TD 4 – Analyse de sensibilité, Sobol

Bertrand Iooss
Polytech Nice Sophia

1 Estimateur pick-freeze

a) Calculez les indices de Sobol du premier ordre du modèle $Y = X_1^2 [1 + \cos(X_2)]^2$ où $X_1 \sim U[0,1]$ et $X_2 \sim N(0,1)$.

L'algorithme est le suivant :
1) construire 2 matrices des entrées (par exemple de taille 1000) indépendantes l'une de l'autre,
2) calculer le dénominateur des indices de Sobol,
3) calculez le numérateur des indices de Sobol par estimation pick-freeze (slide 26 du cours),
4) calculez les indices de Sobol.

b) Calculer les indices de Sobol avec la fonction *sobol()* du package "sensitivity". On calculera également l'indice d'ordre 2 entre X_1 et X_2. On activera également l'option bootstrap (parameter nboot) de la fonction afin d'estimer la variabilité des estimateurs.

c) Prendre des matrices de taille 10000 puis refaites-la question b).

2 Modèle de bras de robot

On étudie à présent quelques fonctions du package "sensitivity" permettant de calculer les indices de Sobol' du 1er ordre et totaux sur la fonction "robot" modélisant un bras de robot à 4 segments. La formule analytique, qui évalue la distance de l'extrémité du bras du robot à l'origine et qui prend comme entrées les longueurs L_i (de loi uniforme sur [0,1]) de chaque segment et les angles θ_i (de loi uniforme sur $[0,2\pi]$) de chaque jonction, est la suivante :

$$f(\mathbf{x}) = \left(u^2 + v^2\right)^{0.5}, \text{where}$$
$$u = \sum_{i=1}^{4} L_i \cos\left(\sum_{j=1}^{i} \theta_j\right)$$
$$v = \sum_{i=1}^{4} L_i \sin\left(\sum_{j=1}^{i} \theta_j\right)$$

Pour utiliser la fonction, il suffira de simuler 8 variables U[0,1].

a) A l'aide d'un échantillon Monte Carlo de taille 1000, calculez les indices SRC^2 (fonction *src()* du package "sensitivity"). Sont-ils des indices de sensibilité valides dans le cas présent ? Pourquoi ?

b) Calculez les indices de Sobol (1er ordre et totaux) avec la fonction *sobol2002()*, qui prend en entrée 2 échantillons Monte Carlo indépendants de taille *n*. Les formules des numérateurs des indices implémentées dans *sobol2002()* sont celles de la slide 19 du cours. Estimez les variabilités des indices de Sobol' par bootstrap. Commencez avec la taille *n*=1000. Quels problèmes observez-vous sur les estimations des indices de Sobol' ? Augmentez la taille à *n*=10000 si la variabilité des estimateurs ne vous semble pas acceptable. Interprétez les résultats.

c) La fonction *sobolmartinez()* utilise une formule mieux conditionnée pour le calcul des indices de Sobol. En effet, elle est basée sur un calcul de coefficient de corrélation (fonction *cor()* de R), l'indice de Sobol étant justement un coefficient de corrélation. Utilisez cette fonction avec *n*=1000 puis *n*=10000.

d) Une autre technique pour tenter d'améliorer la précision des estimateurs des indices de Sobol est de prendre des échantillons quasi-Monte Carlo (par exemple une suite de Sobol, function *sobol()* du package "randtoolbox") plutôt que Monte Carlo. Effectuez ceci avec *n*=10000 et la fonction *sobol2002()* (sans bootstrap car il n'est plus valide dans ce cas). La seule contrainte est de réussir à générer deux échantillons indépendants avec la suite de Sobol. Il y a une astuce à trouver car les tirages d'une suite de Sobol ne sont pas i.i.d (par contre, on sait que ses colonnes sont orthogonales deux à deux).

e) S'il n'était possible que de faire que quelques centaines d'évaluations de la fonction « robot », le recours à un métamodèle serait nécessaire. On se place dans cette situation :

   a. Générez un plan d'expériences space filling (cf cours et TD 3) de taille 400 sur lequel vous évaluez la fonction « robot ».
   b. Construisez un métamodèle de type krigeage (aussi appelé processus gaussien) avec la fonction *km()* du package « DiceKriging ». Utilisez les paramétrages que vous voulez.
   c. Estimez les indices de Sobol' en utilisant le prédicteur du métamodèle plutôt que la fonction « robot », en utilisant la fonction *sobolmartinez()* avec n=10000. Comparez avec les résultats précédents.
   d. L'intérêt du métamodèle des processus gaussiens est de pouvoir utiliser sa structure d'erreur pour quantifier l'impact de l'approximation faite par le métamodèle sur l'estimation des indices de Sobol'. Utilisez la fonction *sobolGP()* du package « sensitivity » qui permet de calculer les indices de Sobol' sur un métamodèle émanant de la fonction *km()*.