

SENSILIB

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

La méthode de Morris et ses extensions

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

Table des matières

I.	Introduction	1
II.	Principe de la méthode.....	1
III.	Optimisation de la dispersion des trajectoires.....	2
IV.	Calcul des indices de sensibilité.....	3
V.	Conclusion.....	3

I. Introduction

La Méthode de Morris, proposée dans Morris (1991), est une méthode d'analyse de sensibilité de type screening permettant d'identifier les facteurs qui peuvent être considérés comme d'effet négligeable et ceux qui doivent être considérés comme importants parmi un grand nombre de facteurs.

II. Principe de la méthode

Cette méthode statistique est basée sur un format de plan d'expérience dans lequel les facteurs en entrée (de nombre k), ont des valeurs comprises entre 0 et 1 inclus et tels que les vecteurs de facteurs X (de taille k donc) dont on souhaite évaluer les effets sont astreints à une *grille régulière* de dimension k et à p niveaux. Les composantes X_i peuvent donc prendre leurs valeurs dans $\{0, 1/(p+1), 2/(p+1), \dots, 1\}$. Une transformation simple permet de ramener ce plan d'expérience dans l'espace des facteurs du modèle, non nécessairement définis entre 0 et 1, et des valeurs différentes de p peuvent être définies selon les facteurs.

Il est alors défini, pour le modèle testé représenté ici par la fonction y , un vecteur d'effets élémentaires de composantes d_i , avec :

$$d_i = \frac{[y(x_1, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k) - y(x)]}{\Delta}$$

où $1 \leq i \leq k$ et $x=(x_1, x_2, \dots, x_k)$ est un point de Ω tel que le point transformé $(x + e_i \Delta)$, avec e_i un vecteur de zéros ayant un 1 comme $i^{\text{ème}}$ composante, soit également situé sur un point de la grille régulière défini (Δ est donc un multiple de $1/(p-1)$). La distribution des effets élémentaires est alors approximée par les distributions finies F_i et G_i valant respectivement la valeur moyenne de d_i et la valeur moyenne des valeurs absolues de d_i sur un échantillon aléatoire de r valeurs x de X . Cet échantillon de r valeurs de X peut être réalisé de manières différentes :

1. Tirer aléatoirement r valeurs de x puis effectuer les simulations correspondantes (celles pour les r x et celles à partir des $(x + e_i \Delta)$). Le coût de l'algorithme est alors de $2rk$ simulations.
2. Par la méthode des trajectoires préconisée par Morris, qui consiste à tirer aléatoirement dans l'espace une valeur de base x^* pour le vecteur X à partir duquel un chemin aléatoire est construit. Le premier point est obtenu en augmentant ou pas x^* de Δ pour une composante (tirée aléatoirement), le deuxième point est obtenu en augmentant ou pas le point précédent de Δ pour une autre composante, etc... Ce type de trajectoires dans lesquelles une seule composante est modifiée à la fois est appelée One-At-a-Time (voir Figure 1). Hormis les points aux extrémités des trajectoires, chacun sert ainsi pour le calcul de 2 taux d'accroissement, ce qui réduit le nombre de simulations à $r(k+1)$.

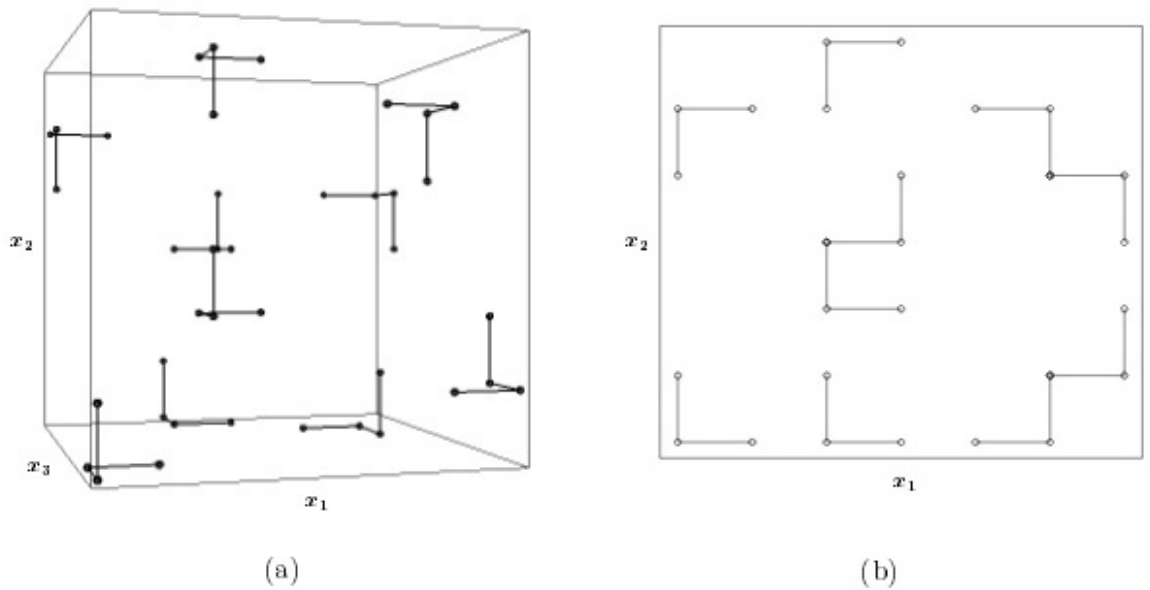


Figure 1 (a) Trajectoires One-At-a-Time de Morris dans une grille régulière de dimension 3 avec $r = 10$ trajectoires. (b) Projections de ces trajectoires dans le plan (x_1, x_2) . Figure extraite de Pujol (2009)

III. Optimisation de la dispersion des trajectoires

Afin de maximiser la dispersion de ces trajectoires dans le plan d'expérience pour bien couvrir l'espace des facteurs, Campolongo et al (2007) propose de générer un grand nombre d'ensembles de r trajectoires (sans faire les simulations associées) puis de sélectionner le groupe ayant la plus forte distance entre ses trajectoires, la distance entre 2 trajectoires m et l étant définie par :

$$d_{ml} = \sum_{i=0}^{k+1} \sum_{m=1}^{k+1} \sqrt{\sum_{z=1}^k [X_i^m(z) - X_j^l(z)]^2}$$

Où $X_i^m(z)$ est la $z^{\text{ème}}$ coordonnée du $i^{\text{ème}}$ point de la $m^{\text{ème}}$ trajectoire. La distance totale D entre les r trajectoires d'un groupe donné vaut alors :

$$D = \sqrt{\sum_{i=1}^r \sum_{j=i+1}^r d_{ij}^2}$$

IV. Calcul des indices de sensibilité

Les indices de sensibilité de la méthode de Morris sont des statistiques sur les distributions F_i et G_i . Trois sont classiquement calculés pour chaque facteur en entrée (et pour chaque variable simulée) : μ , la valeur moyenne de la distribution F_i , c'est-à-dire des effets élémentaires d_i , σ , l'écart-type de la distribution F_i , μ^* , la valeur moyenne de la distribution G_i , c'est-à-dire des valeurs absolues des effets élémentaires d_i . Ces indices permettent : de faire un classement des facteurs par ordre d'importance dans le modèle, d'identifier les facteurs négligeables, et d'avoir des éléments sur la monotonie et la linéarité du modèle.

Si le modèle est monotone, μ mesure l'effet de chaque facteur dans le modèle et permet de classer ceux-ci. Si le modèle est non-monotone, μ ne permet plus de classer les facteurs car les indices d_i peuvent se compenser. μ^* est alors utilisé pour classer les facteurs. Une valeur de σ non-nulle indique soit une non-linéarité du modèle par rapport au facteur concerné, soit la présence d'interactions impliquant ce facteur.

Classiquement, on utilise μ^* pour classer les facteurs par ordre d'importance et on trace le graphe de σ en fonction de μ pour avoir des informations sur la géométrie du modèle (Figure 2).

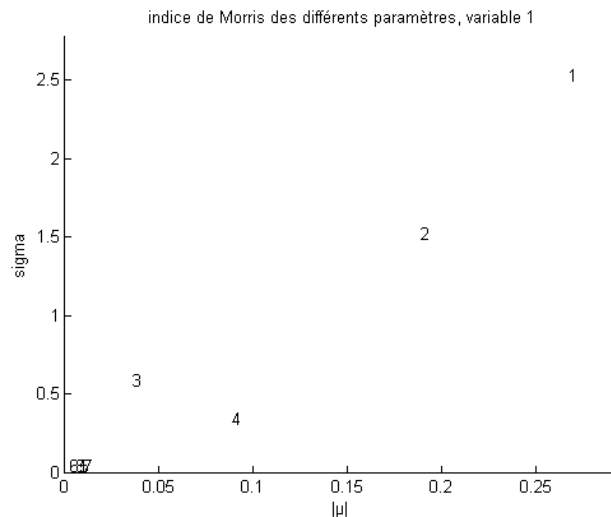


Figure 2 : Exemple de figure permettant d'interpréter les indices de Morris. Des valeurs faibles de μ concomitantes à des valeurs relativement fortes de σ indiquent ici une non-monotonie du modèle analysé.

V. Conclusion

La Méthode de Morris telle qu'elle est présentée dans Morris (1991) est une méthode efficace d'analyse de sensibilité par screening qui connaît un certain succès

notamment depuis les publications de Cariboni et al (2007) et Campolongo et al. (2007).

Comme dans toute analyse de sensibilité, il reste important d'évaluer la fiabilité des indices calculés. Dans la méthode de Morris, en plus du caractère aléatoire de la position des trajectoires dans l'espace (point de départ, ordre des composantes modifiées et sens de déplacement), plusieurs paramètres peuvent impacter les résultats selon la géométrie du modèle, notamment le nombre de modalités par facteur, la taille du pas Δ , et le nombre de trajectoires r . La variation des indices en fonction de celle de ces paramètres est donc importante à étudier dans chaque cas.

Plusieurs extensions ont été proposées depuis Morris (1991). Celle de Campolongo et al. (2007) été décrite ci-avant. Pujol (2009) propose un plan d'expérience basé sur des simplexes permettant d'être ré-utilisé pour la construction de méta-modèles bénéficiant des résultats de la méthode de Morris. Santiago et al (2012) proposent de calculer les indices de Morris sur un plan d'expérience quelconque. Cela est d'un grand intérêt mais peut cependant conduire à des biais dans les indices calculés (ce qui est visible dans les résultats présentés dans cette publication mais non commenté ni discuté).

Bibliographie

- Max D. Morris (1991), "Factorial sampling plans for preliminary computational experiments", *Technometrics* Vol. 33, 161-174
- G. Pujol (2009) "Simplex-based screening designs for estimating metamodels", *Reliability Engineering and System Safety* 94, 1156-1160
- F. Campolongo, J. Cariboni, et A. Saltelli (2007) "An effective screening design for sensitivity analysis of large models" *Environmental Modelling and Software* 22: 1509-1518
- J. Cariboni, D. Gatelli, R. Liska, A. Saltelli (2007). "The role of sensitivity analysis in ecological modelling." *Ecological Modelling* 203(1-2): 167-182.
- J. Santiago, B. Corre, M. Claeys-Bruno, M. Sergent (2012). Improved sensitivity through Morris extension. *Chemometrics and Intelligent Laboratory Systems* 113, 52-57.