

TD 4 – Analyse de sensibilité, Criblage

Bertrand Iooss
Polytech Nice Sophia

Les informations que l'on recherche dans une phase de criblage sont les variables non influentes, le sens d'influence des variables influentes, les interactions entre variables à effets significatifs, les variables à effets non linéaires, ...

1 Méthode de Morris

Nous allons tester la méthode de Morris (fonction `morris()` codée dans le package R « sensitivity »). Son coût en nombre d'évaluations du modèle est $r(d+1)$ où r est le nombre de répliques OAT (on pourra prendre une valeur de $r = 10$) et d le nombre d'entrées.

Le site web <http://www.sfu.ca/~ssurjano/index.html> contient une panoplie de fonctions tests sur les sujets de l'analyse d'incertitudes et de sensibilité. Nous étudions la « Wing weight function » (présente dans l'onglet « Screening » du site web) qui modélise le poids d'une aile d'avion en pound (1 lb = 0.4536 kg) par la formule analytique :

$$f(\mathbf{x}) = 0.036 S_w^{0.758} W_{fw}^{0.0035} \left(\frac{A}{\cos^2(\Lambda)} \right)^{0.6} q^{0.006} \lambda^{0.04} \left(\frac{100t_c}{\cos(\Lambda)} \right)^{-0.3} (N_z W_{dg})^{0.49} + S_w W_p$$

Ses 10 entrées sont les suivantes :

$S_w \in [150, 200]$	wing area (ft ²)
$W_{fw} \in [220, 300]$	weight of fuel in the wing (lb)
$A \in [6, 10]$	aspect ratio
$\Lambda \in [-10, 10]$	quarter-chord sweep (degrees)
$q \in [16, 45]$	dynamic pressure at cruise (lb/ft ²)
$\lambda \in [0.5, 1]$	taper ratio
$t_c \in [0.08, 0.18]$	aerofoil thickness to chord ratio
$N_z \in [2.5, 6]$	ultimate load factor
$W_{dg} \in [1700, 2500]$	flight design gross weight (lb)
$W_p \in [0.025, 0.08]$	paint weight (lb/ft ²)

Cette fonction est codée dans le fichier R du TP. Les entrées y sont normalisées de telle sorte qu'elle prend en entrée un vecteur de 10 variables, chacune dans $[0, 1]$.

Appliquez la méthode de Morris à ce modèle, visualisez et interprétez les résultats.

2 Outils graphiques pour l'analyse de sensibilité

On étudie le jeu de données nommé « BostonHousing2 » du package « mlbench » de R.

- a) On prend en main le jeu de données en regardant ses attributs. On le range dans une data frame qui contient juste les entrées et la sortie (cmedv) que l'on veut étudier :

crim	per capita crime rate by town
zn	proportion of residential land zoned for lots over 25,000 sq.ft
indus	proportion of non-retail business acres per town
nox	nitric oxides concentration (parts per 10 million)
rm	average number of rooms per dwelling
age	proportion of owner-occupied units built prior to 1940
dis	weighted distances to five Boston employment centres
rad	index of accessibility to radial highways
tax	full-value property-tax rate per USD 10,000
ptratio	pupil-teacher ratio by town
lstat	percentage of lower status of the population
cmedv	Corrected median value of owner-occupied homes in USD 1000's

- b) On effectue les analyses unidimensionnelles et bidimensionnelles, ainsi que l'analyse des corrélations entre variables. Effectuez une première interprétation de la sensibilité de la sortie y (variable cmedv qui est le prix des maisons) par rapport aux autres variables (les entrées). Au vu du graphique, à quoi faut-il faire attention ?

3 Calcul d'indices de sensibilité linéaires (mesures d'importance basées sur la variance)

On utilise le package « sensitivity ».

- a) Sur la « Wing weight function », générez un échantillon d'entrées-sortie de taille 100. Calculez et visualisez les indices SRC (avec leur intervalle de confiance obtenu par bootstrap). Interprétez-les en termes de partage de R^2 . Calculez un indicateur afin de valider ou d'invalider l'hypothèse de linéarité sous-jacente (cf. TP 1, section 5).
- b) A présent, sur le jeu de données « BostonHousing2 », calculez et interprétez les SRC, SRC^2 et R^2 du modèle linéaire sous-jacent entre la sortie et les entrées. Vérifiez que vous obtenez les mêmes résultats que sur la dernière slide du cours 4.
- c) Calculez la métrique de multicollinéarité VIF (fonction *vif()* du package). Qu'en déduisez-vous ? Calculez alors les indices de sensibilité LMG (fonction *lmg()* du package « sensitivity ») et interprétez les par rapport aux SRC^2 .
- d) En réduisant le nombre d'entrées à 3 dans le modèle linéaire ajusté sur les données « BostonHousing2 », comparez les R^2 des modèles linéaires obtenus en sélectionnant ce 3 entrées avec les SRC^2 (les 3 plus grands) à une sélection faite avec les LMG (les 3 plus grands).