A minimum variance unbiased (generalized) estimator of total sensitivity indices: an illustration to a flood risk model

Matieyendou LAMBONI[1,2]

[1] University of Guyane (UG), Department of Science and Technology (DFRST), Cayenne, French Guiana

[2] UMR Espace-Dev, University of Guyane, University of Réunion, University of Montpellier, University of Nouvelle-Calédonie, IRD.

matieyendou.lamboni@gmail.com

SAMO 2016: Le Tampon, December 2016

Introduction

- Context: better understand complex models wrt their inputs

$$Y = \mathcal{M}(X_1, X_2, \ldots, X_d)$$

- characterize interactions among inputs

- identify inputs that have the great effects on the model

- Interest in Variance-based-Sensitivity Analysis [1,2,3] and Multivariate Sensitivity Analysis [4,5]

- objective way to evaluate the impact of input factors on the model output (s)

- provide total sensitivity indices (TSIs) or total-effect variances (TEVs)

Introduction

- Objective: provide new estimators of TEV and TSI and their statistical properties

- Outline
  1. Our new estimator of a TEV that generalizes Saltelli-Jansen's estimator
  2. Some properties of the generalized estimator of a TSI
  3. Illustration to a flood risk model
  4. Conclusion

1. New estimator of a total-effect variance (TEV)

Model

$$Y = \mathcal{M}(X_1,\, X_2,\, \ldots,\, X_d)$$

with $\mathbf{X} = (X_1, \ldots, X_d) = (\mathbf{X}_u, \mathbf{X}_{\sim u})$, d independent inputs (A1)

Sobol-Hoeffding's decomposition: if $\mathbb{E}\left(f^2(\mathbf{X})\right) < +\infty$ (A2),

$$\mathcal{M}(\mathbf{X}) \quad = \quad \mathcal{M}_0 + \sum_{j}^{d} \mathcal{M}_j(X_j) + \sum_{i<j}^{d} \mathcal{M}_{ij}(X_i, X_j) + \ldots \qquad (1.1)$$

Re-organized Sobol-Hoeffding's decomposition ([6,7])

$$\mathcal{M}(\mathbf{X}) \quad = \quad \mathcal{M}_0 + g_u(\mathbf{X}) + h_{\sim u}(\mathbf{X}_{\sim u})\,, \qquad (1.2)$$

with $g_u(\mathbf{X}) = \sum_{v, v\cap u \neq \emptyset} \mathcal{M}_v(\mathbf{X}_u)$ and $\mathbb{E}\left[g_u^2(\mathbf{X})\right] = D_u^{tot}$ (TEV).

It is known that the TEV of $\mathbf{X}_u$ is ([8,9]):

$$D_u^{tot} = \mathbb{E}\left(\left[\mathcal{M}(\mathbf{X}) - \mathbb{E}_{\mathbf{X}_u}\left[\mathcal{M}(\mathbf{X})\right]\right]^2 \right) . \qquad (1.3)$$

Estimator of the TEV of $\mathbf{X}_u$

Definition

Consider p independent variables $(\mathbf{X}_u^{(1)}, \mathbf{X}_{\sim u}), \ldots, (\mathbf{X}_u^{(p)}, \mathbf{X}_{\sim u})$ from the probability measure $\mu(\mathbf{X})$. We define the kernel

$$K\left(\mathbf{X}_u^{(1)}, \ldots, \mathbf{X}_u^{(p)}, \mathbf{X}_{\sim u}\right) =$$

$$\frac{1}{(p-1)p^2} \sum_{k=1}^{p} \left(\sum_{\substack{j=1 \\ j \neq k}}^{p} \left[\mathcal{M}(\mathbf{X}_u^{(k)}, \mathbf{X}_{\sim u}) - \mathcal{M}(\mathbf{X}_u^{(j)}, \mathbf{X}_{\sim u})\right] \right)^2 .$$

We use $\sigma_{p,1}^2 = \mathbb{V}\text{ar}\left[K\left(\mathbf{X}_u^{(1)}, \ldots, \mathbf{X}_u^{(p)}, \mathbf{X}_{\sim u}\right)\right]$ if $\mathbb{E}\left[\mathcal{M}(\mathbf{X})^4\right] < \infty$ (A3).

1. New estimator of a total-effect variance (TEV)

Theorem

Consider independent samples $(\mathbf{X}_{i,u}^{(1)}, \mathbf{X}_{i,\sim u}), \ldots, (\mathbf{X}_{i,u}^{(p)}, \mathbf{X}_{i,\sim u})$ $i = 1, \ldots, m$ from $\mu(\mathbf{X})$. If A1, A3 hold then the estimator

$$\widehat{D_u^{tot}} = \frac{1}{m} \sum_{i=1}^{m} K\left(\mathbf{X}_{i,u}^{(1)}, \ldots, \mathbf{X}_{i,u}^{(p)}, \mathbf{x}_{i,\sim u}\right)$$

i) is unbiased:

$$\mathbb{E}\left[\widehat{D_u^{tot}}\right] = D_u^{tot} \,, \tag{1.4}$$

ii) has the minimum variance and optimal rate of convergence:

$$m\mathbb{E}\left(\widehat{D_u^{tot}} - D_u^{tot}\right)^2 = \sigma_{p,1}^2 \,, \tag{1.5}$$

iii) follows asymptotically a normal distribution:

$$\sqrt{m}\left(\widehat{D_u^{tot}} - D_u^{tot}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_{p,1}^2\right) \,. \tag{1.6}$$

1. New estimator of a total-effect variance (TEV)

Proof

Ideas of the proof (comprehensive details are in [9]).

i) Show that $\mathbb{E}\left[K\left(\mathbf{X}_u^{(1)}, \ldots, \mathbf{X}_u^{(p)}, \mathbf{X}_{\sim u}\right)\right] = D_u^{tot}$.

ii) Use the theory of U-statistics [10] to obtain the optimal rate, the minimum variance and the asymptotic distribution.

2. Some properties of the estimator of a TSI

Theorem

Consider independent samples $(\mathbf{X}_{i,u}^{(1)}, \mathbf{X}_{i,\sim u}), \ldots, (\mathbf{X}_{i,u}^{(p)}, \mathbf{X}_{i,\sim u})$ $i = 1, \ldots, m$ from $\mu(\mathbf{X})$. If A1 and A3 hold then the generalized estimator of S_{T_u} is:

$$\widehat{S_{T_u}} = \frac{\widehat{D_u^{tot}}}{\hat{D}}\,, \tag{1.7}$$

with $\hat{D}$ the estimator of the model variance (D).

The estimator $\widehat{S_{T_u}}$ i) is consistent:

$$\widehat{S_{T_u}} \xrightarrow{\mathcal{P}} S_{T_u}\,, \tag{1.8}$$

ii) follows asymptotically a normal distribution

$$\sqrt{m}\left(\widehat{S_{T_u}} - S_{T_u}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma_{p,1}^2}{D^2}\right)\,. \tag{1.9}$$

Proof

Ideas of the proof (comprehensive details are in [9]).

For i): using Slutsky's theorem and the fact that $\widehat{D_u^{tot}} \xrightarrow{\mathcal{P}} D_u^{tot}$ and $\widehat{D} \xrightarrow{\mathcal{P}} D$, we have

$$\widehat{S_{T_u}} \xrightarrow{\mathcal{P}} S_{T_u} \,.$$

For ii): using Slutsky's theorem and the fact that $\sqrt{m}\left(\widehat{D_u^{tot}} - D_u^{tot}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_{p,1}^2\right)$ and $\sqrt{m}\left(\frac{1}{\hat{D}} - \frac{1}{D}\right) \xrightarrow{\mathcal{P}} 0$ the result follows.

3. Illustration to a flood risk model

- Flood risk model ([11])

$$S = Z_v + H - H_d - C_b \quad \text{with} \quad H = \left(\frac{Q}{BK_s \sqrt{\frac{Z_m - Z_v}{L}}} \right)^{0.6},$$

H: maximal annual height of the river (in meters).

S: maximal annual overflow (in meters).

Flooding occurs when $S > 0$.

The model includes 8 input factors.

Table: Input factors of the flood risk model and their probability distributions [11]

Input	Description	Unit	Probability distribution
Q	Maximal annual flowrate	m^3/s	T. Gumbel $\mathcal{G}(1013, 558)$ on $[500, 3000]$
K_s	Strickler coefficient	-	T. normal $\mathcal{N}(30, 8)$ on $[15, +\infty[$
Z_v	River downstream level	m	Triangular $\mathcal{T}(49, 50, 51)$
Z_m	River upstream level	m	Triangular $\mathcal{T}(54, 55, 56)$
H_d	Dyke height	m	Uniform $\mathcal{U}[7, 9]$
C_b	Bank level	m	Triangular $\mathcal{T}(55, 55.5, 56)$
L	Length of the river stretch	m	Triangular $\mathcal{T}(4990, 5000, 5010)$
B	River width	m	Triangular $\mathcal{T}(295, 300, 305)$

3. Illustration to a flood risk model

- Quasi MC sampling of p-fold input values of type $m \times d$

$$\mathcal{X}_1 = \begin{pmatrix} x_{11}^{(1)} & \dots & x_{1j}^{(1)} & \dots & x_{1d}^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1}^{(1)} & \dots & x_{mj}^{(1)} & \dots & x_{md}^{(1)} \end{pmatrix} \dots \mathcal{X}_p = \begin{pmatrix} x_{11}^{(p)} & \dots & x_{1j}^{(p)} & \dots & x_{1d}^{(p)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1}^{(p)} & \dots & x_{mj}^{(p)} & \dots & x_{md}^{(p)} \end{pmatrix}$$

For any input X_j, replace the j^{th} column of $\mathcal{X}_1$ with the j^{th} column of $\mathcal{X}_2, \dots, \mathcal{X}_p$.

$$\mathcal{X}_1 = \begin{pmatrix} x_{11}^{(1)} & \dots & x_{1j}^{(1)} & \dots & x_{1d}^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1}^{(1)} & \dots & x_{mj}^{(1)} & \dots & x_{md}^{(1)} \end{pmatrix} \dots \mathcal{X'}_p = \begin{pmatrix} x_{11}^{(1)} & \dots & x_{1j}^{(p)} & \dots & x_{1d}^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1}^{(1)} & \dots & x_{mj}^{(p)} & \dots & x_{md}^{(1)} \end{pmatrix}$$

- Computational cost for the estimation of d TSIs

$$cost = m((p-1)d + 1)$$

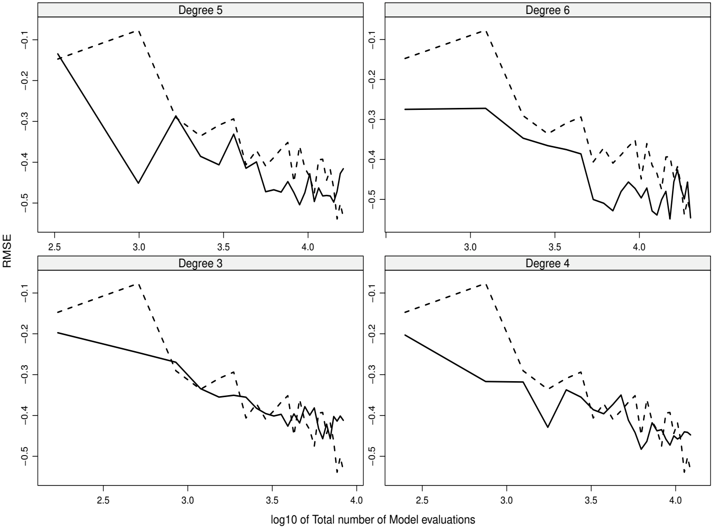
3. Illustration to a flood risk model

- Comparison of different estimators of TSI ($p = 2, 3, 4, 5, 6$)
  Saltelli-Jansen's estimator $p = 2$
  Generalized estimator $p = 3, 4, 5, 6$

- Measure of accuracy: average of Root Mean Square Error (RMSE)

$$RMSE_d = \frac{1}{d} \sum_{j=1}^{d} \sqrt{\frac{1}{R} \sum_{r=1}^{R=30} \left(\widehat{S_{T_j,r}} - S_{T_j} \right)^2},$$

3.Illustration to a flood risk model



4. Conclusion

- Generalized estimator of total-effed variance

  - using p-fold input values

  - has minimum variance and is unbiased

  - reaches the optimal rate of convergence

- Generalized estimator of total sensitivity index

  - is consistent (convergence to the true value)

  - follows asymptotically a normal distribution

  - total cost is $m((p-1)d+1)$

- Illustration to a flood risk model

  - improve the estimation of TSIs when $p = 6$

- Perspective

  - the value of p depends on the model

  - find a way to determine p for a given model

References

[1] Sobol, I. M. (1993) Sensitivity estimates for nonlinear mathematical models, MMCE 1, 407414.
[2] Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Salsana M, Tarantola S (2008) Global sensitivity analysis - The primer. Wiley.
[3] Higdon, D., Ghanem, R. and Owhadi H. (2016) Handbook of Uncertainty Quanti?cation. Springer.
[4] Lamboni M, Monod H, Makowski D (2011) Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. Reliability Engineering and System Safety 96: 450 - 459.
[5] Gamboa F, Janon A, Klein T, Lagnoux A (2014) Sensitivity indices for multivariate outputs. Comptes Rendus de lAcadémie des Sciences p In press.
[6] Sobol IM, Kucherenko S (2009) Derivative based global sensitivity measures and the link with global sensitivity indices. Math Comput Simul 79:30093017.
[7] Lamboni M, Iooss B, Popelin AL, Gamboa F (2013) Derivative-based global sensitivity measures: general links with sobol indices and numerical tests. Mathematics and Computers in Simulation 87:4554
[8] Lamboni M (2016) Global sensitivity analysis: an efficient numerical method for approximating the total sensitivity index. Int J Uncertain Quantif (accepted).
[9] Lamboni M (2016) Global Sensitivity Analysis: a generalized, unbiased and optimal estimation of total-effect variance. Stat. Papers. DOI 10.1007/s00362-016-0768-5
[10] Ferguson TS (1996) A Course in Large Sample Theory. Chapman-Hall, New York.
[11] Iooss B Revue sur l'analyse de sensibilité globale de modèles numériques, Journal de la Société Française de Statistique 152 (2011) 1 - 23.

Thank for your attention !