

Lois asymptotiques des estimateurs des indices de Sobol

Application de la méthode delta

Rapport technique

Client : EDF Adresse : Affaire : PMFRE-01116 Commande n° :
Titre : Lois asymptotiques des estimateurs des indices de Sobol Référence : RT-PMFRE-01116-001C Accessibilité : Confidentiel Mots clés : Indices de Sobol, estimateurs, loi asymptotique, delta methode

Prédiffusion

Diffusion

Bertrand Iooss, Michael Baudin	EDF
--------------------------------	-----

Production

Rédacteur	Vérificateur	Approbateur
A. Dumas 	J. Schueller	T. Yalamas

Suivi des indices

Indice	Date	Évolutions
A	1/12/2017	Création
B	23/01/2018	Intégration notation variable centrée et remarques suite à la réunion du 1/12/2017.
C	13/02/2018	
		Ajout de la section validation.

Table des matières

1	Introduction	4
2	Estimateurs de Saltelli	5
2.1	Estimateur du premier ordre	5
2.2	Estimateur de l'ordre total	6
3	Estimateurs de Jansen	8
3.1	Estimateur du premier ordre	8
3.2	Estimateur de l'ordre total	9
4	Estimateurs de Mauntz-Kucherenko	11
4.1	Estimateur du premier ordre	11
4.2	Estimateur de l'ordre total	12
5	Estimateurs de Martinez	14
5.1	Estimateur du premier ordre	14
5.2	Estimateur de l'ordre total	15
6	Indices agrégés	18
7	Implémentation et validation	20
7.1	Distributions théoriques et empiriques	20
7.1.1	Fonction G-Sobol	20
7.1.2	Fonction Ishigami	25
7.1.3	Fonction poutre	29
7.1.4	Fonction agrégée	31
7.2	Distributions théoriques moyennes et empiriques	35
7.2.1	Fonction G-Sobol	35
7.2.2	Fonction Ishigami	39
7.2.3	Fonction poutre	43
7.2.4	Fonction agrégée	45
7.3	Anomalie dans l'estimateur de Saltelli	49
	Références	50

1 Introduction

L'objectif est de définir théoriquement la variance des estimateurs des indices de Sobol présent dans OpenTURNS (Saltelli, Jansen, Martinez et Mauntz-Kucherenko). La technique utilisée par Janon *et al.* (2014), consistant à utiliser la *méthode delta*, est appliquée pour déterminer la loi asymptotique de ces estimateurs.

Les notations d'OpenTURNS sont reprises pour la définitions des échantillons indépendants :

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n_X} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n_X} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{1,2} & \cdots & a_{n,n_X} \end{pmatrix}, B = \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n_X} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n_X} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n,1} & b_{1,2} & \cdots & b_{n,n_X} \end{pmatrix} \quad (1)$$

ainsi que la matrice de mélange suivante :

$$E^i = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & b_{1,i} & \cdots & a_{1,n_X} \\ a_{2,1} & a_{2,2} & \cdots & b_{2,i} & \cdots & a_{2,n_X} \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{1,2} & \cdots & b_{n,i} & \cdots & a_{n,n_X} \end{pmatrix} \quad (2)$$

On note également :

$$Y^A = G(A) \quad , \quad Y_k^A = G(A_k) \quad (3)$$

$$Y^B = G(B) \quad , \quad Y_k^B = G(B_k) \quad (4)$$

$$Y^E = G(E) \quad , \quad Y_k^E = G(E_k) \quad (5)$$

où G est la fonction du modèle physique. Les échantillons centrés seront annotés d'un c :

$$^c Y^A = Y_i^A - \mathbb{E} [Y^A] \quad (6)$$

Dans OpenTURNS, l'échantillon de sortie est centrée globalement par rapport à l'ensemble de tous les tirages et mélange.

2 Estimateurs de Saltelli

2.1 Estimateur du premier ordre

L'estimateur du premier ordre utilisé dans OpenTURNS est donné par la formule suivante (Saltelli, 2002) :

$$S = \frac{\text{Cov}[cY^B, cY^E]}{\text{Var}[cY^A]} \quad (7)$$

$$= \frac{\frac{1}{n} \sum_{k=1}^n cY_k^B cY_k^E - \left(\frac{1}{n} \sum_{k=1}^n cY_k^B\right) \left(\frac{1}{n} \sum_{k=1}^n cY_k^E\right)}{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2 - \left(\frac{1}{n} \sum_{k=1}^n cY_k^A\right)^2} \quad (8)$$

Attention : Différence avec OpenTURNS : $G_0^2 = \left(\frac{1}{n} \sum_{k=1}^n Y_k^A\right)^2$ utilisé (dans la documentation) et non $\left(\frac{1}{n} \sum_{k=1}^n Y_k^B\right) \left(\frac{1}{n} \sum_{k=1}^n Y_k^A\right)$ dans le membre de droite au numérateur. Dans Saltelli (2002), la formule préconisée est bien cette dernière équation. Dans le code il y a une erreur, il est calculé ainsi : $\left(\frac{1}{n} \sum_{k=1}^n Y_k^B Y_k^A\right)$.

Simplification : Comme les échantillons sont centrées par rapport à leurs moyennes respectives, le numérateur et le dénominateur se simplifient et on obtient l'estimateur suivant :

$$S = \frac{\frac{1}{n} \sum_{k=1}^n cY_k^B cY_k^E}{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2} \quad (9)$$

Démonstration du calcul de la variance de l'estimateur :

On pose alors :

$$U_i = \left(cY_i^B cY_i^E, (cY_i^A)^2\right)^T \quad (10)$$

et

$$\Psi_S(x, y) = \frac{x}{y} \quad (11)$$

ce qui implique :

$$S_N = \Psi_S(\bar{U}_N) \quad (12)$$

Le théorème centrale limite donne :

$$\sqrt{N} (\bar{U}_N - \mu) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_2(0, \Gamma) \quad (13)$$

avec Γ la matrice de covariance de U_1 et :

$$\mu = \begin{pmatrix} \text{Cov}[Y^B, Y^E] \\ \text{Var}[Y^A] \end{pmatrix} \quad (14)$$

La méthode delta implique :

$$\sqrt{N} (S_N - S) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, g^T \Gamma g) \quad (15)$$

où $g = \nabla \Psi_S(\mu)$. Pour tout x, y tel que $y \neq 0$, on a :

$$\nabla \Psi_S(x, y) = \left(\frac{1}{y}, \frac{-x}{y^2} \right)^T \quad (16)$$

On obtient, en utilisant l'équation (7) :

$$g = \left(\frac{1}{\text{Var}[Y^A]}, -\frac{S}{\text{Var}[Y^A]} \right)^T \quad (17)$$

La variance s'écrit alors comme suit :

$$g^T \Gamma g = \frac{\text{Var}[x] + S^2 \text{Var}[y] - 2S \text{Cov}[x, y]}{\text{Var}[Y^A]^2} \quad (18)$$

$$= \frac{\text{Var}[x] + \text{Var}[Sy] - 2\text{Cov}[x, Sy]}{\text{Var}[Y^A]^2} \quad (19)$$

$$= \frac{\text{Var}[x - Sy]}{\text{Var}[Y^A]^2} \quad (20)$$

$$= \frac{\text{Var}[cY^{Bc}Y^E - S(cY^A)^2]}{\text{Var}[Y^A]^2} \quad (21)$$

2.2 Estimateur de l'ordre total

L'estimateur de l'ordre total dans OpenTURNS est donné par la formule suivante :

$$S^T = 1 - \frac{\frac{1}{n} \sum_{k=1}^n cY_k^{Ac}Y_k^E - \left(\frac{1}{n} \sum_{k=1}^n cY_k^A \right)^2}{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2 - \left(\frac{1}{n} \sum_{k=1}^n cY_k^A \right)^2} \quad (22)$$

Attention : Différence avec OpenTURNS et Saltelli (2002) :

- $\frac{1}{n} \sum_{k=1}^n Y_k^B Y_k^C$ est utilisé et non $\frac{1}{n} \sum_{k=1}^n Y_k^A Y_k^E$ au numérateur.

Simplification : Comme pour l'estimateur du premier ordre, l'estimateur se simplifie :

$$S^T = 1 - \frac{\frac{1}{n} \sum_{k=1}^n cY_k^{Ac}Y_k^E}{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2} \quad (23)$$

Démonstration du calcul de la variance de l'estimateur :

On pose alors :

$$U_i = \left(cY_i^{Ac}Y_i^E, (cY_i^A)^2 \right)^T \quad (24)$$

et

$$\Psi_S(x, y) = 1 - \frac{x}{y} \quad (25)$$

ce qui implique :

$$S_N^T = \Psi_S(\bar{U}_N) \quad (26)$$

Le théorème centrale limite donne :

$$\sqrt{N} (\bar{U}_N - \mu) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_2(0, \Gamma) \quad (27)$$

avec Γ la matrice de covariance de U_1 et :

$$\mu = \begin{pmatrix} \text{Cov}[Y^A, Y^E] \\ \text{Var}[Y^A] \end{pmatrix} \quad (28)$$

La méthode delta implique :

$$\sqrt{N} (S_N^T - S^T) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, g^T \Gamma g) \quad (29)$$

où $g = \nabla \Psi_S(\mu)$. Pour tout x, y tel que $y \neq 0$, on a :

$$\nabla \Psi_S(x, y) = \left(\frac{-1}{y}, \frac{x}{y^2} \right)^T \quad (30)$$

On pose :

$$S^{-X} = \frac{\text{Cov}[Y^A, Y^E]}{\text{Var}[Y^A]} \quad (31)$$

On obtient :

$$g = \left(\frac{-1}{\text{Var}[Y^A]}, \frac{\text{Cov}[Y^A, Y^E]}{\text{Var}[Y^A]^2} \right)^T \quad (32)$$

$$= \left(\frac{-1}{\text{Var}[Y^A]}, \frac{S^{-X}}{\text{Var}[Y^A]} \right)^T \quad (33)$$

La variance s'écrit alors comme pour l'ordre total :

$$g^T \Gamma g = \frac{\text{Var}[x] + (S^{-X})^2 \text{Var}[y] - 2S^{-X} \text{Cov}[x, y]}{\text{Var}[Y^A]^2} \quad (34)$$

$$= \frac{\text{Var}[x] + \text{Var}[S^{-X}y] - 2\text{Cov}[x, S^{-X}y]}{\text{Var}[Y^A]^2} \quad (35)$$

$$= \frac{\text{Var}[x - S^{-X}y]}{\text{Var}[Y^A]^2} \quad (36)$$

$$= \frac{\text{Var}\left[\sum_i c_i Y_i^A c_i Y_i^E - S^{-X} \left(\sum_i c_i Y_i^A\right)^2\right]}{\text{Var}[Y^A]^2} \quad (37)$$

3 Estimateurs de Jansen

3.1 Estimateur du premier ordre

L'estimateur du premier ordre dans OpenTURNS est donné par la formule suivante :

$$S = \frac{\text{Var} [cY^A] - \frac{1}{2n} \sum_{k=1}^n (cY_k^E - cY_k^B)^2}{\text{Var} [cY^A]} \quad (38)$$

$$S = 1 - \frac{\frac{1}{2n} \sum_{k=1}^n (cY_k^E - cY_k^B)^2}{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2} \quad (39)$$

Démonstration du calcul de la variance de l'estimateur :

On pose :

$$U_i = \left((cY_i^E - cY_i^B)^2, (cY_i^A)^2 \right)^T \quad (40)$$

et

$$\Psi_S(x, y) = 1 - \frac{0.5x}{y} \quad (41)$$

ce qui implique :

$$S_N = \Psi_S(\bar{U}_N) \quad (42)$$

Le théorème centrale limite donne :

$$\sqrt{N} (\bar{U}_N - \mu) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_2(0, \Gamma) \quad (43)$$

avec Γ la matrice de covariance de U_1 et :

$$\mu = \begin{pmatrix} \text{Var} [Y^E - Y^B] \\ \text{Var} [Y^A] \end{pmatrix} \quad (44)$$

La méthode delta implique :

$$\sqrt{N} (S_N - S) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, g^T \Gamma g) \quad (45)$$

où $g = \nabla \Psi_S(\mu)$. Pour tout x, y tel que $y \neq 0$, on a :

$$\nabla \Psi_S(x, y) = \left(\frac{-0.5}{y}, \frac{0.5x}{y^2} \right)^T \quad (46)$$

On pose :

$$J = \frac{\text{Var} [Y^E - Y^B]}{\text{Var} [Y^A]} \quad (47)$$

On obtient :

$$g = \left(-\frac{0.5}{\text{Var}[Y^A]}, \frac{0.5 \text{Var}[Y^E - Y^B]}{\text{Var}[Y^A]^2} \right)^T \quad (48)$$

$$= \left(-\frac{0.5}{\text{Var}[Y^A]}, \frac{0.5J}{\text{Var}[Y^A]} \right)^T \quad (49)$$

La variance s'écrit alors comme suit :

$$N \times g^T \Gamma g = \frac{(0.25J^2 \text{Var}[y] - 0.5J \text{Cov}[x; y] + 0.25 \text{Var}[x])}{\text{Var}[Y^A]^2} \quad (50)$$

$$= \frac{0.5^2 (\text{Var}[x] + \text{Var}[Jy] - 2 \text{Cov}[x, Jy])}{\text{Var}[Y^A]^2} \quad (51)$$

$$= \frac{1}{4} \frac{\text{Var}[x - Jy]}{\text{Var}[Y^A]^2} \quad (52)$$

$$= \frac{1}{4} \frac{\text{Var}[(cY_i^E - cY_i^B)^2 - J(cY_i^A)^2]}{\text{Var}[Y^A]^2} \quad (53)$$

3.2 Estimateur de l'ordre total

L'estimateur de l'ordre total est donné par la formule suivante :

$$S^T = \frac{\frac{1}{2n} \sum_{k=1}^n (cY_k^A - cY_k^E)^2}{\text{Var}[cY^A]} \quad (54)$$

$$= \frac{\frac{1}{2n} \sum_{k=1}^n (cY_k^A - cY_k^E)^2}{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2} \quad (55)$$

Démonstration du calcul de la variance de l'estimateur :

On pose :

$$U_i = \left((cY_i^A - cY_i^E)^2, (cY_i^A)^2 \right)^T \quad (56)$$

et

$$\Psi_S(x, y) = \frac{0.5x}{y} \quad (57)$$

ce qui implique :

$$S_N^T = \Psi_S(\bar{U}_N) \quad (58)$$

Le théorème centrale limite donne :

$$\sqrt{N} (\bar{U}_N - \mu) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_2(0, \Gamma) \quad (59)$$

avec Γ la matrice de covariance de U_1 et :

$$\mu = \begin{pmatrix} \text{Var}[Y^A - Y^E] \\ \text{Var}[Y^A] \end{pmatrix} \quad (60)$$

La méthode delta implique :

$$\sqrt{N} (S_N - S) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, g^T \Gamma g) \quad (61)$$

où $g = \nabla \Psi_S(\mu)$. Pour tout x, y tel que $y \neq 0$, on a :

$$\nabla \Psi_S(x, y) = \left(\frac{0.5}{y}, \frac{-0.5x}{y^2} \right)^T \quad (62)$$

On pose :

$$J^{-X} = \frac{\text{Var}[Y^A - Y^E]}{\text{Var}[Y^A]} \quad (63)$$

On obtient :

$$g = \left(\frac{0.5}{\text{Var}[Y^A]}, -\frac{0.5 \text{Var}[Y^A - Y^E]}{\text{Var}[Y^A]^2} \right)^T \quad (64)$$

$$= \left(\frac{0.5}{\text{Var}[Y^A]}, -\frac{0.5 J^{-X}}{\text{Var}[Y^A]} \right)^T \quad (65)$$

La variance s'écrit alors comme suit :

$$N \times g^T \Gamma g = \frac{(0.25(J^{-X})^2 \text{Var}[y] - 0.5 J^{-X} \text{Cov}[x; y] + 0.25 \text{Var}[x])}{\text{Var}[Y^A]^2} \quad (66)$$

$$= \frac{0.5^2 (\text{Var}[x] + \text{Var}[J^{-X}y] - 2 \text{Cov}[x, J^{-X}y])}{\text{Var}[Y^A]^2} \quad (67)$$

$$= \frac{1}{4} \frac{\text{Var}[x - J^{-X}y]}{\text{Var}[Y^A]^2} \quad (68)$$

$$= \frac{1}{4} \frac{\text{Var}[(cY_i^A - cY_i^E)^2 - J^{-X} (cY_i^A)^2]}{\text{Var}[Y^A]^2} \quad (69)$$

4 Estimateurs de Mauntz-Kucherenko

4.1 Estimateur du premier ordre

L'estimateur du premier ordre dans OpenTURNS est donné par la formule suivante :

$$S = \frac{\frac{1}{n} \sum_{k=1}^n cY_k^B (cY_k^E - cY_k^A)}{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2} \quad (70)$$

Démonstration du calcul de la variance de l'estimateur :

On pose :

$$U_i = \left(cY_i^B (cY_i^E - cY_i^A), (cY_i^A)^2 \right)^T \quad (71)$$

et

$$\Psi_S(x, y) = \frac{x}{y} \quad (72)$$

ce qui implique :

$$S_N = \Psi_S(\bar{U}_N) \quad (73)$$

Le théorème centrale limite donne :

$$\sqrt{N} (\bar{U}_N - \mu) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_2(0, \Gamma) \quad (74)$$

avec Γ la matrice de covariance de U_1 et :

$$\mu = \begin{pmatrix} \text{Cov}[Y^B, Y^E - Y^A] \\ \text{Var}[Y^A] \end{pmatrix} \quad (75)$$

La méthode delta implique :

$$\sqrt{N} (S_N - S) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, g^T \Gamma g) \quad (76)$$

où $g = \nabla \Psi_S(\mu)$. Pour tout x, y tel que $y \neq 0$, on a :

$$\nabla \Psi_S(x, y) = \left(\frac{1}{y}, \frac{-x}{y^2} \right)^T \quad (77)$$

On pose :

$$M = \frac{\text{Cov}[Y^B, Y^E - Y^A]}{\text{Var}[Y^A]} \quad (78)$$

On obtient :

$$g = \left(\frac{1}{\text{Var}[Y^A]}, -\frac{\text{Cov}[Y^B, Y^E - Y^A]}{\text{Var}[Y^A]^2} \right)^T \quad (79)$$

$$= \left(\frac{1}{\text{Var}[Y^A]}, -\frac{M}{\text{Var}[Y^A]} \right)^T \quad (80)$$

La variance s'écrit alors comme suit :

$$N \times g^T \Gamma g = \frac{-M \text{Cov}[x, y] + M (M \text{Var}[y] - \text{Cov}[x, y]) + \text{Var}[x]}{\text{Var}[Y^A]^2} \quad (81)$$

$$= \frac{\text{Var}[x - My]}{\text{Var}[Y^A]^2} \quad (82)$$

$$= \frac{\text{Var}\left[cY_i^B (cY_i^E - cY_i^A) - M (cY_i^A)^2\right]}{\text{Var}[Y^A]^2} \quad (83)$$

$$(84)$$

4.2 Estimateur de l'ordre total

L'estimateur de l'ordre total dans OpenTURNS est donné par la formule suivante :

$$S^T = \frac{\text{Var}[Y^A] + \frac{1}{n} \sum_{k=1}^n cY_k^A (cY_k^A - cY_k^E)}{\text{Var}[cY^A]} \quad (85)$$

$$= 1 - \frac{\frac{1}{n} \sum_{k=1}^n cY_k^A (cY_k^A - cY_k^E)}{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2} \quad (86)$$

Attention : Différence entre le code OpenTURNS et la documentation : le numérateur dans la documentation est $\frac{1}{n} \sum_{k=1}^n Y_k^A (Y_k^A - Y_k^E)$ tandis que dans le code, on $\text{Var}[y^A] + \frac{1}{n} \sum_{k=1}^n Y_k^A (Y_k^A - Y_k^E)$.

Démonstration du calcul de la variance de l'estimateur :

On pose :

$$U_i = \left(cY_i^A (cY_i^A - cY_i^E), (cY_i^A)^2 \right)^T \quad (87)$$

et

$$\Psi_S(x, y) = 1 - \frac{x}{y} \quad (88)$$

ce qui implique :

$$S_N^T = \Psi_S(\bar{U}_N) \quad (89)$$

Le théorème centrale limite donne :

$$\sqrt{N} (\bar{U}_N - \mu) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_2(0, \Gamma) \quad (90)$$

avec Γ la matrice de covariance de U_1 et :

$$\mu = \begin{pmatrix} \text{Cov} [Y^A, Y^A - Y^E] \\ \text{Var} [Y^A] \end{pmatrix} \quad (91)$$

La méthode delta implique :

$$\sqrt{N} (S_N^T - S^T) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1 (0, g^T \Gamma g) \quad (92)$$

où $g = \nabla \Psi_S(\mu)$. Pour tout x, y tel que $y \neq 0$, on a :

$$\nabla \Psi_S(x, y) = \left(\frac{-1}{y}, \frac{x}{y^2} \right)^T \quad (93)$$

On pose :

$$M^{-X} = \frac{\text{Cov} [Y^A, Y^A - Y^E]}{\text{Var} [Y^A]} \quad (94)$$

On obtient :

$$g = \left(\frac{-1}{\text{Var} [Y^A]}, \frac{\text{Cov} [Y^A, Y^A - Y^E]}{\text{Var} [Y^A]^2} \right)^T \quad (95)$$

$$= \left(\frac{-1}{\text{Var} [Y^A]}, \frac{M^{-X}}{\text{Var} [Y^A]} \right)^T \quad (96)$$

La variance s'écrit alors comme pour l'ordre total :

$$N \times g^T \Gamma g = \frac{\text{Var} [x] + (M^{-X})^2 \text{Var} [y] - 2M^{-X} \text{Cov} [x, y]}{\text{Var} [Y^A]^2} \quad (97)$$

$$= \frac{\text{Var} [x] + \text{Var} [M^{-X} y] - 2\text{Cov} [x, M^{-X} y]}{\text{Var} [Y^A]^2} \quad (98)$$

$$= \frac{\text{Var} [x - M^{-X} y]}{\text{Var} [Y^A]^2} \quad (99)$$

$$= \frac{\text{Var} \left[c_{Y_i^A} (c_{Y_i^A} - c_{Y_i^E}) - M^{-X} (c_{Y_i^A})^2 \right]}{\text{Var} [Y^A]^2} \quad (100)$$

5 Estimateurs de Martinez

5.1 Estimateur du premier ordre

L'estimateur du premier ordre dans OpenTURNS est donné par la formule suivante :

$$S = \rho_n \left({}^cY_k^B, {}^cY_k^E \right) \quad (101)$$

$$= \frac{\text{Cov} \left[{}^cY^B, {}^cY^E \right]}{\sqrt{\text{Var} \left[{}^cY^B \right] \text{Var} \left[{}^cY^E \right]}} \quad (102)$$

$$= \frac{\frac{1}{n} \sum_{k=1}^n \left({}^cY_k^B {}^cY_k^E \right)}{\sqrt{\frac{1}{n} \sum_{k=1}^n \left({}^cY_k^B \right)^2 \frac{1}{n} \sum_{k=1}^n \left({}^cY_k^E \right)^2}} \quad (103)$$

Démonstration

On pose :

$$U_i = \left(\left({}^cY_i^B {}^cY_i^E \right), \left({}^cY_i^B \right)^2, \left({}^cY_i^E \right)^2 \right)^T \quad (104)$$

et

$$\Psi_S(x, y, z) = \frac{x}{\sqrt{y \times z}} \quad (105)$$

ce qui implique :

$$S_N = \Psi_S(\bar{U}_N) \quad (106)$$

Le théorème centrale limite donne :

$$\sqrt{N} \left(\bar{U}_N - \mu \right) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_3(0, \Gamma) \quad (107)$$

avec Γ la matrice de covariance de U_i et :

$$\mu = \begin{pmatrix} \text{Cov} \left[Y^B, Y^E \right] \\ \text{Var} \left[Y^B \right] \\ \text{Var} \left[Y^E \right] \end{pmatrix} \quad (108)$$

La méthode delta implique :

$$\sqrt{N} (S_N - S) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, g^T \Gamma g) \quad (109)$$

où $g = \nabla \Psi_S(\mu)$. Pour tout x, y, z tel que $y, z \neq 0$, on a :

$$\nabla \Psi_S(x, y, z) = \left(\frac{1}{\sqrt{yz}}, -\frac{x}{2y\sqrt{yz}}, -\frac{x}{2z\sqrt{yz}} \right)^T \quad (110)$$

On pose :

$$M = \frac{\text{Cov}[Y^B, Y^E]}{\sqrt{\text{Var}[Y^B] \text{Var}[Y^E]}} \quad (111)$$

On obtient :

$$g = \left(\frac{1}{\sqrt{\text{Var}[Y^B] \text{Var}[Y^E]}}, -\frac{M}{2\text{Var}[Y^B]}, -\frac{M}{2\text{Var}[Y^E]} \right)^T \quad (112)$$

La variance s'écrit alors comme suit :

$$N \times g^T \Gamma g = \frac{2M^2 \text{Cov}[y, z]}{4\text{Var}[Y^B] \text{Var}[Y^E]} + \frac{M^2 \text{Var}[z]}{4\text{Var}[Y^E]^2} + \frac{M^2 \text{Var}[y]}{4\text{Var}[Y^B]^2} \\ - \frac{\text{Cov}[Y^B, Y^E] \text{Cov}[x; y]}{\text{Var}[Y^B]^2 \text{Var}[Y^E]} - \frac{\text{Cov}[Y^B, Y^E] \text{Cov}[x, z]}{\text{Var}[Y^B] \text{Var}[Y^E]^2} \quad (113)$$

$$+ \frac{\text{Var}[x]}{\text{Var}[Y^B] \text{Var}[Y^E]} \quad (114)$$

$$= \frac{M^2 \left(2\text{Var}[Y^B] \text{Var}[Y^E] \text{Cov}[y, z] + \text{Var}[Y^B]^2 \text{Var}[z] + \text{Var}[Y^E]^2 \text{Var}[y] \right)}{4\text{Var}[Y^B]^2 \text{Var}[Y^E]^2} \\ - \frac{\text{Cov}[Y^B, Y^E] \text{Var}[Y^E] \text{Cov}[x; y] + \text{Cov}[Y^B, Y^E] \text{Var}[Y^B] \text{Cov}[x, z]}{\text{Var}[Y^B]^2 \text{Var}[Y^E]^2} \quad (115)$$

$$+ \frac{\text{Var}[x]}{\text{Var}[Y^B] \text{Var}[Y^E]} \quad (116)$$

$$= \frac{M^2 \left(\text{Var}[z \text{Var}[Y^B] + y \text{Var}[Y^E]] \right)}{4\text{Var}[Y^B]^2 \text{Var}[Y^E]^2} \\ - \frac{\text{Cov}[Y^B, Y^E] \text{Cov}[x; y \text{Var}[Y^E] + z \text{Var}[Y^B]]}{\text{Var}[Y^B]^2 \text{Var}[Y^E]^2} \quad (117)$$

$$+ \frac{\text{Var}[x]}{\text{Var}[Y^B] \text{Var}[Y^E]} \quad (118)$$

5.2 Estimateur de l'ordre total

L'estimateur de l'ordre total dans OpenTURNS est donné par la formule suivante :

$$S = 1 - \rho_n(cY_k^A, cY_k^E) \quad (119)$$

$$= 1 - \frac{\text{Cov}[cY^A, cY^E]}{\sqrt{\text{Var}[cY^A] \text{Var}[cY^E]}} \quad (120)$$

$$= 1 - \frac{\frac{1}{n} \sum_{k=1}^n (cY_k^A cY_k^E)}{\sqrt{\frac{1}{n} \sum_{k=1}^n (cY_k^A)^2 \frac{1}{n} \sum_{k=1}^n (cY_k^E)^2}} \quad (121)$$

Démonstration

On pose alors :

$$U_i = \left({}^cY_i^A {}^cY_i^E, \left({}^cY_i^A \right)^2, \left({}^cY_i^E \right)^2 \right)^T \quad (122)$$

et

$$\Psi_S(x, y, z) = 1 - \frac{x}{\sqrt{y \times z}} \quad (123)$$

ce qui implique :

$$S_N = \Psi_S(\bar{U}_N) \quad (124)$$

Le théorème centrale limite donne :

$$\sqrt{N} (\bar{U}_N - \mu) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_3(0, \Gamma) \quad (125)$$

avec Γ la matrice de covariance de U_i et :

$$\mu = \begin{pmatrix} \text{Cov}[Y^A, Y^E] \\ \text{Var}[Y^A] \\ \text{Var}[Y^E] \end{pmatrix} \quad (126)$$

La méthode delta implique :

$$\sqrt{N} (S_N - S) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_1(0, g^T \Gamma g) \quad (127)$$

où $g = \nabla \Psi_S(\mu)$. Pour tout x, y, z tel que $z \neq y^2$, on a :

$$\nabla \Psi_S(x, y, z) = \left(-\frac{1}{\sqrt{yz}}, \frac{x}{2y\sqrt{yz}}, \frac{x}{2z\sqrt{yz}} \right)^T \quad (128)$$

On pose :

$$M = \frac{\text{Cov}[Y^A, Y^E]}{\sqrt{\text{Var}[Y^A] \text{Var}[Y^E]}} \quad (129)$$

On obtient :

$$g = \left(-\frac{1}{\sqrt{\text{Var}[Y^A] \text{Var}[Y^E]}}, \frac{M}{2\text{Var}[Y^A]}, \frac{M}{2\text{Var}[Y^E]} \right)^T \quad (130)$$

La variance s'écrit alors comme pour l'ordre 1 mais avec l'échantillon Y^A . On a :

$$N \times g^T \Gamma g = \frac{2M^2 \text{Cov}[y, z]}{4\text{Var}[Y^A] \text{Var}[Y^E]} + \frac{M^2 \text{Var}[z]}{4\text{Var}[Y^E]^2} + \frac{M^2 \text{Var}[y]}{4\text{Var}[Y^A]^2} \\ - \frac{\text{Cov}[Y^A, Y^E] \text{Cov}[x; y]}{\text{Var}[Y^A]^2 \text{Var}[Y^E]} - \frac{\text{Cov}[Y^A, Y^E] \text{Cov}[x, z]}{\text{Var}[Y^A] \text{Var}[Y^E]^2} \quad (131)$$

$$+ \frac{\text{Var}[x]}{\text{Var}[Y^A] \text{Var}[Y^E]} \quad (132)$$

$$= \frac{M^2 \left(2\text{Var}[Y^A] \text{Var}[Y^E] \text{Cov}[y, z] + \text{Var}[Y^A]^2 \text{Var}[z] + \text{Var}[Y^E]^2 \text{Var}[y] \right)}{4\text{Var}[Y^A]^2 \text{Var}[Y^E]^2} \\ - \frac{\text{Cov}[Y^A, Y^E] \text{Var}[Y^E] \text{Cov}[x; y] + \text{Cov}[Y^A, Y^E] \text{Var}[Y^A] \text{Cov}[x, z]}{\text{Var}[Y^A]^2 \text{Var}[Y^E]^2} \quad (133)$$

$$+ \frac{\text{Var}[x]}{\text{Var}[Y^A] \text{Var}[Y^E]} \quad (134)$$

$$= \frac{M^2 \left(\text{Var}[z \text{Var}[Y^A]] + \text{Var}[y \text{Var}[Y^E]] \right)}{4\text{Var}[Y^A]^2 \text{Var}[Y^E]^2} \\ - \frac{\text{Cov}[Y^A, Y^E] (\text{Cov}[x; y \text{Var}[Y^E]] + z \text{Var}[Y^A])}{\text{Var}[Y^A] \text{Var}[Y^E]^2} \quad (135)$$

$$+ \frac{\text{Var}[x]}{\text{Var}[Y^A] \text{Var}[Y^E]} \quad (136)$$

6 Indices agrégés

Lorsque le nombre de sorties est supérieure à 1, les indices de Sobol agrégés sont évalués. La distribution asymptotique doit donc correspondre à la distribution des indices agrégés. Soit d la dimension de sortie, les expressions des indices agrégés d'ordre 1, noté $S_{i,agg}$, et total, noté $S_{Ti,agg}$, de la i^e variable d'entrée sont données dans les équations 137 et 138.

$$S_{i,agg} = \frac{\sum_{k=1}^d V_{I,ik}}{\sum_{k=1}^d \text{Var}[Y_k]} \quad (137)$$

$$S_{Ti,agg} = \frac{\sum_{k=1}^d V_{TI,ik}}{\sum_{k=1}^d \text{Var}[Y_k]} \quad (138)$$

Afin de pouvoir calculer la variance des indices agrégés directement, la fonction Ψ de chaque estimateur est adaptée pour prendre en compte le nombre de sorties. Elles sont différentes suivant les estimateurs :

- Saltelli et Mauntz-Kucherenko pour la i^e variable d'entrée :

$$\Psi_S = \frac{\sum_{k=1}^d x_k}{\sum_{k=1}^d y_k} \quad (139)$$

$$\Psi_{ST} = 1 - \frac{\sum_{k=1}^d x_k}{\sum_{k=1}^d y_k} \quad (140)$$

- Jansen pour la i^e variable d'entrée :

$$\Psi_S = 1 - \frac{0.5 \sum_{k=1}^d x_k}{\sum_{k=1}^d y_k} \quad (141)$$

$$\Psi_{ST} = \frac{0.5 \sum_{k=1}^d x_k}{\sum_{k=1}^d y_k} \quad (142)$$

- Martinez pour la i^e variable d'entrée :

$$\Psi_S = \frac{\sum_{k=1}^d x_k}{\sum_{k=1}^d \sqrt{y_k \times z_k}} \quad (143)$$

$$\Psi_{ST} = 1 - \frac{\sum_{k=1}^d x_k}{\sum_{k=1}^d \sqrt{y_k \times z_k}} \quad (144)$$

Le vecteur U_i contient également des colonnes supplémentaires correspondant aux échantillons associés à chaque variable de sortie. Le calcul de la variance est donc réalisée directement pour les indices agrégés et non par sorties.

7 Implémentation et validation

Un prototype python du calcul de la loi asymptotique a été développé pour chaque estimateur. Pour l'implémentation, il est plus simple de calculer la variance à partir de la formulation générale :

$$\text{Var}[S] = \frac{g^T \Gamma g}{N} \quad (145)$$

$$= \frac{\nabla \Psi_S(\mu)^T \text{Cov}[U] \nabla \Psi_S(\mu)}{N} \quad (146)$$

avec $\mu = \frac{1}{N} \sum_{k=1}^N U_k$.

Il est alors uniquement nécessaire de définir pour chaque estimateur la fonction Ψ_S ainsi que l'échantillon U . Le calcul du gradient à la moyenne et de la matrice de covariance est réalisée par OpenTURNS.

Le prototype Python permet de calculer la distribution asymptotique des estimateurs du premier ordre et de l'ordre total. Ces distributions théoriques sont comparées aux distributions empiriques obtenues après 1000 répétitions du calcul des indices de Sobol par chaque estimateur. Il ne s'agit pas de bootstrap mais d'une simulation de Monte Carlo avec un nouvel échantillon pour chaque estimation des indices. La distribution empirique est alors affichée dans un histogramme et son écart-type empirique est comparé à l'écart-type de la distribution théorique. Les indices de Sobol sont estimés avec un échantillon de 5000 réalisations.

La validation a été réalisée en évaluant les indices de Sobol et la variance empirique sur 4 modèles mathématiques :

- La fonction G-Sobol à 3 paramètres de loi uniforme compris entre 0 et 1.
- La fonction Ishigami à 3 paramètres de loi uniforme entre $-\pi$ et π .
- Une fonction "poutre" à 5 paramètres de loi lognormale.
- Une fonction "aggrégée" à 3 paramètres de loi uniforme compris entre 0 et 1, et avec 2 sorties.

7.1 Distributions théoriques et empiriques

7.1.1 Fonction G-Sobol

Sur toutes les figures, l'histogramme empirique des indices de Sobol est représenté en bleu. La distribution asymptotique d'un seul calcul d'indices de Sobol est affichée en rouge. La moyenne de la distribution asymptotique présente un biais par rapport à la moyenne empirique. Ce résultat est attendu car il s'agit d'une estimation unique des indices de Sobol. Il est donc possible

SaltelliSensitivityAlgorithm - N=5000 - Repetitions = 1000

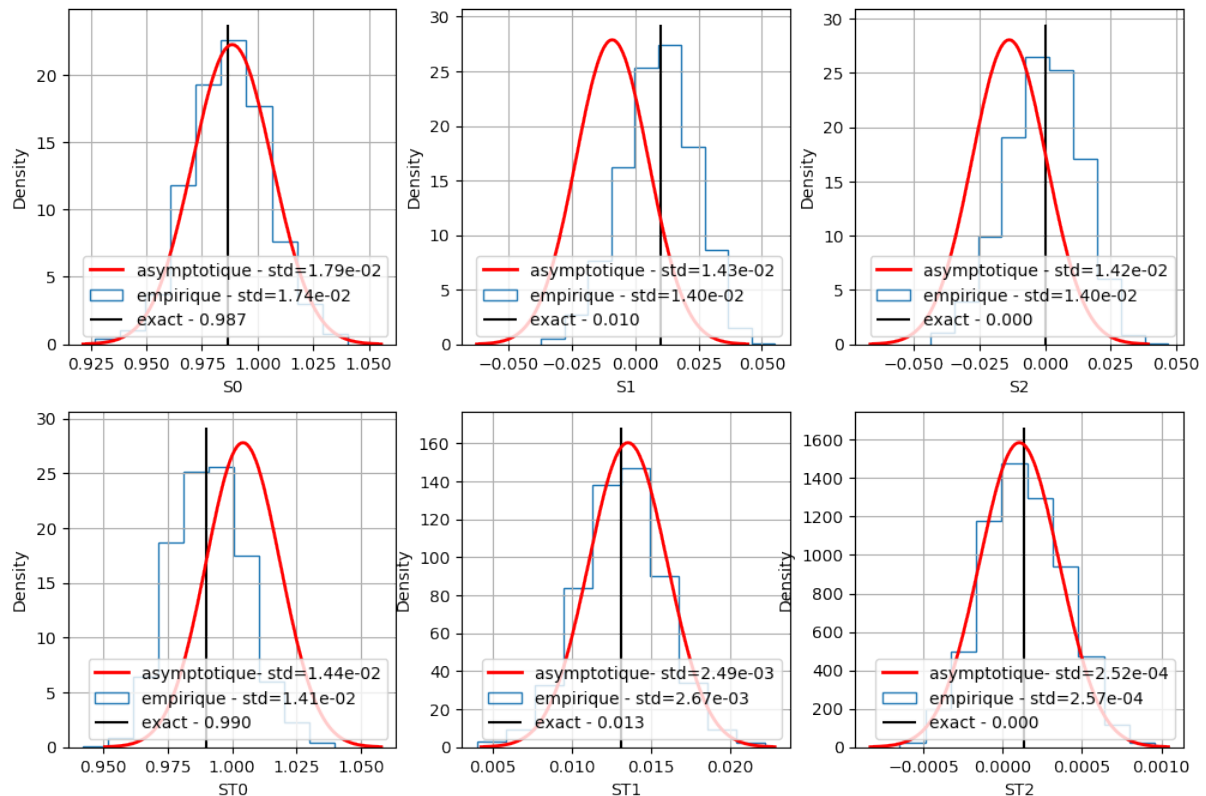


Figure 1 – Distributions des indices pour la fonction G-Sobol avec l'estimateur de Saltelli.

d'avoir un écart par rapport à la valeur moyenne ou exact des indices de Sobol. Par comparaison, les figures de la section 7.2 sont bien centrées avec l'histogramme empirique car il s'agit des distributions moyennes.

Les écart-types (ou variances) des distributions théoriques sont toutes en adéquation avec les écart-types empiriques ce qui permet de valider le calcul théorique des variance pour tous les estimateurs.

Toutes les figures suivantes peuvent être analysées de façon similaire, c'est pourquoi elles ne seront pas commentées.

JansenSensitivityAlgorithm - N=5000 - Repetitions = 1000

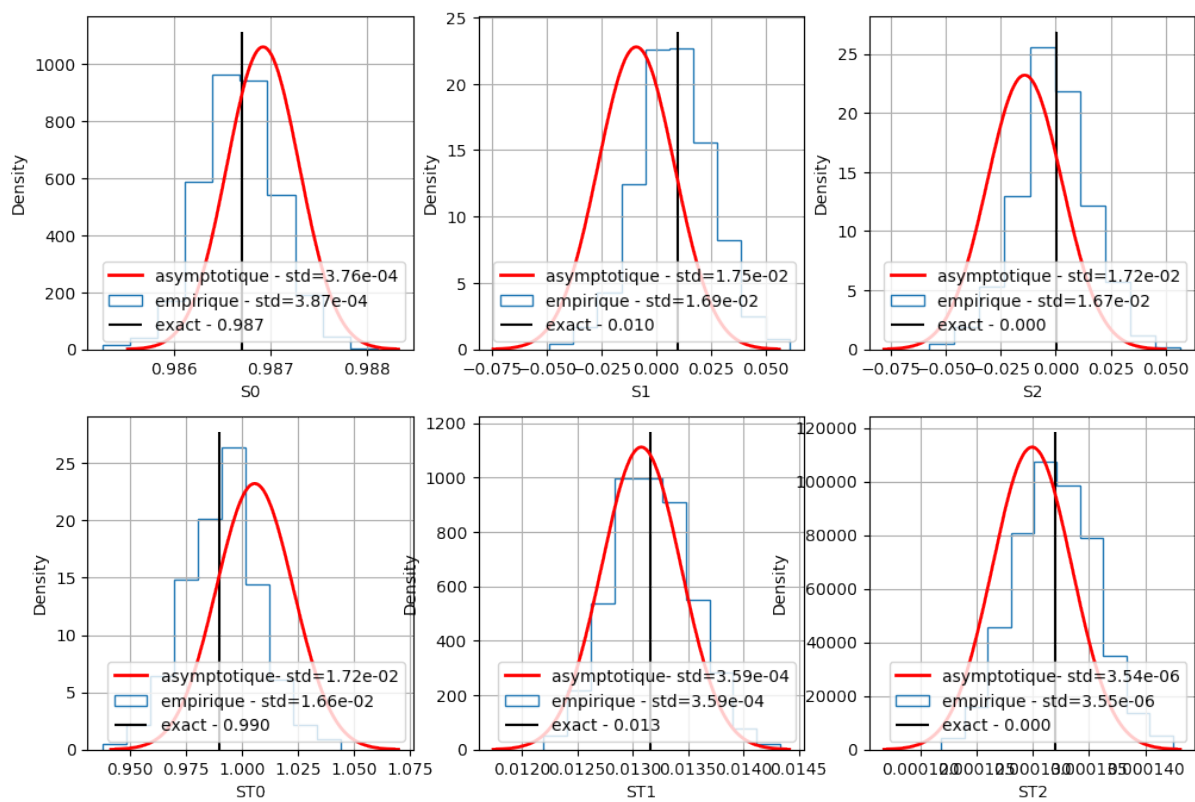


Figure 2 – Distributions des indices pour la fonction G-Sobol avec l'estimateur de Jansen.

MauntzKucherenkoSensitivityAlgorithm - N=5000 - Repetitions = 1000

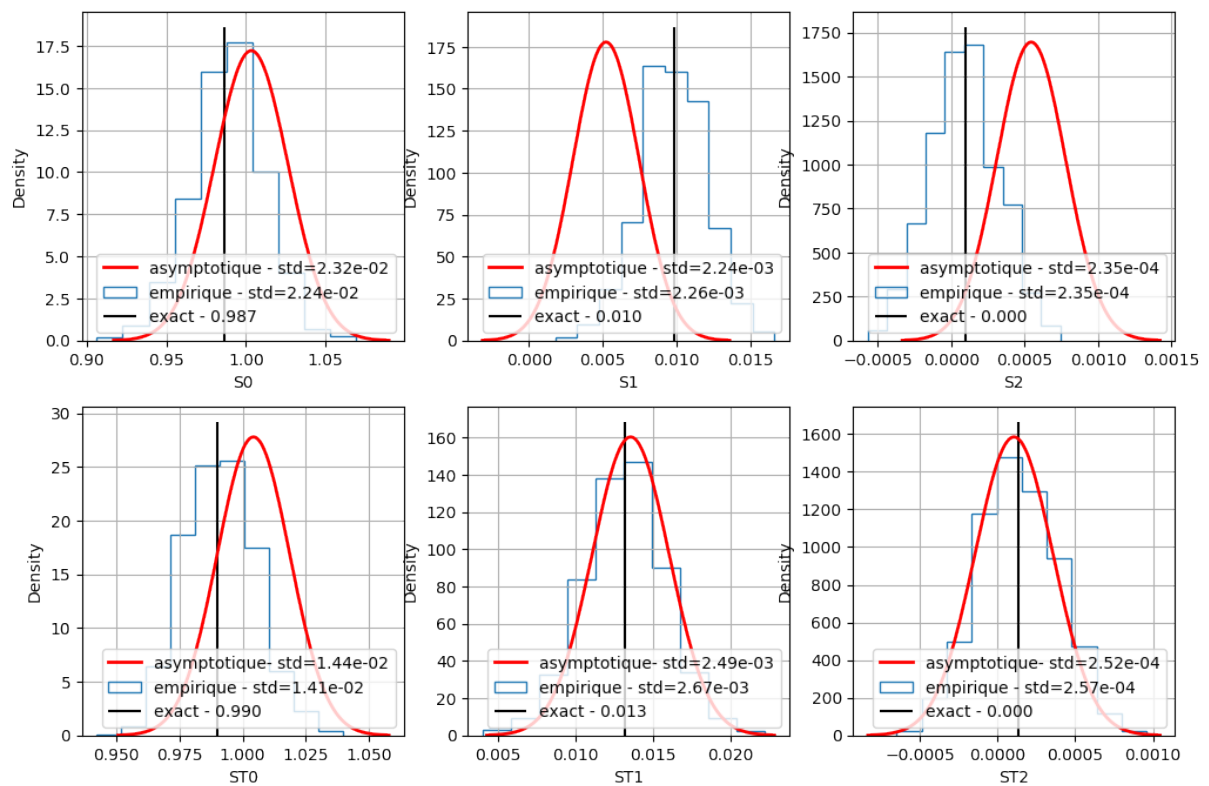


Figure 3 – Distributions des indices pour la fonction G-Sobol avec l'estimateur de Mauntz-Kucherenko.

MartinezSensitivityAlgorithm - N=5000 - Repetitions = 1000

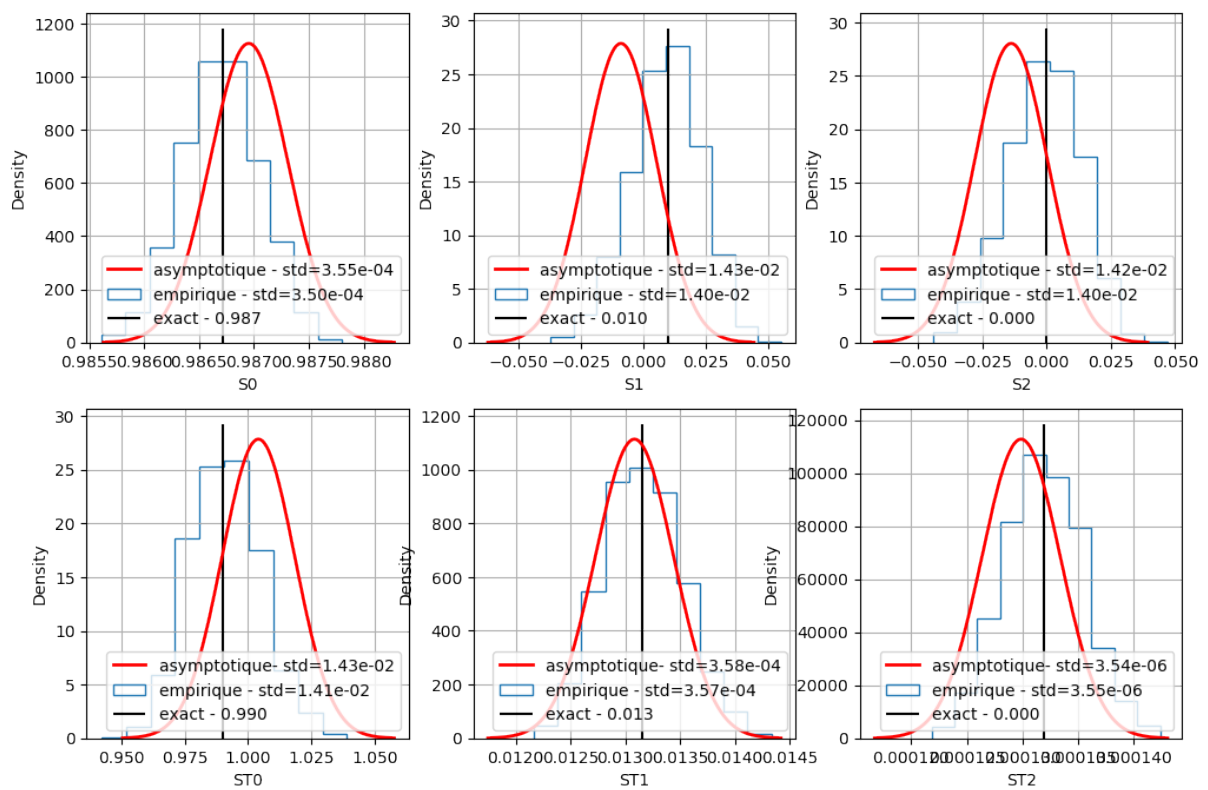


Figure 4 – Distributions des indices pour la fonction G-Sobol avec l'estimateur de Martinez.

7.1.2 Fonction Ishigami

SaltelliSensitivityAlgorithm - N=5000 - Repetitions = 1000

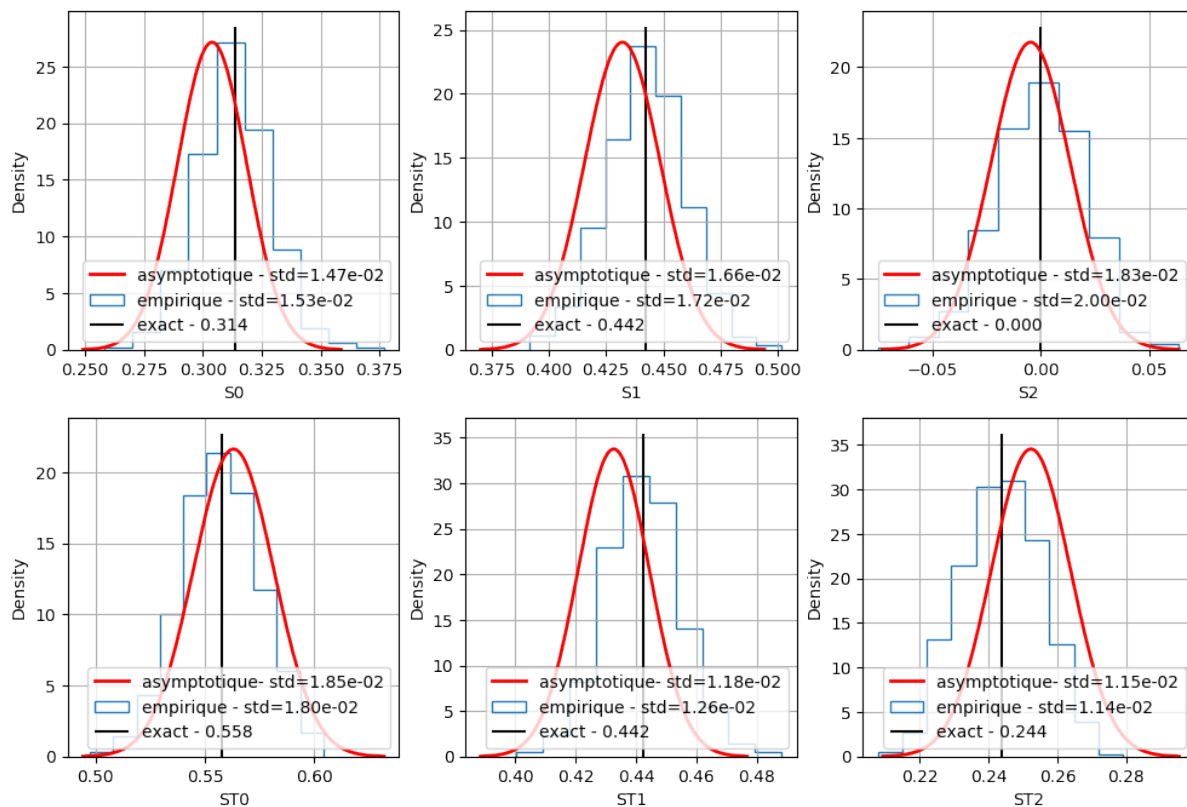


Figure 5 – Distributions des indices pour la fonction Ishigami avec l'estimateur de Saltelli.

JansenSensitivityAlgorithm - N=5000 - Repetitions = 1000

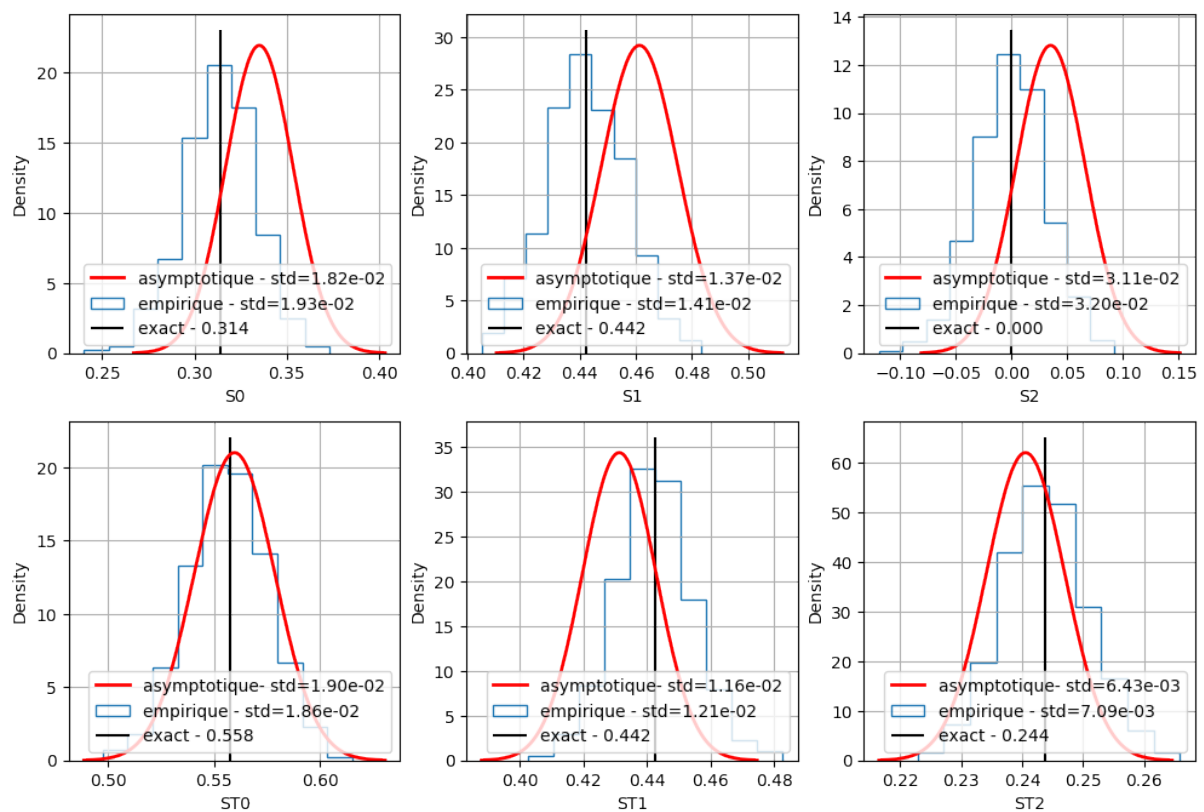


Figure 6 – Distributions des indices pour la fonction Ishigami avec l'estimateur de Jansen.

MauntzKucherenkoSensitivityAlgorithm - N=5000 - Repetitions = 1000

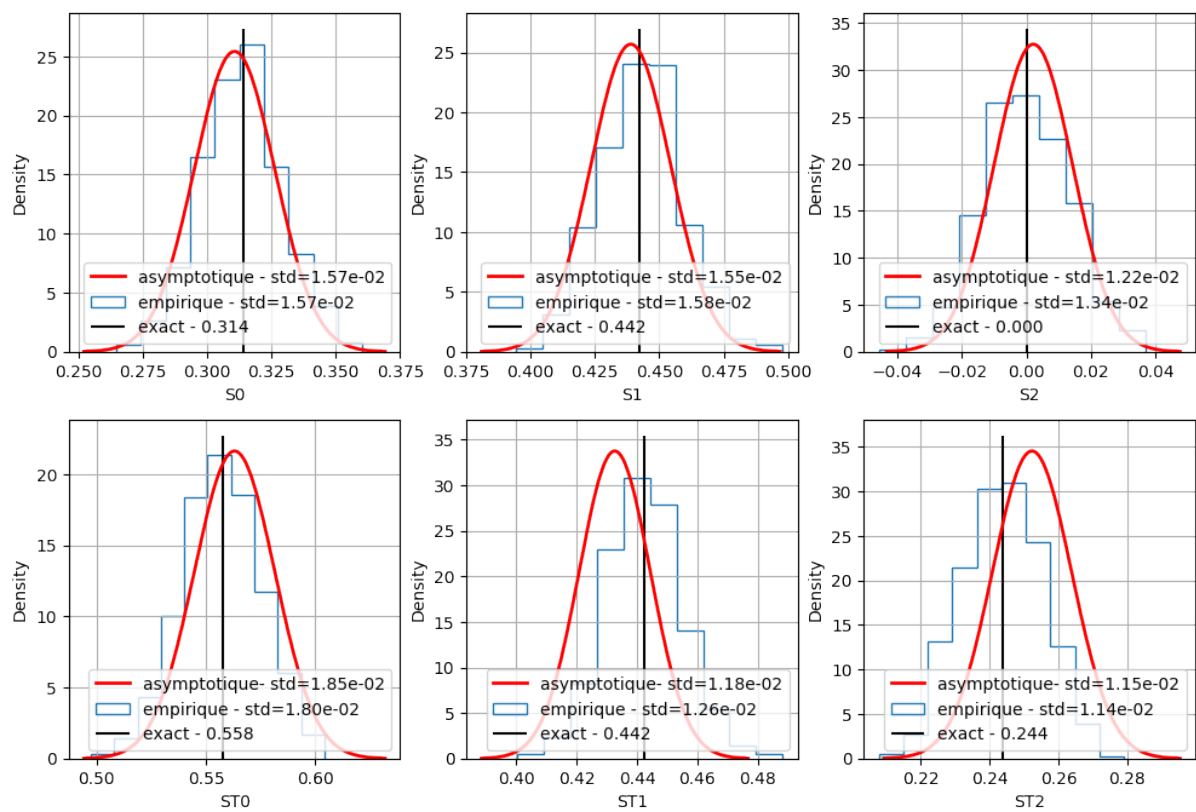


Figure 7 – Distributions des indices pour la fonction Ishigami avec l'estimateur de Mauntz-Kucherenko.

MartinezSensitivityAlgorithm - N=5000 - Repetitions = 1000

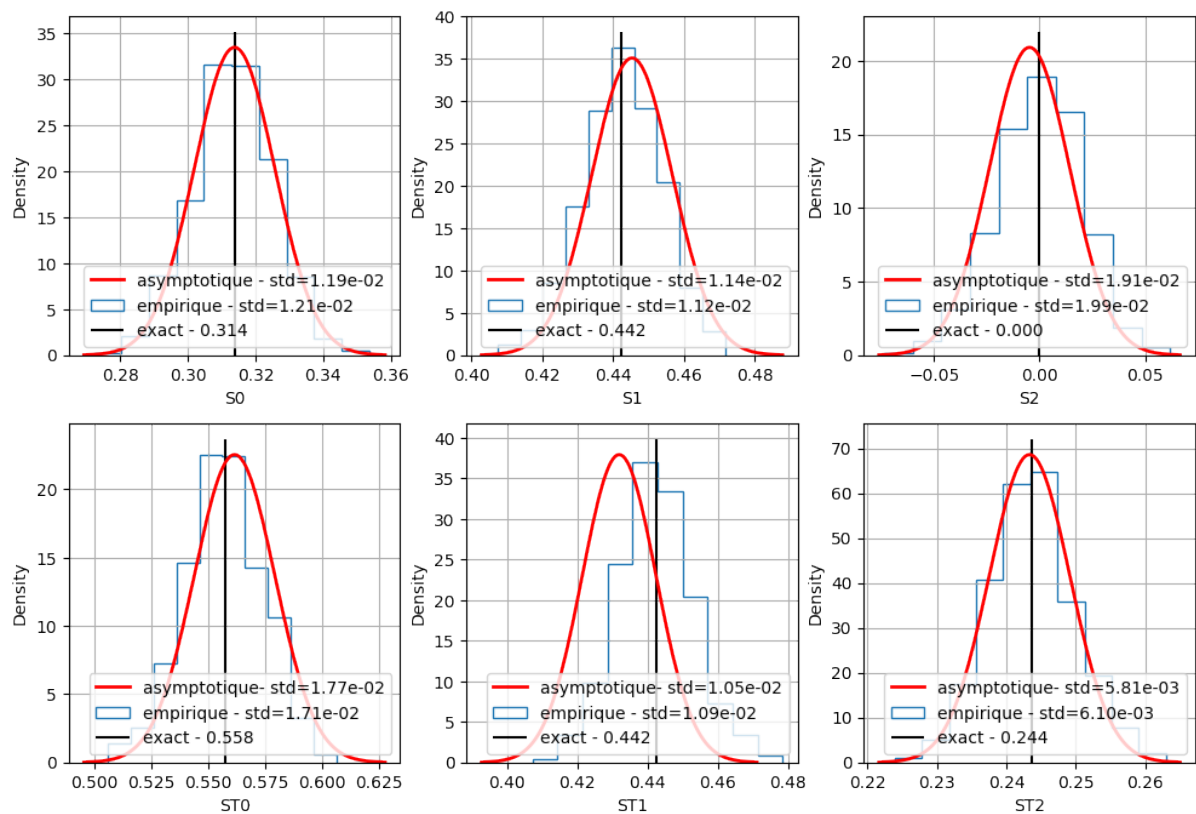


Figure 8 – Distributions des indices pour la fonction Ishigami avec l’estimateur de Martinez.

7.1.3 Fonction poutre

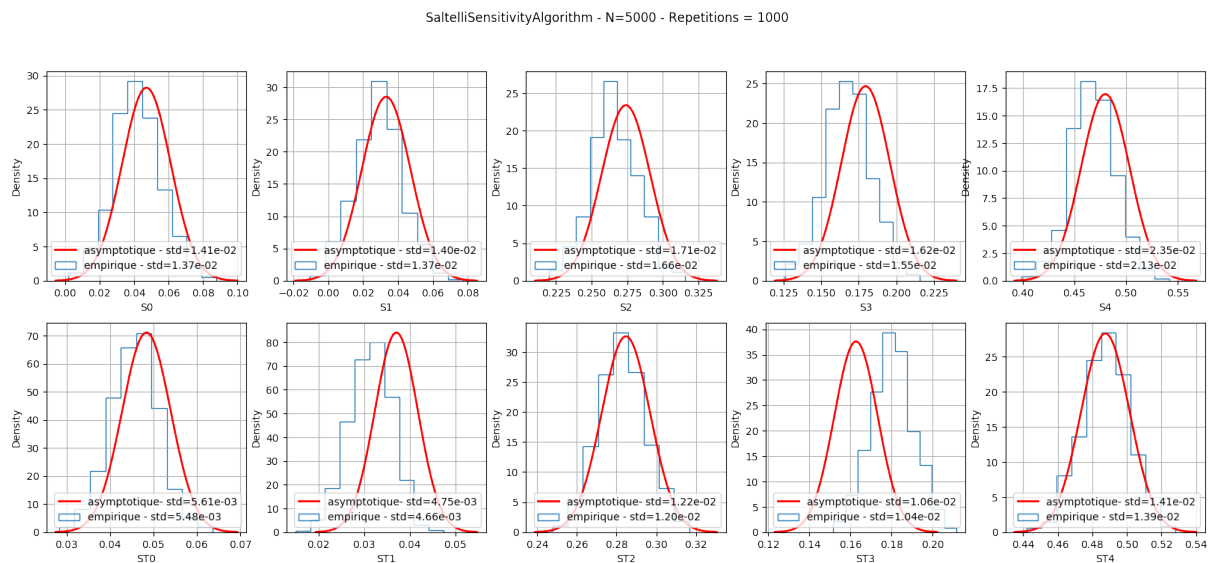


Figure 9 – Distributions des indices pour la fonction poutre avec l'estimateur de Saltelli.

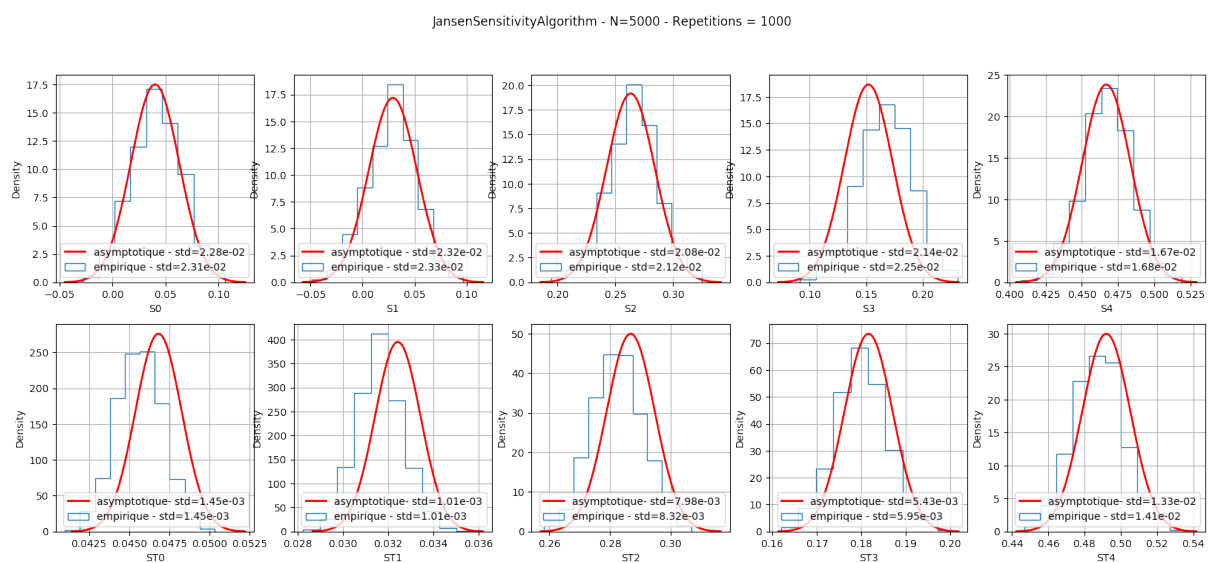


Figure 10 – Distributions des indices pour la fonction poutre avec l'estimateur de Jansen.

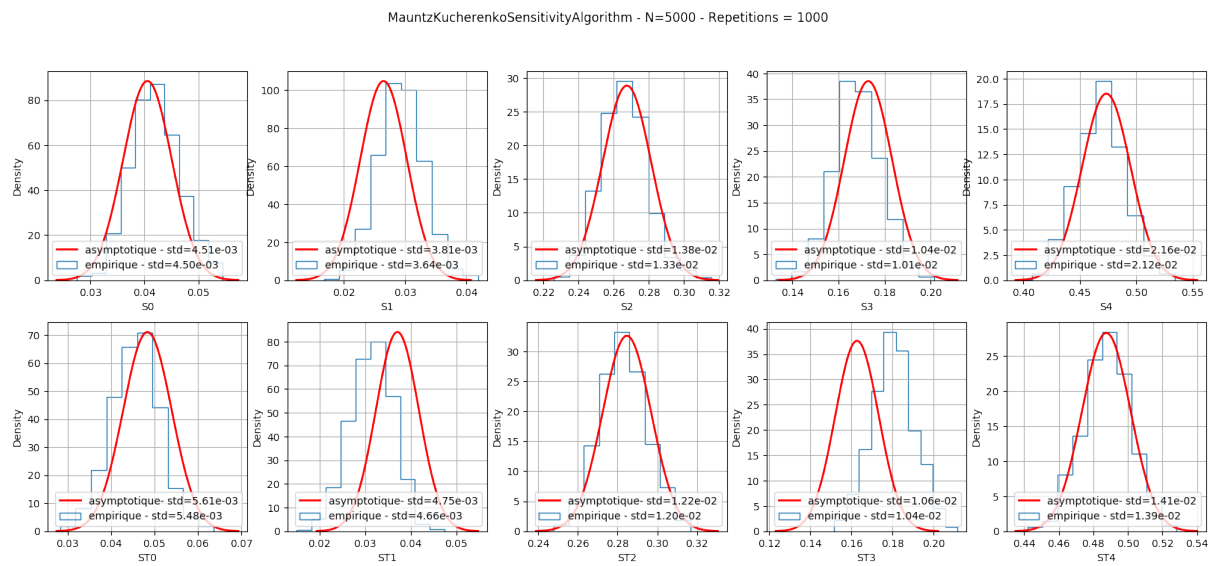


Figure 11 – Distributions des indices pour la fonction poutre avec l'estimateur de Mauntz-Kucherenko.

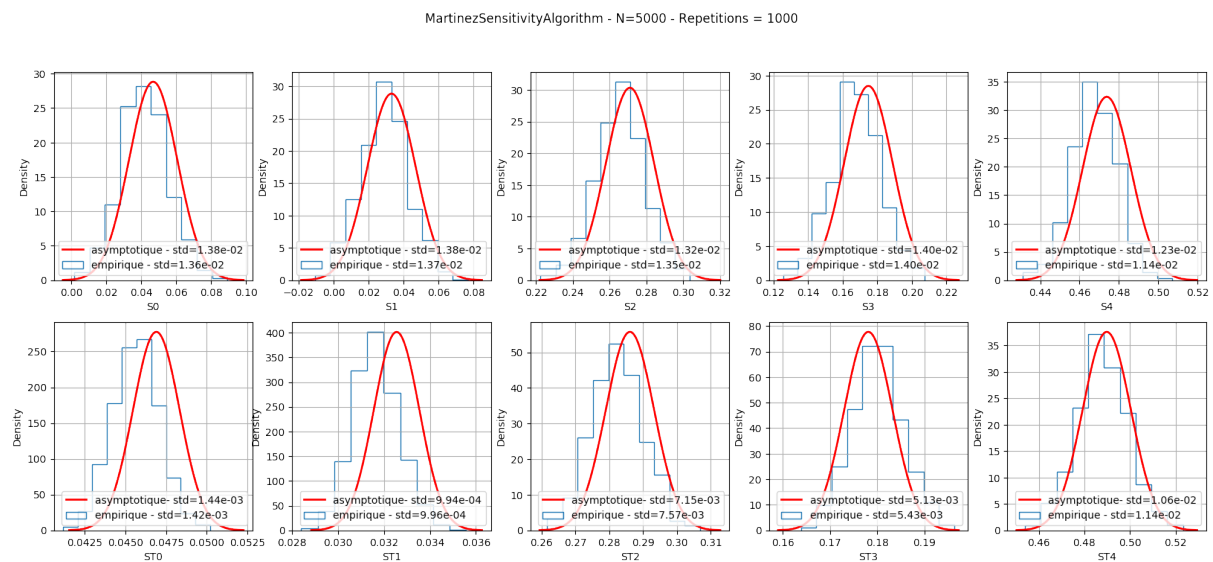


Figure 12 – Distributions des indices pour la fonction poutre avec l'estimateur de Martinez.

7.1.4 Fonction agrégée

SaltelliSensitivityAlgorithm - N=5000 - Repetitions = 1000

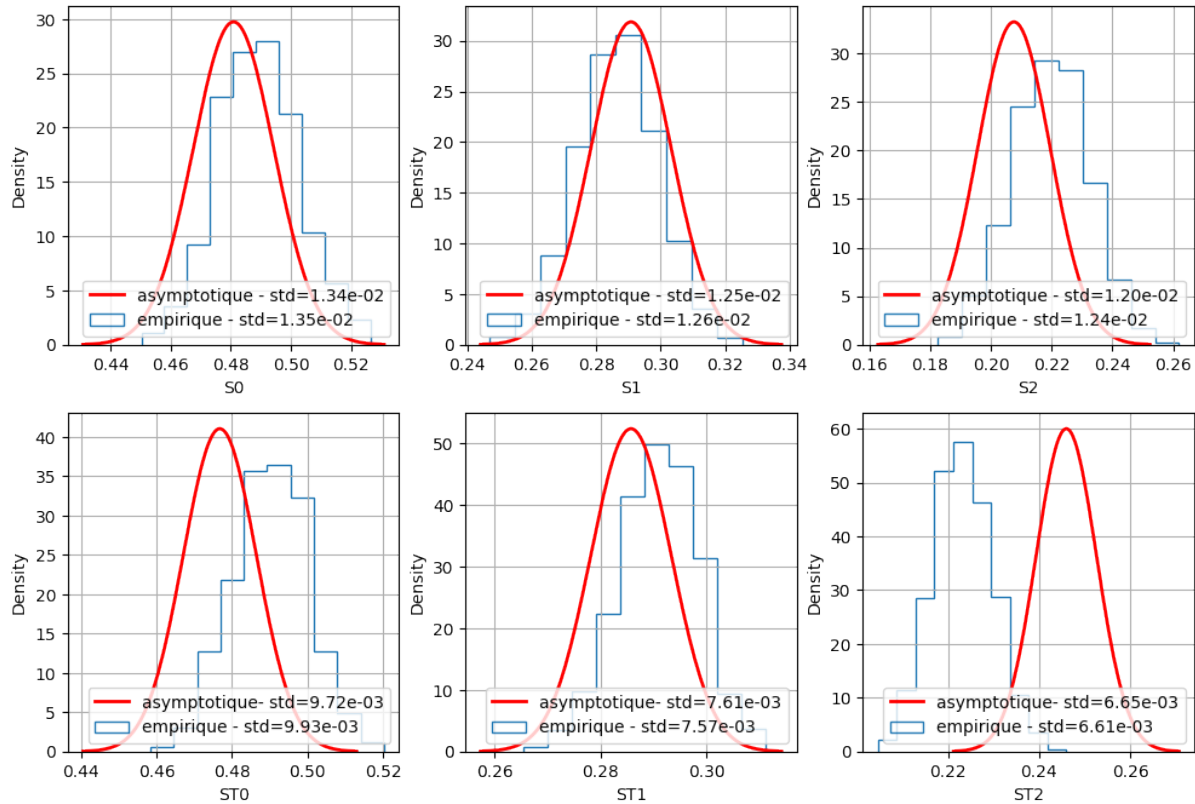


Figure 13 – Distributions des indices pour la fonction agrégée avec l'estimateur de Saltelli.

JansenSensitivityAlgorithm - N=5000 - Repetitions = 1000

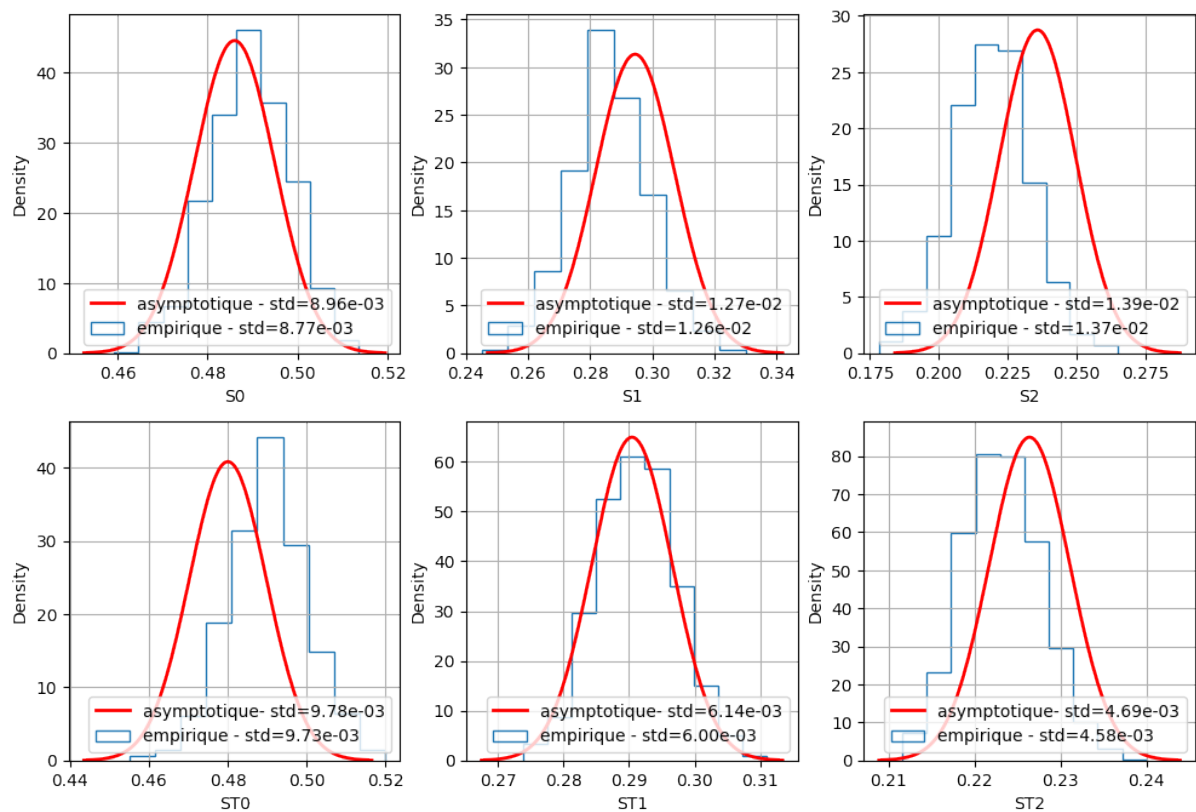


Figure 14 – Distributions des indices pour la fonction agrégée avec l'estimateur de Jansen.

MauntzKucherenkoSensitivityAlgorithm - N=5000 - Repetitions = 1000

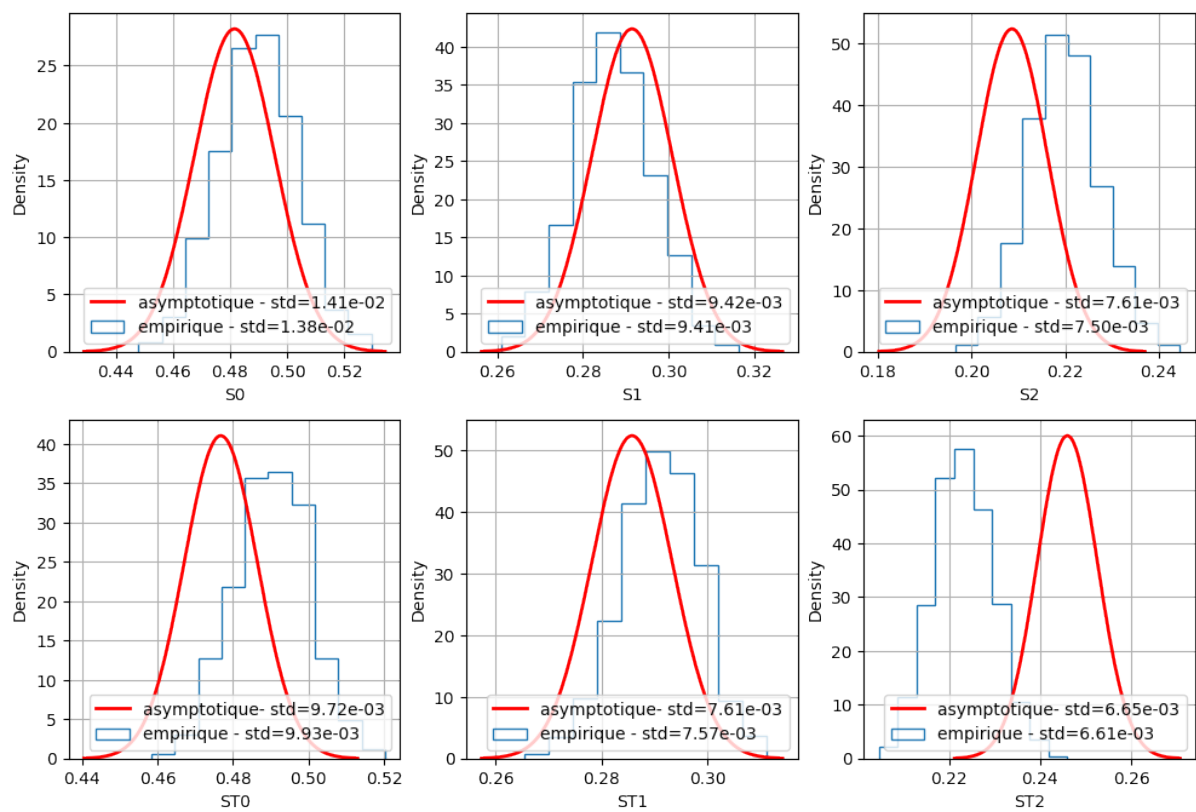


Figure 15 – Distributions des indices pour la fonction agrégée avec l'estimateur de Mauntz-Kucherenko.

MartinezSensitivityAlgorithm - N=5000 - Repetitions = 1000

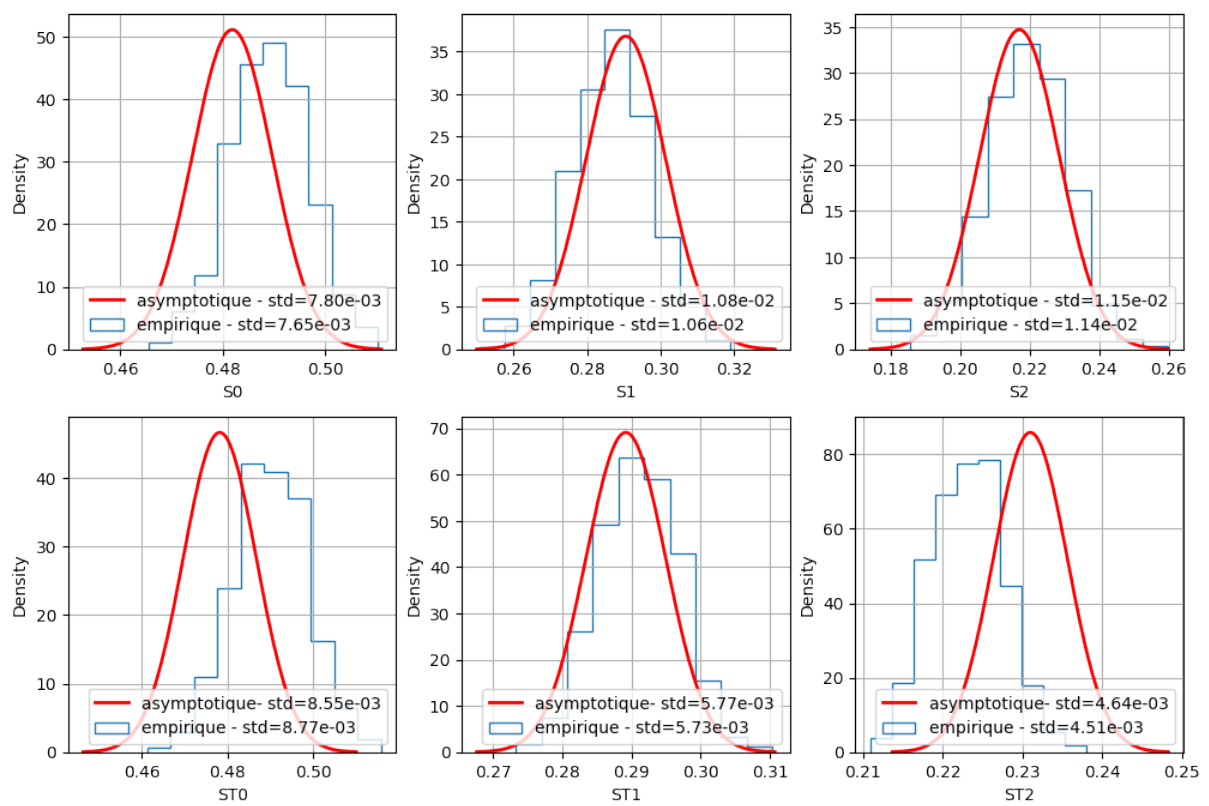


Figure 16 – Distributions des indices pour la fonction agrégée avec l'estimateur de Martinez.

7.2 Distributions théoriques moyennes et empiriques

Dans cette section, les paramètres des distributions asymptotiques sont les moyennes des indices de Sobol et les moyennes des écart-types de toutes les distributions asymptotiques calculées lors des 5000 répétitions. Les distributions asymptotiques sont alors bien centrées avec l'histogramme empirique.

7.2.1 Fonction G-Sobol

SaltelliSensitivityAlgorithm - N=5000 - Repetitions = 1000

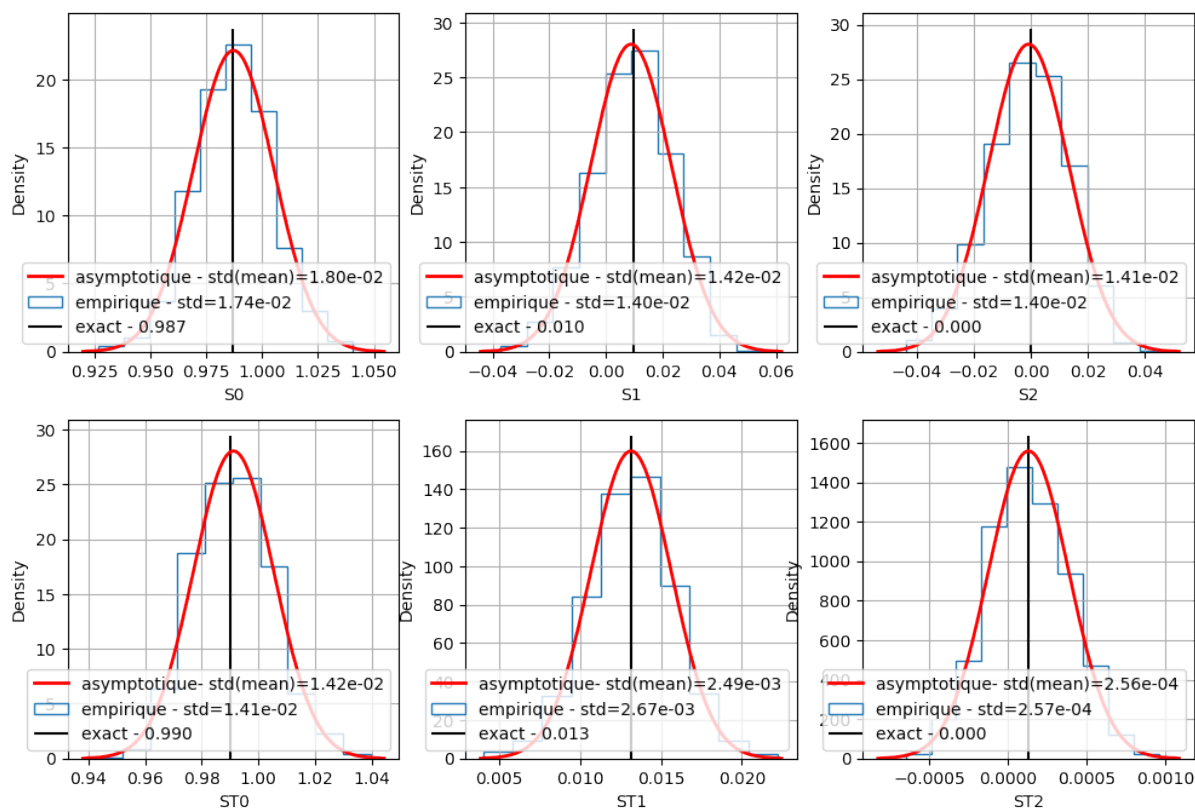


Figure 17 – Distributions moyennes des indices pour la fonction G-Sobol avec l'estimateur de Saltelli.

JansenSensitivityAlgorithm - N=5000 - Repetitions = 1000

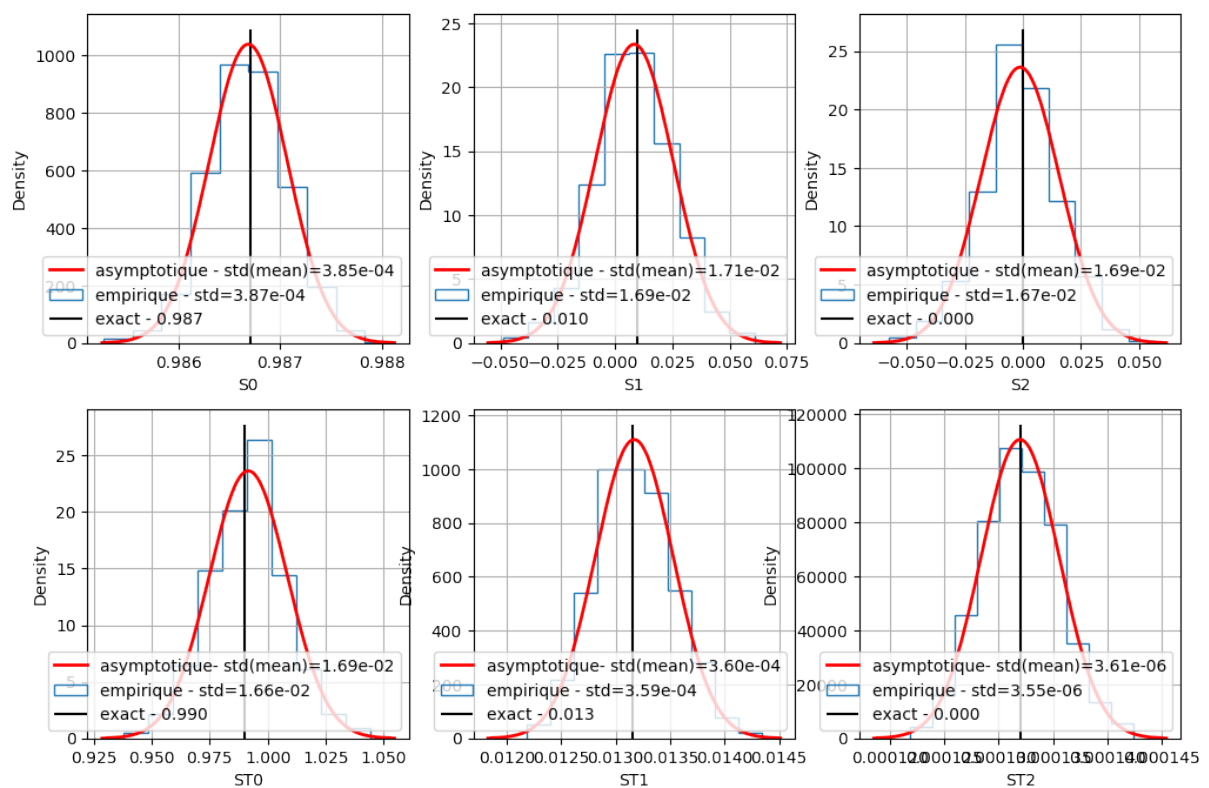


Figure 18 – Distributions moyennes des indices pour la fonction G-Sobol avec l'estimateur de Jansen.

MauntzKucherenkoSensitivityAlgorithm - N=5000 - Repetitions = 1000

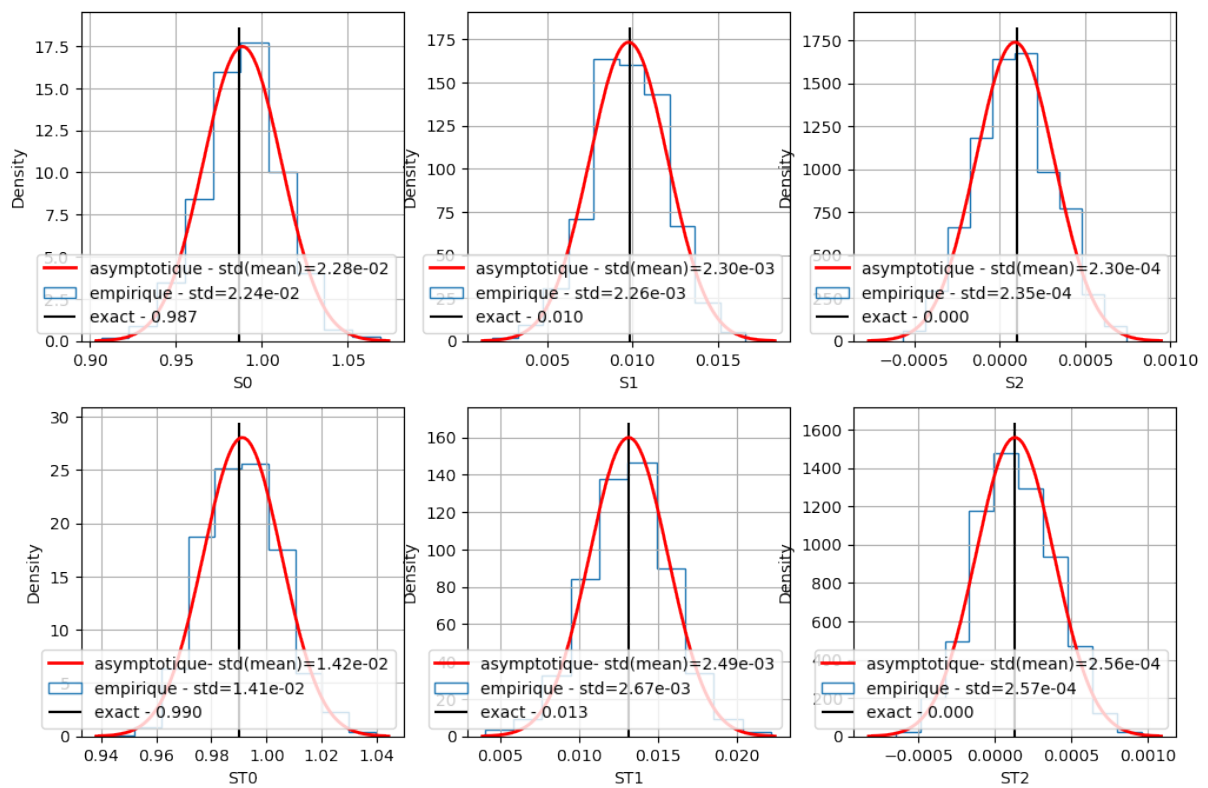


Figure 19 – Distributions moyennes des indices pour la fonction G-Sobol avec l'estimateur de Mauntz-Kucherenko.

MartinezSensitivityAlgorithm - N=5000 - Repetitions = 1000

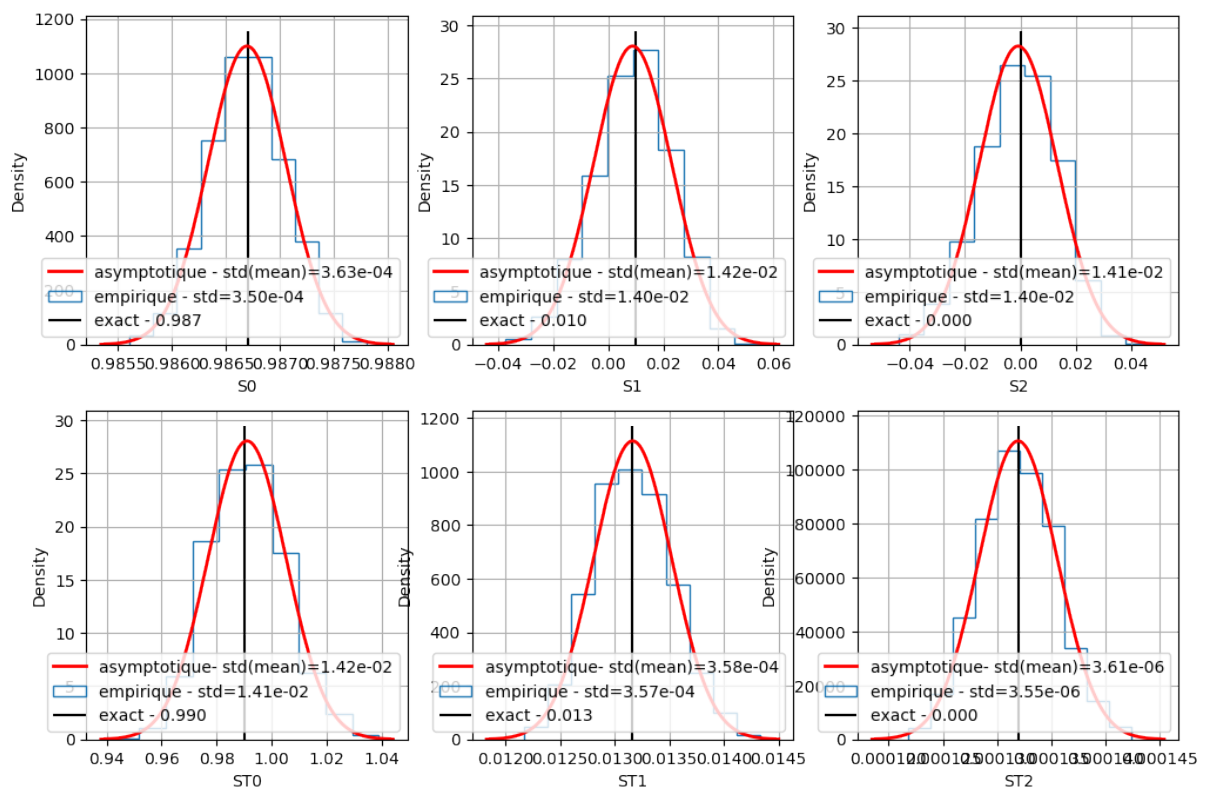


Figure 20 – Distributions moyennes des indices pour la fonction G-Sobol avec l’estimateur de Martinez.

7.2.2 Fonction Ishigami

SaltelliSensitivityAlgorithm - N=5000 - Repetitions = 1000

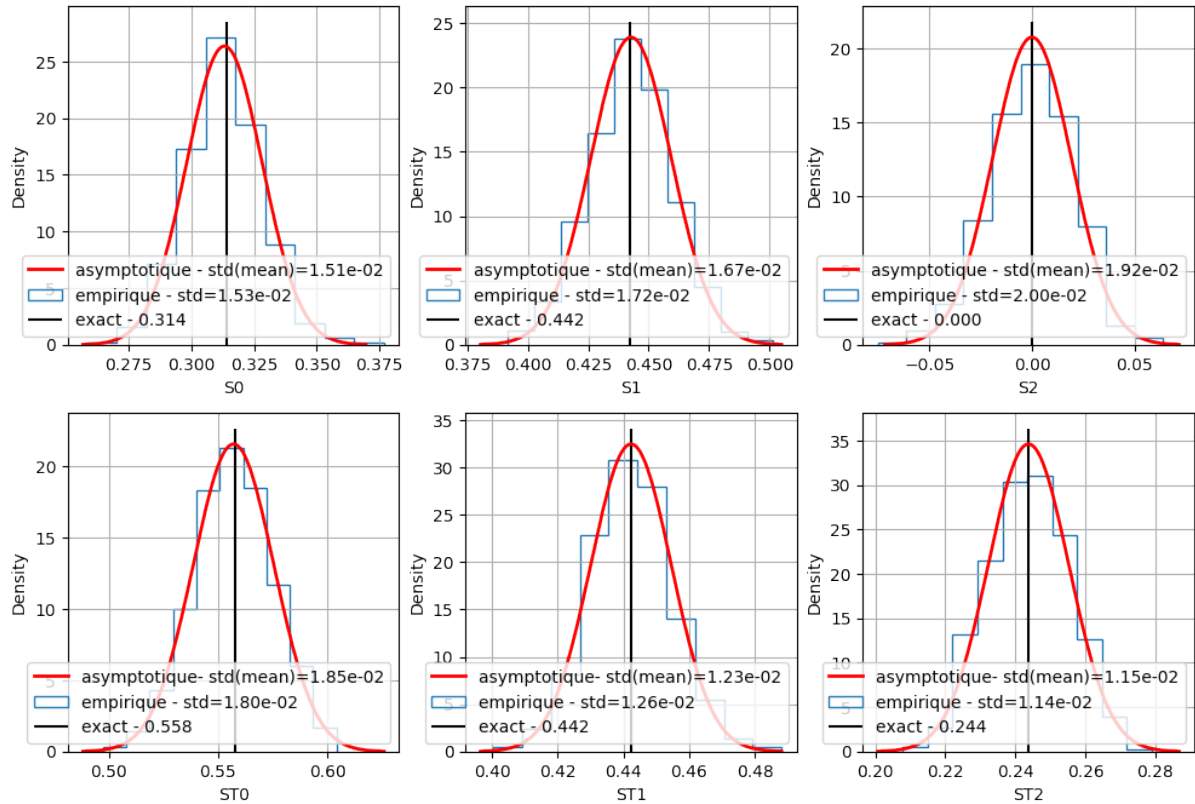


Figure 21 – Distributions moyennes des indices pour la fonction Ishigami avec l'estimateur de Saltelli.

JansenSensitivityAlgorithm - N=5000 - Repetitions = 1000

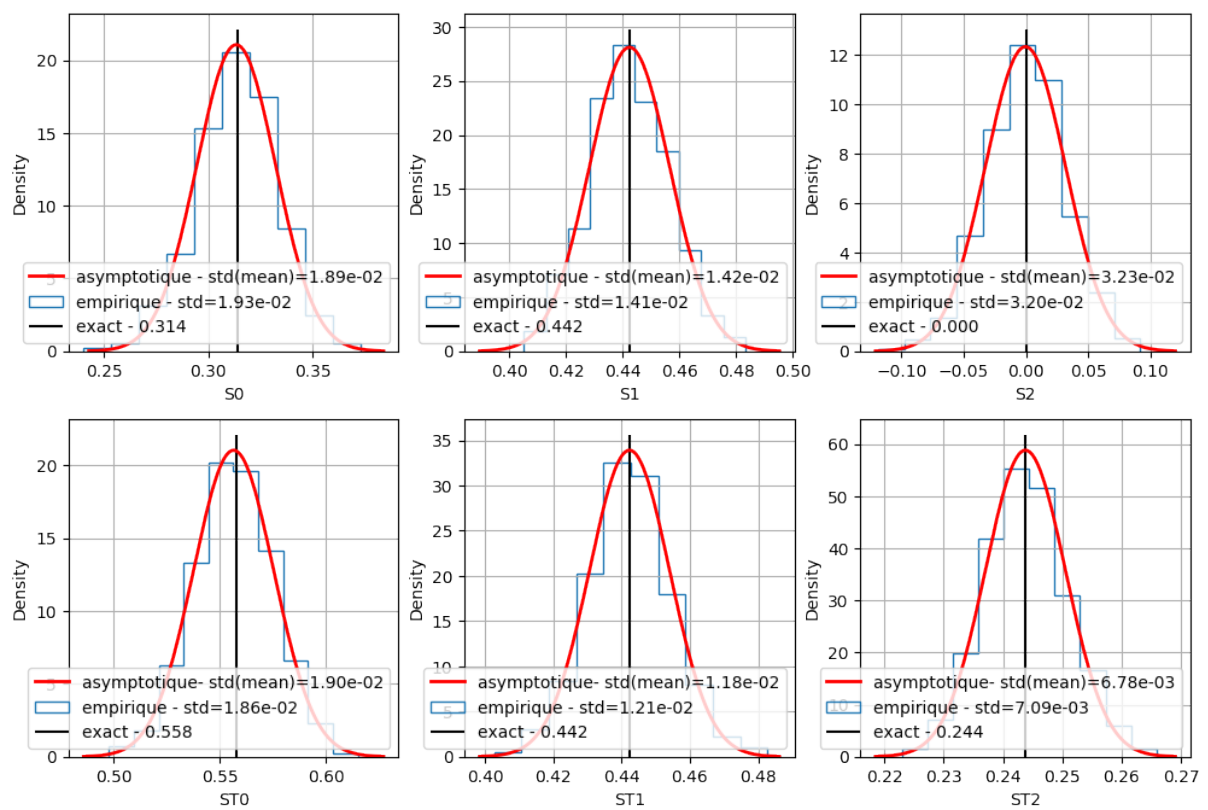


Figure 22 – Distributions moyennes des indices pour la fonction Ishigami avec l'estimateur de Jansen.

MauntzKucherenkoSensitivityAlgorithm - N=5000 - Repetitions = 1000

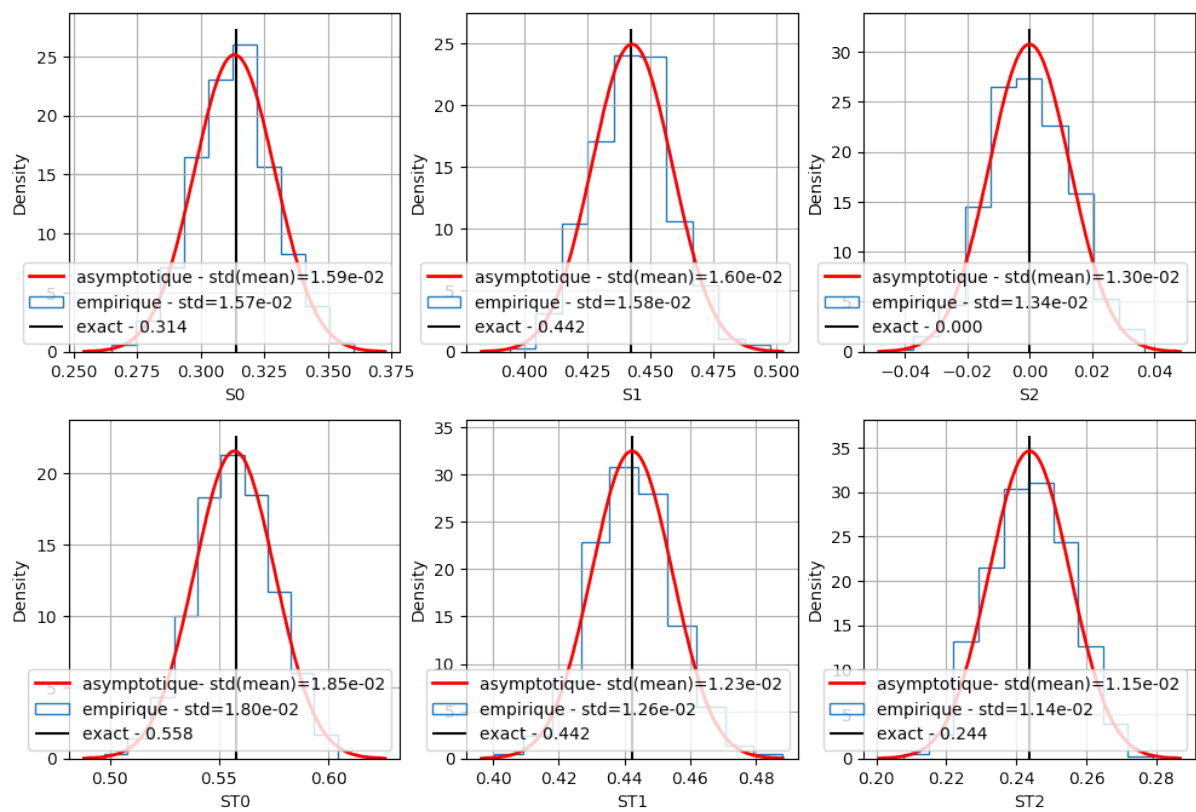


Figure 23 – Distributions moyennes des indices pour la fonction Ishigami avec l'estimateur de Mauntz-Kucherenko.

MartinezSensitivityAlgorithm - N=5000 - Repetitions = 1000

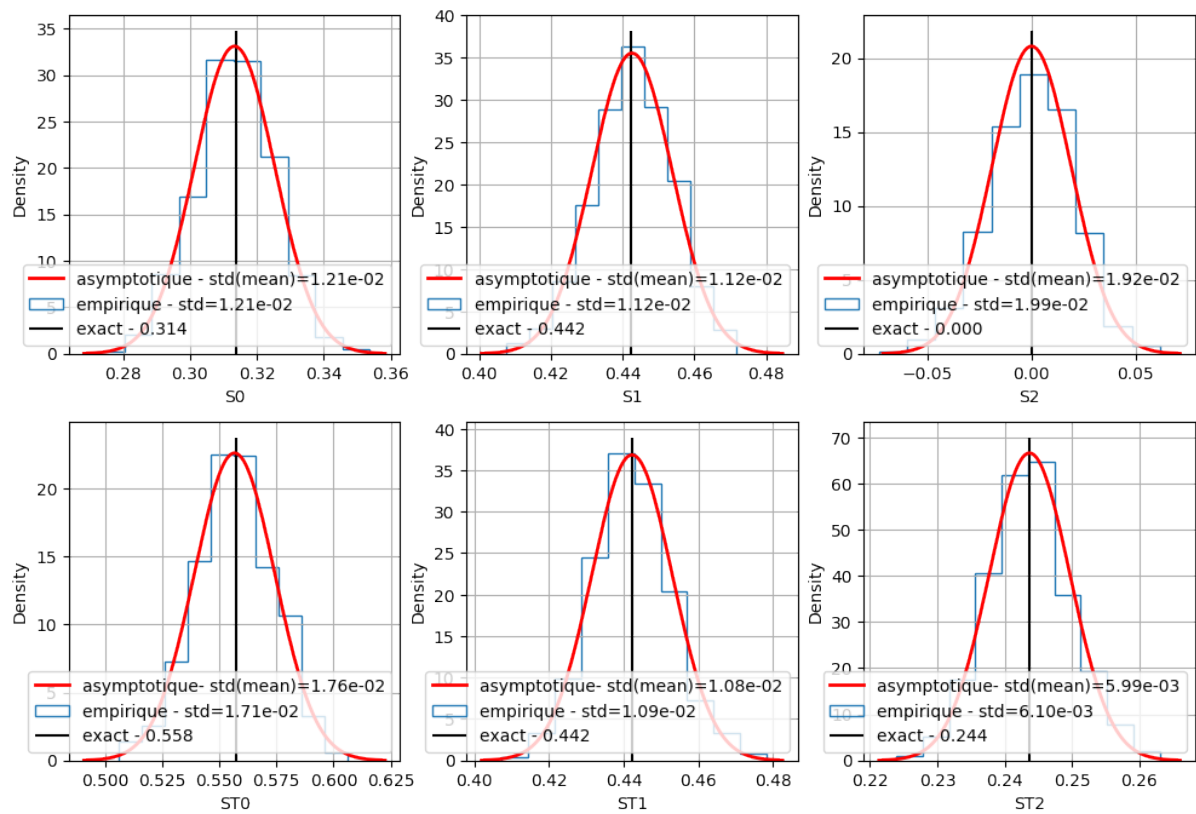


Figure 24 – Distributions moyennes des indices pour la fonction Ishigami avec l’estimateur de Martinez.

7.2.3 Fonction poutre

SaltelliSensitivityAlgorithm - N=5000 - Repetitions = 1000

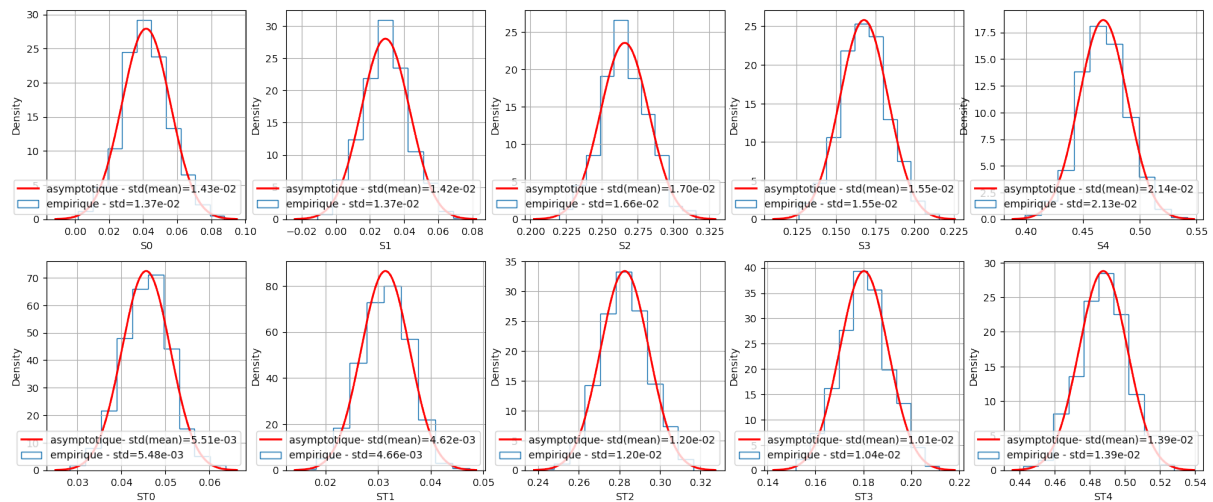


Figure 25 – Distributions moyennes des indices pour la fonction poutre avec l'estimateur de Saltelli.

JansenSensitivityAlgorithm - N=5000 - Repetitions = 1000

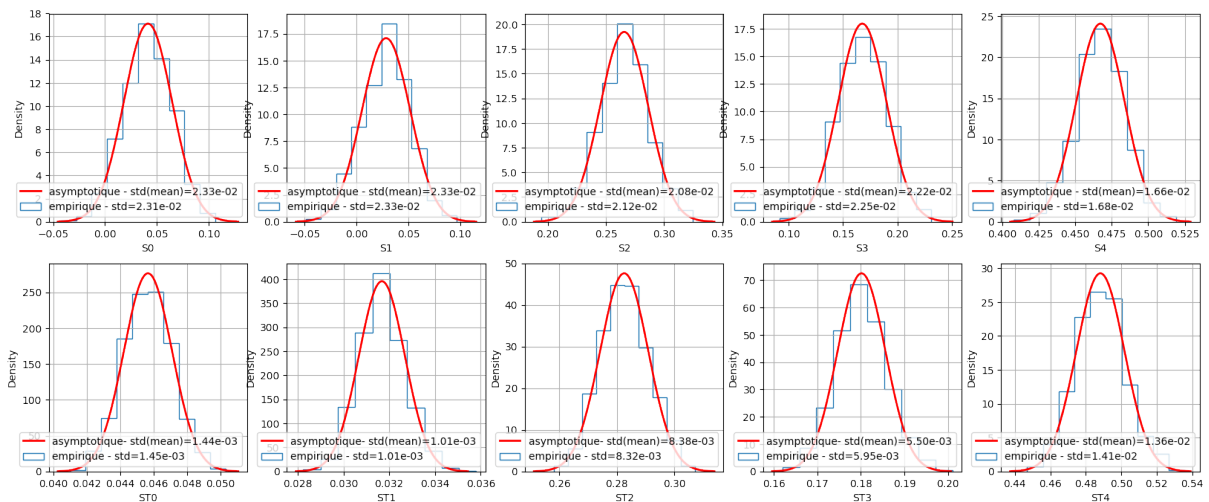


Figure 26 – Distributions moyennes des indices pour la fonction poutre avec l'estimateur de Jansen.

MauntzKucherenkoSensitivityAlgorithm - N=5000 - Repetitions = 1000

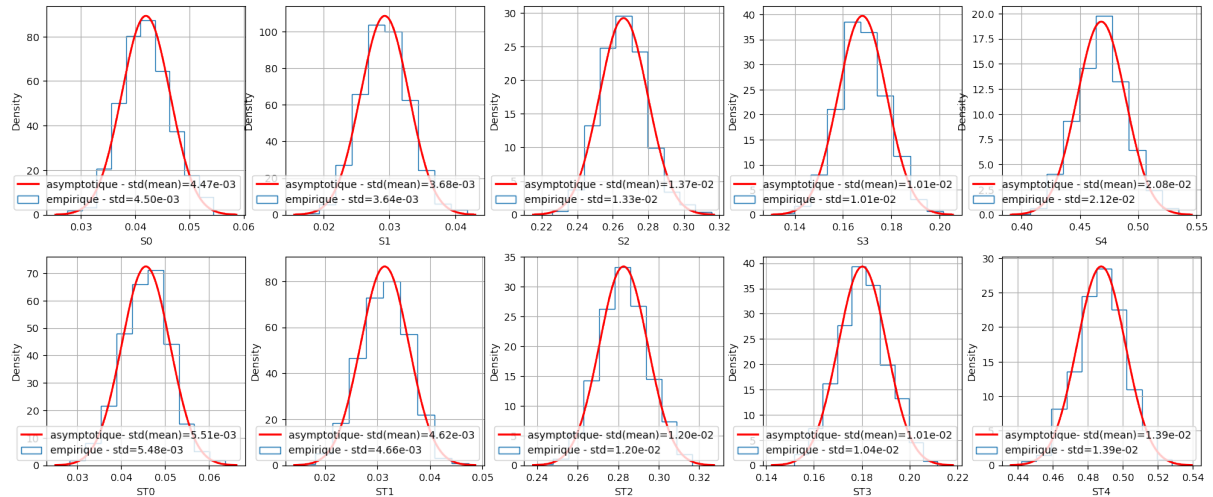


Figure 27 – Distributions moyennes des indices pour la fonction poutre avec l'estimateur de Mauntz-Kucherenko.

MartinezSensitivityAlgorithm - N=5000 - Repetitions = 1000

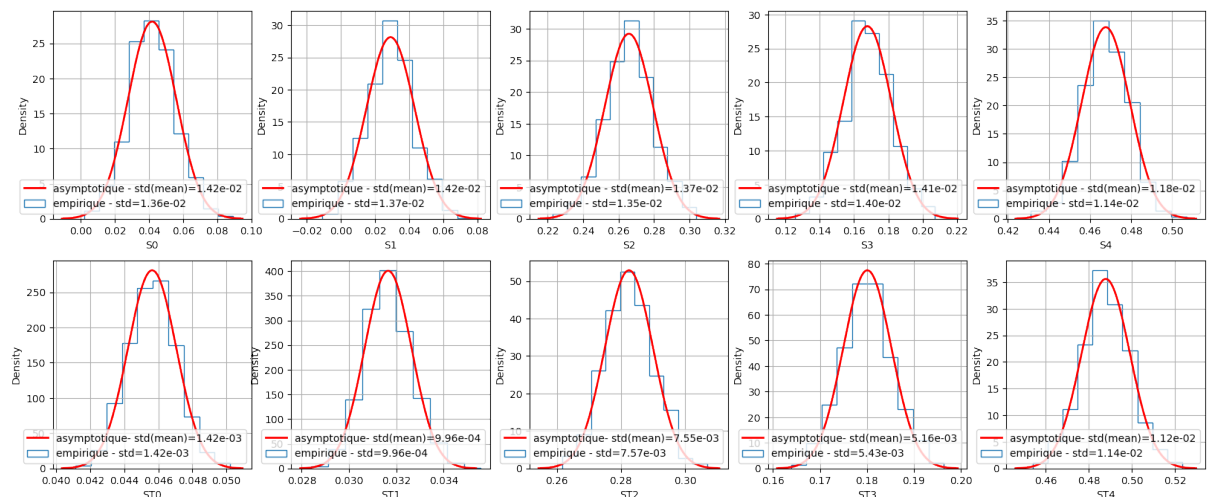


Figure 28 – Distributions moyennes des indices pour la fonction poutre avec l'estimateur de Martinez.

7.2.4 Fonction agrégée

SaltelliSensitivityAlgorithm - N=5000 - Repetitions = 1000

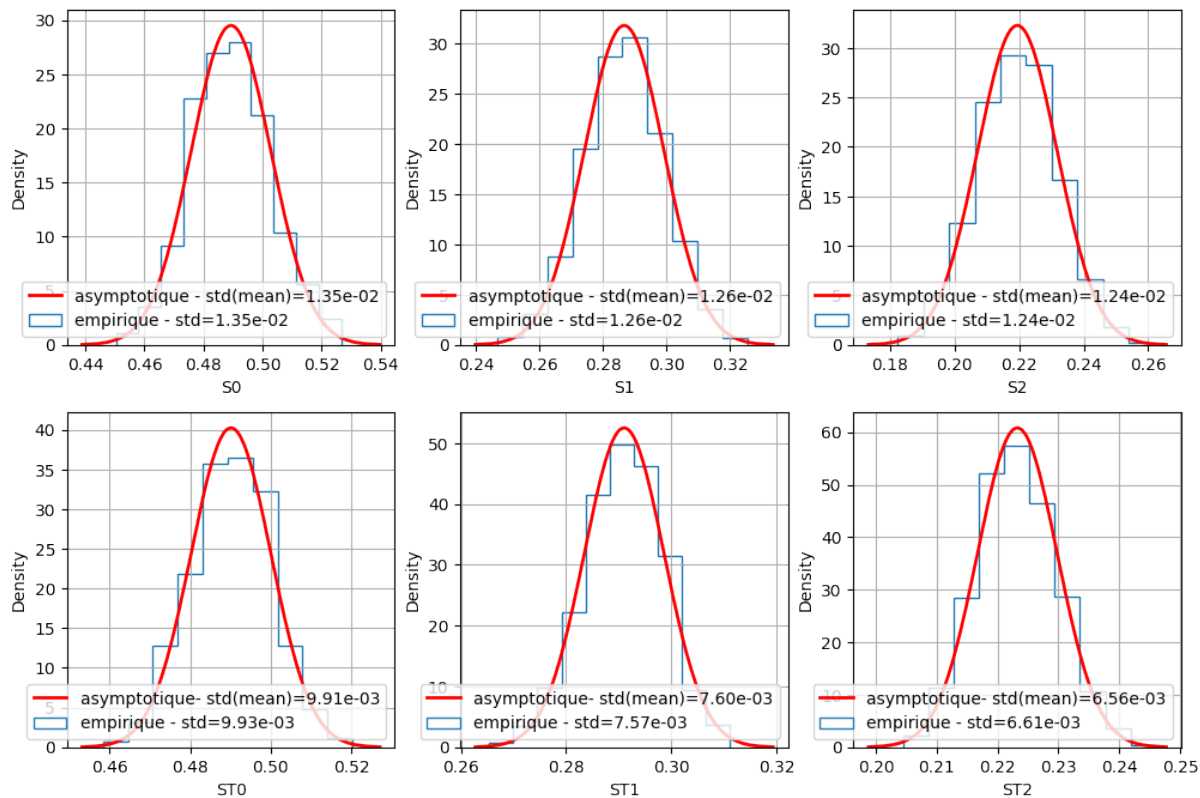


Figure 29 – Distributions moyennes des indices pour la fonction agrégée avec l'estimateur de Saltelli.

JansenSensitivityAlgorithm - N=5000 - Repetitions = 1000

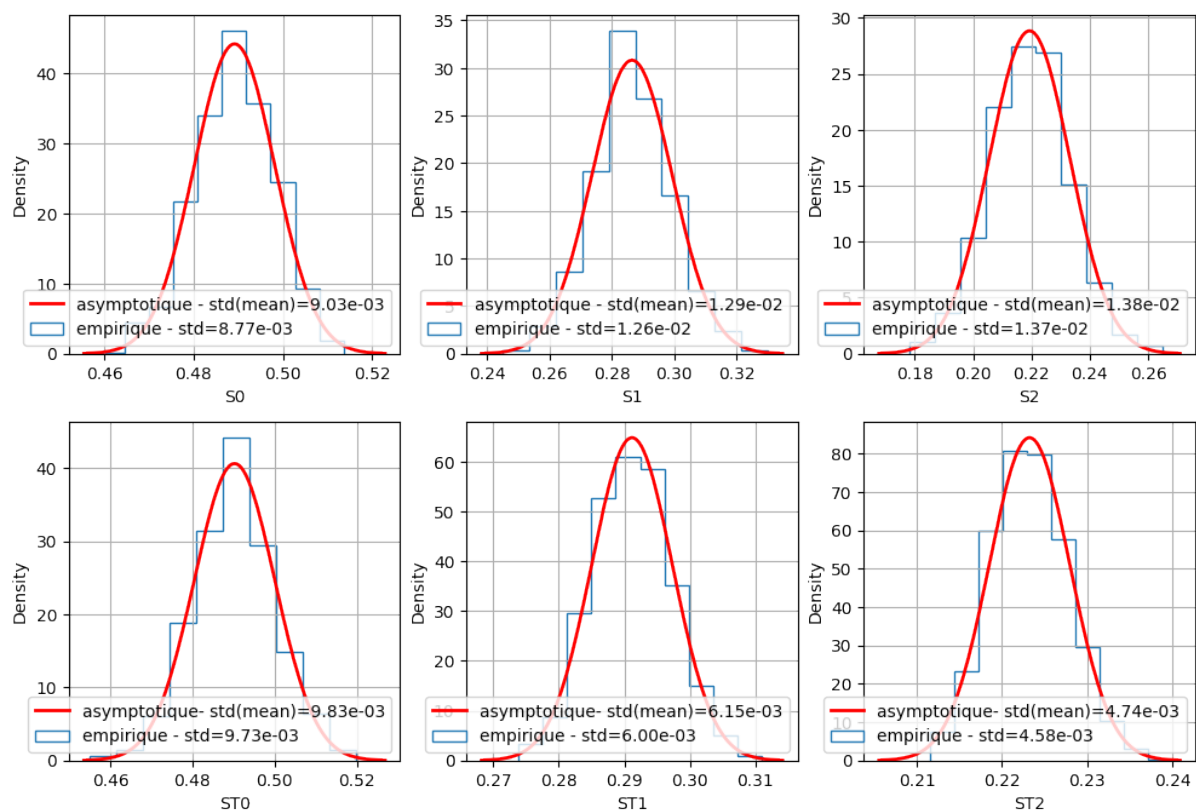


Figure 30 – Distributions moyennes des indices pour la fonction agrégée avec l'estimateur de Jansen.

MauntzKucherenkoSensitivityAlgorithm - N=5000 - Repetitions = 1000

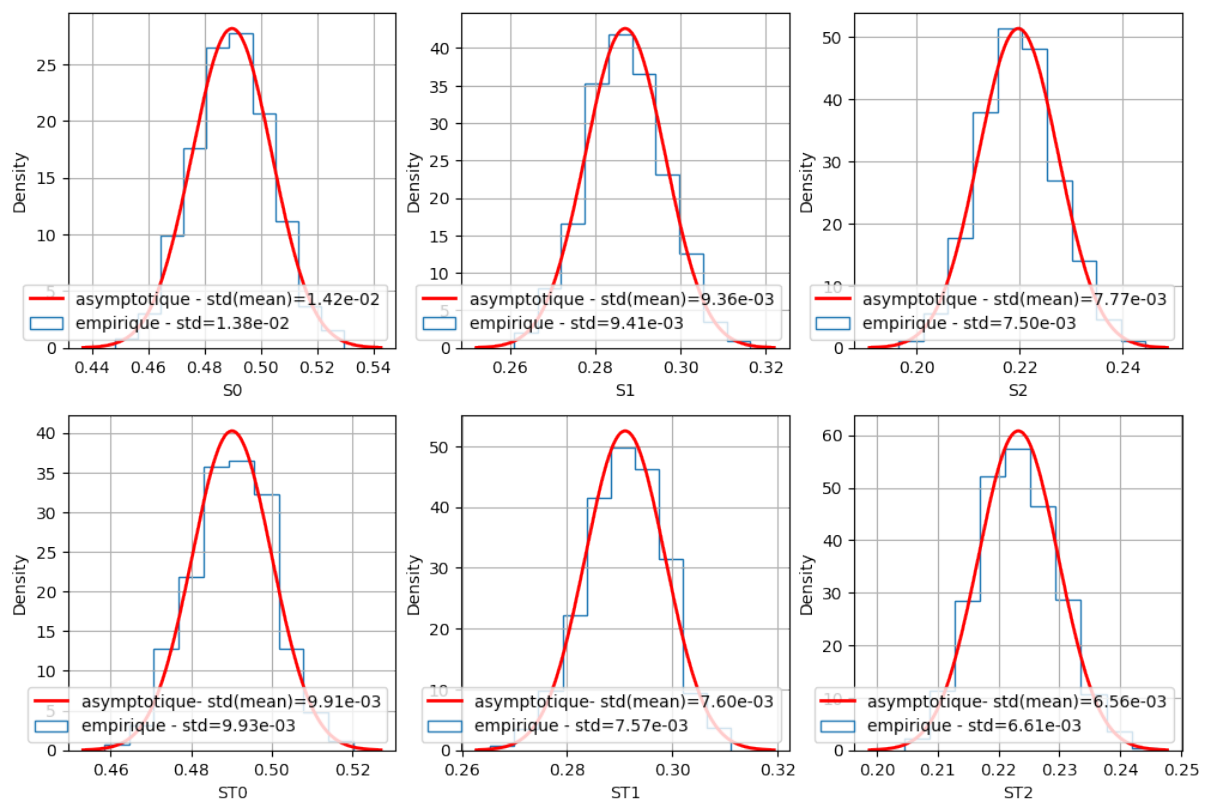


Figure 31 – Distributions moyennes des indices pour la fonction agrégée avec l'estimateur de Mauntz-Kucherenko.

MartinezSensitivityAlgorithm - N=5000 - Repetitions = 1000

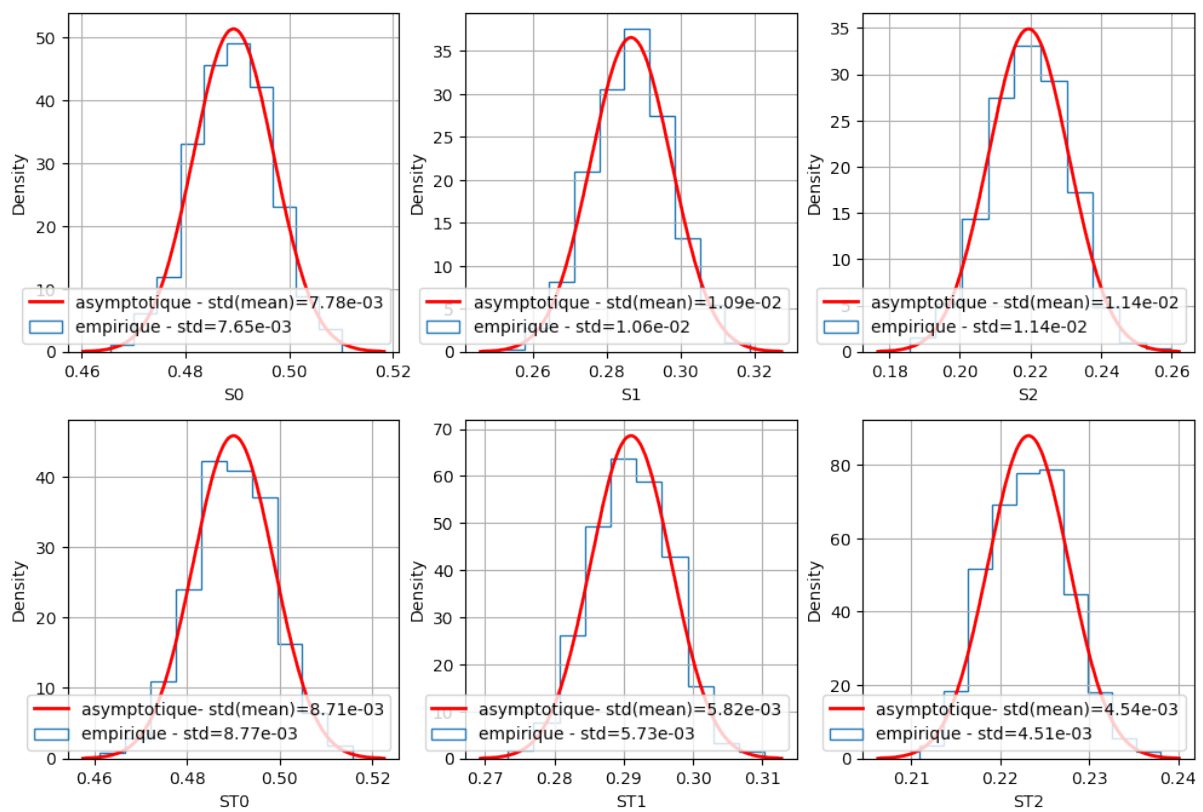


Figure 32 – Distributions moyennes des indices pour la fonction agrégée avec l’estimateur de Martinez.

7.3 Anomalie dans l'estimateur de Saltelli

L'estimateur du premier ordre de Saltelli présente une erreur dans le calcul au numérateur. le membre soustrait à droite au numérateur est défini comme suit : $\left(\frac{1}{n} \sum_{k=1}^n Y_k^B\right) \left(\frac{1}{n} \sum_{k=1}^n Y_k^A\right)$ alors que le code d'OpenTURNS calcule : $\left(\frac{1}{n} \sum_{k=1}^n Y_k^B Y_k^A\right)$. Ce calcul correspond à celui de l'estimateur de Mauntz Kucherenko.

Cette erreur a été découverte car elle entraine un écart entre la distribution asymptotique théorique et la distribution empirique comme montré dans la figure 33.

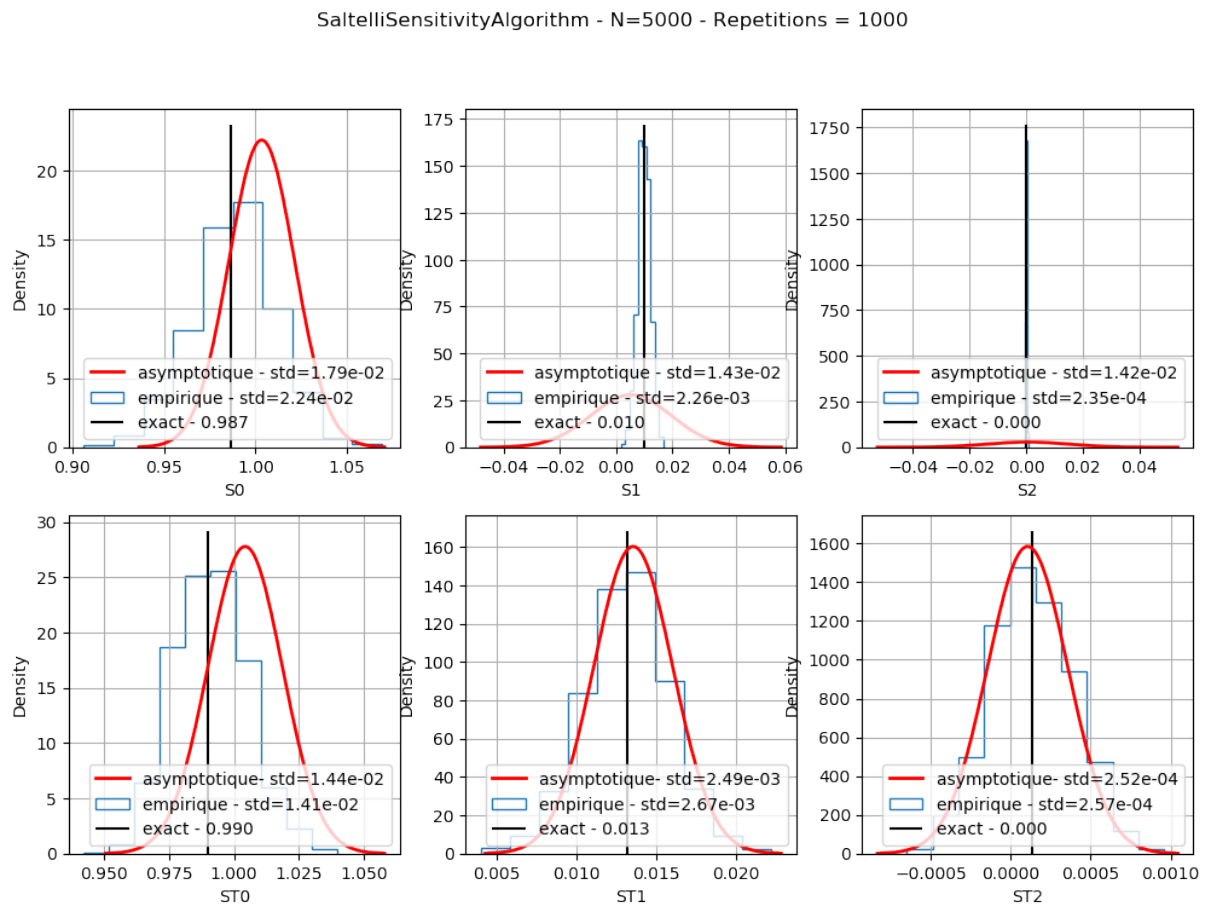


Figure 33 – Distributions des indices pour la fonction G-Sobol avec l'estimateur de Saltelli présentant une erreur de calcul.

Références

Alexandre JANON, Thierry KLEIN, Agnès LAGNOUX, Maëlle NODET et Clémentine PRIEUR : Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM : Probability and Statistics*, 18:342–364, 2014. ISSN 1292-8100, 1262-3318. URL <http://www.esaim-ps.org/10.1051/ps/2013040>.

Andrea SALTELLI : Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, mai 2002. ISSN 00104655. URL <http://doc.openturns.org/papers/saltelli2002.pdf>.