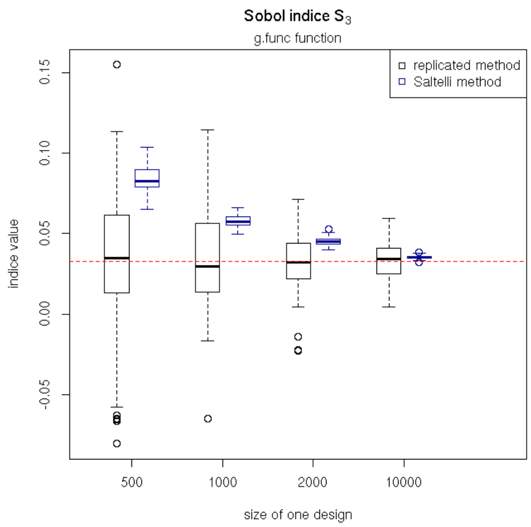
3 Etude de la fonction g-Sobol

- Pour $a = c(1, 4, 9)$, comparer les deux méthodes d'estimation en traçant des boxplots pour différentes valeurs de N (nombre d'évaluations) à l'aide des fonctions *compar* et *compar_plot* du script compar.R du répertoire UTIL du TP2. Si la légende gêne l'affichage graphique utiliser l'argument $legend = FALSE$ dans *compar_plot*.

```
# ---------------
# Comparaison Saltelli sobolroalhs avec a=c(1,4,9)
# ---------------
#ordre 1
# -------
```

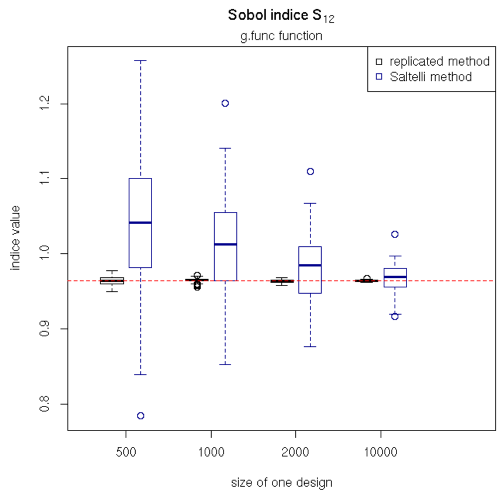


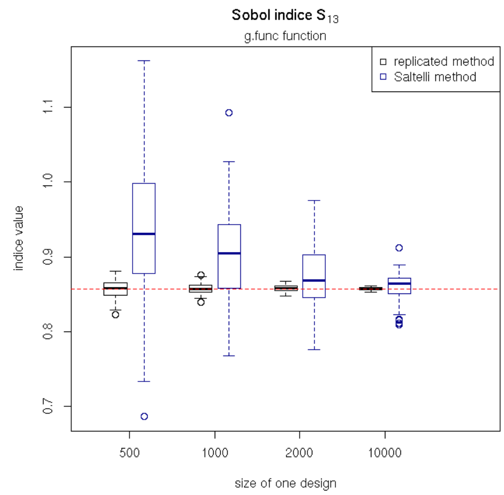


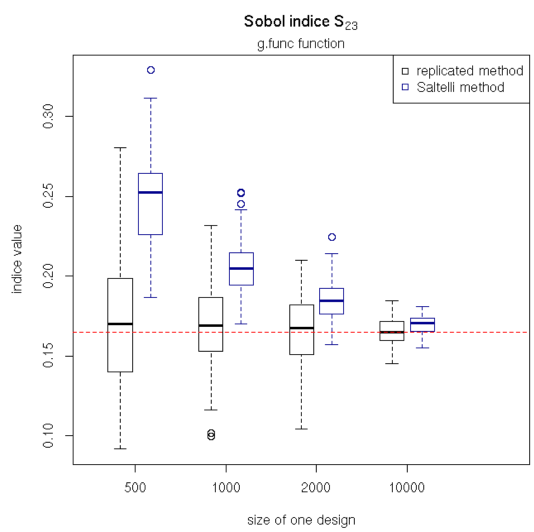


Sur les trois graphes ci-dessus on a calculé pour différentes tailles d'échantillons (valeurs sur axe des abscisses) 100 estimations de chacun des indices afin de pouvoir tracer des boxplots représentatif de ces ensembles d'estimations. Les boxplots bleus correspodent aux estimations données par la méthode de Saltelli, les boxplots noirs par la méthode sobolroalhs. Les lignes en pointillées rouges correspondent aux valeurs théoriques des indices de Sobol. On voit que la méthode sobolroalhs a plus de mal à estimer les indices proches de zéro que les indices proches de 1, mais par contre que l'estimation médiane (traits verticales sur les boxplot) est toujours proche de l'indice théorique. La méthode de Saltelli, elle, converge vers la valeur théorique mais recquiert un plus grand nombre d'évaluations afin de se rapprocher de la valeur théorique.

```
# -------
#ordre 2
# -------
```







On observe le même phénomène que pour les indices d'ordre 1. Pour une estimation précise, il est recommandé d'utiliser la méthode sobolroalhs.

4 Approche spectrales sur fonction g-Sobol et indbis

Le but de cet exercice est d'illustrer que pour des fonctions irrégulières les méthodes de type spectrale peuvent induire un biais qui ne disparaitra pas avec la convergence.
Cependant, une correction systématique de la partie du biais non liée à la régularité est proposée dans la méthode rbd2006.