MACS - Quantification des incertitudes pour la simulation

TD 9 - Année 2022-2023

1 Indices de Sobol de la fonction Ishigami

On considère le modèle suivant :

$$y(\boldsymbol{X}) = \sin(X_1) + 7\sin(X_2)^2 + 0.1X_3^4\sin(X_1), \quad \boldsymbol{X} = (X_1, X_2, X_3, X_4) \sim \sqcap\left([-\pi, \pi]^4\right).$$

1. Rappeler l'expression des indices de Sobol d'ordre 1, $S_1^{(i)}$, et totaux, $S_T^{(i)}$, caractérisant l'influence des entrées X_i sur la variance de y.

2. Donner la PDF de $\boldsymbol{X}$, $f_{\boldsymbol{X}}$.

3. Calculer $\mathbb{E}\left[\sin(X_i)\right]$, $\mathbb{E}\left[\sin^2(X_i)\right]$, $\mathbb{E}\left[\sin^4(X_i)\right]$, $\mathbb{E}\left[X_i^4\right]$ et $\mathbb{E}\left[X_i^8\right]$.

4. Calculer la moyenne de y.

5. Calculer la variance de y.

6. Calculer les indices $S_1^{(i)}$ et $S_T^{(i)}$.

7. Commenter les résultats précédents.

8. Sans refaire tous les calculs, commenter l'influence d'un augmentation du domaine de définition de $\boldsymbol{X}$ à $[0, 2\pi]^4$.

2 Indices de Sobol pour variables corrélées

On considère la fonction :

$$y : \begin{cases} \mathbb{R}^3 & \rightarrow \mathbb{R} \\ \boldsymbol{x} = (x_1, x_2, x_3) & \mapsto y(\boldsymbol{x}) = \sqrt{2}x_1 + x_2 + \dfrac{x_3}{\sqrt{2}}. \end{cases}$$

On suppose que $\boldsymbol{x}$ est un vecteur gaussien **centré** et de matrice de covariance :

$$\boldsymbol{R} := \begin{bmatrix} 1 & 0 & \rho \\ 0 & 1 & 0 \\ \rho & 0 & 1 \end{bmatrix}, \quad -1 \leq \rho \leq 1.$$

1. Rappeler la densité $f_{\boldsymbol{x}}$ de $\boldsymbol{x}$ en fonction de $\boldsymbol{R}$. Peut-on dire que les variables x_1, x_2, x_3 sont indépendantes statistiquement ?

2. Calculer la moyenne et la variance de $y(\boldsymbol{x})$.

On rappelle que si $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_y)$ et $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_z)$ sont deux vecteurs gaussiens centrés corrélés, tels que $\mathrm{Cov}(\boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{R}_{yz}$, alors :

$$\mathbb{E}\left[\boldsymbol{y}|\boldsymbol{z}\right] = \boldsymbol{R}_{yz}\boldsymbol{R}_z^{-1}\boldsymbol{z}.$$

3. Calculer $\mathbb{E}[y(\boldsymbol{x})|x_1]$, $\mathbb{E}[y(\boldsymbol{x})|x_2]$, $\mathbb{E}[y(\boldsymbol{x})|x_3]$.

4. En déduire les indices de Sobol d'ordre 1 associés à x_1, x_2, x_3, respectivement notés S_1, S_2, S_3.

5. Tracer sur un même graphique l'évolution de ces indices en fonction de $\rho \in [-1, 1]$. Commenter le résultat (en particulier les cas $\rho \in \{-1, 0, 1\}$).

6. Calculer la somme des indices d'ordre 1. Que constatez vous comme différence (surprenante) avec le cas des variables indépendantes (on pourra à nouveau se concentrer sur les valeurs $\rho \in \{-1, 0, 1\}$) ? Comment l'expliquez vous ?

3 Approches spectrales et indices de Sobol

On s'intéresse à l'influence des entrées $\boldsymbol{X} = (X_1, \ldots, X_d)$, que l'on suppose indépendantes statistiquement et de même loi (après renormalisation par exemple), sur la variance de $y(\boldsymbol{X})$. On propose alors de passer par une approche spectrale. Pour cela, à partir d'un nombre limité d'évaluations de y, on suppose avoir identifié une approximation de y sous la forme :

$$y(\boldsymbol{X}) \approx \sum_{\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_d)\in\mathcal{A}} c_{\boldsymbol{\alpha}}\psi_{\alpha_1}(X_1) \times \cdots \times \psi_{\alpha_d}(X_d),$$

où $\{\psi_0, \psi_1, \psi_2, \ldots\}$ est une base orthonormée par rapport à la loi des X_i, telle que :

$$\psi_0(X_i) = 1, \quad \mathbb{E}\left[\psi_m(X_i)\psi_n(X_i)\right] = \delta_{nm}, \quad n, m \geq 0, \quad 1 \leq i \leq d,$$

et où $\mathcal{A}$ est un sous ensemble fini de $\mathbb{N}^d$.

1. Calculer la moyenne de y.

2. Calculer la variance de y.

3. Calculer $\mathbb{E}\left[y(\boldsymbol{X})|X_i\right]$.

4. En déduire $S_1^{(i)} = \mathrm{Var}(\mathbb{E}\left[y(\boldsymbol{X})|X_i\right])/\mathrm{Var}(y(\boldsymbol{X}))$.

5. Expliquer alors comment calculer $S_2^{(i,j)}, \ldots, S_d^{(1,\ldots,d)}$ et $S_T^{(i)}$.