DM 5. Indices de Sobol fonction produit

On considère la fonciton :

$$\begin{array}{cccc} f & : & [-1,1]^3 & \rightarrow & \mathbb{R} \\ & & x=(x_1,x_2,x_3) & \mapsto & y(x)=(0.1+x_1)(1+x_2)(5+x_3) \end{array}$$

,

Question 1 – .
Soit $x=(x_1,x_2,x_3) \sim \mathcal{U}([0,1]^3)$. On a alors comme densité :

$$\begin{aligned} f_x(x) &= \mathbb{1}_{[0,1]^3}(x) \\ &= \mathbb{1}_{[0,1]}(x_1)\mathbb{1}_{[0,1]}(x_2)\mathbb{1}_{[0,1]}(x_3) \end{aligned}$$

Donc par théorème, les coordonnées de x sont indépendantes et peuvent être tiré selon la loi uniforme sur [0,1].

Question 2 – .

Calcul de la valeur moyenne de y (par l'espérance) :

$$\begin{aligned} \mathbb{E}(y(x)) &= \mathbb{E}[(0.1+x_1)(1+x_2)(5+x_3)] \\ &= \mathbb{E}(0.1+x_1)\mathbb{E}(1+x_2)\mathbb{E}(5+x_3) && \text{lemme des coalitions} \\ &= (0.1+\frac{1}{2})(1+\frac{1}{2})(5+\frac{1}{2}) \\ &= (\frac{6}{10})(\frac{3}{2})(\frac{11}{2}) \\ &= \boxed{\frac{99}{20}} \end{aligned}$$

Question 3 – .
Soit a,b $\in \mathbb{R}$. Soit Z et W deux variables aléatoires indépendantes et uniformément distribuée sur [0,1].

Calcul de $\mathbb{V}(a+Z)$:

$$\begin{aligned} \mathbb{V}((a+Z)) &= \mathbb{V}((Z)) \\ &= \frac{(1-0)^2}{12} \\ &= \boxed{\frac{1}{12}} \end{aligned}$$

Calcul de $\mathbb{E}((a+Z)^2)$:

$$\begin{aligned}
\mathbb{E}((a+Z)^2) &= \mathbb{V}(a+Z) + (\mathbb{E}(a+Z))^2 \\
&= \frac{1}{12} + (a+\mathbb{E}(Z))^2 \\
&= \frac{1}{12} + (a+\frac{1}{2})^2 \\
&= \boxed{\frac{1}{12} + (\frac{1+2a}{2})^2}
\end{aligned}$$

Calcul de $\mathbb{V}((a+Z)(b+W))$:

$$\begin{aligned}
\mathbb{V}((a+Z)(b+W)) &= \mathbb{V}((b+W))\mathbb{E}((a+Z)^2) + \mathbb{V}((a+Z))(\mathbb{E}(b+W))^2 \quad \text{indépendance de W et Z} \\
&= \boxed{\frac{\frac{1}{12} + (\frac{1+2a}{2})^2 + (\frac{1+2b}{2})^2}{12}}
\end{aligned}$$

Question 4 – .
Calculons maintenant la variance de y(x) :

$$\begin{aligned}
\mathbb{V}(y(x)) &= \mathbb{V}((0.1+x_1)[(1+x_2)(5+x_3)]) \qquad \text{Or } x_1, x_2 \text{ sont indépendant de } x_3 \\
&= \mathbb{V}((1+x_2)(5+x_3))\mathbb{E}((0.1+x_1)^2) + \mathbb{V}(0.1+x_1)(\mathbb{E}((1+x_2)(5+x_3)))^2 \\
&= \frac{\frac{1}{12} + (\frac{3}{2})^2 + (\frac{11}{2})^2}{12}(\frac{1}{12} + (\frac{1+0.2}{2})^2) + \frac{((\frac{3}{2})^2(\frac{11}{2})^2)}{12} \\
&= \frac{\frac{1}{12} + (\frac{3}{2})^2 + (\frac{11}{2})^2}{12}(\frac{1}{12} + (\frac{1+0.2}{2})^2 + 1) - \frac{1}{12^2} \\
&= \frac{\frac{1}{12} + (\frac{3}{2})^2 + (\frac{11}{2})^2}{12}(\frac{13}{12} + (\frac{12}{20})^2) - \frac{1}{12^2} \\
&= \boxed{\frac{3\cdot 11^2 + 3^3 + 1}{12^2}(\frac{13}{12} + (\frac{12}{20})^2) - \frac{1}{12^2}} \\
&\approx 3.836
\end{aligned}$$

Question 5 – .

Calcul de S_1 :

$$\begin{aligned}
S_1 &= \frac{\mathbb{V}(\mathbb{E}(Y|x_1))}{\mathbb{V}(y(x))} \\
&= \frac{\mathbb{V}((0.1 + x_1)(\frac{33}{4}))}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{33}{4})^2\mathbb{V}((0.1 + x_1))}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{33}{4})^2\frac{1}{12}}{\mathbb{V}(y(x))} \\
&= \boxed{\frac{3 \cdot 11^2}{4^3 \cdot \mathbb{V}(y(x))}} \\
&\approx \frac{5.67188}{\mathbb{V}(y(x))}
\end{aligned}$$

Calcul de S_2 :

$$\begin{aligned}
S_2 &= \frac{\mathbb{V}(\mathbb{E}(Y|x_2))}{\mathbb{V}(y(x))} \\
&= \frac{\mathbb{V}((1 + x_2)(\frac{33}{10}))}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{33}{10})^2\mathbb{V}((1 + x_2))}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{33}{10})^2\frac{1}{12}}{\mathbb{V}(y(x))} \\
&= \boxed{\frac{3 \cdot 11^2}{10^2 \cdot 4 \cdot \mathbb{V}(y(x))}} \\
&\approx \frac{0.90749}{\mathbb{V}(y(x))}
\end{aligned}$$

Calcul de S_3 :

$$\begin{aligned}
S_3 &= \frac{\mathbb{V}(\mathbb{E}(Y|x_3))}{\mathbb{V}(y(x))} \\
&= \frac{\mathbb{V}((5 + x_3)(\frac{3^2 \cdot 2}{20}))}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{3^2}{10})^2\mathbb{V}((5 + x_3))}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{3^2}{10})^2\frac{1}{12}}{\mathbb{V}(y(x))} \\
&= \boxed{\frac{3^3}{10^2 \cdot 4 \cdot \mathbb{V}(y(x))}} \\
&\approx \frac{0.06749}{\mathbb{V}(y(x))}
\end{aligned}$$

Question 6 – .
D'après les calculs ci-dessus, la plus grande part de variance expliqué par une seule variable est

celle correspondant à $\boxed{\text{l'indice de Sobol } S_1}$. Cela peut s'expliquer par la présence du coefficient qu'on ajoute aux variables aléatoires avant de les multiplier. En effet pour x_3, comme $5 \geq\geq 1$, on voit que la variable x_3 aura peut d'impact sur le résultat de $5+x_3$ (ordre de grandeur) donc celle qui engendrera le plus de variance sera la variable dont la somme sera représentée au mieux par la valeur trié, donc x_1.

Question 7 – .
Calcul de T_1 :

$$\begin{aligned}
T_1 &= \frac{\mathbb{E}(\mathbb{V}(Y|x_{(-1)}))}{\mathbb{V}(y(x))} \\
&= \frac{\mathbb{E}[((1+x_2)(5+x_3))^2\mathbb{V}(0.1+x_1)]}{\mathbb{V}(y(x))} \\
&= \frac{\frac{1}{12}\mathbb{E}((1+x_2)^2)\mathbb{E}((5+x_3)^2)}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{1}{12}+(\frac{3^2}{2^2}))(\frac{1}{12}+(\frac{11^2}{2^2}))}{12\mathbb{V}(y(x))} \\
&= \frac{(\frac{3^3+1}{12})(\frac{3\cdot 11^2+1}{12})}{12\mathbb{V}(y(x))} \\
&\approx \frac{2.52777}{\mathbb{V}(y(x))}
\end{aligned}$$

Calcul de T_2 :

$$\begin{aligned}
T_2 &= \frac{\mathbb{E}(\mathbb{V}(Y|x_{(-2)}))}{\mathbb{V}(y(x))} \\
&= \frac{\mathbb{E}[((0.1+x_1)(5+x_3))^2\mathbb{V}(1+x_2)]}{\mathbb{V}(y(x))} \\
&= \frac{\frac{1}{12}\mathbb{E}((0.1+x_1)^2)\mathbb{E}((5+x_3)^2)}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{1}{12}+(\frac{12^2}{20^2}))(\frac{3\cdot 11^2+1}{12}))}{12\mathbb{V}(y(x))} \\
&\approx \frac{2.52777}{\mathbb{V}(y(x))}
\end{aligned}$$

Calcul de T_3 :

$$\begin{aligned}
T_3 &= \frac{\mathbb{E}(\mathbb{V}(Y|x_{(-3)}))}{\mathbb{V}(y(x))} \\
&= \frac{\mathbb{E}[((0.1+x_1)(1+x_2))^2\mathbb{V}(5+x_3)]}{\mathbb{V}(y(x))} \\
&= \frac{\frac{1}{12}\mathbb{E}((0.1+x_1)^2)\mathbb{E}((1+x_2)^2)}{\mathbb{V}(y(x))} \\
&= \frac{(\frac{1}{12}+(\frac{12^2}{20^2}))(\frac{3^3+1}{12}))}{12\mathbb{V}(y(x))} \\
&\approx \frac{0.19444}{\mathbb{V}(y(x))}
\end{aligned}$$

Question 8 – .
On sait que si pour un indice i, $S_i \approx T_i$ alors l'influence de X_i sur y(x) sera négligeable. on pourra

alors remplacer x_i par sa valeur moyenne sans risquer de changer considérablement y(x). D'après les calculs précédents, on a $S_3 \approx T_3$ donc on peut considérer que la variable x_3 peut être remplacée par sa moyenne $\frac{1}{2}$ sans avoir beaucoup d'interaction sur le résultat y(x).