

SENSILIB

Bibliothèque de fonctions pour l'analyse de sensibilité de modèles de simulation

Implémentation, validation et test de la méthode de Morris

Cyril Auclair (stagiaire master2 encadré par Samuel Buis mai-septembre 2013), 09/07/2013

Table des matières

I. Introduction	1
II. Implémentation	1
II.1 F_gen_variant_morris : Génération du plan d'expérience	2
II.2 Calcul des indices de Morris	3
III. Validation	3
III.1 Tests de la création du plan d'expérience	4
III.2 Vérification du calcul des indices avec une fonction linéaire	4
III.3 Vérification du calcul des indices avec le modèle test « g-fonction »	5

I. Introduction

Ce document vise à décrire l'implémentation de la Méthode de Morris au sein de la bibliothèque de fonctions d'analyse de sensibilité SENSILIB. Le document intitulé «La méthode de Morris et ses extensions» servira de référence pour la description de la méthode.

Nous décrirons ici rapidement la structure de l'implémentation informatique de la méthode de Morris dans la bibliothèque SENSILIB (les choix relatifs à l'implémentation de cette méthode et les paramètres d'entrée et de sortie sont commentés dans le détail dans le code source des différentes fonctions citées section II), et les tests qui ont été réalisés pour valider cette implémentation.

II. Implémentation

L'implémentation de la méthode de Morris est structurée en 2 fonctions principales :

- F_gen_variant_morris : génère le plan d'expérience numérique d'entrée du modèle à analyser nécessaire pour le calcul des indices ;
- F_sensi_analysis_morris : calcule les indices de Morris à partir des simulations réalisées avec le modèle analysé sur le plan d'expérience fourni par F_gen_variant_morris.

Pour une description détaillée des paramètres d'entrée et de sortie de ces fonctions, reportez-vous directement au code source des fonctions qui est largement commenté et au document «La méthode de Morris et ses extensions» qui sert de référence à ces commentaires.

II.1 F_gen_variant_morris : Génération du plan d'expérience

Dans le formalisme matriciel sous lequel nous avons implémenté l'algorithme de Morris sous MATLAB®, une trajectoire dans le plan d'expérience numérique Ω est une matrice de nombre de colonnes égal à la dimension de Ω . Chaque ligne de cette matrice correspond ainsi à une position dans le plan d'expérience numérique et le nombre de lignes est le nombre de points constituant cette trajectoire (voir Figure 2).

```
>> F_gen_variant_Morris(4,8,4,1,50)

ans =

    0    0.4286    0.5714    0.7143
    0    0.4286         0    0.7143
    0    0.4286         0    0.1429
    0    1.0000         0    0.1429
 0.5714    1.0000         0    0.1429
```

Figure 1 : Exemple de matrice trajectoire dans le formalisme que nous utilisons. Il s'agit bien d'une trajectoire « One-At-a-Time » comme prescrit par la Méthode de Morris puisque, à chaque ligne, une et une seule composante diffère par rapport à celles de la ligne précédente.

Pour obtenir de telles matrices trajectoires, nous effectuons l'algorithme suivant (décrit dans Morris (1991)) :

- Nous prenons d'abord $\Delta = \frac{n}{p-1}$, avec p et n des vecteurs contenant respectivement le nombre de niveaux de la grille régulière pour chaque facteur et le nombre de pas de déplacement.
- Nous effectuons un tirage aléatoire de x^* avec chaque composante choisie aléatoirement dans $\{ 0, \frac{1}{p-1}, \frac{2}{p-1}, \dots, 1-\Delta \}$ (la limite supérieure en $1-\Delta$ sert à éviter que des points de la trajectoire sortent de la fourchette $[0;1]$).
- Nous obtenons ensuite les matrices trajectoires $Btmp^*$ par l'équation suivante :

$$Btmp_i^* = \left(J_{k+1,1} \cdot x^* + \frac{\Delta}{2} \cdot [(2B - J_{k+1,k}) \cdot D^* + J_{k+1,k}] \right) \cdot P^* \quad \text{avec les} \quad \text{définitions de matrice suivantes :}$$

- B est une matrice strictement triangulaire inférieure de taille $(k+1,k)$ dont les éléments non-nuls valent 1 ;
- $J_{k+1,k}$ et $J_{k+1,1}$ sont des matrices de format respectifs $(k+1,k)$ et $(k+1,1)$ dont tous les éléments valent 1 ;
- D^* est une matrice diagonale de rang k dont tous les éléments non-nuls valent 1 ou -1 avec une probabilité égale et indépendante entre chaque élément (elle introduit un facteur aléatoire pour le sens de déplacement dans l'espace) ;
- P^* est une matrice de permutation de format (k,k) dont chaque colonne ne doit contenir qu'un seul élément non-nul valant 1 et qui ne

contient pas 2 fois la même colonne. Elle introduit un facteur aléatoire dans le choix de la dimension à explorer (facteur sur lequel appliquer le déplacement).

Cet algorithme est répété r fois pour obtenir un groupe de r trajectoires et le tout est répété Q fois afin de sélectionner le groupe de r trajectoires de distance totale maximale parmi les Q groupes de r trajectoires générés. Les matrices B_{tmp}^* du groupe sélectionné sont alors concaténées verticalement pour former une seule matrice de format $(r(k+1),k)$.

La matrice obtenue est ensuite transformée dans l'espace des facteurs en fonction des lois de probabilités (bornées) définies par l'utilisateur par application des fonctions inverse des fonctions de distribution cumulatives des lois associées.

Une option permet de répéter cet algorithme m fois afin de calculer une incertitude sur les indices calculés (écart-type des indices calculés sur les m groupe de r trajectoires sélectionnés).

II.2 Calcul des indices de Morris

L'algorithme utilisé est le suivant :

- A partir de la matrice contenant le plan d'expérience, pour chaque trajectoire on obtient une matrice de format (k,k) dont chaque élément de i ème ligne et de j ème colonne contient la valeur du « saut » pour le j ème déplacement et pour le i ème facteur.
- A partir de celle-ci, on obtient la matrice F de format (k,r) , dont chaque élément est égal au taux d'accroissement du facteur correspondant pour la trajectoire correspondante et la matrice G , de même format encore, telle que $G_{i,j} = |F_{i,j}|$.
- On obtient au final les vecteurs μ_i , σ_i et μ_i^* en faisant les opérations suivantes sur l'ensemble des trajectoires :
 - $\mu_i = \overline{F_i}$,
 - $\sigma_i = \text{std}(F_i)$,
 - $\mu_i^* = \overline{G_i}$.

Deux dimensions supplémentaires sont en fait prises en compte dans les tableaux décrit ci-dessus : une dimension pour les différentes variables simulées analysées, et une dimension pour la répétition du plan d'expérience (par réplication ou bootstrap selon le choix de l'utilisateur) si l'utilisateur souhaite avoir une idée de l'incertitude sur les indices due à la répartition des trajectoires dans l'espace.

III. Validation

On s'intéressera dans cette partie à la validation de l'implémentation de la méthode de Morris. Cette validation se décompose en différents tests :

- vérification de l'algorithme pour la création du plan d'expérience numérique
- vérification du calcul des indices et de leur convergence sur des modèles tests

III.1 Tests de la création du plan d'expérience

Dans la Méthode de Morris, l'analyse de sensibilité repose sur la génération d'un plan d'expérience, décrit dans le document « La Méthode de Morris et ses extensions ».

Afin de tester si l'ensemble de trajectoires dans l'espace des phases sélectionné par l'algorithme répondait bien à l'exigence de maximisation de la distance totale D , nous avons représenté dans la Figure 2 les groupes de trajectoires ayant des distances internes minimale et maximale obtenues pour un plan d'expérience à 2 facteurs, un nombre de trajectoire par groupe de $r=10$ et un nombre de groupe de trajectoires de $Q=500$. On peut constater de visu que le plan d'expérience sélectionné a bien des trajectoires mieux réparties dans l'espace que celui qui mène à la distance minimale.

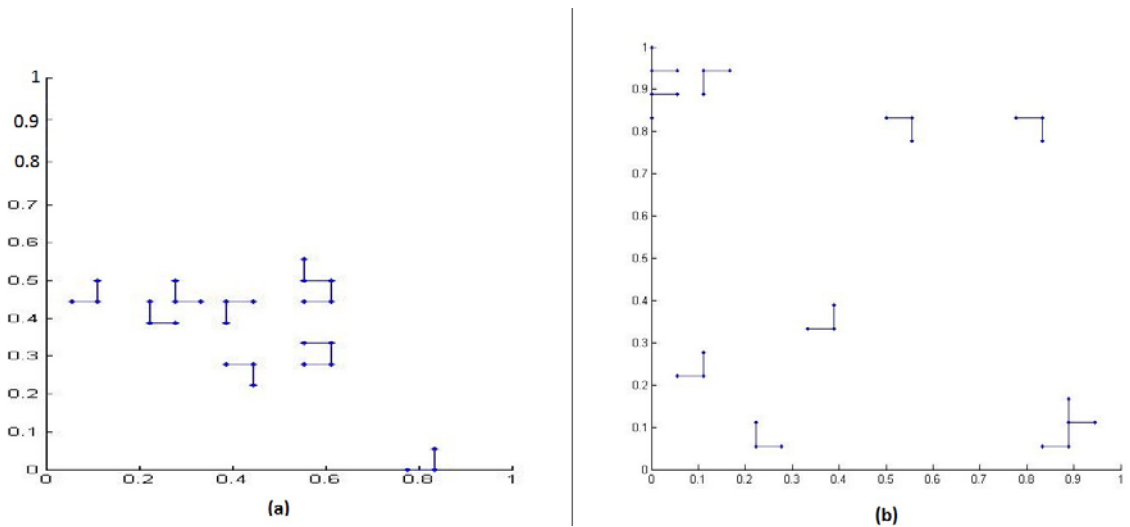


Figure 2 : images des trajectoires dans l'espace des phases ($Q = 500$ et $r = 10$) pour deux cas : (a) ensemble de trajectoires de distance totale D minimale et (b) ensemble de trajectoires de distance totale D maximale

III.2 Vérification du calcul des indices avec une fonction linéaire

Le premier test que nous avons réalisé concernant le calcul des indices a consisté à vérifier si, avec un modèle strictement linéaire, nous retrouvons bien, pour μ et μ^* , des valeurs des taux d'accroissement correspondant aux paramètres multiplicatif de chaque facteur et des valeurs nulles ou quasi-nulles pour σ . Pour ce faire, nous avons appliqué $F_sensi_analysis_Morris$ sur une matrice égale au produit du plan d'expérience généré par une matrice-colonne dont chaque élément est égal au numéro de sa ligne, pour simuler un modèle de type $f(x) = \sum_{i=1}^k i \cdot x_i$. Nous avons bien obtenu le résultat escompté (voir Figure 3).

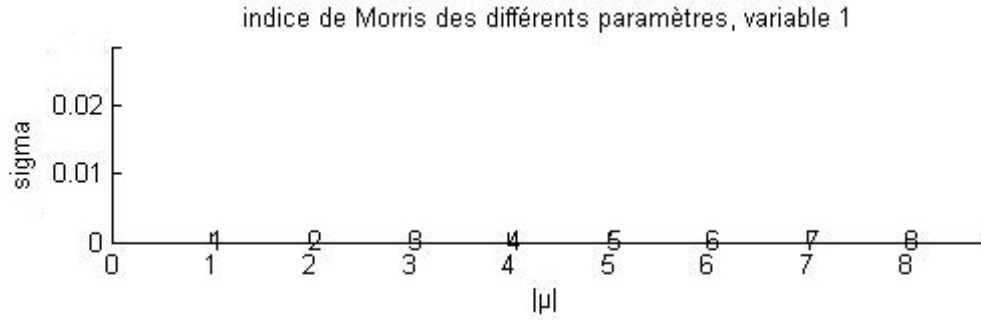


Figure 3: Graphique σ en fonction de $|\mu|$ pour chacun des 8 facteurs en entrée pour la fonction linéaire décrite à la sous-section III.2. Nous pouvons bien observer des valeurs de σ nulles et, pour chaque facteur, des valeurs de $|\mu|$ égale au coefficient multiplicateur correspondant.

III.3 Vérification du calcul des indices avec le modèle test « g-function »

L'autre test que nous avons réalisé est l'application à la fonction g de Sobol', dans la configuration $a_i = \{0, 1, 4.5, 9, 99, 99, 99, 99\}$, qui est classiquement utilisée dans la littérature pour tester les méthodes d'analyse de sensibilité (cf. documentation SensiLib sur les modèles tests). Les valeurs théoriques, pour cette fonction et dans cette configuration, des indices de sensibilité totaux de Sobol', dont le classement doit être comparable à ceux des indices de Morris, sont données dans le tableau suivant ainsi que les résultats obtenus en appliquant `F_gen_variant_Morris` et `F_sensi_analysis_Morris` pour différentes valeurs de p.

	p=2	p=4	p=8	ST_i de Sobol'
X1	0 ($\sigma = 0$)	2.4722 ($\sigma = 1.5337$)	3.5874 ($\sigma = 0.3925$)	0.787
X2	0 ($\sigma = 0$)	1.3420 ($\sigma = 0.8616$)	1.8941 ($\sigma = 0.1755$)	0.242
X3	0 ($\sigma = 0$)	0.5174 ($\sigma = 0.3256$)	0.6747 ($\sigma = 0.0509$)	0.034
X4	0 ($\sigma = 0$)	0.2767 ($\sigma = 0.1731$)	0.3839 ($\sigma = 0.0396$)	0.010
X5	0 ($\sigma = 0$)	0.0261 ($\sigma = 0.0162$)	0.0419 ($\sigma = 0.0043$)	$1.05 \cdot 10^{-4}$
X6	0 ($\sigma = 0$)	0.0246 ($\sigma = 0.0153$)	0.0387 ($\sigma = 0.0059$)	$1.05 \cdot 10^{-4}$
X7	0 ($\sigma = 0$)	0.0275 ($\sigma = 0.0177$)	0.0413 ($\sigma = 0.0045$)	$1.05 \cdot 10^{-4}$
X8	0 ($\sigma = 0$)	0.0294 ($\sigma = 0.0190$)	0.0409 ($\sigma = 0.0046$)	$1.05 \cdot 10^{-4}$

Tableau 1 : Valeurs des indices de Morris μ^* et des incertitudes associées pour 10 répétitions de l'algorithme pour la fonction g de Sobol' avec $r=20$, $Q=100$ et différentes valeurs de p. La dernière colonne contient les valeurs théoriques des indices totaux de Sobol' pour cette fonction.

On voit que le classement des indices de Morris est correct pour $p=4$ et $p=8$. Pour $p=2$, les indices sont nuls, comme attendu, car la fonction est symétrique par rapport à $x = 1/2$ ¹. On note que l'incertitude sur les indices de Morris (σ entre parenthèses sous l'indice) obtenu en calculant l'écart type sur les valeurs des indices calculés pour plusieurs répétitions de la méthode est plus faible pour $p=8$ que pour $p=4$, ce qui était également attendu (si p est paire, plus p est petit mieux on va capter la structure linéaire par symétrie de ce modèle).

Les graphiques de σ en fonction de $|\mu|$ pour chacun des 8 facteurs en entrée pour les 3 valeurs de p (dans tous les graphiques $n = \frac{4}{p-1} = 1$) de la fonction g de Sobol' sont

présentés Figure 4. Nous retrouvons sur les graphiques **(b)** et **(c)** la hiérarchie de l'importance des indices telle qu'elle est donnée par la colonne « STi de Sobol' » du Tableau 1 si l'on considère l'indice σ . Les valeurs de $|\mu|$ sont très faibles, comme attendu étant donné la non-monotonie et la symétrie du modèle (cf. ci-avant), et ce d'autant plus que p est grand. Sur le graphique **(a)**, toutes les valeurs sont confinées autour de $|\mu|=0$ et $\sigma=0$.

¹ En effet, la fonction g vaut un produit de fonctions g_i telles que $g_i = \frac{|4x-2|+a_i}{1+a_i}$ avec $a_i \geq 0$. Or, si $n = p-1$ (ce qui est le cas ici) la grille régulière ne peut prendre que 2 valeurs : 0 ou 1 ; et $\frac{|4 \times 1 - 2| + a_i}{1 + a_i} = \frac{|4 \times 0 - 2| + a_i}{1 + a_i} = \frac{2 + a_i}{1 + a_i}$. Ainsi, nous avons nécessairement $|\mu| = 0$.

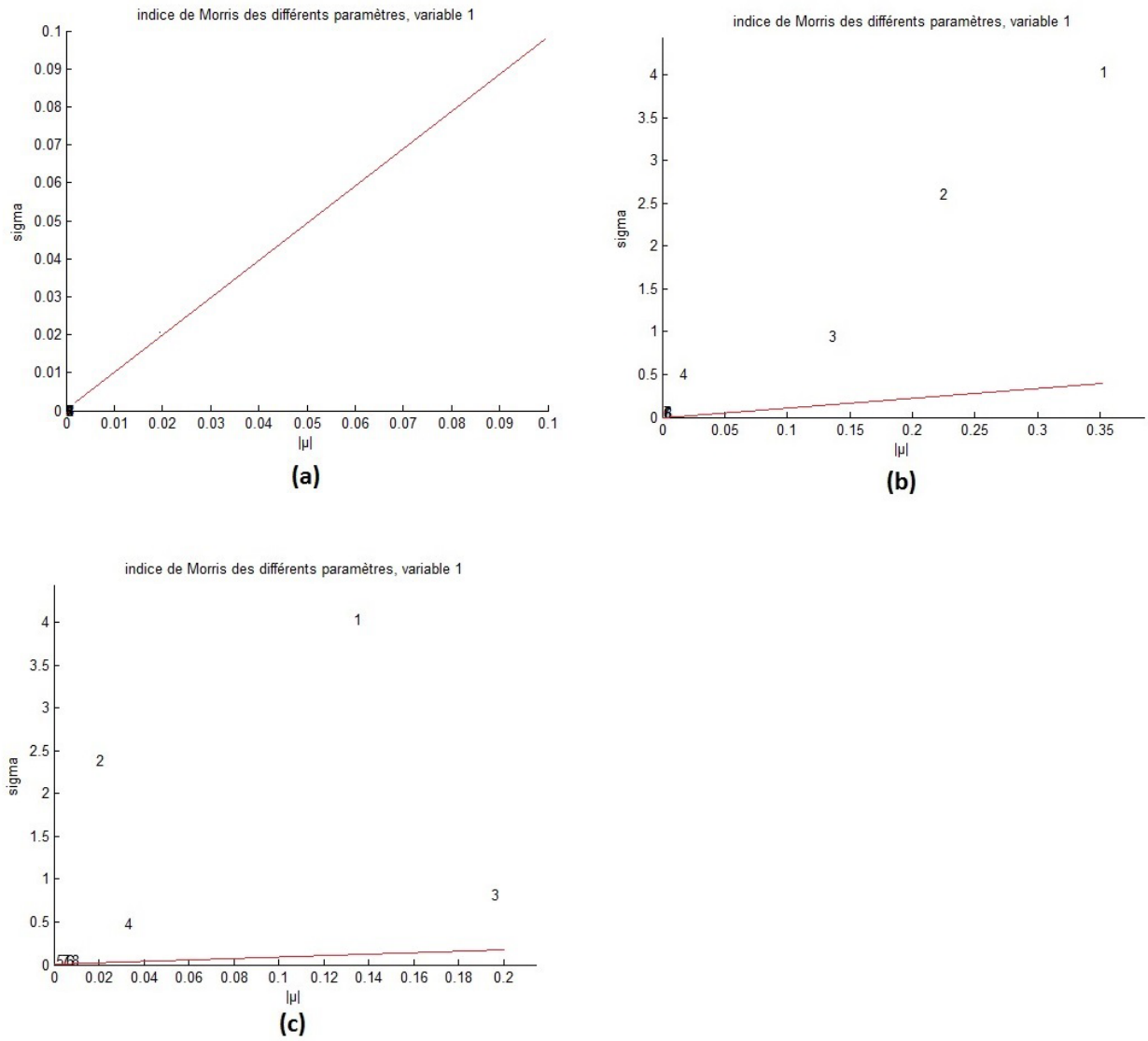


Figure 4 : Graphiques de σ en fonction de $|\mu|$ pour chacun des 8 facteurs en entrée – numérotés de 1 (pour X1) à 8 (pour X8) – pour 3 tests différents (dans tous $n = \frac{A}{p-1} = 1$) de la fonction g de Sobol'. (a) $p=2$. (b) $p = 4$. (c) $p=8$. La droite en rouge est la bissectrice entre les abscisses et les ordonnées.

Bibliographie

Max D. Morris (1991), “Factorial sampling plans for preliminary computational experiments”, *Technometrics* Vol. 33, 161-174